



Title: New techniques and framework for sentiment analysis and tuning of CRM structure in the context of Arabic language

Name: Hamed Saad AL-Rubaiee

This is a digitised version of a dissertation submitted to the University of Bedfordshire.

It is available to view only.

This item is subject to copyright.

NEW TECHNIQUES AND FRAMEWORK FOR SENTIMENT  
ANALYSIS AND TUNING OF CRM STRUCTURE IN THE  
CONTEXT OF ARABIC LANGUAGE

Hamed Saad AL-Rubaiee

Ph.D

2018

UNIVERSITY OF BEDFORDSHIRE

NEW TECHNIQUES AND FRAMEWORK FOR SENTIMENT ANALYSIS AND  
TUNING OF CRM STRUCTURE IN THE CONTEXT OF ARABIC LANGUAGE

by

Hamed Saad AL-Rubaiee

A thesis submitted to the University of Bedfordshire in partial fulfilment of the  
requirements for the degree of Doctor of Philosophy

January 2018

## Declaration

I, Hamed AL-Rubaiee declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

[New Techniques and Framework for Sentiment Analysis and Tuning of CRM Structure in the Context of Arabic Language]

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have cited the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission, or parts of this work have been published as indicated on [P.VI]: “

Hamed AL-Rubaiee

# NEW TECHNIQUES AND FRAMEWORK FOR SENTIMENT ANALYSIS AND TUNING OF CRM STRUCTURE IN THE CONTEXT OF ARABIC LANGUAGE

Hamed Saad AL-Rubaiee

## ABSTRACT

Knowing customers' opinions regarding services received has always been important for businesses. It has been acknowledged that both Customer Experience Management (CEM) and Customer Relationship Management (CRM) can help companies take informed decisions to improve their performance in the decision-making process. However, real-world applications are not so straightforward. A company may face hard decisions over the differences between the opinions predicted by CRM and actual opinions collected in CEM via social media platforms. Until recently, how to integrate the unstructured feedback from CEM directly into CRM, especially for the Arabic language, was still an open question. Furthermore, an accurate labelling of unstructured feedback is essential for the quality of CEM. Finally, CRM needs to be tuned and revised based on the feedback from social media to realise its full potential. However, the tuning mechanism for CEM of different levels has not yet been clarified.

Facing these challenges, in this thesis, key techniques and a framework are presented to integrate Arabic sentiment analysis into CRM. First, as text pre-processing and classification are considered crucial to sentiment classification, an investigation is carried out to find the optimal techniques for the pre-processing and classification of Arabic sentiment analysis. Recommendations for using sentiment analysis classification in MSA as well as Saudi dialects are proposed. Second, to deal with the complexities of the Arabic language and to help operators identify possible conflicts in their original labelling, this study proposes techniques to improve the labelling process of Arabic sentiment analysis with the introduction of neural classes and relabelling. Finally, a framework for adjusting CRM via CEM for both the structure of the CRM system (on the sentence level) and the inaccuracy of the criteria or weights employed in the CRM system (on the aspect level) are proposed. To ensure the robustness and the repeatability of the proposed techniques and framework, the results of the study are further validated with real-world applications from different domains.

## Acknowledgments

First, thanks to Allah Almighty Who helped me to finish this thesis.

I would like to express my gratitude to my supervisory team, Dr Renxi Qiu and Dr prof Dayou Li, for kindness, patience, continuous support and professional guidance that helped me to complete my PhD study.

I would like to thank the Saudi Arabian Cultural Bureau in London which represent the Saudi Arabia government for their support and funding me through this project.

I would like to thank the King Abdelaziz University and especially the deanship of e-learning and distance education for their support and help me to obtain data through their Twitter Account and other systems.

I would like to thank the Mubasher Company for their support and help for providing me the data.

I would like to thank the research graduate school (RGS) in the University of Bedfordshire for supporting me with information during this study, special thanks to Madam Caroline Loma for her moral support and understanding my needs as international students.

I would like to thank all the participants and the interviewees who gave up their time to be part of this study. Additionally, I would like to thank the library staff for their great help and support with learning resources.

Warm thanks go to my friends and colleagues Dr. Khalid Al-Omar, Dr. Walid AL-Ballah, Saleh Al-Zahrani, Mohammed AL-Otibi, Omer AL- Nomani, Samer Basseyy, Naif AL-Otibi, Khalid al-Mazrou, and Mohamed AL-Basri who supported, encouraged and motivated me during my research work and made this experience pleasant and memorable.

Finally, I would like to thank my colleague Iroshan Abeykoon for making our office a place where we can share our feelings, research ideas and have friendly environment.

## **Dedication**

I dedicate my PhD work to my family. A special feeling of gratitude to my loving parents, brothers and sisters. Very special thanks to my waif, who have never left my side and stand with me all the time.

## Publications to Date

The work in this thesis has been published in the following peer-reviewed journals and proceedings:

1. Hamed, A.R., R. Qiu, and D. Li, Analysis of the relationship between Saudi twitter posts and the Saudi stock market. In Intelligent Computing and Information Systems (ICICIS), IEEE Seventh International Conference on (pp. 660–665). 2015. IEEE.
2. Al-Rubaiee, H., R. Qiu, and D. Li, Identifying Mubasher software products through sentiment analysis of Arabic tweets. In Industrial Informatics and Computer Systems (CIICS), International Conference on (pp. 1–6). 2016. IEEE.
3. Hamed, A.R., R. Qiu, and D. Li, The importance of neutral class in sentiment analysis of Arabic tweets. International Journal of Computer Science & Information Technology (IJCSIT), 2016. 8 (2): pp. 17-31.
4. Al-Rubaiee, H., R. Qiu, and D. Li, Visualising Arabic sentiments and association rules in financial text. International Journal of Advanced Computer Science and Applications, 2017. 8 (2): pp.1–7.
5. Al-Rubaiee, H., R. Qiu, K. Alomar, and D. Li, Sentiment analysis of Arabic tweets in e-learning. Journal of Computer Science, 2016. 12 (11): pp. 553–563.
6. Al-Rubaiee, H., and K. Alomar, Clustering students' Arabic tweets using different schemes. International Journal of Advanced Computer Science and Applications, 2017. 8(4): pp. 276–280.



# Table of Contents

Declaration.....	I
Acknowledgments.....	III
Dedication.....	IV
Publications to Date .....	V
List of Tables .....	IX
List of Figures .....	XIV
Chapter 1: Introduction.....	2
1.1 Introduction.....	2
1.2 The scope of the research .....	4
1.3 Research questions & objective .....	4
1.4 Contributions.....	5
1.5 Research approach.....	6
1.6 The Limitations .....	8
1.7 Thesis structure .....	8
Chapter 2: Background and Related Work .....	10
2.1. Introduction.....	10
2.2. Customer Experience Management.....	10
2.3. Sentiment analysis .....	11
2.3.1. Sentiment analysis approaches.....	12
2.3.2. Applications of sentiment analysis.....	13
2.3.3. Applications domains .....	14
2.3.4. Natural language processing (NLP).....	15
2.3.5. Text mining methods.....	16
2.4. Challenges in applying sentiment analysis.....	16
2.4.1. Data mining.....	17
2.4.2. Data pre-processing.....	18
2.4.3. Feature selection advantages for text classification.....	19
2.4.4. Text classification (based on data mining).....	19
2.4.5. Methods and algorithms .....	20
2.4.6. Evaluation .....	21
2.5. Customer Relationship Management (CRM) .....	21

2.5.1. Customer Relationship Management (based on data mining) .....	22
2.5.2. Integrating sentiment analysis into CRM .....	23
2.6. Summary .....	24
Chapter 3: Techniques for Arabic Sentiment Analysis .....	25
3.1. Introduction.....	25
3.2. Challenges of using sentiment analysis in Arabic .....	25
3.3. Arabic language.....	26
3.3.1. Arabic language characteristics.....	27
3.3.2. Arabic sentence structure.....	28
3.3.3. Difference between Standard Arabic and Local Arabic.....	28
3.4. Sentiment analysis in Arabic.....	29
3.5. Arabic text pre-processing and sentiment analysis process .....	31
3.5.1. Preparing the Arabic text.....	32
3.5.2. The weighting schemes.....	33
3.5.3. Sentiment analysis via the machine learning approach .....	34
3.5.4. Techniques for pre-processing and classification of Arabic sentiment analysis and applications.....	34
3.5.5. Application in the financial domain: Saudi stock market prediction .....	35
3.5.6. Application in the educational domain .....	42
3.5.7. Application in the service domain .....	44
3.6. General recommendation .....	45
3.7. Summary .....	45
Chapter 4: Techniques for Improving the Labelling Process of Sentiment Analysis.....	46
4.1. Introduction.....	46
4.2. Improving the labelling process with a neutral class .....	47
4.3. Experiment: Classification into positive, negative, and neutral classes .....	47
4.4. Improve the labelling process with relabelling .....	51
4.4.1. Process of relabelling .....	51
4.4.2. Wordlist process.....	53
4.4.3. Filtering .....	53
4.4.4. Visualisation .....	54
4.4.5. Association rules .....	55
4.4.6. Extracting rules .....	57
4.4.7. Relabelling .....	58
4.4.8. Experiment: Relabelling .....	58
4.5. Summary .....	73

Chapter 5: Tuning of CRM via CEM using sentiment analysis on aspects level .....	74
5.1. Introduction.....	74
5.2. Framework on tuning CRM via CEM.....	75
5.2.1. Sentiment Analysis for CEM.....	76
5.2.2. Classification using CRM .....	76
5.2.3. Tuning on CRM via CEM.....	76
5.3. Case study for Tuning CRM via CEM at the aspect level: Optimisation of CRM of King Abdul-Aziz University.....	77
5.3.1. Background of King Abdul-Aziz University.....	77
5.3.2. Preparing the sentiment analysis for CEM using the criteria from CRM.....	78
5.3.3. CEM based on sentiment on the aspect level.....	78
5.3.4. Experiment on sentiment analysis for CEM .....	78
5.3.5. Classification using CRM .....	87
5.3.6. Experiment on the Classification using CRM .....	88
5.3.7. Tuning of CRM via CEM.....	93
5.3.8. Experiment in tuning on CRM via CEM .....	93
5.3.9 Tuning the CRM weights .....	104
5.4. Summary.....	108
Chapter 6: Tuning of CRM via CEM using sentiment analysis on sentence level .....	109
6.1. Introduction.....	109
6.2. Case study for tuning CRM based on CEM at the sentence level: Optimisation of CRM for Mubasher .....	109
6.2.1. Background of Mubasher’s products .....	109
6.2.2. Preparing the sentiment analysis for CEM using the criteria from CRM.....	110
6.2.3. CEM based on sentiment on the sentence level.....	110
6.2.4. Experiment of sentiment analysis for CEM .....	110
6.2.5. Classification using CRM .....	130
6.2.6. Tuning CRM via CEM .....	134
6.3. Summary.....	141
Chapter 7: Conclusion and Future Work .....	142
7.1 Conclusion .....	142
7.2 Research Contributions .....	143
7.3 Future Work.....	144

## List of Tables

Table 1: The difference between the CRM and CEM (taken from [5]).	2
Table 2: Keywords in Standard Saudi Dialect represent the classes	36
Table 3: Collection of tweets from Saudi stockmarket	37
Table 4: Replacing words in pre-process phase	37
Table 5: Precision and recall for positive and negative classes without N-gram feature using SVM and NB	40
Table 6: Precision and recall for positive and negative classes with N-gram feature using SVM and NB	40
Table 7: Collection of Tweets for the application in the educational domain.	43
Table 8: Precision and recall for positive and negative classes without N-gram feature using SVM and NB	43
Table 9: Precision and recall for positive and negative classes with N-gram feature using SVM and NB	43
Table 10: Number of tweets.	44
Table 11: Precision and recall for positive and negative classes without N-gram feature using SVM and NB	44
Table 12: Precision and recall for positive and negative classes with N-gram feature using SVM and NB	44
Table 13: Collection of tweets about the Saudi stock market with neutral class.	48
Table 14: Precision and recall for positive, negative and neutral classes without N-gram with SVM and NB	48
Table 15: Precision and recall for positive, negative and neutral classes using N-gram with SVM and NB	48
Table 16: Precision and recall for positive, negative, and neutral classes with SVM and NB	49
Table 17: Occurrence of the feature “Earnings” (أرباح)	58
Table 18: Phrases for the feature “earnings” (ارباح) in all the data	59
Table 19: The term “distribute profits” [توزيع-ارباح] in the neutral class documents	60
Table 20: The term “distribute profits” [توزيع-ارباح] in the negative class documents	62
Table 21: Feature “rise” (ارتفاع) occurrence	63
Table 22: Phrases for the feature “rise” (ارتفاع) in all the data	64
Table 23: The phrase “increase – percentage” [ارتفاع-نسبة] in the negative class documents	65
Table 24: Phrases for the feature “losses” (خسائر) in all the data	66
Table 25: Phrases for the feature “accumulated losses” (خسائر-المتراكمه) in all the data	66
Table 26: The phrase “Accumulated losses” [خسائر-المتراكمه] in the positive class documents	68

Table 27: Phrases for the feature “green” (الاخضر) in all the data .....	68
Table 28: Phrases for the feature “green colour” (باللون_الاخضر) in all the data. ....	69
Table 29: Phrase “green colour” [باللون_الاخضر] in the negative class documents .....	70
Table 30: All-class performance accuracy in original classification for SVM with TF-IDF schema .....	71
Table 31: All-class performance accuracy in new classification for SVM with TF-IDF schema. ....	71
Table 32: All-class performance accuracy comparison for SVM with TF-IDF schema without N-gram.....	71
Table 33: All-class performance accuracy in original classification NB with BTO schema with N-gram.....	72
Table 34: All-class performance accuracy in new classification NB with BTO schema with N-gram .....	72
Table 35: All-class performance accuracy comparison for NB with BTO schema with N-gram. ....	72
Table 36: A survey with four questions for e-learning criteria .....	81
Table 37: e-learning criteria.....	81
Table 38: Number of tweets utilized for Sentiment analysis for CEM.....	82
Table 39: Common Arabic words and standard Saudi dialect representing the classes .....	82
Table 40: Replacing words in the pre-processing stage .....	83
Table 41: Accuracy, precision, and recall for all classes using SVM and NB classifiers without N-gram.....	84
Table 42: Accuracy, precision, and recall for all classes using SVM and NB classifiers with N-gram .....	84
Table 43: Accuracy, precision, and recall for all classes using for SVM and NB classifiers without N-gram.....	84
Table 44: Accuracy, precision, and recall for all classes using for SVM and NB classifiers with N-gram.....	85
Table 45: Accuracy, precision and recall for all classes using for SVM and NB classifiers without N-gram.....	85
Table 46: Accuracy, precision and recall for all classes using for SVM and NB classifiers with N-gram.....	86
Table 47: Accuracy, precision, and recall for all classes using for SVM and NB classifiers without N-gram.....	86
Table 48: Accuracy, precision, and recall for all classes using for SVM and NB classifiers with N-gram.....	86
Table 49: Number of CRM utilized for Classification using CRM. ....	88

Table 50: Question 1 data description.....	88
Table 51: Question 1 Criteria.....	89
Table 52: Question 2 data description.....	89
Table 53: Question 2 Criteria.....	89
Table 54: Question 3 data description.....	90
Table 55: Question 3 Criteria.....	90
Table 56: Question 4 data description.....	90
Table 57: Question 4 Criteria.....	91
Table 58: Accuracy, precision, and recall with all classes for SVM and NB classifiers.....	91
Table 59: Accuracy, precision, and recall with all classes for SVM and NB classifiers.....	92
Table 60: Accuracy, precision, and recall with all classes for SVM and NB classifiers.....	92
Table 61: Accuracy, precision and recall with all classes for SVM and NB classifiers.....	93
Table 62: Number of integrated records utilized for tuning of CRM via CEM.....	94
Table 63: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question1.....	104
Table 64: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 2.....	105
Table 65: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 3.....	105
Table 66: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 4.....	106
Table 67: Common Arabic words in standard Saudi dialect representing the classes.....	112
Table 68: Number of tweets.....	113
Table 69: Replacing words in pre-processing.....	113
Table 70: Accuracy, precision, and recall for positive and negative classes for SVM and NB classifiers without N-gram.....	114
Table 71: Accuracy, precision, and recall for positive and negative classes for SVM and NB classifiers with N-gram.....	115
Table 72: Accuracy, precision, and recall for all classes using SVM and NB classifiers without N-gram.....	116
Table 73: Accuracy, precision, and recall for all classes using SVM and NB classifiers with N-gram.....	117
Table 74: Number of tweets.....	118
Table 75: NB & SVM for all classes.....	118
Table 76: NB & SVM with N-gram for all classes.....	118
Table 77: Number of tweets.....	119
Table 78: NB & SVM for all classes.....	119

Table 79: NB & SVM with N-gram for all classes.....	119
Table 80: Occurrence of the feature “the problem” (المشكلة) .....	120
Table 81: Phrases for the feature “the problem” (المشكلة) in all data .....	121
Table 82: Phrase “problem of Mubasher” [المشكلة-مباشر] in the neutral class documents .....	122
Table 83: Phrase “problem of Mubasher” [المشكلة-مباشر] in the positive class documents .....	123
Table 84: Occurrence of the feature “better” (افضل).....	124
Table 85: Phrases for the feature “better” (افضل) in all data.....	124
Table 86: Phrase “users prefer Mubasher’s product” [مباشر - افضل] in the neutral class documents .....	125
Table 87: Phrase “problem of Mubasher” [المشكلة-مباشر] in the negative class documents....	127
Table 88: All class performance accuracy in original classification for SVM with TF-IDF schema without N-gram .....	128
Table 89: All class performance accuracy in new classification for SVM with TF-IDF schema with N-gram .....	128
Table 90: All class performance accuracy comparison for SVM with TF-IDF schema .....	128
Table 91: All class performance accuracy in original classification NB with TF-IDF schema with N-gram. ....	129
Table 92: All class performance accuracy in new classification NB with TF-IDF schema with N-gram.....	129
Table 93: All class performance accuracy comparison for NB with BTO schema with N-gram. ....	129
Table 94: Number of records.....	132
Table 95: Accuracy, precision, and recall with all classes for SVM and NB classifiers.....	132
Table 96: Cluster number and main topic.....	137
Table 97: (1) Real data collected from Saudi Stock Market, and (2) Real data collected from Saudi Stock Market multiplied by constant. ....	156
Table 98: Occurrence of the term “growing” (زيادة). ....	157
Table 99: rules for the term “growing” (زيادة)in the all data.....	158
Table 100: Term “capital increase” [زيادة - رأس] in two documents. ....	159
Table 101: Term “capital increase” [زيادة - رأس] in the neutral class documents. ....	160
Table 102: Occurrence of the term “distribution” (توزيع).....	161
Table 103: Rules for the term “Distribution” (توزيع) in the all data. ....	162
Table 104: Term “Exchange of profits” [صرف-أرباح] in the neutral class documents.....	163
Table 105: Term “distribute profits” [توزيع-ارباح] in the negative class documents. ....	165
Table 106: Occurrence of the feature “reduction” (انخفاض).....	166
Table 107: rules for the term “reduction” (انخفاض) in the all data. ....	166
Table 108: Term “Low losses” [انخفاض - خسائر] in the negative class documents. ....	167

Table 109: Occurrence of the feature “Fine” (غرامة).....	167
Table 110: Rules for the term “Fine” (غرامة) in the all data. ....	167
Table 111: Occurrence of the feature “postponement” (تأجيل).....	167
Table 112: Rules for the term “postponement” (تأجيل) in the all data.....	167
Table 113: Occurrence of the feature “Works” (شغال).....	171
Table 114: phrase for the feature “Works” (شغال) in the all data.....	171
Table 115: phrase “user prefer mubasher product” [مباشر- شغال] in the neutral class documents. .....	172
Table 116: phrase “user prefer mubasher product” [مباشر- شغال] in the positive class documents. .....	174
Table 117: Occurrence of the feature “subscription” (اشتراك).....	174
Table 118: phrase for the feature “subscription” (اشتراك) in the all data.....	175
Table 119: phrase “subscription with mubasher” [مباشر- اشتراك] in the negative class documents .....	176
Table 120: phrase “subscription with mubasher” [مباشر- اشتراك] in the negative class documents. .....	178
Table 121: Occurrence of the feature “Support” (الدعم).....	179
Table 122: phrase for the feature “Support” (الدعم) in the all data. ....	180
Table 123: shows the phrase “Technical support” [الدعم - الفني] happened in neutral class documents.....	181
Table 124: shows the phrase “Technical support” [الدعم - الفني] happened in neutral class documents.....	183
Table 125: Occurrence of the feature “Disrupt” (عطلان).....	183
Table 126: phrase for the feature “Disrupt” (عطلان) in the all data. ....	184
Table 127: phrase “user prefer mubasher product” [عندي - عطلان] in the positive class documents .....	185



## List of Figures

Figure 1: The process of sentiment analysis in an Arabic tweet .....	34
Figure 2: Optimistic NB in red and SVM in blue.....	41
Figure 3: Pessimistic NB in red and SVM in blue .....	42
Figure 4: Neutral NB in red and SVM in blue.....	42
Figure 5: Lift chart for neutral class for SVM using TF-IDF schema without N-gram feature	50
Figure 6: The relation between labelling by human operators and the TASI for the Saudi stock market for positive, negative, and neutral classes. ....	50
Figure 7: Relabelling process.....	52
Figure 8: Wordlist process.....	53
Figure 9: The visualisation generated association rules process. ....	55
Figure 10: FP-tree construction (taken from [24]). ....	57
Figure 11: Visualisation of the association rules for the feature “earnings” (ارباح).....	59
Figure 12: The correlation rules of the feature “earnings” (ارباح) in the neutral class. ....	60
Figure 13: The correlation rules of the feature “earnings” (ارباح) in the negative class. ....	61
Figure 14: The correlations of the “negation term” (عدم) and the feature “earnings” (ارباح) in the negative class. ....	63
Figure 15: Visualisation of the association rules for the feature “rise” (ارتفاع).....	64
Figure 16: The correlation rules of the feature “rise” (ارتفاع) in the negative class. ....	65
Figure 17: Visualisation of the association rules for the feature “losses” (خسائر).....	66
Figure 18: The correlation rules of the feature “accumulated losses” (خسائر-المتراكمة) in the positive class. ....	67
Figure 19: Visualisation of the association rules for the feature “green” (الاخضر).....	68
Figure 20: The correlation rules of the feature “green colour” (باللون_الاخضر) in the positive class. ....	69
Figure 21: The process of the proposed framework of tune CRM via CEM .....	75
Figure 22: Announcement conducted to encourage students to participate in the Hashtag.	79
Figure 23: Comparing CEM labelling with CRM labelling for Question 1. ....	94
Figure 24: Comparing CEM labelling class with CRM labelling class for Question 1. ....	95
Figure 25: Comparing CEM all labelling classes with CRM positive labelling for Question 1. ....	95
Figure 26: Comparing CEM all labelling classes with CRM negative labelling for Question 1. ....	96
Figure 27: Comparing CEM all labelling classes with CRM neutral labelling for Question 1.	96
Figure 28: Comparing CEM labelling with CRM labelling for Question 2. ....	97
Figure 29: Comparing CEM labelling with CRM labelling for Question 2. ....	97

Figure 30: Comparing CEM all labelling classes with CRM positive labelling for Question 2. .....	98
Figure 31: Comparing CEM all labelling classes with CRM negative labelling for Question 2. .....	98
Figure 32: Comparing CEM all labelling classes with CRM neutral labelling for Question 2.	99
Figure 33: Comparing CEM labelling with CRM labelling for Question 3. ....	99
Figure 34: Comparing CEM labelling with CRM labelling for Question 3. ....	100
Figure 35: Comparing CEM all labelling classes with CRM positive labelling for Question 3. .....	100
Figure 36: Comparing CEM all labelling classes with CRM negative labelling for Question 3. .....	101
Figure 37: Comparing CEM all labelling classes with CRM neutral labelling for Question 3. .....	101
Figure 38: Comparing CEM labelling with CRM labelling for Question 4. ....	102
Figure 39: Comparing CEM labelling with CRM labelling for Question 4 ....	102
Figure 40: Comparing CEM all labelling classes with CRM positive labelling for Question 4. .....	103
Figure 41: Comparing CEM all labelling classes with CRM negative labelling for Question 4. .....	103
Figure 42: Comparing CEM all labelling classes with CRM neutral labelling for Question 4. .....	104
Figure 43: Comparing CEM labelling with CRM with the old and new criteria for Question 1. .....	106
Figure 44: Comparing CEM labelling with CRM with the old and new criteria for Question 2. .....	107
Figure 45: Comparing CEM labelling with CRM with the old and new criteria for Question 3. .....	107
Figure 46: Comparing CEM labelling with CRM with the old and new criteria for Question 4. .....	108
Figure 47: The process of sentiment analysis in Arabic tweets (Mubasher case study).....	111
Figure 48: Roc for BTO weighting scheme for both classifiers (NB in red, SVM in blue) ...	115
Figure 49: Roc for TF-IDF weighting scheme for both classifiers (NB in red, SVM in blue)	116
Figure 50: Lift chart for the neutral class for SVM using the TF-IDF schema.....	117
Figure 51: Visualisation of the association rules for the feature “the problem” (المشكلة) .....	120
Figure 52: The correlation rules of the feature “the problem” (المشكلة) in the neutral class...	121
Figure 53: The correlation rules of the feature “the problem” (المشكلة) in the negative class	122
Figure 54: Visualisation of the association rules for the feature “better” (افضل).....	124

Figure 55: The correlation rules of the feature “better” (افضل) in the neutral class .....	125
Figure 56: The correlation rules of the feature “better” (افضل) in negative class .....	126
Figure 57: The process of Mubasher’s CRM classification. ....	130
Figure 58: Comparing CEM labelling with CRM labelling .....	133
Figure 59: Comparing CEM labelling with CRM positive labelling .....	133
Figure 60: Comparing CEM labelling with CRM negative labelling .....	134
Figure 61: Comparing CEM labelling with CRM neutral labelling.....	134
Figure 62: The process of clustering analysis in Arabic tweets.....	136
Figure 63: Topics found by clustering.....	138
Figure 64: Comparing CEM labelling with CRM labelling .....	138
Figure 65: Labelling similarity between CRM and CEM.....	139
Figure 66: Comparing CRM labelling with CEM labelling for Cluster 0 .....	139
Figure 67: Comparing CRM labelling with CEM labelling for Cluster 1 .....	140
Figure 68: Comparing CRM labelling with CEM labelling for Cluster 2 .....	140
Figure 69: Comparing CRM labelling with CEM labelling for Cluster 3 .....	141
Figure 70: Visualize the association rules for the feature “growing” (زيادة) in all class.....	158
Figure 71: The correlation rules of the feature “growing” (زيادة) in neutral class.....	159
Figure 72: The correlation rules of the feature “growing” (زيادة) in negative class. ....	160
Figure 73: Visualize the association rules for the feature “Distribution” (توزيع) in all class. .	162
Figure 74: The correlations of the term “Distribution” (توزيع) in neutral class. ....	163
Figure 75: The correlations of the term “Distribution” (توزيع) in negative class.....	164
Figure 76: Visualize the association rules for the feature “reduction” (انخفاض) in all class..	166
Figure 77: show reports for collected tweets. ....	169
Figure 78: labelling students tweets. ....	170
Figure 79: Visualize the association rules for the feature “Works” “شغال” .....	171
Figure 80: The correlation rules of the feature “Works” ((شغال)) in neutral class.....	172
Figure 81: The correlation rules of the feature “Works” ((شغال)) in negative class. ....	173
Figure 82: Visualize the association rules for the feature “subscription” ((اشتراك)).....	175
Figure 83: The correlation rules of the feature “subscription” “اشتراك” in neutral class.....	176
Figure 84: The correlation rules of the feature “subscription” “اشتراك” in negative class. ....	177
Figure 85: Visualize the association rules for the feature “Support” “الدعم”.....	180
Figure 86: The correlation rules of the feature “Support” “الدعم” in neutral class.....	181
Figure 87: The correlation rules of the feature “Support” “الدعم” in positive class .....	182
Figure 88: Visualize the association rules for the feature “Disrupt” “عطلان”.....	184
Figure 89: The correlation rules of the feature “Disrupt” “عطلان” in positive class. ....	185

## List of Abbreviation

BTO	Binary Term Occurrence
CEM	Customer Experience Management
CMA	Capital Market Authority
Co-LDA	Semi-Supervised Topic Sentiment Mixture Model
CRM	Customer Relationship Management
DJIA	Dow Jones Industrial Average
FP-growth	Frequent-Pattern Tree Algorithm
FPR	False Positive Rate
GPOM	Google Profile of Mood
IDF	Inverse Document Frequency
KDD	Knowledge Discovery in Databases
KDDM	Knowledge Discovery and Data Mining
KNN	K-Nearest Neighbor
LMS	Learning Management Systems
MSA	Modern Standard Arabic
NB	Naive Bayes
NLP	Natural Language Processing
ODUS	Online Demand University System
OLAP	Online Analytical Processing
POS	parts of speech
ROC	Receiver operating characteristic
SIS	Student Information System
SOFNN	Self-Organised Fuzzy Neural Networks
SVM	Support Vector Machine
TASI	Tadawul All Share Index
TF	Term Frequency
TF-IDF	Term Frequency–Inverse Document Frequency
TPR	True Positive Rate
VSM	Vector Space Models

# Chapter 1: Introduction

## 1.1 Introduction

Both Customer Experience Management (CEM) and Customer Relationship Management (CRM) can help companies to take informed decisions that improve their performance in the decision-making process.

CRM is a set of people, processes, and technology that together represent a combined approach to managing relationships by converging on customer retention and relationship development [1]. CRM has different business elements under each customer dimension such as customer development, including customer lifetime value analysis, up-cross-selling, and market basket analysis [2]. It is mainly driven by what a company thinks about its customers. CEM, on the other hand, is focused on people's opinions with regard to service received. It is mainly about what customers think about the company. As social media platforms such as Twitter have given people a place to express themselves and share their thoughts and opinions on all kinds of topics [3], social media platforms have become a key source for CEM. The main differences between CRM and CEM are illustrated in Table 1. Winer [4] agreed that it is of high importance to scale the customer's responses against contents in CRM in order to improve immediate solutions to negative experiences.

Table 1: The difference between the CRM and CEM (taken from [5]).

	What	When	How monitored	Who uses the information
CEM	spread what a customer thinks about a firm	At points of customer interaction: "touch points"	Surveys, targeted studies, observational studies	Business, to make desires and better experiences with products and services
CRM	spread what a company knows about a customer	After a customer interaction is recorded	Point of sales data, market research, web site clicks	Sales, marketing, customer services, to drive more productive and successful execution

Nonetheless, there is a distinction between the assessments anticipated by CRM and the real sentiments gathered from online networking via CEM. On the one hand, CEM helps companies understand people's experiences through receiving largely unstructured feedback from social media platforms. On the other hand, CRM helps businesses extract hidden knowledge or information from an overview of users in general, which results in a structured understanding of people's behaviours.

Mostafa [6] illustrates that social media has recently become a rich resource for mining user sentiments for the purposes of CRM and CEM. Social opinion has been analysed using sentiment analysis. This is basically a natural language processing (NLP) application that uses computational linguistics and text mining to identify text sentiments as positive or negative, as well as neutral in some cases. This technique is known in the text mining field as “emotional polarity analysis”, “opinion mining”, and “review mining”. In addition, to calculate a sentiment score, the sentiment acquired from the text is compared to a dictionary to determine that sentiment’s strength. Studies on sentiment analysis focus on text written in English, using techniques as sentiment lexicons; applying such techniques to other languages will cause domain adaptation problems [7]. On the other hand, using the semantic processes model called “ontology” is another technique for text classification. Ontology is a formal and expository representation that includes the vocabulary referring to the terms in a particular topic domain and the logical statements that describe various expressions, such as how they are related to each other [8].

Sentiment analysis uses sentiment lexicons to focus on text written in English; applying this technique to other languages will be more challenging. Especially, in the case of Arabic text, there is no lexicon for Arabic sentiment words. Beside, the nature of Arabic text is different from that of English and other languages [9]. Arabic texts are written from right to left, and the forms of letters change depending on their location [10]. Moreover, Arabic words are precise with gender and number, and their affixes can change accordingly; capital letters are not used in Arabic text, and special grammatical rules are required to detect nouns, acronyms, and abbreviations [11, 12]. Such issues cause challenges in attempting to analyse Arabic text. Limited research has been conducted on Arabic sentiment analysis, and therefore this field is still in its early stages [10]. Besides, an accurate labelling of unstructured feedback is fundamental for the nature of CEMs. However, the task is unwieldy, with uncertain contributions from human operators, since input may contain both positive and negative keywords, which befuddles humans and makes the labelling process problematic.

Data mining technologies have created a new area in CRM, since they can extract hidden knowledge or information from large organisational databases, which results in understanding customers’ behaviours by distinguishing valuable customers and enables firms to take proactive actions to provide good services [13]. Bodine, Rogowski, and Stone [14] and Trainor [15] suggest that improving the overall interaction with the customer would lead to better service and customer satisfaction. To support that interaction, social media can be used as a tool to develop and maintain CRM. Investing in the resources that integrate social data into their existing customer databases will result in firms’ gaining meaningful data in order to improve their performance.

Facing the aforementioned difficulties, in this study, key techniques and a framework are introduced to integrate Arabic sentiment analysis into CRM. To start with, as text pre-processing and classification are crucial to sentiment classification, an investigation will be carried out to discover the optimal techniques for the pre-processing and classification of Arabic sentiment analysis. Recommendations for using sentiment analysis classification in Modern Standard Arabic (MSA) as well as Saudi dialects will be proposed. Second, to deal with the complexities of the Arabic language and to help operators identify possible conflicts in their original labelling, this study proposes techniques to improve the labelling process of Arabic sentiment analysis with the introduction of a neutral class and relabelling. The proposed techniques will improve human operators' flexibility to express their opinions during the labelling process with ambiguous feedback. Finally, a framework for adjusting CRM via CEM for both the structure of the CRM system (on the sentence level) and the inaccuracy of the criteria or weights utilised in the CRM system (on the aspect level) is proposed. To guarantee the robustness and repeatability of the proposed techniques and framework, the consequences of the investigation are validated and approved with real-word applications from various domains.

## **1.2 The scope of the research**

The focus of this thesis is to integrate Arabic sentiment analysis into CRM, and to help operators identify possible conflicts in their original labelling in order to deal with the complexities of the Arabic language. In addition, the study is focus on to adjust CRM via CEM for both the structure of the CRM system and the inaccuracy of the criteria or weights utilised in the CRM system.

## **1.3 Research questions & objective**

There are two main research questions to be investigated in this thesis:

First, how can hidden information (opinions) significant to CEM be extracted from Arabic social media data?

This question can be divided into two sub-questions:

- 1.1. What are the optimal techniques for pre-processing and classification of Arabic sentiment analysis?
- 1.2. How should ambiguity be tackled during the labelling process in sentiment analysis?

Second, how should CRM be adjusted via CEM to reduce the differences between the opinions predicted by CRM and the actual opinions collected in social media?

By addressing these two questions, this research will explore techniques that mine opinions from Arabic social media data and integrate the result of the opinion mining with existing analytical CRM databases. The main aim of the research is to integrate the techniques into a framework to improve the interaction between CRM databases and social media–based CEM. The framework will be applied to discover patterns that could help in the development of CRM analytical elements such as structure revision and weight optimisation. The objectives of this research are as follow:

- Explores optimal techniques for pre-processing and classification of Arabic sentiment analysis and provides recommendations for using sentiment analysis classification in MSA as well as Saudi dialects
- Proposes a technique for improving the labelling process in Arabic sentiment analysis with a neutral class by enabling a better coverage of the vector space constructed by the training dataset
- Proposes a technique for improving the labelling process of Arabic sentiment analysis with relabelling to help operators identify possible conflicts in their original labelling.
- Adjust CRM via CEM to reduce the differences between the opinions predicted by CRM and actual opinions collected from social media for both the structure of the CRM system and the inaccuracy of the criteria or weights employed in the CRM system.

#### **1.4 Contributions**

The innovation of the research will be implemented through

- New knowledge on identifying optimal techniques for pre-processing and classification of Arabic sentiment analysis.
- New techniques for handling ambiguous information during the labelling process.

A framework that integrates existing customers' data, social data, and applied data mining functions such as classification, association, and clustering into the CRM optimisation process. As a result, new structures and improved semantic schemas (such as criteria and weights of rules) will be identified to improve CRM's performance.

These can be summarised as follows:

In Chapter 3,

1. The thesis explores optimal techniques for pre-processing and classification of Arabic sentiment analysis. The proposed approach will improve the extraction of hidden information. Consequently, pre-processing will occur in several steps, such as tokenisation, stemming, stop words removal, and filtering tokens by length to remove



useless words. In addition, an N-gram feature will also be utilised, since higher-order N-grams are more accurate in detecting context, as they give a better understanding of a word's location. The strategy is to use tokens such as bigrams (N = 2) in the feature space rather than just unigrams (N = 1) [16], since during the pre-processing stage in this experiments study which is for Arabic text it gives a better result in the final accuracy comparing to other N-gram features such as trigrams (N=3).

2. The thesis evaluates the effectiveness of text pre-processing with different combinations of term weighting schemes on Arabic text and investigates the effectiveness of using popular text classification algorithms in Arabic sentiment analysis.
3. The thesis provides recommendations for using sentiment analysis classification in MSA as well as Saudi dialects.

In Chapter 4,

4. The thesis proposes a technique for improving the labelling process in Arabic sentiment analysis with a neutral class by enabling a better coverage of the vector space constructed by the training dataset.
5. The thesis proposes a technique for improving the labelling process of Arabic sentiment analysis with relabelling to help operators identify possible conflicts in their original labelling.

In Chapters 5 and 6,

6. The thesis proposes a framework to adjust CRM via CEM to reduce the differences between the opinions predicted by CRM and actual opinions collected from social media for both the structure of the CRM system (sentence level) and the inaccuracy of the criteria or weights employed in the CRM system (aspect level).

## **1.5 Research approach**

To answer the research questions and fulfil the objective, the research will be carried out by the following main steps:

- Step one: Provide fundamental background knowledge by reviewing the previous works and studies about social data mining (unstructured data). This step emphasises extracting opinions from text from Arabic social media platforms. Therefore, the sentiment analysis field will be investigated and reviewed in detail to
  - Collect unstructured data from social media using Twitter's streaming API.

- Pre-process using NLP to tokenise, remove stop words, lightly stem, filter tokens by length, and implement N-grams.
- Optimise the vectors for semantic schemas using machine learning based on classification accuracy, precision, and recall measures.

The techniques will be deployed in selected applications from different domains, and a recommendation for using sentiment analysis classification in MSA as well as Saudi dialects will be proposed (see contributions 1, 2, and 3) based on the common patterns of the selected Arabic Sentiment Analysis applications.

- Step Two: Identify the sources of ambiguity in the labelling process and then tackle them by corresponding methods, such as identifying possible contradictions and adding a neutral class. Techniques for improving the labelling process of Arabic sentiment analysis by giving humans more flexibility to express their uncertainty will be established (see contributions 4 and 5).
- Step Three: Carry out a detailed investigation about data mining in CRM. The current study considers the following methods to validate and test CRM structured data to achieve the fourth objective for this study:
  - Collecting structured data from a CRM system using CRM Databases API.
  - Labelling relevant records into three classes, namely, positive, negative, and neutral, based on their behaviour patterns and CRM criteria.
  - Building a collection of attributes for each of the classes by pre-processing the attributes of CRM and their combination (e.g., user's personal information, user's transactions, product used).
  - Initialising dictionaries for each of the classes and then optimising them by machine learning approaches based on classification accuracy, precision, and recall measures.
  - Establishing a second semantic schema.
- Step Four: Integrate the CRM and CEM opinions predicted by the above steps to validate the consistency of the opinions from both sources. The following methods will be used to validate and test the consistency of the selected CRM system and Arabic tweets to achieve the fifth and sixth objectives for this study.
  - Tuning on aspect level: Sentiment analysis can be applied to adjust the criteria or the weights utilised in CRM.
  - Tuning on sentence level: Sentiment analysis can be applied to adjust the structure of CRM.

In this stage, it was realised that CRM can be adjusted by sentiment analysis (see contribution 6).

## **1.6 The Limitations**

Although the results can be considered promising and positive, the research has some limitations that should be pointed out.

- The first limitation of this study is the sample size is small. To identify significant relationships larger sample size could have generated more accurate results.
- The second limitation of this study is the lack of available lexicon or dictionary for Arabic Language. This causes Labelling documents conducted manually.
- The third limitation of this study is the Arabic informal language's lacking structure and being difficult to standardize, and NLP is not applicable to Arabic because of its morphological richness.

## **1.7 Thesis structure**

The structure of the rest of this thesis is as follows:

Chapter 2 provides background knowledge for the research. It starts with some general information about text classification and sentiment analysis and then discusses the NLP in general, followed by data mining methods and approaches used to build data mining models and their evaluation measures. Moreover, it offers the necessary background on improving sentiment analysis in the Arabic language and the challenges of using Arabic in sentiment analysis. It discusses Arabic's script, syntax, and morphology. The chapter then describes the techniques that have been used to pre-process Arabic text using NLP and the Arabic sentiment analysis approaches that are currently being built in this research.

Chapter 3 explores techniques for pre-processing and classification of Arabic sentiment analysis. The characteristics of the Arabic language are presented along with the general state of the art for sentiment analysis in Arabic. The chapter then proposes an Arabic text pre-processing and sentiment analysis process. Thereafter, optimal techniques for pre-processing and classification of Arabic sentiment analysis in three different domains are explored. Finally, repeated patterns in Arabic sentiment analysis are summarised based on applications from different domains.

Chapter 4 presents a technique for improving the labelling process in sentiment analysis. The chapter shows why the labelling process is improved with a neutral class, along with the need to improve the labelling process with relabelling. The chapter also illustrates the process of

relabelling by analysing the experimental findings for Saudi stock market data. The final section will constitute the conclusion and recommendations for further work in this area.

Chapter 5 introduces a framework for integrating CRM and CEM. Then the chapter presents a case study for King Abdulaziz University for tuning CRM based on CEM at the aspect level. Then, sentiment analysis for CEM, with experimental results and evaluations, is presented. The chapter also illustrates classification using CRM with results and evaluations. Thereafter, the chapter covers the tuning of CRM via CEM along with experimental results and evaluations. Finally, a summary of this chapter is presented.

Chapter 6 presents the background of the Mubasher product. Then, a framework for integrating CRM and CEM is introduced. Next, the chapter presents a case study with Mubasher for tuning CRM based on CEM at the sentence level. Thereafter, sentiment analysis for CEM, with experimental results and evaluations, is presented. The chapter also illustrates the classification using CRM with results and evaluations. After that, the chapter covers the tuning of CRM via CEM, along with experimental results and evaluations. Finally, a summary of this chapter is presented.

## **Chapter 2: Background and Related Work**

### **2.1. Introduction**

Building relationships with their customers is an important function for modern businesses. Organisations use different processes and modern technology to manage relationships by focusing on retaining their customer bases and developing positive relationships with their customers. CEM helps businesses understand the customer experience through honest appraisals and feedback from customers; knowing what individuals think about the service provided has always been important for making business decisions. On the other hand, CRM utilises data mining technologies to help extract hidden knowledge from an overview of customers in general, which results in understanding customer behaviours and enables firms to take proactive actions to provide good services. Hence, CRM and CEM can both help companies to take immediate and informed decisions to improve their performance.

In today's world, CRM needs support from social media to realise its full potential. Social opinions have been analysed using sentiment analysis, which helps organisations, educational institutions, and businesses to analyse and monitor public opinion towards the services provided. This is important, as an explosion of social networks and microblogging has increased interest in analysing public opinion.

To develop a comprehensive background for this research, one should have a thorough knowledge of three essential areas: (a) sentiment analysis and its related topics, such as NLP and data pre-processing; (b) data mining tasks, including text classification, methods, and algorithms, along with their evaluation; and (c) CRM's definition and CRM based on data mining, integrating sentiment analysis into CRM.

### **2.2. Customer Experience Management**

CEM defined by (Zomerdijk and Voss, 2009) is a holistic concept that encompasses every aspect of a company's offering. In addition, Meyer and Schwager (2007) defined the CEM is the internal and individual reactions that customers have to any direct action such as purchase, use and serves or indirect touch with a company which is unplanned encounter with representation of company product, service or brand which came from word -of- mouth from different resource such as recommendation, criticism, news and blogs. Customer experience has become increasingly important for service organizations. collecting explicit customer feedback are generating increasing volumes of unstructured textual data, making it difficult for managers to analyze and interpret this information. Consequently, text mining, a method

enabling automatic extraction of information from unstructured text data, is gaining in popularity [17]. For example, using text mining for Improving student experience management in higher education [18]. Now days, sentiment analysis and opinions expressed in social networks play significant role in study customer behaviors across different dominos as buying products, providing serves and stock market [6].

### **2.3. Sentiment analysis**

Sentiment analysis is at the heart of social media research, and there have been a number of studies [19-21] in which users' opinions have been elicited in areas such as commerce, management, and politics. Social media is indeed a rich resource for mining users' feelings; social opinion is analysed using sentiment analysis, and studies show that sentiment analysis of various inputs on social media can support business planning [22-24]. For instance, Twitter is an online platform that was created in 2006 and allows its users to send and read text-based posts, known as tweets, with a 140-character limit. The number of Arabic users on platforms such as Twitter has been rapidly increasing, reflecting the rise in young populations in Arab countries. With the rapid growth of social media utilisation in the Arabic-speaking world, it is evident that there is also growth in Arabic reviews, comments, ratings, feedback, and opinions. However, Duncan and Zhang [25, 26] also reported that it is not possible to compare traditional sentiment analysis with Twitter sentiment analysis; due to the limit on the number of characters that can be used in a tweet, there are large numbers of abbreviated words, misspellings, and colloquialisms being used. This is why the application of neural networks has tended to have a low accuracy rate. There have been some attempts to deal with this problem, and there have been some research studies, but those related to Arabic texts have been very limited.

Using social media to provide business information for companies' decision-making is a relatively new concept. Sentiment analysis is an application that combines NLP, computational linguistics, and text analytics; this enables sentiment polarity to be identified and then retrieved from the text by studying the opinions of social media users [27]. The given text can then be classified by determining whether the expressed opinion is positive, negative, or neutral [20]. Most of the open literature on sentiment analysis has focused mainly on the following levels:

- Document level:

The main task of document-level sentiment analysis is to classify the entire document and identify whether it has a positive or negative sentiment [28, 29]. This assumes that the document presents a single sentiment [30, 31]. Consequently, this has been criticised as impractical because particular documents could hold more than one sentiment [32].

- Sentence level:

The sentence level is a closer level of sentiment analysis [33]; on this level, the sentence expresses a single opinion, whether positive, negative, or neutral [28]. Moreover, Liu [34] and Ding, Liu, and Yu [35, 36] stated that sentence-level sentiment analysis is closely related to subjectivity analysis. Solomon [37] explained subjectivity in philosophical terms, referring to the subject's perspective, feelings, beliefs, and desires. Turney [38] as well as Pang and Lee [38, 39] determined that subjectivity examination consists of two tasks: extracting the subjective signs, for example terms and expressions, and then utilising them to classify the considered sentence as subjective or objective. This assignment is the first process of most sentiment analysis applications and is called "subjectivity classification".

- Entity and aspect level:

Entity- and aspect-level sentiment analysis is based on targeted positive or negative opinions describing entities and aspects of them; for example, '*The Samsung call quality is good*' evaluates one aspect of the call quality of the entity Samsung [28]. Its objective is to identify the target of each opinion, and it suggests that every opinion has a defined target [40]. In addition, some researchers [35, 41] have added two more levels in sentiment analysis:

- Comparative sentiment analysis, where the aim is to identify sentences that contain comparative opinions and to extract the preferred entities from each opinion [42].
- Sentiment lexicon acquisition: A manual approach is generally not feasible, as each domain requires its own lexicon, and it would take an enormous amount of time. However, a dictionary-based approach is more effective, as it can start with a small number of sentiment words suitable for the target domain; this set of words can then be expanded by using synonyms and antonyms [42].

### **2.3.1. Sentiment analysis approaches**

Sentiment classification identifies the extent of polarity in texts by reviewing the sentiment expressed. This section investigates the different Machine Learning techniques that are used for identifying sentiment.

Generally speaking, there exist three different approaches for detecting sentiment in text [43]:

- The machine learning approach or corpus-based approach uses vectors to represent positive or negative data; the classifier is trained to deal with these vectors and thereby produce features that will identify them as relating to certain classes [29]. However, it

has been found that sentiment classifiers trained to deal with one domain often fail when applied to other domains [44, 45]. For example, Aue and Gamon [44] reported that an in-domain support vector machine classifier trained to review data related to films (with an accuracy of 90.45%) had a much lower accuracy rate of 70.29% when used for book reviews. This is a clear indication that sentiments are dependent on context and can vary according to the topic.

- The lexicon-based approach depends on constructing linguistic resources such as dictionaries and lexicons and keeps track of words. Once the lexicon or dictionary is built, a simple count of the different lexical items will reveal the polarity of the text [29, 46, 47]. The major challenges for building lexicons are that these are very hard to build manually and they are domain dependent, but they do not require an annotated dataset to detect sentiment [48].
- The linguistic approach or so-called hybrid approach is a combination of both previously mentioned approaches. It is also referred to as a semi-supervised or weakly supervised approach, and it uses a large amount of unlabelled data and partially annotated data to build better classifiers. It normally uses the syntax of words and phrases, as well as the structure of the text, to make a judgement of the text's orientation [43, 49]. One method used in the linguistic approach is based on part-of-speech (POS) [33], which are the grammatical category labels assigned to each token based on the linguistic and contextual information within a sentence [50].

These three approaches can be used independently or combined. The model classifies unlabelled data using a supervised classifier trained with labelled data; an example of this is when there is a combination of machine learning and linguistic approaches and the selected features for training are focused on just one POS [33]. This means that a lexicon can be constructed from adjectives [33], which can then be classified as either positive or negative [29]. Nevertheless, Dasgupta and Ng [51] have reported that there are difficulties in analysing both supervised and semi-supervised approaches, and it is likely that unsupervised or weakly-supervised approaches create even more challenges.

### **2.3.2. Applications of sentiment analysis**

Sentiment analysis remains an open research field because expanding application domains, linguistic nuances, differing contexts, and even cultural factors make it challenging to automatically assess a piece of text's sentiment. The need to know other people's sentiments about objects has always been part of humans' information needs. Traditionally, when an individual needs sentiment information, he or she asks friends and family, whereas



organisations, companies, and governments conduct surveys and opinion polls [52]. Sentiment analysis in reviews entails examining product reviews to determine the overall opinion of a specific product [33]. Reviews reflect content generated by the actual users of a product, and therefore it is important for professionals to receive this feedback [53]. Sentiment analysis applications are endless. Work with sentiment analysis in different domains such as products/services, stock market prices and their fluctuations, and even students' opinions can prove a breakthrough for complete business success. Three such examples are the focus for this research: products reviews, stock market, and student feedback.

### **2.3.3. Applications domains**

In the first instance, a sentiment analysis system can be developed to determine consumer attitudes on products/services from review data. For example, when Dehkharghani and Yilmaz [54] conducted a review using a logistical classifier, they found that the average accuracy was 66.6% [9]. Bross and Ehrig [55] found that an aspect-based review to detect individual opinions and expressions about specific aspects of a product had a high accuracy [9]. Indhuja and Reghu [56] used an approach with novel fuzzy functions and achieved 85.58% accuracy. Wang [57] found that his model using a combined sentiment Co-LDA and topic Co-LDA was more effective than just topic sentiment analysis. Zhang and Varadarajan's [58] models, which incorporated features to predict utility scores of product reviews, achieved high performance, indicating that this is an effective approach [9].

The second example looks at the influence of Twitter; in terms of any relationship between sentiment analysis and the stock market, Bollen, Mao, and Zeng [22] reported that Twitter moods were used to predict the Dow Jones stock market index [10]. In their approach, public moods were measured using two tools: Google Profile of Mood (GPOM) and Opinion Finder [10] and a predictive model based on Self-Organised Fuzzy Neural Networks (SOFNN). The study found that the accuracy of the standard stock prediction model was significantly improved when mood dimensions and the value of the Dow Jones Industrial Average (DJIA) were included [10]. In addition, a study by Martin, Bruno, and Murisasco [23] showed a correlation between public opinion expressed on Twitter and the French stock market, using a neural network to find association patterns. They found that adding the sentiment feature on tweets two days before stock market closing values could improve accuracy [10].

Simsek and Ozdemir [59] also found a relationship between Turkish Twitter posts and the stock market index. By using the most common Turkish words representing happiness and unhappiness collected from tweets over a period of a month and a half, they worked out the frequencies of these two classes of words [10]. They found that the terms happy and trouble were commonly used in the emotional word database, and when the Twitter post contained

stock market–related words, the average emotional value of the tweets changed from happiness to unhappiness by approximately 45% [10]. In a further study, Khatri, Singhal, and Johri [60] tried to train the neural network with words such as happy, hope, sad, and disappointing as input to predict the Bombay Exchange index. Their study indicated that an artificial neural network provided optimum results when set up with one hidden layer containing nine neurons [10]. Gao et al. [61] tried using a sentiment classification approach for Chinese stock news and found that pre-processing and a relevant sentiment dictionary affected the classification. Zhang and Skiena [24] showed that news sentiment can have an influence on market trading algorithms; they also found that news sentiment had a much sooner impact on stock markets than sentiment expressed on Twitter, which could take up to three days [10]. All of these studies, however, point to a strong correlation between sentiments expressed on Twitter and global stock market indices.

In the third example, student feedback can highlight any issues students may have with the services provided by their colleges or universities. An example of this is when students cannot understand a lecturer or do not avail themselves of specific online services. Students have a habit of regularly using social media to express their opinions and describe activities they are involved in [62]. Therefore, universities utilise social media as a way of improving their teaching processes and generally to find out more about student experiences [62]. This is especially useful when finding out about online distance-education students, who do not give feedback face to face [62]. For instance, Tian et al. [63] developed an e-learner questionnaire and compared emotion words to measure the intensity of sentiment in each category. This approach enabled a positive result in dealing with challenges faced in the analysis of texts, such as those in Chinese, which are characterised by the richness of emotions. Wang, Zuo, and Diao [64] worked with the essential function of sentiment feedback in education over the Internet [62]. Wang [65] set up a Student Feedback Mining System (SFMS) to carry out an in-depth analysis of qualitative student feedback, which allowed insight into teaching practices, thus significantly improving student learning [62]. Donovan, Mader, and Shinsky [66] found that online student comments were much more detailed and informative than traditional paper-based feedback but are more time-consuming to analyse [62]. However, sentiment classification is important, as it gathers attitudes and opinions of users by mining and analysing personal information [65].

#### **2.3.4. Natural language processing (NLP)**

NLP encompasses a range of techniques that enable computers to analyse and understand human languages [67, 68]. Therefore, it is clearly relevant in sentiment analysis, since sentiment is typically expressed in the form of unstructured textual data. Sentiment analysis

is closely related to some of the techniques developed in NLP, such as tokenization, stemming, removing stop words, negation handling, and feature selection. In addition, the process of marking up words corresponding to a particular POS defines the syntactic patterns or categories of the words [69]. N-grams are utilised to describe these patterns, and higher-order N-grams are more accurate in detecting context, since they give a good understanding of a word's position [62]. The approach is to utilise tokens like trigrams in the feature space rather than unigrams [16, 29].

### **2.3.5. Text mining methods**

Text mining identifies patterns in data [70]. There are a number of different text mining methods, such as text classification and information retrieval, but the addition of machine learning has improved results. When applied to sentiment classification, text mining's main aim is to determine attitudes and opinions through mining and analysing the level of interest shown by users, as well as other subjective information [65].

### **2.4. Challenges in applying sentiment analysis**

Sentiment analysis can cause a classification problem. It involves to a certain extent NLP and accordingly reflects many of the challenges inherent in such an approach [9]. The challenge of this field is to improve machines' ability to understand texts in the same way as human readers can, as it is imperative for companies in a competitive world to note what their users are saying about their products or services [9].

Working with sentiment analysis has some challenges because text pre-processing has a significant role in sentiment classification performance [33]. Data preparation can be more time-consuming and challenging than data mining [71, 72]. Moreover, it is highly domain and context dependent. A piece of text that is positive in one domain can be negative in another. For example, "go and read the book" is positive in the domain of books but negative in the domain of movies [73].

Additionally, text properties present peculiar challenges for sentiment analysis. For instance, social media text (e.g., Twitter) users often employ non-standard spelling/grammar and sarcasm to express their opinions. Bharti et al. [74] defined sarcasm as a type of sentiment in which people express their negative feelings using positive or extremely positive words in the text. It can be observed that sarcasm is essentially a device for expressing sentiment, and it is difficult to handle, as the literal meaning affected by the use of sarcasm must be detected and treated as the opposite of what is said. Existing research work on sarcasm detection typically focuses on detecting the presence or absence of sarcasm but not how to handle it for effective sentiment analysis [75].

The research study by Da Silva et al. [76] showed that Twitter sentiment analysis needs to be treated differently, as the frequency of misspelt or abbreviated words and slang in tweets is amplified due to the 140-character limit [9]. Similarly, Maynard and Greenwood [75] found that text on these platforms is typically short, thus presenting high levels of ambiguity. Gokulakrishnan et al. [77] discussed the challenges of and presented strategies for using steps to help with the classification process. These steps are as follows: replacing emoticons, upper case, and lower case; URL extraction; detection of pointers (usernames and hashtags); identification of punctuation; removal of stop words; and compression of words.

#### **2.4.1. Data mining**

Data mining is an important field and a core process within Knowledge Discovery in Databases (KDD). KDD is a process used to identify any patterns or models that exist within data and help identify their potential use. While data mining is only one step in the KDD process, data mining includes the use of specific algorithms that make use of suitable computational efficiency restrictions in order to find patterns or models in databases [78, 79].

Furthermore, the process of data mining incorporates the automated extraction of interesting information from large databases (relational databases, data warehouses, etc.). The overall goal of data mining is to search for relationships and global patterns present within databases that may be hidden amongst the high volume of data. Many data mining techniques are implemented directly in the data sources that produce the results, and their various patterns can then be evaluated using specific rules.

The two classes or methods of data mining objectives can be identified as descriptive and prescriptive. The descriptive task is a method of summarising the properties of data. In other words, this allows it to be understood and interpreted by humans. That is, descriptive tasks identify a set of categories that describe the data; an example is the clustering method. The prescriptive task is a method of inference performed on data to produce predictions based on historical data and is designed to separate items into discrete classes [80-82].

Bharati and Ramageri [83, 84] and Liao et al. [84] reported that there are different types of data mining techniques, such as

1. Classification.
2. Clustering.
3. Association rules.

Classification is an important technique that has been studied widely within the data mining and machine learning literature. The two primary goals of classification are to analyse the

history of data and to develop a model that describes the correlation between the objects and the class labels within the historical data. The model produced from this technique represents a set of rules, also known as classifiers, which are used to forecast class labels of previously unseen objects.

Clustering refers to the process of grouping objects into a subset of objects that share a similar pattern. This technique is known as unsupervised learning and differs from classification based on its use of predefined classes.

The association rules technique is one of the most researched areas of data mining. The technique involves finding a relationship amongst items in a transactional database and is a mixture of both descriptive and predictive aspects. By producing a set of records, it can predict the occurrence of each item based on patterns identified in the data.

One of the most important processes after building data mining models is evaluation. This serves two purposes: First, it predicts how well the final model will work in future, and second, it helps find the model that best represents the training data [85].

Classification rules are used to predict unseen data. Therefore, the measure used for evaluating quality is predictive accuracy; this is defined as the number of testing examples correctly classified by the rule set, divided by the total number of testing examples [86]. The widely known performance metrics used for evaluating classification results are precision, recall, and accuracy [87]. Higher precision means fewer false positives, and higher recall means fewer false negatives [88-90].

#### **2.4.2. Data pre-processing**

The main purpose of the data pre-processing stage is to recognize the quality and format of data. Data pre-processing is important for data mining algorithms' discovery of clear patterns among the data [71]. The data pre-processing stage focuses on cleaning and constructing the data: for example, creating derived variables, discretising where relevant, integrating if necessary, and converting data to the format that satisfies the requirements of the given data mining tool. According to Sharma et al., there are several potential outcomes in developing the data, regardless of the derived variables or the ways in which the data can be discretised or integrated. Transforming the data to the format needed for a specific data mining task requires tools and expertise to understand and identify how it can be done [91].

Data preparation is used to direct the discovery of knowledge and to help develop effective systems [71]. It is important because real-world data are not pure, yet high-performance

mining systems require high-quality data; having such high-quality data allows more concentrated patterns to be formed [92].

In this research, we must discuss the most important breakthrough in text analytics, that of sentiment analysis. Preparing the information in textual data helps to identify the subjective domain. Pre-processing consists of different steps—such as removing stop words, removing white space, stemming, extending abbreviations, negation handling, and feature selection [62]—and is designed to extract just the relevant content from the text [93]. A later chapter in this thesis on specific problems such as data pre-processing for classification will contain detailed discussions on text features.

#### **2.4.3. Feature selection advantages for text classification**

All the pre-process steps (normalisation, tokenization, and so on) are called “feature selection”, and this technique has important advantages [94]:

1. It reduces the running time of classifiers, as unnecessary features are eliminated, and therefore the feature vectors used by the classifiers are smaller [66].
2. It yields more accurate results, as the elimination of unnecessary features ensures that the meaningful features are retained [66].
3. It limits the expected memory to deal with the documents' vectors by decreasing the quantity of features represented by the vectors [66].

#### **2.4.4. Text classification (based on data mining)**

Text classification is the task of assigning one or more predefined categories to a text document. An example of this is an email system’s categorisation of messages as "junk email" or "inbox email" [95]. Text mining usually starts with a data pre-processing phase [96].

Many applications have used text mining techniques to discover information and knowledge from unstructured data, ranging from the manufacturing to the construction industry. Textual data may be unstructured, as free text in natural language, or semi-structured, as in HTML or XML documents [97]. Many machine learning methods are designed for sets with vectors representing attribute values, so the text data have to be transformed into document vectors [98]. The order of words is not taken into account; the frequency of each word in a document is taken as the value of the attribute [97], and words that occur on frequently but do not add to interpretations are often removed. These Term Frequency (TF) values are often weighted to take frequent words into consideration [99]. Sentiment analysis is one of the typical tasks carried out in text mining, with its value being in its ability to measure sentiments by comparing

them with a dictionary [33]. However, it should be noted that sentiment analysis has focused mainly on texts in the English language, and applying sentiment lexicons to any other language may well bring about mistakes in adaptation [100].

#### 2.4.5. Methods and algorithms

In this research study, it was decided to follow Knowledge Discovery and Data Mining (KDDM) processes, which involve business understanding, information understanding and preparation, and creating modelling, assessment, and deployment [91].

Attempting to classify text in Arabic and apply sentiments expressed on Twitter may be a challenge. Firstly, it requires building a set of models that can accurately predict class. These can then be used to predict information about class for objects that have previously not been classified. Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers are widely used in sentiment analysis.

- Naïve Bayes

NB depends on the precise nature of the probability model; therefore, the classifier can be trained very efficiently in a supervised learning setting and utilise the method of maximum probability [101], [102]. The following provides a description for NB:

By assigning to a given document  $d$  the class  $c^* = \arg \max_c P(c | d)$ , the NB classifier is driven by first overlooking the Bayes rule,

$$P(c | d) = \frac{P(d | c)P(c)}{P(d)}, \quad (2.1)$$

where  $p(d)$  plays no role in electing  $c^*$  to guess the term  $P(d | c)$ . NB decomposes it by assuming  $f_i$ 's are conditionally independent, given  $d$ 's class:

$$P(c | d) = \frac{P(c)(\prod_{i=1}^m P(f_i | c)^{n_i(d)})}{P(d)} \quad (2.2)$$

In other terms, the NB model has many advantages, such as treating each feature independently, being relatively fast to compute and easy to construct, and having no need for any complex schemes and less over-fitting compared to other models [103].

- Support vector machines

SVM have been applied successfully in many sentiment analysis tasks. This technique has outperformed other machine learning techniques because of its primary advantages.

“First, they are robust in high dimensional spaces; second, any feature is relevant; third, they are robust when there is a sparse set of samples; and, finally, most text categorization problems are linearly separable.” [89]

The basic idea of SVM is to find a hyperplane represented by vector  $\vec{w}$  that separates the document vectors in one class from those in other documents. Suppose  $c_j \in \{1, -1\}$  is identical to positive and negative in the true class of document  $d_j$ :

$$\vec{w} := \sum_j \alpha_j c_j \vec{d}_j, \alpha_j \geq 0, \quad (2.3)$$

where the  $\alpha_j$ 's are obtained by solving a double optimization dilemma. Those  $\vec{d}_j$  such that  $\alpha_j$  is greater than zero are called support vectors, since they are the only document vectors shared in  $\vec{w}$ . Classification of test instances consists simply of deciding which side of the  $\vec{w}$ ,s hyperplane they enrol. In other words, the SVM is a classification prediction tool that uses machine learning theory to maximize predictive accuracy while automatically eliminating over-fitting of the data [80, 104-106].

#### 2.4.6. Evaluation

The performance metrics widely used to evaluate classification results are accuracy, precision, and recall [107] [90]. The accuracy chosen as the main criteria because it is the proportion of both true positive and true negative unlike the precision and recall, which is the fraction of true positive only. Where tp is the number of true positive predictions, tn is the number of true negative prediction, fp is the number of false predication, fn is the number of false negative, and tn is the number of true negative predication.

- Accuracy =  $(tp + tn)/(tp + fp + fn + tn)$  (2.4)

Accuracy involves calculating the ratio of true results (positives and negatives).

- Precision =  $\frac{tp}{tp+fp}$  (2.5)

Precision is the number of true positives out of all positively predicted documents.

- Recall =  $\frac{tp}{tp+fn}$  (2.6)

Recall is the number of true positives out of the actual positive documents.

### 2.5. Customer Relationship Management (CRM)

CRM is a strategic approach that aims at improving shareholder value by developing effective relationships with key customers [108]. CRM uses IT alongside relationship marketing strategies to promote potentially long-term relationships that result in profitable alliances [109].



CRM provides increased opportunities by using data to develop better customer understanding and marketing strategies. Moreover, to improve customer relations and understand the information generated from customer data, enterprises need a cross-functional integration of people, operations, and marketing efficiencies as well as technology and applications. The same authors determined that the CRM mission is to build a strong relationship with customers, thus increasing customer lifetime and company profitability and potentially making competitive progress. However, the authors identified five processes needed for a successful CRM: strategy development, value creation, channel integration, information management, and performance assessment [109].

Another definition of CRM is that it is a process providing added value to customers that can motivate them to remain loyal or to buy again [110]. Dyche [110] divided the architecture of CRM into two important sections: operational CRM and analytical CRM. Operational CRM is known as “front office” and indicates the areas of direct customer contact, such as calls to a company or an email promotion using touch-point interaction. Analytical CRM is known as “back office” or “strategic CRM” and refers to the understanding of a customer’s activities and behaviours in the front office to help firms identify resources needed for valuable customer segmentation [110].

For some, CRM is the consolidation of marketing, sales, and service functions through IT, allowing business process automation so that the value of each customer interaction can be increased. Others have proposed CRM as the means to extract accurate and sensitive customer information to identify customer segmentation or to improve product innovation and gain considerable market value [111].

Winer [4] illustrated that the market depends on two dimensions: customer acquisition and customer retention. Distinguishing between the two activities can be difficult, but it is acquisition that dominates retention.

### **2.5.1. Customer Relationship Management (based on data mining)**

Due to the competitiveness in today’s markets, it has become clear that it is essential to take into account the importance of customer support and customer retention in industry [112]. CRM has expanded into a significant sector in its own right and has made considerable technological progress [112]. A number of computer science algorithms, including optimisation and scheduling, have been developed with CRM in mind [113, 114]. CRM is a strategy that businesses can employ to build long-term relationships with their customers [115]. The foundation for a successful CRM strategy nowadays depends on customer data and information technology. This has been further aided by the internet, which has significantly

increased marketing opportunities and changed the way businesses keep contact with their customers [116].

The application of data mining has resulted in its becoming an important field of CRM. Many studies have been conducted to explain customer behaviour. For example, Rygielski et al. [13] explained the importance of the customer lifecycle framework and its relation to understanding customer behaviour. Guta and Aggarwal [117] stated that operational CRM needs analytical CRM with predictive data mining models. Consequently, a successful business needs a marketing manager who understands their customers and what they need and implements data mining with an effective and appropriate model. Achabal et al. [118] reported that CRM is motivated by the most important area of research, namely, data mining, which consists of evaluating a very large volume of data; this leads to more time being needed, so to reduce this time, efficient data mining algorithms need to be implemented. In this way, the computational time, as well as efficiency, increases. Mostafa [107] and Galitsky et al. [6, 112] agreed that CRM must employ machine learning in the domains of NLP approaches, public opinion tracking, and text filtering. In the past year, the concept of the traditional CRM has been changed to social CRM, and this may help in managing customer relationships due to the growth of social media over the internet [119-121].

### **2.5.2. Integrating sentiment analysis into CRM**

In general, a social CRM conversation between customers and enterprise agents over social media is called social CRM. Social CRM can influence the customer community to solve a customer's problem and turn a bad opinion into a good one [122].

Furthermore, social media has been used extensively by enterprises in the recent past to get insights about what users think about their products or services; this is typically achieved in a "listening" mode, i.e., a large amount of data from multiple social media sites is analysed in offline mode to extract aggregate-level business insights [123, 124]. Usually, the relational model database is the backbone for most companies or organizations; it stores the data in a structured order with rows and columns. However, a huge percentage of data in organizations or companies are located in unstructured data such as text data. This shows the need to analyse the unstructured data for companies' benefit [125].

Many researchers have shared their experiences with this subject. Ajmera et al. [126] built a social CRM that enabled firms to engage with customers by presenting analytical methods to identify actionable posts and analysing them. They presented novel features such as user intent and severity of issues in a customer complaint to determine a post's priority. In addition, Yaakub and Zhang [127] proposed a multi-dimensional model for opinion mining to integrate

customers' characteristics and their related opinions about products. They used POS tagging to pre-process their data, trying to capture three parts from each document: nouns, which describe the name of the product, and adverbs and adjectives, which describe the sentiment toward that product. However, it was hard to evaluate their model's conclusions about product attributes based on customers' opinions. In other words, it is hard to cover all details in CRM using only sentiment analysis. The baseline models for the above study were developed by trawling the reviews before putting them into a review database [128]. They used association mining and POS to isolate the reviews into categories. They tried to identify features by finding nouns or noun phrases and then identified infrequent features by looking for adjectives. Then they extracted the opinion based on the adjective close to the feature word before generating the review summary. The same concept for integrating social data into a structured database was applied by [129], who presented an approach to integrate opinion mining that extracted from web feeds into corporate Online Analytical Processing (OLAP) in order to allow analysts to perform new analysis tasks with both sources of information. They faced problems with product features extracted from their customer opinion forums, since the result is a rich dataset that contains synonyms and cannot be used for complex queries. To conclude, working only with sentiment analysis is not enough to fulfil the integration because the CRM details are absent. On the other hand, using opinion mining to support traditional CRM is more effective.

## **2.6. Summary**

This chapter has discussed the needs of text mining and CRM in the context of knowledge discovery. It consequently provides a background review to address the key question in the thesis, "How can we analyse customers' sentiments and opinions in text CEM to support CRM?" The basic tools for knowledge discovery include text mining tools to handle classification techniques for both unstructured data (e.g., textual data) and structured data (e.g., transactional data), which are estimated to provide more knowledge discovery for enterprise solutions. These are capable of handling two sources of data (textual data and CRM data) to add more information to discover useful knowledge relationships.

## **Chapter 3: Techniques for Arabic Sentiment Analysis**

### **3.1. Introduction**

There is a growing body of literature that recognises the importance of social media research, and several studies have been applied to obtain users' opinions in fields such as e-business, e-learning, politics, and trading. Social media is becoming a rich resource for mining user opinions. Social opinion has been analysed using sentiment analysis. For example, some studies on trading strategies have shown that sentiment analysis of news, documents, quarterly reports, and blogs can be useful. In this chapter, Twitter has been selected as a platform to carry out opinion mining on Saudi tweets (that is, Standard and Arabian Gulf dialects) about trading strategy in the Saudi stock market.

This chapter explores techniques for pre-processing and classification of Arabic sentiment analysis. In section 3.2, the characteristics of the Arabic language are presented; the general state of the art for sentiment analysis in Arabic is described in section 3.3; an Arabic text pre-processing and sentiment analysis process is proposed in section 3.4; Section 3.5 explores optimal techniques for pre-processing and classification of Arabic sentiment analysis in three different domains. Finally, repeated patterns are extracted for future applications within the field.

### **3.2. Challenges of using sentiment analysis in Arabic**

This section shows the structural complexity of the Arabic language and the challenges related to Arabic content analysis. These difficulties are likewise connected with the growth of social media networking and the utilisation of colloquial Arabic on these social platforms. It is clear that the structure of the Arabic language differs from that of other languages, such as the European ones. For instance, Arabic is written from right to left, and the letters have different forms depending on their location. Gender also determines the affixes of words, and capital letters do not exist in Arabic [65]. Such issues cause challenges in attempting to analyse Arabic text. In addition, there is a lack of research to provide direction on how to proceed with sentiment analysis in Arabic. The informal (colloquial) language is diverse and would be very difficult to standardise. In addition, there is the challenge of social media, where emotions can range from anger to optimism in texts and large numbers of words can have ambiguous meanings [94], especially when young people start inventing their own words. Moreover, Arabizi, Arabish, or Romanized Arabic, which refers to the writing of Arabic using Latin characters, is widely used to write Modern Standard Arabic (MSA) as well as Arabic dialects

on social media platforms. Dealing with this form of writing has been the subject solely of studies that aim to detect Arabizi and convert it into Arabic. As for sentiment analysis, the published works have not dealt with this problem, as texts are pre-processed to filter out all Latin letters [130]. Duwairi et al. [130] reported that Arabic is a morphologically rich language, which creates challenges, some of which are presented in the following examples:

- A given root can take several forms depending on the context, such as “أحب-يحب-أحبو”, which means in English “love, he loves, they love”.
- Many people tend to use the dialect of their country instead of using MSA; for example, “اهلين” and “سلام” both means, “welcome”.
- Compression of words; for example, “ممتازاًالاز”, which means “extremely perfect”, is written in MSA as “ممتاز”.
- Arabic has various diacritics; the meaning of words can be very different depending on them. For instance, knowledge “عَلِمَ” and flag “عَلَمَ” can both be read as the same when written without diacritics “علم”.
- Negation words, which change the meaning of the verb to the opposite, e.g., “لم أعجب” “بهذه السيارة”, which means, “I did not like this car”.

Such complexities result in difficulties in analysing Arabic text, and this is intensified by the lack of research that has been conducted in this field. Authors have tried to analyse Arabic texts by different methods such as NB [102] and maximum entropy [131]. However, their approaches have many limitations, such as Arabic informal language’s lacking structure and being difficult to standardise [130], and NLP is not applicable to Arabic [132] because of its morphological richness [29].

### **3.3. Arabic language**

Arabic is a Semitic language spread throughout North Africa and the Middle East and spoken by more than 400 million individuals in the Arab world and in excess of 1.5 billion Muslims overall [133]. The Semitic language family incorporates Arabic, Aramaic, Hebrew, and Amharic and has assumed a vital part in the social scene of the Middle East for over 4,000 years [10].

Today, Arabic individuals compose their writings in two structures: MSA and Colloquial Arabic [129]. Korayem et al. [133] announced that MSA is the formal composed standard ordinarily utilised all through the Arab world. MSA is not the native language spoken in any Arabic nation, as its basis is traditional Arabic, the language used in the Qur’an [133].

Colloquial Arabic is the language spoken in everyday communications and is not the taught language [133]. It changes significantly based upon locale and social class, among other components; nonetheless, although it is basically a spoken dialect, it can likewise be found in written form [133]. This is probably in online networking discourse, as well as in some plays and poetry [134]

### **3.3.1. Arabic language characteristics**

Arabic is rich in morphology and syntax. It is normally written with optional diacritics and without the idea of capitalisation. These attributes make dealing with Arabic a challenge for both learners and researchers [135]. This section will introduce the elements of Arabic script, morphology, and syntax.

#### **3.3.1.1. Elements of the Arabic script**

The Arabic abjad contains 28 letters and three different kinds of diacritics; most letters within a word are connected in writing [136]. There are no uppercase letters in Arabic, and it is read from right to left [136]. The 28 essential letters represent consonants and can take different forms depending on their position in the word, whether at the beginning, middle, or end [136]. Only six letters (ا, د, ذ, ر, ز, و) do not have a particular shape medially and should utilise the initial form, although in this case they are not connected to the succeeding letter [136].

Diacritical marks are aimed at reducing any ambiguity in words [136]. For example, knowledge “عَلْمٌ” and flag “عَلَمٌ” both can be read as the same word when written without diacritics “علم”. Diacritical marks are important for helping with the pronunciation of words, although Arabic words can be written with or without diacritical signs [135, 137].

#### **▪ Arabic morphology**

In linguistics, morphology relates to the study of word and sentence structure, and Arabic is characterised by a complex, though productive, morphology [136]. A brief description of Arabic morphology follows, with examples for clarification.

- Stemming is one of many tools used in information retrieval; it is the process by which a word is reduced to its stem to obtain the most accurate meaning. In English and many other western European languages, stemming is mainly a process of removing a suffix [138]. A stem is a part of a word that cannot be divided further into smaller parts and expresses the basic meaning [139]. For instance, the stem of the word “مهاجرون” “mohajeron” immigrants is “هجره” “hijrah” migration. In addition, the stem refer to part of word that is common in all of its inflected language [134].

- Affixes connect to the stem. There are three sorts of Arabic affixes: (a) prefixes, which link before the stem; (b) suffixes, which link after the stem; and (c) circumfixes, which include the stem. Prefixes and suffixes are predictable and are used to denote gender, number, person, aspect, and so on [140]. For example, “بالمدينة” “bi'lmdynh” in the city has a stem “مدينة” mdyneh city and two prefixes; the first is “ب” “b” in, which is a preposition, and the second is “ال” “al” the, which is the definite article [141].
- Clitics are attached to a stem or to each other and are linguistic units that are pronounced and written like an affix but are grammatically independent [142]. From a linguistic perspective, if one can identify the grammatical components of an Arabic linguistic unit attached to a stem, it should be considered a clitic [142]. Some Arabic clitics are identifiable based on their position in the word; for example, in “ت + كتب = تكتب” she writes, the letter “ت” “t” can be considered a clitic on the verb “كتب” write [135].
- Derivation is the process of new words being built from other words [33], for example, “درس” “daras” to study, “دارس” “daaris” student. Inflection is the variation of a word with the same meaning in different grammatical contexts (e.g., in English: write, wrote, written) [143].

### 3.3.2. Arabic sentence structure

Arabic sentences are categorised as either verbal or nominal sentences [33]. The Arabic verbal structures are made out of two classes, a noun and a verb. A noun is a name or a word that describes a person, thing, or idea, while a verb is a word that signifies an activity. In terms of tense, the verb could be past, present, or future [144]. A nominal sentence is defined as a sentence that begins with a noun; the structure of the nominal sentence consists of two parts: the subject and the predicate [33]. A verbal sentence starts with a verb, followed by a subject and one or more objects. The verb and subject have the same gender, number, and person, and in each case the sentence can be straightforward, or complex [135].

### 3.3.3. Difference between Standard Arabic and Local Arabic

The term “Standard Arabic” used in this study refers to the officially endorsed variety of Arabic that is mostly associated with education and literacy. On the other hand, the term “Local Arabic” is used in this study to refer to a number of regional varieties that are often used by Arabic speakers in everyday communication and are mostly associated with informal situations [145, 146]. Ferguson [147] stated that in many speech communities, two or more varieties of the same language are used by some speakers, depending on different conditions.

There are similar instances in other languages where there is a standard version and regional dialects that are used in everyday life; in Italian, for example, many speakers use the local dialect in their own family or home environments but a standard version when communicating with others from outside their region or on public occasions; there are basically different social functions of language [148]. There are also different spoken conventions between Muslim and Christian Arabs living in the same communities; Bruweleit [149] explored this with Lebanese communities and Hann et al. [150] with Baghdadi communities. Language can therefore reflect a cultural identity as well as a regional identity.

After discussing the characteristics of Arabic that may influence the process of sentiment analysis, the next section reviews how sentiment analysis of the Arabic language has been approached.

### **3.4. Sentiment analysis in Arabic**

Sentiment analysis, also referred to as “opinion mining”, involves the use of NLP and machine learning to identify and extract subjective information from a piece of text. Sentiment analysis is extremely useful, as it enables us to gain an overview of public opinions or attitudes towards certain topics that relate to products or services.

In general, subjectivity refers to parts of language that are utilised to express sentiment, emotions, and assessment—in other words, how people perceive things. The process of subjectivity classification involves classifying texts as being either objective or subjective and deciding whether or not a judgement is being made. Subjective content can additionally be assigned to an opinion by identifying whether it is expressing positive, negative, or neutral emotions [151]. In addition, Liu [28] and Duwairi et al. [152] observed that sentiment analysis is a research field that analyses people’s opinions towards entities such as products and services, and this is also supported by Haddi [33].

The nature of Arabic text is different from that of English and other languages [9]. Arabic words are precise with gender and number, and their affixes can change accordingly; only small letters are used in Arabic text, and special grammatical rules are required to detect nouns, acronyms, and abbreviations [11, 143]. Moreover, Arabic texts are written from right to left, and the forms of letters change depending on their position in a word [10].

Limited research has been conducted on Arabic sentiment analysis, so this is a field that is still in its early stages [10]. However, Boudad et al. [143, 153] reviewed the challenges and open issues that need to be addressed and explored in more depth to improve Arabic sentiment analysis, finding that these include domain, method of sentiment classification, data pre-processing, and level of sentiment analysis. They show that, in contrast to work on the



English language, work on Arabic sentiment analysis is still in the early stages, and there are a lot of potential approaches and techniques that have not yet been explored.

Nevertheless, sentiment analysis has been receiving some attention in the NLP community, particularly with regard to polarity classification, whose aim is to decide whether or not a sentiment is expressed in a text. On the other hand, Ibrahim et al. [132] have presented a multi-genre tagged corpus of MSA and colloquial language, with a focus on Egyptian dialects [9]. Interestingly, they suggested that NLP supplements, which have been applied to other languages like English, are not valid for processing Arabic directly [9]. The main limitation of this approach is that it does not capture the morphological richness of Arabic.

Abdulla et al. [152] explored the polarity of 2,000 collected tweets on various topics, such as politics and art. The authors compared a lexicon-based approach as an unsupervised approach with machine learning or a corpus-based approach as a supervised approach. They used SVM, NB, KNN, and D-tree for their documents' sentiment classification. They showed that SVM and NB have better accuracy than other classifiers in a corpus-based approach. Their results reported that the average accuracy of SVM was 87.2%, while the average accuracy of NB was 81.3%. The main reason for this result is that they tried to build sentiment lexicon-based dictionary in their second experiment, whose sample was a very small size and included only the original word for their dataset. They claimed that they built tools (ATC Tool) that can work as NLP, but they did not explain how they labelled their data (manual or automated), given there is no Arabic sentiment dictionary.

El-Halees's [131] combined approach classified documents using lexicon-based methods, used these as a training set, and then applied k-nearest neighbour to classify the rest of the documents [9]. Such approaches enabled the achievement of a higher level of accuracy: 80.40%. Al-Harbi [154] used two classification algorithms (SVM and C5.0) to predict text polarity and reported average SVM accuracy of 68.65%, while the average accuracy of C5.0 was 78.42% [9]. Finally, El-Kordi et al.'s [155] Arabic text web documents from the Al Jazeera website were classified into five main categories, using the machine learning or corpus-based approach. They found that some of the misclassifications made when using NB contained a large number of words belonging to other categories, indicating that documents associated with a category containing numerous words have a higher frequency in another category's domain.

The main reasons for the difference of the classification results with the given examples above is domain diversity, as well as the way that data were pre-processed, the size of the dataset, and of course the approaches that were used in each article. Moreover, In terms of the

accuracy of the last survey, which studied recent work in Arabic sentiment analysis, SVM was applied successfully in several sentiment analysis tasks [143], [153].

### **3.5. Arabic text pre-processing and sentiment analysis process**

This research study will explore the impact of text pre-processing on Arabic text classification, using popular text classification algorithms such as SVM, and NB. Text pre-processing includes applying different term weighting schemes to Arabic text. Zhang et al. [71] argued that the reasons for data preparation cover three aspects: (1) real-world data are impure because they include incomplete and noisy data that will lead to errors or outliers and inconsistent data that may contain discrepancies in codes or names; (2) high-performance mining systems require high-quality data, which means the data pre-processing can reduce the original dataset but will improve the efficiency of the data and will lead to high-performance mining systems, which require high-quality data; this means that data pre-processing can reduce the original dataset but will improve the efficiency of the data mining which will lead to (3) high-quality patterns. There are ways of dealing with these issues, for instance, by recovering incomplete data and filling the values missed, or by reducing ambiguity. In addition, the impact of pre-processing the text data in terms of unstructured data may contain rich semantic relationships between various entities. One method is to utilise entity extraction to specify linguistic relationships. Other approaches are to remove stop words, stem the data, and utilise a bag-of-words representation.

The effectiveness of text pre-processing with different term weighting combinations on Arabic text using popular text classification algorithms has not been studied adequately in the literature. For example, in two studies, Duwairi et al. [94, 156] compared three dimensionality reduction techniques: stemming, light stemming, and word cluster. They used the k-nearest neighbours (KNN) classifier to perform the comparison. Their performance metrics were time, accuracy, and the size of the vector; light stemming was found to be the best in terms of classification accuracy. Thabtah [157] investigated different variations of Vector Space Models (VSM) and term weighting methods, using a KNN algorithm; experimental results showed that the distance function with Term Frequency–Inverse Document Frequency (TF-IDF) achieved the highest average score. In addition, Said et al. [158] carried out an evaluation of several morphological tools for Arabic text classification using SVMs, finding that a light stemmer and good feature selection method optimised Arabic text classification for small datasets and yielded small threshold values for large datasets [159]. Additionally, they reported that whilst using raw text led to the worst performance in small datasets, it actually performed best for large datasets. It can be seen that the performance of pre-processing tools depends on the characteristics of the dataset being used.

### 3.5.1. Preparing the Arabic text

#### 3.5.1.1. Normalization

It has been mentioned that diacritics are one of the elements of the Arabic script. Moreover, diacritics and some special characters involved in Arabic text can be more challenging in this phase. It is necessary to replace a word or term that includes diacritics with the same word without any diacritics. In addition, one of the challenges that has been mentioned is Arbaaz, or using English words or code to indicate entities or names inside an Arabic statement. NLP tools perform many pre-processing steps to normalise and prepare Arabic text. For example, they remove punctuation, diacritics, and non-Arabic letters, and specified Arabic letters are normalised: for instance, ا, آ, and إ are converted to ا (bare Alif); ي and, ع (Yaa') are replaced by ع (Yaa'); and ة, ه (Taa', Haa') are replaced by ه (Haa'). Taghva [160] argued that removing the Hamza in this case does not affect the root. However, because of similarity in appearance, a more challenging type of spelling variation is that certain symbols joining Hamza or Madda with Alif (e.g., اأ and إا) are sometimes written as a plain Alif (ا).

#### 3.5.1.2. Tokenization

In the case of Arabic, where a single word can comprise up to four independent tokens, it is essential that morphological knowledge be incorporated into the tokenizer [161]. A tokenizer relies mainly on white spaces and punctuation marks to identify word boundaries (or main tokens). Punctuation marks are used in Arabic, such as the comma '،', question mark '؟', and semicolon '؛', and are also considered main tokens, along with numbers [162]. Thus, tokenization requires knowledge of the constraints on how affixes may be linked, and clitics within Arabic words and Arabic text must be cleansed and normalized before the tokenization phase [141].

#### 3.5.1.3. Stemming

Stemming is a technique that helps with text classification. Various types of stemming methods can help manage different languages, for example, the Khoja stemmer for Arabic and the Porter stemmer for English [146, 159, 163].

In Arabic, the light and the root-based stemmers are the best-known stemming approaches.

- Root-based stemming involves the removal of prefixes and suffixes to produce the root of a given Arabic word and therefore affect the semantics of words. It reduce words to their roots since several words with different meaning might have the same root. The Khoja stemmer is the best-known of these approaches [94].

- Light stemming allows effective information retrieval; it is utilised to remove the suffixes and prefixes and doesn't affect the semantics of words as it doesn't extract the root [94].

The Khoja stemmer (a root-based stemmer) does both pattern checking and stem lists [34], but light stemming has outperformed other approaches [80]. This study will carry out a comparison of both types of stemmers in the experiment chapter.

#### **3.5.1.4. Stop words**

In Arabic written text, some words are very common but add no additional meaning to the content of the text (such as في "in", على "on", أين "where"); these are called stop words [34, 164]. They serve only a syntactic function but do not indicate subject matter. The reason to remove stop words is that they can affect the weighting process, as they have a high frequency. The removal of such words can be carried out before indexing of an input document begins. The benefit of excluding them is that these characters, words, and numbers do not help with determining the document topic. It may also be useful to exclude rare words, which can be defined as words that occur in a low percentage of the processed documents [161], [165], [166].

#### **3.5.1.5. N-gram**

The work of Xu et al. [161] used N-grams, with and without stemming, for text searching. Their results indicate that a combination with stemming improved the performance of tri-grams in search retrieval; however, it was not statistically significant. In addition, Keselj et al. [167] reported that using N-gram models in NLP is a generally basic concept that has been effective in different applications: for instance, character-level N-gram dialect models, which can be effectively related to any language.

### **3.5.2. The weighting schemes**

The aim of weighting schemes is to give higher weight to the most distinctive words or terms in the text. In other words, weighting schemes aim to represent high-quality text. The less the term is useful for segregation between texts as relating to different classes, the more the term shows in different texts representing various classification. Two schemes are utilised in this study, TF-IDF and BTO (Binary Term Occurrence). TF-IDF describes the importance of a word for a document and consists of two parts: Term Frequency (TF) and Inverse Document Frequency (IDF) [16, 168, 169], where  $F_{ij}$  is the number of occurrence of term  $i$  in document  $j$ ,  $F_{dj}$  is the total number of term occurring in document  $j$ , and  $F_{ti}$  is the total number of documents in which term  $i$  appear at least 1.

$$TF - IDF = \left(\frac{f_{ij}}{f_{dj}}\right) \log\left(\frac{1}{f_{ti}}\right) \quad (3.1)$$

BTO is defined as a binary value, and a word or term gets 1 if present in a document or 0 otherwise:

$$BTO = 1 \text{ for } f_{ij} > 0 \text{ and } = 0 \text{ otherwise} \quad (3.2)$$

The architecture of Arabic text has thus been defined; this study is going to describe how a NLP can play a significant role in taking out the noise from the Arabic text and consequently generating high quality text.

### 3.5.3. Sentiment analysis via the machine learning approach

Since the Arabic language still has no sentiment lexicon, sentiment analysis via the machine learning approach is the preferred option in the literature. The first step in the sentiment analysis process is a labelling process carried out by a human. Figure 1 summarizes the sentiment analysis process of Arabic tweets after the labelling process. The machine learning approach is a supervised method in which a set of data labelled as positive or negative is represented by feature vectors. After that, the classifier uses those vectors as training data in order to determine the combination of specific features that relate to a specific class.

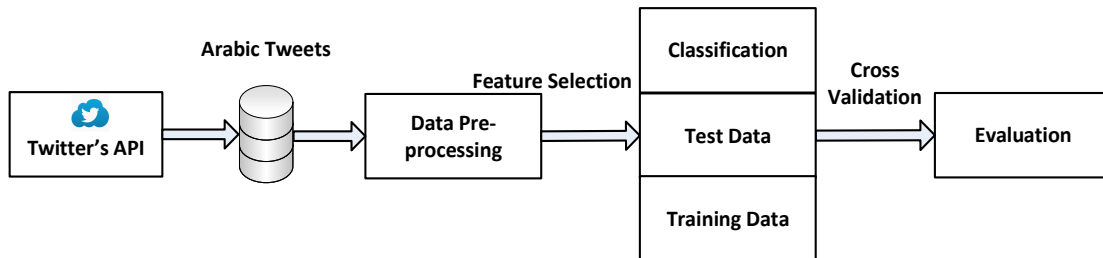


Figure 1: The process of sentiment analysis in an Arabic tweet

### 3.5.4. Techniques for pre-processing and classification of Arabic sentiment analysis and applications

As mentioned in 3.3, most of the existing research focuses on a specific application domain with its own pre-processing approach. To explore optimal techniques for pre-processing and classification of Arabic sentiment analysis, the process of sentiment analysis in Arabic tweets, which is described in section 3.4, will be deployed into three different domains—the Saudi stock market, educational domain, and service domain—to evaluate the effectiveness of text pre-processing with different combinations of term weighting schemes on Arabic text and to learn the effectiveness of using popular text classification algorithms for Arabic sentiment

analysis. Finally, the repeated patterns will be extracted to provide a recommendation for future applications with similar approach.

### **3.5.5. Application in the financial domain: Saudi stock market prediction**

This experiment presents the proposed framework for classification of financial Arabic text to positive and negative classes within a corpus collection. However, mining unstructured data with NLP and machine learning techniques can be challenging, especially with Arabic because of limited research conducted in this area, as mentioned previously in this chapter.

In the late fifties, Saudi Arabia unofficially started the Capital Market Authority<sup>1</sup> (CMA) for regulating and developing the Saudi Arabian capital market. This was achieved by issuing provisions for rules and regulations by implementing the Capital Market Law. The CMA was established to protect investors and the public in Saudi Arabia from unfair and fraudulent trading practices. Currently, Tadawul is the only Saudi Stock Exchange in Saudi Arabia. The CMA regulates Tadawul, and within Tadawul there are many sectors, such as banks, financial services, petrochemical industries, retail, telecommunication, and information technology. Recently, the Saudi government decided to allow foreign institutions to buy and sell shares in the Saudi market in accordance with the times and dates established by the CMA. For this study, Twitter was chosen as a platform on which to carry out opinion mining on Saudi tweets (that is, Standard and Arabian Gulf dialects) on trading strategy in the Saudi stock market.

#### **3.5.5.1. Collecting the data**

To collect the corpus of Arabic tweets, a small desktop application (Twitter Data Grabber) was developed using C# with Twitter's official Developers API.<sup>2</sup> It allows us to label and save relevant tweets and discard irrelevant ones. The tweets' source data were obtained from Mubasher's company website<sup>3</sup> in Saudi Arabia. The company is a leading stock analysis software provider in the Gulf region. A collection of tweets was compiled over a 53-day period between 17 March and 10 May 2015. The dataset includes 3,335 tweets covering all share sectors for the Saudi stock market that appear on the Tadawul website.

First, keywords and idioms in Arabic were selected from the targeted corpus such as increase, growth, decline, fall, rise, cash dividends, distribution of bonus shares, and not to distribute. Then the keywords were divided into the following two classes: positive and negative. Table 2 shows the most common Arabic keywords in the two classes in the standard Saudi dialect.

---

<sup>1</sup> <https://cma.org.sa>

<sup>2</sup> <https://dev.twitter.com/rest/public>

<sup>3</sup> <http://twitter.com/mubashersa>

Table 2: Keywords in Standard Saudi Dialect represent the classes

Arabic Positive Sentiment	English Translation	Arabic Negative Sentiment	English Translation
نمو	Growth	تراجع	Decline
ارتفع	Rise	انخفض	Fall
ازدهار	Prosperity	تقلص	Shrink
ربح	Profit	خسارة	Loss
ارباح	Earnings	خسائر	Losses
توزيع أرباح	Dividend	ارتفاع خسائر	High losses
مغري	Seductive	غير مغري	Uninteresting
اخضر	Green	احمر	Red

The users' privacy issues were handled as follows. Tweets on Twitter are set to public and can be viewed by any person regardless of membership of Twitter. The tweets used in this study were obtained from public timelines. The Twitter API has several methods, such as GET statuses, user\_timeline, home\_timeline, and GET search / tweets. This study implemented two methods to download tweets:

- **GET statuses/mentions\_timeline** returns the 20 most recent mentions (tweets containing a user's @screen\_name) for the authenticating user. The timeline returned is the equivalent of the one seen when you view your mentions on twitter.com.
- **GET statuses/user\_timeline** returns a collection of the most recent tweets posted by the user indicated by the screen\_name or user\_id parameters. User timelines belonging to protected users may only be requested when the authenticated user either "owns" the timeline or is an approved follower of the owner. The timeline returned is the equivalent of the one seen when you view a user's profile on twitter.com.

### 3.5.5.2. Assigning labels

Operators who are experienced with Saudi stock shares labelled the data manually. Negative tweets were set the label "-1", while Positive tweets were set the label "1". Irrelevant tweets were deleted from the database. Table 3 shows the number of tweets in our database; 1,392 tweets were deleted in the labelling process.





Finally, 'Process Documents from Data' operator generates a word vector from the dataset after pre-processing and represents the text data as a matrix to show the frequency of each term's occurrence. TF-IDF and BTO utilised as a weighting scheme to create the word vector, after which relevant data mining approaches were explored, e.g., NB and SVM, and for building models that classify tweets according to their sentiment polarity into positive and negative. The evaluation is carried out using precision and recall methods.

#### **3.5.5.4. Experiment parameter settings**

The proposed data model was examined and validated utilizing the Rapidminer platform which is an agile enterprise platform for predictive analytics. The SVM and NB settings followed the default values for most of the parameters in order to get more accurate results.

##### **▪ Support vector machine (SVM) operator parameter settings**

SVM operator generates an SVM classification model. This model can be used for classification and provides good results for many learning tasks. In addition, it supports various kernel types, including dot, radial, polynomial, and neural [170].

- SVM type: For this study, C-SVM was chosen to handle the classification task.
- Kernel Type: For this study, a linear kernel type was chosen because the relation between class labels and attributes is linear.
- C is the penalty parameter of the error term and set to its default value, which is zero.
- Cache size is an expert parameter. It specifies the cache size in megabytes and for this study was set to its default value, which is 80.
- Epsilon: This parameter specifies the tolerance of the termination criterion and was set to the default value, which is 0.001 [170].

##### **▪ Naïve Bayes (NB) operator parameter settings**

NB operator generates a NB classification model. The NB classifier is independent, so the presence or absence of a feature of a class is unrelated to any other feature [170].

- Laplace correction:

This parameter indicates whether Laplace correction should be used to prevent high influence of zero probabilities. Assume that our training set is so large that one is added to each count, which will make a small difference in the estimated likelihoods [170].

- **Details of using the closer classifiers with different schemes**

NB and SVM are applied with different weighting schemes like TF-IDF and BTO. These schemes are often used in information retrieval and text mining [16].

- **Other experiment parameter settings**

1. Cross-validation was utilised to evaluate predictive models by splitting the dataset into 10 parts, nine for training data and one for testing data. The cross-validation process was then repeated 10 times. The model was trained using the training data, and the accuracy was evaluated using the test data. The results of the 10 repetitions were averaged to produce a single evaluation [171] [172].
2. The stratified sampling creates random subsets, which ensures that every subset in the training data will be in the testing data [170].
3. Prune contains two parameters: prune below and prune above. For example, if prune below equals 4 and prune above is 500, it will ignore the documents or the tweets that have words appearing in less than four documents or more than 500 documents. For this experiment, prune was set to equal none to ignore the too-frequent and too-infrequent words [170].

### **3.5.5.5. Experiment for classification into positive and negative classes**

The goal of the experiment was to provide the best accuracy in sentiment analysis for Arabic text in relation to the Saudi stock market for positive and negative classes (see Table 5 and Table 6). Two classifiers used to explore the polarity of given data were NB and SVM. In this stage, the pre-processing was set up using the first four steps mentioned in the 'Process Documents from Data' operator in RapidMiner, and light stemming was examined versus deep stemming. Using a root stemmer or a light stemmer will yield two different accuracy levels, regardless of the weighting schemes. For instance, in the stock market data pre-processing stage, the accuracy was 67.77% with the NB classifier and the light stemmer, while the accuracy with the root stemmer was 44.52%. In addition, with the SVM classifier, the accuracy with the light stemmer was 86.24%, while the accuracy with the root stemmer was 64.60%. For that reason, choosing the right stemmer is an essential step to keep relevant results in text mining. Moreover, the N-gram feature was not involved to show the impact of N-grams in the next phase of this experiment. Table 5 shows the class precision and recall for the SVM and NB classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling.

Table 5: Precision and recall for positive and negative classes without N-gram feature using SVM and NB

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	67.77%	70.69%	69.32%
		TF-IDF	67.70%	69.64%	68.19%
	SVM	BTO	86.24%	83.42%	86.39%
		TF-IDF	86.32%	83.81%	86.19%

After applying the N-gram feature during the pre-processing stage, accuracy improved, especially for the NB classifier. Table 6 shows the class precision and recall for the SVM and NB classifiers with the N-gram feature: Crosse-validation=10, sampling type=stratified sampling.

Table 6: Precision and recall for positive and negative classes with N-gram feature using SVM and NB.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	78.21%	78.62%	77.05%
		TF-IDF	75.97%	75.22%	74.41%
	SVM	BTO	85.86%	82.60%	86.45%
		TF-IDF	85.24%	81.70%	85.95%

To conclude, the analysis shows promising results for the SVM classifier with both BTO and TF-IDF schemes that outperform the NB classifier. As shown in the results from Table 6, the SVM algorithm worked better to classify the targeted documents with the TF-IDF scheme. Moreover, the NB classifier worked with the BTO scheme better than with the TF-IDF scheme. Table 6 shows the classifiers' different results in terms of the class accuracy, recall, and precision when the document was pre-proceed using the N-gram feature. To sum up this stage of the classification experiment, in general, there was a slight drop in accuracy when SVM used the N-gram feature to classify the targeted documents. The best accuracy was achieved by SVM with the BTO schema (85.86%), along with the best recall and precision and the least classification error. Furthermore, there was a big improvement in NB when using the N-gram feature to classify the collected documents with the BTO schema. However, in order to find the performance differences between both classifiers, receiver operating characteristics were utilised.

### 3.5.5.6. Receiver operating characteristics (ROC)

Receiver operating characteristics (ROC) graphs are a utilisable technique to evaluate and visualise the predictive power of a binary classifier. ROC graphs represent the difference of likelihood level between the True Positive Rate (TPR), which is plotted on the y-axis, and the

False Positive Rate (FPR), which is plotted on the x-axis [173]. The best point in the ROC space is in the upper left corner (0, 1). Figures 2,3, and 4 show the ROC curves for the SVM and NB classifiers with different parameters determining how the ROC are evaluated. The visualisation of the various techniques indicates the trade-offs among them. As can be inferred from the graph, NB was not as good as SVM. Using a curved line to divide the graph resulted in SVM outperforming the NB. However, the high selection rate of the false positives offsets the good prediction of the true positives when displayed in ROC space. Neutral selections could arguably be a better option to represent the ROC graph.

- Optimistic: The correctly classified examples are considered before visually examining the erroneous classification.

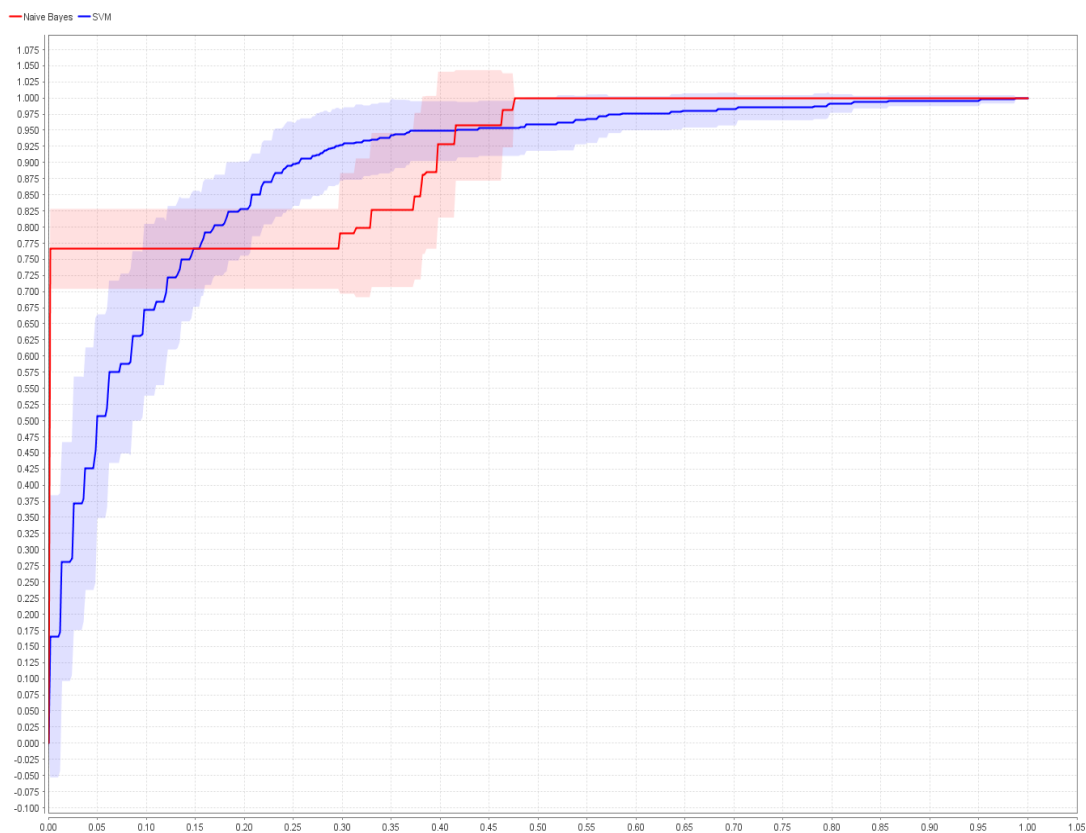


Figure 2: Optimistic NB in red and SVM in blue

- Pessimistic: Erroneous classifications are considered before optically canvassing the correct classification.

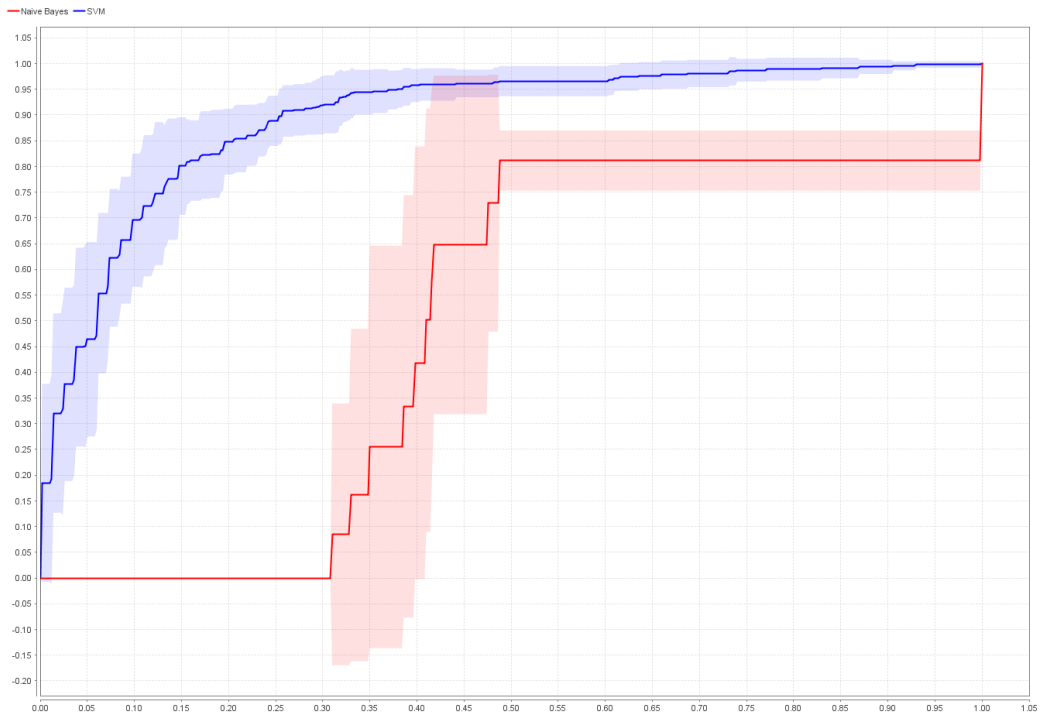


Figure 3: Pessimistic NB in red and SVM in blue

- Neutral: A combination of both optimistic and pessimistic calculation methods. Here, correct true and erroneous classifications are considered alternately.

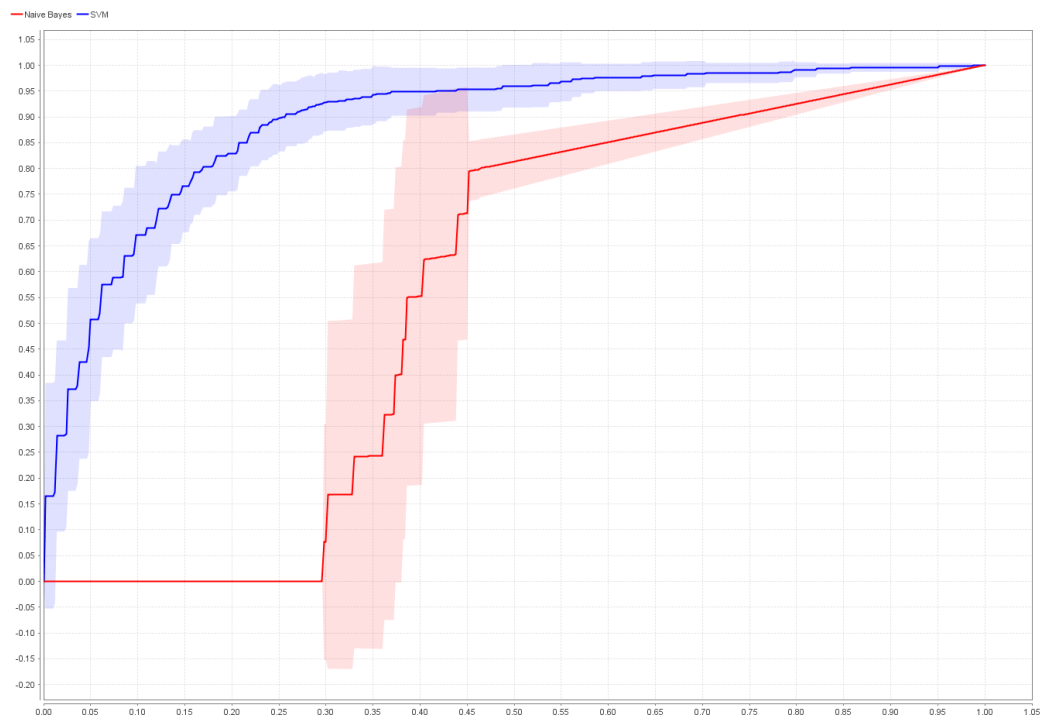


Figure 4: Neutral NB in red and SVM in blue

### 3.5.6. Application in the educational domain

The aim of this experiment was to emphasize that the finding and recommendation obtained in the financial domain in the opinion mining of the Saudi stock market is repeatable in the

educational domain. The same approach, with the same experiment parameter settings, opinion mining process, labelling process, pre-processing documents, and classification method as were used in the application in the financial domain in the Saudi stock market was applied in the educational domain. The Twitter account of the Deanship of e-Learning and Distance Education at King Abdelaziz University in Saudi Arabia was chosen. Table 7 shows the number of tweets in our database: 2000 Arabic tweets were collected, of which 494 tweets were annotated for the training dataset, and 879 irrelevant tweets were deleted from the database.

Table 7: Collection of Tweets for the application in the educational domain.

Total	Positive	Negative
494	144	350

This experiment assigned Arabic text to positive and negative classes within a corpus collection in e-learning as an education domain with different algorithms such as NB and SVM to test the performance of the sentiment classification method, which can help improve the e-learning system process. The result and the finding of this experiment shows that the application on the financial domain in Saudi Stock Market in opinion mining can be repeatable in the educational domain and have the same recommendation and finding. Tables 8 and 9 show that the best accuracy was achieved by SVM without N-gram feature. On the other hand, the best accuracy was achieved by NB when the N-gram feature is involved.

Table 8: Precision and recall for positive and negative classes without N-gram feature using SVM and NB.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	79.74%	77.26%	76.39%
		TF-IDF	79.73%	76.40%	76.11%
	SVM	BTO	82.40%	72.45%	83.71%
		TF-IDF	84.61%	74.19%	89.57%

Table 9: Precision and recall for positive and negative classes with N-gram feature using SVM and NB.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	81.96%	78.40%	78.61%
		TF-IDF	81.75%	77.81%	78.10%
	SVM	BTO	80.78%	68.83%	83.98%
		TF-IDF	83.19%	71.33%	89.77%

### 3.5.7. Application in the service domain

The same approach as was applied in the financial domain in opinion mining of the Saudi stock market was applied in the service domain, including the same experiment parameter settings, opinion mining process, labelling process, pre-processing documents, and classification method. Table 6810 shows the number of tweets in our database; 1,133 tweets were annotated for the training dataset, of which 727 tweets were deleted in the labelling process.

Table 10: Number of tweets.

Total	Positive	Negative
1133	378	755

This experiment assigned Arabic text to positive and negative classes within a corpus collection in the Mubasher product as a service domain with different algorithms such as NB and SVM to test the performance of the sentiment classification method. The result and the finding of this experiment shows that the application on the financial domain in Saudi Stock Market in opinion mining can be repeatable in the service domain and have the same recommendation and finding. Tables 11 and 12 show that the best accuracy was achieved by SVM without N-gram feature. On the other hand, the best accuracy was achieved by NB when the N-gram feature is involved.

Table 11: Precision and recall for positive and negative classes without N-gram feature using SVM and NB.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	83.12%	82.23%	80.83%
		TF-IDF	82.50%	80.81%	80.17%
	SVM	BTO	83.83%	77.26%	85.81%
		TF-IDF	84.90%	79.10%	86.34%

Table 12: Precision and recall for positive and negative classes with N-gram feature using SVM and NB.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	84.19%	81.92%	82.25%
		TF-IDF	83.92%	81.52%	82.00%
	SVM	BTO	83.93%	76.38%	88.20%
		TF-IDF	84.55%	77.19%	89.03%

### **3.6. General recommendation**

SVM classifiers with TF-IDF schemes outperform NB classifiers in general. SVM has outperformed other machine learning techniques because of its primary advantages. For instance, generality of text categorisation dilemmas is linearly separable, and it's robust in high-dimensional spaces and powerful when there is a sparse set of samples and any feature is pertinent [89]. However, if SVM used the N-gram feature to classify the targeted documents, there was a performance drop, as shown in the previous experiment. On the other hand, NB performed its best when using the N-gram feature to classify the collected documents with the BTO schema.

### **3.7. Summary**

In the present chapter, an Arabic sentiment analysis model was designed and implemented to classify documents regarding Saudi stock market opinions through SVM and NB algorithms. Moreover, the results of this study show that text pre-processing is an essential factor in opinion mining classification. Finally, this study also demonstrated applications in two different domains—education and the service industry—of sentiment analysis of positive and negative opinions in Saudi tweets. The results and findings of these two experiments show that opinion mining's application in the financial and educational domains is repeatable in the service domain and has the same recommendation and finding.



## Chapter 4: Techniques for Improving the Labelling Process of Sentiment Analysis

### 4.1. Introduction

As described in Chapter 3, the first step of the proposed sentiment analysis of the Arabic language is the labelling process, which is normally carried out by human operators. The focus of this labelling process is on the classification of sentiment polarity into positive and negative. This could be a challenging task due to the subjectivity factor of sentiment polarity. For example, opinionated text could carry some speech informality, such as sarcasm, subjective language, and emoticons, that makes the opinion ambiguous.

For the classification of sentiment polarity, it is natural to have cases of known knowns (these are things we know that we know), known unknowns (there are things that we know we don't know), unknown knowns (things we don't know we don't know), and unknown unknowns (things we are neither aware of nor understand) [90]. An accurate labelling process normally requires the human operators to have a deeper understanding of a subjective than an objective text [33]. This deeper understanding would expand the territory of known knowns and thus improve the quality of the labelling process. However, this kind of knowledge might not always be available. In this chapter, techniques for handling cases of known unknowns, unknown knowns, and unknown unknowns will be explored to assist human operators with the labelling process.

In the case of known unknowns, e.g., one tweet might contain both positive and negative keywords, which confuses the human, causing the labelling process to be unreliable. Adding a neutral class would give the human more flexibility to express his or her uncertainty. In addition, the neutral class would improve the training and classification process by enabling better coverage of the vector space constructed by the training dataset. In this chapter, the effectiveness of deploying the neutral class during the labelling process will be explored.

Due to the cases of unknown unknowns and unknown knowns, human-based labelling generates different kinds of mistakes within the labelling process. Thus, the inaccurate classification is created by the operators unintentionally. This kind of inaccuracy could be reflected as noises during the classification process in a later stage of sentiment analysis. To tackle the problem, this chapter introduces a relabelling process for Arabic tweets to identify the noise generated by labelling. The main goal of this relabelling process is to identify the potential cases of unknown unknowns and unknown knowns via the noises visualised during the classification. This will lead to updating experts' knowledge about the domain and may

lead to improved classification. It has been noted that labelled texts usually contain noise, especially comments that are relatively long and consist of multiple sentences, such as blogs [174]. The proposed relabelling approach will not eliminate the classification noise but will help operators identify possible contradictions in their labelling and reveal cases of unknown unknowns and unknown knowns.

This chapter presents techniques for improving the labelling process of sentiment analysis. Section 4.2 focuses on the effectiveness of a labelling process with a neutral class. Section 4.3 shows the need to improve the labelling process with relabelling and demonstrates the process of relabelling. The final section will constitute the conclusion and recommendations for further work in this area.

## **4.2. Improving the labelling process with a neutral class**

In contrast to positive or negative classes, researchers used to ignore the neutral class under the hypothesis that there is less sentiment in neutral texts—in other words, because the insufficient amount of neutral training data leads to poor performance, and they consider the neutral class generally to mean no opinion [175] [28]. However, The neutral class is useful for real-life applications due to the nature of sentiment, since sentiment is sometimes neutral, and excluding it forces instances into other classes (positive or negative) [176]–[20].

In this research, we believe that the neutral class needs more investigation for the language used to extract the sentiment and domain. In addition, sentiment polarity might contain positive and negative keywords that can make the labelling process unreliable. This occurred frequently in the neutral class, where one tweet might contain both positive and negative keywords. Labelling carried out by hand can cause human mislabelling of sentiments. Therefore, adding the neutral class can give humans more flexibility and options in the labelling process. However, this might cause less accurate results, since the data dictionary, which consists of all vectors that belong to positive, negative, and neutral, becomes larger.

## **4.3. Experiment: Classification into positive, negative, and neutral classes**

Arabic sentiment analysis classification was applied to only positive and negative classes in Chapter 3. To evaluate the effectiveness of introducing the neutral class into the labelling process, a new experiment will be carried out with the neutral class, with and without the N-gram feature. The rest of the configuration will be the same as in Chapter 3. The classifiers utilised to explore the polarity of all classes' data were NB and SVM. Two different weighting schemes (TF-IDF and BTO) were used for all classes (positive, negative, and neutral). Table 13 shows the number of tweets in the targeted dataset.

Table 13: Collection of tweets about the Saudi stock market with neutral class.

Positive	Negative	Neutral	Total
828	466	649	1943

Tables 14 and 15 show a comparison between the classifiers with the neutral class in terms of the class accuracy, recall, and precision. In addition, Tables 14 and 15 show that SVM with TF-IDF without the N-gram feature works better for classifying the targeted documents when the neutral class is added. However, NB works better with the N-gram feature to classify the same documents.

Table 14: Precision and recall for positive, negative and neutral classes without N-gram with SVM and NB.

Classes	Classifier name	Weighting schemes	Accuracy	Class recall	Class precision
All classes	NB	BTO	62.63%	64.27%	63.83%
		TF-IDF	64.13%	65.68%	64.25%
	SVM	BTO	84.04%	82.29%	85.27%
		TF-IDF	84.71%	83.54%	85.63%

Table 15: Precision and recall for positive, negative and neutral classes using N-gram with SVM and NB.

Classes	Classifier name	Weighting schemes	Accuracy	Class recall	Class precision
All classes	NB	BTO	76.27%	75.94%	75.82%
		TF-IDF	74.88%	73.91%	74.22%
	SVM	BTO	83.94%	82.10%	85.44%
		TF-IDF	84.25%	82.15%	86.46%

In conclusion, the best accuracy was achieved by SVM with TF-IDF, and it was equal to 84.71%; this combination also had the best recall and the least classification error but not precision. Moreover, the analysis shows almost identical results for SVM with both schemas with and without N-gram and only slightly different results between recall and precision. Table 16 shows a comparison between the classifiers in terms of the class accuracy, recall, and precision. The main indication from Table 16 is that the best accuracy result occurred with SVM using the TF-IDF schema without the N-gram feature.

Table 16: Precision and recall for positive, negative, and neutral classes with SVM and NB.

Classifier	Accuracy	Recall	Precision
NB using BTO schema with N-gram feature	76.27%	75.94%	75.82%
SVM using TF-IDF schema without N-gram feature	84.71%	83.54%	85.63%

- Lift chart

The lift chart is easier to read and clearer than the ROC chart; the lift chart is used to represent and evaluate classifier performance using the TF-IDF schema without the N-gram feature. The number of bins setup up in advanced by RapidMiner which set to 10 bins. Then, in order to target neutral class, split data operator used to produce the desired number of subset of the given data, split operator will partition the data set into different random parts for training data and it was setup by 70 % and testing data by 30%. For example, in stock market experiment, the total data size was 1944 records and by multiplying that with 70% and dived it by ten which the number of bins, the result will be 136 records for 10 parts. Crosse validation ten-fold was implemented one time for each part to determine the confidence value for neutral documents and their count in each part. Figure 5 shows an illustration of the performance of the neutral class as a target class for the lift chart, using the SVM classifier with TF-IDF and the N-gram feature. The x-axis of the chart clarifies the percentage of confidence values for each prediction. The y-axis of the chart demonstrates the percentage of the test dataset that accumulated to represent the performance predictions. The highest-confidence numbers are shown first, after which they decrease. An obvious pattern is that classifier pay attention to the neutral class.

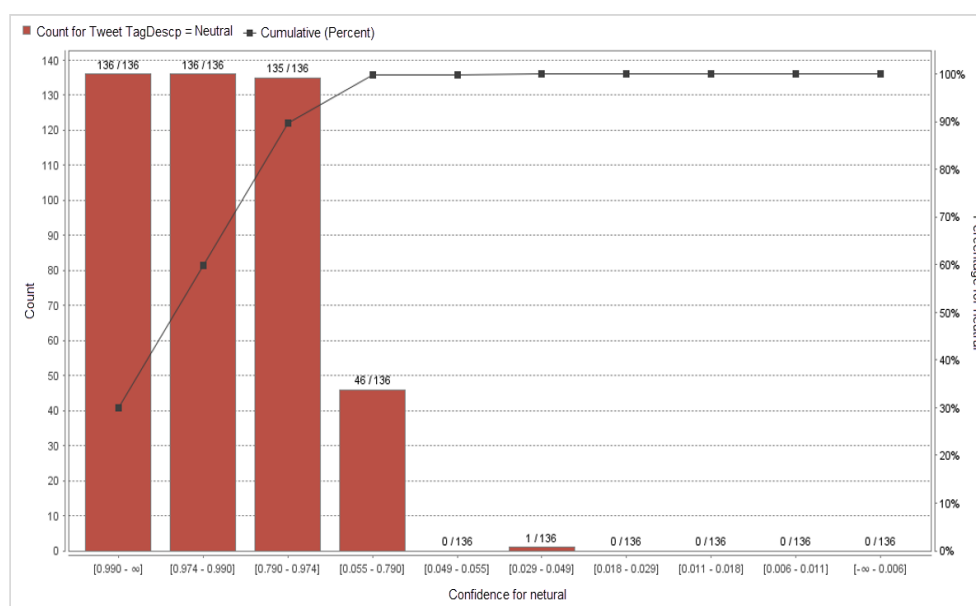


Figure 5: Lift chart for neutral class for SVM using TF-IDF schema without N-gram feature

A one-to-one model was built to show the relation between labelling by human operators for positive, negative, and neutral classes and the Tadawul All Share Index (TASI)<sup>5</sup>, with accuracy equal to 84.71 for SVM using the TF-IDF schema without the N-gram feature, and with accuracy equal to 76.32 for NB using the BTO schema with the N-gram feature. The built model illustrates results in labelling by showing the positive, negative, and neutral opinions as well as the TASI closing values. The Mubasher platform was used to download the closing values of the Saudi stock market from the TASI during the data collection period, except for the weekends (because the stock market is closed at the weekend) (see Appendix A). Figure 6 shows the relation between labelling by human operators and the TASI for the Saudi stock market for positive negative, and neutral classes.

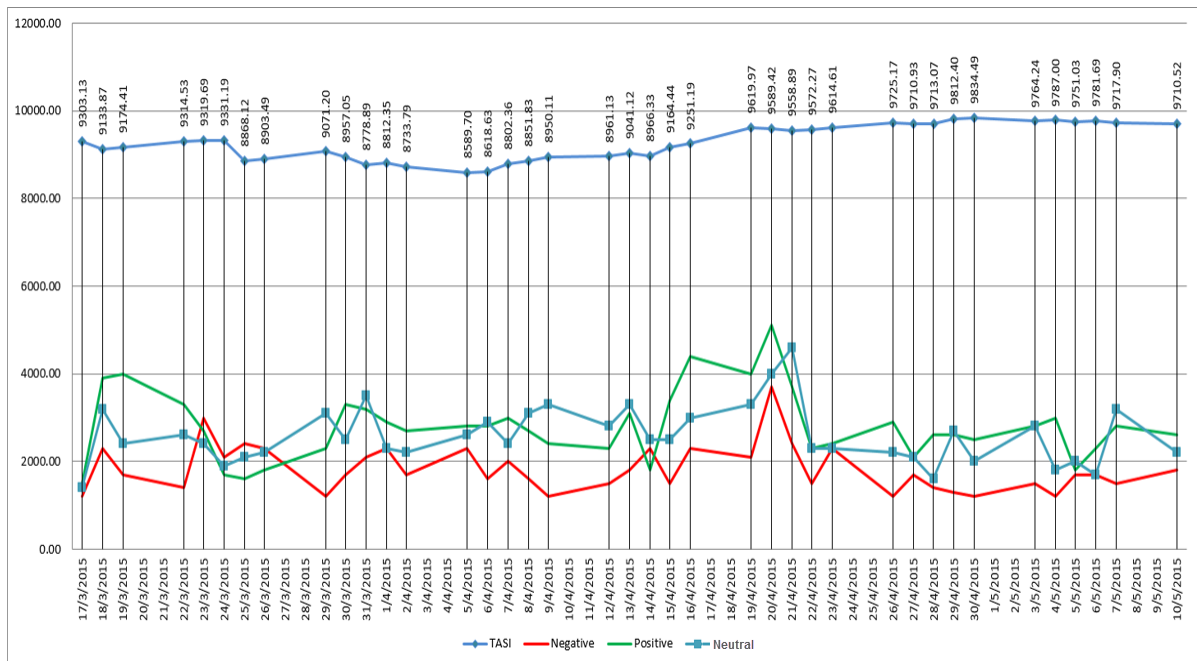


Figure 6: The relation between labelling by human operators and the TASI for the Saudi stock market for positive, negative, and neutral classes.

It can clearly be seen that the positive, negative, and neutral classes rise and fall with each other over the time. The neutral class, even considered to have no opinion, play a significant role in term of categorisation. The greatest score for neutral classes occurred in 21/4/2015, and the lowest neutral class score occurred on 28/4/2015; whereas the lowest negative class score occurred in 29/3/2015. Turning to details, one time only the neutral class was lower than the negative class that occurred for four days between 23/3/2015 to 26/3/2015. On the contrary, many times the neutral class went above the positive class for 15 days, but TASI remain the same for 12 days. However, research tends to ignore the neutral class because it

<sup>5</sup> <https://www.tadawul.com.sa>

reduces the accuracy. It can be demonstrated by the experiment here that, the accuracy was not reduced significantly with the introduction of the neutral class. In Chapter 3, the accuracy achieved for positive and negative classes with N-gram using SVM with weighting schema TF-IDF was 85.24, while when the neutral class was added, the accuracy became 84.25. In contrast, adding the neutral class would simplify the labelling process of the human operator and improve the training and classification process by enabling better coverage of the vector space constructed by the training dataset. Therefore, improving the labelling process by the proposed relabelling will lead to updating experts' knowledge about the domain and may be considered as a preferred approach.

#### **4.4. Improve the labelling process with relabelling**

High dimensionality in texts makes text pre-processing significant in text classification problems including sentiment analysis [177, 178]. This problem slightly increases once the dimensionality becomes higher, such as when adding the neutral class data. For example, in the previous experiment to improve the labelling process with a neutral class, there was approximately 17% misclassification when SVM was used and approximately 28% misclassification when NB was used to classify the documents. In addition, manually labelling the documents may cause some mistakes, even when the neutral class is added. Thus, inaccurate classification creates the need to construct a relabelling process for Arabic tweets to identify contradictions based on the noise in the original labelling. Therefore, a relabelling process for Arabic sentiment analysis is proposed as follows.

##### **4.4.1. Process of relabelling**

Figure 7 demonstrates the relabelling process for Arabic sentiment analysis. The process starts with the same step as the Arabic sentiment analysis process in Chapter 3, by collecting the corpus of data to label and save the relevant training data and discard the irrelevant ones. The second step is pre-processing the data by cleaning up the hashtags, duplicate tweets, retweets, URLs, and special characters and preparing the collected data for the labelling process. The next step is labelling the cleaned data as positive, negative, or neutral by the expert in the domain. After that, the relabelling process starts, which consist of several steps: wordlist process, filtering, visualisation, extracting rules, and relabelling.

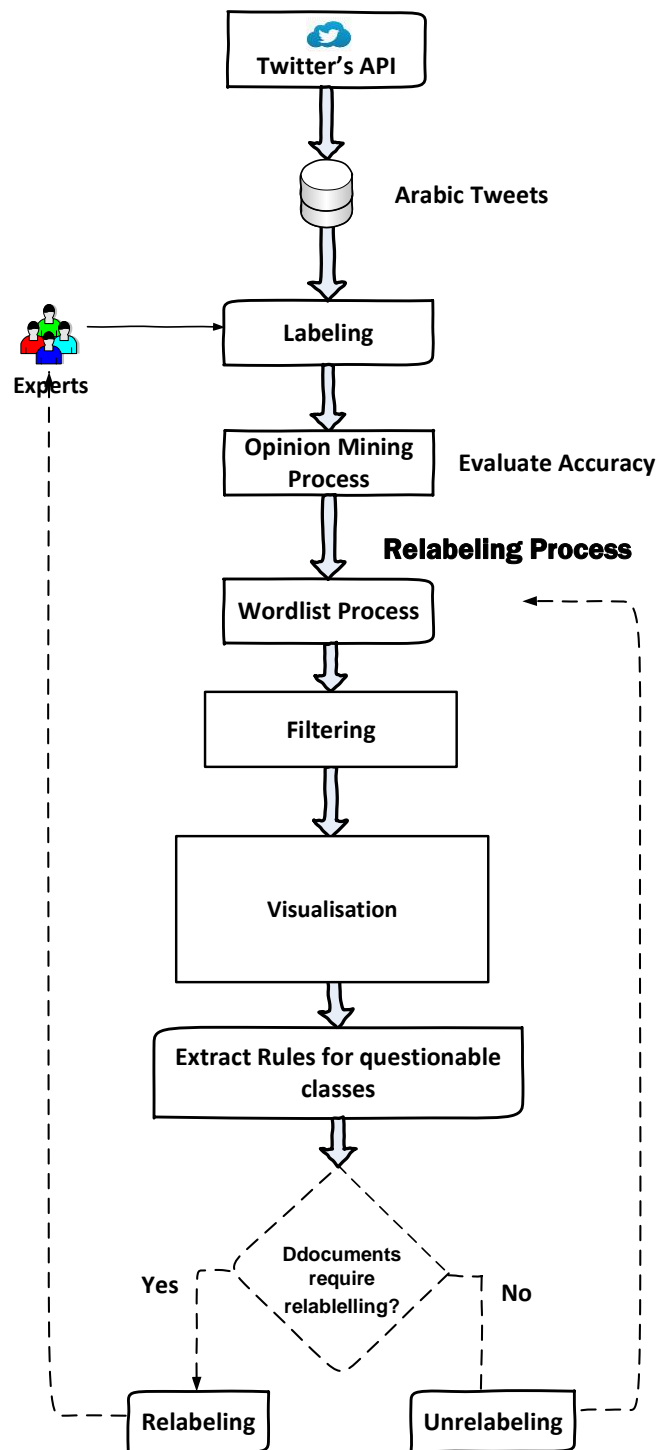


Figure 7: Relabelling process.

The most challenging part of the process is feature selection, since any feature can occur in all classes (positive, negative, and neutral). In addition, the main difficulty is to find out how many times the feature occurred in each class. Therefore, the wordlist technique is utilised to represent the text data as a matrix to show the frequency of occurrence of each term within the three classes. Next, the features are filtered to select the higher frequency features presented by the wordlist process. Then, in order to understand the sentences' structures and

the sentiments behind them, the visualisation technique is utilised on all data within the corpus. In other words, visualisation of the text shows the vital importance of the correlation between terms involved in the textual contents in general. However, visualisation only shows the features with all the related terms in their textual contents without showing the classes that belong too. Using the wordlist technique with visualisation can produce the important features created by the wordlist technique during the pre-processing stage. Overcoming the limitation of visualisation for the important features in the targeted text is an essential step in the relabelling process. After that, association rules are extracted from the documents that have features in questionable classes. Association rules are generated regardless of the minimum support and minimum confidence thresholds using the visualisation technique for the features that belong to the questionable class. By following these processes, the documents that have features in questionable classes can be relabelled again, and the noise of the original labelling will be reduced.

#### 4.4.2. Wordlist process

Figure 8 shows the wordlist process. In this phase, the same corpus was classified whether positive, negative, or neutral was used, and the same data pre-processing procedure was utilised as in the opinion mining process. The goal of visualising association rules as a wordlist is to have a dataset that contains a row for each word, attributes for the word itself, and the number of labelled documents in each class for each term or word in the training data [170]. In other words, it represents the text data as a matrix to show the frequency of occurrence of each term within the three classes. The key feature in this process was the N-gram, which represents the correlation between the feature selection and other terms with their frequency of occurrence for just two nodes within all classes' data.

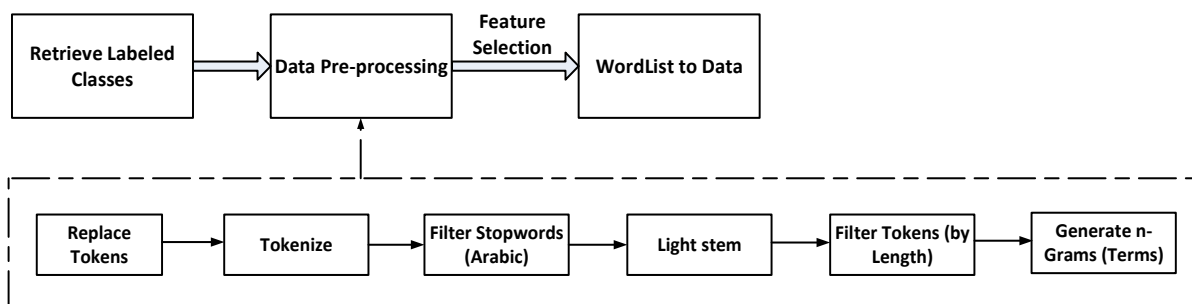


Figure 8: Wordlist process.

#### 4.4.3. Filtering

Features in the context of opinion mining are the words, phrases, or terms that strongly express an opinion as one of three polarities: positive, negative, or neutral. In other words, features are the keywords for deciding whether a text is positive or negative. That means that



features have a higher impact on the orientation of a text than other words in the same texts. The impact of feature selection is to help to reduce the dimensionality of a text to increase the classification accuracy. Features in the text are considered as explicit or implicit. Features that appear in a text are explicit, whereas features that do not appear but it are implied are implicit [35]. In the proposed process, only the explicit features are considered.

#### **4.4.4. Visualisation**

The importance of visualising a text is that it helps to understand the sentences' structure and the sentiments behind them. Visualising the text shows the vital importance of the correlations between terms involved in the textual contents. The first step of the visualisation technique is producing the important features created by the wordlist process. Then, one of the features that appears in the dictionary created by the wordlist is selected. Selection of the feature is random to cover high-, average-, and low-frequency features to prove that the relabelling can work with different frequency levels. The next step is visualising the selected feature in all classes' data as a wordlist representation. The wordlist shows how frequently the selected feature occurs in the positive, negative, and neutral classes. If the selected feature has a positive sentiment and occurs in other classes, such as neutral or negative, then the other classes (neutral, negative) are considered questionable. In other words, if the selected feature is from the positive list, then it should only occur in the positive class; otherwise, this feature's incidence in different classes leads them to be considered questionable, since a strong, positive keyword should classify the text to as positive unless there is negation. Besides, it should not occur in the neutral class unless there are other words in the text that affect the sentiment. However, features in a questionable class need further investigation to confirm the correctness of the labelling.

Figure 9 illustrates the visualisation association rules process. In this phase, again the same corpus classified as either positive, negative, or neutral was used, and the same data pre-processing procedure utilised in the opinion mining process was carried out. After that, FP-Growth was used to discover frequent items above the minimum support and minimum confidence thresholds. Then, association rules were generated to expose the relationships between seemingly unrelated data. The output of visualisation is the association of the high-frequency terms with selected features presented previously from the wordlist process.

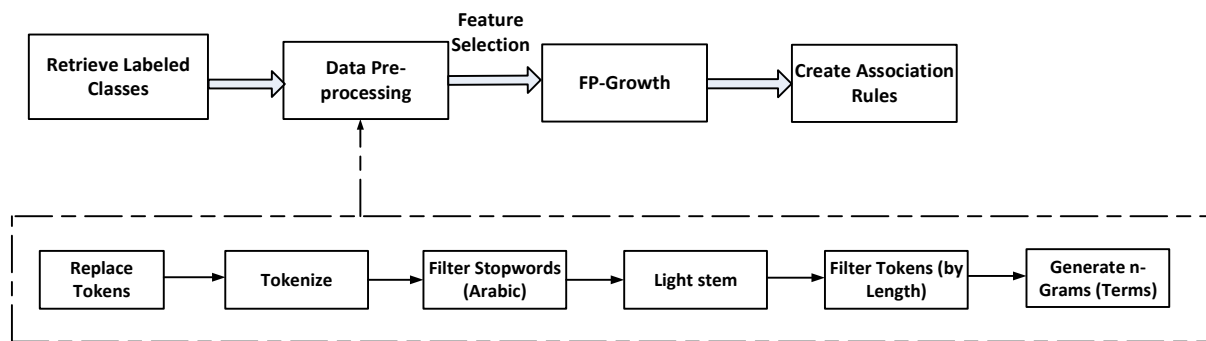


Figure 9: The visualisation generated association rules process.

#### 4.4.5. Association rules

One of the most essential and popularly researched techniques of data mining is association rule mining, which was first introduced in [179]. Association rules are widely utilised in several areas, for example, telecommunication networks, markets, risk management, and inventory control [180]. Association rules aim to extract interesting correlations, frequent patterns, associations, or casual structures among sets of items in transaction databases or other data repositories.

Association rules are if/then statements that help to discover relationships between unrelated data. An association rule has two components, an antecedent (if) and a consequent (then). An antecedent is an item set found in the dataset, and a consequent is an item set found in incorporation with the antecedent. In other words, an association rule is an implicative insinuation of the form  $A \Rightarrow B$ , where  $A$  and  $B$  are frequent item sets in a transaction database and  $A \cap B = \emptyset$ . In practical usage, the rule is  $A \Rightarrow B$  [181].

Lopes et al. expounded the quandary of mining association rules from text [182]. They commenced by representing the text as a bag of words: Let  $I = i_{(1)}, i_{(2)} \dots, i_{(m)}$ , and let  $D$  be a set of transactions, where each transaction  $T$  is a set of items that represent the document, so  $T \subseteq I$ . An association rule is an involvement of the form  $X \Rightarrow Y$ , where  $X \subset I$  and  $Y \subset I$ , and  $X \cap Y = \emptyset$ . The rule  $X \Rightarrow Y$  holds confidence if the document  $D$  contains  $X$  and  $Y$  and support if the document contains  $X \cup Y$ . The left of the rule is the head of the rule and the set of residual words is the rule's body.

##### 4.4.5.1. Association rules methods

Text mining is the knowledge extracted from textual databases. However, association rules are produced by analysing data for frequent if/then patterns and utilising the criteria for support and confidence to identify the most consequential relationships. Support means how frequently the items appear in the database, and confidence indicates the number of times the if/then representations have been found to be true. The frequent if/then patterns are mined

utilising methods such as the Apriori algorithm and the FP-Growth algorithm [21], [183]. Association rule mining is divided into two phases. In the first phase, frequent patterns are mined above the minimum support threshold. In the second one, association rules are generated based on the minimum confidence threshold [184]. Experts in data mining claim that in the dataset, some terms or words frequently appear, while others rarely occur. In this case, the values of the minimum support will control the rule detection. For example, if the minimum support value is set low, rules that frequently occur will appear. In contrast, if the minimum support value is set high, rules that infrequently occur will not appear. This causes a problem called a “rare item”, where good rules with high confidence might be ignored because they have very little support [185, 186].

#### **4.4.5.2. Apriori algorithm**

The Apriori algorithm is one of the earliest algorithms for discovering association rules from data. It was proposed by Agrawal in 1993 [179]. The process of the Apriori algorithm starts with finding all frequent one-item sets that pass the minimum support threshold. The next step is to generate a potential candidate item set of size two from the outcome of the previous step (largest item or most frequent item). Apriori uses “downward closure” to improve the searching process by reducing the size of the candidate item set list during each iteration [80, 186].

#### **4.4.5.3. Frequent-pattern tree algorithm**

The Frequent-Pattern Tree Algorithm (FP-growth) is an important algorithm in the data-mining field that does not require candidate generation. FP-growth works by storing relevant item set information and then allowing an efficient novel structure to determine the frequent item sets. FP-growth has a way of decomposing the mining process into small tasks on a conditional FP-tree. First, it considers the dataset to find the frequent items at level 1 by computing the support for frequent items. Those frequent one-item sets are stored in descending order of their supports. Then, the dataset is scanned again to build an FP-tree using the head table with a null label root. The database scanning process continues for each transaction T to re-sort the frequent items in the header table according to their frequencies of occurrence and insert them into the FP-tree [80]. FP-growth has been applied successfully in several sentiment classification method mining tasks. However, for this study, the FP-growth method was used due to its advantage over the Apriori method—for instance, only two passes over the dataset, no candidate generation, and compresses\_data-set [187]. Figure 10 shows how to discover frequent item sets by the FP-growth algorithm.

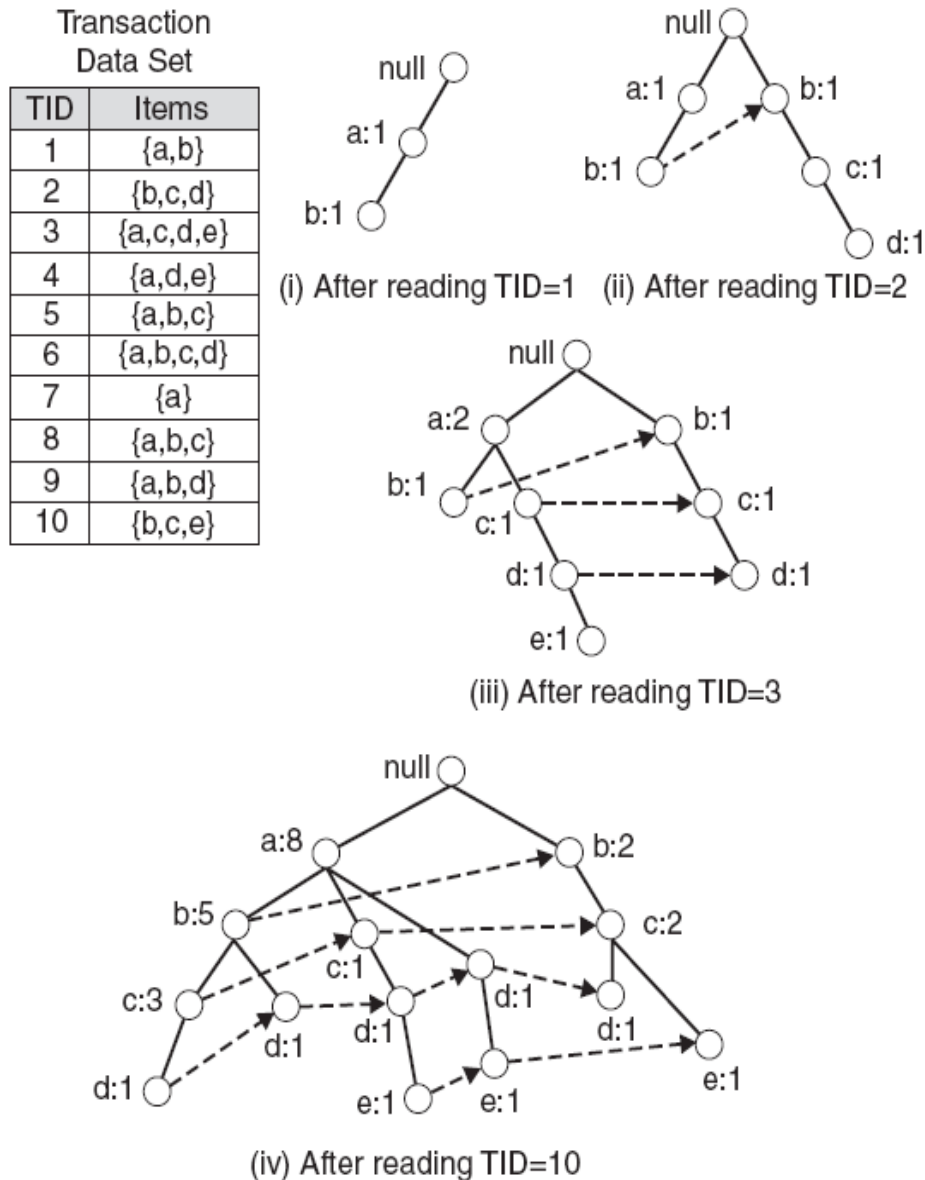


Figure 10: FP-tree construction (taken from [24]).

#### 4.4.6. Extracting rules

In this phase, association rules are extracted from the collection of documents based on the features created by the wordlist. Association rules are generated about the minimum support and minimum confidence thresholds using the previous process of visualisation; the only difference here is that the data we are going to use belong to the questionable class. This step focuses on extracting the rule that occurs less frequently in the questionable class within a specific document.

#### 4.4.7. Relabelling

In this step, the training data are searched for a sentiment that contains less frequently occurring terms in other classes that the feature does not belong to according to the wordlist matrix. To ensure reliability, the relabelling is performed by an expert on a specific document. Then, the sentiment with labelling noise is sent as a recommendation to an expert to check its labelling.

#### 4.4.8. Experiment: Relabelling

The relabelling process starts by representing the text data as a matrix to show the frequency of occurrence of each term within the three classes. The relabelling process focuses on representing the correlation between feature selection and other highly frequent terms for just two nodes within all classes' data. Table 17 shows the feature “earnings” (ارباح) as a positive sentiment in the Saudi stock market domain. In addition, Table 17 shows the occurrence of the feature “earnings” (ارباح) in the positive, negative, and neutral classes.

Table 17: Occurrence of the feature “Earnings” (أرباح).

Feature	Occurrence	Neutral	Positive	Negative
ارباح	304	14	223	67

Figure 11 shows the association rules related to the feature “earnings” (ارباح) in all classes with respect to the minimum support and minimum confidence thresholds. The feature “earnings” (ارباح) entailed sharing the profits of firms in the Saudi stock market. In addition, Figure 11 shows the most important rules for the feature “earnings” (ارباح), which are [ارباح] → [توزيع] (support: 0.031 confidence: 1), [ارباح] → [ارتفاع] (support: 0.017 confidence: 1), and [ارباح] → [تراجع] (support: 0.013 confidence: 1). The term “sharing out” [توزيع] correlated with the feature “earnings” [ارباح] to compose positive phrases about distributing profits in the sentence. Further, the term “rising” [ارتفاع] correlated with the feature “earnings” [ارباح] to compose positive phrases about rising profits in the sentence. On the other hand, the term “decline” [تراجع] correlated with the feature “earnings” [ارباح] to compose negative phrases about profit decline in the sentence.

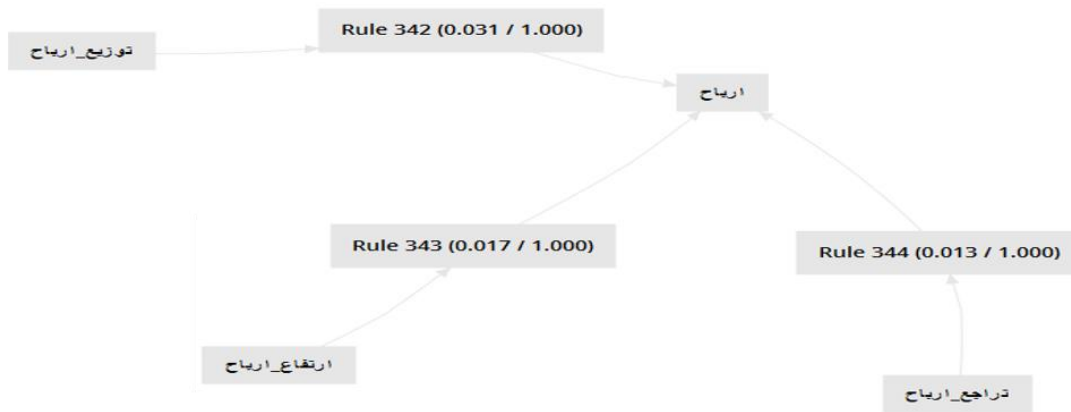


Figure 11: Visualisation of the association rules for the feature “earnings” (ارباح).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases correlated with the feature “earnings” (ارباح). Table 18 shows that the phrase “rising profits” [ارتفاع-ارباح] occurred 79 times: three in the neutral, 66 in the positive, and 10 in the negative class.

Table 18: Phrases for the feature “earnings” (ارباح) in all the data.

Phrase - Terms	Occurrence	Neutral	Positive	Negative
توزيع-ارباح	79	3	66	10
ارتفاع-ارباح	37	0	37	0
تراجع-ارباح	31	0	0	31

From Table 18, the phrase “rising profits” [ارتفاع-ارباح] does not need further investigation since it only occurred in the positive class. In addition, the phrase “profits decline” [تراجع-ارباح] does not need further investigation since it only occurred in the negative class. However, since the phrase “distribute profits” [توزيع-ارباح] occurred in both the negative and neutral classes, both classes become questionable. Therefore, the feature “earnings” (ارباح) needs further investigation in order to find out the association rules in the both classes. As result, two scenarios will be followed: The first scenario is extracting association rules for the feature “earnings” (ارباح) in the neutral class. The second scenario is extracting association rules for the feature “earnings” (ارباح) in the negative class. The association rules are generated with regard to the minimum support and minimum confidence thresholds using the previous process of visualisation.

Figure 12 shows that the feature “earnings” (ارباح) can occur with many rules that appear in the premises column with the minimum support and minimum confidence values. However, according to the first scenario, the interested rule here is [ارباح] → [توزيع-ارباح] (support: 0.005

confidence: 1), which represents the phrase “distribute profits” [توزيع-ارباح] illustrated in the wordlist matrix in the neutral class.

No.	Premises	Conclusion	Support	Confidence
1898	شركات	ارباح	0.003	0.667
3203	توزيع	ارباح	0.005	0.750
3893	بالرئيس الاول	ارباح	0.008	0.833
195...	توقعات	ارباح	0.008	1
195...	مصرف_ارباح	ارباح	0.005	1
195...	توزيع_ارباح	ارباح	0.005	1
195...	المادر_الذي	ارباح	0.005	1
195...	وطريقة	ارباح	0.003	1
195...	التكنولوجيا_طريقة	ارباح	0.003	1
195...	طريقة_مصرف	ارباح	0.003	1
195...	طريقة	ارباح	0.003	1
195...	سابقه	ارباح	0.003	1
195...	تاريخ_وطريقة	ارباح	0.003	1
195...	النصف_الثاني	ارباح	0.003	1
195...	النصف	ارباح	0.003	1
195...	الداية_ارباح	ارباح	0.003	1
195...	ارباح_شركة	ارباح	0.003	1
195...	ارباح_سابقه	ارباح	0.003	1
195...	ارباح_المساهمين	ارباح	0.003	1
195...	ارباح_الربع	ارباح	0.003	1
195...	اهدات_ارباح	ارباح	0.003	1
195...	اهدات	ارباح	0.003	1

Figure 12: The correlation rules of the feature “earnings” (ارباح) in the neutral class.

According to Figure 12, the rule [ارباح] → [توزيع-ارباح] (support: 0.005 confidence: 1), which represents the phrase “distribute profits”, occurred in the neutral class. Therefore, the next step is to search for the phrase “distribute profits” [توزيع-ارباح] in the neutral class documents. Table 19 shows that the phrase “distribute profits” [توزيع-ارباح] occurred in three documents.

Table 19: The term “distribute profits” [توزيع-ارباح] in the neutral class documents.

Original labelling class	Original Arabic tweets with English translation	New labelling class
Neutral	تعلن شركة اتحاد مصانع الأسلاك أسلاك عن تاريخ وطريقة توزيع أرباح النصف الثاني من عام 2014	Positive
	Union Wire Mills Company <u>announces</u> the date and method of <u>distributing dividends</u> for the second half of 2014	
Neutral	تعلن شركة مكة للإنشاء والتعمير عن توزيع أرباح على المساهمين عن السنة المالية المنتهية في 1436/4/30 هـ	Positive
	Makkah Construction & Development Company <u>announces</u> the <u>distribution of dividends</u> to shareholders for the fiscal year ended 30/4/1436 H.	
Neutral	تعلن شركة فواز عبدالعزيز الحكير وشركاه عن توزيع أرباح على المساهمين عن النصف الثاني من العام 2014	Positive
	Fawaz Abdul-Aziz Al Hokair & Co. <u>announces a dividend distribution</u> for the second half of 2014	

Figure 13 shows that the feature “earnings” (ارباح) can occur with many rules that appear in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [ارباح] → [توزيع-ارباح] (support: 0.021 confidence: 1), which represents the phrase “distribute profits” [توزيع-ارباح] illustrated in the wordlist matrix in the negative class.

Show rules matching

all of these conclusions: ▼

- تراجع
- نسبه
- الأول
- ارباح
- بالربع
- بالربع الأول
- شركة
- تعلن
- خسائر
- السوق
- عدم
- السعودي
- صعودية
- تراجع
- القطر
- السوق\_السعودي
- تأجيل
- العمامة
- عدم\_لحفظ
- الحقد
- السعودية
- تأجيل\_صعودية
- الجمعية\_العمامة
- الجمعية
- مدالية
- العمامة\_العمامة
- العمامة
- لقتنين

No.	Premises	Conclusion	Support	Confidence
147	نيكي	ارباح	0.004	0.667
148	مجلس_دارة	ارباح	0.004	0.667
149	سافولا	ارباح	0.004	0.667
150	المركزية_البريد	ارباح	0.004	0.667
151	اصحاء	ارباح	0.004	0.667
152	لمعتت_تويته	ارباح	0.004	0.667
153	استرا_الصناعية	ارباح	0.004	0.667
154	استرا	ارباح	0.004	0.667
854	تراجع_نسبه	ارباح	0.013	0.750
855	بتراجع	ارباح	0.006	0.750
856	دارة	ارباح	0.006	0.750
972	مجلس	ارباح	0.009	0.800
973	للموافقة	ارباح	0.009	0.800
1014	الأول_العام	ارباح	0.011	0.833
2234	تراجع_ارباح	ارباح	0.054	1
2235	توزيع_ارباح	ارباح	0.021	1
2236	توزيع	ارباح	0.021	1

Figure 13: The correlation rules of the feature “earnings” (ارباح) in the negative class.

According to Figure 13, the rule [ارباح] → [توزيع-ارباح] (support: 0.021 confidence: 1), which represents the phrase “distribute profits”, occurred in the negative class. Therefore, the next step is to search for the phrase “distribute profits” [توزيع-ارباح] in the negative class documents. Table 20 shows that the phrase “distribute profits” [توزيع-ارباح] happened in 10 documents. In addition, ten documents has been found to satisfy the rule [ارباح] → [توزيع-ارباح]; therefore, it was resent to the expert who labelled the document in the first stage. Moreover, it can be seen from the structure of nine documents that the phrase “distribute profits” توزيع أرباح has occurred with negation [توزيع أرباح] → [عدم], so the right place for this term is the negative class. Moreover, only one document has been relabelled from negative to positive.



Table 20: The term “distribute profits” [توزيع-ارباح] in the negative class documents.

Original labelling class	Original Arabic tweets with English translation	New labelling class
Negative	تأجيل عمومية فيبكو المتضمنة الموافقة على توزيع أرباح Vipco generality agreed to <u>postpone</u> the <u>dividend distribution</u>	Positive
Negative	مجلس إدارة العالمية للتأمين يوصي بعدم توزيع أرباح Global Insurance Board recommend <u>not to distribute dividend</u>	Negative
Negative	21 مايو عمومية الدرع العربي للموافقة على عدم توزيع أرباح May 21 general Arabian Shield for approval on <u>not distributing dividend</u>	Negative
Negative	غداً التصويت عن بُعد على بنود عمومية كيمانول والمتضمنة عدم توزيع أرباح Tomorrow remote voting on the general terms of Kimanol, including on <u>not distributing dividend</u>	Negative
Negative	ايس للتأمين تقر عدم توزيع أرباح وتنتخب أعضاء مجلس إدارتها ACE Insurance recognises <u>not distributing dividend</u> and elects its board of directors	Negative
Negative	تعلن شركة اتحاد إتصالات موبايلي عن توصية مجلس الإدارة بعدم توزيع أرباح عن الربع الأول من العام المالي 2015م Etihad Etisalat announces the recommendation of the Board of Directors to <u>not distribute dividend</u> for the first quarter of the fiscal year 2015	Negative
Negative	إدارة موبايلي تقر عدم توزيع أرباح عن الربع الأول من هذا العام Mobily's management confirms Directors to <u>not distribute dividend</u> for the first quarter of this year	Negative
Negative	9 يونيو عمومية موبايلي للموافقة على عدم توزيع أرباح June 9 Mobily's approval for <u>not distributing dividend</u>	Negative
Negative	609 % خسائر المتراكمة حتى 30 أبريل 9 يونيو عمومية للموافقة على عدم توزيع أرباح 609% accumulated losses until 30 April 9 June general to approve <u>not distributing dividend</u>	Negative
Negative	شركه صافولا توصي 22 يونيو عمومية للموافقة على عدم توزيع أرباح Savola recommends a June 22 general meeting to approve <u>not distributing dividend</u>	Negative

Figure 14 shows the correlation rules that can happen with the feature “earnings” (ارباح) in the negative class. Moreover, this process can identify the negation terms such as عدم, which means the opposite of positive, to help solve the negation problem in Arabic sentiment analysis.

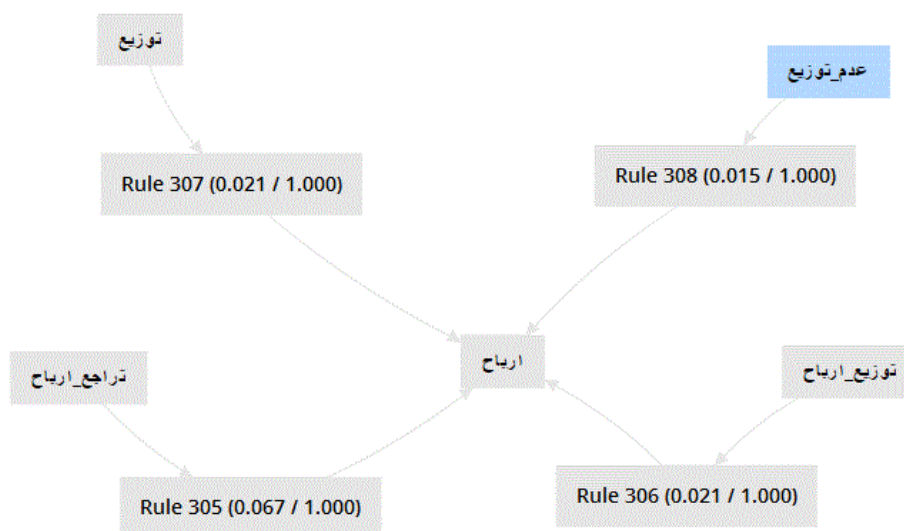


Figure 14: The correlations of the “negation term” (عدم) and the feature “earnings” (ارباح) in the negative class.

The second example verifies our hypothesis with another positive sentiment, “rise” (ارتفاع). Table 21 shows the feature “rise” (ارتفاع) as a positive sentiment in the Saudi stock market domain. In addition, Table 21 shows the occurrence of the feature “rise” (ارتفاع) in the positive, negative, and neutral classes.

Table 21: Feature “rise” (ارتفاع) occurrence.

Feature	Occurrence	Neutral	Positive	Negative
ارتفاع	163	3	113	47

Figure 15 shows the association rules related to the feature “rise” (ارتفاع) in all classes with respect to the minimum support and minimum confidence thresholds. The feature “rise” (ارتفاع) refers to obtaining a financial advantage or benefit from an investment of some company in the Saudi stock market. In addition, Figure 15 shows the most important rules for the feature “rise” (ارتفاع), which are [ارباح] → [ارتفاع] (support: 0.091 confidence: 1) and [نسبه] → [ارتفاع] (support: 0.013 confidence: 1). The term “earnings” [ارباح] correlated with the feature “rise” (ارتفاع) to compose positive phrases about rising profits in the sentence. Further, the term “percentage” [نسبه] correlated with the feature “rise” (ارتفاع) to compose positive phrases about high ratios in the sentence.

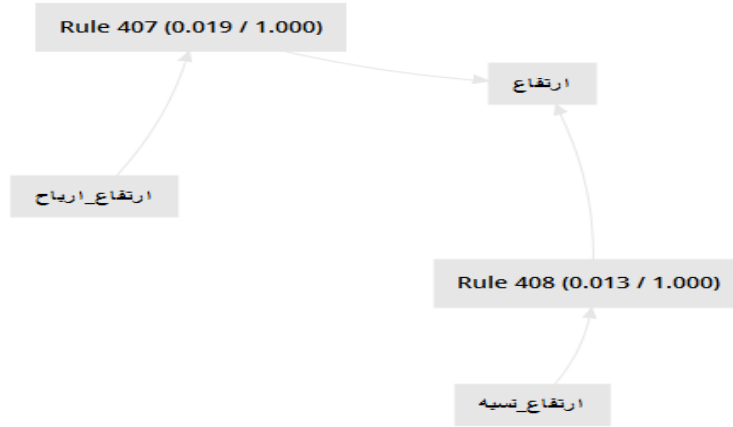


Figure 15: Visualisation of the association rules for the feature “rise” (ارتفاع).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases correlated with the feature “rise” (ارتفاع). Table 22 shows that the phrase “increase – percentage” [ارتفاع – نسبة] occurred 30 times: zero in the neutral, 23 in the positive, and seven in the negative class.

Table 22: Phrases for the feature “rise” (ارتفاع) in all the data.

Phrase -Terms	Occurrence	Neutral	Positive	Negative
ارتفاع – ارباح	37	0	37	0
ارتفاع – نسبة	30	0	23	7

From Table 22, the phrase “rising profit” [ارتفاع-ارباح] does not need further investigation, since it only occurred in the positive class.

The phrase “increase – percentage” [ارتفاع – نسبة] occurred in the negative class, the negative class becomes questionable. Therefore, the feature “increase” (ارتفاع) in the negative class needs further investigation to find out the association rules in the negative class. As a result, the second scenario will be followed: extracting association rules that occurred for the feature “increase” (ارتفاع) in the negative class. Association rules are generated about the minimum support and minimum confidence thresholds using the previous process of the visualisation.

Figure 16 shows that the feature “increase” (ارتفاع) occurred with many rules in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [ارتفاع] → [ارتفاع – نسبة] (support: 0.015 confidence: 1), which represents the phrase “increase – percentage” [ارتفاع – نسبة] illustrated in the wordlist matrix in the negative class.

Premises	Conclusion	Support	Confidence
صافي	ارتفاع	0.004	0.667
شركاته	ارتفاع	0.004	0.667
نتج	ارتفاع	0.004	0.667
الذهب يواصل	ارتفاع	0.004	0.667
الثانية	ارتفاع	0.004	0.667
الأمريكية يتبع	ارتفاع	0.004	0.667
وسط غرب	ارتفاع	0.006	0.750
وسيلة	ارتفاع	0.006	0.750
قراب	ارتفاع	0.006	0.750
رغم	ارتفاع	0.009	0.800
الذهب ارتفاع	ارتفاع	0.015	1
ارتفاع نسبة	ارتفاع	0.015	1

Figure 16: The correlation rules of the feature “rise” (ارتفاع) in the negative class.

According to Figure 16, the rule [ارتفاع] → [ارتفاع – نسبة] (support: 0.015 confidence: 1), which represents the phrase “increase – percentage” [ارتفاع – نسبة], occurred in the negative class. Therefore, the next step is to search for the phrase “increase – percentage” [ارتفاع – نسبة] in the negative class documents. Table 23 shows that the phrase “increase – percentage” [ارتفاع – نسبة] occurred in three documents. Moreover, three documents have been found to satisfy the rule [ارتفاع] → [ارتفاع – نسبة], therefore they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from the structure of the three documents that Arabic is a derivation language which can create new words from other words. For example, ارتفاع can create مرتفعاً, يرتفع, يرتفع. All those words have the same meaning (“raise” in English), but each term has a different shape; this is one of the problems that can be solved by light stemming in the pre-processing phase. Moreover, replacing the token operator was used to replace % with the term نسبة during the pre-processing phase. For this reason, the term “نسبة” cannot be seen in the original documents.

Table 23: The phrase “increase – percentage” [ارتفاع-نسبة] in the negative class documents.

Original labelling class	Original Arabic tweets with English translation	New labelling class
Negative	صافي أرباح قطاع التجزئة ترتفع 12% رغم تراجع ثلث شركاته Net profit for the retail sector <u>increased</u> 12% despite a third of its companies <u>falling</u>	Positive
Negative	السوق السعودي يرتفع 04% بنهاية الأسبوع وصافولا تتراجع بقطاعها The Saudi market <u>is up</u> 4% at the end of the week, and Savola is <u>falling</u> in its sector	Positive
Negative	المؤشر العام يغلق مرتفعاً وميدغلف يتراجع بأكثر من 8% General index closes <u>higher</u> and Medgulf <u>went down</u> more than 8%	Positive

The third example is for “losses”, (خسائر) which is a negative sentiment. First, we should visualise the most important rules for the term “losses” (خسائر). Table 24 shows the feature “losses” (خسائر) as a negative sentiment in the Saudi stock market domain. In addition, Table 24 shows the occurrence of the feature “losses” (خسائر) in the positive, negative, and neutral classes.

Table 24: Phrases for the feature “losses” (خسائر) in all the data.

feature	Occurrence	Neutral	Positive	Negative
خسائر	87	0	25	62

Figure 17 shows the association rules related to the feature “losses” (خسائر) in all class with respect to the minimum support and minimum confidence thresholds. The feature “losses” (خسائر) meaning the fact of losing in investment of some business in the Saudi stock market. In addition, Figure 17 shows the most important rule for the feature “losses” (خسائر), which is [خسائر] → [المتراكمة] (support: 0.013 confidence: 1). The term “losses” [خسائر] correlated with the term “Accumulated” (المتراكمة) to compose the positive phrase “accumulated losses” in the sentence.



Figure 17: Visualisation of the association rules for the feature “losses” (خسائر).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases correlated with the feature “losses” (خسائر). Table 25 shows that the phrase “accumulated losses” [خسائر-المتراكمة] occurred 14 times: zero in the neutral, four in the positive, and 10 in the negative class.

Table 25: Phrases for the feature “accumulated losses” (خسائر-المتراكمة) in all the data.

Phrase -Terms	Occurrence	Neutral	Positive	Negative
خسائر-المتراكمة	14	0	4	10

Since the phrase Accumulated losses [خسائر-المتراكمة] occurred in the positive class, the positive class becomes questionable. Therefore, the feature “losses” (خسائر) in the positive class needs further investigation in order to find out the association rules in the positive class. As result, the second scenario will be followed, extracting association rules that occurred for the feature

“losses” (خسائر) in the positive class. Association rules are generated with regard to the minimum support and minimum confidence thresholds using the previous process of visualisation. Figure 18 shows that the feature “accumulated losses” (خسائر-المتراكمة) occurred with many rules in the premises column with the minimum support and minimum confidence values.

Premises	Conclusion	Support	Confidence
تصريحات	خسائر	0.002	0.667
تراجع_خسائر	خسائر	0.007	1
خسائر_المتراكمة	خسائر	0.005	1
انخفاض_خسائر	خسائر	0.005	1
المتراكمة	خسائر	0.005	1
نسيه_رسمالها	خسائر	0.004	1
خسائر_تصريحات	خسائر	0.002	1
خسائر_الابوجية	خسائر	0.002	1
المتراكمة_اقل	خسائر	0.002	1
اقل_نسيه	خسائر	0.002	1
اقل	خسائر	0.002	1
بواصل_خسائر	خسائر	0.001	1
يسبق_احداث	خسائر	0.001	1
يتكبدان_خسائر	خسائر	0.001	1
يتكبدان	خسائر	0.001	1
يتكبد خسائر	خسائر	0.001	1

Figure 18: The correlation rules of the feature “accumulated losses” (خسائر-المتراكمة) in the positive class.

According to Figure 18, the rule [خسائر] → [خسائر-المتراكمة] (support: 0.005 confidence: 1), which represents the phrase “accumulated losses” [خسائر-المتراكمة], occurred in the positive class. Therefore, the next step is to search for the phrase “accumulated losses” [خسائر-المتراكمة] in the positive class documents.

Table 26 shows that the phrase “accumulated losses” [خسائر-المتراكمة] occurred in four documents. Moreover, four documents have been found to satisfy the rule [خسائر] → [خسائر-المتراكمة]; therefore, they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from the structure of the three documents that the term “descent” (إنخفاض), a negative sentiment, came before the phrase “accumulated losses” [خسائر-المتراكمة], another negative phrase that makes the three documents positive.

Table 26: The phrase “Accumulated losses” [خسائر-المتراكمة] in the positive class documents.

Original labelling class	Original Arabic tweets with English translation	New labelling class
Positive	تعلن الشركة السعودية الهندية للتأمين التعاونيه وفا للتأمين عن إنخفاض خسائرها المتراكمة إلى أقل من 50 % من رأسمالها Saudi Indian Cooperative Insurance Company (Wafa) announces its decrease in accumulated losses to less than 50% of its capital	Positive
Positive	انخفاض الخسائر المتراكمة لـ وفا للتأمين عن 50% من رأسمالها Wafa Insurance's decreased accumulated loss of 50% of its capital	Positive
Positive	تعلن الشركة المتحدة للتأمين التعاوني عن إنخفاض خسائرها المتراكمة إلى أقل من 50 % من رأسمالها United Cooperative Insurance Company announces a decrease in its accumulated losses to less than 50% of its capital	Positive
Positive	انخفاض الخسائر المتراكمة لـ وفا للتأمين عن 50% من رأسمالها Decreased loss of Wafa Insurance accumulated 50% of its capital	Positive

Finally, green represents a positive sentiment in the stock market domain, and red represents a negative sentiment in the HMI field in computing. In other words, colour can be a sentiment in this domain. Table 27 shows the feature “green” (الاحضر) as a positive sentiment in the Saudi stock market domain. In addition, Table 27 shows the occurrence of the feature “green” (الاحضر) in the positive, negative, and neutral classes.

Table 27: Phrases for the feature “green” (الاحضر) in all the data

Feature	Occurrence	Neutral	Positive	Negative
الاحضر	15	0	10	5

Figure 19 shows the association rules related to the feature “green” (الاحضر) in all class with respect to the minimum support and minimum confidence threshold. The feature “green” (الاحضر) means that the Saudi stock market values have closed in the green. In addition, Figure 19 shows the most important rule for the feature “green” (الاحضر), which is [الاحضر] → [باللون] (support: 0.006 confidence: 1). The feature “green” (الاحضر) correlated with the term “colour” (اللون) to compose the positive phrase “green colour” in the sentence.



Figure 19: Visualisation of the association rules for the feature “green” (الاحضر).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases correlated with the feature “green” (الاحضر). Table 28 shows that the phrase “green

colour” [باللون\_الاخضر] occurred 14 times: zero in the neutral, eight in the positive, and five in the negative class.

Table 28: Phrases for the feature “green colour” (باللون\_الاخضر) in all the data.

Phrase -Terms	Occurrence	Neutral	Positive	Negative
باللون_الاخضر	14	0	8	5

Since, the phrase “green colour” [باللون\_الاخضر] occurred in the negative class, the negative class becomes questionable. Therefore, the feature “green” (الاخضر) in the negative class needs further investigation in order to find out the association rules in the positive class. As a result, the second scenario will be followed: extracting association rules that occurred for the feature “green” (الاخضر) in the negative class. Association rules are generated with regard to the minimum support and minimum confidence thresholds using the previous process of visualisation. Figure 20 shows that the feature “green colour” (باللون\_الاخضر) occurred with many rules in the premises column with the minimum support and minimum confidence values.

Premises	Conclusion	Support	Confidence
بغلق_اللون	الاخضر	0.004	0.667
باللون_الاخضر	الاخضر	0.011	1
وخطا_اللون	الاخضر	0.004	1
وخطا_واحد	الاخضر	0.002	1
وخطا_واحد	الاخضر	0.002	1
واحد_اللون	الاخضر	0.002	1
واحد	الاخضر	0.002	1
وابرظي_اللون	الاخضر	0.002	1
وابرظي	الاخضر	0.002	1
نقطه_وخطا	الاخضر	0.002	1
دسه_وخطا	الاخضر	0.002	1
عمليات_جلي	الاخضر	0.002	1
عمليات	الاخضر	0.002	1
خلفية_عاصفة	الاخضر	0.002	1
تغلق_السوق	الاخضر	0.002	1

Figure 20: The correlation rules of the feature “green colour” (باللون\_الاخضر) in the positive class.

According to Figure 20 the rule [الاخضر] → [باللون\_الاخضر] (support: 0.011 confidence: 1), which represents the phrase “green colour” [باللون\_الاخضر], occurred in the negative class. Therefore, the next step is to search for the phrase “green colour” [باللون\_الاخضر] in the negative class documents. Table 29 shows that the phrase “green colour” [باللون\_الاخضر] occurred in five



documents. Moreover, five documents have been found to satisfy the rule [الاخضر] → [باللون\_الاخضر]; therefore, they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from the structure of the five documents that the term “decline” (تراجع) as a negative sentiment came sometimes before and after the phrase “green colour” [باللون\_الاخضر], making the five documents negative. In addition, the same process was carried out for random features (positive or negative), namely, “growing” (زياده), “distribution” (توزيع), “decrease” (انخفاض), “financial penalty” (غرامة), and “delay” (تأجيل). The result shows that, after completing the process, 23 documents were sent to experts to check the labelling (see Appendix A). In total, 48 tweets were examined, and 20 tweets were relabelled. In last stage, the original data were updated according to the new labelling. Then, the updated data were loaded to run a new classification process. A comparison was carried out between the original classification and the new classification.

Table 29: Phrase “green colour” [باللون\_الاخضر] in the negative class documents

Original labelling class	Original Arabic tweets with English translation	New labelling class
Negative	المؤشر العام يغلق باللون الأخضر والتطوير العقاري يتراجع 3% Index closes in green and real estate development falls 3%	Positive
Negative	الأسواق الخليجية تتراجع على خلفية عاصفة الحزم وأبوظبي باللون الأخضر Gulf markets <u>retreat</u> against the drop of the Al-Hazm Storm and Abu Dhabi <u>in green</u>	Negative
Negative	السوق السعودي يغلق متراجعاً بنسبة 17% وقطاع واحد باللون الأخضر The Saudi market closed <u>down</u> 17% and one sector <u>in green</u>	Negative
Negative	عمليات جني ارباح تغلق السوق السعودي دون 9750 نقطة وقطاعاً واحداً باللون الأخضر Profit <u>taking</u> closes the Saudi market below 9750 points and one sector <u>in green</u>	Negative
Negative	السوق السعودي يغلق باللون الأحمر وقطاع واحد باللون الأخضر The Saudi market closes in red and one sector in green	Negative

Tables 30 and 31 show the performance accuracy for the SVM with TF-IDF schema without the N-gram feature for both the original classification and the new classification, respectively. For the neutral class, the precision for the original classification was 93.86%, which rose to 96.99% for the new classification after the relabelling process. Likewise, the recall for the original classification for the neutral class was 84.75%, which rose to 85.49% for the new classification after the relabelling process.

Table 30: All-class performance accuracy in original classification for SVM with TF-IDF schema

	true neutral	true positive	true negative	class precision
pred. neutral	550	24	12	93.86%
pred. positive	90	739	97	79.81%
pred. negative	9	65	357	82.83%
class recall	84.75%	89.25%	76.61%	

Table 31: All-class performance accuracy in new classification for SVM with TF-IDF schema.

	true neutral	true positive	true negative	class precision
pred. neutral	548	10	7	96.99%
pred. positive	87	788	134	78.10%
pred. negative	6	45	318	86.18%
class recall	85.49%	93.48%	69.28%	

Table 32 shows the comparison carried out between the original and the new classification. The obtained result shows that there was improvement when using SVM with the TF-IDF schema without the N-gram feature for the new classification.

Table 32: All-class performance accuracy comparison for SVM with TF-IDF schema without N-gram.

Classifier	Accuracy	Recall	Precision
SVM with the original labelling dataset	84.71%	83.54%	85.63%
SVM with the new labelling dataset	85.13%	82.74%	87.28%

On the other hand, a better result took place with NB using the BTO schema with the N-gram feature to examine the new labelling dataset. Tables 33 and 34 show the performance accuracy for NB using the BTO schema with the N-gram feature for the new classification. For example, the neutral class precision for the original classification was 75.90%, which rose to 79.66% for the new classification after the relabelling process. On the other hand, the recall for the original classification for the neutral class was 87.37%, which fell to 87.36% for the new classification after the relabelling process.

Table 33: All-class performance accuracy in original classification NB with BTO schema with N-gram.

	true neutral	true positive	true negative	class precision
pred. neutral	567	117	63	75.90%
pred. positive	67	596	84	79.79%
pred. negative	15	115	319	71.05%
class recall	87.37%	71.98%	68.45%	

Table 34: All-class performance accuracy in new classification NB with BTO schema with N-gram

	true neutral	true positive	true negative	class precision
pred. neutral	560	85	58	79.66%
pred. positive	65	646	91	80.55%
pred. negative	16	112	310	70.78%
class recall	87.36%	76.63%	67.54%	

Table 35 shows the comparison carried out between the original and the new classification. The obtained result shows that there was improvement when using NB with the BTO schema with the N-gram feature for the new classification.

Table 35: All-class performance accuracy comparison for NB with BTO schema with N-gram.

Classifier	Accuracy	Recall	Precision
NB with the original labelling dataset	76.27%	75.94%	75.82%
NB with the new labelling dataset	78.02%	77.17%	77.11%

This study has shown that our process can easily classify Arabic tweets and can handle many text association rules for positive, negative, and neutral classes. The analysis shows the importance of a neutral class in sentiment analysis of Arabic documents, as adding a neutral class leads to an improvement in classification accuracy. The results are different because the new vectors dictionary for the text data consists of all the words that belong to positive and negative classes as well. On the other hand, the relabelling process might lead to improving the classification accuracy results. Moreover, the results show that good understanding of the text structure is necessary, as it can affect a full document. Finally, all experiments can be mutually beneficial for both developing expert knowledge and helping machine learning methods to obtain better predictions in future.

#### **4.5. Summary**

In this chapter, techniques for improving the labelling process of sentiment analysis have been explored and validated by classifying Arabic texts about the Saudi stock market. A neutral class has been introduced to improve the sentiment analysis, with a positive result. In addition, visualisation of financial Arabic texts was conducted to learn the sentences' structure along with the sentiments behind them. Besides, misclassification could be identified and eliminated during the relabelling process. Further work is required in the Arabic financial domain to compare and address other metrics such as lift, correlation, and collective strength. These metrics can be utilised to extract and examine interesting association patterns from textual data. However, in order to select the right measures for the correlations between textual data attributes, several key properties should be considered in the examination. Moreover, to enhance the relabelling system, it is necessary to develop software that can work with the same relabelling concept by using more nodes to determine questionable classes and check even the small rules that occurred in the text. This may help to build robots dictionary in the future.

## **Chapter 5: Tuning of CRM via CEM using sentiment analysis on aspects level**

### **5.1. Introduction**

Today, social media has become a method for maintaining strong relationships between users and companies [111]. To meet users' expectations and understand their opinions via social media platforms, companies are keen to adjust their CRM based on the feedback from social media platforms. In other words, CRM can be fine-tuned based on the differences between its own prediction of users' opinions and the actual feedback from CEM. The differences between the opinions predicted by CRM and the actual opinions collected from social media are important to improve CRM further. The inconsistency in the opinions comes from two sources of error in CRM:

- The criteria or weights employed in CRM are not accurate.
- The structure of CRM is not optimised.

In the former case, CRM and CEM cover similar aspects in a given domain; their difference is solely due to the inaccuracy of the criteria or weights employed in the CRM system. In the latter case, CRM could have missed a few key aspects that are important for CEM. The difference in the latter case is inherent in the structure of the CRM system.

Facing the problem, tuning CRM via CEM needs to be organised in two levels:

- **Tuning on the aspect level:** In this case, CRM and CEM have been constructed from similar aspects; sentiment analysis could be applied to adjust the criteria or weights in the CRM via the difference between CRM and CEM. This part of the problem will be investigated in this chapter. A framework for adjusting CRM via CEM will be proposed and validated by a real-word example from the educational sector.
- **Tuning on the sentence level:** In this case, CRM and CEM have been constructed separately with different aspects. Opinions can only be collected on the sentence level for certain subjects in this domain. Tuning for this kind of system will be investigated in the next chapter. The framework above will be revised and validated by a selected case study from the service sector.

This chapter organised as follows: In Section 5.2, a framework for integrating CRM and CEM is introduced. Section 5.3 contains the CRM tuning based on CEM at the aspect level—a case study from King Abdul-Aziz University. Section 5.3.4 presents sentiment analysis for CEM along with experimental results and evaluations. In Section 5.3.6, classification using CRM along with results and evaluations is shown. In Section 5.3.8, tuning of CRM via CEM is

presented along with experimental results and evaluations. Finally, a summary of this chapter is presented in Section 5.4.

## 5.2. Framework on tuning CRM via CEM

This study proposes a framework that combines a supervised machine learning and a semantic orientation approach to tune CRM via CEM. The framework extracts data from social media first and then integrates CRM and CEM by tuning and optimising CRM to reflect the needs and expectations of users on social media. In other words, in order to reduce the gap between the users' predicted opinions in CRM and their opinions on social media, the existing data from CEM will be applied to determine the similar behavioural patterns of customers towards similar outcomes within CRM. The framework will lead to a quantitative approach to uncover relationships between the extracted data from social media and the CRM data.

Figure 21 illustrates the proposed framework for integrating CRM with CEM. The framework consists of three processes: (1) sentiment analysis for CEM, (2) classification using CRM, and (3) data tuning of CRM via CEM. In terms of data modelling, the main components of the three processes are very similar. The difference between process one and process two is in the input: one takes unstructured social media input, while the other takes a structured customer database. In process three, CRM data can be labelled automatically by CEM or vice versa. In this work, we will focus on the fine-tuning of CRM through CEM.

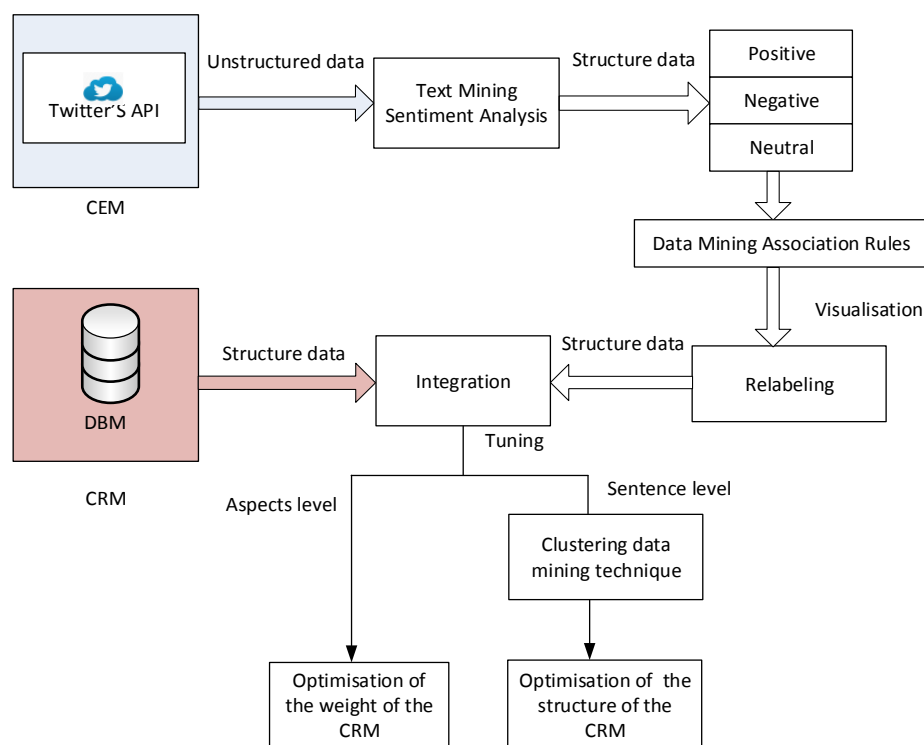


Figure 21: The process of the proposed framework of tune CRM via CEM

### **5.2.1. Sentiment Analysis for CEM**

This is a process that extracts key features for CEM using social media. The extracted features will be represented by a semantic schema. The schema can be applied directly to the social media input. The process

- Collects tweets based on a set of keywords that describe the case study using Twitter's streaming API.
- Performs text-processing techniques based on the proposed ontology model to reduce the amount of noise and extracts key features.
- Apply supervised learning method. By exploring relevant data mining approaches such as sentiment analysis and NLP on building models, which have been described in Chapter 3 and 4, the component classifies tweets according to their sentiment polarity into one of the classes of positive, negative, and neutral.

### **5.2.2. Classification using CRM**

This is a process that extracts key features for CRM using an existing customer database. The extracted features are also represented by a semantic schema, which can be applied directly to the consumer database. The process

- Collects structured data for the CRM using the database's API.
- Performs pre-processing techniques based on the existing CRM model to reduce the amount of noise and extracts key features.
- Apply supervised learning method, which is again similar to the CEM process.

### **5.2.3. Tuning on CRM via CEM**

A model has been developed that cross-validates features extracted from both CEM and CRM. In this case, if CRM and CEM are not consistent, CRM's semantic schema can be updated by CEM's output directly. In this process,

- Statistical algorithms (such as Index and Matching utilised where Match locates and feeds a position to Index of records) will be applied to discover patterns and correlations in features extracted from both CEM and CRM.
- The confidence of the discovery will be examined automatically by comparing validation between the outputs. False positives, negatives, and neutrals will be identified during the process.
- CRM's semantic schema will be revised iteratively. If CRM and CEM are constructed from similar aspects, the tuning will be focused on the criteria or weights in CRM.

Otherwise, CRM and CEM will need to be interpreted on the sentence level, so the structure of the CRM will be optimised based on its difference from CEM at this level.

- **Tuning the weight of CRM**

By asking social media users for their opinions on specific aspects or details about entities of an existing CRM, the CEM can be constructed on the aspect level. In this way, CEM will have coverage and complete matching for details regarding the CRM information. In order to reflect the social media users' needs, the weight of some aspects of the CRM criteria can be tuned to close the gap between CRM and CEM.

- **Tuning the structure of CRM**

By asking social media users for their opinions on general aspects or details about entities of an existing CRM, the CEM can be constructed on the sentence level. In this way, CEM will lack coverage and will have incomplete matching for details regarding the CRM information. In order to reflect the social media users' needs, the structure of some aspects of the CRM can be tuned to close the gap between CRM and CEM.

### **5.3. Case study for Tuning CRM via CEM at the aspect level: Optimisation of CRM of King Abdul-Aziz University**

The Deanship of e-Learning and Distance Education at King Abdul-Aziz University's Twitter account was chosen as a platform for opinion mining. The investigation was conducted on students' sentiments on the aspect level, with students being asked specific questions about e-learning criteria. The aim of this case study was to validate the proposed framework by adjusting the criteria or weights employed in the targeted CRM system.

#### **5.3.1. Background of King Abdul-Aziz University**

Students use social media and social networking sites on a daily basis to express their views and activities. Therefore, it follows that universities utilise social media in educational practice for the purpose of improving their teaching processes, as well as for the purpose of analysing opinions and learning about experiences. This is especially relevant for students who mostly depend on the web, such as online distance education students. Therefore, this chapter aims to investigate whether social media could change experts' knowledge about their users and help them understand them more. For example, in universities, some students, such as distance education students, fail not because they are not good enough but because the university does not make the right efforts to help them. To achieve this aim, our goal is to combine the results of social network users' sentiments with existing students' data based on their behaviour patterns. Therefore, in this chapter, an experiment is carried out to understand distance education students' learning experiences by analysing students' opinions from the Twitter account of the Deanship of e-Learning and Distance Education at King Abdul-Aziz



University in Saudi Arabia. The experiment examines the aspect level for students' opinions of King Abdul-Aziz University.

The Deanship of e-Learning and Distance Learning at King Abdul-Aziz University's Twitter account, which, with 38,000 followers, is a leading account in the Gulf region, has been chosen as a platform for opinion mining. The Deanship's Twitter account has won the Khalifa Educational Award in the Field of New Media at the Arab World Level, and the Deanship also received the Khalifa Prize for Education in the Field of New Media and Distance Education in the Arab World for spreading e-learning culture through new media (Facebook, Twitter, YouTube, and other new media tools) on 1 April 2015. The award is widely recognized in the educational and academic communities and granted annually under the patronage of HH Sheikh Khalifa bin Zayed, President of the UAE, and Sheikh Mohammed bin Zayed, Crown Prince of Abu Dhabi.

### **5.3.2. Preparing the sentiment analysis for CEM using the criteria from CRM**

In order to prepare the sentiment and to clean the text to hand it off to the classifier, links, URLs, and hashtags were removed from the sentiments. In the following case study, the experimental collection process was based on certain hashtags that represent important topics, using Twitter hashtags as the domain.

### **5.3.3. CEM based on sentiment on the aspect level**

In this experiment, a hashtag was created to ask users for opinions on and reactions toward distance education criteria in order to identify students' positive, negative, and neutral opinions. The aspect level of sentiment analysis was followed for the students' opinions on the Twitter platform. The collection process was based on certain hashtags that represent the criteria, such as “# Enhance Distance education criteria” “#تطوير آليات التعليم عن بُعد”. Following techniques developed in Chapters 3 and 4, the experiment aimed to illustrate the relationship between the sentiments conveyed in Arabic tweets and the students' learning experiences at universities.

### **5.3.4. Experiment on sentiment analysis for CEM**

During this stage, a hashtag was created to ask users for opinions on and reactions toward distance education criteria in order to identify students' positive, negative, and neutral opinions. Figure 22 shows the announcement conducted to encourage students to participate in the hashtag.



Figure 22: Announcement conducted to encourage students to participate in the Hashtag.

### 5.3.4.1. Experimental parameter settings

The same experimental parameter settings used in Chapter 3 for sentiment analysis of CEM were utilised in this experiment. For example, the classifier settings followed the default values for most of the parameters to get more accurate results. Ten-fold cross-validation was used one time to find the best values of these parameters. In addition, both SVM and NB had the same operator parameter settings used.

### 5.3.4.2. Data collection and description

To collect opinions that were comprehensive for the time period on the targeted objectives, tweets from different students on the Deanship of e-Learning and Distance Education page were obtained using Twitter's official developer's API. The data were collected over four months between 14 April and 13 August 2017. The data distribution depended on the methods Twitter's API utilised and the number of tweets posted on the distance education Twitter account page. Downloaded tweets were marked manually by employees of the distance education deanship as positive, negative, or neutral. In addition, the labelled tweets were stored in a database for experiments (see Appendix B).

Distance education students' CRM data were obtained from the Learning Management System (LMS) Blackboard at King Abdul-Aziz University, which consists of the most important fields that can describe the students' activities. In this experiment, the collection process was based on students' data from one course chosen randomly from the student schedule of the second semester in 2017. Distance education students were labelled as "neutral" who had total marks in a course between 60 and 75 and attended 9 to 10 out of 12 lectures. For students who had total marks in a course less than 60 and attended fewer than 9 out of 12 lectures, the label was "negative". If a student had total marks above 75 and attended more than 10 out of 12 lectures, then he or she was considered an active student and labelled as "positive". Table 36 shows the number of tweets that were involved in the classification experiment. One hundred and forty-three students' details were collected from the Blackboard CRM. All of them were anonymised. After that, the data were labelled depending on distance education criteria, and then the data were stored. Rapidminer was used to load the dataset in normalized form with no duplicate records and no null attribute values. After the data collection stage, the data model layout was created. This included a tweets description table (Tweet ID, Tweet Original, Tweet Filtered, Tweet Time, Tweet User, Tweet Label) and selected attributes from Blackboard databases. Again, this information was anonymised to protect privacy. Each student record contained the individual student's positive, negative, and neutral opinions, as well as the original tweet and the tweet labelled as an opinion from the CEM data for each individual student. Each student's record thus had two labels. The CRM label depended on the CRM criteria and the CEM label, which came from the sentiment analysis model.

Ten thousand six hundred and three students viewed the hashtag and agreed to start the experiment. Only 567 students registered their Twitter accounts and allowed the university to follow their tweets. One of the obstacles to letting students continue the survey was that the authentication from Twitter is in English. However, 242 of the 567 students have records in Blackboard, since only distance education students must use Blackboard regularly, whereas regular students and external students in some colleges still do not use Blackboard. Of these 242 students, 143 completed the survey.

The similar proposed model for sentiment analysis of Saudi Arabic (Standard and Arabian Gulf dialects) tweets was applied to extract feedback features from King Abdul-Aziz University data. The main idea was to examine the aspect level in sentiment analysis. Moreover, we proved in Chapter 4 that generalization of our classification with a neutral class shows the importance of a neutral class in sentiment analysis. In addition, the neutral class is important in the Arabic sentiment analyses, so our experiment was carried out with the neutral class as well as the negative and positive classes. The following tables show the type of tuning to

provide the best accuracy in sentiment analysis for Arabic text in relation to King Abdul-Aziz University. For this experiment, 143 students' tweets responding to four questions were utilised. Table 36 shows the four questions asked of participants. A survey with four questions for e-learning criteria<sup>6</sup> was created and made available on Twitter via hashtags (see Table 37). Three questions asked a specific question on the criteria, while the last one was an open question asking students about their opinions in general about e-learning. Table 37 shows the total number of tweets for the 143 students utilized for this experiment.

Table 36: A survey with four questions for e-learning criteria

Question No.	Question
Q1	What do you think about the following criteria: The student evaluates from 100 degrees: the quarterly work (30 degrees of the total) and the final test (70 degrees of the total).
Q2	What do you think about the following criteria: The student is denied entry to the final exam of the distance learning course if his absence exceeds 3 (25%) of the lectures.
Q3	What do you think about the following criteria: The quarterly work is divided as follows: four duties for each subject and the calculation of a single assignment three degrees, two periodic tests and each test six degrees, participation in the forums (discussion board) three posts, and each class two degrees (2).
Q4	Are there any comments you would like to share with us about the distance learning mechanism?

Table 37: e-learning criteria.

Number	E-learning criteria
1	Classwork: includes interaction with the instructor through the given activities available in the learning management system (30 marks of the grand total). Final examination: 70 marks of the grand total.
2	If student absence exceeded three lectures (equivalent to 25% of the synchronous online lectures) offered throughout the semester, then the student is prevented from taking the final exam of the distance learning course.
3	The quarterly work is divided as follows: assignments 4 (3 marks per assignment), activities 2 (6 marks per activity), forums (discussion board) 3 (2 marks per forum).
4	Any comments about the distance learning mechanism.

<sup>6</sup> [http://elearning.kau.edu.sa/Content.aspx?Site\\_ID=214&lng=EN&cid=241446](http://elearning.kau.edu.sa/Content.aspx?Site_ID=214&lng=EN&cid=241446)

Table 38: Number of tweets utilized for Sentiment analysis for CEM.

Question	Positive	Negative	Neutral	Total
Q1	73	11	59	143
Q2	54	80	9	143
Q3	87	37	19	143
Q4	39	42	62	143
Total	253	170	149	572

### 5.3.4.3. Data pre-processing

A selection of 28 words and idioms in Arabic from the emotion corpus such as growth, good, excellent, problem, and inappropriate (see Table 399) was then formed into the following three classes: positive, negative, and neutral. The most common Arabic words in the standard Saudi dialect represent positive and negative classes.

Table 39: Common Arabic words and standard Saudi dialect representing the classes

Arabic Positive Sentiment	English Translation	Arabic Negative Sentiment	English Translation
فتح تخصصات جديده	open new courses	مشكلة	problem
مناسب	suitable	تطوير بلاك بورد	develop Blackboard
جيد	good	لا يوجد تجاوب	no response
متكامل	complementary	غير مناسب	unsuitable
ممتاز	excellent	غير ملائم	inappropriate
سريع جدا	very fast	غير كافي	inadequate,
نظام جيدة	good platform	حرمان	absence exceeds limits
موافق	agree	زيادة النسبة	Increase percentage

Two distance education employees who have experience in e-learning labelled the data manually. Positive tweets were given the label "1", while negative tweets were given the label "-1". Neutral tweets were given the label "0", and irrelevant tweets were deleted from the database. After data labelling was completed, the data were stored in our system in normalized form, as mentioned in the framework, with no hashtags, no duplicate tweets, no retweets, no URLs, and no special characters. After loading the dataset, the data were pre-processed with RapidMiner. The first step was to replace some Arabic words taking different shapes and icons. For example, Table 40 shows replacement of words in the pre-processing phase.



Table 41: Accuracy, precision, and recall for all classes using SVM and NB classifiers without N-gram

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	39.24%	43.95%	48.18%
		TF-IDF	55.19%	54.59%	54.54%
	SVM	BTO	77.57%	55.06%	52.73%
		TF-IDF	83.19%	59.65%	56.74%

Table 42 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature, which is set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 42: Accuracy, precision, and recall for all classes using SVM and NB classifiers with N-gram

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	55.24%	53.36%	54.42%
		TF-IDF	67.86%	61.65%	62.66%
	SVM	BTO	78.95%	61.73%	60.42%
		TF-IDF	81.05%	63.59%	61.97%

In conclusion, the experiment shows that NB performance was better when we used the N-gram feature with both schemas (TF-IDF and BTO). On the other hand, there was a slight performance increase when SVM used the same feature. However, the best accuracy was achieved by SVM with the TF-IDF schema when the N-gram feature was not involved.

#### 5.3.4.6. Sentiment Analysis Classification for Question 2

Table 43 shows Q2 classification accuracy, precision, and recall for the NB and SVM classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 43: Accuracy, precision, and recall for all classes using for SVM and NB classifiers without N-gram

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	77.62	66.44	63.18
		TF-IDF	76.24	65.47	62.26
	SVM	BTO	78.29	56.11	52.81
		TF-IDF	72.76	49.67	50.48

Table 44 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature, which is set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 44: Accuracy, precision, and recall for all classes using for SVM and NB classifiers with N-gram

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	78.33	66.86	63.54
		TF-IDF	77.67	66.31	63.06
	SVM	BTO	74.90	54.53	51.81
		TF-IDF	70.00	46.94	49.80

In conclusion, the experiment shows that NB performance was better when we used the N-gram feature with both schemas (TF-IDF and BTO). On the other hand, there was a performance decrease when SVM used the same feature. However, the best accuracy was achieved by SVM with the BTO schema when the N-gram feature was not involved.

### 5.3.4.7. Sentiment Analysis Classification for Question 3

Table 45 shows Q3 classification accuracy, precision, and recall for the NB and SVM classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 45: Accuracy, precision and recall for all classes using for SVM and NB classifiers without N-gram.

Classes	Classifier Name	Weighting schemes	Accuracy	class recall	Class precision
All Class	NB	BTO	72.05	64.58	66.34
		TF-IDF	76.95	67.22	67.20
	SVM	BTO	80.33	61.85	62.91
		TF-IDF	82.52	65.19	68.74

Table 46 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature, which is set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.



Table 46: Accuracy, precision and recall for all classes using for SVM and NB classifiers with N-gram.

Classes	Classifier Name	Weighting schemes	Accuracy	class recall	Class precision
All Class	NB	BTO	80.48	70.46	67.66
		TF-IDF	81.86	71.76	71.15
	SVM	BTO	79.67	62.69	63.60
		TF-IDF	81.05	61.39	61.32

In conclusion, the experiment shows that NB performance was better when we used the N-gram feature with both schemas (TF-IDF and BTO). On the other hand, there was a performance drop when SVM used the N-gram feature with both schemas (TF-IDF and BTO).

#### 5.3.4.8. Sentiment Analysis Classification for Question 4

Table 47 shows Q6 classification accuracy, precision, and recall for the NB and SVM classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 47: Accuracy, precision, and recall for all classes using for SVM and NB classifiers without N-gram.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	60.00%	61.51%	59.15%
		TF-IDF	69.10%	71.34%	70.53%
	SVM	BTO	58.00%	53.27%	65.65%
		TF-IDF	68.43%	67.48%	74.74%

Table 48 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-Gram feature, which is set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 48: Accuracy, precision, and recall for all classes using for SVM and NB classifiers with N-gram.

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	59.29%	61.03%	58.54%
		TF-IDF	69.81%	71.90%	71.59%
	SVM	BTO	49.81%	42.33%	48.96%
		TF-IDF	62.86%	62.89%	72.66%

In conclusion, the experiment shows that NB performance was better when we used the N-gram feature with both schemas (TF-IDF and BTO). On the other hand, there was a reduction when SVM used the N-gram feature with both schemas (TF-IDF and BTO).

### **5.3.5. Classification using CRM**

Component two presented the design and implementation of students' CRM data (or classification through different algorithms, such as SVM and NB. In this part of the experiment, the collection process was based on students' Blackboard attributes assigned according to their points of view on different classes in Blackboard data. It was labelled according to the university's e-learning criteria.

#### **5.3.5.1. Details of using the classifiers with RapidMiner for CRM classification**

The classifier settings followed the default values for most of the parameters to get more accurate results. Ten-fold cross-validation was used several times to find the best values of these parameters.

#### **5.3.5.2. Support vector machine (SVM) operator parameter settings**

The SVM operator generates the SVM classification model. This model can be used for classification and provides good results for many learning tasks. In addition, it supports various kernel types, including dot, radial, polynomial, and neural [170].

- SVM type: C-SVM, which is for classification tasks.
- linear: A linear classifier works based on the value of a linear combination.
- C is the penalty parameter of the error term and was set to its default value, which is zero.
- Cache size is an expert parameter. It specifies the cache size in megabytes and was set to default value, which is 80.
- Epsilon: This parameter specifies the tolerance of the termination criterion and was set to the default value, which is 0.001 [170].

#### **5.3.5.3. Naïve Bayes (NB) operator parameter settings**

The NB operator generates a NB classification model. It is a probabilistic classifier based on applying Bayes' theorem with powerful independence assumptions. The NB classifier assumes that the presence or absence of a particular feature of a class is unrelated to any other feature[170].

- Laplace correction:

This parameter indicates whether Laplace correction should be used to prevent high influence of zero probabilities. Assume that our training set is so large that adding one to each count will make a small difference in the estimated likelihoods [170].

### 5.3.6. Experiment on the Classification using CRM

Similar approach for CRM in the previous chapter applied to classify student's data structured inputs from King Abdul-Aziz university blackboard database each student record labelled according to the university e-learning criteria. For this experiment 567 student registered their twitter account and allow the university to follow their tweets. However, only 242 students are distance education students and have full record in Blackboard, where other students do not use Blackboard. **Error! Reference source not found.**49 shows the total number of students record for the 242 students utilised for this experiment. For this experiment 567 student registered their twitter account and allow the university to follow their tweets. However, only 242 students are distance education students and have full record in Blackboard, where other students do not use Blackboard.

Table 49: Number of CRM utilized for Classification using CRM.

Question	Positive	Negative	Neutral	Total
Q1	84	65	93	242
Q2	90	122	30	242
Q3	84	89	69	242
Q4	80	93	69	242
Total	338	369	261	968

#### 5.3.6.1. Data Description for Question 1

Table 50 shows the data description for question 1, and Table 51 show the criteria utilised for question 1. Table 51 shows the criteria utilised to label record as positive, negative, or neutral for question 1.

Table 50: Question 1 data description.

SEM	Semester
STUD_NO	Student ID
CRN_NO	Division number
BB_F_MARK_NO	Student Mark at Blackboard (total mark is 30)
ODS_MARK	Student Final Exam Mark at Online Demand University System (ODUS) from (total mark is 70)
TOT_MARK	Student Final Mark in ODUS from (total mark is 100)

Table 51: Question 1 Criteria.

	Labeled	Criteria
Student Mark at Blackboard (total mark is 30)	Positive	>20
	Neutral	15-20
	Negative	<15
Student Final Exam Mark in ODUS from (total mark is 70)	Negative	<35
	Neutral	35-45
	Positive	>45
Student Final Mark in ODUS from (total mark is 100)	Negative	<50
	Neutral	50-75
	Positive	>75
TOT_MARK = BB_F_MARK_NO + ODS_MARK		

### 5.3.6.2. Data Description for Question 2

Table 52 shows the data description for question 2, and Table 53 show the criteria utilised for question 2.

Table 53 shows the criteria utilised to label record as positive, negative, or neutral for question 2.

Table 52: Question 2 data description.

SEM	Semester
STUD_NO	Student ID
CRN_NO	Division number
REC_MARK_COUNT	Number of recorded attendance
ONLINE_MARK	Number of online attendance
ATTEND_MARK	Total student attendance

Table 53: Question 2 Criteria

	labeled	Criteria
Number of recorded attendance	Negative	>5 days
	Neutral	4-5 days
	Positive	<4 days
Number of online attendance	Negative	<9 lectures
	Neutral	9-10 lectures
	Positive	>10 lectures
ATTEND_MARK = REC_MARK_COUNT + ONLINE_MARK		

### 5.3.6.3. Data Description for Question 3

Table 54 shows the data description for question 3, and Table 55 show the criteria utilised for question 3.

Table 55 shows the criteria utilised to label record as positive, negative, or neutral for question 3.

Table 54: Question 3 data description

SEM	Semester
STUD_NO	Student ID
CRN_NO	Division number
FORUM_TEST_MARK_FORUM	Number of times student participate in the discussion
FORUM_TEST_MARK_TEST	Number of student participation in quizzes and assignments
FORUM_TEST_MARK_TOTAL	Total posts of discussions, quizzes, and assignments

Table 55: Question 3 Criteria

	labeled	Criteria
Number of times student participate in the discussion, quizzes and assignments	Negative	<5
	Neutral	5-6
	Positive	>6
FORUM_TEST_MARK_TOTAL = FORUM_TEST_MARK_FORUM + FORUM_TEST_MARK_TEST		

### 5.3.6.4. Data Description for Question 4

Table 56 shows the data description for question 4, and Table 57 show the criteria utilised for question 4. Table 57 shows the criteria utilised to label record as positive, negative, or neutral for question 4.

Table 56: Question 4 data description.

SEM	Semester
STUD_NO	Student ID
CRN_NO	Division number
BB_F_MARK_NO	Student Mark at Blackboard (total mark is 30)
ODS_MARK	Student Final Exam Mark in ODUS from (total mark is 70)
TOT_MARK	Student Final Mark in ODUS from (total mark is 100) TOT_MARK = (BB_F_MARK_NO+ ODS_MARK)

REC_MARK_COUNT	Number of recorded attendance
ONLINE_MARK	Number of online attendance
ATTEND_MARK	Total student attendance ATTEND_MARK = (REC_MARK_COUNT+ ONLINE_MARK)
FORUM_TEST_MARK_FORUM	Number of times the discussion took place
FORUM_TEST_MARK_TEST	Number of test sessions
FORUM_TEST_MARK_TOTAL	Total posts of discussions and tests FORUM_TEST_MARK_TOTAL = (FORUM_TEST_MARK_FORUM+ FORUM_TEST_MARK_TEST)
FORUM_TEST_MARK_RESULT	Evaluate the total number of posts with discussions, quizzes, and assignments

Table 57: Question 4 Criteria.

	labeled	Criteria
Total Mark	Negative	<4
	Neutral	5-4
	Positive	>5
TOTAL_EVA_CHAR = (TOT_MARK + ATTEND_MARK + FORUM_TEST_MARK_RESULT) / 3		

### 5.3.6.5. Experimental results and evaluations

The results were divided into groups for each question to show the Blackboard students' data classification accuracy, precision, and recall for the NB and SVM classifiers.

### 5.3.6.6. CRM classification for Question 1

The following table shows the type of tuning to provide the best accuracy in King Abdul-Aziz University's CRM classification. Table 58 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling. The best accuracy was achieved by NB due to the advantages of NB, such as its simplicity, ease of implementation, and combination of efficiency with acceptable accuracy, as is stated in Chapter 2.

Table 58: Accuracy, precision, and recall with all classes for SVM and NB classifiers.

Classes	Classifier name	Accuracy	Class recall	Class precision
Positive, negative, and neutral	NB	93.43%	93.95%	94.12%
	SVM	87.25%	86.55%	89.74%

### 5.3.6.7. CRM classification for Question 2

The following table shows the type of tuning to provide the best accuracy in King Abdul-Aziz University's CRM classification. Table 59 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling. The best accuracy was achieved by NB due to the advantages of NB, such as its simplicity, ease of implementation, and combination of efficiency with acceptable accuracy, as is stated in Chapter 2.

Table 59: Accuracy, precision, and recall with all classes for SVM and NB classifiers.

Classes	Classifier name	Accuracy	Class recall	Class precision
Positive, negative, and neutral	NB	95.05%	96.69%	96.48%
	SVM	81.80%	62.82%	55.83%

### 5.3.6.8. CRM classification for Question 3

The following table shows the type of tuning to provide the best accuracy in King Abdul-Aziz University's CRM classification. Table 60 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling. The best accuracy was achieved by NB due to the advantages of NB, such as its simplicity, ease of implementation, and combination of efficiency with acceptable accuracy, as is stated in Chapter 2.

Table 60: Accuracy, precision, and recall with all classes for SVM and NB classifiers.

Classes	Classifier name	Accuracy	Class recall	Class precision
Positive, negative, and neutral	NB	98.33%	72.44%	72.31%
	SVM	96.72%	65.44%	64.52%

### 5.3.6.9. CRM classification for Question 4

The following table shows the type of tuning to provide the best accuracy in King Abdul-Aziz University's CRM classification. Table 61 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling. The best accuracy was achieved by NB due to the advantages of NB, such as its simplicity, ease of implementation, and combination of efficiency with acceptable accuracy, as is stated in Chapter 2.

Table 61: Accuracy, precision and recall with all classes for SVM and NB classifiers.

Classes	Classifier name	Accuracy	Class recall	Class precision
Positive, negative, and neutral	NB	85.13%	84.61%	85.88%
	SVM	83.05%	81.47%	83.12%

### 5.3.7. Tuning of CRM via CEM

This process aims to find a way to support CRM (or Blackboard in this case study). For instance, what exactly do students want or think about experiences? This is especially relevant for students who mostly depend on the Web, such as online distance education students, also proving that the value of social media information can bring better understanding of a student's study situation. Therefore, similar tweets are grouped such that tweets within the same group bear similarity to each other, while tweets in different groups are dissimilar from each other. This will help to understand students' behaviours and find out the most common problems. Moreover, this will give the university the ability to learn about and validate students' data with more support from social media to develop new e-learning criteria to match the inputs from social media.

This case study investigates an alternative solution for supporting CRM by social media inputs on the aspect sentiment level and tuning CRM weights for some aspects of the e-learning criteria that students need, according to their posts on social media. Optimisation of the CRM weights was applied to update some aspects' values in CRM. The results show closely CRM's student labels match CEM's, especially in the negative and neutral classes. Furthermore, they show that optimising CRM's weights can enhance classification accuracy in the Blackboard data and help to understand more students' studying statuses.

### 5.3.8. Experiment in tuning on CRM via CEM

To be included in this experiment, students should have records in CRM and have completed the survey. Out of the 143 with records on Blackboard, only 79 completed the survey. Table 62 shows the difference between CRM and CEM in this case study before the tuning.



Table 62: Number of integrated records utilized for tuning of CRM via CEM.

Question	CEM/CRM	Positive	Negative	Neutral
Q1	CEM	46	5	28
	CRM	31	23	25
Q2	CEM	29	46	4
	CRM	37	24	18
Q3	CEM	28	26	25
	CRM	36	13	30
Q4	CEM	18	26	35
	CRM	31	20	28
Total CEM		121	103	92
Total CRM		135	80	101

### 5.3.8.1. Experiment results and evaluations

The results were divided into four groups for each question to show the comparison between the CRM and CEM labelling classes.

### 5.3.8.2. Integration results for Question 1

Figure 23 shows a comparison between CRM labelling and CEM labelling. Thirty-one tweets were labelled similarly by CRM and CEM. On the other hand, 48 tweets were labelled dissimilarly by CEM and CRM. In other words, 62% of collected tweets were labelled differently by CRM and CEM, and only 38% were similarly labelled by CEM and CRM.

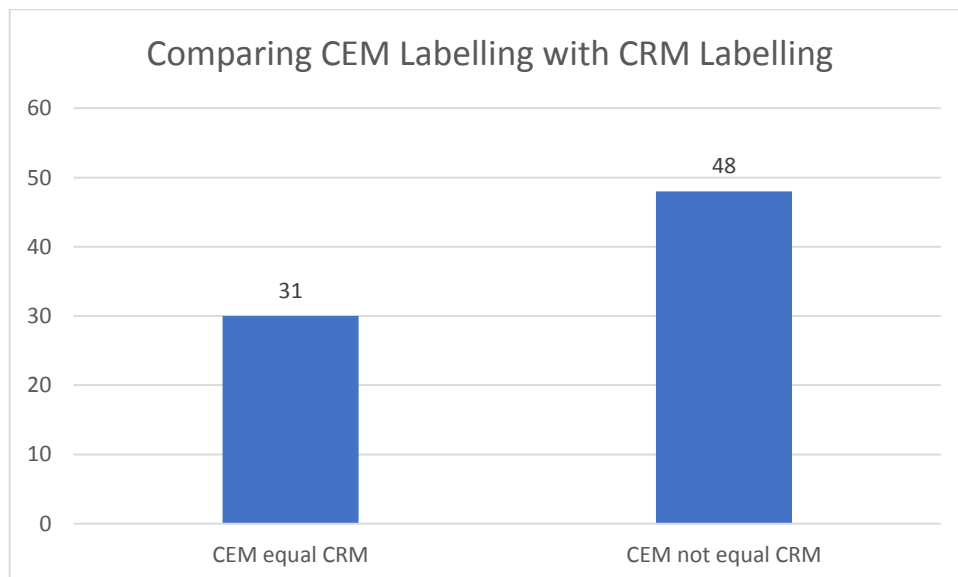


Figure 23: Comparing CEM labelling with CRM labelling for Question 1.

From Figure 24, there are 46 positive tweets in CEM, compared to 31 tweets in CRM. In contrast, there are only five negative tweets in CEM, compared to 23 negative tweets in CRM. For the neutral class, there are 28 tweets in CEM and 25 in CRM.

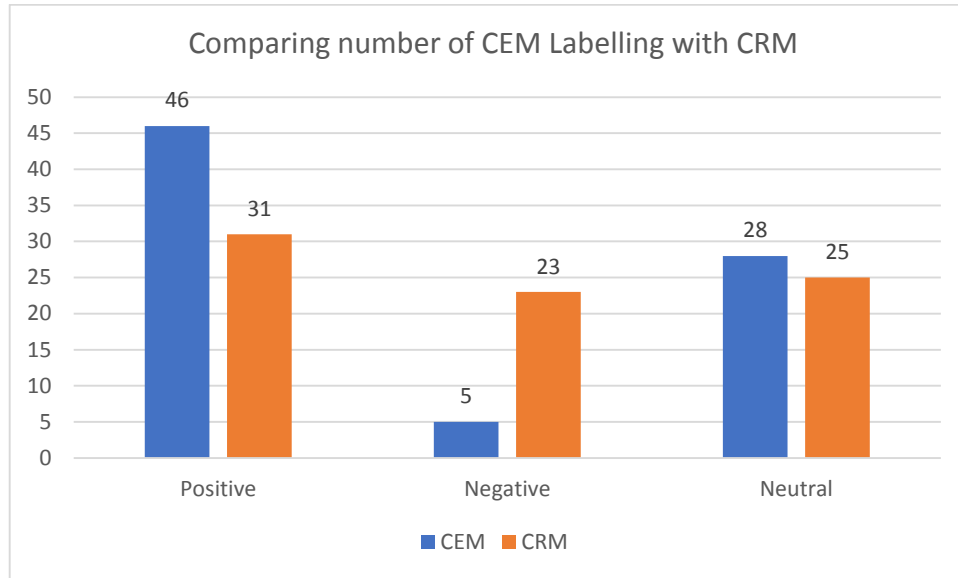


Figure 24: Comparing CEM labelling class with CRM labelling class for Question 1.

Figure 25 shows the number of tweets with similar labelling between CRM positive class and the CEM (positive, negative and neutral) labelling for Question 1. The total number of tweets labelled as positive in the CEM is 22. In contrast, 10 tweets were labelled as negative and 14 tweets as neutral. Overall, 52% of the labelling was inconsistent between CRM and CEM for CEM's positive class.

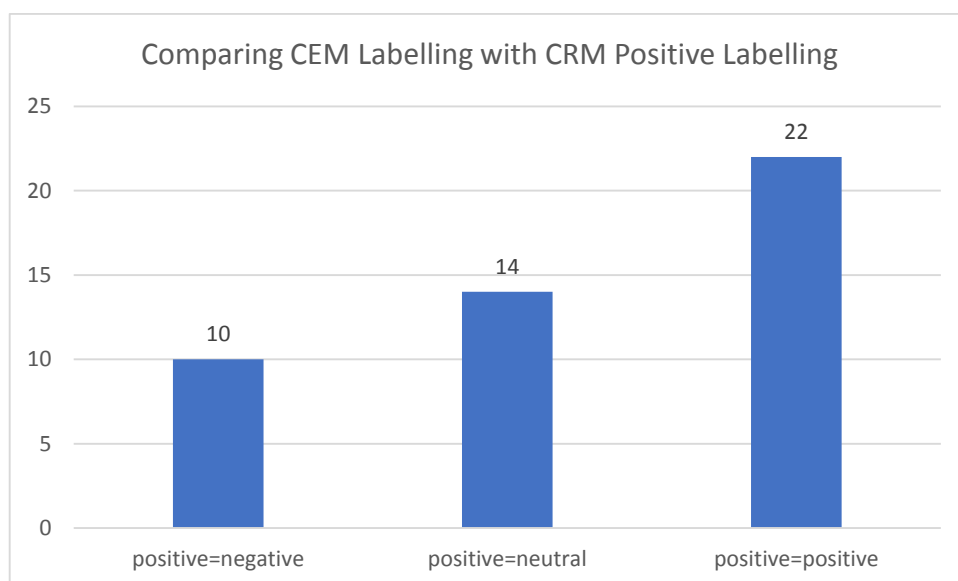


Figure 25: Comparing CEM all labelling classes with CRM positive labelling for Question 1.

Figure 26 shows the number of tweets with similar labelling between the CRM negative class and the CEM (positive, negative, and neutral) labelling for Question 1. The number of tweets labelled as negative in the CEM is only one tweet. In contrast, 1 tweet were labelled as positive and 3 tweets as neutral.

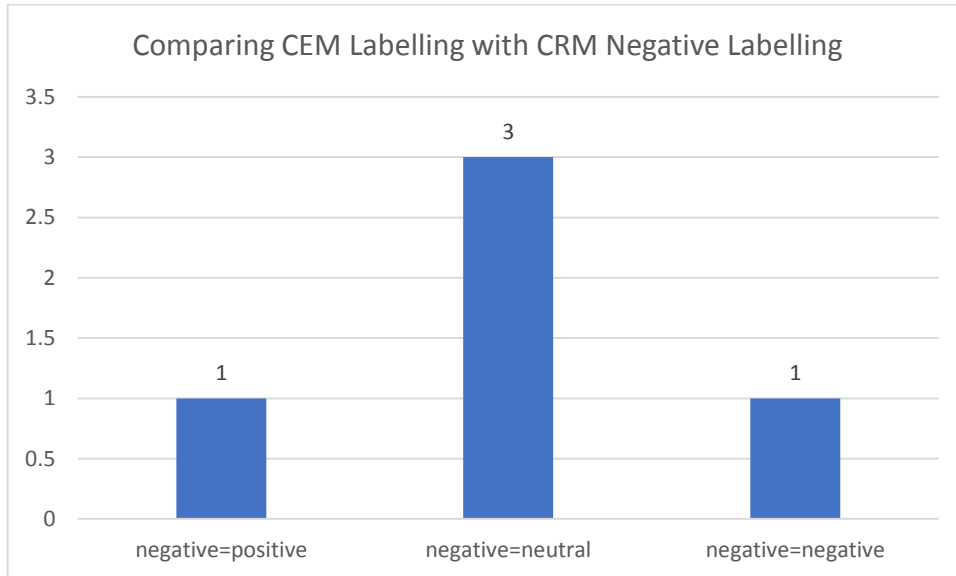


Figure 26: Comparing CEM all labelling classes with CRM negative labelling for Question 1.

Figure 27 shows the number of tweets with similar labelling between the CRM neutral class and the CEM (positive, negative, and neutral) labelling for Question 1. In CEM, 28 tweets were labelled as neutral. By contrast, CRM labelled eight of these tweets as neutral, eight as positive, and 12 as negative.

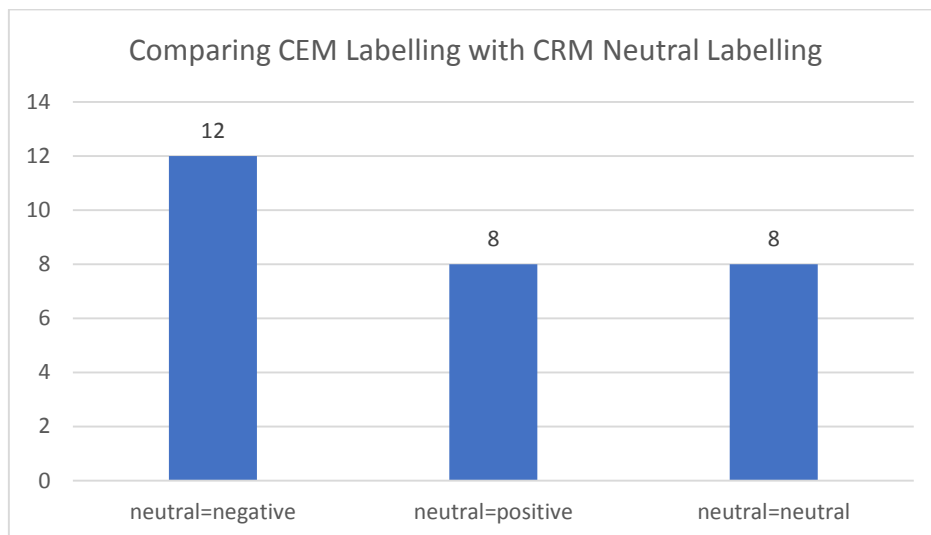


Figure 27: Comparing CEM all labelling classes with CRM neutral labelling for Question 1.

### 5.3.8.3. Integration Results for Question 2

Figure 28 shows a comparison between CRM labelling and CEM labelling. Twenty-three tweets were labelled similarly between CRM and CEM. On the other hand, 56 tweets were labelled dissimilarly between CEM and CRM. In other words, 71% of collected tweets were labelled differently between CRM and CEM, and only 29% had the same labelling between CEM and CRM.

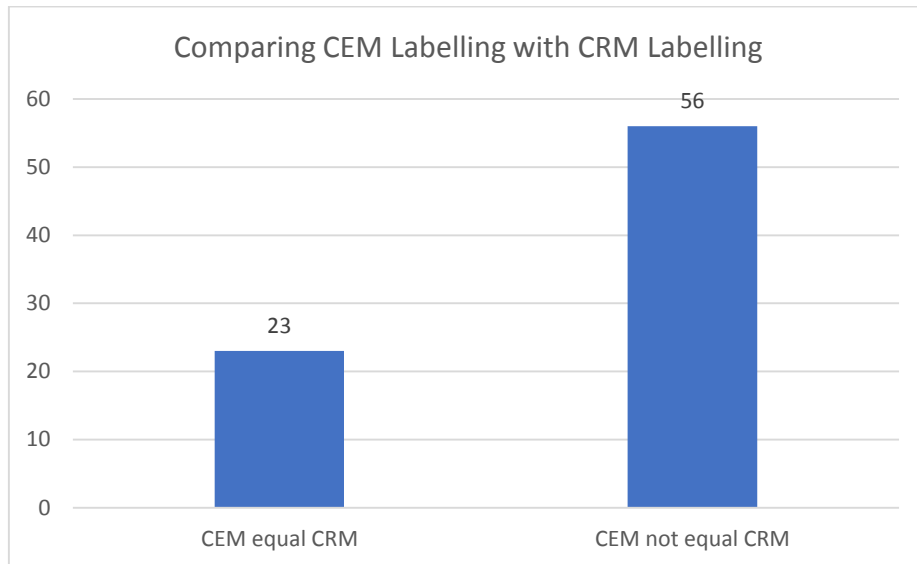


Figure 28: Comparing CEM labelling with CRM labelling for Question 2.

From Figure 29, there are 29 positive tweets in CEM compared to 37 tweets in CRM, 24 negative tweets in CEM compared to 46 in CRM, and four neutral tweets in CEM compared to 18 in CRM.

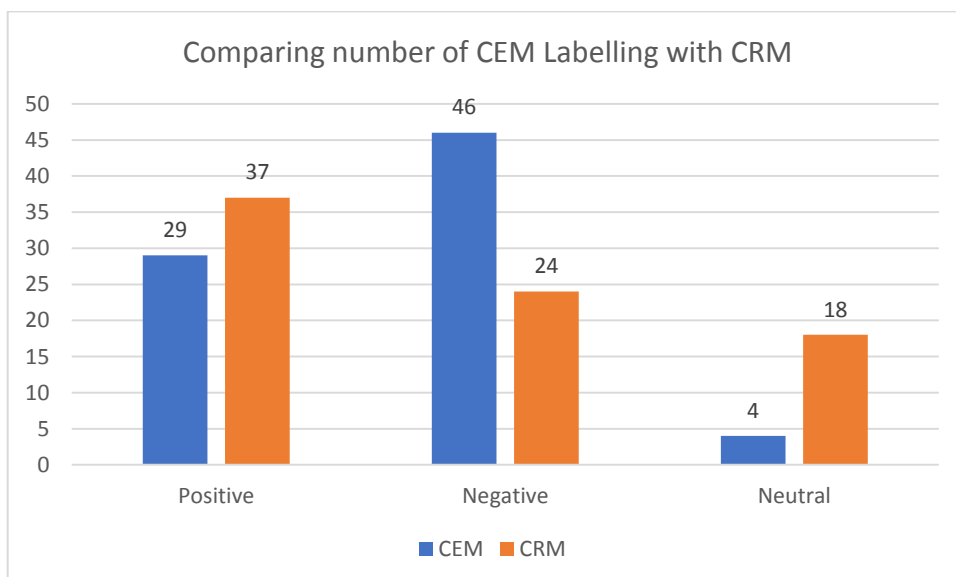


Figure 29: Comparing CEM labelling with CRM labelling for Question 2.

Figure 30 shows the number of tweets with similar labelling between the CRM positive class and to the CEM (positive, negative, and neutral) labelling for Question 2. In CEM, 29 tweets were labelled as positive. By contrast, CRM labelled 11 of these tweets as negative, 10 as positive, and eight as neutral.

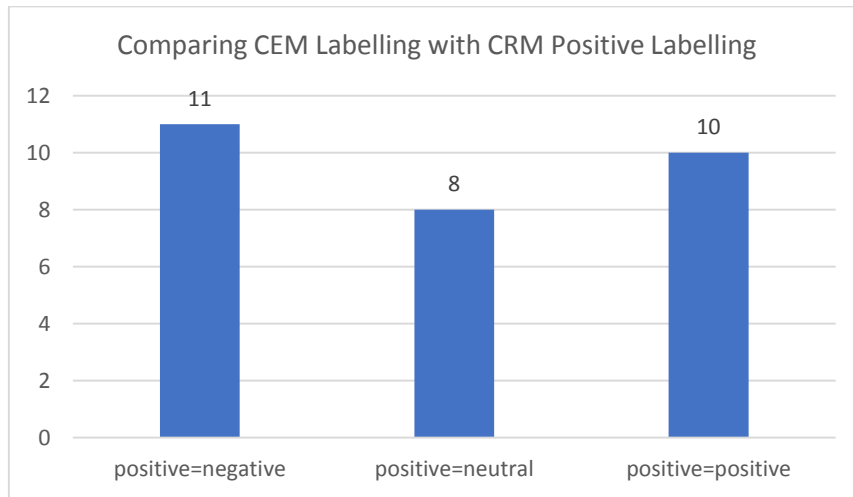


Figure 30: Comparing CEM all labelling classes with CRM positive labelling for Question 2.

Figure 31 shows the number of tweets with similar labelling between the CRM negative class and the CEM (positive, negative, and neutral) labelling for Question 2. In CEM, 46 tweets were labelled as negative. By contrast, CRM labelled 11 of these tweets as negative, 27 as positive, and eight as neutral.

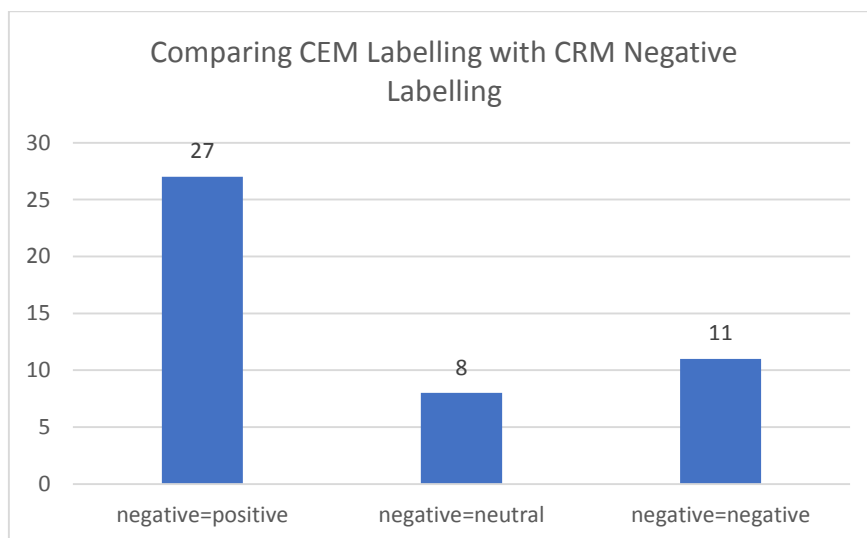


Figure 31: Comparing CEM all labelling classes with CRM negative labelling for Question 2.

Figure 32 shows the number of tweets with similar labelling between the CRM neutral class and the CEM (positive, negative, and neutral) labelling for Question 2. In CEM, only four tweets were labelled as neutral. By contrast, CRM labelled two of these tweets as negative, two as neutral, and none as positive.

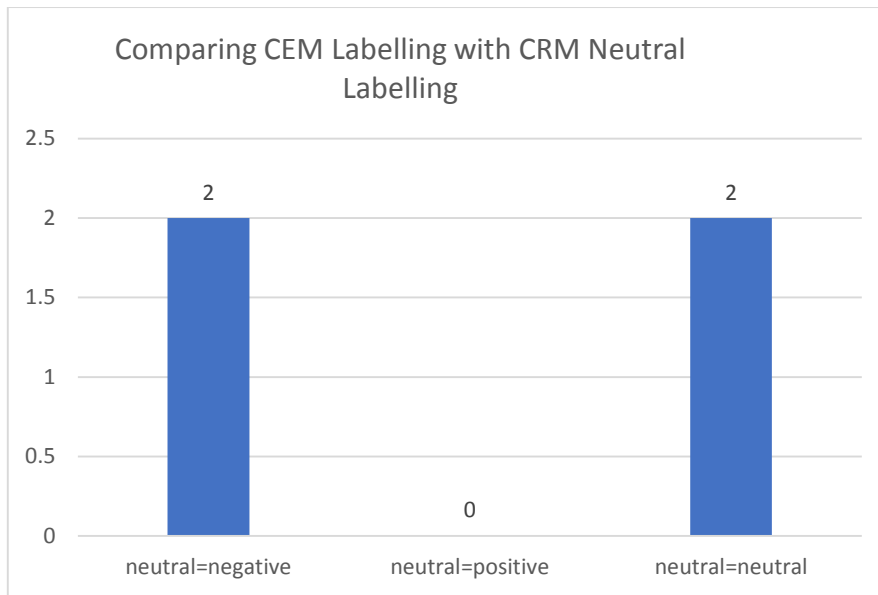


Figure 32: Comparing CEM all labelling classes with CRM neutral labelling for Question 2.

#### 5.3.8.4. Integration Results for Question 3

Figure 33 shows a comparison between CRM labelling with CEM labelling. Twenty tweets were labelled similarly between CRM and CEM. On the other hand, 59 tweets were labelled dissimilarly between CEM and CRM. In other words, 75% of collected tweets were labelled differently between CRM and CEM, and only 25% had similar labelling.

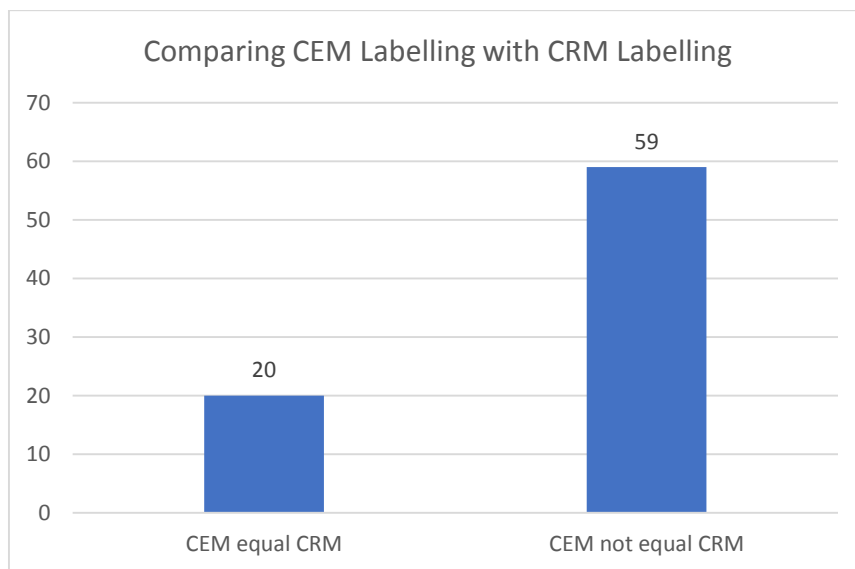


Figure 33: Comparing CEM labelling with CRM labelling for Question 3.

Figure 34 there were 28 positive tweets in CEM, compared to 36 tweets in CRM. In contrast, CEM had only 13 negative tweets, compared to 26 in CRM. The neutral class contained 25 tweets in CEM and 30 in CRM, a difference of only five tweets.

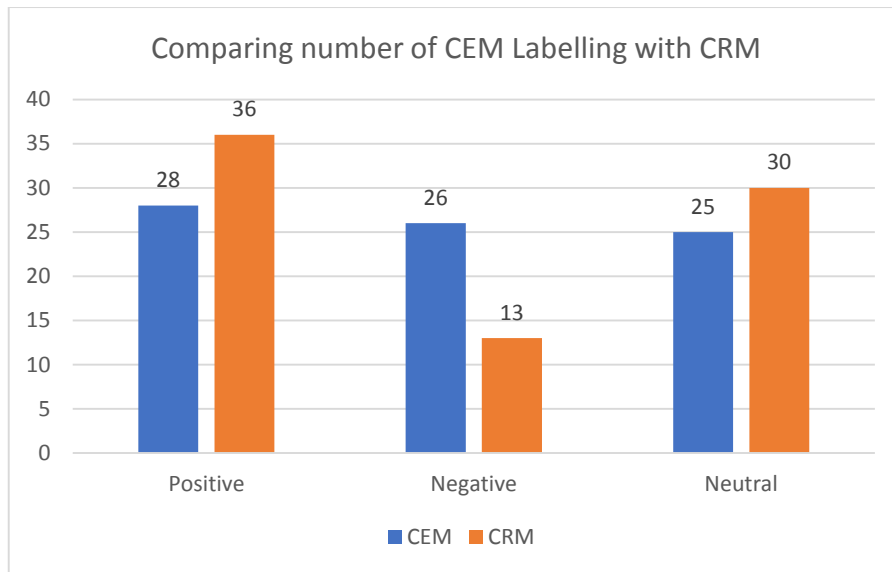


Figure 34: Comparing CEM labelling with CRM labelling for Question 3.

Figure 35 shows the number of tweets with similar labelling between the CRM positive class and the CEM (positive, negative, and neutral) labelling for Question 3. In CEM, only 28 tweets were labelled as positive. By contrast, CRM labelled seven of these tweets as negative, eight as positive, and 13 as neutral.

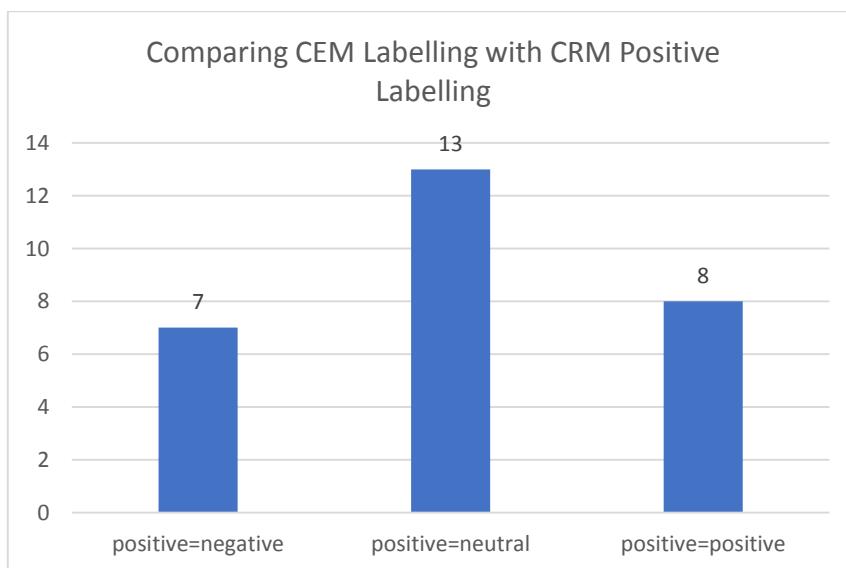


Figure 35: Comparing CEM all labelling classes with CRM positive labelling for Question 3.

Figure 36 shows the number of tweets with similar labelling between the CRM negative class and the CEM (positive, negative, and neutral) labelling for Question 3. In CEM, 26 tweets were labelled as negative. By contrast, CRM only labelled one of these tweets as negative, six as neutral, and 19 as positive. Overall, 73% of the labelling was the exact opposite in CEM and CRM.

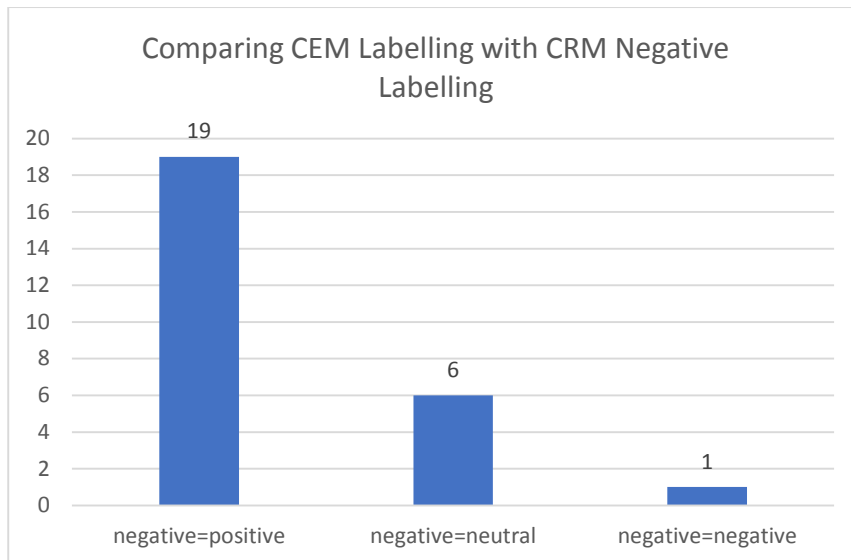


Figure 36: Comparing CEM all labelling classes with CRM negative labelling for Question 3.

Figure 37 shows the number of tweets with similar labelling between the CRM neutral class and the CEM (positive, negative, and neutral) labelling for Question 3. In CEM, 25 tweets were labelled as neutral. By contrast, CRM labelled five of these tweets as negative, 11 as neutral, and nine as positive. Overall, 56% of the tweets labelled neutral in CRM were labelled as either positive or negative in CEM.

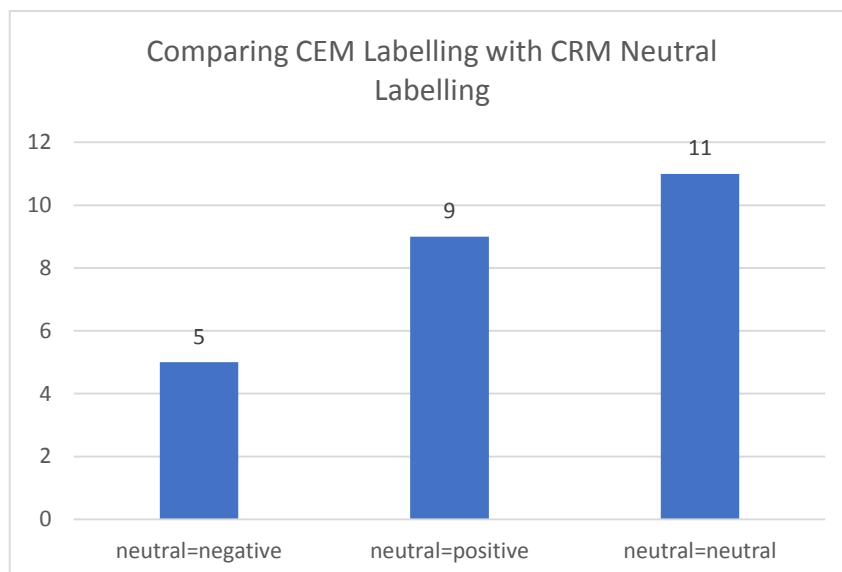


Figure 37: Comparing CEM all labelling classes with CRM neutral labelling for Question 3.

#### 5.3.8.5. Integration Results for Question 4

Figure 38 shows a comparison between CRM labelling and CEM labelling. Twenty-four tweets were labelled similarly between CRM and CEM. On the other hand, 55 tweets were labelled dissimilarly between CEM and CRM. In other words, 70% of the collected tweets were labelled differently between CRM and CEM, and only 30% had the same labelling.



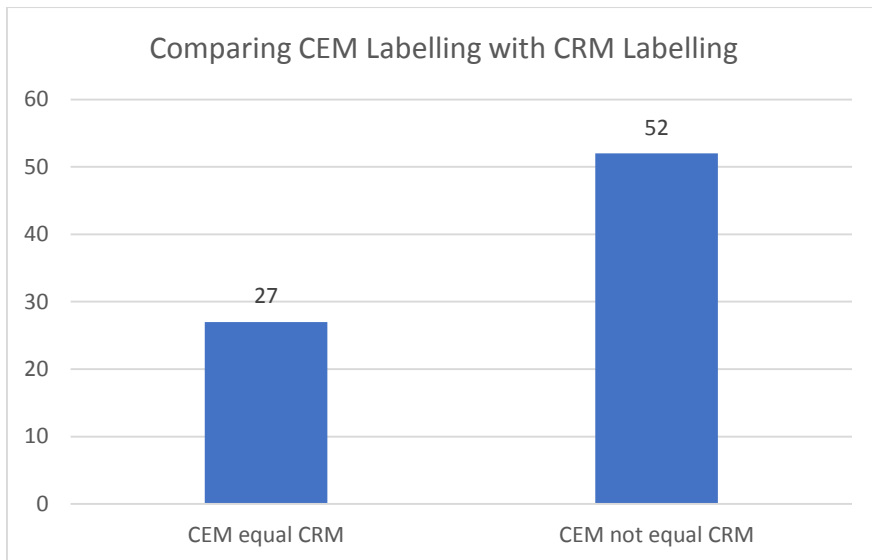


Figure 38: Comparing CEM labelling with CRM labelling for Question 4.

From Figure 39, there were 18 positive tweets in CEM, compared to 31 tweets in CRM. In contrast, there were 26 negative tweets in CEM, compared to 20 negative tweets in CRM. For the neutral class, there were 35 tweets in CEM, compared to 28 in CRM.

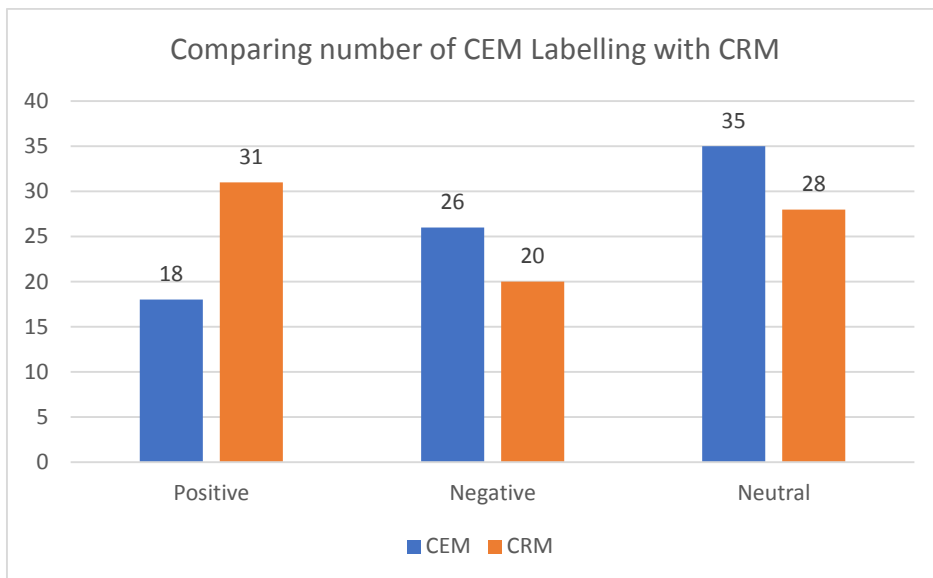


Figure 39: Comparing CEM labelling with CRM labelling for Question 4

Figure 40 shows the number of tweets with similar labelling between the CRM positive class and the CEM (positive, negative, and neutral) labelling for Question 4. In CEM, 17 tweets were labelled as positive. By contrast, CRM labelled five of these tweets as negative, four tweets as neutral, and eight as positive.

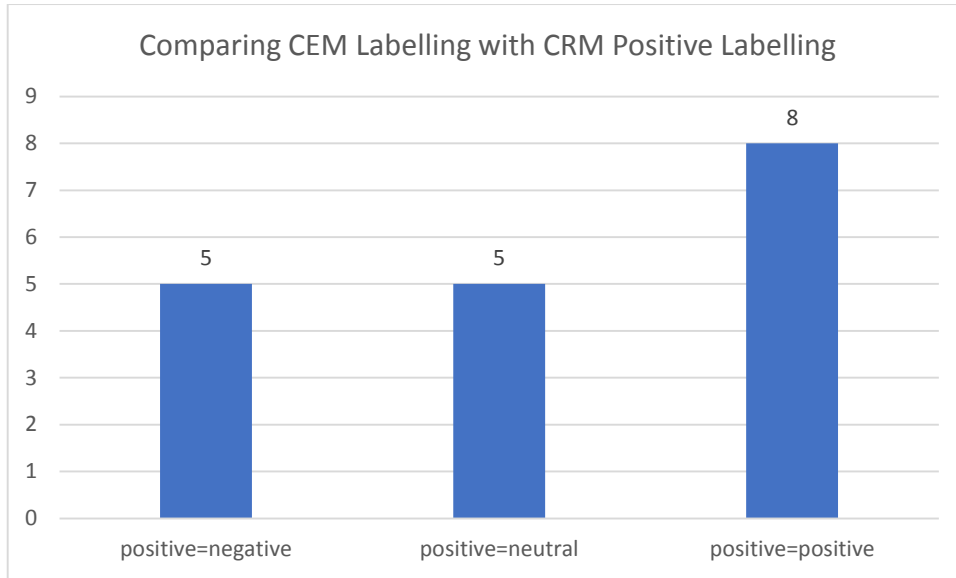


Figure 40: Comparing CEM all labelling classes with CRM positive labelling for Question 4.

Figure 41 shows the number of tweets with similar labelling between the CRM negative class and the CEM (positive, negative, and neutral) labelling for Question 4. In CEM, 26 tweets were labelled as negative. By contrast, CRM labelled five of these tweets as negative, nine as neutral, and twelve as positive.

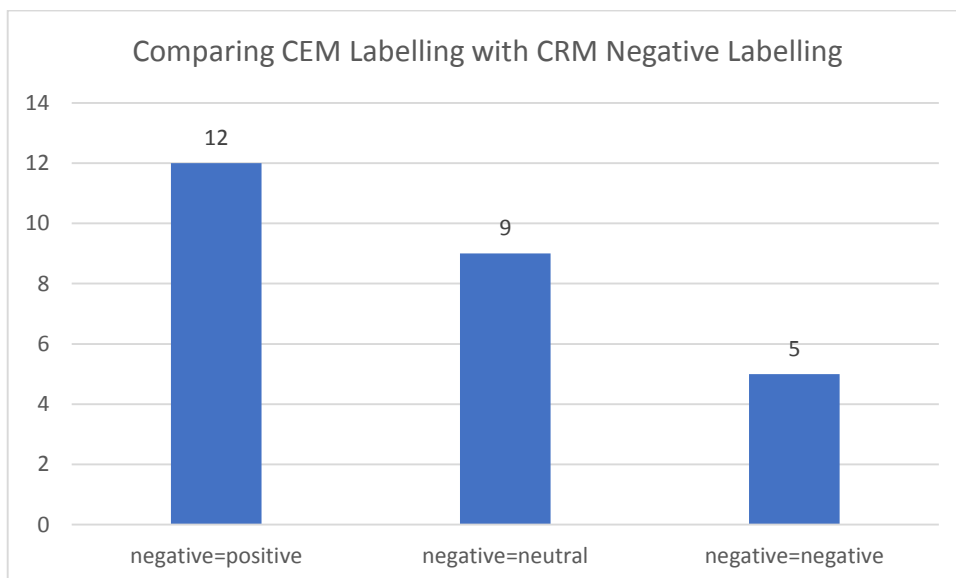


Figure 41: Comparing CEM all labelling classes with CRM negative labelling for Question 4.

Figure 42 shows the number of tweets with similar labelling between the CRM neutral class and the CEM (positive, negative, and neutral) labelling for Question 4. In CEM, 32 tweets were labelled as neutral. By contrast, CRM labelled 10 of these tweets as negative, 11 as neutral, and 11 as positive.

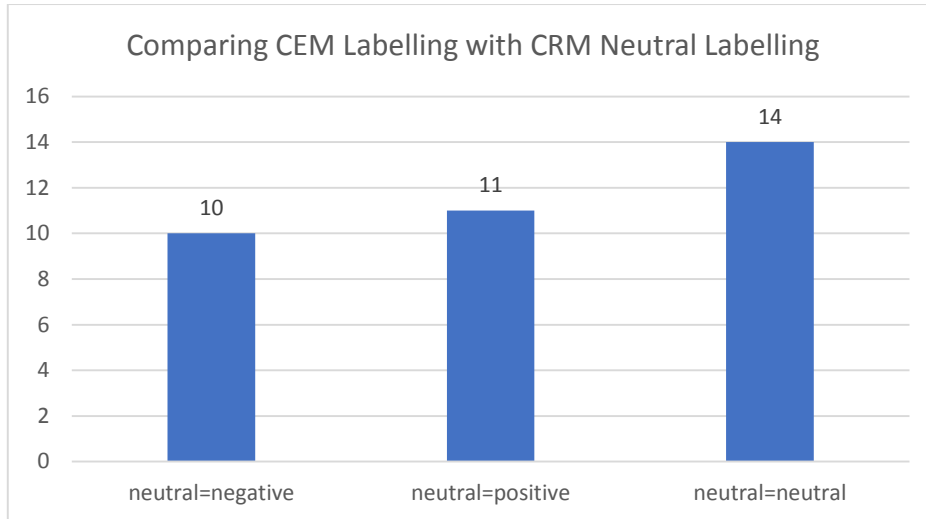


Figure 42: Comparing CEM all labelling classes with CRM neutral labelling for Question 4.

### 5.3.9. Tuning the CRM weights

The aim of this experiment was to tune CRM weights for some aspects of e-learning criteria that students need, according to their social media posts. To achieve this aim, first, CRM weights were updated by changing the values of some attributes that reflected the aspect-level opinions in CEM. After that, the criteria were also updated according to the input of the opinions in CEM. Then, the validation was carried out for the new criteria. Last, classification was carried out in order to compare the original criteria with the updated ones. The weight of CRM changed based on the input from social media. For this study, the values of the criteria were changed for aspects of Q1. Changes occurred for students with Blackboard marks between 30 and 40 and student final exam marks in ODUS from 70 to 60. After updating the weight of CRM, another classification was carried out in order to evaluate the accuracy, precision, and recall of all classes for the SVM and NB classifiers. After that, a comparison was carried to find out the accuracy, precision, and recall of all Classes for the SVM and NB classifiers before and after CRM weight tuning. Table 63 shows that the accuracy before the tuning is higher than after the tuning since the matching between the CRM criteria weights and students' opinions on social media are nearly the same for Q1. This indicates that there was no need to update the criteria weights for the aspects in Q1.

Table 63: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question1.

Tuning	Classifier name	Accuracy	Class recall	Class precision
Before	NB	93.43%	93.95%	94.12%
	SVM	87.25%	86.55%	89.74%
After	NB	92.58%	92.38%	92.59%
	SVM	85.18%	83.37%	84.70%

Table 64 shows the validation of the classification accuracy, precision, and recall for the NB and SVM classifiers for Question 1: Crosse-validation=10, sampling type=stratified sampling. Values of the criteria were changed for aspects in Q2: Changes occurred in student attendance in CRM from seven days to 14 days. Table 64 shows the comparison between the evaluation of the classification before and after the tuning. The results show that the accuracy after the tuning is higher than before for the aspects in Q2.

Table 64: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 2.

Tuning	Classifier name	Accuracy	Class recall	Class precision
Before	NB	95.05%	96.69%	96.48%
	SVM	81.80%	62.82%	55.83%
After	NB	95.85%	96.09%	95.76%
	SVM	84.70%	65.96%	57.98%

Values of the criteria were changed for aspects in Q3: Changes occurred in evaluating the total number of posts, including discussions and tests. Table 65 shows the comparison between the evaluation of the classification before and after the tuning. The results show that the accuracy after the tuning is almost the same.

Table 65: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 3.

Tuning	Classifier name	Accuracy	Class recall	Class precision
Before	NB	98.33%	72.44%	72.31%
	SVM	96.72%	65.44%	64.52%
After	NB	98.35%	72.44%	72.37%
	SVM	95.92%	64.96%	63.92%

Q4 combines the aspects in Q1, Q2, and Q3. Table 66 shows the comparison between the evaluation of the classification before and after the tuning. The results show that the accuracy after the tuning is higher than before. This indicates that changing some aspects of the criteria can help to enhance the classification accuracy in the CRM data.

Table 66: Evaluation for accuracy, precision, and recall with all classes for SVM and NB classifiers for Question 4.

Tuning	Classifier name	Accuracy	Class recall	Class precision
Before	NB	85.13	84.61	85.88
	SVM	83.05	81.47	83.12
After	NB	92.58	92.38	92.59
	SVM	85.18	83.37	84.70

To sum up King Abdul-Aziz University’s validation CRM experiments, the best accuracy was achieved by NB due to the advantages of NB such as its simplicity, ease of implementation, and combination of efficiency and acceptable accuracy, as is stated in Chapter 2.

Figure 43 shows a comparison between the CEM labelling and the distance education criteria with suggestion one for Question 1. The number of positive tweets increases, as does the number of negative tweets, while the number of neutral tweets sharply decreases.

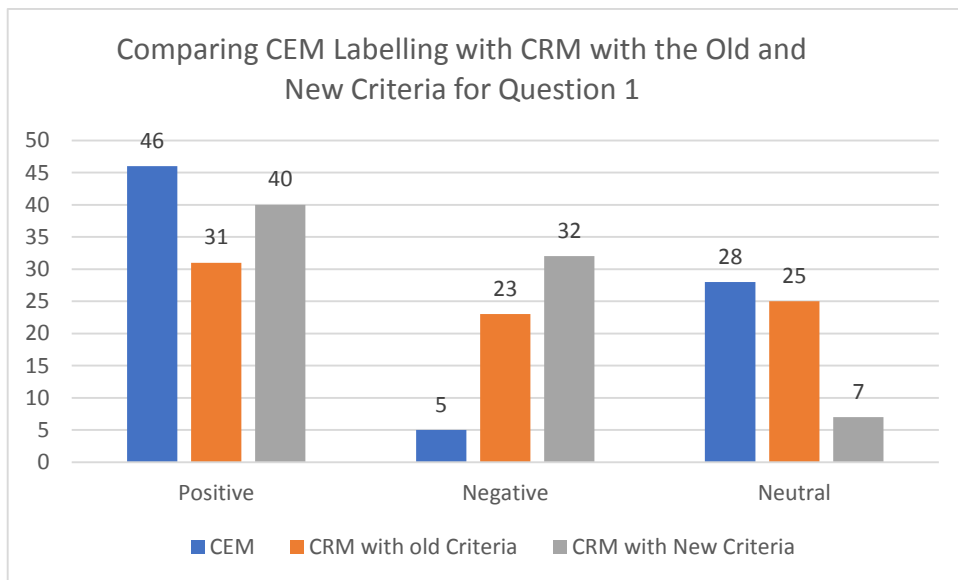


Figure 43: Comparing CEM labelling with CRM with the old and new criteria for Question 1.

Figure 44 shows a comparison between the CEM labelling and the distance education criteria with suggestion one for Question 1. The number of positive tweets increases, and the number of negative tweets decreases, as does the number of neutral tweets.

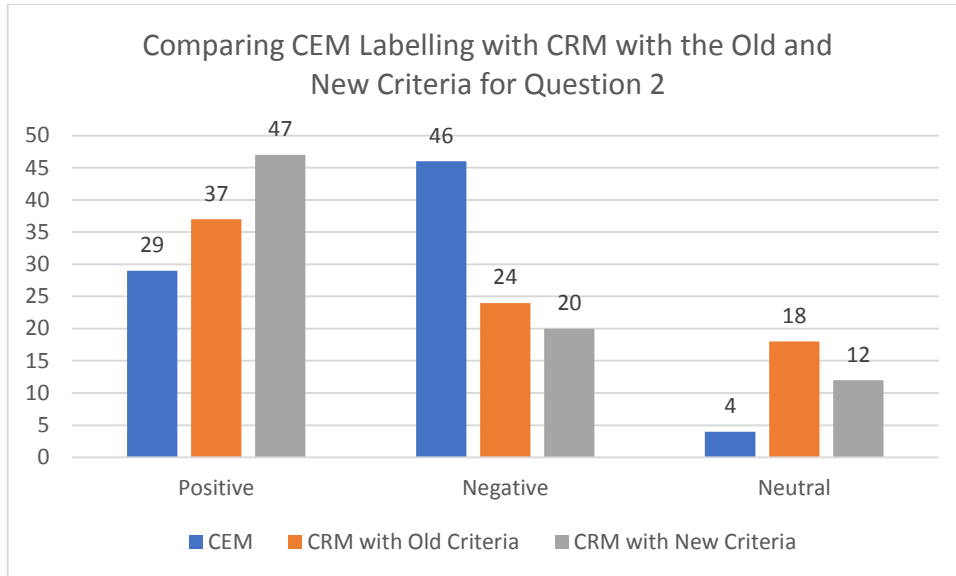


Figure 44: Comparing CEM labelling with CRM with the old and new criteria for Question 2.

Figure 45 shows a comparison between the CEM labelling and the distance education criteria with suggestion one for Question 1. The number of positive tweets approximately doubles, and the number of negative tweets approximately halves. In addition, the number of neutral tweets drops to zero.

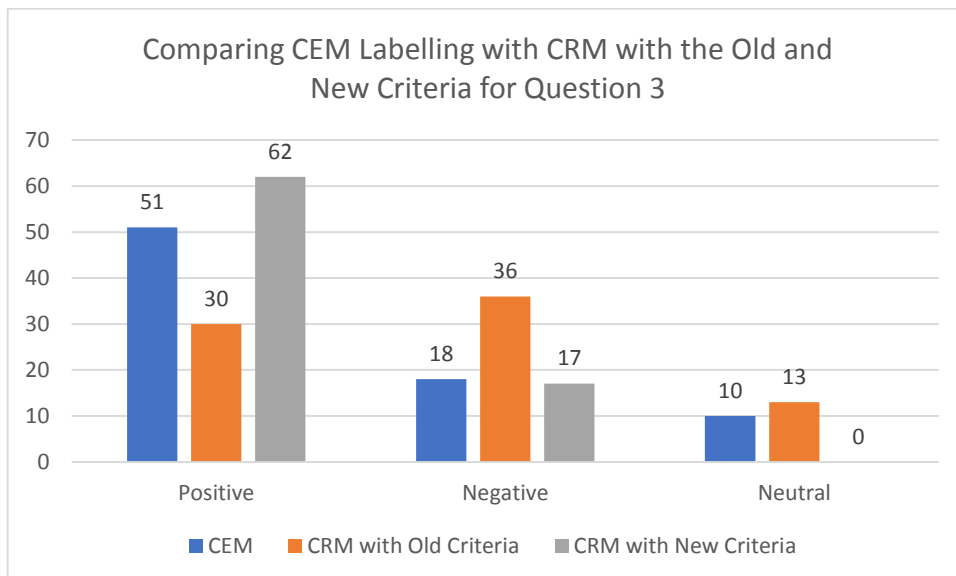


Figure 45: Comparing CEM labelling with CRM with the old and new criteria for Question 3.

Figure 46 shows a comparison between the CEM labelling and the distance education criteria with suggestion one for Question 1. The number of positive and neutral tweets increases, and the number of negative tweets decreases. In conclusion, the result of this experiment shows that social media can support CRM with more details. The main aim of this study was achieved by showing the gap between the criteria and students' needs. The validation result

demonstrates the gap between the students' perspectives and the criteria. This might help universities to adapt the distance education criteria in a way that helps the students to deal with the university. The validation results confirm that tuning the criteria will help students.

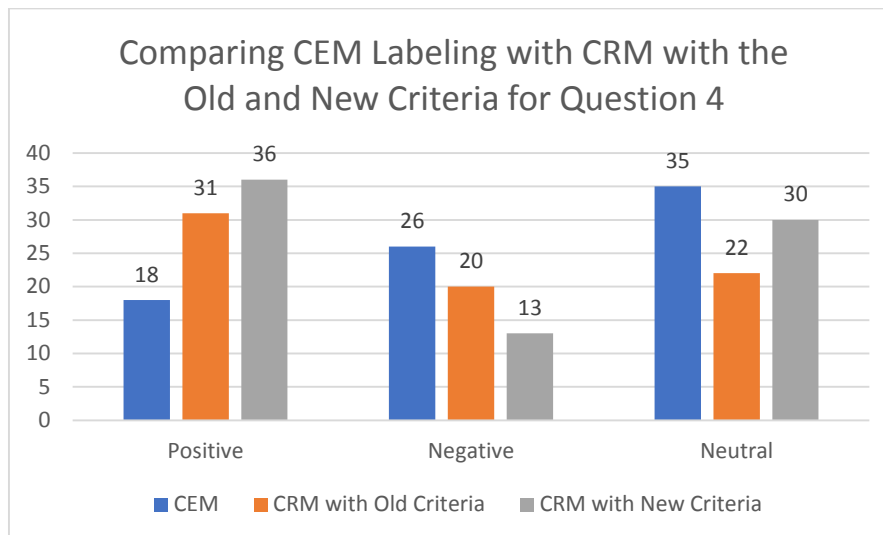


Figure 46: Comparing CEM labelling with CRM with the old and new criteria for Question 4.

#### 5.4. Summary

This chapter has documented the first set of experiments carried out to investigate whether social media could help experts to understand users' perspectives and could support academics' knowledge about their students. Therefore, a framework was proposed for integrating CRM with CEM on the aspect level. The framework consists of three components: sentiment analysis for CEM, classification using CRM Blackboard data, and tuning of CRM via CEM, which integrates both results to study the level of matching between both resources' information, namely, social media and Blackboard data. In other words, there is good consistency between the CRM structure and students' opinions on the aspect level. This takes the study further by investigating King Abdul-Aziz University in e-learning and distance criteria, which brings the similarity between CEM and CRM opinions closer. Moreover, the final stage of this experiment shows an interesting result after changing the e-learning criteria according to the necessary requirements of input on social media requested in students' feedback comments through the Twitter platform. The results show that changing some aspects of the e-learning criteria that were required by students in their social media posts can help to enhance the classification accuracy in the Blackboard data and to understand more students' studying statuses. Furthermore, the results show matching between students' opinions in CRM and CEM, especially in the negative and neutral classes. In other words, applying changes in the criteria in the Blackboard data can help the university to keep students' loyalty and develop its education business in future.

## **Chapter 6: Tuning of CRM via CEM using sentiment analysis on sentence level**

### **6.1. Introduction**

In real-world applications, CRM and CEM might not always have same structures and aspects. CRM could lack a few key aspects important for CEM and vice versa. The difference comes from the structure of the CRM system. In this case, opinions can only be collected on the sentence level for some subjects in the domain. Sentiment analysis can be applied to optimise the structure of CRM by adding the missing key aspects into CRM and removing unhelpful and confusing aspects. This part of the problem will be investigated in this chapter. The framework for adjusting CRM via CEM in Chapter 5 will be revised and validated by a selected case study from service sector.

This chapter is organised as follows: In Section 6.2 Case study for tuning CRM based on CEM at the sentence level is presented. In Section 6.2.1, the background of Mubasher products is presented. Section 6.2.2 contains the preparing the sentiment analysis for CEM using the criteria from CRM. CEM based on sentiment on the sentence level is presented in Section 6.2.3. Section 6.2.4 Sentiment analysis for CEM along with experimental results and evaluations. In Section 6.2.5, Classification for CRM is presented along with results and evaluations. In Section 6.2.6, Tuning CRM via CEM along with experimental results and evaluations. Finally, a summary of this chapter is presented in Section 6.3.

### **6.2. Case study for tuning CRM based on CEM at the sentence level: Optimisation of CRM for Mubasher**

Mubasher is a major financial service company in the Middle East. It has a traditional legacy CRM, which contains very simple information about the users and their transactions and activities. To modernise this CRM, customers have been asked general questions on different aspects regarding their services. The aim of this experiment is to add the missing key aspects to their existing CRM from the user opinions. Due to the difference between the legacy CRM and CEM, the tuning of CRM can only be carried out on the sentence level.

#### **6.2.1. Background of Mubasher's products**

The Capital Market Authority (CMA) was established in order to protect investors and the public in Saudi Arabia from unfair and fraudulent trading practices. Currently, Tadawul is the only Saudi Stock Exchange in Saudi Arabia. The CMA regulates Tadawul, and within Tadawul, there are many sectors, such as banks, financial services, petrochemical industries, retail,



telecommunications, and technology. Tadawul authorises Mubasher to access real-time price dissemination for a wide variety of products. Mubasher produces customised products for all types of investment needs, such as Mubasher Pro 10, Mubasher Pro 9, Mubasher NetPlus, Mubasher iPhone, Mubasher Android Phone, Mubasher iPad, and Mubasher Touch for Android Tab.

### **6.2.2. Preparing the sentiment analysis for CEM using the criteria from CRM**

Twitter is a relatively new method of mass communication that has arguably become popular because it operates in real time and was designed for mobility [188]. Twitter has textual features such as retweet, mention, link or URL, and hashtag, which have been pre-processed during the first stage of the opinion mining model to clean the text before handing it to the classifier. Hashtags are an important feature of social networking sites that can be inserted anywhere within a message: before, within, or after the body of a message [103]. In addition, topic hashtags are used as search key on Twitter's track interface to proactively search Twitter for more tweets belonging to a particular topic [189]. In the following case study experiments, the collection process was based on certain hashtags that represent important topics using Twitter hashtags for the domain.

### **6.2.3. CEM based on sentiment on the sentence level**

In this experiment, a hashtag was created to ask users for opinions and reactions toward Mubasher's products in order to identify positive, negative, and neutral opinions. The sentence level of sentiment analysis was followed to analyse customers' opinions on the Twitter platform. The collection process was based on certain hashtags that represent important events for Mubasher's products, such as “#مباشر\_برو\_15\_عام” (# Mubasher\_Pro\_15 Years).

A similar approach could be extended to CEM analysis, which was introduced in Chapter 5 for the Arabic opinion mining process.

### **6.2.4. Experiment of sentiment analysis for CEM**

During the first stage, a hashtag was created to ask users for opinions and reactions toward Mubasher products in order to identify positive, negative, and neutral opinions. Figure 47 summarises the sentiment analysis process of Arabic tweets. The machine learning approach is a supervised method in which a set of data labelled as positive or negative is represented by feature vectors. After that, the classifier uses those vectors as training data in order to identify the combination of specific features that relate to a specific class.

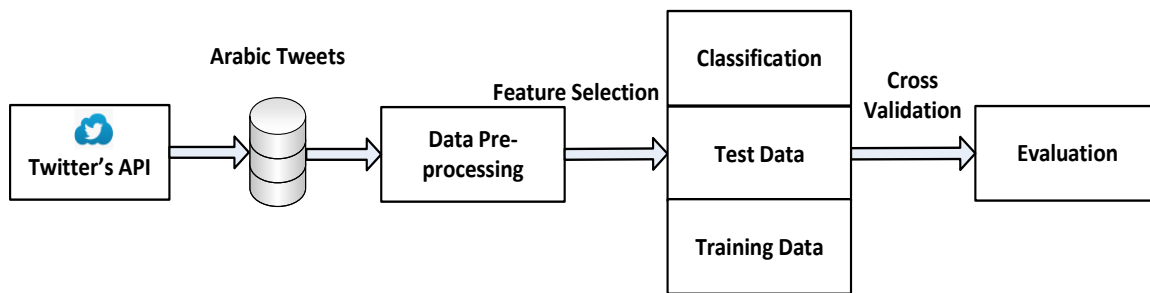


Figure 47: The process of sentiment analysis in Arabic tweets (Mubasher case study)

Since the first task is loading the dataset, the “Retrieve” operator was used; then, “Process documents from data” pre-processed the text using available operators to create the word vector list. After that, a set role operator was used to change the role of the attributes, which is, in this case, labelling the text into positive, negative, and neutral classes. Finally, an X-validation operator was used to evaluate the accuracy. The X-validation operator is a nested operator. It has two sub processes: a training process and a testing process. The training process is used for training a model. Then, the trained model is applied in the testing process. The performance of the model is also measured during the testing phase. RapidMiner was applied for text processing. It has a list of text processing operators. Those operators processed Arabic text in order to create our vector list, such as tokenisation and removal of stop words [170].

#### 6.2.4.1. Experiment parameter settings

The same experiment parameter settings used in Chapter 3 for sentiment analysis for CEM were utilised in this experiment. For example, the classifier settings followed the default values for most of the parameters to get results that are more accurate. Ten-fold cross-validation was used several times to find the best values of these parameters. In addition, both SVM and NB had the same operator parameter settings used.

#### 6.2.4.2. Data description

To collect opinions that were comprehensive for the time on the targeted objectives, tweets from different users on Mubasher’s page were obtained using Twitter’s official developers’ API. The data were collected over a 57-day period between 15 July and 12 September 2015. The data distribution depended on the Twitter API’s methods utilised and the number of tweets posted on Mubasher page. Downloaded tweets were marked manually by humans as positive, negative, and neutral. In addition, the marked tweets are stored in a database for experiments.

The Twitter API has several methods, such as GET statuses / user-timeline, GET statuses / home-timeline, and GET search / tweets, which return a timeline of tweet data. In addition,

timelines can grow very large, so there are limits to how much of a timeline a client application may fetch in a single request.

The preliminary data were received from the company and it contained two attributes (Tweet-Text, data-type=text, Tweet-Label, data-type=text). The RapidMiner wizard was used to import the dataset; RapidMiner studio uses strongly typed attributes. The Tweet-Text data type remained the same. On the other hand, the Tweet-Label data type changed to nominal, since nominal scales were used for labelling variables, without any quantitative value [190]. After that, the data were loaded, and the set role operator was used to change the role of one or more attributes to specify the label columns for RapidMiner.

#### 6.2.4.3. Data pre-processing

The dataset included 2,051 tweets. A selection of over 20 words and idioms in Arabic from the emotion corpus (including “liberality”, “growth”, “magnificent”, “abruption”, “ineligible”, and “not accurate”) was then formed into the following three classes: positive, negative, and neutral. The most common Arabic words in the standard Saudi dialect representing the positive and negative classes are in Table 67.

Table 67: Common Arabic words in standard Saudi dialect representing the classes.

Arabic Positive Sentiment	English Translation	Arabic Negative Sentiment	English Translation
عطاء	liberality	انقطاع	abruption
تطور	growth	مشكلة	trouble
جميل	magnificent	بطيئ	slow
الأكثر ترشيحا	most rated	برنامج الاتجاه أفضل	Etjah product better
جيد	good	برنامج تكرشارت أفضل	ticker chart product best
متكامل	complementary	لم استقد	uninteresting
محترف	professional	مو مضبوط	not accurate
ممتاز	excellent	غالي السعر	the price is high
سريع جدا	very fast	غير جدير	ineligible
منصة احترافية	professional platform	فاشل	unsuccessful

Two Mubasher employees, who have experience with Mubasher products, labelled the data manually. Positive tweets were given the label "1", while negative tweets were given the label "-1". Neutral tweets were given the label "0", and irrelevant tweets were deleted from the database. Table 68 shows the number of tweets in our database; 727 tweets were deleted in the labelling process.



- Filter token by length removes useless words, and it is set to two.
- The N-gram feature is a subsequence of N items from a given text.

Then, the 'Process Documents from Data' operator generates a word vector from the dataset after pre-processing and represents the text data as a matrix to show the frequency of occurrence of each term [16]. TF-IDF (term frequency–inverse document frequency) and BTO (binary term occurrence) were used as weighting schemes to create the word vectors, after which relevant data mining approaches, i.e., NB and SVM, were explored for building models that classify tweets according to their sentiment polarity into positive, negative, and neutral. Finally, the evaluation was carried out using precision and recall methods.

#### 6.2.4.4. Experimental results and evaluations

The proposed model for sentiment analysis of Saudi Arabic (standard and Arabian Gulf dialects) tweets was applied to extract feedback about Mubasher products. A hybrid of natural language processing and machine learning approaches for building CEM were used to classify tweets according to their sentiment polarity into one of the classes of positive or negative. First, a document's pre-processing was explored on the dataset. Second, NB and SVM were applied with different weighting schemes (TF-IDF and BTO). Third, the proposed model for sentiment analysis was expanded to obtain the results for the N-gram's term of tokens, which is a subsequence of N items from a given text [16]. Finally, humans labelled the data, which may have introduced some mistakes into the labelling process. At this moment, a neutral class with generalisation of our classification will improve results' classification accuracy.

The following tables show the type of tuning to provide the best accuracy in sentiment analysis for Arabic text in relation to Mubasher's products. Table 70 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 70: Accuracy, precision, and recall for positive and negative classes for SVM and NB classifiers without N-gram.

Classes	Classifier name	Weighting schemes	Accura cy	Class recall	Class precision
Positive and negative	NB	BTO	83.12%	82.23%	80.83%
		TF-IDF	82.50%	80.81%	80.17%
	SVM	BTO	83.83%	77.26%	85.81%
		TF-IDF	84.90%	79.10%	86.34%

Table 71 shows the class precision and recall for the NB classifier with the N-gram feature set to two.

Table 71: Accuracy, precision, and recall for positive and negative classes for SVM and NB classifiers with N-gram

Classes	Classifier name	Weighting schemes	Accuracy	Class recall	Class precision
Positive and negative	NB	BTO	84.19%	81.92%	82.25%
		TF-IDF	83.92%	81.52%	82.00%
	SVM	BTO	83.93%	76.38%	88.20%
		TF-IDF	84.55%	77.19%	89.03%

To sum up the first experiments, the best accuracy was achieved by SVM without the N-gram feature. On the other hand, almost the same accuracy was attained by NB and SVM when the N-gram feature was involved.

ROC graphs are a utilisable technique to evaluate and visualise both classifiers' performance. The SVM and NB classifiers with the N-gram feature were chosen to visualise their classification performance due to the better results in terms of accuracy, precision, and recall for the positive and negative classes only. Figure 48 and Figure 49 show the performance of both techniques when TF-IDF and BTO weighting schemes were used. Moreover, it was decided to choose neutral selections, which could be a better option to represent the ROC graph.

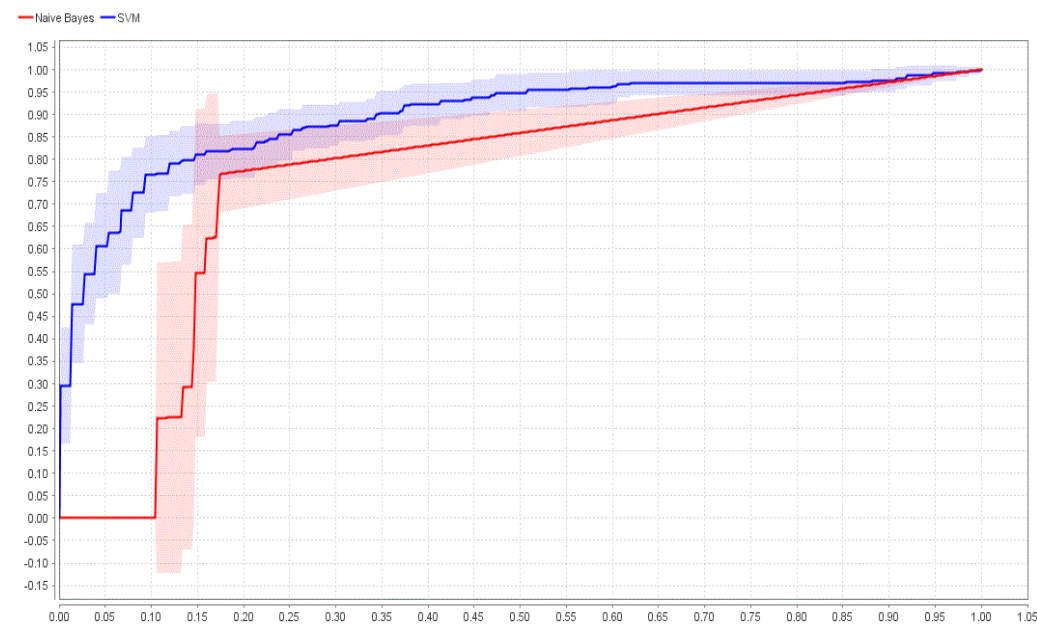


Figure 48: Roc for BTO weighting scheme for both classifiers (NB in red, SVM in blue)

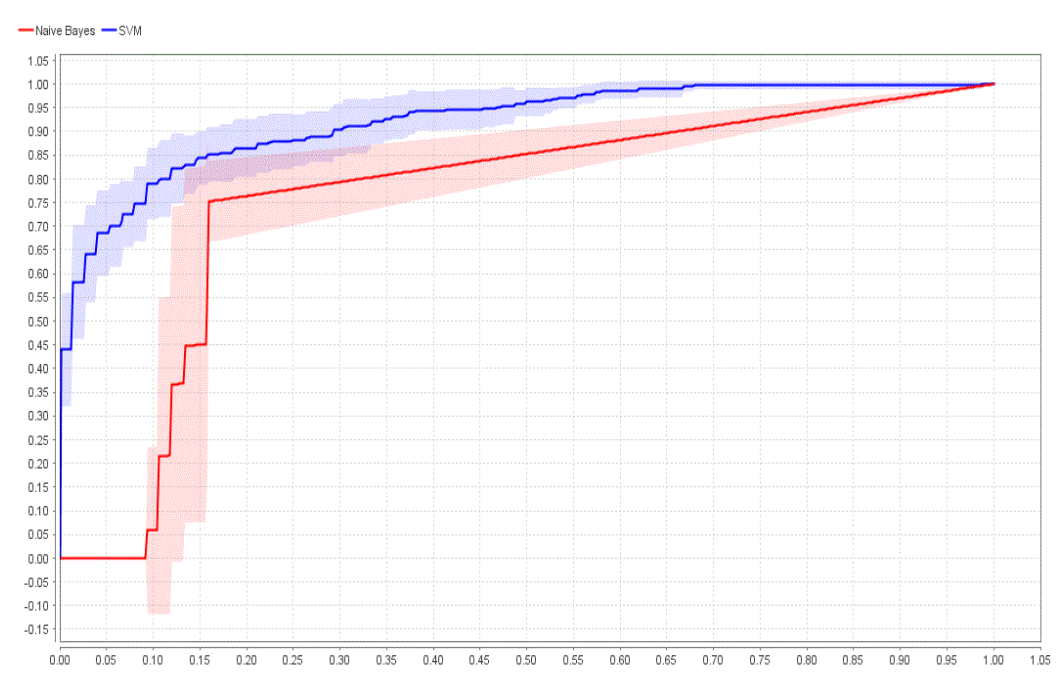


Figure 49: Roc for TF-IDF weighting scheme for both classifiers (NB in red, SVM in blue)

Moreover, humans labelled the data, and this may have introduced some mistakes into the labelling process; only for this reason, generalisation of our classification with a neutral class will show the importance of a neutral class in sentiment analysis by improving the results of classification accuracy. The neutral class was added here to be classified among the positive and negative classes. The SVM technique paid attention to the new neutral class.

Table 72 shows the class accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 72: Accuracy, precision, and recall for all classes using SVM and NB classifiers without N-gram

Classes	Classifier name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	67.30%	63.13%	61.17%
		TF-IDF	67.52%	62.52%	60.99%
	SVM	BTO	74.09%	59.51%	72.91%
		TF-IDF	76.74%	64.40%	77.26%

Table 73 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 73: Accuracy, precision, and recall for all classes using SVM and NB classifiers with N-gram

Classes	Classifier name	Weighting schemes	Accuracy	Class recall	Class precision
All Class	NB	BTO	71.90%	64.90%	64.82%
		TF-IDF	72.05%	64.48%	64.85%
	SVM	BTO	74.62%	58.91%	80.17%
		TF-IDF	75.76%	59.89%	81.56%

In conclusion, the experiment shows that NB performance was better when we used the N-gram feature with both schemas (TF-IDF and BTO). On the other hand, there was a slight performance drop with TF-IDF and rise with BTO when SVM used the same feature. Furthermore, both experiments demonstrate the robustness of the proposed data model. Using social media input, the proposed model could extract key features required by CEM with good accuracy.

In general, the lift chart is a way to discretely represent and visualise classifier performance. In this stage, it was decided to generate a lift chart for the neutral class with SVM using the TF-IDF schema without the N-gram feature as result of its better accuracy. Figure 50 shows a depiction of the performance of the SVM with TF-IDF without the N-gram feature using a lift chart.

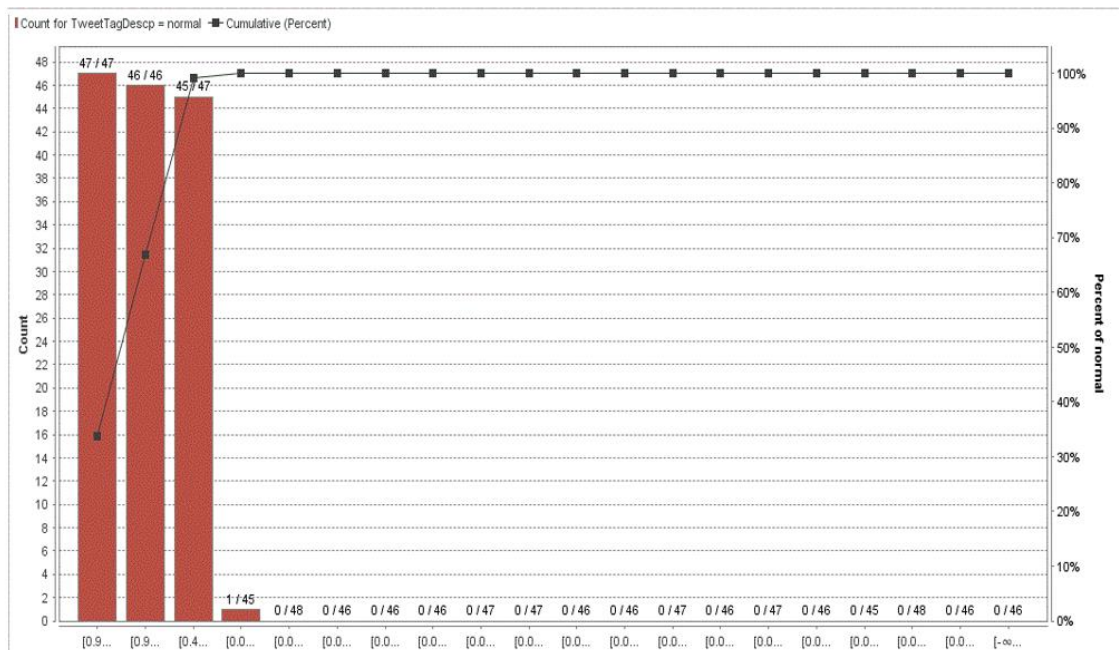


Figure 50: Lift chart for the neutral class for SVM using the TF-IDF schema



### 6.2.4.5. Reproducing the experiment

The main purpose of the existing experiment was to demonstrate the success of the proposed method on sentiment analysis of the Arabic language, particularly when the neutral class is added. This is essential for the success of the first proposed component. Moreover, the same model found a clear relationship between Twitter messages and the Saudi Stock Market Index (see Chapter 4). Regarding the comparison between SVM and the Bayesian method, we have split the data into two independent subsets from different periods, and the experiments were carried out for each subset respectively.

#### **First part:**

Table 74 shows the class accuracy, precision, and recall for the NB and SVM classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 74: Number of tweets

Total	Positive	Negative	Neutral
665	189	377	99

Table 75 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 75: NB & SVM for all classes

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision	Classification Error
All Class	NB	BTO	69.95	65.11	63.80	30.05
		TF-IDF	70.10	63.63	62.94	29.90%
	SVM	BTO	78.07	62.45	80.24	21.93%
		TF-IDF	78.53	63.45	78.09	21.47%

Table 76 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature for all classes.

Table 76: NB & SVM with N-gram for all classes

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision	Classification Error
All Class	NB	BTO	73.41	67.18	66.38	26.59%
		TF-IDF	73.56	66.20	65.90	26.44%
	SVM	BTO	75.97	58.66	76.00	24.03%
		TF-IDF	75.67	58.03	75.52	24.33%

**Second part:**

Table 77 shows the class accuracy, precision, and recall for the NB and SVM classifiers without the N-gram feature: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 77: Number of tweets

Total	Positive	Negative	Neutral
666	189	378	99

Table 78 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature set to two: Crosse-validation=10, sampling type=stratified sampling, prune=none.

Table 78: NB & SVM for all classes

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision	Classification Error
All Class	NB	BTO	59.87	56.62	57.02	40.13%
		TF-IDF	60.93	57.11	57.26	39.07%
	SVM	BTO	68.40	59.45	67.23	31.60%
		TF-IDF	69.76	61.04	68.09	30.24%

Table 79 shows the class accuracy, precision, and recall for the NB and SVM classifiers with the N-gram feature for all classes.

Table 79: NB & SVM with N-gram for all classes

Classes	Classifier Name	Weighting schemes	Accuracy	Class recall	Class precision	Classification Error
All Class	NB	BTO	64.90	60.44	60.57	35.10%
		TF-IDF	64.74	59.82	59.90	35.26%
	SVM	BTO	66.57	56.30	60.46	33.43%
		TF-IDF	69.92	59.51	71.95	30.08%

At this stage, the data in this experiment were split into two independent parts for different time periods, and the experiment was reproduced for each part to compare the NB and SVM methods in terms of accuracy with the N-gram feature and the neutral class as well as in their absence. The analysis shows similar results when we split the data, though adding a neutral

class yielded different results in classification accuracy. The author believes that the reason the results are different is that the pre-processed text data consist of all the words that belong to the positive and negative classes as well. However, the best accuracy was attained by SVM in both cases, with the absence and presence of the N-gram feature.

#### 6.2.4.6. Relabelling experiment

Chapter 4 mentioned that there was a need to construct a relabelling framework for Arabic tweets to remove noise in the labelling. This would lead to updating experts' knowledge about labelling and may lead to better classification. The relabelling process starts by representing the text data as a matrix to show the frequency of occurrence of each term within the three classes. The relabelling process focuses on representing the correlation between the feature selection and other terms with highly frequent occurrence for just two nodes within all classes' data. Table 80 shows the feature "the problem" (المشكلة) as a negative sentiment in Mubasher's product review domain. In addition, Table 80 shows the occurrence of the feature "the problem" (المشكلة) in the positive, negative, and neutral classes.

Table 80: Occurrence of the feature "the problem" (المشكلة)

Feature	Occurrence	Neutral	Positive	Negative
المشكلة	161	6	29	126

Figure 51 shows the association rules related to the feature "the problem" (المشكلة) in all classes with respect to the minimum support and minimum confidence threshold. In addition, Figure 51 shows the most important rule for the feature "the problem" (المشكلة), which is [المشكلة] --> [مباشر] (support: 0.0;17 confidence: 1). The term "Mubasher" was correlated with the feature "the problem" [المشكلة] to compose the negative phrase "problem of Mubasher" in a sentence.



Figure 51: Visualisation of the association rules for the feature "the problem" (المشكلة)

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases related with the feature "the problem" (المشكلة). Table 81 shows that the phrase "problem of Mubasher" [المشكلة-مباشر] occurred 25 times: three in the neutral, four in the positive, and 18 in the negative class.

Table 81: Phrases for the feature “the problem” (المشكلة) in all data

Phrase -Terms	Occurrence	Neutral	Positive	Negative
المشكلة-مباشرة	25	3	4	18

Since the phrase “problem of Mubasher” [المشكلة-مباشرة] occurred in the positive and neutral classes, both classes become questionable. Therefore, the feature “the problem” (المشكلة) needs further investigation to find out the association rules in both classes. As a result, two scenarios will be followed: The first scenario is extracting association rules that occurred for the feature “the problem” (المشكلة) in the neutral class. The second scenario is extracting association rules that occurred for the feature “the problem” (المشكلة) in the positive class. Association rules are generated with regard to the minimum support and minimum confidence thresholds using the previous process of visualisation in Chapter 4.

Figure 52 shows that the feature “the problem” (المشكلة) can occur with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the first scenario, the interested rule here is [المشكلة] --> [المشكلة-مباشرة] (support: 0.005 confidence: 1), which represents the phrase “problem of Mubasher” [المشكلة-مباشرة] illustrated in the wordlist matrix in the neutral class.

No.	Premises	Conclusion	Support	Confidence
6973	بيت_مباشر	المشكلة	0.005	1
6974	بيت	المشكلة	0.005	1
6975	المشكلة_شخص	المشكلة	0.005	1
6976	المشكلة	المشكلة	0.005	1
6977	المشكلة_بايت	المشكلة	0.005	1
6978	المشكلة_مباشرة	المشكلة	0.005	1
6979	المشكلة_لبيهم	المشكلة	0.005	1
6980	المشكلة_بيت	المشكلة	0.005	1
6981	المشكلة_الإمداد	المشكلة	0.005	1
6982	المشكلة_لنظارات	المشكلة	0.005	1
6983	المشكلة_وطني	المشكلة	0.005	1
6984	المشكلة_جدار	المشكلة	0.005	1

Figure 52: The correlation rules of the feature “the problem” (المشكلة) in the neutral class

According to Figure 52, the rule [المشكلة] --> [المشكلة-مباشرة] (support: 0.005 confidence: 1), which represents the phrase “problem of Mubasher”, occurred in the neutral class. Therefore, the next step is to search for the phrase “problem of Mubasher” [المشكلة-مباشرة] in the neutral class documents. Table 82 shows that the phrase “problem of Mubasher” [المشكلة-مباشرة] occurred in three documents.

Table 82: Phrase “problem of Mubasher” [المشكلة-مباشر] in the neutral class documents

Original labelling class	Original Arabic tweets with English translation	New labelling class
Neutral	مشاكل مباشر اغلب الاحيان تكون بسبب مزود خدمة الانترنت <u>Mubasher problems</u> are often caused by an ISP	Negative
Neutral	مع مباشر ياليت ترسل لي رقم التواصل ع الخاص اذا استمرت <u>المشكلة</u> If the <u>Mubasher problem</u> continues please send me your contact number	Negative
Neutral	لحل هذه <u>المشاكل</u> بحيث يكون مباشر متصل وابدا خطوه خطوه وهي اضغط 1 ابدا 2 لوحة التحكم 3 جدار الحماية To resolve these <u>problems</u> so that it is <u>Mubasher</u> connected and never takes a step, press 1 Start 2 Control Panel 3 Firewall	Negative

Therefore, they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from their structure that the above documents have negative meanings, so the right class for those documents is the negative class. Figure 53 shows that the feature “the problem” (المشكلة) can occur with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [المشكلة-مباشر] --> [المشكلة] (support: 0.017 confidence: 1), which represents the phrase “problem of Mubasher” [المشكلة-مباشر] illustrated in the wordlist matrix in the positive class.

Premises	Conclusion	Support	Confidence
ونحن	المشكلة	0.005	0.667
فيه	المشكلة	0.005	0.667
عندك	المشكلة	0.005	0.667
حل	المشكلة	0.005	0.667
الجديدة	المشكلة	0.005	0.667
الجزيري	المشكلة	0.005	0.667
الآن مباشر	المشكلة	0.005	0.667
المشكلة مباشر	المشكلة	0.008	1
بعمل اليوم	المشكلة	0.005	1
بعمل فيه	المشكلة	0.005	1
وسرعة الجواب	المشكلة	0.005	1
وسرعة	المشكلة	0.005	1
ونحن لسنا	المشكلة	0.005	1

Figure 53: The correlation rules of the feature “the problem” (المشكلة) in the negative class

According to Figure 53, the rule [المشكلة-مباشر] --> [المشكلة] (support: 0.008 confidence: 1), which represents the phrase “problem of Mubasher”, occurred in the positive class. Therefore, the next step is to search for the phrase “problem of Mubasher” [المشكلة-مباشر] in the positive class documents. Table 83 shows that the phrase “problem of Mubasher” [المشكلة-مباشر] occurred in four documents.

Table 83: Phrase “problem of Mubasher” [المشكلة-مباشر] in the positive class documents

Original labelling class	Original Arabic tweets with English translation	New labelling class
Positive	كان فيه تحديث عالبرنامج وبعدين اتصال مافيه اي مشاكل واحب اضيف كمان افضل برنامج حتى الان مباشر برو It was a great update the program and then the connection has no <u>problems</u> and I would like to, the best program so far is <u>Mubasher Pro</u>	Positive
Positive	كنت اعاني من مشاكل عده في مباشر برو ولكن الان لم يعد للمشاكل وجود بتحميل هذا البرنامج الى الجهاز وتنصيبه ثم الدخول الى مباشر برو I was having several <u>problems</u> with <u>Mubasher Pro</u> but now no longer having the same <u>problems</u>	Positive
positive	مباشر عشرة على عشرة لكن الاتصال له دور ممكن اتصالك بطيئ اتصالي قوي ولا يوجد اي مباشر برو مشاكل مع <u>Mubasher</u> ten is good but the connection has a role as possible Your connection is slow.my internet communication is strong and there are no <u>problems</u> with <u>Mubasher Pro</u>	Positive
positive	بخصوص مباشر فهو الشيء الوحيد الايجابي في سوقنا عندما ننظر لموقع تداول الرسمي او بنوكنا نجده متميزا في خدمته وان وجد له بعض المشاكل فهي قليلة Regarding Mubasher is the only thing positive in our market when we look at Tadawul official website or our banks, we find <u>Mubasher</u> outstanding in his service and if it has some <u>problems</u> it is few	Positive

Therefore, they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from their structure that the above documents have positive meanings, so the right class for those documents is the positive class.

The second example to verify our experiment with another sentiment was mentioned in Chapter 2—comparative sentiment analysis—and the goal of this sentiment is to identify the sentences that contain comparative opinions. Therefore, “better” (افضل) is comparative sentiment to determine which product users prefer to use in this domain. Table 84 shows the feature “better” (افضل); as a comparative sentiment, it can be either positive or negative. In addition, Table 84 shows the occurrence of the feature “better” (افضل) in the positive, negative, and neutral classes.

Table 84: Occurrence of the feature “better” (افضل)

Feature	Occurrence	Neutral	Positive	Negative
افضل	103	5	41	57

Figure 54 shows the association rules related to the feature “better” (افضل) in all classes with respect to the minimum support and minimum confidence threshold. In addition, Figure 54 shows the most important rules for the feature “better” (افضل), such as [افضل] --> [مباشر] (support: 0.011 confidence: 1. The term “Mubasher” correlated with the feature “better” [افضل] to compose positive phrases in the sentences indicating that the user prefers Mubasher’s products.



Figure 54: Visualisation of the association rules for the feature “better” (افضل)

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases related to the feature “better” (افضل). Table 85 shows phrases that occurred in all classes with the feature “better” (افضل).

Table 85: Phrases for the feature “better” (افضل) in all data

Phrase -Terms	Occurrence	Neutral	Negative	Positive
مباشر- افضل	25	1	11	12

From Table 85, the phrase “Mubasher is better” [مباشر-افضل] occurred in the positive, negative and neutral classes, and the feature “better” (افضل) is a positive sentiment towards Mubasher’s product. As a result, the neutral and negative classes become questionable. Therefore, the feature “better” (افضل) needs further investigation to find out the association rules in both classes. As a result, two scenarios will be followed: The first scenario is extracting association rules that occurred for the feature “better” (افضل) in the neutral class. The second scenario is extracting association rules that occurred for the feature “better” (افضل) in the negative class. Association are rules generated about the minimum support and minimum confidence thresholds using the previous process of visualisation.

Figure 55 shows that the feature “better” (افضل) can occur with many rules that appeared in the premises column with the minimum support and minimum confidence values. However,

according to the first scenario, the interested rule here is [افضل] --> [مباشر- افضل] (support: 0.005 confidence: 1), which represents the phrase “users prefer Mubasher’s product” [مباشر- افضل] illustrated in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
مركب	افضل	0.005	1
مباشر_كجا	افضل	0.005	1
سوق_رس	افضل	0.005	1
رسو_التشارك	افضل	0.005	1
رايك_افضل	افضل	0.005	1
خبراء_برنامج	افضل	0.005	1
خبراء	افضل	0.005	1
برامج_مركب	افضل	0.005	1
ايهم_افضل	افضل	0.005	1
ايهم	افضل	0.005	1
ان_سوق	افضل	0.005	1
افضل_مباشر	افضل	0.005	1
افضل_برنامج	افضل	0.005	1
افضل_ان	افضل	0.005	1

Figure 55: The correlation rules of the feature “better” (افضل) in the neutral class

According to Figure 55, the rule [افضل] --> [مباشر- افضل] (support: 0.005 confidence: 1), which represents the phrase “users prefer Mubasher’s product”, occurred in the neutral class. Therefore, the next step is to search for the phrase “users prefer Mubasher’s product” [مباشر- افضل] in the neutral class documents.

right class for that documents.

Table 86 shows the phrase “users prefer Mubasher’s product” [مباشر- افضل] occurred in one document. Therefore, it was resent to the expert who labelled the document in the first stage. Moreover, it can be seen from its structure that the above document has a meaning of discussion among users and comparison between three products, which makes the neutral class the right class for that documents.

Table 86: Phrase “users prefer Mubasher’s product” [مباشر- افضل] in the neutral class documents

Original labelling class	Original Arabic tweet with English translation	New labelling class
Neutral	أيهم أفضل لمن جرب مباشر والاتجاه برو و التكرتشارت	Neutral
	Which is <u>better</u> for those who try <u>Mubasher</u> and Ettijah Program and tekiertchart	



Figure 56 shows that the feature “better” (افضل) can occur with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [افضل] --> [مباشر- افضل] (support: 0.011 confidence: 1), which represents the phrase “users prefer Mubasher’s product” [مباشر- افضل] illustrated in the wordlist matrix in the negative class.

Premises	Conclusion	Support	Confidence
لجيت_التراك	الفضل	0.003	0.667
لجيت	الفضل	0.003	0.667
صبيط	الفضل	0.003	0.667
سوف_مباشر	الفضل	0.003	0.667
سفت	الفضل	0.003	0.667
توجد	الفضل	0.003	0.667
التراكك	الفضل	0.003	0.667
الجا_الفضل	الفضل	0.013	1
بكتير	الفضل	0.012	1
الفضل_مباشر	الفضل	0.011	1
تكرشات_الفضل	الفضل	0.007	1
الفضل_بكتير	الفضل	0.007	1

Figure 56: The correlation rules of the feature “better” (افضل) in negative class

The next step is to search for the phrase “users prefer Mubasher’s product” [مباشر- افضل] in the negative class documents. Table 87 shows that the phrase “users prefer Mubasher’s product” [مباشر- افضل] occurred in 11 documents. Therefore, they were resent to the expert who labelled the documents in the first stage. Moreover, it can be seen from their structure that the above documents have meanings of users’ judgment and comparison among three products, which makes the labelling clearer than any statements in the targeted corpus. Moreover, the same process was carried out for random features (positive or negative), namely, “works” (شغال), “subscription” (اشتراك), “support” (الدعم), and “disrupt” (عطلان). The results show that, after completing the process, 19 documents were sent to experts to check the labelling (see Appendix C). In total, 42 tweets were examined, and 18 tweets were relabelled. In last stage, the original data were updated according to the new labelling. Then, the updated data were loaded to run a new classification process. A comparison was carried out between the results with the original and the new classification.

Table 87: Phrase “problem of Mubasher” [المشكلة-مباشر] in the negative class documents

Original labelling class	Original Arabic tweets with English translation	New labelling class
Negative	مباشر برو جيد وليس الأفضل سواء في المنتج أو السعر مبروك Mubasher Pro is good and not the best either in product or price congratulations	Positive
Negative	افضل تقليل الفراغ بين مباشر برو و البرنامج المستخدم في الهواتف الذكية ، بالاصح تقريبا المزايا والخدمات المقدمة بينهم وشكرا Better reduce the gap between Mubasher Pro and the program used in smart phones, bring the advantages and services provided by them and thank you	Neutral
Negative	مباشر برو ولا بريال تكرر شارت الافضل Mubasher Pro is not the best	Negative
Negative	اتوقع ان النسخه الاخيريه فيها بعض المشاكل هذا الي لحظته بالنسبه لي برنامج سمارت شارت اتوقع افضل من مباشر I expect the latest version to have some problems to this moment for me Smart Chart I expect better than Mubasher	Negative
Negative	جاري تحميل أفضل نسخة لمباشر برو 9 نسخة 10 سيء جداً جداً Download the best version of Mubasher Pro 9 version 10 very, very, very bad	Negative
Negative	تعلم شركة اتحاد اتصالات موبيلي عن توصية مجلس الإدارة بعدم توزيع أرباح عن الربع الأول من العام المالي 2015م both and Mubasher Pro together have a slightly better Ettijah, but their breakout problem is like Pro's	Negative
Negative	الاتجاه أفضل منه وسعره أقل بكثير من مباشر The Ettijah product is better than it and its price is much lower than Mubasher	Negative
Negative	ضحكتني مباشر برو يرفع أسعاره لو يصير سعره 100 ورب الكعبة ما أشتريه برنامج غبي جدا ويقطع وتوقف انتهى وقته الإتجاه والمال أفضل منه بكثير Mubasher Pro raises its prices if it becomes a price of 100 and the Lord of the Kaaba not going to buy a program is very stupid and interrupts and stopped so many time and Ettijah much better	Negative
Negative	مباشر غثيث ويقطع ويرفع الضغط وانا أخوك المؤشر أفضل Mubasher grinds and cuts and raises pressure and the best smart chart	Negative
Negative	خلص اشتراكي واحب اغير افيدوني عن برنامج افضل من مباشر برو I would like to change the program better than Mubasher	Negative
Negative	اشترك في مباشر برو عن طريق البنك 6 شهور افضل لك انا مشترك عن طريق بنك الرياض روووعه لاتقطع ولا شحن ولا شي Subscribe to Mubasher Pro through the bank 6 months better for you I am a subscriber through Riyadh Bank Rowaoh for cutting and not shipping and nothing.	Positive

Tables 88 and 89 show the performance accuracy for the SVM with TF-IDF schema with the N-gram feature for both the original classification and the new classification, respectively. For the neutral class, the precision for the original classification was 82.76%, which dropped to 80.77% for the new classification after the relabelling process. On the other hand, the recall for the original classification for the neutral class was 24.24%, which rose to 21.76% for the new classification after the relabelling process. Table 88 shows performance accuracy with the original labelling. In the last stage, the original data were updated according to the new labelling. Then, the updated data were loaded to run a new classification process. A comparison was carried out between the original classification and the new classification.

Table 88: All class performance accuracy in original classification for SVM with TF-IDF schema without N-gram

	true neutral	true positive	true negative	class precision
pred. neutral	48	4	6	82.76%
pred. positive	129	744	154	72.44%
pred. negative	21	7	211	88.28%
class recall	24.24%	98.54%	56.87%	

Table 89: All class performance accuracy in new classification for SVM with TF-IDF schema with N-gram

	true neutral	true positive	true negative	class precision
pred. neutral	42	2	8	80.77%
pred. positive	131	749	156	72.30%
pred. negative	20	10	206	87.29%
class recall	21.76%	98.42%	55.68%	

Table 90 shows a performance accuracy comparison for SVM with the TF-IDF schema before and after the relabelling process. The obtained result shows that there is no improvement using SVM with the TF-IDF schema with the N-gram feature for the new classification

Table 90: All class performance accuracy comparison for SVM with TF-IDF schema

Classifier	Accuracy	Recall	Precision	Classification Error
SVM with the original labelling dataset	76.74	64.40	77.26	23.26 %
SVM with new labelling Dataset	75.31	58.63	81.85	24.69%

On the other hand, a better result was obtained when NB with the TF-IDF schema with the N-gram feature was used to examine the new labelling dataset. Tables 91 and 92 show the performance accuracy for NB using the TF-IDF schema with the N-gram feature for the new classification. For example, the neutral class precision for the original classification was 43.75%, which rose to 54.26% for the new classification after the relabelling process. On the other hand, the recall for the original classification for the neutral class was 45.96%, which rose to 36.27% for the new classification after the relabelling process.

Table 91: All class performance accuracy in original classification NB with TF-IDF schema with N-gram.

	true neutral	true positive	true negative	class precision
pred. neutral	91	70	47	43.75%
pred. positive	62	621	82	81.18%
pred. negative	45	64	242	68.95%
class recall	45.96%	82.25%	65.23%	

Table 92: All class performance accuracy in new classification NB with TF-IDF schema with N-gram.

	true neutral	true positive	true negative	class precision
pred. neutral	70	36	23	54.26%
pred. positive	72	657	91	80.12%
pred. negative	51	68	256	68.27%
class recall	36.27%	86.33%	69.19%	

Table 93 shows the comparison carried out between the results with the original and the new classification. The obtained results show that there is an improvement using NB with TF-IDF schema with the N-gram feature for the new classification.

Table 93: All class performance accuracy comparison for NB with BTO schema with N-gram.

Classifier	Accuracy	Recall	Precision
NB with the original labelling dataset	72.05%	64.48%	64.85%
NB with the new labelling dataset	74.25%	63.92%	67.40%

To sum up the experiment in Arabic text classification regarding Mubasher's product, the results show that NB has greater performance improvements than SVM after the relabelling

process, which can enhance classification accuracy as well as take out some noise that occurred during the original labelling.

### 6.2.5. Classification using CRM

A similar approach could also be extended to CRM analysis with structured inputs from a consumer database. Furthermore, the research proven that there is a strong link between behaviour patterns and customer attitudes. For example, form the data obtained for this experiment 205 Mubasher customers out of 472 started trading and then stopped for some reason such as session interruption, high price, and lack of customer services.

#### 6.2.5.1. Experiment for classification using CRM

The interest of CRM investment is to build applications to automate a firm’s sales. Moreover, having consumer information in databases and applying data mining techniques can help to identify customer patterns. Thus, companies can distinguish their loyal customers and develop new marketing strategies. In this experiment, the collection process was based on customer data divided according to business points of view into different classes in CRM. For example, Mubasher’s data was labelled according to the following criteria:

The label was “neutral” for users who had never done any trading on Mubasher’s platform. For a user who started trading and then stopped for some reason, the label was “negative”. If the user continued to trade throughout Mubasher’s platforms, the label was “positive”. The label was created by seeing when each user had started to trade and comparing that with his last order. If his last order was before the last week (i.e., he had not placed a single trade in the last week of the period when we collect the data), then the customer must have stopped trading at a certain point before. Figure 57 summarises the classification process of Mubasher’s CRM data.

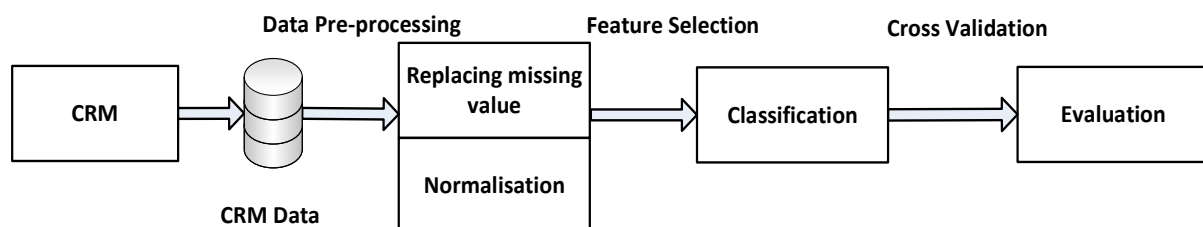


Figure 57: The process of Mubasher’s CRM classification.

Since the first task is loading the dataset, the “Retrieve” operator was used to load Mubasher’s CRM data, then the “Replace missing values” operator was used to replace the missing values with the average value of the attributes used for the classification process. After that, the

“Normalise” operator was used to pre-process and normalise the selected attributes from Mubasher’s CRM. In addition, the “Set role” operator was used to change the role of a specific field to the target variable or class for each record in the data. Finally, the X-validation operator was used to evaluate the accuracy using the SVM algorithm. The X-validation operator is a nested operator. It has two sub processes: a training process and a testing process. The training process is used for training a model. Then, the trained model is applied in the testing process. The performance of the model is also measured during the testing phase.

#### **6.2.5.2. Experiment parameter settings**

The same experiment parameter settings used in Chapter 5 for classification using CRM were utilised in this experiment. For example, the classifier settings followed the default values for most of the parameters to get results that are more accurate. Ten-fold cross-validation was used several times to find the best values of these parameters. In addition, both SVM and NB had the same operator parameter settings used.

#### **6.2.5.3. Data description**

Mubasher’s CRM data contain the most important fields that describe the customers’ patterns or behaviour. For example,

- Customer No: A unique identifier for each user.
- Status: The status of the user’s account. It can either be “Active”, “Locked”, or “Deleted”. The users who sign up start with the status “Active”. “Locked” accounts are those that were stopped by Mubasher staff, because their accounts were either changed to a different one or eliminated. “Deleted” is self-explanatory.
- Trading: “Yes” if the customer had done at least one trade in the past three months, otherwise “No”.
- Relationship Manager: This field was used to track the different signup channels for the Mubasher platform. The channels include Facebook, Twitter, and our new hub, Mubasher.info.
- Signup Date: The signup date of the customer.
- Preferred Platform: The platform in which the user executed most of his orders.

The RapidMiner wizard was used to import the dataset; RapidMiner studio uses strongly typed attributes. The numerical data type remains the same, and the string data are changed to text. On the other hand, the CRM-Label data type was changed to nominal, since nominal scales are used for labelling variables without any quantitative value [190]. After that, the data were

loaded, and the set role operator was used to change the role of one or more attributes to specify the label columns for Rapidminer.

#### 6.2.5.4. Experimental results and evaluation

The proposed model for classifying Mubasher’s CRM data needed to understand the data and then to classify them in order to learn the importance of the data in the targeted CRM. It was decided to label each customer record depending on the company criteria in different classes (positive, negative, and neutral) to prove whether the results of CRM classification are consistent from the previous result in the first experiment related to opinion mining. Mubasher’s CRM classification model collects structured CRM data using its database’s API. It performs normalisation techniques based on the existing CRM data to rescale the selected attribute values to fit in a specific range. Finally, it extracts key features with supervised learning, which is again like the opinion mining model. The number of records collected from Mubasher’s CRM was 488; each record was labelled based on Mubasher’s criteria, and then the data were stored. Rapidminer was used to load the dataset in normalised form, as mentioned in the framework, with no duplicate records and no null attribute values. Table 94 shows the number of raw data involved in the classification experiment.

Table 94: Number of records

Total	Positive	Negative	Neutral
488	156	119	213

Table 95 shows the classification accuracy, precision, and recall for the NB and SVM classifiers: Crosse-validation=10, sampling type=stratified sampling. To sum up Mubasher’s CRM experiments, the best accuracy was achieved by NB due to the advantages of NB stated in Chapter 2.

Table 95: Accuracy, precision, and recall with all classes for SVM and NB classifiers

Classes	Classifier name	Accuracy	Class recall	Class precision
Positive, Negative and Neutral	NB	86.89%	84.99%	86.14%
	SVM	73.37%	66.19%	74.92%

Figure 58: Comparing CEM labelling with CRM shows a comparison between CRM labelling and CEM labelling. One hundred thirty-four tweets were labelled similarly between CRM and CEM. On the other hand, 354 tweets were labelled dissimilarly between CEM and CRM. On other words, 72.5% of collected tweets labelled differently by CRM and CEM, and only 27.5% had similar labelling between CEM and CRM.

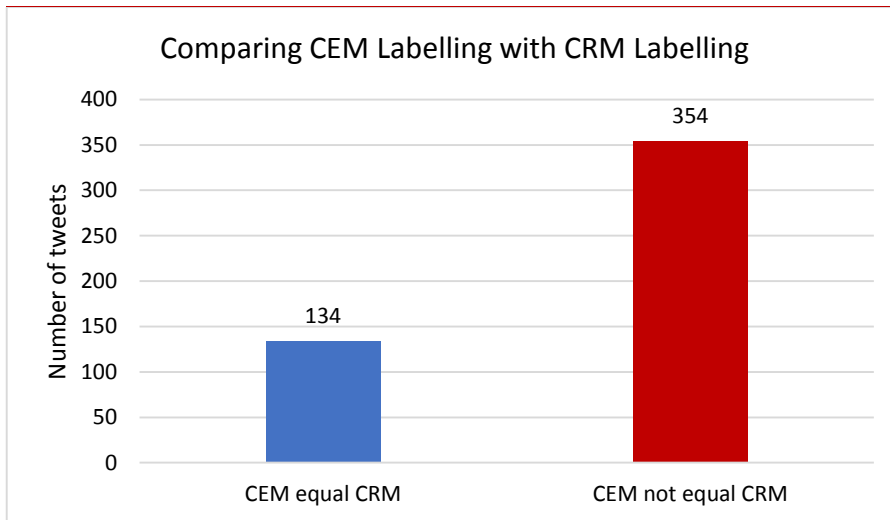


Figure 58: Comparing CEM labelling with CRM labelling

Figure 59 shows the number of tweets with similar labelling between CRM's positive labelling and CEM's (positive, negative, and neutral) labelling. The total number of tweets labelled positive in CRM is 156. Only 25 tweets had the same labels between CRM and CEM, whereas CEM labelled 20 of the tweets as neutral and 111 as negative. Overall, 84% of the labelling in CRM's positive class was different from CEM.

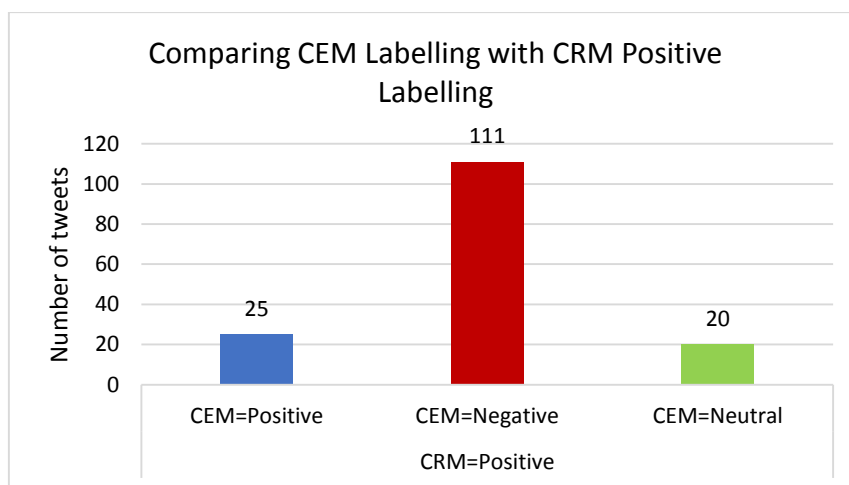


Figure 59: Comparing CEM labelling with CRM positive labelling

Figure 60 shows the number of tweets with similar labelling between CRM's Negative Labelling and CEM's (positive, negative, and neutral) labelling. The total number of tweets with labelled negative in CRM is 119 tweets. Eighty-six tweets have the same labelled between CRM and CEM, and 12 tweets labelled as neutral. On the other hand, 21 tweets labelled positive in CEM, while it is negative in CRM. Overall, 72% of the labelling in CRM's negative class labelled similar with CEM.



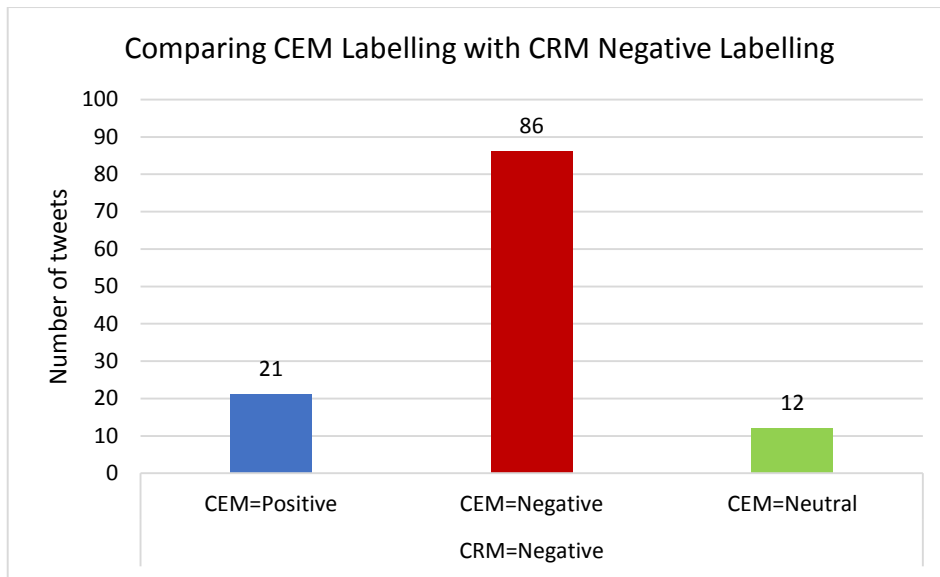


Figure 60: Comparing CEM labelling with CRM negative labelling

Figure 61 shows the number of tweets with similar labelling between CRM's neutral labelling and CEM's (positive, negative, and neutral) labelling. The total number of tweets labelled neutral in CRM is 23. On the other hand, CEM labelled 49 of the tweets as positive and 141 as negative. Overall, 90% of the labelling in CRM's neutral class was labelled differently than in CEM.

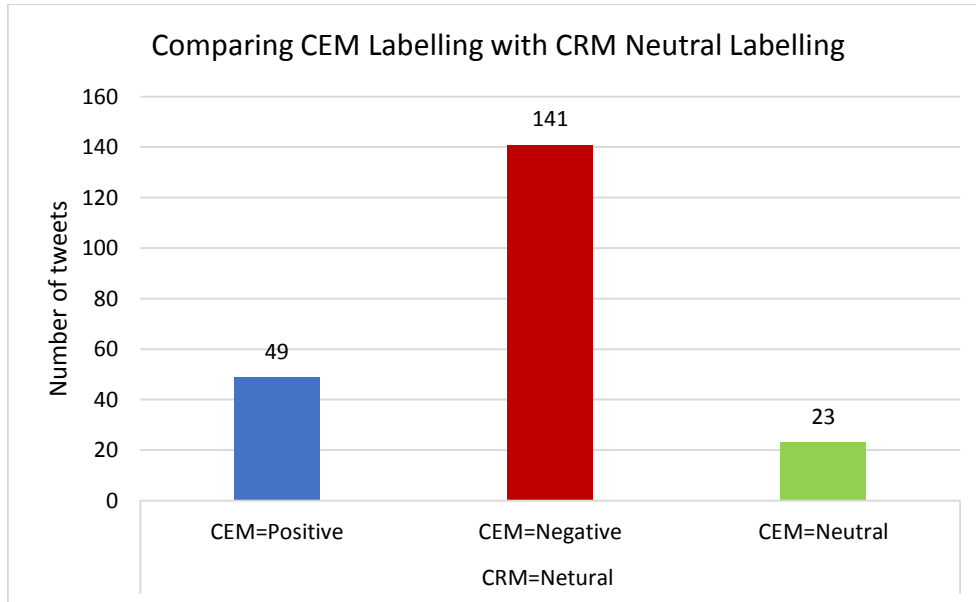


Figure 61: Comparing CEM labelling with CRM neutral labelling

### 6.2.6. Tuning CRM via CEM

Generally, schema integration is the process by which structures from different databases are conceptually integrated into a single consistent schema [191]. Integration is performed when multiple sources of data must be combined to retrieve all data that do not have the same

schemas. For example, in this research, first information about the tweet (text document in which the user can express his or her opinion) is obtained from the Twitter platform (CEM Data). In addition, a second database stores information about the customer, which can be obtained from Mubasher's CRM.

The aim of this study is to find a way to support CRM. For instance, what exactly customers want or think about a particular product also proves that the value of social media information can bring better business decisions. Therefore, similar tweets were grouped such that tweets within the same group bore similarity to each other, while tweets in different groups were dissimilar from each other. This helps to understand customers' behaviours, find out the most common problems, and contact customers within the same group and answer their inquiries faster. Moreover, this will give companies the ability to learn and to make strategic plans about what to sell, how and where to advertise, and what price to set.

After the data collection stage, the data model layout was created. This included a tweet description table (Tweet ID, Tweet Original, Tweet Filtered, Tweet Time, Tweet User, Tweet Label) and selected attributes from Mubasher's CRM databases (Customer ID, Comments). Please note that all information above was anonymised by the data provider. Then, the integration between CRM and CEM was identified by customer number and Twitter user for each customer record with the individual customer's positive, negative, and neutral opinions. Moreover, the same record also has the original tweet and the tweet label as an opinion from the CEM data for each individual customer. Each customer record now has two labels: a CRM label dependent on CRM's criteria and a CEM label from the sentiment analysis model. Then, design and implementation of Arabic text clustering over social media were presented to Mubasher's customers in order to identify the absent aspects, since the sentence level is more general, and it is not based on targeted positive or negative opinions like the entity and aspect level. In addition, users can express their opinions in different topics that makes determining opinions challenging. Therefore, a clustering technique was applied to group similar and dissimilar tweets as well as tweets' opinions into groups. This will help Mubasher to understand the missing aspects in their CRM structure.

#### **6.2.6.1. Optimising the structure of CRM**

Traditional CRM contains three modules, marketing, sales, and support, which rely on the customer relationship and profiling information input by an operator [192]. The missing part for the traditional CRM is how to benefit from the inputs of social media. This case study for Mubasher's products tried to find a relationship between customers and their opinions over social media platforms. This information can be used to enrich and direct CRM to perform better in supporting business objectives. Extracting customers' opinions from platforms such

as Twitter requires Arabic sentiment analysis to be developed and then integrating those opinions with CRM data. This study provides a model of data mining that can utilise and tune traditional CRM to reflect the needs and expectations of customers on social media. This study contributes to CRM enhancement by optimising and updating the CRM structure to become integrated with customers' opinions on social media using a clustering technique to determine the main topics that need to be added in the CRM structure.

#### 6.2.6.2. Experiment in tuning of CRM via CEM

Text mining tools have been proposed in the form of a framework after the integration processes are performed through the clustering technique to group users' text details to find out those missing aspects related to each CRM customer record, which can lead to updating the CRM structure. Figure 62 outlines the grouping procedure for Arabic tweets. Integrated tweets were saved in the main dataset. After that, there were five steps for pre-processing documents using Rapidminer. The first was to replace English terms with Arabic equivalent terms; all Hamza forms such as *أ, إ, ؤ, ة* were removed because of diacritics and replaced with initial, *أ, آ, إ* and with Alif *ا*. The second was the tokenisation step, in which each tweet was divided into multiple token-based whitespace characters. The third was removing the Arabic stop words. In the fourth step, the suffixes and prefixes were removed, and words were returned to their roots by the Arabic stemmer. The fifth step was the filtering of tokens by length, which removed useless words and was set to four. Finally, the N-gram set of words within a given text was set to two. Then, clustering techniques were applied when there was no class predicted. Clustering is a type of unsupervised learning used when one has unlabelled data. The classical clustering technique is called K-means. The following steps describe how K-means works [152]: (1) The user determines in advance how many clusters are needed by the set-up K parameter; (2) cluster centres are randomly initialised using K points; (3) Clusters are assigned to their closest centre according to the Euclidean distance metric; (4) Centroids are moved by recalculating the position of the centroid of the instances in each cluster; and (5) Steps 3 and 4 are reiterated until the centroids no longer move.

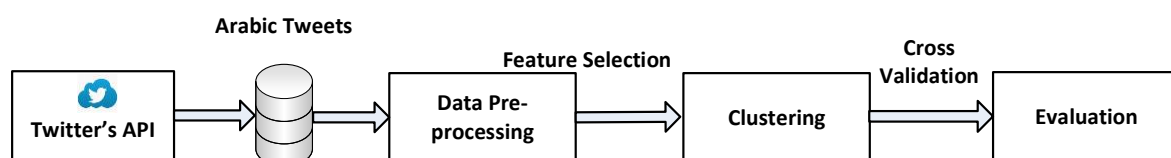


Figure 62: The process of clustering analysis in Arabic tweets.

The idea here is to find the semantic similarities between words according to the information in the collection of tweets—in other words, to calculate semantic similarity between words or concepts in a sentence. It was decided to use TF-IDF followed by obtaining the smaller distances within centroids for the targeted corpus.

### 6.2.6.3. Details of using clustering with Rapidminer

- K: The number of clusters that should be detected.
- Measure Types: Numerical measure with the default setting for Rapidminer.
- Education Distance was the smaller distance within centroids
- Max Run is the maximal number of runs of K-means with random initialisation and set to the default, which is 10.
- Max Optimisation Steps specifies the maximal number of iterations performed for one run of the K-means and is set to the default, which is 100.

### 6.2.6.4. Experimental results and evaluation

After the clustering process was implemented in four clusters as requested by Mubasher, humans interfered by looking carefully into each cluster's documents. They came up with the main topics discussed among the customers.

Table 96 shows the cluster numbers with the records and the main topic for each cluster.

Table 96: Cluster number and main topic

Cluster	Number of records	Topic
0	23	Customer service (training)
1	61	Price
2	323	Session interruption (Tec)
3	81	Comparison with other product

Figure 63 shows all the tweets grouped and organised into four clusters. The cluster distance performance task was applied to determine the average distance between the centroid clusters. The results lead us to conclude that the best clustering works were the smaller distances within centroids. It was achieved with TF-IDF using the N-gram feature. Moreover, humans interfered by looking carefully into our data cluster to make sure of the main topic into each cluster.

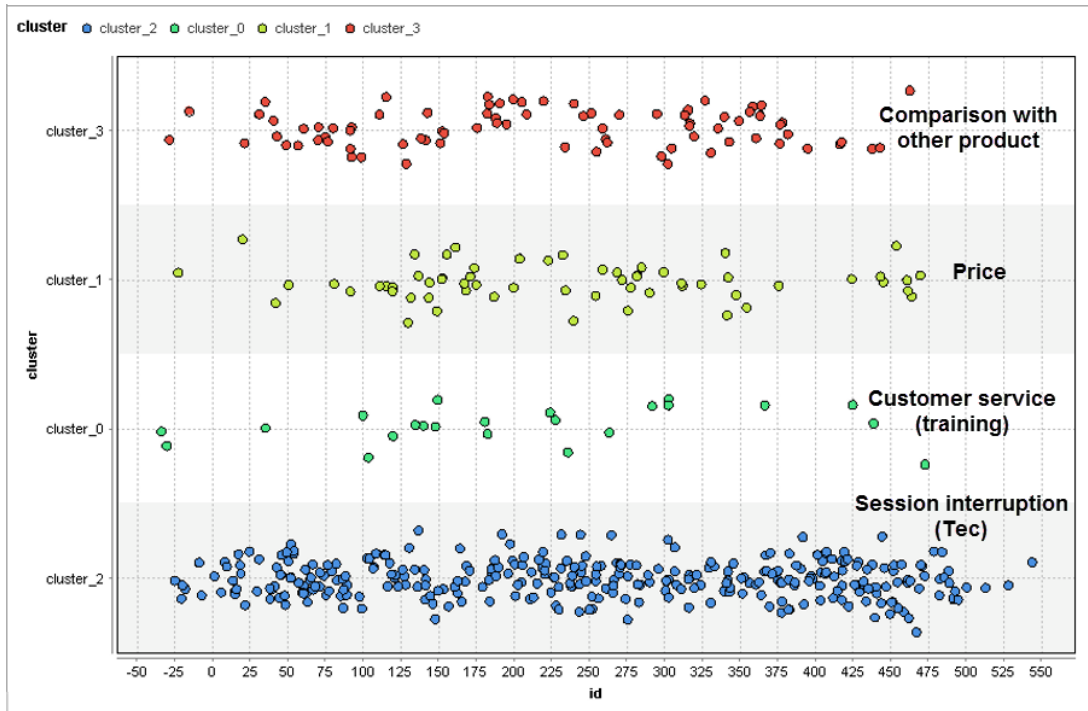


Figure 63: Topics found by clustering

Figure 64 shows a comparison between CRM labelling and CEM labelling for the four clusters. In Cluster 0, 36 tweets were labelled the same in both CRM and CEM, and 77 tweets were labelled otherwise. In Cluster 1, 16 tweets were labelled the same in CRM and CEM, and 45 were different. In Cluster 2, 59 tweets were labelled the same in CRM and CEM, and 174 were different. In Cluster 3, 17 tweets were labelled the same in CRM as in CEM, and 64 were dissimilar. Figure 65 shows the percentage of labelling similarity between CRM and CEM. Cluster 0 has the highest similarity scores with 32% (47% difference between CRM's labelling and CEM's labelling), followed by Cluster 1 (26%), Cluster 2 (25%), and Cluster 3 (21%). Overall, approximately 70% of the labelling is dissimilar between CRM and CEM.

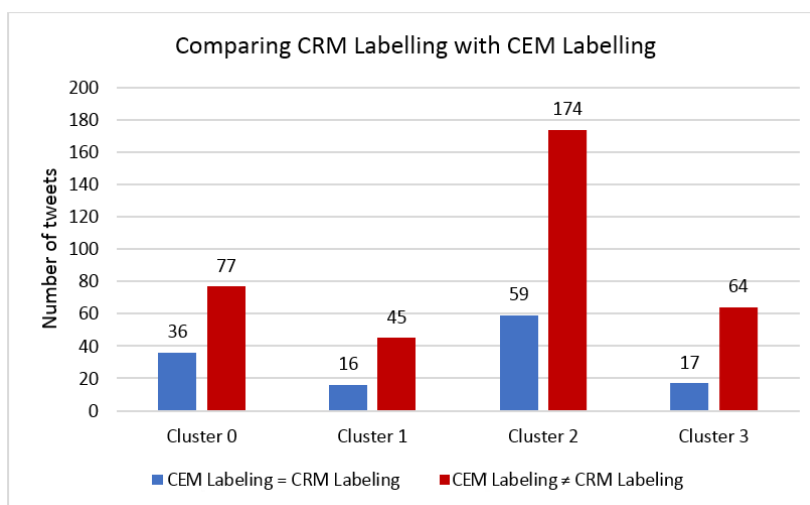


Figure 64: Comparing CEM labelling with CRM labelling

Figure 66 illustrates the number of tweets in Cluster 0 labelled the same in CRM and in CEM. The most labelling similarity between CRM and CEM is in the negative labelling (11 tweets dissimilar and 25 similar), followed by the neutral labelling. Only one same labelling occurred between CEM and CRM in the positive labelling. In addition, the most dissimilarity occurred between CRM and CEM in CRM with the neutral class (47 tweets dissimilar and 10 similar).

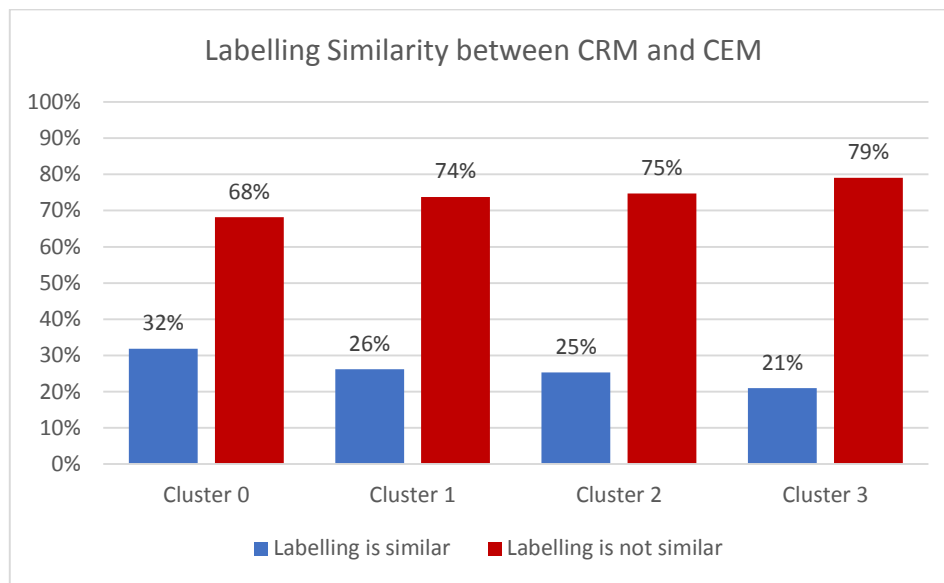


Figure 65: Labelling similarity between CRM and CEM

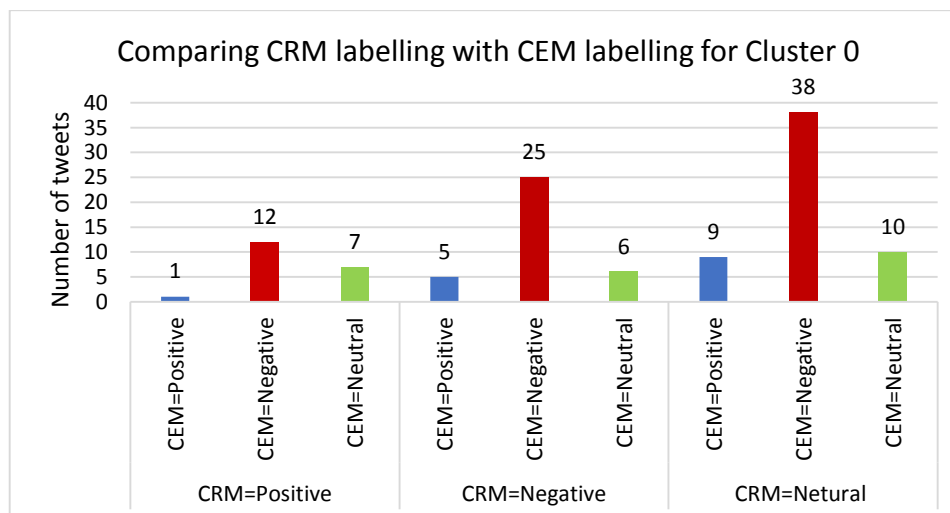


Figure 66: Comparing CRM labelling with CEM labelling for Cluster 0

Figure 67 illustrates the number of tweets in Cluster 1 whose labelling in CRM and CEM was and was not equal. The most labelling similarity between CRM and CEM is in the negative labelling (three tweets dissimilar and 13 similar), followed by the positive labelling. In cluster 1, zero neutral labelling occurred in CEM, indicating that most people had clear opinions on the topic. In addition, the most dissimilarity occurred between CRM and CEM in CRM's neutral class (23 tweets dissimilar and only three similar).

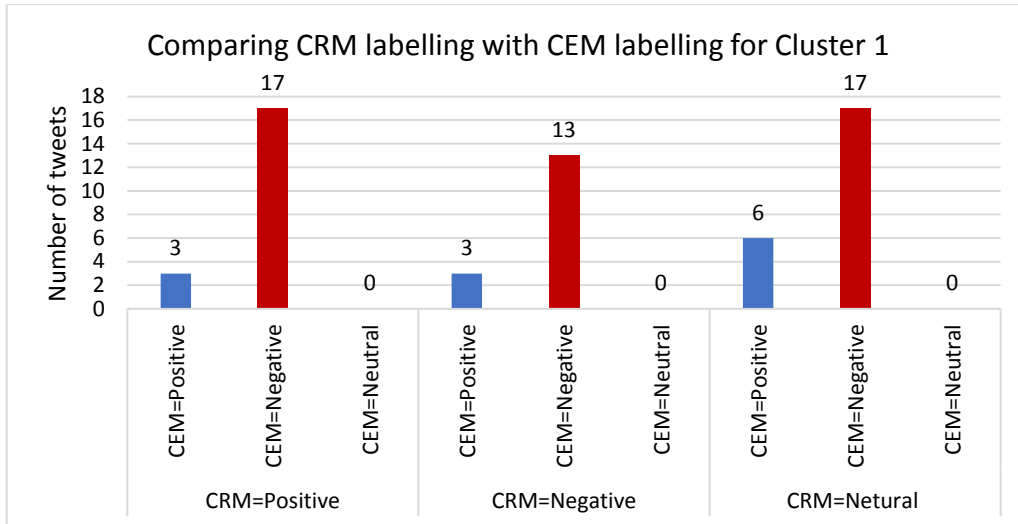


Figure 67: Comparing CRM labelling with CEM labelling for Cluster 1

Figure 68 shows a comparison between the tweet labelling in CRM and CEM for Cluster 2. The most dissimilarity occurred between CRM and CEM in CRM's neutral class (85 tweets dissimilar and only seven similar). On the other hand, the most similarity occurred in CRM's negative class (10 tweets dissimilar and 38 similar).

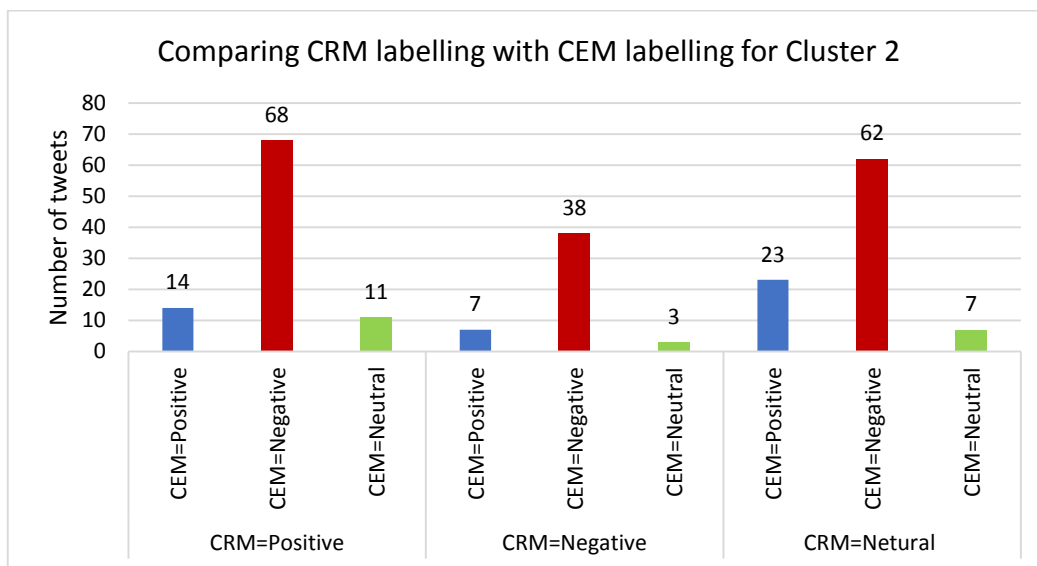


Figure 68: Comparing CRM labelling with CEM labelling for Cluster 2

Figure 69 illustrates the number of tweets labelling in CRM is equal and not equal to CEM for cluster 3. The most labelling similarity between CRM and CEM is in the negative labelling, followed by the positive labelling. Cluster 3 is similar to cluster 1, a zero-labelling occurred on CEM as neutral, that indicate that most people have a clear opinion for the topic. In addition, the most dissimilarity occurred between CRM and CEM is in CRM's neutral class (35 tweets dissimilar). On the other hand, the most similarity occurred between CRM and CEM is in CRM's negative class (six tweets dissimilar and 10 similar).

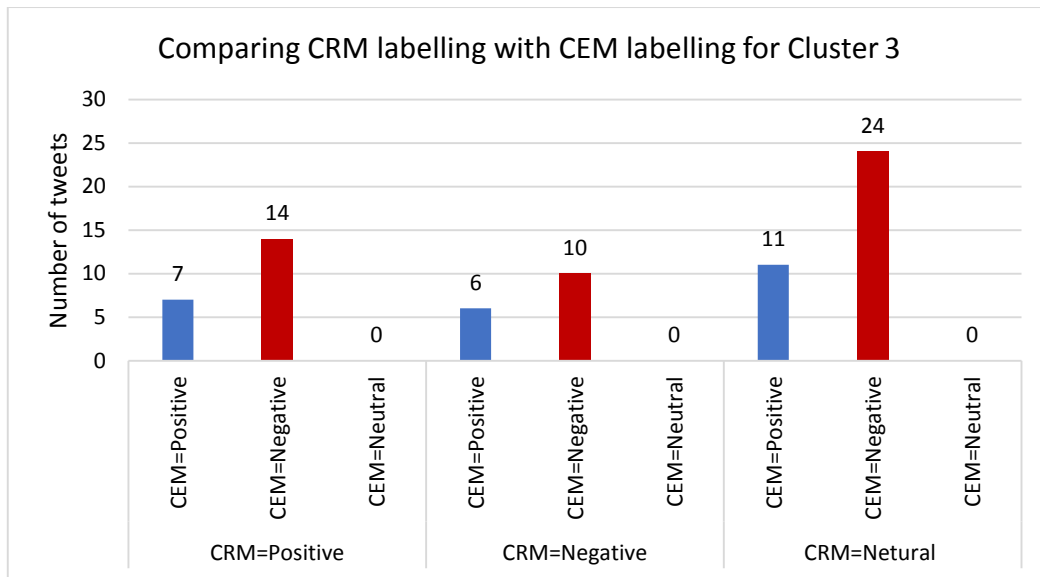


Figure 69: Comparing CRM labelling with CEM labelling for Cluster 3

In conclusion, the previous results lead us to conclude that the most label matching occurred between CRM and CEM in the negative class since negative opinions are usually clear for people. In addition, the neutral class is essential for ambiguous topics such as customer service. On the other hand, it's not important for some certain topics, such as whether Mubasher's price is high or low and comparing Mubasher with other products. Therefore, for certain topics, expert should not label sentiments as neutral.

### 6.3. Summary

This chapter has documented the second set of experiments carried out to investigate whether social media could help experts understand users' perspectives and could change experts' knowledge about their users. Therefore, a framework was proposed for integrating CRM with CEM. It can be concluded that social media can indeed change experts' knowledge and support CRM by updating its structure to cover more aspects that customers have mentioned in their social media posts. Consequently, this will make any new prediction study on customer patterns more realistic and better corresponding to another information source, such as social media platforms.



## Chapter 7: Conclusion and Future Work

### 7.1 Conclusion

This work attempted to move the state-of-the-art in Arabic sentiment analysis, labelling techniques, and CRM optimisation techniques forward. Returning to the first main questions posed at the beginning of this research, it is now possible to state that CRM could benefit from the inputs of social media and emerging data mining techniques. With the proposed framework, CRM can be tuned and revised based on feedback from social media to realise its full potential. The proposed framework involves three processes, using supervised and unsupervised methods that require less human involvement: first, Arabic sentiment analysis with the new method of relabelling to reduce the conflicts in the original labelling (the key enabling techniques have been addressed in Chapters 3 and 4); second, CRM classification techniques according to the best understanding of users' behaviour in each case study's data structure; and finally, integration of the information from the previous two processes to address users' satisfaction and understand their sentiments, likes, and dislikes. It helps to give actionable insights for the targeted domain. These two processes are addressed in Chapters 5 and 6.

Returning to question 1.1 posed in Chapter 1, it is now possible to state that the most favourable techniques for pre-processing are tokenization, stemming, stop words removal, and filtering tokens by length and N-gram feature. In addition, for the classification of Arabic sentiment analysis, the SVM classifier performs better without N-grams with both schemas TF-IDF and BTO, while with N-grams, with both schemas TF-IDF and BTO, performance drops. The NB classifier performs better with N-grams with both schemas TF-IDF and BTO, while without N-grams, with both schemas TF-IDF and BTO, performance drops. In general, the SVM classifier with both schemas TF-IDF and BTO outperforms the NB classifier. In other words, using the SVM classifier without the N-gram feature with TF-IDF schema is the first recommendation, and the NB classifier with the N-gram feature with the BTO schema is the second recommendation. To explore optimal techniques for pre-processing and classification of Arabic sentiment analysis, the process of sentiment analysis in Arabic tweets has been deployed in real-world applications from three different domains—the Saudi stock market, product reviews (Mubasher products), and education (E-learning in King Abdulaziz University)—to ensure the repeatability of the findings.

As the Arabic sentiment analysis model is a supervised learning technique requiring annotated training data, it is now possible to state that adding a neutral class would give humans more flexibility to express their uncertainty. In addition, the neutral class would improve the training and classification process by enabling better coverage of the vector space constructed by the training data set. In addition, a relabelling technique for Arabic tweets to identify the noise generated by the original labelling has been introduced to identify the potential contradictions via the noise visualised during the classification. Altogether, they help us to tackle the ambiguous information in the feedback collected from the CEM.

Returning to the second main question posed at the beginning of this study, it is now possible to state that CRM can be fine-tuned based on the differences between the opinions predicted by CRM and the actual opinions collected via social media for both the structure of the CRM system (sentence level) and the inaccuracy of the criteria or weights employed in the CRM system (aspect level). To meet users' expectations and understand their opinions through social media platforms, CRM can now be adjusted based on feedback from social media platforms. Two case studies have been carried out, the first one to integrate CRM with CEM on the aspect level in the educational domain. The results show that changing some aspects of the criteria required by students in their texts over social media can help enhance the overall classification accuracy in the LMS data and improve the understanding of students' studying status. The second experiment carried out integrated CEM into CRM on the sentence level. This experiment was carried out in the service domain. The results show that social media can change expert knowledge and support CRM by updating its structure, which can cover all the aspects that customers have mentioned in their posts on social media.

## **7.2 Research Contributions**

This section illustrates the contributions for each research question introduced in Chapter 1.

### **• *Research contributions regarding the first research question***

"First, how can hidden information (opinions) significant to CEM be extracted from Arabic social media data? This question can be divided into two sub-questions."

The first sub question is "*What are the optimal techniques for pre-processing and classification of Arabic sentiment analysis?*"

- The innovation of this research question implemented through new knowledge on identifying optimal techniques for pre-processing and classification of Arabic sentiment analysis. The novelty of this question provide is using the SVM classifier without the N-

gram feature with TF-IDF schema is the first recommendation, and the NB classifier with the N-gram feature with the BTO schema is the second recommendation. The limitation of this contribution was the labelling conducted manually and NLP is not applicable to Arabic because of its morphological richness.

The second sub question is “*How should ambiguity be tackled during the labelling process in sentiment analysis?*”

- The thesis contributes to literature by introducing neutral class to improve the training and classification process by enabling better coverage of the vector space constructed by the training data set. In addition, a relabelling technique for Arabic tweets introduced to identify the potential conflict via the noise visualised during the classification. Altogether, they help us to tackle the ambiguous information in the feedback collected from the CEM. The limitation was in the relabelling process that valid only for two nodes to determine the questionable class and check even the small rules that occurred in the text.

#### • **Research contributions regarding the second research question**

The second main question is “*How should CRM be adjusted via CEM to reduce the differences between the opinions predicted by CRM and the actual opinions collected in social media?*”.

- The thesis contributes to literature by fine-tuned CRM based on the differences between the opinions predicted by CRM and the actual opinions collected via social media for both the structure of the CRM system and the criteria or weights employed in the CRM system. That carried out by using the sentence level of the sentiment analysis to update the structure of the CRM system, and using the aspect level of the sentiment analysis to adjust the criteria or weights employed in the CRM system. Setting new benchmark for this contribution was difficult because the difficulties to obtain real data for both CRM and CEM.

### **7.3 Future Work**

The scope of future work is wide. There is plenty of room for more research in different domains using these research approaches, for example, in health services and the communications business.

- Building a dictionary application for Arabic sentiment analysis will solve the manual input and human interference problems, since the main issue with sentiment analysis is the domain adaptation problem.

- The relabelling system could be enhanced by developing applications to do the same job carried out in Chapter 4 (relabelling process) automatically. This software would need to use more nodes to determine the questionable class and check even the small rules that occurred in the text. This may help to build a robots dictionary in the future.
- More investigation is required into the influence of posts or tweets about products or services by users with more followers.
- The current study should be repeated using sentence-level and aspect-level sentiment analysis for more companies or organizations regardless of domain in order to cover both the structure of the CRM system and the inaccuracy of the criteria or weights employed in the CRM system at the same time.

## References

1. Chen, I.J. and K. Popovich, *Understanding customer relationship management (CRM): People, process and technology*. Business process management journal, 2003. 9(5): p. 672-688.
2. Ngai, E.W.T., L. Xiu, and D.C.K. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification*. Expert Systems with Applications, 2009. 36(2): p. 2592-2602.
3. Saif, H., Y. He, and H. Alani, *Semantic sentiment analysis of twitter*. The Semantic Web—ISWC 2012, 2012: p. 508-524.
4. Winer, R.S., *Customer relationship management: a framework, research directions, and the future*. Haas School of Business, 2001.
5. Meyer, C. and A. Schwager, *Customer experience*. Harvard business review, 2007. 85(2): p. 116-126.
6. Mostafa, M.M., *More than words: Social networks' text mining for consumer brand sentiments*. Expert Systems with Applications, 2013. 40(10): p. 4241-4251.
7. Cambria, E., et al., *New avenues in opinion mining and sentiment analysis*. IEEE Intelligent Systems, 2013(2): p. 15-21.
8. Sam, K.M. and C.R. Chatwin. *Ontology-based text-mining model for social network analysis*. in *Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on*.
9. Al-Rubaiee, H., R. Qiu, and D. Li. *Identifying Mubasher software products through sentiment analysis of Arabic tweets*. in *Industrial Informatics and Computer Systems (CIICS), 2016 International Conference on*. 2016. IEEE.
10. Hamed, A.-R., R. Qiu, and D. Li. *Analysis of the relationship between Saudi twitter posts and the Saudi stock market*. in *Intelligent Computing and Information Systems (ICICIS), 2015 IEEE Seventh International Conference on*. 2015. IEEE.
11. Ahmad, K. and Y. Almas. *Visualising sentiments in financial texts?* in *Information Visualisation, 2005. Proceedings. Ninth International Conference on*. 2005. IEEE.
12. Boudad, N., et al., *Sentiment analysis in Arabic: A review of the literature*. Ain Shams Engineering Journal, 2017: p. 1-12.
13. Rygielski, C., J.-C. Wang, and D.C. Yen, *Data mining techniques for customer relationship management*. Technology in society, 2002. 24(4): p. 483-502.
14. Bodine, K., R. Rogowski, and A. Stone, *Top 10 Ways To Improve Digital Experiences*. 2012, Forrester: Cambridge, MA.
15. Trainor, K.J., *Relating social media technologies to performance: A capabilities-based perspective*. Journal of Personal Selling and Sales Management, 2012(3): p. 317-331.
16. Tripathi, G. and S. Naganna, *Feature selection and classification approach for sentiment analysis*. Machine Learning and Applications: An International Journal, 2015. 2(2): p. 1-16.
17. Ordenes, F.V., et al., *Analyzing customer experience feedback using text mining: A linguistics-based approach*. Journal of Service Research, 2014. 17(3): p. 278-295.
18. Yu, C.H., S.A. DiGangi, and A. Jannasch-Pennell, *Using Text Mining for Improving Student Experience Management in Higher Education*, in *Cases on Innovations in Educational Marketing: Transnational and Technological Strategies*. 2011, IGI Global. p. 196-213.
19. Pak, A. and P. Paroubek. *Twitter as a corpus for sentiment analysis and opinion mining*. in *LREc*. 2010.
20. Agarwal, A., et al. *Sentiment analysis of twitter data*. in *Proceedings of the Workshop on Languages in Social Media*. 2011. Association for Computational Linguistics.
21. Deng, Z.-H., K.-H. Luo, and H.-L. Yu, *A study of supervised term weighting scheme for sentiment analysis*. Expert Systems with Applications, 2014. 41(7): p. 3506-3513.
22. Bollen, J., H. Mao, and X. Zeng, *Twitter mood predicts the stock market*. Journal of Computational Science, 2011. 2(1): p. 1-8.

23. Martin, V., E. Bruno, and E. Murisasco, *Predicting the French Stock Market Using Social Media Analysis*. International Journal of Virtual Communities and Social Networking (IJVCSN), 2015. 7(2): p. 70-84.
24. Zhang, W. and S. Skiena. *Trading Strategies to Exploit Blog and News Sentiment*. in *Icwsn*. 2010.
25. Duncan, B. and Y. Zhang. *Neural networks for sentiment analysis on twitter*. in *Cognitive Informatics & Cognitive Computing (ICCI\* CC), 2015 IEEE 14th International Conference on*. 2015. IEEE.
26. Al-Kabi, M.N., et al., *Opinion mining and analysis for arabic language*. International Journal of Advanced Computer Science and Applications (IJACSA), SAI Publisher, 2014. 5(5): p. 181-195.
27. Mouthami, K., K.N. Devi, and V.M. Bhaskaran. *Sentiment analysis and classification based on textual reviews*. in *Information communication and embedded systems (ICICES), 2013 international conference on*. 2013. IEEE.
28. Liu, B., *Sentiment analysis and opinion mining*. Synthesis lectures on human language technologies, 2012. 5(1): p. 1-167.
29. Pang, B. and L. Lee, *Opinion mining and sentiment analysis*. Foundations and Trends® in Information Retrieval, 2008. 2(1–2): p. 1-135.
30. Pang, B., L. Lee, and S. Vaithyanathan. *Thumbs up?: sentiment classification using machine learning techniques*. in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. 2002. Association for Computational Linguistics.
31. Maas, A.L., et al. *Learning word vectors for sentiment analysis*. in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. 2011. Association for Computational Linguistics.
32. Liu, B., *Sentiment analysis and opinion mining: Synthesis lectures on human language technologies [M].[sl]: Morgan & Claypool Publishers, 2012: 1–167*. Google Scholar.
33. Haddi, E., *Sentiment analysis: text, pre-processing, reader views and cross domains*. 2015, Brunel University London.
34. El-Khair, I.A., *Effects of stop words elimination for Arabic information retrieval: a comparative study*. International Journal of Computing & Information Sciences, 2006. 4(3): p. 119-133.
35. Liu, B., *Sentiment Analysis and Subjectivity*. Handbook of natural language processing, 2010. 2: p. 627-666.
36. Ding, X., B. Liu, and P.S. Yu. *A holistic lexicon-based approach to opinion mining*. in *Proceedings of the 2008 international conference on web search and data mining*. 2008. ACM.
37. Solomon, R.C., *Subjectivity*. 2005.
38. Turney, P.D. *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. in *Proceedings of the 40th annual meeting on association for computational linguistics*. 2002. Association for Computational Linguistics.
39. Pang, B. and L. Lee. *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*. in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. 2004. Association for Computational Linguistics.
40. Hu, M. and B. Liu. *Mining and summarizing customer reviews*. in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004. ACM.
41. Feldman, R., *Techniques and applications for sentiment analysis*. Communications of the ACM, 2013. 56(4): p. 82-89.
42. Ganapathibhotla, M. and B. Liu. *Mining opinions in comparative sentences*. in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. 2008. Association for Computational Linguistics.

43. Thelwall, M., K. Buckley, and G. Paltoglou, *Sentiment in Twitter events*. Journal of the Association for Information Science and Technology, 2011. 62(2): p. 406-418.
44. Aue, A. and M. Gamon. *Customizing sentiment classifiers to new domains: A case study*. in *Proceedings of recent advances in natural language processing (RANLP)*. 2005.
45. Blitzer, J., M. Dredze, and F. Pereira. *Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification*. in *ACL*. 2007.
46. Hu, X., et al. *Unsupervised sentiment analysis with emotional signals*. in *Proceedings of the 22nd international conference on World Wide Web*. 2013. ACM.
47. Kaji, N. and M. Kitsuregawa. *Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents*. in *EMNLP-CoNLL*. 2007.
48. Denecke, K. *Are SentiWordNet scores suited for multi-domain sentiment classification?* in *Digital Information Management, 2009. ICDIM 2009. Fourth International Conference on*. 2009. IEEE.
49. Tan, L., et al., *Sentence-level sentiment polarity classification using a linguistic approach*. Digital Libraries: for cultural heritage, knowledge dissemination, and future creation, 2011: p. 77-87.
50. Mitkov, R., *The Oxford handbook of computational linguistics*. 2005: Oxford University Press.
51. Dasgupta, S. and V. Ng. *Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification*. in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. 2009. Association for Computational Linguistics.
52. Liu, B., *Sentiment analysis: Mining opinions, sentiments, and emotions*. 2015: Cambridge University Press.
53. Tang, H., S. Tan, and X. Cheng, *A survey on sentiment detection of reviews*. Expert Systems with Applications, 2009. 36(7): p. 10760-10773.
54. Dehkharghani, R. and C. Yilmaz. *Automatically identifying a software product's quality attributes through sentiment analysis of tweets*. in *Natural Language Analysis in Software Engineering (NaturaLiSE), 2013 1st International Workshop on*. 2013. IEEE.
55. Bross, J. and H. Ehrig. *Generating a context-aware sentiment lexicon for aspect-based product review mining*. in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. 2010. IEEE.
56. Indhuja, K. and R.P. Reghu. *Fuzzy logic based sentiment analysis of product review documents*. in *Computational Systems and Communications (ICCSC), 2014 First International Conference on*. 2014. IEEE.
57. Wang, W. *Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model*. in *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*. 2010. IEEE.
58. Zhang, Z. and B. Varadarajan. *Utility scoring of product reviews*. in *Proceedings of the 15th ACM international conference on Information and knowledge management*. 2006. ACM.
59. Şimşek, M.U. and S. Özdemir. *Analysis of the relation between Turkish twitter messages and stock market index*. in *Application of Information and Communication Technologies (AICT), 2012 6th International Conference on*. 2012. IEEE.
60. Khatri, S.K., H. Singhal, and P. Johri. *Sentiment analysis to predict Bombay stock exchange using artificial neural network*. in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on*. 2014. IEEE.
61. Gao, Y., et al. *Sentiment classification for stock news*. in *Pervasive Computing and Applications (ICPCA), 2010 5th International Conference on*. 2010. IEEE.
62. Hamed AL-Rubaiee, R.Q., Khalid Alomar and Dayou Li, *Sentiment Analysis of Arabic Tweets in e-Learning*. Journal of Computer Science. 12(11): p. 553-563.

63. Tian, F., et al. *Can e-Learner's emotion be recognized from interactive Chinese texts?* in *Computer Supported Cooperative Work in Design, 2009. CSCWD 2009. 13th International Conference on.* 2009. IEEE.
64. Wang, P., M. Zuo, and L. Diao. *Design of the monitoring and analysis system of the Internet Education Public Sentiment.* in *Future Computer and Communication (ICFCC), 2010 2nd International Conference on.* 2010. IEEE.
65. Fu, G. and X. Wang. *Chinese sentence-level sentiment classification based on fuzzy sets.* in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters.* 2010. Association for Computational Linguistics.
66. Donovan, J., C.E. Mader, and J. Shinsky, *Constructive Student Feedback: Online vs. Traditional Course Evaluations.* *Journal of Interactive Online Learning*, 2010. 9(3): p. 283-296.
67. Bandyopadhyay, S., *Emerging Applications of Natural Language Processing: Concepts and New Research: Concepts and New Research.* 2012: IGI Global.
68. Cambria, E. and B. White, *Jumping NLP curves: A review of natural language processing research.* *IEEE Computational intelligence magazine*, 2014. 9(2): p. 48-57.
69. Charniak, E., *Statistical language learning.* 1996: MIT press.
70. Hamed, A.-R., R. Qiu, and D. Li, *The importance of neutral class in sentiment analysis of Arabic tweets.* *Int. J. Comput. Sci. Inform. Technol*, 2016. 8: p. 17-31.
71. Zhang, S., C. Zhang, and Q. Yang, *Data preparation for data mining.* *Applied Artificial Intelligence*, 2003. 17(5-6): p. 375-381.
72. Pérez, J., et al., *A Data Preparation Methodology in Data Mining Applied to Mortality Population Databases.* *Journal of medical systems*, 2015. 39(11): p. 152.
73. Paltoglou, G., *Sentiment analysis in social media*, in *Online Collective Action.* 2014, Springer. p. 3-17.
74. Bharti, S., et al., *Sarcastic sentiment detection in tweets streamed in real time: A big data approach.* *Digital Communications and Networks*, 2016. 2(3): p. 108-121.
75. Maynard, D. and M.A. Greenwood. *Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis.* in *LREC.* 2014.
76. Da Silva, N.F., E.R. Hruschka, and E.R. Hruschka, *Tweet sentiment analysis with classifier ensembles.* *Decision Support Systems*, 2014. 66: p. 170-179.
77. Gokulakrishnan, B., et al. *Opinion mining and sentiment analysis on a twitter data stream.* in *Advances in ICT for emerging regions (ICTer), 2012 International Conference on.* 2012. IEEE.
78. Fayyad, U.M., et al., *Advances in knowledge discovery and data mining.* Vol. 21. 1996: AAAI press Menlo Park.
79. Frawley, W.J., G. Piatetsky-Shapiro, and C.J. Matheus, *Knowledge discovery in databases: An overview.* *AI magazine*, 1992. 13(3): p. 57.
80. Witten, I.H., et al., *Data Mining: Practical machine learning tools and techniques.* 2016: Morgan Kaufmann.
81. Zhao, Q. and S.S. Bhowmick, *Sequential pattern mining: A survey.* *ITechnical Report CAIS Nanyang Technological University Singapore*, 2003: p. 1-26.
82. Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques.* 2011: Elsevier.
83. Bharati, M. and M. Ramageri, *Data mining techniques and applications.* 2010.
84. Liao, S.-H., P.-H. Chu, and P.-Y. Hsiao, *Data mining techniques and applications—A decade review from 2000 to 2011.* *Expert systems with applications*, 2012. 39(12): p. 11303-11311.
85. Souza, J., S. Matwin, and N. Japkowicz. *Evaluating data mining models: a pattern language.* in *Proceedings of the 9th Conference on Pattern Language of Programs, Illinois, USA.* 2002.
86. Mendes, R.R., et al. *Discovering fuzzy classification rules with genetic programming and co-evolution.* in *European Conference on Principles of Data Mining and Knowledge Discovery.* 2001. Springer.



87. Huq, M.R., A. Ali, and A. Rahman, *Sentiment Analysis on Twitter Data using KNN and SVM*. IJACSA International Journal of Advanced Computer Science and Applications, 2017. 8(6): p. 19-25.
88. Geng, L. and H.J. Hamilton, *Interestingness measures for data mining: A survey*. ACM Computing Surveys (CSUR), 2006. 38(3): p. 9.
89. Rushdi-Saleh, M., et al., *OCA: Opinion corpus for Arabic*. Journal of the Association for Information Science and Technology, 2011. 62(10): p. 2045-2054.
90. Khan, F.H., S. Bashir, and U. Qamar, *TOM: Twitter opinion mining framework using hybrid classification scheme*. Decision Support Systems, 2014. 57: p. 245-257.
91. Sharma, S., K.-M. Osei-Bryson, and G.M. Kasper, *Evaluation of an integrated Knowledge Discovery and Data Mining process model*. Expert Systems with Applications, 2012. 39(13): p. 11335-11348.
92. Pyle, D., *Data preparation for data mining*. Vol. 1. 1999: morgan kaufmann.
93. Haddi, E., X. Liu, and Y. Shi, *The role of text pre-processing in sentiment analysis*. Procedia Computer Science, 2013. 17: p. 26-32.
94. Duwairi, R., M. Al-Refai, and N. Khasawneh. *Stemming versus light stemming as feature selection techniques for Arabic text categorization*. in *Innovations in Information Technology, 2007. IIT'07. 4th International Conference on*. 2007. IEEE.
95. Lodhi, H., et al., *Text classification using string kernels*. Journal of Machine Learning Research, 2002. 2(Feb): p. 419-444.
96. Jun-Shan, T., H. Wei, and Q. Yan. *Application of genetic algorithm in data mining*. in *Education Technology and Computer Science, 2009. ETCS'09. First International Workshop on*. 2009. IEEE.
97. Jindal, R. and S. Taneja, *U-STRUCT: A Framework for Conversion of Unstructured Text Documents into Structured Form*, in *Advances in Computing, Communication, and Control*. 2013, Springer. p. 59-69.
98. Salton, G., A. Wong, and C.-S. Yang, *A vector space model for automatic indexing*. Communications of the ACM, 1975. 18(11): p. 613-620.
99. Salton, G. and C. Buckley, *Term-weighting approaches in automatic text retrieval*. Information processing & management, 1988. 24(5): p. 513-523.
100. Cambria, E., et al., *New avenues in opinion mining and sentiment analysis*. IEEE Intelligent Systems, 2013. 28(2): p. 15-21.
101. Rish, I. *An empirical study of the naive Bayes classifier*. in *IJCAI 2001 workshop on empirical methods in artificial intelligence*. 2001. IBM.
102. Lewis, D.D. *Naive (Bayes) at forty: The independence assumption in information retrieval*. in *European conference on machine learning*. 1998. Springer.
103. Alsaedi, N., P. Burnap, and O.F. Rana, *A combined classification-clustering framework for identifying disruptive events*. 2014.
104. Suykens, J.A. and J. Vandewalle, *Least squares support vector machine classifiers*. Neural processing letters, 1999. 9(3): p. 293-300.
105. Joachims, T., *Text categorization with support vector machines: Learning with many relevant features*. Machine learning: ECML-98, 1998: p. 137-142.
106. Jakkula, V., *Tutorial on support vector machine (svm)*. School of EECS, Washington State University, 2006. 37.
107. Khreisat, L., *Arabic Text Classification Using N-Gram Frequency Statistics A Comparative Study*. DMIN, 2006. 2006: p. 78-82.
108. Bolton, R.N. and C.O. Tarasi, *Managing customer relationships*, in *Review of marketing research*. 2007, Emerald Group Publishing Limited. p. 3-38.
109. Knox, S., et al., *Customer relationship management*. 2007: Routledge.
110. Dyche, J., *The CRM handbook: A business guide to customer relationship management*. 2002: Addison-Wesley Professional.
111. Chen, I.J. and K. Popovich, *Understanding customer relationship management (CRM) People, process and technology*. Business process management journal, 2003. 9(5): p. 672-688.

112. Galitsky, B. and J.L. de la Rosa, *Concept-based learning of human behavior for customer relationship management*. Information Sciences, 2011. 181(10): p. 2016-2035.
113. Zirtiloğlu, H. and P. Yolum. *Ranking semantic information for e-government: complaints management*. in *Proceedings of the first international workshop on Ontology-supported business intelligence*. 2008. ACM.
114. Yuksel, A., U. Kilinc, and F. Yuksel, *Cross-national analysis of hotel customers' attitudes toward complaining and their complaining behaviours*. Tourism management, 2006. 27(1): p. 11-24.
115. Lee, C.-H., Y.-H. Kim, and P.-K. Rhee, *Web personalization expert with combining collaborative filtering and association rule mining technique*. Expert Systems with Applications, 2001. 21(3): p. 131-137.
116. Ngai, E.W., *Customer relationship management research (1992-2002) An academic literature review and classification*. Marketing intelligence & planning, 2005. 23(6): p. 582-605.
117. Gupta, G. and H. Aggarwal, *Improving customer relationship management using data mining*. International Journal of Machine Learning and Computing, 2012. 2(6): p. 874.
118. Achabal, D.D., et al., *A decision support system for vendor managed inventory*. Journal of retailing, 2000. 76(4): p. 430-454.
119. Choudhury, M.M. and P. Harrigan, *CRM to social CRM: the integration of new technologies into customer relationship management*. Journal of Strategic Marketing, 2014. 22(2): p. 149-176.
120. Budiardjo, E.K., *Building Social CRM Framework on Enterprise Architecture Framework Using Value Chain Process Approach*. 2014.
121. Woodcock, N., A. Green, and M. Starkey, *Social CRM as a business strategy*. Journal of Database Marketing & Customer Strategy Management, 2011. 18(1): p. 50-64.
122. Mohan, S., E. Choi, and D. Min. *Conceptual modeling of enterprise application system using social networking and Web 2.0 "social CRM system"*. in *Convergence and Hybrid Information Technology, 2008. ICHIT'08. International Conference on*. 2008. IEEE.
123. Biere, M., *Business intelligence for the enterprise*. 2003: Prentice Hall Professional.
124. Spangler, S., et al., *COBRA—Mining web for corporate brand and reputation analysis*. Web Intelligence and Agent Systems: An International Journal, 2009. 7(3): p. 243-254.
125. Sukumaran, S. and A. Sureka, *Integrating structured and unstructured data using text tagging and annotation*. Business Intelligence Journal, 2006. 11(2): p. 8.
126. Ajmera, J., et al. *A CRM system for social media: challenges and experiences*. in *Proceedings of the 22nd international conference on World Wide Web*. 2013. ACM.
127. Yaakub, M.R., Y. Li, and J. Zhang, *Integration of sentiment analysis into customer relational model: the importance of feature ontology and synonym*. Procedia Technology, 2013. 11: p. 495-501.
128. Hu, M. and B. Liu, *Mining and summarizing customer reviews*, in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004, ACM: Seattle, WA, USA. p. 168-177.
129. Gharehchopogh, F.S. and Z.A. Khalifelu. *Analysis and evaluation of unstructured data: text mining versus natural language processing*. in *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*.
130. Duwairi, R., et al. *Sentiment analysis in arabic tweets*. in *Information and communication systems (icics), 2014 5th international conference on*. 2014. IEEE.
131. El-Halees, A.M., *Arabic text classification using maximum entropy*. IUG Journal of Natural Studies, 2015. 15(1).
132. Ibrahim, H.S., S.M. Abdou, and M. Gheith. *MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis*. in *Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd International Conference on*. 2015. IEEE.
133. Gharehchopogh, F.S. and Z.A. Khalifelu. *Analysis and evaluation of unstructured data: text mining versus natural language processing*. in *Application of Information and*

- Communication Technologies (AICT), 2011 5th International Conference on.* 2011. IEEE.
134. Kroeger, P.R., *Analyzing grammar: An introduction.* 2005: Cambridge University Press.
  135. Habash, N.Y., *Introduction to Arabic natural language processing.* Synthesis Lectures on Human Language Technologies, 2010. 3(1): p. 1-187.
  136. Althobaiti, M., *Minimally-supervised Methods for Arabic Named Entity Recognition.* 2016, University of Essex.
  137. Elkateb, S., et al. *Arabic WordNet and the challenges of Arabic.* in *Proceedings of Arabic NLP/MT Conference, London, UK.* 2006.
  138. Larkey, L.S., L. Ballesteros, and M.E. Connell. *Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis.* in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval.* 2002. ACM.
  139. Payne, T., *Exploring language structure: a student's guide.* 2006: Cambridge University Press.
  140. Farber, B., et al. *Improving NER in Arabic Using a Morphological Tagger.* in *LREC.* 2008.
  141. Farghaly, A. and K. Shaalan, *Arabic natural language processing: Challenges and solutions.* ACM Transactions on Asian Language Information Processing (TALIP), 2009. 8(4): p. 14.
  142. Alotaiby, F., S. Foda, and I. Alkharashi. *Clitics in Arabic language: a statistical study.* in *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation.* 2010.
  143. Boudad, N., et al., *Sentiment analysis in Arabic: A review of the literature.* Ain Shams Engineering Journal, 2017.
  144. El Hadj, Y., I. Al-Sughayeir, and A. Al-Ansari. *Arabic part-of-speech tagging using the sentence structure.* in *Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt.* 2009.
  145. Albirini, A., *Modern Arabic Sociolinguistics: Diglossia, variation, codeswitching, attitudes and identity.* 2016: Routledge.
  146. Abuata, B. and A. Al-Omari, *A rule-based stemmer for Arabic Gulf dialect.* Journal of King Saud University-Computer and Information Sciences, 2015. 27(2): p. 104-112.
  147. Ferguson, C.A., *Diglossia.* word, 1959. 15(2): p. 325-340.
  148. McCrea, B., *Languages of the Night: Minor Languages and the Literary Imagination in Twentieth-Century Ireland and Europe.* 2015: Yale University Press.
  149. Bruweleit, S., *Aspect, Tense and Action in the Arabic Dialect of Beirut.* 2015: Brill.
  150. Hann, G., K. Dabrowska, and T.T. Greaves, *Iraq: The Ancient Sites and Iraqi Kurdistan.* 2015: Bradt Travel Guides.
  151. Khasawneh, R.T., et al. *Sentiment analysis of arabic social media content: a comparative study.* in *Internet Technology and Secured Transactions (ICITST), 2013 8th International Conference for.* 2013. IEEE.
  152. Abdulla, N.A., et al. *Arabic sentiment analysis: Lexicon-based and corpus-based.* in *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on.* 2013. IEEE.
  153. Kaseb, G.S. and M.F. Ahmed, *Arabic Sentiment Analysis approaches: An analytical survey.*
  154. Al-Harbi, S., et al., *Automatic Arabic text classification.* 2008.
  155. El Kourdi, M., A. Bensaid, and T.-e. Rachidi. *Automatic Arabic document categorization based on the Naïve Bayes algorithm.* in *Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages.* 2004. Association for Computational Linguistics.
  156. Duwairi, R., M.N. Al-Refai, and N. Khasawneh, *Feature reduction techniques for Arabic text categorization.* Journal of the Association for Information Science and Technology, 2009. 60(11): p. 2347-2352.

157. Thabtah, F., *VSMs with K-Nearest Neighbour to categorise Arabic text data*. 2008.
158. Said, D., et al. *A study of text preprocessing tools for arabic text categorization*. in *The second international conference on Arabic language*. 2009.
159. Thabet, N. *Stemming the Qur'an*. in *Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages*. 2004. Association for Computational Linguistics.
160. Taghva, K., R. Elkhoury, and J. Coombs. *Arabic stemming without a root dictionary*. in *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*. 2005. IEEE.
161. Xu, J., A. Fraser, and R. Weischedel. *Empirical studies in strategies for Arabic retrieval*. in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. 2002. ACM.
162. Attia, M.A. *Arabic tokenization system*. in *Proceedings of the 2007 workshop on computational approaches to semitic languages: Common issues and resources*. 2007. Association for Computational Linguistics.
163. Al-Kabi, M., E. Al-Shawakfa, and I. Alsmadi, *The Effect of Stemming on Arabic Text Classification: An Empirical Study*. *Information Retrieval Methods for Multidisciplinary Applications*, 2013: p. 207.
164. Al-Shalabi, R., et al. *Stop-word removal algorithm for Arabic language*. in *Proceedings of 1st International Conference on Information and Communication Technologies: From Theory to Applications, CTTA*. 2004.
165. McNamee, P., C.D. Piatko, and J. Mayfield. *JHU/APL at TREC 2002: Experiments in Filtering and Arabic Retrieval*. in *TREC*. 2002.
166. Lewicki, P. and T. Hill, *Statistics: methods and applications*. Tulsa, OK. Statsoft, 2006.
167. Kešelj, V., et al. *N-gram-based author profiles for authorship attribution*. in *Proceedings of the conference pacific association for computational linguistics, PACLING*. 2003.
168. Ramos, J. *Using tf-idf to determine word relevance in document queries*. in *Proceedings of the first instructional conference on machine learning*. 2003.
169. Lan, M., et al., *Supervised and traditional term weighting methods for automatic text categorization*. *IEEE transactions on pattern analysis and machine intelligence*, 2009. 31(4): p. 721-735.
170. Rapidminer. *Rapidminer*. 2016 [cited 2017 1 December]; Available from: <https://rapidminer.com>.
171. Shao, J., *Linear model selection by cross-validation*. *Journal of the American statistical Association*, 1993. 88(422): p. 486-494.
172. Duwairi, R.M. and I. Qarqaz. *Arabic sentiment analysis using supervised classification*. in *Future Internet of Things and Cloud (FiCloud), 2014 International Conference on*. 2014. IEEE.
173. Fawcett, T., *ROC graphs: Notes and practical considerations for researchers*. *Machine learning*, 2004. 31(1): p. 1-38.
174. Yoshida, S., et al. *Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection*. in *Computational Intelligence in Big Data (CIBD), 2014 IEEE Symposium on*. 2014. IEEE.
175. Pak, A. and P. Paroubek. *Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives*. in *Proceedings of the 5th International Workshop on Semantic Evaluation*. 2010. Association for Computational Linguistics.
176. Go, A., R. Bhayani, and L. Huang, *Twitter sentiment classification using distant supervision*. *CS224N Project Report, Stanford*, 2009. 1(2009): p. 12.
177. Das, S.R. and M.Y. Chen, *Yahoo! for Amazon: Sentiment extraction from small talk on the web*. *Management science*, 2007. 53(9): p. 1375-1388.
178. Sebastiani, F., *Machine learning in automated text categorization*. *ACM computing surveys (CSUR)*, 2002. 34(1): p. 1-47.
179. Agrawal, R., T. Imieliński, and A. Swami. *Mining association rules between sets of items in large databases*. in *Acm sigmod record*. 1993. ACM.

180. Kotsiantis, S. and D. Kanellopoulos, *Association rules mining: A recent overview*. GESTS International Transactions on Computer Science and Engineering, 2006. 32(1): p. 71-82.
181. Chen, M.-S., J. Han, and P.S. Yu, *Data mining: an overview from a database perspective*. IEEE Transactions on Knowledge and data Engineering, 1996. 8(6): p. 866-883.
182. Lopes, A., et al., *Visual text mining using association rules*. Computers & Graphics, 2007. 31(3): p. 316-326.
183. Nasukawa, T. and J. Yi. *Sentiment analysis: Capturing favorability using natural language processing*. in *Proceedings of the 2nd international conference on Knowledge capture*. 2003. ACM.
184. Wu, X., C. Zhang, and S. Zhang, *Efficient mining of both positive and negative association rules*. ACM Transactions on Information Systems (TOIS), 2004. 22(3): p. 381-405.
185. Liu, B., Y. Ma, and C.K. Wong. *Improving an association rule based classifier*. in *European Conference on Principles of Data Mining and Knowledge Discovery*. 2000. Springer.
186. Karthikeyan, T. and N. Ravikumar, *A survey on association rule mining*. International Journal of Advanced Research in Computer and Communication Engineering, 2014. 3(1): p. 2278-1021.
187. Verhein, F., *Frequent Pattern Growth (FP-Growth) Algorithm*. School of Information Studies, The University of Sydney, Australia, 2008: p. 1-16.
188. Chang, H.-C., *A new perspective on Twitter hashtag use: Diffusion of innovation theory*. Proceedings of the American Society for Information Science and Technology, 2010. 47(1): p. 1-4.
189. Sankaranarayanan, J., et al., *TwitterStand: news in tweets*, in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2009, ACM: Seattle, Washington. p. 42-51.
190. Beniwal, S. and J. Arora, *Classification and feature selection techniques in data mining*. International Journal of Engineering Research & Technology (IJERT), 2012. 1(6).
191. Gal, A., et al. *A model for schema integration in heterogeneous databases*. in *Database Engineering and Applications Symposium, 2003. Proceedings. Seventh International*. 2003. IEEE.
192. Karna, N., I. Supriana, and U. Maulidevi. *Social CRM using web mining*. in *Information Technology Systems and Innovation (ICITSI), 2014 International Conference on*. 2014. IEEE.

## Appendices

## Appendix A

Table 97: (1) Real data collected from Saudi Stock Market, and (2) Real data collected from Saudi Stock Market multiplied by constant.

( 1 )				( 2 )				
tweetTagDes c	Column Labels			Convert from current unite to ( Unite * 100 )				
Row Labels	Negative	Neutra l	positiv e	Date	TASI	Negativ e	Neutra l	Positiv e
03/17/15	2	4	5	03/17/15	9303.13	1200	1400	1500
03/18/15	13	22	29	03/18/15	9133.87	2300	3200	3900
03/19/15	7	14	30	03/19/15	9174.41	1700	2400	4000
03/22/15	4	16	23	03/22/15	9314.53	1400	2600	3300
03/23/15	20	14	17	03/23/15	9319.69	3000	2400	2700
03/24/15	11	9	7	03/24/15	9331.19	2100	1900	1700
03/25/15	14	11	6	03/25/15	8868.12	2400	2100	1600
03/26/15	13	12	8	03/26/15	8903.49	2300	2200	1800
03/29/15	2	21	13	03/29/15	9071.20	1200	3100	2300
03/30/15	7	15	23	03/30/15	8957.05	1700	2500	3300
03/31/15	11	25	22	03/31/15	8778.89	2100	3500	3200
04/01/15	13	13	19	04/01/15	8812.35	2300	2300	2900
04/02/15	7	12	17	04/02/15	8733.79	1700	2200	2700
04/05/15	13	16	18	04/05/15	8589.70	2300	2600	2800
04/06/15	6	19	18	04/06/15	8618.63	1600	2900	2800
04/07/15	10	14	20	04/07/15	8802.36	2000	2400	3000
04/08/15	6	21	17	04/08/15	8851.83	1600	3100	2700
04/09/15	2	23	14	04/09/15	8950.11	1200	3300	2400
04/12/15	5	18	13	04/12/15	8961.13	1500	2800	2300
04/13/15	8	23	21	04/13/15	9041.12	1800	3300	3100
04/14/15	13	15	8	04/14/15	8966.33	2300	2500	1800
04/15/15	5	15	24	04/15/15	9164.44	1500	2500	3400
04/16/15	13	20	34	04/16/15	9251.19	2300	3000	4400
04/19/15	11	23	30	04/19/15	9619.97	2100	3300	4000
04/20/15	27	30	41	04/20/15	9589.42	3700	4000	5100
04/21/15	14	36	27	04/21/15	9558.89	2400	4600	3700
04/22/15	5	13	13	04/22/15	9572.27	1500	2300	2300

04/23/15	13	13	14	04/23/15	9614.61	2300	2300	2400
04/26/15	2	12	19	04/26/15	9725.17	1200	2200	2900
04/27/15	7	11	11	04/27/15	9710.93	1700	2100	2100
04/28/15	4	6	16	04/28/15	9713.07	1400	1600	2600
04/29/15	3	17	16	04/29/15	9812.40	1300	2700	2600
04/30/15	2	10	15	04/30/15	9834.49	1200	2000	2500
05/03/15	5	18	18	05/03/15	9764.24	1500	2800	2800
05/04/15	2	8	20	05/04/15	9787.00	1200	1800	3000
05/05/15	7	10	8	05/05/15	9751.03	1700	2000	1800
05/06/15	7	7	13	05/06/15	9781.69	1700	1700	2300
05/07/15	5	22	18	05/07/15	9717.90	1500	3200	2800
05/10/15	8	12	16	05/10/15	9710.52	1800	2200	2600
<b>Grand Total</b>	<b>327</b>	<b>620</b>	<b>701</b>		<b>362162</b>	<b>71700</b>	<b>101000</b>	<b>109100</b>

## Relabeling for Saudi Stock Market

### ▪ Feature “growing” (زيادة)

Table 98 shows the feature “growing” (زيادة) as positive sentiment in Saudi Stock market domain. In addition, Table 98 shows the occurrence of the feature “growing” (زيادة) in positive, negative, and neutral class.

Table 98: Occurrence of the term “growing” (زيادة).

Feature	Occurrence	Neutral	Positive	Negative
زيادة	117	2	106	9

Figure 70 shows the association rules that related to the feature “growing” (زيادة) in all class with respect to the minimum support and minimum confidence threshold. The feature “growing” (زيادة) meaning some company in the Saudi stock market raise share capital by giving existing shareholders the right to subscribe to new shares for cash. In addition, Figure 70 shows the most important rules for the feature “growing” (زيادة) which is [زيادة] --> [رأس] (support: 0.042 confidence: 1). The term (share capital) [رأس] correlated with the feature “growing” [زيادة] to compose positive phrases capital increase in the sentence.





Figure 70: Visualize the association rules for the feature “growing” (زيادة) in all class.

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “growing” (زيادة). Table 99 shows that phrases capital increase [زيادة - رأس] occurred 99 times, 2 in neutral, 92 in positive class, and 5 in negative class.

Table 99: rules for the term “growing” (زيادة) in the all data.

phrase - Terms	Occurrence	Neutral	Positive	Negative
زيادة - رأس	99	2	92	5

Since, the phrase “capital increase” [زيادة - رأس] occurred in negative and neutral class, then the both classes becomes questionable classes. Therefore, the feature “growing” (زيادة) needs further investigation in order to find out the association rules in the both classes. As result, two scenarios will be followed: First scenario is extracting association rules that occurred for the feature “growing” (زيادة) in the neutral class. Association rules generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Figure 71, shows that the feature “growing” (زيادة) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [زيادة] --> [زيادة - رأس] (support: 0.003 confidence: 1) which represent the phrase “capital increase” [زيادة - رأس] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
زيادة_رأس	زيادة	0.003	1
مبالغ_بيع	زيادة	0.002	1
كسور_الاسهم	زيادة	0.002	1
كسور	زيادة	0.002	1
عائد_مبالغ	زيادة	0.002	1
عائد	زيادة	0.002	1
سلامة_العائنه	زيادة	0.002	1
حسابات_المساهمين	زيادة	0.002	1
حسابات	زيادة	0.002	1
تداول_زيادة	زيادة	0.002	1
بيع_كسور	زيادة	0.002	1
إيداع_عائد	زيادة	0.002	1
التحقيه_زيادة	زيادة	0.002	1
التحقيه	زيادة	0.002	1

Figure 71: The correlation rules of the feature “growing” (زيادة) in neutral class.

According to Figure 71 the rule [زيادة] --> [زيادة - رأس] (support: 0.003 confidence: 1) which represent the phrase “capital increase” [زيادة - رأس] that illustrate in the wordlist matrix in the neutral class. Therefore, the next step is to search for the phrase “capital increase” [زيادة - رأس] in the neutral class documents. Table 100 shows the phrase “capital increase” [زيادة - رأس] happened in two documents.

Table 100: Term “capital increase” [زيادة - رأس] in two documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	تعليّن مجموعة سامبا المالية عن إيداع عائد مبالغ بيع كسور الأسهم الناتجة عن زيادة رأسمال البنك في حسابات المساهمين	Positive
	Samba Financial Group <u>announces</u> the deposit of the proceeds of the sale of the <u>fractional shares</u> resulting from the increase of the Bank's capital in the shareholders' accounts	
Neutral	تعليّن شركة دله الصحية عن اخر التطورات بخصوص توصية الزيادة في رأس المال للاستحواذ على كامل الحصص في شركة مستشفى باقدو	Positive
	Dallah Health Company <u>announces</u> the latest developments regarding the capital <u>increase recommendation</u> for the acquisition of full stake in Bagdo Hospital Company	

Figure 72, shows that the feature “growing” (زيادة) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values in the negative class. However, according to the Second scenario, the interested rule here is [زيادة] -> [زيادة - رأس] (support: 0.011 confidence: 1) which represent the phrase “capital increase” [زيادة - رأس] that illustrate in the wordlist matrix in the negative class.

Premises	Conclusion	Support	Confidence
رأس_المال	زياده	0.006	0.750
المال	زياده	0.006	0.750
زياده_رأس	زياده	0.011	1
يطلق_القرار	زياده	0.002	1
بعادل_تليس	زياده	0.002	1
والمتضمنه_زياده	زياده	0.002	1
موباي_تخر	زياده	0.002	1
مصاري_الاستهلاكات	زياده	0.002	1
مصاري	زياده	0.002	1
مال_الشركه	زياده	0.002	1
للمصاحف_المتطوره	زياده	0.002	1
لشركه_والمتضمنه	زياده	0.002	1

Figure 72: The correlation rules of the feature “growing” (زيادة) in negative class.

According to Figure 72 the rule [زيادة] --> [زيادة - رأس] (support: 0.003 confidence: 1) that illustrate in the wordlist matrix in represent the phrase “capital increase” occurred in the negative class. Therefore, the next step is to search for the phrase “capital increase” [زيادة - رأس] in the negative class documents. Table 101 shows the phrase “capital increase” [زيادة - رأس] happened in five documents which is satisfy the rule [زيادة] --> [زيادة - رأس], therefore it has been send again to the expert who labelled the document in the first stage. Moreover, it can have been seen form the structure three documents have been relabelled again to positive class the rest of the documents which is documents the phrase “capital increase” [زيادة - رأس] has been occurred with the negation [عدم] [زيادة - رأس] --> [عدم] which is the right place for this terms in the negative class.

Table 101: Term “capital increase” [زيادة - رأس] in the neutral class documents.

original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	تعلن شركة الخزف السعودي عدم انعقاد اجتماع الجمعية العامة غير العادية للشركة والمتضمنة زيادة رأس المال	Positive
	Saudi Ceramic Company <u>announces</u> that the Extraordinary General Meeting of the Company <u>will not be held</u> , including <u>capital increase</u>	
Neutral	تعلن الشركة السعودية للصناعات المتطورة عدم انعقاد اجتماع الجمعية العامة غير العادية المتضمنة زيادة رأس مال الشركة	Positive
	Saudi Advanced Industries Company <u>announces</u> that the Extraordinary General Meeting (EGM) <u>will not be held</u>	

Neutral	ساما تطلب استيفاء شروط لزيادة رأسمال اتحاد الخليج	Negative
	SAMA <u>seeks to meet conditions to increase the capital</u> of the Gulf Union	
Neutral	زيادة رأسمال بترورابغ يعادل تأسيس شركة جديدة وخبير يعلق القرار سلبي	positive
	Petro Rabigh's <u>capital increase</u> is equivalent to establishing a new company and an expert suspends the <u>negative decision</u>	
Neutral	تهامة تؤكد عدم قدرتها على زيادة رأس مالها	Negative
	Tihama confirms its <u>inability</u> to <u>increase its capital</u>	

Five documents have been found to satisfy the rule above and only three documents have been relabelled again to positive class. Moreover, the terms [زيادة –رأس] has been occurred in the rest of the documents which is 2 documents with the negation [عدم] which is the right place for this terms in the negative class.

▪ **Feature “distribution” (توزيع)**

Table 102 shows the feature “distribution” (توزيع) as positive sentiment in Saudi Stock market domain. In addition, Table 102 shows the occurrence of the feature “distribution” (توزيع) in positive, negative, and neutral class.

Table 102: Occurrence of the term “distribution” (توزيع).

Term	Occurrence	Neutral	Positive	Negative
توزيع	167	5	152	10

Figure 73 shows the association rules that related to the feature “distribution” (توزيع) in all class with respect to the minimum support and minimum confidence threshold. The feature “distribution” (توزيع) meaning entailed sharing out the profits of some company in the Saudi stock market. In addition, Figure 73 shows the most important rules for the feature “distribution” (توزيع) which is [توزيع] --> [ارباح] (support: 0.041 confidence: 1). The term (earnings) [ارباح] correlated with the feature “distribution” [توزيع] to compose positive phrases "distribute profits" in the sentence. The term (agree) [توافق] correlated with the feature “distribution” [توزيع] to compose positive phrases "Agree to distribution" in the sentence.

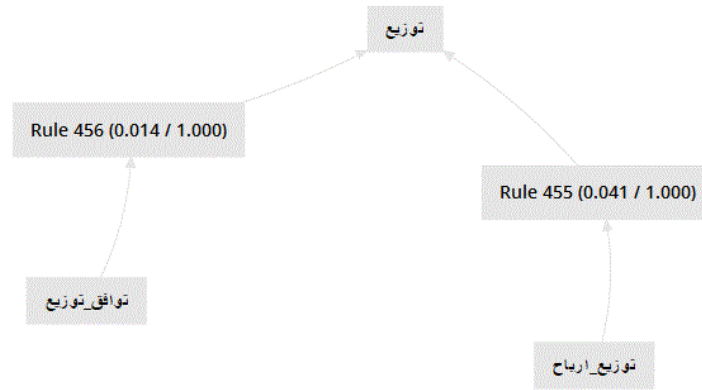


Figure 73: Visualize the association rules for the feature “Distribution” (توزيع) in all class.

Table 103 shows the phrase “Agree to distribution” phrase [توافق-ارباح] does not need further investigation since only occurred in the positive class. While, the “distribute profits” [توزيع-ارباح] needs more further investigation since its occurred 10 times in the negative class and three times in the neutral class. Since, the phrase “distribute profits” [توزيع-ارباح] occurred in negative and neutral class, then the both classes becomes questionable classes. Therefore, the feature “Distribution” (توزيع) needs further investigation in order to find out the association rules in the both classes. As result, two scenarios will be followed: First scenario is extracting association rules that occurred for the feature “Distribution” (توزيع) in the neutral class. Association rules generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Table 103: Rules for the term “Distribution” (توزيع) in the all data.

phrase - Terms	Occurrence	Neutral	Positive	Negative
توزيع-ارباح	91	3	78	10
توافق-ارباح	29	0	29	0

Figure 74, shows that the feature “Distribution” (توزيع) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [توزيع] --> [توزيع-ارباح] (support: 0.005 confidence: 1) which represent the phrase “distribute profits” [توزيع-ارباح] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
نصف_ان	توزيع	0.003	0.667
نصف	توزيع	0.003	0.667
توزيع_ارباح	توزيع	0.005	1
ارباح_مساهم	توزيع	0.003	1
بمبارسون_حظهم	توزيع	0.002	1
بمبارسون	توزيع	0.002	1
نظر_لمساهم	توزيع	0.002	1
مساهم_بمبارسون	توزيع	0.002	1
مساهم_نصف	توزيع	0.002	1
مساهم_ين	توزيع	0.002	1
مبالغ_حقوق	توزيع	0.002	1
مبلغ_ممت	توزيع	0.002	1
لمساهم_فان	توزيع	0.002	1

Figure 74: The correlations of the term “Distribution” (توزيع) in neutral class.

According to Figure 74 the rule [توزيع-ارباح] --> [توزيع] (support: 0.005 confidence: 1) that which represent the phrase “distribute profits” [توزيع-ارباح] that illustrate in the wordlist matrix in the neutral class. Therefore, the next step is to search for the phrase “distribute profits” [توزيع-ارباح] in the neutral class documents.

Table 104: Term “Exchange of profits” [صرف-أرباح] in the neutral class documents.

original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	تعلن شركة مكة للإنشاء والتعمير عن توزيع أرباح على المساهمين عن السنة المالية المنتهية في 1436/4/30 هـ Makkah Construction & Development Company <u>announces the distribution of profits</u> to shareholders for the fiscal year ended 30/4/1436 H.	positive
Neutral	تعلن شركة فواز عبدالعزيز الحكير وشركاه عن <u>توزيع أرباح</u> على المساهمين عن النصف الثاني من العام 2014 Fawaz Abdulaziz Al Hokair & Co. <u>announces a dividend distribution</u> for the second half of 2014	positive
Neutral	تعلن شركة اتحاد مصانع الأسلاك أسلاك عن تاريخ وطريقة توزيع أرباح النصف الثاني من عام 2014 م Union Wire Mills Company <u>announces the date and method of distributing dividends</u> for the second half of 2014	positive
Neutral	تعلن الشركة المتقدمة للبتروكيماويات عن طريقة <u>صرف أرباح</u> الربع الأول من العام المالي 2015م Advanced Petrochemical Company <u>announces the payment of profits</u> for the first quarter of fiscal year 2015	positive
Neutral	تعلن الشركة السعودية للنقل الجماعي سابكو عن تاريخ وطريقة <u>صرف أرباح</u> العام المالي 2014م والأرباح الاستثنائية Saudi Mass Transit Company (SAPTCO) <u>announces the date and method of paying the profits</u> for the fiscal year 2014 and the extraordinary profits	positive
Neutral	تعلن الشركة المتقدمة للبتروكيماويات عن طريقة <u>صرف أرباح</u> الربع الرابع من العام المالي 2014م Advanced Petrochemical Company <u>announces the payment of profits</u> for the fourth quarter of fiscal year 2014	positive

Table 104 shows the phrase “distribute profits” [توزيع-ارباح] happened in six documents which is satisfy the rule [توزيع] --> [توزيع-ارباح], therefore it has been send again to the expert who labelled the document in the first stage. Moreover, it can have been seen form the structure six documents have been relabelled again to positive class. moreover, there are three documents the phrase “Exchange of profits” [أرباح-صرف] has been occurred with them which has the same meaning the phrase “distribute profits” [توزيع-ارباح].in other words, in Arabic language the word “distribution” (توزيع) has the same meaning for the term “cashing” (صرف).

Six documents have been found to satisfy the rule above and all three documents have been relabelled again to positive class. Moreover, the terms توزيع-ارباح has been occurred in the the three documents above with “ Announces” تعلن which is very important term in the neutral class and we believe this is the main reason that make the expert label the three documents above as neutral class. Moreover, the term “cashing” “صرف” has the same meaning in Arabic language for the term “distribution” “توزيع” and this was done by replacing “cashing” “صرف” by “Distribution” “توزيع”.

Figure 75, shows that the feature “distribution” (توزيع) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [توزيع] --> [توزيع-ارباح] (support: 0.021 confidence: 1) which represent the phrase “distribute profits” [توزيع-ارباح] that illustrate in the wordlist matrix in the negative class.

Premises	Conclusion	Support	Confidence
مشارك_ابريل	توزيع	0.004	0.667
موافق	توزيع	0.011	0.714
توزيع_ارباح	توزيع	0.021	1
عدم_توزيع	توزيع	0.015	1
موافق_عدم	توزيع	0.009	1
عموم_موافق	توزيع	0.004	1
تق_عدم	توزيع	0.004	1
تقر	توزيع	0.004	1
بعده_توزيع	توزيع	0.004	1
بعده	توزيع	0.004	1
ابريل_عموم	توزيع	0.004	1
يوص_بعده	توزيع	0.002	1
يوص	توزيع	0.002	1

Figure 75: The correlations of the term “Distribution” (توزيع) in negative class.

According to Figure 75 the rule [توزيع-ارباح] --> [توزيع] (support: 0.041 confidence: 1) that illustrate in the wordlist matrix in represent the phrase “capital increase” occurred in the negative class. Therefore, the next step is to search for the phrase “distribute profits” [توزيع-ارباح] in the negative class documents. Table 105 shows the phrase “distribute profits” [توزيع-ارباح] happened in ten documents which is satisfy the rule [توزيع] --> [توزيع-ارباح], therefore it has been send again to the expert who labelled the document in the first stage. Moreover, it can have been seen form the structure just one document have been relabelled again to positive class. Moreover, it can be seen form the structure in the rest of the documents which is 9 documents that phrase “distribute profits” توزيع أرباح has been occurred with the negation [توزيع] --> [عدم] which is the right place for this terms in the negative class.

Table 105: Term “distribute profits” [توزيع-ارباح] in the negative class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Negative	تأجيل عمومية فيفيكو المتضمنة الموافقة على توزيع أرباح Postpone the Vipco generality agreed on <u>dividend distribution</u>	Positive
Negative	مجلس إدارة العالمية للتأمين يوصي بعدم توزيع أرباح Global Insurance Board recommend on <u>not distribute dividend</u>	Negative
Negative	21 مايو عمومية الدرع العربي للموافقة على عدم توزيع أرباح May 21 general Arabian Shield for approval on <u>not distribute dividend</u>	Negative
Negative	غداً التصويت عن بُعد على بنود عمومية كيمانول والمتضمنة عدم توزيع أرباح Tomorrow remote voting on the general terms of Kimanol, including on <u>not distribute dividend</u>	Negative
Negative	ايس للتأمين تقر عدم توزيع أرباح وتنتخب أعضاء مجلس إدارتها ACE Insurance recognizes on <u>not distribute dividend</u> and elects its board of directors	Negative
Negative	تعلن شركة إتحد اتصالات موبايلى عن توصية مجلس الإدارة بعدم توزيع أرباح عن الربع الأول من العام المالي 2015م Etihad Etisalat announces the recommendation of the Board of Directors on <u>not distribute dividend</u> for the first quarter of the fiscal year 2015	Negative
Negative	إدارة موبايلى تقر عدم توزيع أرباح عن الربع الأول من هذا العام Mobily's management confirms Directors on <u>not distribute dividend</u> for the first quarter of this year	Negative
Negative	9 يونيو عمومية موبايلى للموافقة على عدم توزيع أرباح June 9 Mobily's approval for <u>not distribute dividend</u>	Negative
Negative	609 % خسائر المتراكمة حتى 30 أبريل 9 يونيو عمومية للموافقة على عدم توزيع أرباح %609 accumulated losses until 30 April 9 June general to approve on <u>not distribute dividend</u>	Negative
Negative	شركه صافولا توصي 22 يونيو عمومية للموافقة على عدم توزيع أرباح Savola recommends a June 22 general meeting to approve <u>not distribute dividend</u>	Negative

▪ Feature “reduction” (انخفاض)



Table 106 shows the feature “reduction” (انخفاض) as positive sentiment in Saudi Stock market domain. In addition, Table 106 shows the occurrence of the feature “reduction” (انخفاض) in positive, negative, and neutral class.

Table 106: Occurrence of the feature “reduction” (انخفاض).

Feature	Occurrence	Neutral	Positive	Negative
انخفاض	22	1	9	12

Figure 76 shows the association rules that related to the feature “reduction” (انخفاض) in all class with respect to the minimum support and minimum confidence threshold. The feature “reduction” (انخفاض) meaning some company in the Saudi stock market made their losses in stock market lower. In addition, Figure 76 shows the most important rules for the feature “reduction” (انخفاض) which is [انخفاض] --> [خسائر] (support: 0.042 confidence: 1). The term (Losses) [خسائر] correlated with the feature “reduction” [انخفاض] to compose positive phrases Low losses in the sentence.

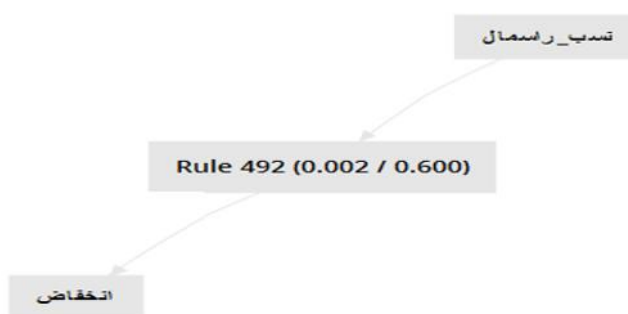


Figure 76: Visualize the association rules for the feature “reduction” (انخفاض) in all class.

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “reduction” (انخفاض). Table 107 shows that phrase Low losses [انخفاض-خسائر] occurred 5 times, 0 in neutral, 4 in positive class and 1 in negative class.

Table 107: rules for the term “reduction” (انخفاض) in the all data.

phrase -Terms	Occurrence	Neutral	Positive	Negative
انخفاض-خسائر	5	0	4	1

Same process has been done to find out the association rules that related to the feature “reduction” (انخفاض) in the negative class with respect to the minimum support and minimum confidence threshold. Table 108 shows the phrase “Low losses” [انخفاض-خسائر] happened in one documents which is satisfy the rule [انخفاض] --> [انخفاض-خسائر]. Therefore, it has been send

again to the expert who labelled the document in the first stage. one document has been found to satisfy the rule above and it is having been relabelled again to positive class

Table 108: Term “Low losses” [انخفاض - خسائر] in the negative class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Negative	انخفاض خسائر أنابيب 7% بالربع الأول <u>Loss of 7% pipe losses</u> in the first quarter	Positive

On the other hand, same process carried out for another negative feature “Fine” [غرامة]. Table 109 and Table 110 shows the number of Occurrence of the feature “Fine” (غرامة).

Table 109: Occurrence of the feature “Fine” (غرامة).

Term	Occurrence	Neutral	Positive	Negative
غرامة	26	0	0	26

Table 110: Rules for the term “Fine” (غرامة) in the all data.

Rule-Terms	Occurrence	Neutral	Positive	Negative
غرامة_مالية	26	0	0	26
فرض_غرامة	26	0	0	26

Table 111 shows in this negative sentiment feature, it could not find any noise because the rule has been filtering by the wordlist process and all the rules happened in the negative class which the right place. Same process carried out for other negative feature “postponement” (تأجيل).

Table 111: Occurrence of the feature “postponement” (تأجيل).

Term	Occurrence	Neutral	Positive	Negative
تأجيل	36	4	0	23

Table 112 shows that in this negative sentiment, it could not find any noise because the rule has been filter by the wordlist process and all the rules happened in the negative class which the right place.

Table 112: Rules for the term “postponement” (تأجيل) in the all data.

Rule-Terms	Occurrence	Neutral	Positive	Negative
تأجيل-عموميه	29	0	0	29

## **Appendix B – e-Learning Application**

The Deanship of e-learning and distance education at King Abdulaziz University develop application to collect tweets from their twitter account (see Figure 77, and 78). The data distribution depended on the methods Twitter's API utilised and the number of tweets posted on the distance education Twitter account page. The Twitter API has several methods, such as GET statuses, user\_timeline, home\_timeline, and GET search / tweets. This study implemented two methods to download tweets GET statuses/mentions\_timeline, and GET statuses/user\_timeline in order to collect the data from twitter platform. The application help the university to collect data from twitter, allow operators to label tweets as positive, negative, and neutral without the need to access the database. In addition, it helps to filter users according to their program types (such as e-learning students, full time student, part time students, or instructors).



عرض التقرير	السؤال	م
<a href="#">عرض التقرير</a>	يُقيم الطالب من (100) درجة: الأعمال الفصلية (30) درجة من المجموع الكلي، والاختبار النهائي: يمثل (70) درجة من المجموع الكلي.	1
<a href="#">عرض التقرير</a>	يُحرم الطالب من دخول الاختبار النهائي لمقرر التعليم عن بعد إذا تجاوزت نسبة غيابه 3 محاضرات (25%) من المحاضرات التزامنية.	2
<a href="#">عرض التقرير</a>	يجوز رفع الحرمان عن الطالب إذا ثبت حضوره لـ (5) محاضرات تزامنية، و(4) محاضرات مسجلة على الأقل.	3
<a href="#">عرض التقرير</a>	احتساب إعادة تشغيل المحاضرة المسجلة من قبل الطالب حضوراً إذا كان ذلك خلال أسبوع من المحاضرة المباشرة.	4
<a href="#">عرض التقرير</a>	الأعمال الفصلية مقسمة كالتالي: 4 اجابات لكل مادة ويحسب الواجب الواحد ب 3 درجات، اختبارين دوريين ولكل اختبار 6 درجات، المشاركة في المنتديات (لوحة المناقشات) عدد 3 مشاركات ولكل مشاركة درجتين (2) ليصبح المجموع الكلي 30 درجة للأعمال الفصلية.	5
<a href="#">عرض التقرير</a>	أي ملاحظات تود مشاركتنا بها عن آلية التعليم عن بُعد	6

First Previous 1 Next Last

Figure 77: show reports for collected tweets.

احتساب إعادة تشغيل المحاضرة المسجلة من قبل الطالب حضوراً إذا كان ذلك خلال أسبوع من المحاضرة المباشرة.

النوع : الكل.. ▼

بحث

التقييم : الكل.. ▼

تغيير التقييم	التقييم	الاجابة	الطالب/عضو هيئة تدريس	م
▼ اختر..	Normal	اسبوع قليل افضل لو كانت المدة مفتوحة طوال الفصل الدراسي اغلب طلاب وطالبات التعليم عن بعد لم يتجهوا لهذا النظام الدراسي الا لانشغالهم .	1397581	1
▼ اختر..	Postive	ممتاز	1454285	2
▼ اختر..	Postive	اتفق	1779839	3
▼ اختر..	Normal	ثلاثة أيام يكون مناسب ..	1650589	4
▼ اختر..	Postive	الوقت كافي لإستعادة تشغيل المحاضرات	1443407	5
▼ اختر..	Postive	ممتاز ر الافضل زيادة المدة اكثر من اسبوع لدرجة الحضور	1605445	6
▼ اختر..	Postive	موافق	1766551	7
▼ اختر..	Negative	اكثر تغيير اعتبره سلبي جدا ترانا موظفين ومنزوحين وارباب اسر ، حملي يدفعني مرات لالتداب في منطقته نائيه لمدة اسبوع ، او تمر بطرف اسري يمنعك	1548224	8
▼ اختر..	Negative	اسبوع غير كافي يجب ان تكون الترم كامل مثل النظام الاول ولكن لا ادري لماذا تم تغييره الى اسبوع ويجب الاخذ في الحسبان الانترنت ومشاكله	1440799	9
▼ اختر..	Postive	جيدة	1647563	10
▼ اختر..	Postive	ممتاز	1725994	11
▼ اختر..	Postive	قرار رائع ويتيح فيه التعويض	1738765	12
▼ اختر..	Postive	ممتاز	1710194	13
▼ اختر..	Postive	اتفق	1441571	14
▼ اختر..	Postive	نعم	1767298	15

Figure 78: labelling students tweets.

## Appendix C – Relabeling for Mubasher Case study

Third example, when users try to describe mubasher in case of it is interruption. The feature “Works” (شغال). Table 113 shows the feature “Works” (شغال) as negative sentiment and it can be happened in positive and neutral class. In addition, Table 113 shows the occurrence of the feature “Works” (شغال) in positive, negative, and neutral class.

Table 113: Occurrence of the feature “Works” (شغال).

Feature	Occurrence	Neutral	Positive	Negative
شغال	80	8	13	59

Figure 79 shows the association rules that related to the feature “Works” (شغال) in all class with respect to the minimum support and minimum confidence threshold. In addition, Figure 79 shows the most important rules for the feature “Works” (شغال) such as [شغال] --> [مباشر] (support: 0.011 confidence: 1). The term (Mubasher) correlated with the feature “Works” [شغال] to compose positive phrases mubasher is working during the trading time in the sentence.



Figure 79: Visualize the association rules for the feature “Works” (شغال).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “Works” (شغال). Table 114 shows that phrases that happened in all class with the feature “Works” (شغال).

Table 114: phrase for the feature “Works” (شغال) in the all data.

Phrase-Terms	Occurrence	Neutral	Positive	Negative
مباشر- شغال	24	4	7	13

From Table 114 the phrase “Mubasher is working” [مباشر- شغال] occurred in positive, negative and neutral class, and the feature “Works” (شغال) is consider as negative sentiment towards the mubasher product. As result, the neutral and positive classes become questionable classes. Therefore, the feature “Works” (شغال) needs further investigation in order to find out the association rules in the both classes. As result, two scenarios will be followed: First

scenario is extracting association rules that occurred for the feature “Works” (شغال) in the neutral class. Association rules generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Figure 80, shows that the feature “Works” (شغال) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [شغال] --> [مباشر- شغال] (support: 0.011 confidence: 1) which represent the phrase “mubasher is working” [شغال- مباشر] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
ياح_ظن	شغال	0.005	1
ياح	شغال	0.005	1
نظام_ماك	شغال	0.005	1
موبايل_تحميل	شغال	0.005	1
ممکن_استفسار	شغال	0.005	1
مطلوب_تصنيع	شغال	0.005	1
مطلوب	شغال	0.005	1
مهلك_وقت	شغال	0.005	1
مهلك	شغال	0.005	1
مباشر_ماهر	شغال	0.005	1
مباشر_فان	شغال	0.005	1
مباشر_شغال	شغال	0.005	1

Figure 80: The correlation rules of the feature “Works” (شغال) in neutral class.

According to Figure 80 the rule [افضل] --> [مباشر- شغال] (support: 0.005 confidence: 1) which represent the phrase “mubasher is working” occurred in the neutral class. Therefore, the next step is to search for the phrase “mubasher is working” [شغال- مباشر] in the neutral class documents. Table 115 shows the phrase “mubasher is working” [شغال- مباشر] happened in four documents.

Table 115: phrase “user prefer mubasher product” [مباشر- شغال] in the neutral class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	مباشر ماهر شغال عندي زين كم مطلوبه التصنيع يا اهل مباشر	Negative
	<u>Mubasher</u> is not <u>working</u> with me I need to know how much Zain share price.	
Neutral	هل يوجد برنامج مباشر برو يشتغل علي نظام الماك	Neutral
	Is there a <u>mubasher</u> program <u>running</u> on the Mac?	
Neutr Neutral	ياخي علمني وشلون يشتغل مباشر على الايباد	Negative

	He taught me how to <u>work mubasher</u> on the iPad	
Neutral	انا مثلك ودقيبت على <u>مباشر</u> قالي اتصال موبايلي تحميله ضعيف جدا وقال غير الخط الى الاتصالات او اي خط اخر <u>يشتغل</u>	Negative
	I'm just like you and I called on <u>mubasher</u> , they told me Mobily's connection is very weak and they said I have to change the line to STC line that can <u>works</u> .	

Table 115 shows the phrase “mubasher is working” [مباشر - شغال] happened in four documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above documents has meaning of discussion among users about mubasher products and its platforms which make the right class for those documents is the neutral class.

Figure 81, shows that the feature “Works” (شغال) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the second scenario, the interested rule here is [شغال] --> [مباشر - شغال] (support: 0.011 confidence: 1) which represent the phrase “mubasher is working” [مباشر - افضل] that illustrate in the wordlist matrix in the negative class.

Premises	Conclusion	Support	Confidence
رياض	شغال	0.005	0.667
كساح	شغال	0.005	0.667
عند_مباشر	شغال	0.016	0.750
شغال_عند	شغال	0.013	1
مباشر_شغال	شغال	0.011	1
نسخة_الذويدي	شغال	0.005	1
نسخة	شغال	0.005	1
نزلت_نسخة	شغال	0.005	1
نزلت	شغال	0.005	1
موقع_تداول	شغال	0.005	1
مباشر_نزلت	شغال	0.005	1
مباشر_موقع	شغال	0.005	1

Figure 81: The correlation rules of the feature “Works” (شغال) in negative class.

The next step is to search for the phrase “mubasher is working” [مباشر - شغال] in the negative class documents. Table 116 shows the phrase “mubasher is working” [مباشر - شغال] happened in 7 documents.



Table 116: phrase “user prefer mubasher product” [مباشِر- شغال] in the positive class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Positive	على نوت 4 انا عندي مباشِر برو 10 ونزلت نسخت الاندرويد وشغال I have a mubsher Pro 10 and I downloaded the version of the android and its work on Note 4	Positive
Positive	انا شغال مثل العسل على مباشِر برو9 أنصحك قفله وقفل النت وشغلهم الاثنين من جديد It works like honey on mubasher Pro 9 I advise you to have to lock it and lock the net and lunch them again	Positive
Positive	سلام انا عندي مباشِر من موقع تداول الراجحي شغال 100% Salam, I have mubasher from the site of trading Al Rajhi 100% working	Positive
Positive	شغال عندي مع مباشر من قبل الافتتاح I was working with mubasher by opening	Positive
Positive	مباشِر برو شغال عندي من بنك الرياض من سنين زي اللوز Mubasher Pro has been working for me from Riyadh Bank since the years of almonds	Positive
Positive	مباشِر برو شغال مثل العسل الحمد لله إذا واجهتك أي مشكلة أو عندك أي أسفاسر أتصل على الدعم الفني ويتجاوب معك سريع Mubasher Pro is <u>working</u> like honey if you have any problem or you have any questions contact technical support and respond quickly	Positive
Positive	والله ياخوي ما اقدر أفيد وشو الحل بس مباشِر شغال عندي ممتاز my brother, I do not know what I can say, but <u>mubasher</u> working with me excellent	Positive

Table 116 shows the phrase “mubasher is working” [مباشِر- شغال] happened in 7 documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above document has meaning of user’s trying to write about their experience about mubasher product during daily trading which make the labelling process very clear and there was no mislabeling has happened in the above documents.

Fourth example, when users try to describe mubasher in case of it is work or interruption. The feature “subscription” (اشتراك). Table 117 shows the feature “subscription” (شغال) as positive sentiment and it can be happened in negative and neutral class. In addition, Table 117 shows the occurrence of the feature “subscription” (اشتراك) in positive, negative, and neutral class.

Table 117: Occurrence of the feature “subscription” (اشتراك)

Feature	Occurrence	Neutral	Positive	Negative
اشتراك	142	14	75	53

Figure 82 shows the association rules that related to the feature “subscription” (اشتراك) in all class with respect to the minimum support and minimum confidence threshold. In addition, Figure 82 shows the most important rules for the feature “subscription” (اشتراك) such as [اشتراك] --> [مباشر] (support: 0.034 confidence: 1. The term (Mubasher) correlated with the feature “subscription” [اشتراك] to compose positive phrases subscription with mubasher in the sentence.

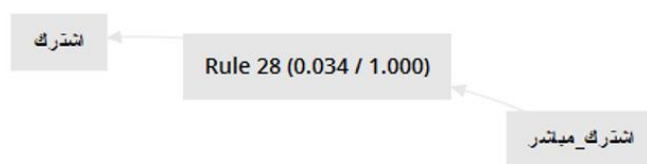


Figure 82: Visualize the association rules for the feature “subscription” (اشتراك).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “subscription” (اشتراك). Table 118 shows that phrases that happened in all class with the feature “subscription” (اشتراك).

Table 118: phrase for the feature “subscription” (اشتراك) in the all data

Phrase-Terms	Occurrence	Neutral	Positive	Negative
اشتراك- مباشر	47	7	25	15

From Table 118 the phrase “subscription with mubasher” [اشتراك - مباشر] occurred in positive, negative and neutral class, and the feature “subscription” (اشتراك) is consider as positive sentiment towards the mubasher product. As result, the neutral and negative classes become questionable classes. Therefore, the feature “subscription” (اشتراك) needs further investigation to find out the association rules in the both classes. As result, two scenarios will be followed: First scenario is extracting association rules that occurred for the feature “subscription” (اشتراك) in the neutral class. Association rules generated regarding the minimum support and minimum confidence threshold using the previous process of the visualisation.

Figure 83, shows that the feature “subscription” (اشتراك) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [اشتراك] --> [مباشر- اشتراك] (support: 0.005 confidence: 1) which represent the phrase “subscription with mubasher” [مباشر- اشتراك] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
مباشِر يفتحه	اشترك	0.005	1
مباشِر يعمل	اشترك	0.005	1
مباشِر رياض	اشترك	0.005	1
مباشِر راجح	اشترك	0.005	1
مباشِر تقدر	اشترك	0.005	1
مباشِر اطم	اشترك	0.005	1
مباشِر اشكال	اشترك	0.005	1
مباشِر اشترك	اشترك	0.005	1
ماحصل ادب	اشترك	0.005	1
ماحصل	اشترك	0.005	1
ماجتن رسلكم	اشترك	0.005	1
ماجتن	اشترك	0.005	1
مؤشر بيان	اشترك	0.005	1
لتداول راجح	اشترك	0.005	1

Figure 83: The correlation rules of the feature “subscription” “اشترك” in neutral class.

According to Figure 83 the rule [مباشِر - اشترك] --> [اشترك] (support: 0.005 confidence: 1) which represent the phrase “mubasher is working” occurred in the neutral class. Therefore, the next step is to search for the phrase “subscription with mubasher” [مباشِر - اشترك] in the neutral class documents. Table 119 shows the phrase “subscription with mubasher” [مباشِر - اشترك] happened in seven documents.

Table 119: phrase “subscription with mubasher” [مباشِر - اشترك] in the negative class documents

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	من عنده علم يفيدنا هل فيه طريقه اخرى للاشتراك في مباشر برو الراجحي From him aware Is it tells us another way to make <u>subscription</u> with mubasher in the -Rajhi	Negative
Neutral	السلام عليكم ياخوان ما عندي اشترك لافي مباشر برو ولا تكثر تشتارت من الافضل Salam alaikoum. I do not have <u>subscription</u> to <u>mubasher</u>	Neutral
Neutral	شباب كيف اشغل مباشر على الايباد .. اعرف مافيه جافا بس فيه ناس تقول اشترك من موقع مباشر مدري يحملون برامج How to work with <u>mubasher</u> on iPad. Some people say <u>subscription</u> from the site direct	Negative
Neutral	ذا قصدك مباشر الرياض فقد تكون مشترك لمدة شهر وهي مده تجريبه من البنك What do you mean mubasher Riyadh may be a joint for a month is an extended period	Neutral
Neutral	ابو عزام بالنسبة لتداول الراجحي مباشر تقدر تشوف المؤشرات والرسم البياني مثل مباشر برو اشتركي في مباشر برو قرب ينتهي وافكر بالراجحي Abu Azzam for Al Rajhi mubasher trading estimated to show indicators and chart like <u>Mubasher</u> Pro, my	Negative

	<u>subscription</u> in Mubasher Pro near ends and think of Al Rajhi	
Neutral	السلام عليكم تواصلت مع الدعم الفني على رقمكم افادوني قبل عشرة ايام بأن الاشتراك للايباد في مباشر فيه اشكالية هل تم حلها؟	Negative
	Hi, I have continued with the technical support on your number. They told me ten days ago that <u>the subscription</u> to the iPad in <u>Mubasher</u> is problematic. Has it been resolved?	
Neutral	صباح الخير سددت اشتراك ف مباشر ايباد وماجتني رسالتكم ع الايميل ولا ع الجوال	Negative
	Good morning, I paid a mubasher subscription to IBAD and did not got any e-mail or text message	

Table 119 shows the phrase “subscription with mubasher” [مباشر- اشتراك] happened in 7 documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above document has meaning of user’s trying to write about their normal experience about mubasher product. However, some of those documents has been labelled wrong due to unclearly negative experience among the text.

Figure 84, shows that the feature “subscription” (اشتراك) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [مباشر- اشتراك] --> [اشتراك] (support: 0.005 confidence: 1) which represent the phrase “subscription with mubasher” [مباشر- اشتراك] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
مباشر_جريت	اشتراك	0.003	0.667
فكرت	اشتراك	0.003	0.667
التهير	اشتراك	0.003	0.667
الشر	اشتراك	0.003	0.667
الشر_مباشر	اشتراك	0.021	1
مباشر_الشر	اشتراك	0.004	1
لغيت_الشر	اشتراك	0.004	1
لغيت	اشتراك	0.004	1
الشر_الفضل	اشتراك	0.004	1
يقول_مجرد	اشتراك	0.003	1
يخلص_الشر	اشتراك	0.003	1
يخلص	اشتراك	0.003	1

Figure 84: The correlation rules of the feature “subscription” “اشتراك” in negative class.

According to Figure 84 the rule [مباشر- اشتراك] --> [اشتراك] (support: 0.021 confidence: 1) which represent the phrase “mubasher is working” occurred in the neutral class. Therefore, the next

step is to search for the phrase “subscription with mubasher” [مباشر- اشتراك] in the neutral class documents. Table 120 shows the phrase “subscription with mubasher” [مباشر- اشتراك] happened in seven documents.

Table 120: phrase “subscription with mubasher” [مباشر- اشتراك] in the negative class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Negative	رفض تطبيق مباشر برو العمل جربت اكثر من ثلاث اشتراكات برضه عطلان هل تم ايقاف من قبل الهيئه او ماذا mubasher Application does not work, I have tried more than three subscriptions and all of them have been suspended	Negative
Negative	عندي اعطي توثيق غير صحيح اذل فتحت مباشر مع العلم ان الاشتراك انتهى I have given incorrect authentication, when I opened mubasher with the knowledge that the subscription is over	Neutral
Negative	عندي اشتراك مباشر موبايلي ايفون وتفاجاة انه عطلان لا أعلم ماهو السبب I have joint mubasher Mobily iPhone and surprise that it is a failure I do not know why	Negative
Negative	كان اشتراكي من سنين مع مباشر برو دقة البيانات جيدة لكنها تتأخر اللحظية ومجلد الهميستوري مشاكل عندي ما يحدث زين وتطلع الشموع رايحة فيها I have been a subscribed for years with Mubasher Pro. The accuracy of the data is good, but the delay is temporary and the folder is hysterical.	Negative
Negative	عد انتهاء اشتراكي بمباشر برو انتقلت لتكرتشارت وانتهت المشكلة وأصبحت الشمعات تخرج بشكل سليم After the end of my subscription, Mubasher Pro moved to Tkrchat, and the problem ended and the candles came out properly	Negative
Negative	طبعاً البرنامج ينفع للي ماعنده اشتراك مباشر والا الاتجاه وتكرتشارت افضل منه بكثير Of course, the program can benefit who has no mubasher involvement the trend, and it is much better	Negative
Negative	ياخوان ياليت تفيدوني عن تكرتشارت كم اسعار الاشتراك هل هو افضل من مباشر برو Brothers, tells me about Tkrchart how much subscription prices are better than mubasher Pro	Neutral
Negative	يا ابن الحلال فيه احد يتداول عبر مباشر الينوك شف لك اشتراك شهري 170 مستشاري او 215 مستثمر او برو وريح راسك If you are a trader, you will have a monthly subscription of 170 advisors, 215 investors or mubasher	Neutral
Negative	نتظر يخلص اشتراك مباشر برو ونجرب الاتجاه We wait for mubasher subscription to end and we will try the trend	Negative
Negative	لكثير اشتكوا من مباشر سابقا وما زالوا يشتكون الشركة توزع اشتراكات مجانية وهذا دليل تخبط Many complained of Mubasher previously and still complain about the company distributing free subscriptions and this guide is floundering	Negative
Negative	خلص امس الليل اشتراكي في مباشر وما مداني اجدد نحين اشتكرت في تكرتشارت Last night, my subscription with mubasher and i participated in Tkrchat	Negative

Negative	خلص اشتر اكي واحب اغير افيدوني عن برنامج افضل من مباشر برو	Neutral
	My <u>subscription</u> end and I would like to change the program better than <u>Mubasher</u>	
Negative	اشترك في مباشر برو عن طريق البنك 6 شهور افضل لك انا مشترك عن طريق بنك الرياض روووعه لاتقطع ولا شحن ولا شي	Positive
	<u>Subscribe</u> with <u>Mubasher</u> Pro through the bank 6 months better for you I am a subscriber through Riyadh Bank Rowaah for cutting and not shipping and nothing	
Negative	لشركه المنتجه للميتاستوك رفضت تزويد البيانات التاريخيه من قبل مباشر برو لاجبارك على الاشتراك في الميتاستوك الاصليه	Negative
	The <u>Metastock</u> company refused to provide historical data by <u>Mubasher</u> To force you to <u>subscribe</u> to the original Stock Market	
Negative	مباشر برو بيخلص الاشتراك بعد يومين وبأكتسلهم	Negative
	<u>Mubasher</u> <u>subscription</u> will end after two days and I will leave them	

Table 120 shows the phrase “subscription with mubasher” [مباشر- اشترك] happened in 15 documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above documents has meaning of user’s trying to write about their observation about mubasher product subscription and most of that was negative experience. On the other hand, labelling process also has been done wrongly in some situation due to the complexity among the text to decide in which class should be belong to.

Fifth example, when users try to contact mubasher technical team in order to solve some technical issues. The feature “Support” (الدعم). Table 121 shows the feature “Support” (الدعم) as negative sentiment and it can be happened in positive and neutral class. In addition, Table 121 shows the occurrence of the feature “Support” (الدعم) in positive, negative, and neutral class.

Table 121: Occurrence of the feature “Support” (الدعم)

Feature	Occurrence	Neutral	Positive	Negative
الدعم	34	7	8	19

Figure 85 shows the association rules that related to the feature “Support” (الدعم) in all class with respect to the minimum support and minimum confidence threshold. In addition, Figure 85 shows the most important rules for the feature “Support” (الدعم) such as [الفنى] --> [الدعم] (support: 0.017 confidence: 1. The term (Technical) correlated with the feature “Support” [الدعم] to compose positive phrases technical support in the sentence.



Figure 85: Visualize the association rules for the feature “Support” “الدعم”.

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “Support” (الدعم). Table 122 shows that phrases that happened in all class with the feature “Support” (الدعم).

Table 122: phrase for the feature “Support” (الدعم) in the all data.

phrase -Terms	Occurrence	Neutral	Positive	Negative
الدعم- الفني	25	4	4	17

From Table 122 the phrase “Technical support” [الدعم - الفني] occurred in positive, negative and neutral class, and the feature “Support” (الدعم) consider as negative sentiment towards the mubasher product. As result, the neutral and positive classes become questionable classes. Therefore, the feature “Support” (الدعم) needs further investigation in order to find out the association rules in the both classes. As result, two scenarios will be followed: First scenario is extracting association rules that occurred for the feature “Support” (الدعم) in the neutral class. Association rules generated with regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Figure 86, shows that the feature “Support” (الدعم) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the First scenario, the interested rule here is [الدعم] --> [الدعم - الفني] (support: 0.017 confidence: 1) which represent the phrase “Technical support” [الدعم - الفني] that illustrate in the wordlist matrix in the neutral class.

Show rules matching

all of these conclusions: ▼

- تجربة
- التطبي
- فني
- الضغط
- الشكر
- الدولة
- الهدم
- الدعم الفني
- الدعم**
- الشاول
- البيك
- الايباد
- الاشترك
- اعرف
- لرجو
- ابي
- ألفينس
- ...

Premises	Conclusion	Support	Confidence
الفني	الدعم	0.020	1
<b>الدعم الفني</b>	الدعم	0.020	1
يقوم فريق	الدعم	0.005	1
يقوم	الدعم	0.005	1
بصلحوا الأخطاء	الدعم	0.005	1
بصلحوا	الدعم	0.005	1
ويستلمون رواتب	الدعم	0.005	1
ويستلمون	الدعم	0.005	1
موظفين الدعم	الدعم	0.005	1
موظفين	الدعم	0.005	1
مباشرة تكاليفه	الدعم	0.005	1

Figure 86: The correlation rules of the feature “Support” “الدعم” in neutral class.

According to Figure 86 the rule [الدعم] --> [الدعم - الفني] (support: 0.020 confidence: 1) which represent the phrase “Technical support” occurred in the neutral class. Therefore, the next step is to search for the phrase “Technical support” [الدعم - الفني] in the neutral class documents.

Table 123: shows the phrase “Technical support” [الدعم - الفني] happened in neutral class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Neutral	الرجاء الإتصال على الرقم المجاني 920019333 و سوف يقوم فريق <u>الدعم الفني</u> بمساعدتك	Neutral
	Please call toll free number 920019333 and the <u>technical support</u> team will assist you	
Neutral	تراهم موظفين حول 40 واحد في الدعم الفني ويستلمون رواتب كاملة	Negative
	see about 40 employees in <u>technical support</u> and receive full salaries	
Neutral	دقوا على <u>الدعم الفني</u> يصلحوا لكم الأخطاء	Neutral
	call technical support in order to correcting your mistakes	
Neutral	تواصلت مع <u>الدعم الفني</u> على رقمكم افادوني قبل عشرة ايام بأن الاشتراك للايباد في مباشر فيه اشكالية هل تم حلها؟	Negative
	I have continued with the <u>technical support</u> on your number. They told me ten days ago that the subscription to the iPad in Mubasher is problematic. Has it been resolved?	



Table 123 shows the phrase “Technical support” [الدعم - الفني] happened in 4 documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above documents has meaning of user’s trying to write about their observation about mubasher technical support team and all most of that was negative experience. On the other hand, labelling process also has been done wrongly in some situation due to the complexity among the text to decide in which class should be belong to.

Figure 87, shows that the feature “Support” (الدعم) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the Second scenario, the interested rule here is [الدعم] --> [الدعم - الفني] (support: 0.017 confidence: 1) which represent the phrase “Technical support” [الدعم - الفني] that illustrate in the wordlist matrix in the neutral class.

Premises	Conclusion	Support	Confidence
لبرنامج	الدعم	0.005	0.667
الحساب	الدعم	0.005	0.667
الدعم الفني	الدعم	0.013	1
وكل ما يخص	الدعم	0.005	1
وكل	الدعم	0.005	1
ما يخص عروض	الدعم	0.005	1
ما يخص	الدعم	0.005	1
للتشرك الدعم	الدعم	0.005	1
لبرنامج مباتر	الدعم	0.005	1
عروض الإنترنت	الدعم	0.005	1
برو وكل	الدعم	0.005	1
الفني لبرنامج	الدعم	0.005	1

Figure 87: The correlation rules of the feature “Support” “الدعم” in positive class

According to Figure 87 the rule [الدعم] --> [الدعم - الفني] (support: 0.013 confidence: 1) which represent the phrase “Technical support” occurred in the positive class. Therefore, the next step is to search for the phrase “Technical support” [الدعم - الفني] in the positive class documents. Table 124 shows the phrase “Technical support” [الدعم - الفني] happened in 4 documents.

Table 124: shows the phrase “Technical support” [الدعم - الفني] happened in neutral class documents.

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Positive	اتصل على الدعم الفني 920019333 وستجد مايسرك	Positive
	Call <u>Technical Support</u> 920019333 and you will find your help	
Positive	للاشتراك أو الدعم الفني لبرنامج مباشر برو وكل ما يخص عروض الاشتراكات على الحساب	Neutral
	To subscribe or <u>technical support</u> to Mubasher Pro and all the benefits of subscription offers on the account	
Positive	خدمة ما بعد البيع من أهم اهدافنا وتواصلنا السريع مع عملائنا هو من جعلنا في القمة تواصل معنا هنا او على هاتف الدعم الفني 920019333	Positive
	After sales service is one of our main goals Our fast connection with our customers is who made us at the top Contact us here or call our <u>technical support</u> number 920019333	
Positive	مباشر برو شغال مثل العسل الحمد لله إذا واجهتك أي مشكلة أو عندك أي أسئلة اتصل على الدعم الفني ويتجاوب معك سريع	Positive
	Mubasher Pro is working like honey if you have any problem or you have any questions contact <u>technical support</u> and respond quickly	

Table 124 shows the phrase “Technical support” [الدعم - الفني] happened in 4 documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above documents has meaning of user’s trying to write about their observation about mubasher technical support team and all most of that was negative experience. On the other hand, labelling process also has been done wrongly in some situation due to the complexity among the text to decide in which class should be belong to.

Finally, when users try to describe mubasher in case of it is interruption. The feature “Disrupt” (عطلان). Table 125 shows the feature “Disrupt” (عطلان) as negative sentiment and it can be happened in positive class. In addition, Table 125 shows the occurrence of the feature “Disrupt” (عطلان) in positive, negative, and neutral class.

Table 125: Occurrence of the feature “Disrupt” (عطلان)

Feature	Occurrence	Neutral	Positive	Negative
عطلان	82	0	2	80

Figure 88 shows the association rules that related to the feature “Disrupt” (عطلان) in all class with respect to the minimum support and minimum confidence threshold. In addition, Figure 88 shows the most important rules for the feature “Disrupt” (عطلان) such as [عطلان] --> [مباشر] (support: 0.017 confidence: 1. The term (Mubasher) correlated with the feature “Disrupt” [عطلان] to compose negative phrases mubasher is not working during the trading time in the sentence. Moreover, the rule [عطلان] --> [عندي] (support: 0.011 confidence: 1. The term (Disrupt) correlated with the feature “Disrupt” [عندي] to compose negative phrases mubasher is not working with me during the trading time in the sentence. Moreover,

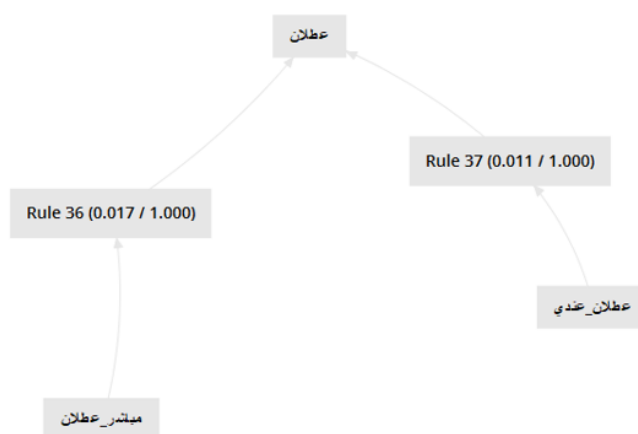


Figure 88: Visualize the association rules for the feature “Disrupt” (عطلان).

The next step is to find out from the wordlist representation the occurrence of the most frequent phrases that related with feature “Disrupt” (عطلان). Table 126 shows that phrases that happened in all class with the feature “Disrupt” (عطلان).

Table 126: phrase for the feature “Disrupt” (عطلان) in the all data.

phrase -Terms	Occurrence	Neutral	Positive	Negative
مباشر- عطلان	28	0	0	28
عطلان- عندي	17	0	1	16

From Table 126 the phrase “Mubasher is not working” [عطلان - مباشر] occurred in negative class only, the feature “Disrupt” (عطلان) is consider as negative sentiment towards the mubasher product. As result, no need for more investigation. On the other hand, the phrase “Mubasher is not working with me” [عطلان- عندي] occurred in negative class and positive class.as result, the positive class become questionable class. Therefore, the feature “Disrupt” (عطلان) needs further investigation in order to find out the association rules in the positive classes. As result, second scenarios will be followed: second scenario is extracting association rules that occurred for the feature “Disrupt” (عطلان) in the neutral class. Association rules generated with

regard to the minimum support and minimum confidence threshold using the previous process of the visualisation.

Figure 89, shows that the feature “Disrupt” (عطلان) can occurred with many rules that appeared in the premises column with the minimum support and minimum confidence values. However, according to the Second scenario, the interested rule here is [عطلان] --> [عندي - عطلان] (support: 0.011 confidence: 1) which represent the phrase “mubasher is not working with me” [عندي - عطلان] that illustrate in the wordlist matrix in the positive class.

Premises	Conclusion	Support	Confidence
ممتاز_عندي	عطلان	0.003	1
فترة_بسيطة	عطلان	0.003	1
فترة	عطلان	0.003	1
غريب_مباشر	عطلان	0.003	1
غريب	عطلان	0.003	1
عندي_عطلان	عطلان	0.003	1
عطلان_فترة	عطلان	0.003	1
بسيطة	عطلان	0.003	1
الراجحي_ممتاز	عطلان	0.003	1

Figure 89: The correlation rules of the feature “Disrupt” “عطلان” in positive class.

According to Figure 89 the rule [عطلان] --> [عندي - عطلان] (support: 0.003 confidence: 1) which represent the phrase “mubasher is not working with me” occurred in the positive class. Therefore, the next step is to search for the phrase “mubasher is not working with me” [عندي - عطلان] in the neutral class documents. Table 127 shows the phrase “mubasher is not working with me” [عندي - عطلان] happened in one document.

Table 127: phrase “user prefer mubasher product” [عندي - عطلان] in the positive class documents

Original Labelling class	Original Arabic tweets with English translation	New Labelling class
Positive	غريب مباشر الراجحي ممتاز <u>عندي تعطل</u> فترة بسيطة فقط	Negative
	<u>Mubasher</u> is not <u>working</u> with me I need to know how much Zain share price,	

Table 127 shows the phrase “mubasher is not working with me” [عندي - عطلان] happened in one documents. Therefore, it has been sent again to the expert who labelled the document in the first stage. Moreover, it can have been seen from the structure in the of the above documents has meaning of discussion among users about mubasher interruption which make the right class for those documents is the negative class.