Using deep learning networks to predict telecom company customer satisfaction based on Arabic tweets

Latifah AbdulRahman Almuqren
Princess Nourah bint Abdulrahman University/Durham University
Riyadh/Saudi Arabia *Laalmuqren@pnu.edu.sa*

Monzer Moh'd Qasem
Princess Nourah bint Abdulrahman University
Riyadh/Saudi Arabia *mmqaseem@pnu.edu.sa*

Alexandra I. Cristea
Durham university
Durham/United Kingdom *alexandra.i.cristea@durham.ac.uk*

Abstract

Information systems are transforming businesses, which are using modern technologies towards new business models based on digital solutions, which ultimately lead to the design of novel socio-economic systems. Sentiment analysis is, in this context, a thriving research area. This paper is a case study of Saudi telecommunications (telecom) companies, using sentiment analysis for customer satisfaction based on a corpus of Arabic tweets. This paper compares, for the first time for Saudi social media in telecommunication, the most popular machine learning approach, support vector machine (SVM), with two deep learning approaches: long short-term memory (LSTM) and gated recurrent unit (GRU). This study used LSTM and GRU with two different implementations, adding attention mechanism and character encoding. The study concluded that the bidirectional-GRU with attention mechanism achieved a better performance in the telecommunication domain and allowed detection of customer satisfaction in the telecommunication domain with high accuracy.

Keywords: Telecommunications, Customer Satisfaction, Arabic, sentiment analysis, Deep learning.

1. Introduction

Global competition in the labour market drives companies to enhance their customer satisfaction [9]. Traditionally, customer satisfaction has been measured using customer interviews and questionnaires [17], but this is changing [3]. It is influenced, amongst other factors, by the advent of social media, which has rapidly become a key part of many people's lives. Supported by big data and the social web, social media has become an easy tool to share opinions, sentiments and moods [22]. Therefore, marketers have started to use social media to monitor and measure customer satisfaction with their services.

The Arabic language is used by a large segment of the world's population, as the fourth most widespread language [5]. Despite that, Arabic Sentiment Analysis (ASA) is still under-researched [5]. This is in part due to the challenging nature of the Arabic language, as each Arabic dialect has its own syntax and vocabulary, which complicates, for instance, any attempts to build an Arabic lexicon [6].

Thus, our work seeks to answer the following research questions (RQs) related to predicting customer satisfaction:

RQ1: *What is the best method to automatically measure and make automatic predictions about customer satisfaction for telecom companies in Saudi Arabia from tweets?*
RQ2: *Are there any interesting sentiment-related novel patterns that can be extracted for telecom companies from the Twitter corpus?*

This paper's main original contributions are:
1. The first work to evaluate customer satisfaction for telecommunications (telecom) companies in Saudi Arabia by using *both* supervised learning approach and different deep learning networks on social media, as well as *comparing* these approaches.
2. The first corpus of Saudi tweets related to telecom companies, comprising 16,000 tweets for training data and 4,000 tweets for testing.

2. Related Researches

There is little research on customer satisfaction in terms of sentiment analysis in the literature. Below we group the limited research in this area, and then we visit potential other research areas, such as deep learning, which are promising to be used in sentiment analysis for Arabic corpora (ASA).

2.1. Customer Satisfaction and Sentiment Analysis

Customer satisfaction (CS), including CS in information systems, is attained by examining customer expectations towards a company's product [13]. After reviewing case studies on the applications and methods used to analyse CS, we found that the survey is still the most popular tool to measure CS [6], [17]. From an analysis of the literature, it appears that only very few studies measure customer satisfaction, particularly in the telecommunications industry using social media mining, as shown in Table 1. All relevant studies used sentiment as a variable that links Twitter features and customer satisfaction. Our work followed the sentiment analysis approach to measure customer satisfaction towards telecom companies, striving however to achieve it in real time.

Table 1. Gap analysis in literature review

Author [Ref. No.]	**Aim**	**Technology**	**Data Set**	**Findings**	**Gap Identified**
[12]	Measured customer satisfaction for two online transportation service providers in Indonesia.	SA using SVM, NB and DT.	9,191 tweets	Customers preferred to express bad sentiments on the companies' Twitter account, instead of positive; SVM and DT had the highest performance.	They did not use features in pre-processing and classifying the data, which could have given better classifier results.
[31]	Measured customer satisfaction towards telecommunication companies in Saudi Arabia using different algorithms.	SA using KNN, NB and ANN.	1331 tweets	KNN was superior to the other algorithms with 75.6% for F-measure.	The data set included only English tweets, although that the majority of customer tweets about Saudi telecommunication companies are in Arabic, limiting the capture of customers' real sentiments.
[27]	Analysed Jordanian telecommunication companies' customers' comments on the Facebook.	SA using KNN, SVM, NB, and DT	14332 customer posts on Facebook	SVM classifier outperformed the other three classifiers with 95% accuracy.	They classified the comments into positive, negative, other, or question, considered the negative and positive comments and discarded the 'other' and 'question' classifications.

SA sentiment analysis, KNN k-nearest neighbour, ANN artificial neural networks, NB naïve Bayesian

Several studies highlighted the benefits of sentiment analysis for organisations [36]. According to Sohangi et al., [36] sentiment analysis can help the organisations to support decision makers in predicting stock market through identifying the feelings of financial social network users. Other researchers considered mining social media data is important for marketers and customers for several reasons: producing an abundance of useful data, which provides a wealth of information about customers for the company [20], it helps to develop a recommendation system to maintain existing customers or gain new ones, and it is also useful in building confidence among customers and stakeholders [38].

2.2. Deep Learning and Sentiment Analysis

The approaches used for SA are machine-learning methods [19], lexicon-based approaches [2] and hybrid approaches that combine the two techniques [10]. Some of the popular supervised learning methods that are used with Arabic Sentiment Analysis (ASA) include NB, SVM and KNN. These methods are the most popular in building effective corpora, especially when downloading from social media, such as Twitter and Facebook [24]. There have been several studies based on the idea that SVM is competitively effective in supervised sentiment classification, especially in an Arabic context [3], highly accurate, and widely accepted, and that it renders high precision, accuracy and recall [19] and achieves better accuracy than NB and K-NN [15]. It is clear from the literature that SVM is a high-performing method for ASA, because it helps in efficient classification of texts, eliminating possible biases and mix-ups, by virtue of forming maximised margins between two classes represented by a hyperplane. In addition, SVM has the ability to handle some text-classification problems, such as avoiding noise in microblogging [40].

Deep learning is a process of training neural networks with multiple hidden layers using multiple representations of data with multiple levels of abstraction, such that generalizations are much more accurate than for shallow neural networks [30]. The most prominent deep learning network is the recurrent neural network (RNN). In a deep learning RNN, the outputs of a forward hidden layer may feedback to a prior hid-den layer, rendering the learning process internally interpretative through an algorithmic process called backpropagation [30]. Backpropagation was introduced in RNNs, where feedback loops from hidden layers feed an error function from a forward layer to the previous layers [39]. RNNs are capable of deep learning that can discover hidden organizations and structures through backpropagation in massive-scale data structures that have time-dependent (sequential) relationships [30]. Historically, the key problems faced in RNN's backpropagation were the explosion and vanishing of the gradients of error functions, when a very deep learning process was executed [39]. This problem was solved in LSTM networks, which were capable of very deep learning without the exponential explosive or shrinking effects [39].

The role of LSTM is crucial in conducting sentiment analysis of a corpus that has multiple aspects with deep contexts [30], [18]. For example, the contexts of long-running discussions such as political discussions running for many months on media or on Twitter, require the capability of LSTM. LSTM is very useful when the aspects and contexts are very deep, hierarchical, sequentially time-dependent or are following a time series [18]. The training data may comprise corpora with multiple contexts with changing aspects over time [18]. In short textual corpora, such as Twitter data, is a prime candidate for LSTM, as here the contextual classifications are much more prominent.

More recently, GRU were proposed, and can be considered as a variation on the LSTM, because both are designed similarly and, in some cases, produce equally excel-lent results. Both LSTM and GRU have a great impact on improving sentiment analysis, not only because they solved the challenge of vanishing and exploding gradients in RNNs, but also because they have opened many new ways of designing deep machine learning architecture for different scenarios of sentiment analysis.

Several studies use deep learning in the sentiment analysis. Alwehaibi and Roy [11] used Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) on three Arabic data sets collected from Twitter AvaVec, ArabicNews and AraFT. The results showed that AraVec achieved 88% in term of accuracy, ArabicNews scored 91% and AraFT achieved 93.5%. Additionally, the results showed that pre-trained Word Embedding enhanced the performance of the model.

In addition, Al-Smadi et al., [4] used two implementations of LSTM; the first one is bidirectional LSTM with conditional random field classifier (Bi-LSTM-CRF) for aspect opinion target expressions, and the second one is an aspect-based LSTM for aspect sentiment analysis. These approaches are trained on Arabic hotels' review. They used character and word embedding features. The result showed that their approach outperformed when compared with the state of art.

Moreover, Sohangi et al., [36] used deep learning to enhance the performance of sentiment analysis in the financial social network Stock Twits. They used LSTM, convolutional neural networks and doc2vec. Their results showed that deep learning raised the accuracy of the financial sentiment analysis. In addition, the convolutional neural network model outperformed the other models.

3. Corpus Collection and Cleaning

To build the data set, we used Python to interact with Twitter's search application programming interface (API) [23] and fetch Arabic tweets based on certain search keys. The Python language and its libraries are one of the most flexible and popular approaches to data analytics, especially for machine learning. The hashtags used in the search were the ones that mention the different Saudi telecom companies, such as STC, Mobily and Zain, as follows: #stc, #STC, #Mobily, #mobily, #Zain, #zain, #موبايلي# , الاتصالات_السعوديه and #زين_السعودية; additionally, the companies' Twitter accounts were used as keywords. The aim was to monitor the telecom customers' sentiments continuously. Data were collected from April 2017 until May 2017. This period was particularly interesting, as it included the reaction to the Saudi Communications and Information Technology Commission news on the compensation of some of the affected customers from telecom companies and the entry of a new telecom company to the Saudi market. Our golden corpus comprised 20,000 tweets to use in the training and testing of the proposed ASA methods (Table 2). Previous studies used fewer than 20,000 tweets, which was shown to be sufficient to produce state-of-the-art systems for SA for Twitter [26], [35]. To analyse the corpus, noise, non-Arabic tweets and re-tweets were eliminated, by cleaning and pre-processing the tweets in the corpus. Mubarak and Darwish [26] showed that most daily Arabic tweets (60%) are from Saudi Arabia. Therefore, we filtered the tweets based on tweet location, to identify Saudi tweets. The tweets were processed using the natural language toolkit (NLTK) library in Python for normalisation and tokenisation, and the sentences were segmented into words for easy analysis. Normalisation in Arabic involves unifying the shapes of some letters with different shapes [6].

Table 2. Statistics of the Saudi tweet corpus

Corpus	Training data	Test data
STC	6,072	1,518
Mobily	5,168	1,292
Zain	4,760	1,190
Total	16,000	4,000

The ethical issue of using social media data in terms of confidentially and privacy has stirred an ongoing controversy in research communities. The availability of social media data is thought to potentially expose a social media user to risks. Business owners using profiling for business purposes raises also such issues. In Twitter, a user phone and address are not made public, to provide some level of privacy. Additionally, in our current research, we have deleted any phone numbers or names that were included in tweets. Also, we collected just time and location of the tweet, without any other information about the tweet author.

4. Annotation

Before the SA process, the classifier needs to be trained to create a machine-readable version of the corpus through annotation. Mohammad defined annotation as providing opinions and sentiments towards the target [16]. In our application, the target was customer satisfaction. At the sentiment level of annotation, several studies have used three labels in the annotation process (positive, negative and neutral) to express the sentiment orientation [1], [5]. The output from the classification is based on the labels used in the annotation.

To ensure the high quality of the manual annotation process, the annotation process needs clear guidelines to obtain consistency between annotators. For this task, three annotators, who were Computer Science graduates (experts in annotation and Microsoft Excel), native Arabic speakers and experienced annotators as used by Al-twairesh [5] and Refae [32], were used. The annotation passed through three annotation stages. The goal was to build a gold standard Arabic corpus that is trustworthy, allowing for meaningful evaluations [32]. Thus we used a 2-way classification (Positive and Negative) that is consistent with the many studies that used binary sentiment classification with Arabic text [32], [25]. The annotators assigned one label only as the emotion expressed per tweet, as noted in many studies [7], [33]. To identify the reliability of the agreement of the annotation task, Fleiss' kappa was used [25]; its value of 0.50 points to a moderate level of agreement according to Landis and Koch [21].

5. Experiments and Results Analysis

5.1. Evaluation Metric

The metrics used to evaluate the performance of the classifiers included accuracy, precision, recall, F-measure [2]. The F-measure is the harmonic average between precision and recall. Here, we use the average of all F-measures (F-avg) for each class, which is considered a better measure than accuracy [1], [13].

5.2. Classification using SVM

Feature Selection

In feature extraction, feature sets are selected to be used with a classifier, and their utility in the text analysis is examined. Feature selection entails choosing the feature subset that achieves a superior performance in a classification [7]. Features selected here were based on the ASA literature [7], [32], [8] as follows:

1. *Term frequency* and *term presence*: These features include terms and their frequency counts
2. *Syntactic features*: This feature set includes word-stem and n-gram (sequence of the words in a text) classifiers.
3. *Language style*: This feature set involves some features that characterise the language typically used in social media, including the 'Is-Sarcastic' feature. This feature was assigned in our manual corpus annotation process by human annotators.
4. *Language style features*: This feature set involves some features that characterise the language typically used in social media, including: *Stylistic features*: This feature set checks the number of informal sentiment indicators on social media and some quantitative features, such as tweet length (characters).

Additionally, some features were selected, based on the nature of the corpus, such as:

5. *Affective-cue features*: This feature set contained six binary features, indicating whether a tweet has any of the following social signals: consent, dazzle, laughs, regret, prayer and sigh. The motivation for using this feature set was finding a set of simple

features that can correlate with the users' culture and, at the same time, can be utilised as a means of conveying sentiments. Due to the fact that there were many examples of *du'a'* (prayers) in telecommunication tweets (Figure 1), the Has-Prayer feature was used.

6. *Tweet-Topic*: This feature evaluates the role of the SA topic. The aim of using this feature was to study the correlations between the services provided by the Saudi telecommunication companies and the sentiment conveyed in a tweet, e.g., whether users tended to have negative attitudes when discussing Internet issues.

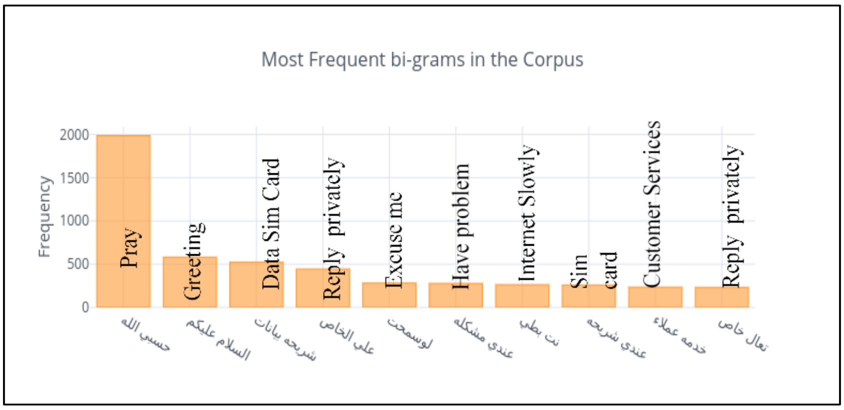

Figure 1. Most frequent bi grams in the corpus

Model Construction

Some studies have stated that the linear kernel with an SVM classifier is the best kernel in text classification [6], [32], [2]. Therefore, we used the linear kernel as one of the potentially best models. First, we created a baseline to compare the model with (Table 3). The baseline includes the basic features, term- and n-gram models. We evaluated the n-gram (unigram, bigram or trigram) and term presence models to establish which performs best. The results showed that the term presence model achieved the best F-avg. This is due to the lack of repetition of a term within a short text, such as a tweet. Pang and Lee noted that using term presence leads to a best performance in sentiment analysis for movie reviews [29].

Regarding the n-gram model, we found the combination of the unigram and bigram models to be the best for our corpus. This result was consistent with what was found in the literature regarding the superiority of combining the unigram and bigram models over the n-gram model in ASA [2], [3], [32] and in English SA [34]. The rationale behind combining the unigram and bigram models was to provide more information than the unigram model alone and to be less sparse than the n-gram model [28]. The baseline for the experiment is 0.853 and the features that included are Term presence + unigram and bigram models.

In terms of the impact of each feature set on the classification model using SVM, we experimented with subsets of the initial feature set (Figure 2). As removing some features increased the classifier's performance, these were interpreted as features harming the classifier and thus were removed from the classification model. When removing some features decreased the classifier's performance, they were kept in the classification model. Removing the *Tweet-topic* feature caused the greatest decrease in the model's performance. This means this feature is the most important. Therefore, the *Tweet-topic* and *Is-sarcastic* features were retained. Removing the *Tweet-length* and *Has-prayer* features increased the performance of the model and we removed it from the model. The result of '*Has-prayer*' feature is somewhat surprising, as it is a specific characteristic of Arabic Tweets. We think that the result is due to the classifier misunderstanding between negative and positive tweets that used prayer, because both types of tweets contain the same word "ﷲ", which means God. After applying the generating model on the test set, the F-avg became 0.908.

Thus, the model's performance increased from the baseline.

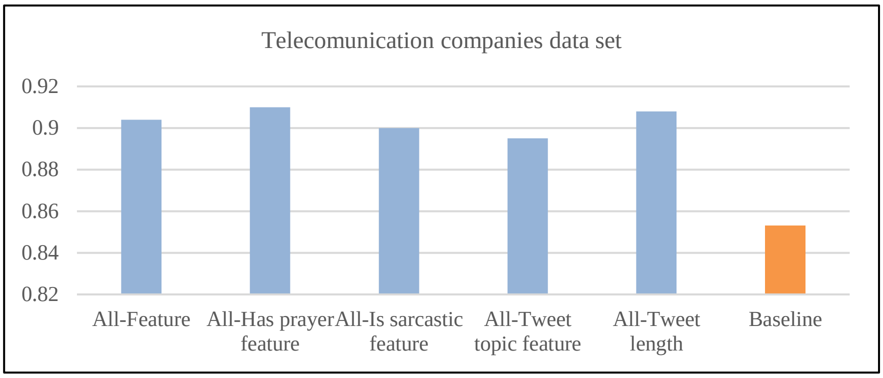

Figure 2. SVM: F-avg of all features in the corpus, and F-avg when a feature is removed

5.3. Classification Using LSTM and GRU

Due to the nature of our sequential data, there was a time/position dimension in the data, e.g. the word you see in the future is not independent of the words you have seen before. We used the most popular deep-learning-based models, LSTM and GRU with two different implementations, simple LSTM and GRU and bidirectional LSTM and GRU.

The model started with word embedding to represent each word in a tweet as a vector for creating 300-dimensional word vectors for each word in a tweet. Then we fed the LSTM layer with this embedding, using a 128-dimensional hidden state. We applied a dropout of 0.5 fraction rate over the batch of sequences, then fed it to another LSTM layer with a 128-dimensional hidden state that returns a single hidden state. Finally, we applied a dense layer with 2 units with 2 possible classes followed by Sigmoid activation. Also, we used back-propagation in a default implementation bundle with the TensorFlow library.

In the bidirectional LSTM or GRU, a future context was included in the model beside the past context. We added the attention mechanism implemented in the Keras library with a context/query vector for temporal data to handle the long sequence.

We put the attention mask on top of a recurrent neural network layer (LSTM or GRU/ Simple RNN) with `return_sequences=True`. The dimensions are inferred based on the output shape of the RNN.

Example:

```
model.add(GRU(64, return_sequences=True))
      model.add(AttentionWithContext())
```

We used a context vector to assist the attention as follows:

Input shape

3D tensor with shape: `(samples, steps, features)`.

Output shape

2D tensor with shape: `(samples, features).

After that, we fed the LSTM/GRU with attention mechanism models with two inputs (word embedding and character embedding) to represent each word/character in a tweet as a vector.

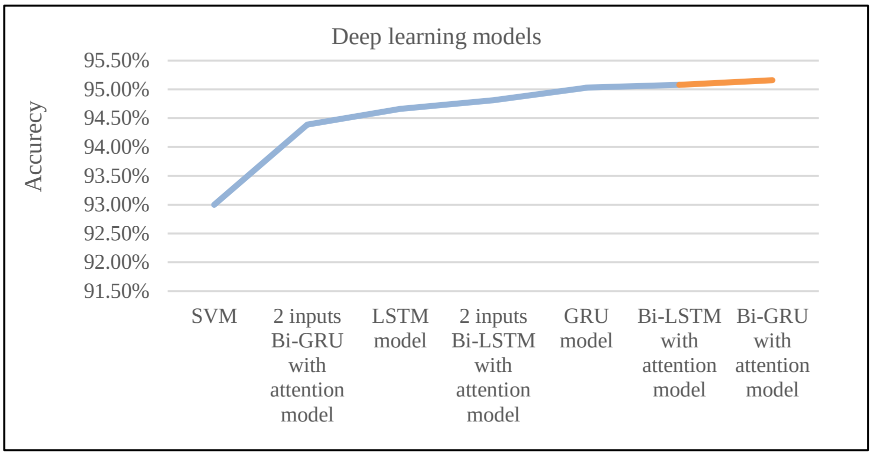

Figure 3. Accuracy of LSTM and GRU models on the corpus

It has been shown from Figure 3 that Bi-GRU with attention mechanism performed better than other models with 95.16% accuracy. The 2 inputs Bi-GRU with attention model achieved less accuracy than the others model with 94.39% accuracy. For that, we used the Bi-GRU with attention model in this research to measure customer satisfaction.

6. Discussion

Comparing the results of the deep learning models with SVM (Table 3), it shows the superiority of deep learning models other than SVM for the reasons mentioned in Section 2 and because of the applicability of deep learning approaches to adapt to the continuously dynamic nature of Twitter (***RQ1***). In addition, we noticed the superiority of GRU with 95.16% accuracy compared to the LSTM model. This was because GRU performs better with a small data set. Adding the attention mechanism to the bi-directional model enhanced the performance of the models, while adding two inputs decreased the performance of the models.

Table 3. Models comparison

Model	Accurac y	F1	Recall	Precision
LSTM model	94.66%	0.966	0.966	0.966
Bi-LSTM with attention model	95.08%	0.951	0.951	0.951
GRU model	95.03%	0.950	0.950	0.950
Bi-GRU with attention model	95.16%	0.952	0.952	0.952
2 inputs Bi-LSTM with attention model	94.81%	0.948	0.948	0.948
2 inputs Bi-GRU with attention model	94.39%	0.944	0.944	0.944
SVM	93.0%	0.930	0.930	0.930

7. Observation

From our observations, there were many examples of *du'a* (prayer mentions) (Figure 1, 1986 tweets). The majority of tweets used *du'a* to confirm a negative sentiment. In addition, we observed more negative tweets than positive tweets in the corpora (Figure 4) (RQ2). The rationale, we believe, was the difficult economic circumstances for all Arab countries in recent years, as discussed by Abdul-Mageed and Diab [2]. In addition, there were many mentions both in the positive tweets and the negative to Communication and

Information Technology Commission (CITC) as shown in Figure 5. Therefore, Twitter analysis should be considered by the CITC as a means of monitoring the Saudi telecom companies.

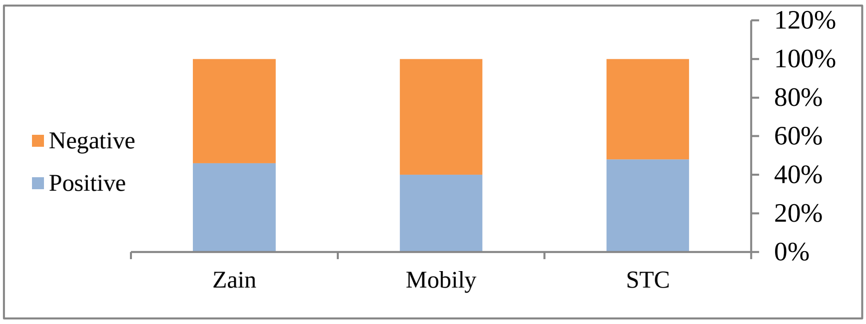

Figure 4. Percentage of positive and negative tweets on the three corpora

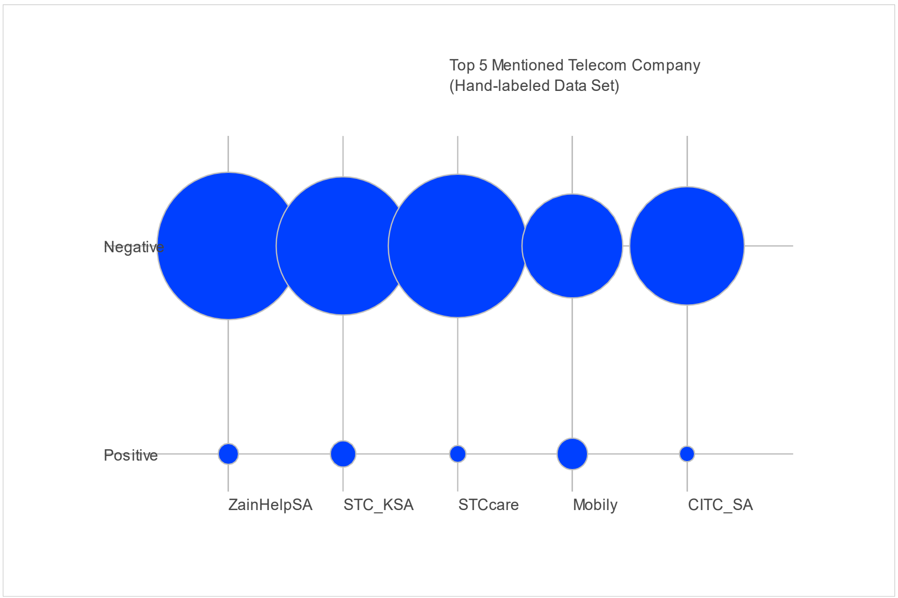

Figure 5. Top 5 mentioned Saudi Telecom companies

8. Predicting Customer Satisfaction

We used our model which achieved the higher result (Bi-GRU with attention mechanism) to predict the customer satisfaction for the corpus based on the predefined companies STC, Mobily and Zain.

First, we calculated the customer satisfaction as follows:

cust_sat = total_ratings /(2*num_customers) (5)

where:

num_customers = len(ratings)
total_ratings = sum(ratings) (the summation of all ratings)
rating: binary rating.

Then we divided the corpus based on the company. We calculated the average accuracy of predicted customer satisfaction using the model with one input (word embedding) based on the equation:

```
pred_ratings = model.predict([all_wrds]).argmax(-1)
```
(6)

where:

```
all_wrds = tokenizer.texts_to_sequences(df.Tweet.values)
all_wrds = pad_sequences(all_wrds,maxlen = max_len)
```

Comparing the predicted customer satisfaction (using the model) with actual customer satisfaction (using the mathematical calculations) (Table 4) showed that the accuracy was very close. Our model achieved the goal of predicting customer satisfaction of telecom companies based on Twitter analysis. These results will give insight to the decision makers in these companies about the percentage of customer satisfaction and help to improve the services provided by these companies. Notably, the customer satisfaction percentage for the three companies, 31.06%, 34.25% and 32.06%, were below 50%. Perhaps that was because a customer tended to post a negative tweet rather than a positive tweet in Twitter, as previously observed. These results should encourage the makers to consider Twitter analysis to measure customer satisfaction and to include it as a new way to evaluate their marketing strategies.

Table 4. Average accuracy of predicted customer's satisfaction vs. actual customer's satisfaction

Company	Predicted Customer's Satisfaction	Actual Customer's Satisfaction
STC	0.3106	0.3134
Mobily	0.3425	0.3401
Zain	0.3206	0.3208

9. Comparison and Implications

After applying our model to the Arabic data set provided in the SemEval 2017 Task 4, subtask A: tweet classification according to a three-point scale of Twitter [33], our mod-el achieved 0.797 in terms of accuracy. Comparing the result of our model to the result of the team NileTMRG [14], which placed the first team among the other top 10 teams in Subtask A, they achieved 0.581. Our model achieved a clearly higher accuracy. This is promising progress in terms of ASA on Tweets.

Social media considers an ease of use of a platform. The number of internet users that use social media has increased in 2019 to 2.77 billion users [37]. Therefore, there is a high probability of people using social media platform for exposing their feelings. This may be one of the reasons why our model is important and has obtained such a high accuracy (95.16%), We believe that our model is suitable for the high volume of respondents and represents a cost-effective tool to monitor a customer satisfaction on social media. Additionally, because of its dependence on text mining, there is a possibility to generalise this model to different social media platforms.

10. Conclusion and limitations of the research

This research conducted Arabic Sentiment Analysis to determine customer satisfaction with Saudi Arabian Telecom companies based on a golden corpus of Arabic tweets. This study resulted in the construction of the first golden corpus of Saudi tweets related to telecom companies, consisting of 20,000 tweets and evaluating customer satisfaction for telecommunications (telecom) companies in Saudi Arabia by testing the SVM, LSTM and GRU models. The results favoured binary-GRU with attention mechanism measured by the measure statistic of the models tested (RQ1). In addition, the results showed that our model is highly accurate, when comparing with other models that monitor a customer satisfaction on social media. Moreover, there is the possibility to generalise our model for

different social media platforms. Interestingly, contrary to our initial expectations, although there are a great number of prayers in the Arabic tweets corpus, the has-prayer feature had to be removed, due possibly to both positive and negative tweets using prayers, often the word “God” (RQ2).
Since the analysis is done on Twitter, the sample contains just comments from customers who use Twitter and may be missing comments of customers who use other ways of communication with the company. The most challenging struggle in this research was the disclosure of the real statistics of customer satisfaction ration from the Telecom companies. Instead, we used the reports of evaluating the performance of telecommunications companies as published by Saudi Communication and Information Commission to evaluate our results. Having additional insight information from companies may increase the overall precision of these results. Future research is needed to test GRU models with different implementations and to test more features to potentially further raise the accuracy.

References

1. Abbasi, A. and H. Chen, *Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums.* ACM Transactions on Information Systems, 2008.
2. Abdul-Mageed, M. and M.T. Diab, *Sana: A large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis.*, in *LREC*2014. p. 1162-1169.
3. Ahmed, S., et al., *Key issues in conducting sentiment analysis on Arabic social media text.* 2013 9th International Conference on Innovations in Information Technology (IIT), 2013. **47**: p. 72-77.
4. Al-Smadi, M., et al., *Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews.* 2018: p. 1-13.
5. Al-Twairesh, N., *Sentiment Analysis of Twitter: A Study on the Saudi Community* in *computer science*: Saudi Arabia.
6. Al-Twairesh, N., H. Al-Khalifa, and A. Al-Salman, *Subjectivity and sentiment analysis of Arabic: trends and challenges*, in *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*2014, IEEE. p. 148-155.
7. Al-twairesh, N.S., *Sentiment Analysis of Twitter : A Study on the Saudi Community.* 2016.
8. Aldahawi, H.A., *Mining and Analysing Social Network in the Oil Business : Twitter Sentiment Analysis and Prediction Approaches.* 2015.
9. Allola, A., *Measuring the quality of services provided by Jawwal Co. from the viewpoint of customers in the provinces of Gaza Strip*, 2011, University of Alazhar
Gaza, Palestine.
10. Alnawas, A. and N. Arıcı, *The Corpus Based Approach to Sentiment Analysis in Modern Standard Arabic and Arabic Dialects: A Literature Review.* Journal of Polytechnic, 2018. **0900**: p. 461-470.
11. Alwehaibi, A. and K. Roy. *Comparison of Pre-Trained Word Vectors for Arabic Text Classification Using Deep Learning Approach*. in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2018. IEEE.
12. Anastasia, S. and I. Budi, *Twitter sentiment analysis of online transportation service providers*, in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*2016, IEEE. p. 359-365.
13. Athanassopoulos, A.D. and A. Iliakopoulos, *Modeling customer satisfaction in telecommunications: assessing the effects of multiple transaction points on the perceived overall performance of the provider.* Production and Operations Management, 2003. **12**: p. 224-245.
14. El-Beltagy, S.R., M.E. Kalamawy, and A.B. Soliman, *NileTMRG at SemEval-2017 Task 4: Arabic Sentiment Analysis.* 2017: p. 1-6.
15. ElSahar, H. and S.R. El-Beltagy, *Building large arabic multi-domain resources for sentiment analysis*, in *International Conference on Intelligent Text Processing and Computational Linguistics*2015, Springer. p. 23-34.
16. Erik Cambria, D.D., Sivaji Bandyopadhyay, Antonio Feraco, *A Practical Guide to Sentiment Analysis.* 2017. **5**: p. 76-77.

17. Gustafsson, A., M.D. Johnson, and I. Roos, *The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention.* Journal of marketing, 2005. **69**: p. 210-218.
18. Huang, M., Y. Cao, and C. Dong, *Modeling Rich Contexts for Sentiment Classification with LSTM.* 2016.
19. Ibrahim, H.S., S.M. Abdou, and M. Gheith, *MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis*, in *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*2015, IEEE. p. 353-358.
20. Kennedy, H.J.J.o.B. and E. Media, *Perspectives on sentiment analysis.* 2012. **56**(4): p. 435-450.
21. Landis, J.R. and G.G. Koch, *Landis_Jr__Koch_Gg_1977_Kappa_and_Observer_Agreement.* Biometrics, 1977. **33**: p. 159-174.
22. Marcus, A., et al., *Processing and visualizing the data in tweets.* ACM SIGMOD Record, 2012. **40**: p. 21-27.
23. McKinney, W., *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* 2012.
24. Mountassir, A., H. Benbrahim, and I. Berrada, *A cross-study of Sentiment Classification on Arabic corpora*, in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*2012, Springer. p. 259-272.
25. Mourad, A. and K. Darwish, *Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs.* ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011. **2**: p. 587-591.
26. Mubarak, H. and K. Darwish, *Using Twitter to collect a multi-dialectal corpus of Arabic*, in *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*2014. p. 1-7.
27. Najadat, H., A. Al-Abdi, and Y. Sayaheen, *Model-based sentiment analysis of customer satisfaction for the Jordanian telecommunication companies.* 2018 9th International Conference on Information and Communication Systems, ICICS 2018, 2018. **2018-Janua**: p. 233-237.
28. Pak, A. and P. Paroubek, *Twitter as a corpus for sentiment analysis and opinion mining.*, in *LREc*2010. p. 1320-1326.
29. Pang, B. and L. Lee, *LR..references..Opinion Mining and Sentiment Analysis.* Foundations and Trends® in Information Retrieval, 2008. **2**: p. 1-135.
30. Patterson, J. and A. Gibson, *Deep learning: A practitioner's approach.* 2017.
31. Qamar, A.M., S.A. Alsuhibany, and S.S. Ahmed, *Sentiment classification of twitter data belonging to saudi arabian telecommunication companies.* International Journal of Advanced Computer Science and Applications (IJACS), 2017. **1**: p. 395-401.
32. Refaee, E., *Sentiment analysis for micro-blogging platforms in arabic.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017. **10283 LNCS**: p. 275-294.
33. Rosenthal, S., N. Farra, and P. Nakov, *SemEval-2017 Task 4: Sentiment Analysis in Twitter.* Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017), 2017: p. 502-518.
34. Ruder, S., P. Ghaffari, and J.G. Breslin, *INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis.* 2016: p. 330-336.
35. Severyn, A. and A. Moschitti, *Unitn: Training deep convolutional neural network for twitter sentiment classification*, in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*2015. p. 464-469.
36. Sohangir, S., et al., *Big Data: Deep Learning for financial sentiment analysis.* Journal of Big Data, 2018. **5**(1).
37. Statista. 2019 [cited 2019 June 20]; Available from: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users.
38. Tundjungsari, V.J.I.J.o.C.S. and E.E. Volume, *Business Intelligence with Social Media and Data Mining to Support Customer Satisfaction in Telecommunication Industry.* 2013. **1**.
39. Young, T., et al., *Recent trends in deep learning based natural language processing.* ieee Computational intelligenCe magazine, 2018. **13**: p. 55-75.
40. Zhao, J., K. Liu, and L. Xu, *Sentiment analysis: mining opinions, sentiments, and emotions*, 2016, MIT Press.