Classification of Encouragement (Targhib) And Warning (Tarhib) Using Sentiment Analysis on Classical Arabic

Hatem AlHasani[#], Saidah Saad[#], Junaidah Kassim[#]

[#]Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia.
E-mail: hatemalhassani@gmail.com, saidah@ukm.edu.my, junaidah@ukm.edu.my

Abstract— The Holy Qur'an is the main religious text of Islam. The Qur'an has its own methods of Targhib (encouragement) and Tarhib (warning), which are important features of the Qur'an. Most of the Quranic verses would urge and encourage people to do right and good deeds, and also warn them from committing evil and bad deeds. The method of classifying a text into two opposing opinions has been applied previously in solving the problem of sentiment analysis. Currently, it is applied in identifying between Targhib (encouragement) and Tarhib (warning) verses in the Qur'an. Each verse of the Qur'an can be treated as either an encouragement, warning or neutral. The language of the Holy Qur'an is one of the most challenging natural languages in sentiment analysis. The aim of this work is to classify the verses of encouragement and warning using sentiment analysis and NLP techniques. Several approaches are used in the Sentiment Analysis classification, such as the machine learning approach, the lexicon-based approach and the hybrid approach. In carrying out this aim, the applied machine learning approach was used, where the impact of the use of different techniques such as POS tagging, N-Gram and Feature selection with correlation based were evaluated and investigated. 95.6% accuracy was achieved using Naïve Bayes (NB) and 91.5% accuracy was achieved using the Support Vector Machines (SVM). This study is a significant study in extracting information and knowledge from the Holy Qur'an. It is significant for both researchers in the field of Islamic studies as well as non-specialized researchers.

Keywords— sentiment analysis; NLP; ML; classical arabic; Qur'an.

I. INTRODUCTION

Sentiment Analysis (SA), which is the computational treatment of opinions, sentiments and subjectivity of a text, is a very popular field of research in text mining. Furthermore, several definitions of sentiment analysis have emerged in the literature over the last few years. [1] for example, has adopted a general definition of sentiment analysis where it is seen as the extraction, identification, or otherwise characterization of the sentiment content of a text unit through the use of natural language processing (NLP), statistics, or machine learning methods.

According to [2], sentiment analysis can be performed at the document level, the sentence level, or the aspect level. The document level classifies the overall document which contains the sentiment words expressed by the author [3]. At the sentence level, each sentence is classified as either a positive or a negative sentiment using the sentiment analysis which involves more tasks compared to the classification at the document level [4]. At the aspect level, finer-grained analysis is performed where it directly looks at the opinion itself rather than looking at the language constructs (documents, paragraphs, sentences, clauses or phrases) [5].

Several approaches are used in the Sentiment Analysis classification, such as the machine learning approach, the lexicon-based approach and the hybrid approach. The Machine Learning (ML) approach is typically a supervised approach whereby a set of data is labelled with its class such as positive and negative class. The input dataset is classified by the algorithms with the help of a trained classifier. The examples of classification algorithms mostly used in SA are the SVM, the NB Classifier and the Maximum Entropy (ME) [6]. This study used the most recommended supervised classifiers, which are SVM [7]–[10] and NB [8], [9] classifiers. The Support Vector Machine (SVM) is a well-known technique in machine learning. It is a supervised learning method which involves other learning techniques and is used in analysing data for classification and analysis purposes. As a fast, accurate, simple and easy to implement classifier, NB was thus chosen as one of the classifiers in this case. The Lexicon-Based (LB) approach is an unsupervised approach in which a sentiment lexicon is created with each word having its score as a number representing its class [11]. The hybrid approach utilizes both the lexicon-based and machine learning approaches [12].

The datasets used in sentiment analysis are a significant issue in this field. The data are mainly sourced from the

social media such as Twitter and Facebook, stock markets, news articles, as well as political debates [13], [14]. Subjectivity is the way that emotions and opinions can be expressed in the language, while objectivity refers to the factual phrases used in the language. Polarity classification is the basic task in sentiment analysis. Polarity classification occurs when a piece of text stating an opinion on a single issue is classified as one of two opposing sentiments. "Thumbs up" versus "thumbs down," or "like" versus "dislike" reviews are examples of polarity classification. Apart from polarity classification, sentiment analysis also has a number of other tasks, for example to distinguish between subjective and objective texts [15]. Moreover, a piece of text might have a polarity without necessarily containing an opinion. A news article for instance, can be classified as either a good or bad news without being subjective [16].

Sentiment Analysis has been addressed in several languages such as English, Arabic, and Indo-European [17]. The Arabic language itself can be classified into three forms of language, which are the Classical Arabic (CA), Modern Standard Arabic (MSA) and the Colloquial Arabic Dialect (DA). CA is the language used in Muslim religious resources, such as the Quran and the Hadith. CA, although it is the foundation of MSA, has some differences when compared to MSA, for example in terms of the lexical meaning of the words, some grammatical structures, and style [18].

Most studies have been carried out in the English language or the Indo-European languages [19]. However, a number of studies have been made on the Arabic language, where the early Arabic studies focused on sentiment analysis in newswire [20], [21], and on the social media, especially Twitter [8], [22]–[24]. The Arabic language is a rich morphological language, specifically the Classical Arabic since it is the language of the Holy Qur'an [18].

There are a lot of synonyms for encouragement and warning in the English language such as "instigation/intimidation", "invitation/intimidation" and "encouragement/intimidation". Encouragement in the Arabic language is called *Targhib* "الترغيب" , while warning is called *Tarhib* "الترهيب".

The Holy Qur'an is the main religious text of Islam. The Qur'an has its own methods of Targhib (encouragement) and Tarhib (warning), which are important features of the Qur'an. Most of the Quranic verses would urge and encourage people to do right and good deeds, also warn them from committing evil and bad deeds. The method of classifying a text into two opposing opinions has been applied previously in solving the problem of sentiment analysis. Currently, it is applied in identifying between Targhib (encouragement) and Tarhib (warning) verses in the Qur'an.

Based on the verses of encouragement and warning contained in the Qur'an, it is found that the verses could be categorized into three types. Firstly, singular verses which talk about either encouragement or warning. Secondly, singular verses which talk about encouragement and warning at the same time in the same verse. Thirdly, the words of encouragement and warning are presented and talked about in a sequence of verses. This study deals with the first type

of verse where it is based on singular verses that are treated as either encouragement or warning. Table I shows the example of both classes.

The extraction of information and knowledge from the Holy Qur'an is highly beneficial for people who specialize in Islamic studies, as well as for the non-specialized people since the Holy Qur'an is extremely important to Muslims all around the world. To the best of our knowledge. This study has been carried out in analysing the Arabic text of the Holy Qur'an using the sentiment analysis technique. The aim of this study is to classify each verse of the Qur'anic Arabic as either an encouragement or warning using the sentiment analysis technique. In addition, this study evaluated the most of suitable sentiment analysis pre-processing techniques with correlation-based feature selection.

TABLE I
EXAMPLE OF ENCOURAGEMENT AND WARNING VERSES

Class	Verse	English Translation
Encouragement	نَبِّئْ عِبَادِي أَنِّي أَنَا الْغَفُورُ الرَّحِيمُ	[O Muhammad], inform My servants that it is I who am the Forgiving, the Merciful
Warning	وَأَنَّ عَذَابِي هُوَ الْعَذَابُ الْأَلِيمُ	And that it is My punishment which is the painful punishment.

II. MATERIAL AND METHOD

Abbasi et al [7] performed opinion mining for both English and Arabic web forums. Opinion classification was made using both the syntactic and stylistic features. The syntactic feature included the words, n-grams and POS tags. The stylistic feature on the other hand, included the length of the review, the existence of special characters and the repetition of some of the special words. The Support Vector Machine (SVM) classifier was then used with the entropy-weighted genetic algorithm (EWGA) as a feature selection technique on the English movie review and on the English and Arabic web forums.

In the work of [8], a supervised sentiment classification system was applied for 1000 Arabic balanced tweets. Feature vectors were used to compare and choose the classifier with the highest accuracy. Based on the comparison of the results between SVM and NB using unigram and bigram terms in both cases, it was clear that SVM produced better results compared to NB. The success rate of their classification system was reported to be around 73%.

Ghadeer et al [9] also analyzed a collection of Arabic tweets to determine whether the tweets had positive or negative sentiments by classifying their polarities. Different supervised machine learning techniques such as SVM, NB and Decision Tree (DT) were applied. The results of the experiment showed the impact of the pre-processing techniques in attaining better results. DT attained the highest result for accuracy without the filtration of stop words and with the use of the unigram. NB showed the highest result for accuracy with the filtration of the stop words, and with the use of stemming and also bigram and trigram. The

results of SVM showed the highest accuracy without the filtration of the stop words and with the use of the unigram.

The Arabic Jordanian Twitter corpus was introduced by Alomari et al [10] where 1,800 tweets were gathered and annotated as having either positive or negative sentiment. Different approaches of the supervised machine learning sentiment analysis were investigated with regard to their application on the general subjects of the Arabic users' social media that were found in either the MSA or the Jordanian dialect. Experiments were carried out in evaluating the use of different weight schemes, stemming and N-grams terms techniques and scenarios. The results of the experiments provided the best scenario for each classifier and indicated that the SVM classifier with the use of TF-IDF weighting scheme with stemming through the bi-grams feature performed better than the NB classifier best scenario performance results.

Sabri & Saad [25] evaluated three Feature Selection Methods (Information Gain (IG), Chi-square (CHI) and Gini Index (GI)). They proved that the use of the feature selection method was able to increase the performance of sentiment classification of the Arabic language. Furthermore, the results of the experiment showed that the use of the CHI feature selection resulted in the best performance for feature selection. The performance of the meta-classifier combination approach outperformed the other approaches for sentiment classification of the Arabic language. A 90.80% accuracy was attained with the use of the combination approach of the meta-classifier with the chi-square feature selection method.

The methodology is divided into seven steps as shown in Figure 1. The Quranic corpus is the dataset collected from the Holy Quran. Pre-processing is the step which was used to prepare the dataset, while feature extraction was used to generate new subset features. The correlation weights were used to correlate weights of the related features while feature selection was used to identify the important features. The classifier model was used based on the supervised machine learning techniques and finally the system was evaluated based on different techniques. All these steps are described in detail in the following sections.
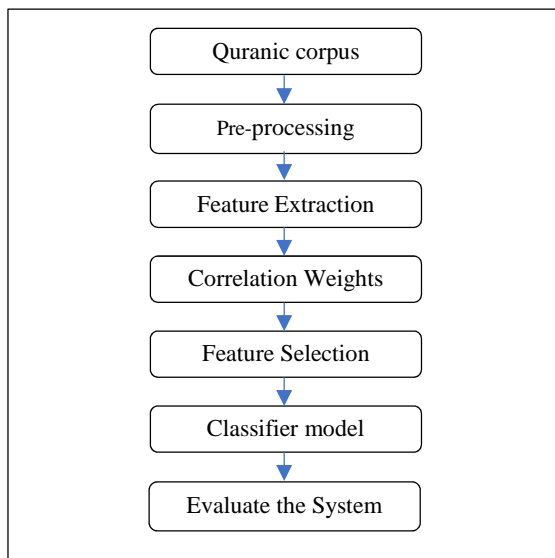

Fig. 1 Methodology Diagram

A. Qur'anic Corpus

The data used in this study were collected from the Arabic text of the Holy Qur'an. In achieving the goal of classifying the corpus as either encouragement or warning, the corpus was manually annotated by the experts of the Islamic domain. In this study, the focus was placed on a balanced corpus which consisted of 2,000 verses where 1,000 were encouragement verses while the other 1,000 were warning verses.

B. Pre-processing

The pre-processing step is important in achieving a high level of accuracy. In this study, normalization, tokenization, lemmatization, light stemming, filter stop words and Part of Speech Tagging (POS) were used. In order to reduce errors and increase accuracy, some letters were normalized to their normal ways and diacritic characters (short vowels) were removed from the text. The set of normalization rules that is mostly used includes [26]: to strip away the Diacritics (for example: الْعَرَبِيَّة to العربية), to normalize the Hamza (for example: ئ to ء and ؤ to و), to normalize the Alef (for example: آ or إ or أ to ا), to normalize the Yeh (for example: ي to ى), and to normalize the Heh (for example: ة to ه).

The Arabic language is a morphologically complex language. This complexity thus requires for the development of appropriate systems that are able to deal with tokenization.

Lemma is a lexical entry recorded in the dictionaries which represents only the static lexicon at a fixed point in time [27]. Lemma is the specific form that represents the lexeme, for example the lemmatization approach maps the word (المدرسة) to (مدرسة). The light stemming technique is used in reducing the different forms of the word to one form (root or stem) [27].

Stop words are very common words, for example prepositions, articles, and pronouns, which appear in the text and carry little meaning. Thus, stop words are unlikely to help text mining.

Part Of Speech (POS) tagging assigns a tag to each word in a text and classifies the word to a specific morphological category such as the noun, verb, adjective and so forth. POS tagging is often used in sentiment analysis and [2] indicated that adjectives are very important in determining the sense of the text. Adjectives can be used both as the main features as well as filters for selecting other features. In this study however, the MADAMIRA tool was used to produce the tokenization, lemmatization, light stemming and parts-of-speech tag for each word in determining whether the word is a noun, verb, adjective and so forth [28]. All processes mentioned are summarized in Table II.

TABLE II
PRE-PROCESSING

process	Verse
Text	مَّن ذَا الَّذِي يُقْرِضُ اللَّهَ قَرْضًا حَسَنًا فَيُضَاعِفَهُ لَهُ أَضْعَافًا كَثِيرَةً ۚ وَاللَّهُ يَقْبِضُ وَيَبْسُطُ وَإِلَيْهِ تُرْجَعُونَ
Normalization	من ذا الذي يقرض الله قرضا حسنا فيضاعفه له ضعافا كثيرة والله يقبض ويبسط واليه ترجعون
Tokenization	من| ذا| الذي |يقرض| الله |قرضا |حسنا |فيضاعفه | له | ضعافا | كثيرة |والله| يقبض | ويبسط| واليه | ترجعون

Lemmatization	من | ذو | الذي | أقرض | الله | قرض | حسن | ضاعف | ل | ضعف | كثير | الله | قبض | بسط | إلى | رجع
Light Stemming	من | ذا | الذي | قرض | الله | قرض | حسن | ضاعف | له | أضعاف | كثير | الله | قبض | بسط | إلي | رجع
POS	من|prep ذا|noun الذي|pron_rel قرض|verb الله|noun_prop قرض|noun حسن|noun ضاعف|verb له|prep أضعاف|noun كثير|noun الله|noun_prop قبض|verb بسط|verb إلي|prep رجع|verb

C. Feature Extraction

Feature extraction is the process where properties are extracted from the data. These properties, called features, are characteristics of the data. The features should be discriminative, to describe the original data as well as possible. On the other hand, the features should reduce the space to prevent redundancies and a high dimensionality of the data. The following features are therefore discussed: n-grams and the TF-IDF measure.

1) N-Gram Model

N-gram is used to define a subsequence of n tokens from a given sequence and is used in various fields of natural language processing and text analysis. An n-gram model defines the method for finding a set of n-gram words from a given document. The commonly-used n-gram models include the unigrams (n=1), bigrams (n=2) and trigrams (n=3) [29].

2) Term Frequency–Inverse Document Frequency (TF-IDF)

TF-IDF which stands for Term Frequency–Inverse Document Frequency is a popular statistical technique used in indexing terms. It is based on documents and term vectors which represent the term frequency as well as term presence. TF-IDF of term is calculated using term frequency. The larger the value of a term frequency, hence the more prominent it is in a given document.

D. Feature Selection

There are two main purposes of feature selection. Firstly, it aims to increase computational competence by finding a subset of features that achieve as well as the much larger set. Secondly, it aims to filter out the noise and the less relevant features to avoid overfitting. According to [30], feature selection could be mainly categorized into the filter method and the wrapper method. The filter method would generally evaluate the features by assigning them a ranking score based on the distributional statistics in the data. The wrapper method on the other hand, would identify the optimal subset of the features using held-out data. However, since the number of subsets is exponential, the wrapper method is tremendously inefficient when a large feature set is involved, even with greedy algorithms [30]. Besides that, [31] pointed out that the filter method is generally faster compared to the wrapper method. It is for these reasons that we have placed more focus on the filter method. The filter method can be based on information theory, statistical tests, or more common principles.

1) Correlation Weights

A correlation is the relationship between features or data attributes. Features may be correlated with one another or could be redundant. A subset of relevant features is selected by applying the correlation weight which has the highest value. The correlation of each attribute is computed with respect to the label attribute. In this study, the weight by correlation operation's RapidMiner tool was used to calculate the correlation weights.

2) Filter Methods

The filter methods are relatively fast and are independent of the classifiers. Hence, the Select by Weight operator in the RapidMiner tool was used in ranking the feature set in order to select the top K from the dataset. In this study, the top K of between 1000 to 6000 features were evaluated.

E. Classifier Model

The classifier model step is the core step where two popular supervised classifiers were utilized in this study, namely the Naïve Bayes and the support vector machine. All classifiers are well known for their text mining research, as well as successful sentiment analysis applications. The important step is the optimization process which is required for the use of the cross-validation techniques in comparing the accuracy between the training and the testing datasets for all classifiers. In this study, the 10-fold cross validation was applied.

III. RESULT AND DISCUSSION

A common practice in classification analysis system is to evaluate and measure the performance of the system in order to obtain the highest accuracy. This section discusses the outcome of the experimental results. The experiment was implemented as a comparison with four experimental results.

A. The Results of POS Tagging

The aim of evaluating POS tagging is to measure the important features of the research dataset. The experimental results of the Classical Arabic sentiment analysis, which are dependent on the adjectives, verbs and noun tags were described. In this study, the evaluation of the three tags was made individually while the evaluation of the combination of these tags without using the N-gram model and feature selection was performed using lemmatization. Figure 2 shows the accuracy of both the SVM and NB classifiers, where the verbs and nouns play an important role in the encouragement and warning datasets. The experimental results show that the combination of three tags gives better accuracy on both classifiers. 56.85%, 70.90%, 74.90% and 81.45% accuracy were attained for the adjectives, verbs, nouns and the combination of three tags respectively for the SVM classifier. 55.95%, 68.35%, 71.85% and 77.05% accuracy were attained for the adjectives, verbs, nouns and the combination of three tags respectively for the NB classifier.

B. The Results of N-gram

The aim of evaluating the use of N-gram is to measure the impact of accuracy of the research dataset. The unigram (single word) which is the baseline accuracy of lemmas, and bigram (two words) beside unigram, and trigram (three words) beside unigram and bigram without using feature selection were evaluated. Figure 3 shows the accuracy of

both the SVM and NB classifiers, where the investigation was carried out to determine whether the N-gram model enhances the accuracy of both classifiers. The experiment shows that the combination of the unigram and bigram gives better accuracy on the SVM classifier, while the combination of unigram, bigram and trigram gives better accuracy on the NB classifier. 81.65%, 83.95% and 82.00% accuracy were attained for the unigram, the combination of the unigram and bigram and the combination of unigram, bigram and trigram respectively for the SVM classifier. 77.05%, 81.25% and 81.35% accuracy were attained for the unigram, the combination of the unigram and bigram and the combination of unigram, bigram and trigram respectively for the NB classifier.
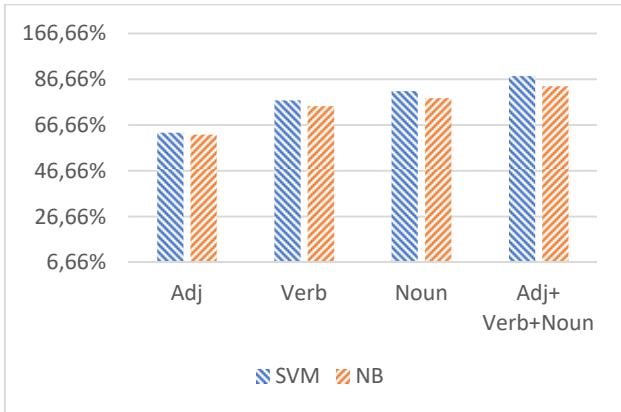

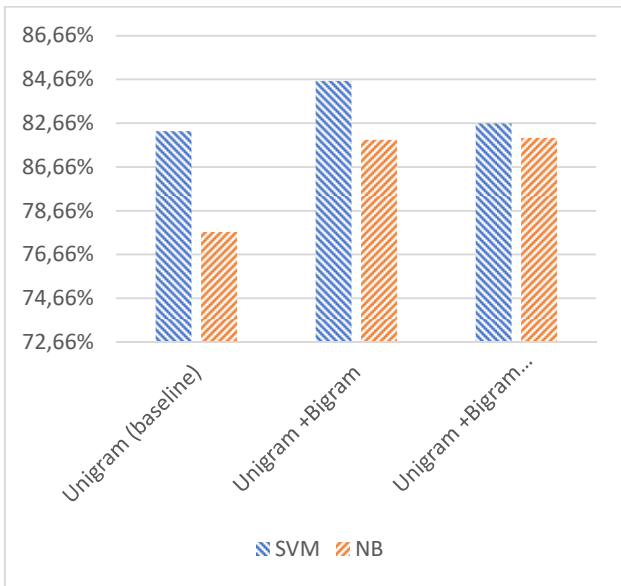Fig. 2 Experimental Results Using POS tagging

Fig. 3 Experimental Results Using N-Gram

C. The Results of the Combination of POS Tagging, N-gram and Feature Selection

This section describes the evaluation of the combination of POS tagging, N-Gram and feature selection for both SVM and NB classifiers. Experiments were conducted to determine the effects of using POS tagging and Feature Selection, and the results are shown in Figure 4. The experiments show the accuracy of both classifiers, which are 81.65%, 87.40% and 89.10% accuracy for feature selection and lemma feature with POS tagging (lemma 2826, 1000

and 2000 respectively) for the SVM classifier. On the other hand, 77.80%, 89.30% and 89.65% accuracy were attained for feature selection and lemma feature with POS tagging (lemma 2826, 1000 and 2000 respectively) for the NB classifier.

Besides that, we experimented on the effects of using the N-Gram and Feature Selection as shown in Figure 5. The experiments show the accuracy of both classifiers which are 81.65%, 84.45%, 88.90%, 90.45%, 91.40%, 91.50% and 91.75% accuracy for feature selection and lemma feature with N-gram (lemma 2826, 1000, 2000, 3000, 4000, 5000 and 6000 respectively) for the SMV classifier. On the other hand, 77.80%, 88.90%, 93.05%, 94.20%, 95.15%, 95.60% and 95.45% accuracy were attained for feature selection and lemma feature with N-gram (lemma 2826, 1000, 2000, 3000, 4000, 5000 and 6000 respectively) for the NB classifier.

In addition, we experimented on using POS tagging, N-Gram model and Feature Selection as shown in Figure 6. The experiments show the accuracy of both classifiers which are 81.65%, 86.20%, 88.85%, 91.55%, 90.85%, 90.90% and 90.75% accuracy, which were achieved for Feature Selection and lemma feature with N-gram and POS tagging (lemma 2826, 1000, 2000, 3000, 4000, 5000 and 6000 respectively) for the SVM classifier. On the other hand, 77.80%, 89.05%, 91.45%, 93.95%, 93.30%, 93.25% and 93.15% accuracy were achieved for feature selection and lemma feature with N-gram and POS tagging (lemma 2826, 1000, 2000, 3000, 4000, 5000 and 6000 respectively) for the NB classifier.
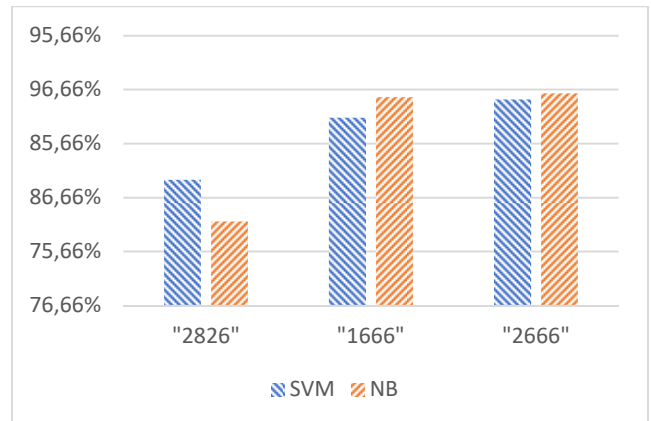

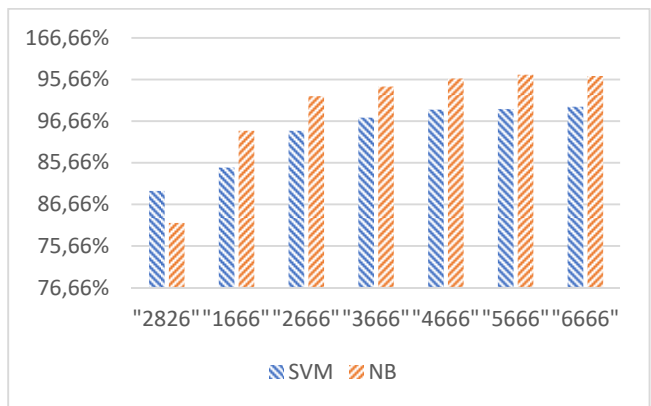Fig. 4 Experimental Results Using POS Tagging and Feature Selection

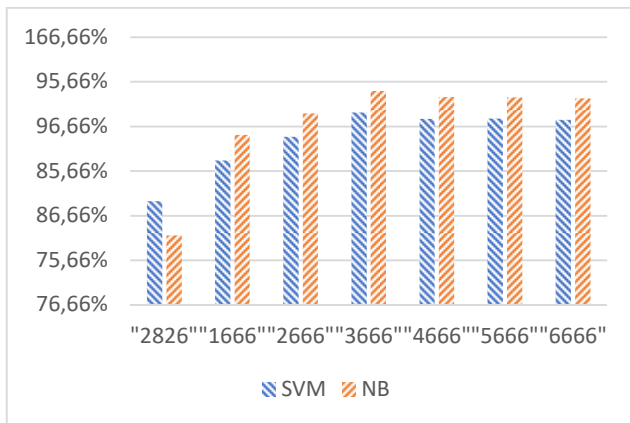Fig. 5 Experimental Results Using N-Gram and Feature Selection

Fig. 6 Experimental Results Using N-Gram, POS Tagging and Feature Selection

IV. CONCLUSION

Feature selection is used in sentiment analysis to increase computational efficiency by finding a subset of features that perform as well as the much larger set, and to filter out noise and less relevant features to avoid overfitting. A subset of relevant features is selected by applying the correlation weight which has the highest value. Our model is evaluated using the filter method in finding the top k which is between 1000 features to 6000 features. The feature selection was evaluated with lemma, POS tagging, N-gram, the combination of lemma and POS tagging and the combination of POS tagging and N-gram model. Based on the observation of the results for this model, it is noticed that feature selection is a good way of enhancing the accuracy of sentiment analysis. Besides that, POS tagging played a good role in both models. In addition, the N-gram model with feature selection achieved the highest accuracy. Finally, the results indicate that the accuracy of the SVM and NB classifiers achieved similar accuracies for most of the experiments. However, the best accuracy was obtained by the NB classifier with feature selection and N-Gram, whereby the SMV classifier achieved 91.75% accuracy for feature selection and lemma feature with N-gram when the top 6000 features were selected, while the NB classifier achieved 95.45% accuracy. For the purpose of the experiment, a balanced corpus was used. The corpus was annotated from the Holy Qur'an. It consisted of 2000 verses, where 1000 were encouragement verses while the other 1000 were warning verses. The lemma was used to evaluate the proposed experiments which are POS tagging and n-gram with feature selection filter methods. Besides that, the combination of these experiments was evaluated, whereby two classifiers, which are the SVM and NB, were used. The experiments show that feature selection is a good way of enhancing the accuracy of sentiment analysis. Furthermore, the results indicate that the accuracy of SVM and NB classifiers has achieved similar accuracy as most of the experiments. However, the best accuracy was achieved by the NB classifier with feature selection and N-Gram feature. In this research study, the encouragement and warning verses were classified from the text of the Holy Qur'an. This initial attempt will enable for future works to be extended to the fields of the *Hadith* and *Sunnah*. Apart from that, we wish to apply multi classification in solving the third state,

which is the "encouragement and warning" state. We need to apply our method using different techniques in studying the improvements of the performance. Moreover, we plan to extend the presented Classical Arabic sentiment analysis corpus using the lexicon-based approach.

REFERENCES

[1] A. Z. Syed, "Applying sentiment and emotion analysis on brand tweets for digital marketing," in *Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on*, 2015, pp. 1–6.

[2] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," *ACM Comput. Surv.*, vol. 50, no. 2, p. 25, 2017.

[3] S. V Wawre and S. N. Deshmukh, "Sentiment classification using machine learning techniques," *Int. J. Sci. Res*, vol. 5, no. 4, pp. 1–3, 2016.

[4] X. Ding and B. Liu, "Resolving object and attribute coreference in opinion mining," in *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 268–276.

[5] N. Boudad, R. Faizi, R. O. H. Thami, and R. Chiheb, "Sentiment analysis in Arabic: A review of the literature," *Ain Shams Eng. J.*, 2017.

[6] M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine Learning Techniques for Sentiment Analysis: A Review," *Int. J. Multidiscip. Sci. Eng*, vol. 8, no. 3, pp. 27–32, 2017.

[7] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums," *ACM Trans. Inf. Syst.*, vol. 26, no. 3, p. 12, 2008.

[8] E. Refaee and V. Rieser, "Subjectivity and sentiment analysis of Arabic twitter feeds with limited resources," in *Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme*, 2014, p. 16.

[9] A.-S. Ghadeer, I. Aljarah, and H. Alsawalqah, "Enhancing the Arabic Sentiment Analysis Using Different Preprocessing Operators," *New Trends Inf. Technol.*, p. 113, 2017.

[10] K. M. Alomari, H. M. ElSherif, and K. Shaalan, "Arabic Tweets Sentimental Analysis Using Machine Learning," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2017, pp. 602–610.

[11] S. A. Morsy, "Recognizing contextual valence shifters in document-level sentiment classification," 2011.

[12] H. D. Kim, K. Ganesan, P. Sondhi, and C. Zhai, "Comprehensive review of opinion summarization," 2011.

[13] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared, "Approaches to cross-domain sentiment analysis: a systematic literature review," *IEEE Access*, vol. 5, pp. 16173–16192, 2017.

[14] B. Saberi and S. Saad, "Sentiment analysis or opinion mining: A review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1660–1666, 2017.

[15] M. M. Altawaier and S. Tiun, "Comparison of machine learning approaches on arabic twitter sentiment analysis," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1067–1073, 2016.

[16] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends® Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.

[17] N. Omar, M. Albared, T. Al-Moslmi, and A. Al-Shabi, "A comparative study of feature selection and machine learning algorithms for arabic sentiment classification," in *Asia information retrieval symposium*, 2014, pp. 429–443.

[18] R. E. Salah and L. Qadri binti Zakaria, "A Comparative Review of Machine Learning for Arabic Named Entity Recognition," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 2, pp. 511–518.

[19] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in *Future Internet of Things and Cloud (FiCloud), 2014 International Conference on*, 2014, pp. 579–583.

[20] M. Abdul-Mageed and M. T. Diab, "Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire," in *Proceedings of the 5th Linguistic Annotation Workshop*, 2011, pp. 110–118.

[21] M. Elarnaoty, S. AbdelRahman, and A. Fahmy, "A machine learning approach for opinion holder extraction in Arabic language," *arXiv Prepr. arXiv1206.1011*, 2012.

[22] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for Arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, 2014.

[23] A. Mourad and K. Darwish, "Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs.," 2013.

[24] M. Salameh, S. Mohammad, and S. Kiritchenko, "Sentiment after Translation: A Case-Study on Arabic Social Media Posts.," in *HLT-NAACL*, 2015, pp. 767–777.

[25] B. Sabri and S. Saad, "Arabic Sentiment Analysis with Optimal Combination of Features Selection and Machine Learning Approaches," *Res. J. Appl. Sci. Eng. Technol.*, vol. 13, no. 5, pp. 386–393, Sep. 2016.

[26] R. M. Sallam, H. M. Mousa, and M. Hussein, "Improving Arabic Text Categorization using Normalization and Stemming Techniques," 2016.

[27] I. Zeroual and A. Lakhouaja, "Arabic information retrieval: Stemming or lemmatization?," in *2017 Intelligent Systems and Computer Vision (ISCV)*, 2017, pp. 1–6.

[28] A. Pasha *et al.*, "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic.," in *LREC*, 2014, vol. 14, pp. 1094–1101.

[29] G. Tripathi and S. Naganna, "Feature selection and classification approach for sentiment analysis," *Mach. Learn. Appl. An Int. J.*, vol. 2, no. 2, pp. 1–16, 2015.

[30] S. Kübler, C. Liu, and Z. A. Sayyed, "To use or not to use: Feature selection for sentiment analysis of highly imbalanced data," *Nat. Lang. Eng.*, pp. 1–35, 2017.

[31] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinformatics*, vol. 2015, 2015.