Original Research Paper

Social Media Sentiment Analysis: The Hajj Tweets Case Study

[1]Mohammad Ashraf Ottom and [2]Khalid M.O. Nahar

[1]*Department of Information Systems,*
Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan
[2]*Department of Computer Sciences,*
Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan

Abstract: About forty five percent of the world's population use social networks, thinking of using these platforms seemed to find people's opinions and feelings on various topics. Companies that offer their services and products to customers focus on the subject for future improvement. Thus, serious thinking began to analyze the views of people across different social platforms and also to develop the best ways to analyze these views. In this study, we focused on finding the best way for sentiment analysis by using a series of Hajj-related tweets, which is one of the most important rituals performed by Muslims, where the companies responsible for the pilgrimage season seek to complete the season in best way every year. We used the Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) as supervised algorithms for machine-learning approach and Text Blob analyzer for lexicon-based approach. Finding shows that, machine learning techniques worked better than the lexicon approach in the classification and analysis of Hajj related tweets. Even the limited availability of Hajj tweets corpus dataset, SVM reaches the best accuracy which was 84%.

Keywords: Sentiment Analysis, Text Mining, Twitter Analysis, Feature Extraction, Tweets Classification

Introduction

Hajj is an Islamic terminology meaning pilgrimage. Hajj is an annual Islamic pilgrimage to Mecca in the Arabian Peninsula and it is mandatory worship and duty for all Muslims who are financially and physically capable, once in a lifetime. Every Islamic calendar year in the month of Du-Al-Hijjah, millions of Muslims visit Mecca to perform Hajj. Nowadays and with advancement in communication and technology, people can use social media platforms such as Twitter to provide opinion and experience during this journey. Data mining techniques especially opinion mining tools can be used to obtain knowledge and pattern about the overall people experiences and opinions about the services and facilities in Mecca, which can be used later by official authorities to enhance future experience and services in the following years (Clingingsmith *et al.*, 2009; Peters, 1996).

Technological development and emergence of many social networking platforms have given people the opportunity to express themselves and their interests and to share their thoughts and feelings on different topics. Many service organizations and businesses are interested in following customer feedback and customer satisfaction or dissatisfaction with the products and services offered to them in order to make better decisions about improvement. Social networking platforms may be one of the best choices for such companies and organizations.

According to the latest statistics published by the social media management platform "Hootsuite", about 3.48 billion people are now using social media, in another way: 45% of the total world population are using social networks. It should be noted that these percentages have been measured through various social media such as Twitter, where it has 321 million monthly active users. Twitter offers a small blogging service that allows users to send "Tweets" to others, with a maximum of 280 characters per message which makes Twitter an ideal choice for the research reported in this study (Khanna, 2018).

Sentiment Analysis (SA) is the study of natural language to systematically extract knowledge and useful patterns for knowledge-based decision. Sentiment

Analysis, also called opinion mining, is the process of determining whether a text has been written to express different view. One of common applications for this technology is to learn how individuals believe about a certain topic (Hussein, 2018). In this study, we want to focus on knowing the most accurate analysis method from the main sentiment analysis methods depending on our data collected from the last pilgrimage season.

The manuscript is structured into the following sections: Literature review, the suggested methodology and after all, conclusions of this study and future work were discussed.

Literature Review

The literature showed many approaches of sentiment analysis. Overall, there are two main methods, firstly is a lexicon-based approach which is presented in this study and the other one is Machine-learning-based Approach Which is often used different classification techniques that can be used for text classification in supervised learning, such as Naïve Bayes, K-Nearest Neighbor classifier (KNN) and Support Vector Machines (SVM) (Alessia *et al.*, 2015)

Naïve Bayes

Naïve Bayes Classifier is a probabilistic classifier based on applying Bayes theorem written in the Eqs. (1a) and (1b):

$$P(C \mid A) = \frac{P(C \mid A)P(C)}{P(A)} \tag{1a}$$

$$P(A_1, A_2, ..., A_n \mid C) = P(A_1 \mid C_J)P(A_2 \mid C_J)...P(A_n \mid C_J) \tag{1b}$$

where each attribute set $A = \{A_1, A_2, ..., A_n\}$ consists of n attributes, given the class label C.

The assumption made here is that the attributes are independent. That is the presence of one particular attribute does not affect the other. Naïve Bayes works well on text categorization (Ahmad, 2013).

Support Vector Machine (SVM)

Another algorithm that can be used for text classification is a support vector machine. The main idea in this classifier is to find the hyperplane which separates the data as optimally as possible (Demidova *et al.*, 2016; Ottom and Alshorman, 2019) as shown in Fig. 1.

K-Nearest Neighbor (KNN)

KNN does not build a model from the training data. To classify any test example A, define K- Neighborhood as K nearest Neighbors of A. Computing the distance between A and every example in the data set. Choosing the K examples in the dataset that are nearest to A. Assign A the class that is the most frequent class in K-nearest classes (Ashraf *et al.*, 2013; Navlani, 2018) as shown in Fig. 2.

The author in (Bagheri and Islam, 2017) showed the application of sentimental analysis and how to connect to Twitter and run sentimental analysis queries using python and some packages like Tweepy and TextBlob packages. Then, he compared the sentimental analysis results from different queries including movie, politics, fashion and fake news.
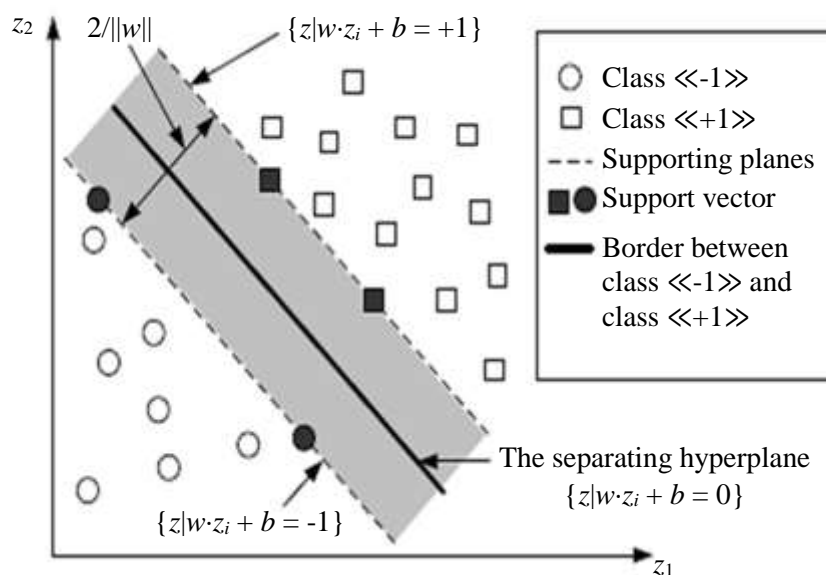

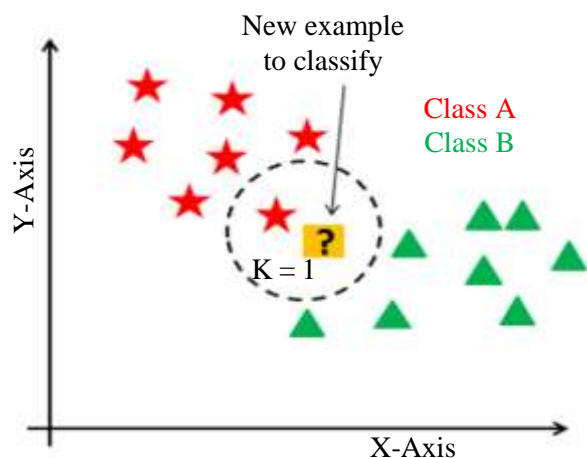**Fig. 1:** SVM Linear separation of 2 classes (Demidova *et al.*, 2016)

Fig. 2: K-Nearest Neighbor (KNN) (Navlani, 2018)

Authors in the paper (Madhoushi *et al.*, 2015) aimed to classify SA methods away from the level or task and to examine the major research problems in recent articles shown in the field. They found that ML methods are the most frequent used approaches. On the other side, they discussed the ongoing issues such as some methods are still not suitable in some domains, due to small number of labeled data records in some applications. Another issue with SA is the need for more researches of SA in Arabic and other languages rather than English, to deal with complicated sentences and words that requires advance sentiment and parsing.

Hasan *et al.* (2018) created a hybrid approach that involves three sentiment analyzers named SentiWordNet, TextBlob and Word Sequence Disambiguation (WSD) to analysis political opinions using machine learning methods. Alsaeedi and Khan (2019) explored the various sentiment analysis applied to Twitter data and their outcomes using different approaches such as bigram, n-gram and ensemble method. The experiments obtained 80% accuracy using n-gram and bigram approaches, whereas ensemble method obtained about 85% of classification accuracy.

Authors in the paper (Gautam and Yadav, 2014) extracted dataset from Twitter using Twitter API and made preprocessing using some ML libraries such as Scikit-learn in Python. The SVM gave better accuracy, but its long execution was the main drawback. The Naive Bayes performed satisfactorily but does not exceed expectations, on the other hand, Logistic Regression performed a bit similar to SVM and needs similar execution time as Naive Bayes.

Sarlan *et al.* (2014) showed the importance of sentiment analysis and its use in analyzing customer opinions and increasing competition in the business market. Their paper focused on the design of sentiment analysis, extracting a vast number of tweets, to provide competitive advantages to the business and increase business revenue using ML and sentiment analysis methods.

Gupta *et al.* (2017) built a hybrid approach consisting of two machine learning algorithms K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). The experiments showed that the proposed hybrid approach achieved better results in regard to accuracy and F-measure. The accuracy result of hybrid approach was about 76% Where the accuracy of the measurement 71 compared to the results of KNN and SVM, which amounted to 67% each of them separately.

For our contribution to this aspect, we will make a comparison of the results of the main methods of analysis using a collection of Hajj related tweets and determine the accuracy of each method in sentiment analysis.

Methodology

Our framework explains the work from the data collection phase to the results analysis phase. We have compiled the Hajj-related tweets using the Twitter API. After keeping data that we need about tweets, we filtered out the tweets, kept the most appropriate tweets and removed the re-tweets, then polarity and subjectivity were calculated using TextBlob Python library.

We applied several ML methods such as Naïve Bayes and SVM classifier on training set using Orange data mining tool to build the classification model. The model was evaluated using the testing dataset to obtain the accuracy result for the classification model, then comparing the result for each analysis in terms of accuracy. The following subsections detail the techniques that help us in sentiment analysis. Furthermore, a clear view of the sentiment-analysis framework is illustrated below in Fig. 3.

Dataset Collection

In this phase, we generated our dataset by collecting tweets using API. API stands for Application Programming Interface. It is a tool that makes interaction with computer programs and web services easily. Many web services provide APIs for developers to interact with their services and to access data programmatically, the following steps demonstrate the process of data collection:

Step 1: Getting Twitter API keys: In order to access Twitter Streaming API, we need to obtain the following information from Twitter: API key, API secret, Access token and Access token secret, therefore, we followed twitter guidelines to obtain the mentioned keys and tokens.

Step 2: Connect to Twitter API and download data after getting the keys and tokens information from Twitter. Twitter API enables us to access data

for the last seven days or the last 30 days from the requested date or get full archive data. Because the goal of the project is to analyze the Hajj tweets, we need to get the Hajj-related tweets on a certain date (the date of the Hajj season), so we need an access to the full archive data API.

Step 3: We identified the date of one of the Hajj seasons (from 1-8-2018 to 25-8-2018) then we made 50 requests, each request gives 100 tweets. The responses were in the form of JSON files containing all information about tweets, such as creation date, tweet_ id, tweet_ text, user_ id, re-tweet _count, etc.

Step 4: To make it easier to work with multiple JSON files, we have merged these files into one JSON file using Python code, choosing the most important tweets attributes and converted them to a single CSV file.
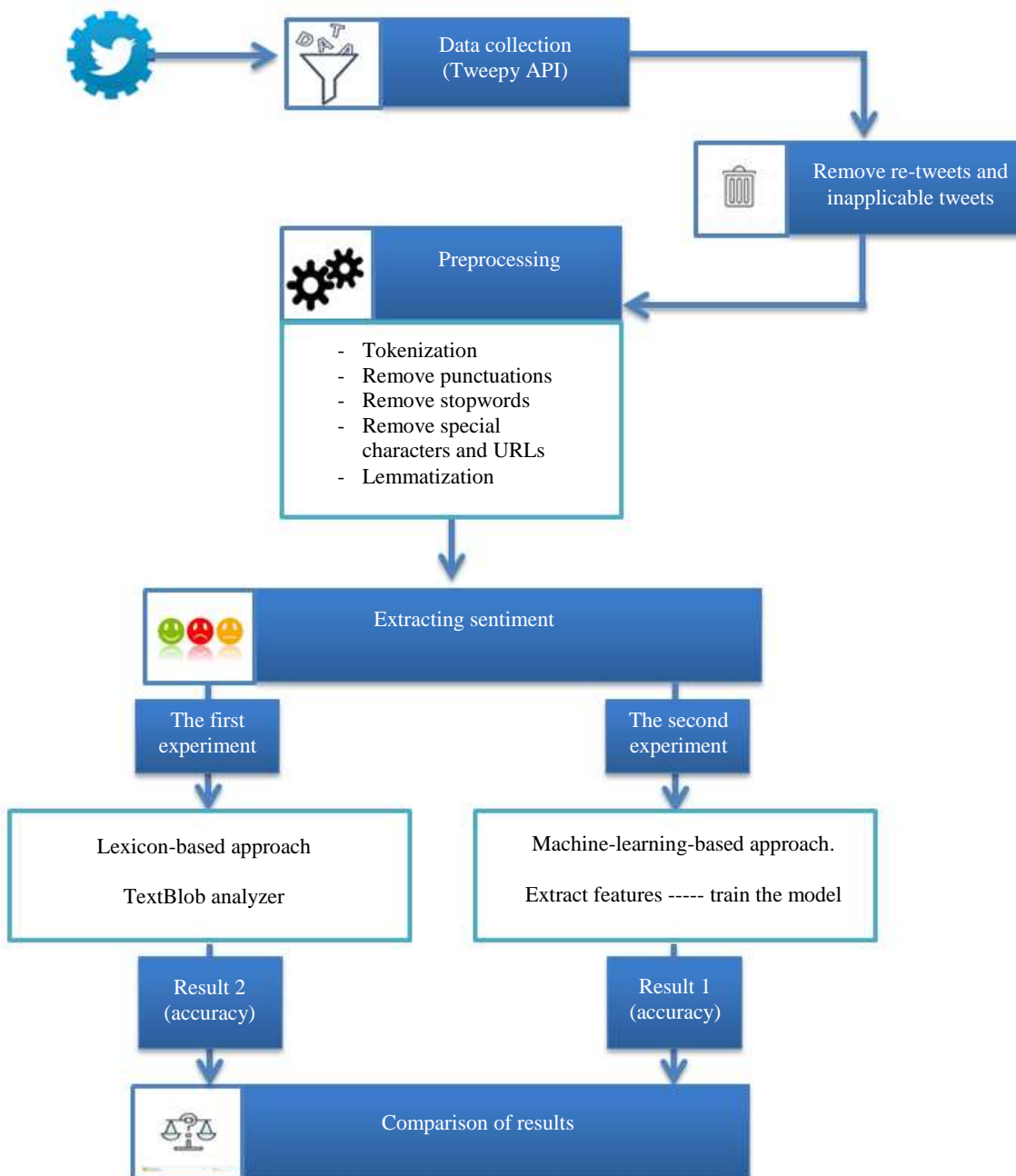

Fig. 3: Overall methodology workflow

268

Dataset Preprocessing

To analyze tweets well and to achieve the project goal, it was important to clean up tweets and remove any things that were not useful and does not add much meaning to the sentence like stop words and remove any character which may hinder the analysis of these tweets such as punctuation and special symbol. For data preprocessing, we did the following steps using Python programming language version 3.6:

- Read the CSV file that contains the data. In the reading process, we only care about the column that contains the tweet text and store the tweet text in the list. We need to import 'string' and 'CSV' modules
- Import 'nltk' module and use it to download English stop words, where these words already collected in the corpus
- Now we can pass on each tweet text within the list using a while loop and apply the cleaning commands on tweets as follow:

  - Import 're' module and use it to remove each "mention" and "URL" in the tweet
  - Split the text into smaller words and remove all the punctuation from the tweet text and remove each re-tweet "RT" from the list
  - The last step in the cleaning process is to remove stop words from the text, so we use a for loop to pass each word in each tweet and check whether the word is a stop word or not. if the word is not a stop word, we insert this word into a new list that contains what we need from the tweet

- Lemmatization: Replaces words with their base form. For example, the words "caring" to "care"
- Now, we can make a join process for all the meaningful words again and get a new pure tweet

In the next step of our work, we need to use all the pre-processed tweets in the sentiment analysis process, so we have compiled all the outputs in 'txt' files using the 'stdout' command. The following Table 1 shows some statistics about data after preprocessing.

Table 1: Statistics about the data used for our experiments

Total number of tweets	3175
Number of re-tweets	2661
Number of tweets after removing "RT" and inapplicable tweets	293
Number of Positive tweets	93
Number of Negative tweets	37
Number of Neutral tweets	163

Sentiment Extraction

There are two main approaches for extracting sentiment which is the lexicon-based approach and machine-learning-based approach, their working mechanism is summarized as follows:

- Lexicon-based methods use a predefined list of words where each word is associated with a specific sentiment, then each sentiment is combined by summing or averaging sentiment, or the majority of sentiment. The lexicon methods have a basic pattern which are:

  I. Preprocess each tweet
  II. Initialize a total polarity score (P) equal 0
  III. Check if the token is present in a dictionary, then:

  - If the token is positive, p will be positive (+)
  - If the token is negative, p will be negative (-)
  - If the token is neutral, p will be neutral (0)

  IV. look at total polarity:

  - If p>0, the text is positive
  - If p<0, the text is negative
  - If p = 0, the text is neutral

- Machine-learning-based Approach: in this approach we built a labeled dataset, extract features from them and then trained the algorithms on these features so that we can predict the correct label as much as possible

Features Extraction

Feature extraction process was carried out using:

- Bag Of Words (BOW): This technique creates a vocabulary of all the unique words occurring in all the documents (tweets), where Counting the occurrences of words in each Document (tweet)
- Term Frequency-Inverse Document Frequency (TF-IDF): This technique aims to represent the number of times a given word appears in a document relative to the number of documents in the corpus that the word appears in
- N-Grams: N-gram is a contiguous sequence of n terms from a given text, where n-gram of size 1 is referred to as a unigram, n-gram of size 2 is a bigram, an n-gram of size 3 is a trigram and so on

For our project, we conducted two experiments using these methods, then, the results of each experiment were extracted separately. The experiments and their results were as follows:

269

Experiments

Experiment 1 (lexicon-based approach): In this experiment, we used the analyzer powered by TextBlob library within Python language, which analysis the polarity and subjectivity tweets text and therefore determine whether it gives a negative, positive or neutral impression. In the beginning, we pass each tweet that has already been processed to the analyzer and then the total polarity is calculated, thus giving an analysis of each tweet. Figure 4 illustrates the details of this experiment.

The result of this experiment was promising and achieved an accuracy of 0.621, it was able to predict 182 tweets correctly from 293 tweets.

Experiment 2 (Machine-learning-based Approach): In this method, we used a set of pre-processed tweets and fed them to the models to classify tweets as positive, negative or neutral based on model learning. The accuracy of these supervised algorithms and their ability to predict the correct label is then calculated as shown in Fig. 5.
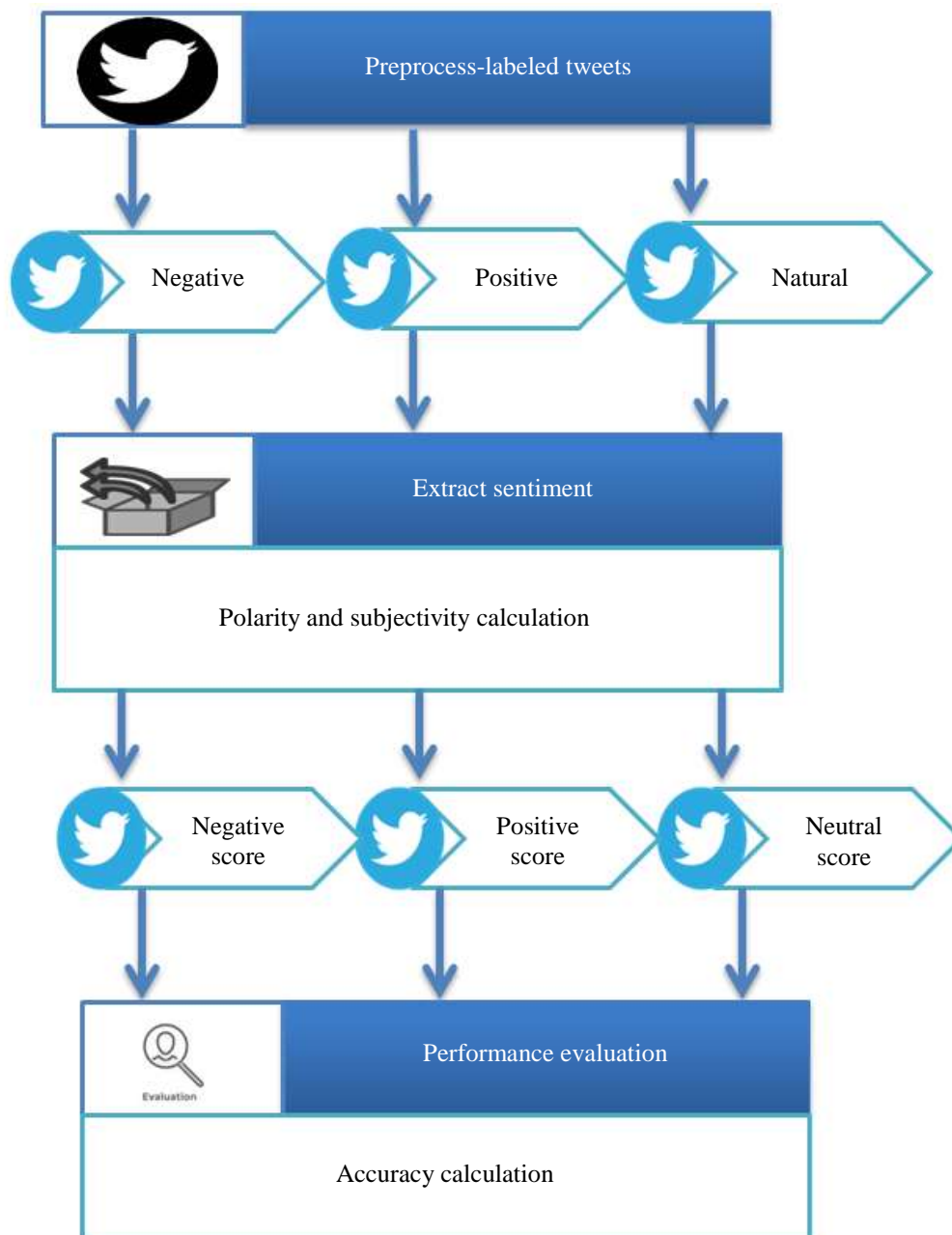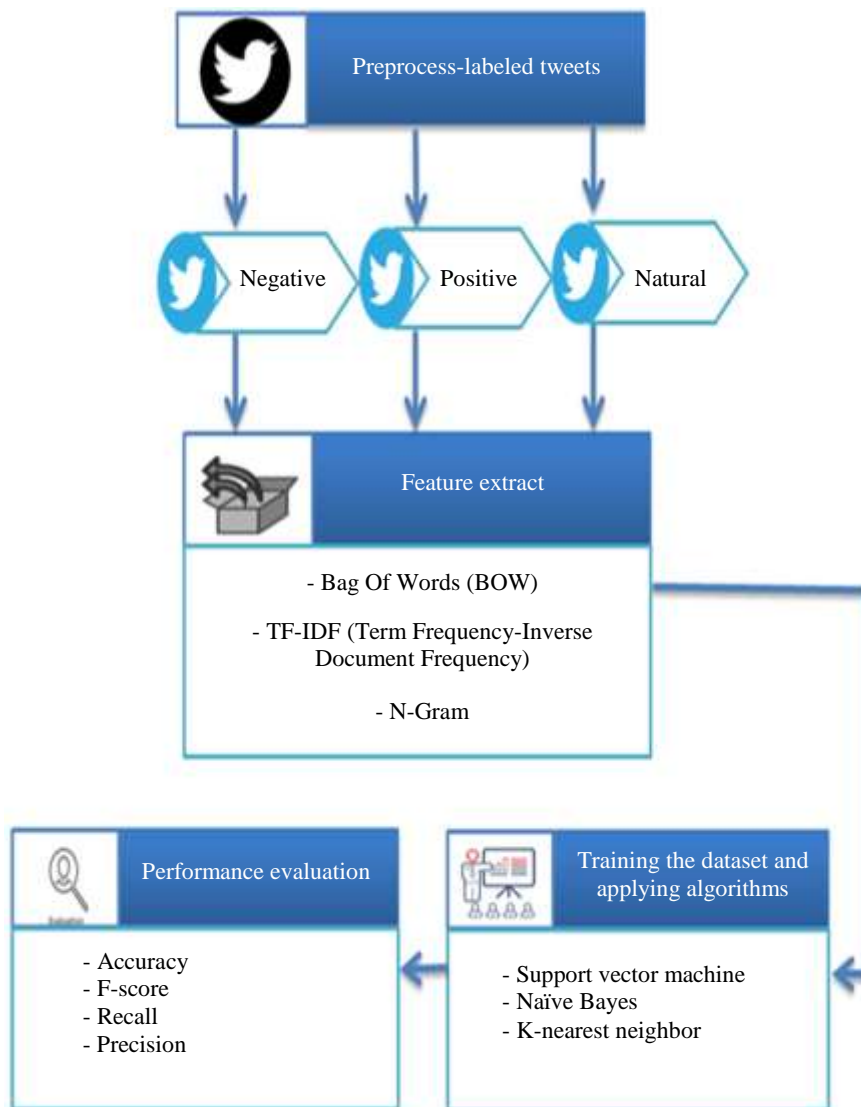

Fig. 4: Lexicon-based approach workflow

270

Fig. 5: Machine-learning-based approach workflow

Results

The dataset is divided into training and testing set using 10-Fold cross-validation. The project uses 3 algorithms for preparing and training the model (Naïve Bayes NB, Support Vector Machines SVM and k Nearest Neighbors kNN). Each machine learning algorithms is evaluated using the following well-knowns measurements in Eqs. 2-5 (Nahar, 2018; Nahar *et al.*, 2019).

Accuracy is simply the number of correct observations to the total observations, as shown in Eq. 2:

$$Accuracy = (TP + TN) / (P + N) \tag{2}$$

Precision is the number of correctly classified positive observations divided by the total number of observations that are classified as positive, as shown in Eq. 3:

$$Precision = TP / (TP + FP) \tag{3}$$

Recall is the number of correctly classified positive observations divided by the total number of actual positive observations in the test set, as shown in Eq. 4:

$$Recall = TP / (TP + FN) \tag{4}$$

F1- Score is the weighted average of Precision and Recall, as shown in Eq. 5:

$$F1 = 2TP / (2TP + FP + FN) \tag{5}$$

Where:
TP: The number of correct classifications of positive observations (true positive).

271

FN: The number of incorrect classifications of positive observations (false negative).

FP: The number of incorrect classifications of negative observations (false positive).

TN: The number of correct classifications of negative observations (true negative)

Accuracy, recall, precision and F1 score are measured for each experiment (lexicon-based approach and machine learning approach) and each classification algorithm (Naive Bayes, Support Vector Machines and k Nearest Neighbors).

Tables 2-5 show the results of applying the dataset on all classification algorithms. Also, each classification algorithm uses n-gram, a contiguous sequence of n items (unigram, bigram and trigram). Also, each classification uses Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) technique which aims to represent the number of times a given word/phrase appears in a document relative to the number of documents in the corpus that the word appears in. Figure 6 shows evaluation and comparison between algorithms used in this study.

Table 2: Accuracy results of each classifier

	Accuracy		
Grams	NB	SVM	KNN
TF-IDF Unigram	0.71	0.84	0.62
Bigram	0.66	0.75	0.61
Trigram	0.62	0.71	0.60
TF Unigram	0.70	0.77	0.60
Bigram	0.61	0.74	0.60
Trigram	0.60	0.72	0.61

Table 3: F-score results of each classifier

	F1-Measure		
Grams	NB	SVM	KNN
TF-IDF Unigram	0.71	0.83	0.63
Bigram	0.65	0.74	0.62
Trigram	0.61	0.70	0.61
TF Unigram	0.70	0.78	0.61
Bigram	0.60	0.72	0.60
Trigram	0.61	0.71	0.60

Table 4: Precision results of each classifier

	Precision		
Grams	NB	SVM	KNN
TF-IDF Unigram	0.73	0.83	0.62
Bigram	0.63	0.73	0.60
Trigram	0.62	0.71	0.62
TF Unigram	0.71	0.79	0.61
Bigram	0.62	0.70	0.61
Trigram	0.61	0.71	0.61

Table 5: Recall results of each classifier

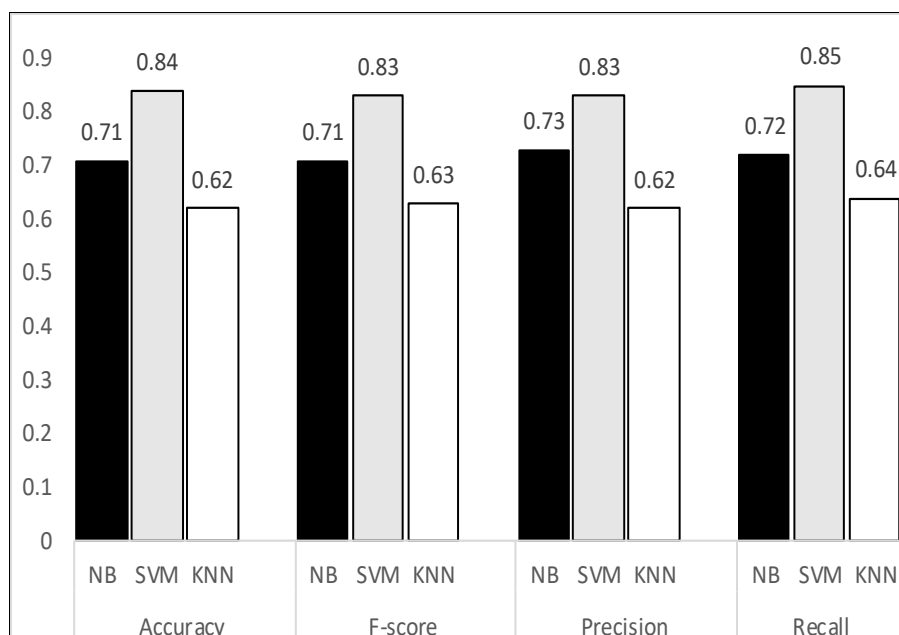|	Recall		
Grams	NB	SVM	KNN
TF-IDF Unigram	0.72	0.85	0.64
Bigram	0.62	0.74	0.62
Trigram	0.61	0.73	0.63
TF Unigram	0.72	0.80	0.62
Bigram	0.63	0.72	0.62
Trigram	0.62	0.73	0.63

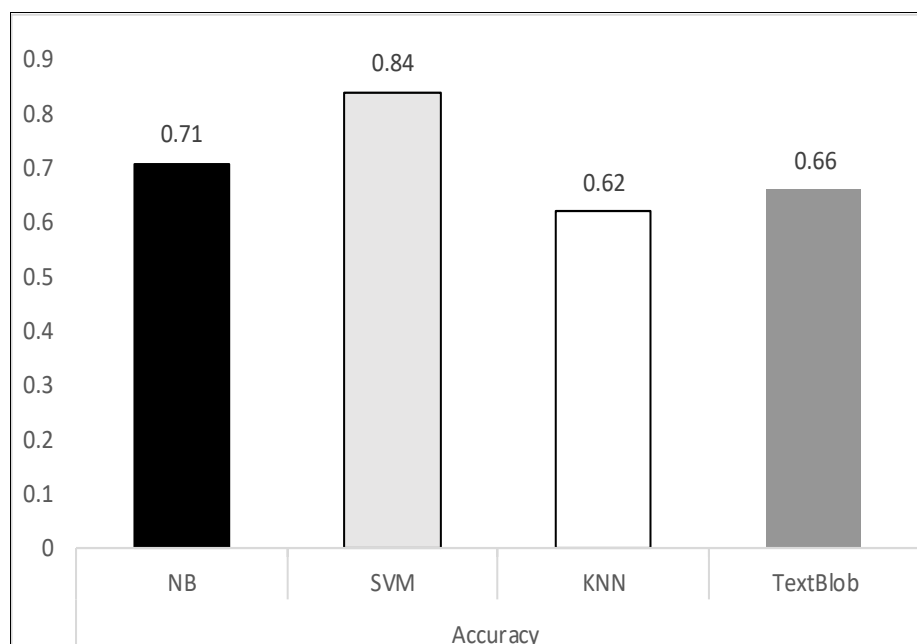**Fig. 6:** The top results for each classifier

272

Fig. 7: Accuracy obtained from the two experiments

Discussion

After reviewing the results of the classification experiment, it was found that SVM achieved the highest accuracy rate of about 0.84 followed by NB in the second degree with an accuracy of about 0.71 and then kNN by about 0.62. According to the lexicon-based approach classification accuracy obtained was 0.66 using Python package analyzer, which is close to the highest accuracy achieved by the kNN classifier, which was 0.620 and less than the accuracy of both the NB and SVM which was 0.715 and 0.84 respectively. It was therefore concluded that the second experiment (Machine-learning-based Approach) was better and with more accurate results than the first experiment (lexicon-based approach). The results were as shown in Fig. 7.

Conclusion and Future Work

Due to the rapid and wide usage of social media, sentiment analysis has become an important topic to be studied and focused on. Emphasis should also be placed on developing methods and techniques used in the analysis process. In this study, we have done two experiments to determine the accuracy of the methods of analysis and this was done using a set of Hajj-related tweets. We made the filtering process of these data and then analyzed using the TextBlob analyzer and a collection of supervised algorithms. we have been reached a set of results that indicate that machine learning was better and achieved greater accuracy than

the TextBlob Where the SVM gave the accuracy of around 0.80 while the accuracy reached by the analyzer did not exceed 0.660. For future work, it is possible to think of creating a model that integrates the two main methods of analysis and takes advantage of the set of characteristics that can be obtained through them and thus obtain higher accuracy in the process of analyzing sentiment, beside collecting more tweets when Hajj resume and after Corona pandemic ends, hopefully.

Acknowledgment

The authors would like to thank all the people who have supported this work.

Author's Contributions

All authors have equally contributed to this work.

Ethics

This article is original and contains unpublished material, the corresponding author confirms that no ethical issues involved.

References

Ahmad, M. A. (2013). Mining Health Data for Breast Cancer Diagnosis Using Machine Learning. https://books.google.com.au/books?id=5gGZoAEA CAAJ

Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. International Journal of Computer Applications, 125(3), 26-33. https://doi.org/10.5120/ijca2015905866

Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. International Journal of Advanced Computer Science and Applications, 10(2), 361-374. https://doi.org/10.14569/IJACSA.2019.0100248

Ashraf, M., Chetty, G., & Tran, D. (2013). Feature selection techniques on thyroid, hepatitis and breast cancer datasets. International Journal on Data Mining and Intelligent Information Technology Applications, 3(1), 1-8. https://doi.org/10.4156/ijmia.vol3.issue1.1

Bagheri, H., & Islam, M. J. (2017). Sentiment analysis of twitter data. ArXiv Preprint ArXiv:1711.10377. https://arxiv.org/abs/1711.10377

Clingingsmith, D., Khwaja, A. I., & Kremer, M. (2009). Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. The Quarterly Journal of Economics, 124(3), 1133-1170. https://doi.org/10.1162/qjec.2009.124.3.1133

Demidova, L., Nikulchev, E., & Sokolova, Y. (2016). The svm classifier based on the modified particle swarm optimization. ArXiv Preprint ArXiv:1603.08296. https://arxiv.org/abs/1603.08296

Gautam, G., & Yadav, D. (2014, August). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh International Conference on Contemporary Computing (IC3) (pp. 437-442). IEEE. https://doi.org/10.1109/IC3.2014.6897213

Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., Badhani, P., & Tech, B. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications, 165(9), 29-34. https://doi.org/10.5120/ijca2017914022

Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. Mathematical and Computational Applications, 23(1), 1-15. https://doi.org/10.3390/mca23010011

Hussein, D. M. E.-D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338. https://doi.org/10.1016/j.jksues.2016.04.002

Khanna, P. (2018). Social Media Security for the New Age Business (NZ Specific). http://researcharchive.wintec.ac.nz/6320/

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015, July). Sentiment analysis techniques in recent works. In 2015 Science and Information Conference (SAI) (pp. 288-291). IEEE. https://doi.org/10.1109/SAI.2015.7237157

Nahar, K. (2018). Off-line arabic hand-writing recognition using artificial neural network with genetics algorithm. The International Arab Journal of Information Technology, 15(4), 701-707. http://iajit.org/PDF/July%202018,%20No.%204/120 02.pdf

Nahar, K. M. O., Al Eroud, A. F., Barahoush, M., & Al-Akhras, A. M. (2019). SAP: Standard Arabic Profiling Toolset for Textual Analysis. International Journal of Machine Learning and Computing, 9(2), 222-229. https://doi.org/10.18178/ijmlc.2019.9.2.790

Navlani, A. (2018). KNN Classification using Scikit-learn, [Online Report]. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

Ottom, M.-A., & Alshorman, W. (2019). Heart diseases prediction using accumulated rank features selection technique. Journal of Engineering and Applied Sciences, 14, 2249-2257. https://doi.org/10.36478/jeasci.2019.2249.2257

Peters, F. E. (1996). The Hajj: The Muslim pilgrimage to Mecca and the holy places. Princeton University Press.

Sarlan, A., Nadam, C., & Basri, S. (2014, November). Twitter sentiment analysis. In Proceedings of the 6th International conference on Information Technology and Multimedia (pp. 212-216). IEEE. https://doi.org/10.1109/ICIMU.2014.7066632