Twitter Sentiment Analysis on Activities of Saudi General Entertainment Authority

Sara Alkhaldi, Sultana Alzuabi, Ryoof Alqahtani, Amjad Alshammari,
Fatimah Alyousif, Dabiah A. Alboaneen, Modhe Almelihi
Computer Science Department, College of Science and Humanities Imam Abdulrahman Bin Faisal University
P.O.Box 31961, Jubail, Kingdom of Saudi Arabia
Email: {dabuainain@iau.edu.sa}

***Abstract*—Sentiment analysis can be defined as a natural language process to determine the individual's sentiment or opinion towards something. It helps institutions, companies and governments to gain a deeper understanding and supports decision-making. This paper aims to analyse individuals' opinions in Twitter on the activities of the Saudi General Entertainment Authority (GEA) using machine and deep learning techniques. To achieve this aim, 3,817 tweets were collected using RapidMiner. To classify tweets into supporters and opposers, three machine learning algorithms were used namely, Multi-Layer Percptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and one deep learning algorithm, which is Recurrent Neural Network (RNN). Two test options were applied to evaluate the classification model, percentage split and K-fold validation tests. The results show that the people are happy and agree with the GEAs' activities. As for the gender, the support rate of females was higher than males. In addition, RF algorithm outperforms other algorithms in terms of the classification accuracy and the error rate.**

***Index Terms*—Entertainment, sentiment analysis, Twitter, machine learning, deep learning.**

I. Introduction

In an era of big data, users around the world express their daily opinions online on social media platforms such as Facebook, YouTube, Twitter and other platforms. Today, big companies and governments are investing in analysing these opinions/views for evaluating its products, services, or trends, whether political or economic, as well as social issues, this process is known as sentiment analysis.

Nowadays, social networking media possess a huge amount of data, and we can use these platforms to collect and analyse data to produce the required knowledge. Twitter is one of the leading social media platforms that allows the user to share posts that can be up to 280 characters long which are called tweets.

Entertainment is currently one of the most important aspects of the Kingdom's vision 2030. The General Entertainment Authority (GEA) was founded in 2016 to lead and improve the entertainment sector in the Saudi Arabia. This study, as the best of our knowledge, is the first study that focuses on analysing the opinions of Twitter users on the activities of GEA using machine and deep learning techniques.

The main contributions of this paper are:

1) To create Arabic dataset on activities of Saudi GEA.
2) To analyse the opinions on activities of Saudi GEA.
3) To compare between machine and deep learning algorithms in terms of classification accuracy, error rate, precision and recall.
4) To compare the results when using two test options, percentage split and k-fold cross-validation.
5) To study the effect of increasing the number of hidden layers in the Multi-Layer Percptron (MLP) algorithm.

The remainder of this paper is organised as follows. The current studies on sentiment analysis are reviewed in section II. Next, sentiment analysis model is discussed in Section III. Experiments and results are discussed in sections IV and V. Final conclusions are presented in Section VI.

II. Related Work

In [1], the authors have analysed the sentiments of Twitter users regarding Bitcoin. The negative and positive sentiment scores were combined with Bitcoin's historical price and then fed it to the Recurrent Neural Network (RNN) model in order to predict the future price of Bitcoin. The sentiment classification accuracy was 81.39%, while the accuracy of Bitcoin's future price prediction was 77.62%.

In [2], a dataset of 2000 sentiments in formal Arabic about politics and arts were collected from Twitter (1000 positive and 1000 negative). Three machine learning algorithms were applied for classification which are Naive Bayes (NB), Decision Tree (DT) and Support Vector Machine (SVM). To improve the classification performance, Term Frequency-Inverse Document Frequency (TF-IDF) and Khoja stemmer algorithm were used in pre-processing stage. The results showed that DT algorithm has exceeded the other algorithms by obtaining 78% in terms of F-measure.

In [3], the authors have addressed some of the challenges that faces Arabic sentiment analysis. Moreover, the authors have focused on the issues that may occur while analysing informal Arabic opinions. 100 opinions were collected from Twitter. After that, they created a simple system to conduct an experiment on these opinions. The system detected positive and negative words with 82.5% and 71.01% accuracy, respectively.

In [4], the authors have implemented both supervised and unsupervised approaches to sentiment analysis for the Arabic language. 2000 tweets were collected from Twitter (1000

978-1-7281-4213-5/20/$31.00 ©2020 IEEE

positive and 1000 negative) on different topics. For the supervised approach, several experiments were done using four algorithms, namely SVM, NB, DT, and K-Nearest Neighbour (KNN). The findings confirmed that the classification accuracy of SVM and NB exceeded the others with 87.2% for SVM and 81.3% for NB. For the unsupervised approach, the authors built a lexicon which contains 3479 words (1262 positive and 2217 negative). Furthermore, a lexicon-based tool was created to determine the polarity of the text which can handle both intensification and negation. The classification accuracy of the tool was 16.5% when the lexicon contained only 1000 words. However, when the lexicon was expanded to 3479 words, the tool's accuracy was improved and reached 59.6%.

In [5], 350,000 Arabic tweets were collected. After that 25000+ tweets were labeled as positive, negative or neutral by crowdsourcing. The authors used RapidMiner and to overcome some of the issues in the pre-processing stage they built three dictionaries Jordanian dialect dictionary, Negation dictionary, Arabizi dictionary and integrated them to RapidMiner. The following classifiers were used NB, KNN, and SVM. Because of memory issues in RapidMiner when using the whole dataset, a small portion of the dataset was used (1000 tweets). The results showed that NB outperformed the other classifiers and its best accuracy which is 76.78% was achieved when 5-fold cross validation was used and both stemming and stopword filter were not used.

In [6], a tweet analysing approach of two main phases; feature selection and tweets classification was proposed. Feature selection was used to reduce the feature dimensions by choosing the best set of features. Tweets classification is used to classify the sentiments using different meta-heuristic algorithms, namely Glowworm Swarm Optimisation (GSO) based MLP, Genetic Algorithm (GA) based MLP and Biogeography-Based Optimisation (BBO) based MLP algorithms. The results showed that GSO based MLP algorithm outperforms most algorithms.

III. Sentiment Analysis

Twitter sentiment analysis contains four main stages as shown in Fig 3. Stage 1: tweets collection, stage 2: tweets preparation, stage 3: tweets labelling, stage 4: tweets classification.

A. Tweets Collection

To collect our dataset, we use RapidMiner[1] program through the Twitter application. Thirteen keywords used to collect the dataset, which are the most popular hashtags posted in Twitter related to the GEA's activities. As the domain of our dataset is Arabic tweets, we filter the tweets by setting the language "Arabic". We have collected around 6000 tweets for all keywords.

[1]https://rapidminer.com/

Class	Gender	Text
y	Male	الرجل كثر خيره عمل نهضة في البلد ما عمل بها أحد من قبله فلذلك إبقاء تركي ال الشيخ مطلب
y	Female	التنمية والتجدد تكمن في هيئة الترفيه وسعيها للتطور والوصل للمستوى المطلوب منها

Fig. 1. Example of positive tweets in the dataset.

Class	Gender	Text
x	Male	شوف احنا نبغى ملاهي نبغى اماكن سياحيه دايمه مو زي هذه الاشياء نبغى دزني لاند السعوديه نبغى اسعار في متناول الجميع خلاص بأريحك سو لنا زي الامارات زي مدنهم زي فنادقهم سو لنا منتجعات ومسابح في متناول الجميع فكنا من الحفلات الغنائيه راحت فلوسنا على الفاضي لو أبغى اسمعهم اروح اشتري شريط
x	Male	لست من مرتادي الحفلات الغنائية لكن عندي استفسار هل الاسعار ترونها مناسبة ام مبالغ فيها انا أراها مبالغ بها جدا

Fig. 2. Example of negative tweets in the dataset.

B. Tweets Preparation

The collected tweets from the previous stage include high number of duplicate tweets; these may be the result of re-tweeting, also some tweets are empty or contain the address of the sender only or URL only. These unwanted tweets may effect on the classification accuracy, hence these tweets were removed from our dataset. The total number of tweets in the dataset are 3817 tweets, which contains 1906 positive tweets and 1911 negative tweets. Preparation stage contains the following steps:

1) Remove punctuation marks.
2) Correct the errors spelling.
3) Remove URL links and mentions.
4) Remove stop words.
5) Replace the words with its roots.
6) Divide the tweet into words, which is called "n-gram" [7]. In this paper, we use a uni-gram to divide a tweet into single words.

C. Tweets Labelling

The tweets in our dataset are manually labelled as positive or negative tweets. Tweets are considered positive if containing positive opinion or feelings, support or agree words. However, tweets are labelled negative. This step has been done by two members. If there is any difference of opinions between them, then it requires a third member to decide the final class for the tweet. In the dataset, positive tweets are given label "y" and negative tweets are given label "x" as shown in figures 1 and 2, respectively. The dataset is available on the UC Irvine Machine Learning Repository website[2].

[2]https://archive.ics.uci.edu/ml/index.php

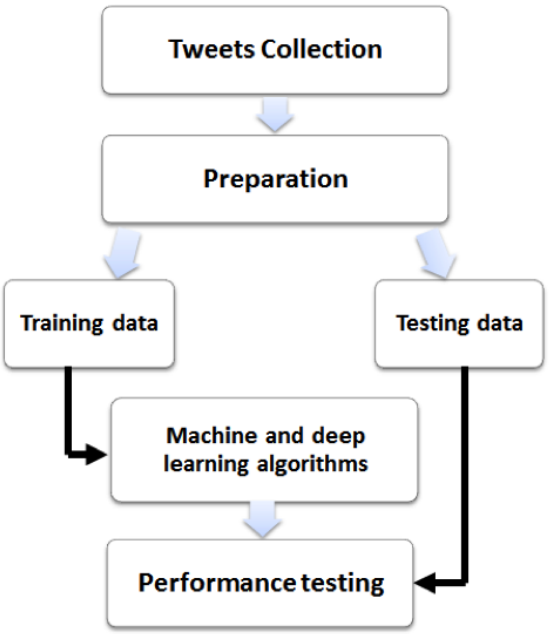

Fig. 3. Twitter sentiment analysis process.

D. Tweets Classification

In our model, machine and deep learning algorithms are applied on the testing data. The classification algorithms i.e., machine and deep learning algorithms, consider a tweet as an input, execute some process on it, and categorise it as positive or negative tweet to be presented as its output. After the classification model is created and tested, performance test metrics are calculated, which are classification accuracy, Mean Squared Error (MSE), precision and recall.

IV. Experimental Settings

In this paper, WEKA is used for classifying the tweets into supporters and opposers to the GEAs' activities. WEKA is developed at the University of Waikato. WEKA provides a comprehensive set of learning methods. It contains tools for dataset preprocessing, data classification, data clustering, and visualising the results [8]. It allows the user to compare different learning methods through flexible GUI. Deep learning package, Deeplearning4j, is developed to integrate the deep learning techniques into WEKA [9]. WEKA does not deal directly with datasets written in Arabic. To do so, the encoded of the dataset is adjusted to UTF-8. The algorithms that are compared include MLP, SVM, RF, and RNN. Parameters setting of classification algorithms are presented in Table I.

A. Performance Test Metrics

Different performance test metrics are used to evaluate the performance of classification algorithms.

Classification Accuracy: is the number of the supporting and opposing tweets which are correctly classified over the total number of tweets, which can be calculated using equation 1.

TABLE I
Parameters Setting

Algorithm	Parameter	Value
MLP	Learning rate	0.1
	Momentum	0.2
	Hidden Layers	5, 7, 10
SVM	Kernel	PolyKernel
	Num Decimal Places	2
	Num of folds	1
RF	Num Decimal Places	2
RNN	Activation Layer	LSTM

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Precision: is the number of the supporting tweets which are correctly classified over the total of the number of the supporting tweets which are correctly and wrongly classified, which can be calculated using equation 2.

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

Recall: is the number of the supporting tweets which are correctly classified over the total of the number of the supporting tweets which are correctly classified and the number of opposing tweets which are wrongly classified, which can be calculated using equation 3.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is false negative.

Mean Squared Error (MSE): is the average of the difference between the true values and the models' values, which can be calculated using equation 4.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{4}$$

where MSE is the Mean Squared Error, y is the model's value and x is the true value.

V. Results and Discussion

A. Case 1: The analysis of supporters and opposers

The percentage of opposers was higher than the supporters In the beginning of our research, after that the percentage of supporters was increasing after the Riyadh season and wonderland activity. As a result, positive tweets are 1906 and negative tweets are 1911, which means that the opinions on GEA are almost equal divided between supporters and opposers. At the end, we analysed the percentage of positive and negative opinions of females as well as males as shown in Fig 4.

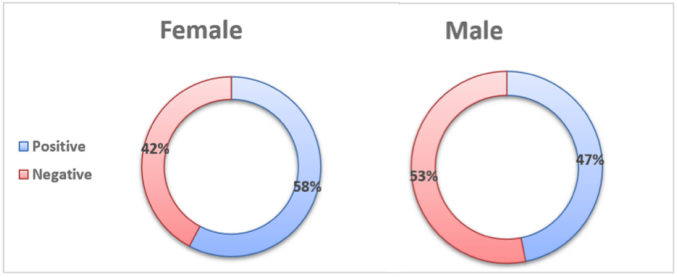

Fig. 4. Supporters and opposers (Females and males).

TABLE II
THE PERFORMANCE OF DIFFERENT ALGORITHMS WITH PERCENTAGE SPLIT TEST

Algorithm	Classification Accuracy %	MSE	Precision %	Recall %
MLP	49.2576	0.2548	0.492	0.493
SVM	79.214	0.2078	0.793	0.792
RF	**80.262**	**0.1504**	0.804	**0.803**
RNN	79.3013	0.1611	**0.812**	0.793

B. Case 2: The effect of using percentage split test

In this test option, 70% of the dataset is used for training and the rest is used for testing. Three machine learning algorithms were applied to classify the tweets, namely MLP, SVM, and RF and one deep learning algorithm, which is RNN. It was figured out of Table II that the RF is the most precise algorithm among the rest algorithms, as well as the lowest MSE of algorithms. In addition, it was concluded that the MLP algorithm has the lowest classification accuracy and the highest MSE rate.

C. Case 3: The effect of using K-fold validation test

We used 10-fold validation test. In K-fold validation, the data is set randomly, then the data is divided into 10 equal folds. For each fold, one fold is taken as a test dataset and the remaining folds as a training dataset. Is has been shown from the Table III that the RF has the highest classification accuracy results as well with the lowest MSE rate. It was concluded that the 10-fold in K-fold validation gives a slightly better results than the percentage split test.

D. Case 4: The effect of increasing the number of hidden layers in MLP

In this case, hidden layers of MLP algorithm were increased to study the impact of this increasing on the results. when using percentage split and K-fold validation tests. Results have been shown that increasing the number of hidden layers leads to increase the percentage of classification accuracy and decrease the rate of MSE when using percentage split test as shown in Fig. 5.

TABLE III
THE PERFORMANCE OF DIFFERENT ALGORITHMS WITH K-FOLD VALIDATION TEST

Algorithm	Classification Accuracy %	MSE	Precision %	Recall %
MLP	62.1692	0.2225	0.622	0.622
SVM	80.7178	0.1928	0.807	0.807
RF	**81.8968**	**0.1377**	**0.820**	**0.819**
RNN	79.3031	0.1616	0.809	0.793

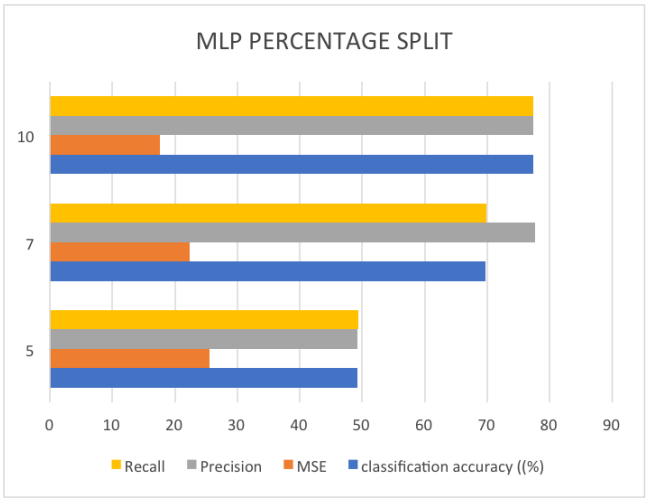

Fig. 5. MLP results in percentage split test with (5,7,10) hidden layers

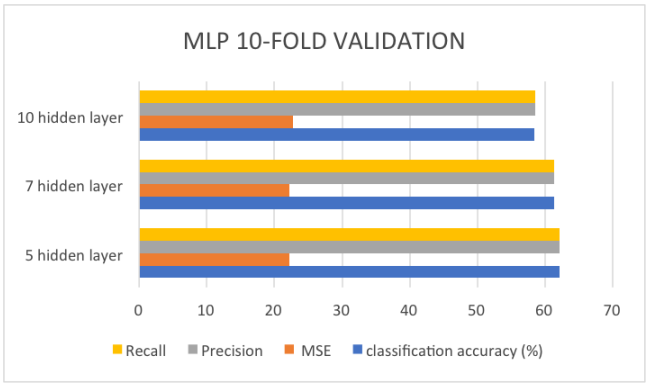

Fig. 6. MLP results in K-fold validation test with (5,7,10) hidden layers.

However, Fig. 6 illustrates that the k-fold validation at k=10 it differs from percentage split test where the relation of this test is inverse relationship. When the hidden layers get increased in MLP algorithm, the classification accuracy decreased.

VI. CONCLUSION

In this paper, the opinions of Twitter users regarding GEAs' activities have been analysed using machine and deep learning algorithms. RapidMiner was used to collect tweets about the activities of GEA from Twitter, which resulted in creating a dataset of 3,817 tweets. The results have showed that people support the GEAs' activities especially the females. Moreover, RF gives the highest classification accuracy and lowest error rate when using cross different tests among machine and deep learning algorithms. In addition, MLP algorithm works better when increasing the number of its hidden layers. In future,

more features can be included and analysed regarding the opinions of people regarding Saudi GEAs' activities.

REFERENCES

[1] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, "Recurrent neural network based bitcoin price prediction by twitter sentiment analysis," in *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, pp. 128–132, IEEE, 2018.

[2] M. M. Altawaier and S. Tiun, "Comparison of machine learning approaches on arabic twitter sentiment analysis," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 1067–1073, 2016.

[3] L. Albraheem and H. S. Al-Khalifa, "Exploring the problems of sentiment analysis in informal arabic," in *Proceedings of the 14th international conference on information integration and web-based applications & services*, pp. 415–418, ACM, 2012.

[4] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in *2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*, pp. 1–6, IEEE, 2013.

[5] R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis in arabic tweets," in *2014 5th International Conference on Information and Communication Systems (ICICS)*, pp. 1–6, IEEE, 2014.

[6] D. A. Alboaneen, H. Tianfield, and Y. Zhang, "Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 4630–4635, IEEE, 2017.

[7] K. O. Ogada, *N-grams for Text Classification Using Supervised machine learning algorithms*. PhD thesis, Jomo Kenyatta University of Agriculture and Technology, 2016.

[8] I. H. Witten, E. Frank, L. E. Trigg, M. A. Hall, G. Holmes, and S. J. Cunningham, "Weka: Practical machine learning tools and techniques with java implementations," 1999.

[9] D. Team *et al.*, "Deeplearning4j: Open-source distributed deep learning for the jvm," *Apache Software Foundation License*, vol. 2, 2016.