Mansoura Journal For Computers and Information Sciences

Proposed Model for Opinion Mining in Arabic Social Media Networks

Taher Hamza[1], Aya M. Al-Zoghby[2], Reem Salama[3]
[1]*[Computer Science Department, Faculty of Computer & Information systems, Mansoura University, Egypt*
Email: taher_hamza@yahoo.com]
[2]*[*Computer Science Department, Faculty of Computer & Information systems, Mansoura University, Egypt
Email: elzoghby.aya@gmail.com; aya_el_zoghby@mans.edu.eg*]*
[3]*[Computer Science Department, Faculty of Computer & Information systems, Mansoura University, Egypt*
Email: reemsalama888@gmail.com]

ABSTRACT

Many research classified opinions into positive and negative with ignoring the neutral classes. Ignoring neutral classes in opinion mining had shown to be an inaccurate practice. Therefore, some previous studies recommended increasing this third class in future works to get better performance and higher accuracy. This work aims to investigate the social opinion mining in regards to Arabic Twitters. It uses neutral training examples in learning because it enables a better division between positive and negative examples, improve pre-processing stage for Arabic language because Arabic language itself is a big challenge, and develop a model for opinion mining in Arabic social media networks. The proposed model was established using a number of classifiers to classify tweets. It is built based on using machine learning on collected data from Twitter to classify tweets on two levels. In the first level, the tweets were organized into positive and negative. In the second, the neutral samples were used in the classification process to distinguish between positive and negative samples. Text pre-processing is the key factor to the sentiment analysis and classification, especially for highly complicated languages (with rich morphology) such as the Arabic language. When the tweets have various approaches of pre-processing, the results showed dissimilar levels of accuracy and also showed the importance of using neutral training examples to facilitate learning. Different experiments had been conducted, using 2,000 identified tweets (1000 positive tweets and 1000 negative ones) on different subjects matters. According to the outcomes from these experiments, the proposed model shows an enhancement in the classification results comparing with some previous works.

Keywords

Sentiment Analysis, Arabic Language, Twitter, Machine Learning, Neutral Class

1. INTRODUCTION

Nowadays, due to the advanced technologies, such as those related to the information and communication technologies, the world is quickly transforming to digitalization [1]. The Internet has become a basic requirement in everybody's personal and business life. The popularity of social network applications allows people to voice their opinions on various topics on different aspects of everyday life [2]. The Internet has made Micro-blogging a widespread communication tool among people online. Micro-blogging websites are rich with data for opinion mining and Sentiment Analysis (SA). SA is a type of *"natural language processing"* to monitor the community's sense of a particular topic. This involves creating a system for gathering and categorizing views on an item or a product. SA often uses Machine Learning (ML) techniques to mine the text for opinion mining [1].

Twitter is an example of a popular micro-blog, which clients may voice their opinions [2], [3]. Twitter allows people to communicate with small texts. On this social media, people use different words and abbreviations, which might cause difficulty to extract their sentiment by the current Natural Language Processing (NLP) systems. Therefore, SA for short texts like Twitter's posts is more challenging [4]. Opinion mining on Twitter is portrayed as unstructured data because of its noisiness. When people post on Twitter, their messages are usually informally written. In can contain spelling mistakes, grammatical errors and improper punctuations. In addition, the maximum limit of characters that are allowed in Twitter is 140. So, people usually do not write accurate full sentences. Thus, all these noises emerge because the micro-blog post is short and uses slang sentences. With such text message condition, traditional opinion mining algorithms do not work well [5],[6],[7]. These problems are common in the Middle East. Arabs interaction on social media is common than any other website online. Therefore, the opinion mining can be used for marketing a particular idea or particular trend, such as political trends, social trends, etc. Researches have showed that opinion mining in Arabic is a further difficult task because of its rich morphology. It is considered a complex task because of these informalities [1]. This is the main holdup for planning text-mining system.

Ignoring the neutral class in opinion mining is not an accurate method as it may be effective in a particular subject at a particular time. The main reason for ignoring neutral class has been dependent on two factors: solving binary (positive vs. negative) drawback because this solves the three-class problem automatically as neutral documents will simply lie close to the binary classifier boundary [8].

Most prior researches explained that data classification is either negative or positive is not accurate .So, finding a ratio in the classification leads to higher accuracy of the classification. Thus, better performance in the opinion mining in social media appears. Furthermore, other prior sentiment-related research ignores the neutral examples, only learning from examples of polarity (positive or negative). They had shown that using neutral samples in polarity training for a variety of reasons is important. What's more, a stronger differentiation between positive and negative examples is encouraged by the use of neutral learning examples [8]. Accordingly, learning from negative and positive samples isolated will not support the correct categorization of neutral samples.

In this paper, we propose a model for opinion mining in Arabic social media networks. This model is based on using machine learning on data from Twitter to classify tweets on two levels. In the first level, the tweets are classified as positive and negative. In the second, the neutral samples are used in the classification process to differentiate between specimens that are positive and negative. Also, there are three main objectives sought in this investigation as: (1) using neutral training examples in learning because it facilitates a better distinction between positive and negative examples; (2) improving pre-processing for Arabic language because Arabic language itself is a big challenge; (3) developing a proposed model for opinion mining in Arabic social media networks.

This paper is organized as follows. The review of prior studies related to Arabic Sentiment Analysis (SA). The methodology used in the study, the illustration of the results that emerged from the analyzing of the experiments, the discussion of the conclusions, and the outline for future work to be carried out.

2. RELATED WORK

In this section, the most prominent papers related to social opinion mining are explored, especially those related to the Arabic language. In [8], Koppel and Schuler outcomes showed that the best solution to polarity is to solve it according to three classes, in which combining classifiers leads to higher accuracy. But it is necessary to use pairwise coupling and combining results in a distinctive way. On the other hand, Abdulla et al. [9] used two methods to evaluate sentiments; the corpus-based and the lexicon-based. They recommended the benefit of adding a third class for the classification. *Accordingly, this study adopted the mentioned approach in the research.* According to Abdulla et al. [9], when the volume of data increase in lexicon-based, it will lead to better performance and higher accuracy. Abdulla et al. [9] mentioned that some sarcasm words lead to wrong polarity classification for the misunderstanding it causes. *This was one of the challenges that were considered in the current research to find a solution with higher accuracy.* Dragoni et al. [10] used the fuzzy logic in sentiment analysis for modeling knowledge. They proved its effectiveness in different domains. This approach inspired the current study to add polarity in order to improve the classification accuracy in different domains.

El Sahar and El-Beltagy [11] had another idea for building multi-domain lexicon from multi reviewing websites, such as movies, hotels, restaurants, and products to be used in the sentiment analysis process. Their experimental results on lexicon showed a fairly good performance for sentiment analysis. They also had some trials through several experiments to test the validity and performance measurement for this data to solve sentiment analysis problems for both two and three classes. In their research, they deduced that the best classifier was Support Vector Machine (SVM) while the worst classifier was K-Nearest Neighbors (KNN). Due to their conclusions, they recommended the importance and the benefit of sentiment analysis in social media. *Their analysis inspired this research to apply sentiment analysis in social media networks*.

Assiri et al. [12] surveyed about Arabic SA. They deduced that the Arabic sentiment analysis contains defects due to the different dialects in the Arabic language, which are not processed sufficiently. This represents a serious challenge as most available Arabic texts on social media are in different dialects and are morphologically complex. Correspondingly, the study found that same classifiers can be used, but on different data. So, they concluded that it is necessary to use creativity in experiment designing and in developing classifiers techniques. Finally, their study revealed that using deep learning and big data techniques such as Hadoop and Mapreduce to develop Arabic SA and enhance the performance to reach the best results.

Mohammad et al. [13] showed the importance of using lexicons in enhancing a system's performance. They translated existing English sentiment lexicons into Arabic, using Google. They showed the importance of Arabic Dialectal Hash-tag Lexicon. They were able to support that adding some features from the translated lexicon leads to a better performance of classification accuracy. This conclusion deemed important and useful for the upcoming researches.

Alayba et al. [14] focused on the "Health Service" and designed an Arabic SA system. The importance of this research lies in the benefit of how to collect data from twitter, how to purify data, and how to apply the pre-processing stage of the Arabic text. This study was considered useful in shedding valuable insights regarding distinguishing the different experiments that can be applied, for example the Deep Neural Networks (DNN) and (ML) algorithms, such as Naïve Bayes (NB), Logistic Regression (LR), SVM and Convolutional Neutral Networks (CNNs).

Howells and Ertugan [15] used fuzzy logic to develop samples which can analyze micro-blog, such as tweet on Twitter. This study provided comprehensions on how to use model fuzzy logic-based with classifying tweets into five categories: strongly positive, positive, neutral, negative, and strongly negative. *This approach was adapted in the current study's methodology.*

Badaro et al. [16] presented a complete literature survey about Arabic opinion mining. They mentioned future directions for researchers interested in studying further opinion mining in Arabic. In their research, the latest advances in the field (such as the learning advances in Arabic Opinion Mining) was reviewed and critiqued. They concluded some discussions on developing Arabic-specific visualizations. This discussion was compared with visualizations of the English opinion-mining systems.

Alharbi and Khan [17] designed a system that sought to detect Arabic comparative opinions from YouTube comments. In their study, they used ML algorithms. They showed that Arabic language is complex. Thus, the language contains many noises and problem difficulties when compared to other languages, such as the Latin languages. Problems found in the Arabic language include: *"the short vowels, Alhamzah, prefixes, suffixes, colloquial, etc."* [17]. *This conclusion was useful for the current study to identify problems of the Arabic language, and to deal with the difficulty of processing it in order to reach high accuracy results.*

Ghallab et al. [18] used a systematic literary review. In their paper, they filtered and reviewed 108 published researches. These researches came from 11 journals and 22 conference proceedings. They conclude that the Arabic sentiment analysis needs more research to be conducted. There is a lack of research on Arabic language characteristics. There is limited research on building standardized datasets. Future analysis should relate to preprocessing process, features selection, and classification methods. Future papers should also apply promising classification methods.

Through the presentation of the above prior studies and literature, it can be concluded that there are many techniques for SA such as Lexicon-based approaches, supervised Approaches (ML Technique) and Deep Supervised approaches (Deep Learning technique). These preceding works have clarified the importance of the Arabic language inside social media. The Arabic language is the fourth most-used language on the Internet; yet there is a lack of research on the opinion mining in Arabic social media networks. Prior studies have showed that the Arabic sentiment analysis still needs more research [16]. Furthermore, prior studies elucidated the importance of pre-processing of the Arabic language to avoid some of the problems of processing, such as the sarcasm words and the multiplicity of dialects [9], [12]. These outcomes encouraged the conduction of this research. This study takes into consideration the pre-processing to reach high accuracy in the results. Enrich the research field in searching for opinions mining in Arabic language and the optimal use of the neutral class to distinguish between positive and negative examples, using ML techniques and reaching high-accuracy results.

3. THE PROPOSED MODEL

In this section, the methodology in this research is outlined and the proposed framework is illustrated.

A. Dataset

We collected 2,000 tweets from Twitter (1,000 positive tweets and 1,000 negative tweets) in various topics that people usually post on. These topics relate to different aspects of everyday life, such as politics, arts, and leisure. The reviewed tweets in this study are in Modern Standard Arabic (MSA) and/ or the Jordanian dialect. The selected reviewed tweets convey certain feelings (positive or negative). We sought to obtain the needed valuable information from such tweets. We aimed to determine the emotion orientation of the posted text. Table 1 shows a summary of dataset information [9].

Table1. Summary of dataset information

	Positive	**Negative**
Total tweets	1000	1000
Total words	7189	9769
Avg. words in each tweet	7.19	9.97
Avg. characters in each tweet	40.04	59.02

B. Pre-processing

Pre-processing is the first step used to transfer Arabic documents to a form that is suitable for classification tasks. In this step, a few linguistic tools are applied (tokenization, normalization, stop word removal, and stemming). The linguistic tools reduce the ambiguity abstraction concepts and the uncertainty of words. These tools should lead to the accuracy of the classification system. After reading the available data and merging it in one data frame, many pre-processing methods are applied to it automatically.

1- *Tokenization*

Tokenization (also known as text segmentation or lexical analysis) is a step of separating longer text strings into smaller pieces or tokens. Segmentation breakdowns and organizes large chunk of text into parts larger than words (e.g. paragraphs or sentences). Tokenization is kept for the breakdown process, resulting in words exclusively. The processing is achieved after a piece of text had been tokenized.

2- *Tweet Cleaning and Normalization*

In this step, the defined parameters are used in data cleaning such as emojis, punctuations, and letters. The data cleaning is normalized in several steps. The steps are:

a. Remove numbers, strange characters, non-Arabic characters, extra spaces, punctuations, and the stop words that do not have any significance or indications from the data. For example:
    1. Conjunctions and pronouns (he هو, she هى)
    2. Prepositions (to الى ,in فى)
    3. Interrogatives (where اين ,when متى)

b. Remove (tashkeel) diacritization problem Such as (fatha 'َ', dammah 'ُ', kasrah 'ِ'), because it causes the creation of multiple forms from the same word [22], which leads to a lack of understanding. For example: (/ كَتَبkataba) (كُتُب /kotob).
c. Remove repeated letters; Table 2 shows some examples:

Table2. Examples of the repetition letters

Tweets/Sentences	English Translate	After pre-Processing
مسكيييييييين Miskin	poor	مسكين
كتييييير رخيص katir rakhis	Very cheap	كتير رخيص

d. Remove duplicate rows.
e. Normalization:

The normalization process is the next step. Its characters suggest putting all words on an equal footing, requiring uniform processing. Normalization contains a series of tasks that place all text on the same level. For example, it can change all texts to the same case (upper or lower), remove punctuation, and convert numbers to their word equivalents. Normalization for the letters in Arabic language (أ and ة) done as all types of the letter "Alif" (ا إ أ) are normalized to (ا) and the letter "ta'a" (ة) is normalized into (ه). Table 3 presents the examples of Normalization. Many Arabic messages posted online use these characters interchangeably between these similar letters [12]. Thus, it leads to the significance for introducing such a system. Table 4 presents a cleaned Arabic tweet.

Table3. Examples for Normalization

Letter	After normalization
ا , إ, أ/ Alif	ا
ة/ , ه ta'a	ه
ى , ي/ ya'a	ي

Table4. Example of Cleaned Tweet

Original tweet	يا جمااااعه يقولون بشار الأسد مااات
Cleaned tweet	يا جماعه يقولون بشار الاسد مات
Tokens	يا , جماعه , يقولون , بشار , الاسد , مات

3- *Checking for emojis*

In this step, the Filter of the emoticons is conducted. Meaningful names are given to each symbol. We checked for positive and negative emojis, and give an initial weight of 1 and -1 to each emotion. Then, we kept track of the number of emotions.

Emotion weight = sum (emotion initial weight) * count of emotions.
When the number of emojis per tweet is increased, we calculated it by the previous equation and normalizes value between +1 to -1. Table 5 and Figure 1 present an example of the converting emotion.

Table 5 Examples of the converting emotion

Emotion	Tweets	English Translate	After converting the Emotion	Emotion weight
:)	أهلا وسهلا بحضرتك :) Ahlan wa sahlan :)	Welcome :)	أهلا وسهلا بحضرتك smileys	+1
:(	ازعج شئ الكذب :(aizeaj shay alkazib :(	Annoying lying :(	ازعج شئ الكذب sad	-1
-_-	-_- طبعا واقع Taban waqia -_-	Of course, reality -_-	طبعا واقع neu	+1

Fig 1*:* Examples of converting emotion

4- *Checking for hash-tag existence*

We assigned weight of 1 or 0 to each sentence whether it contains hash tag or not.

5- *Checking for url existence*

We assigned weight of 1 or 0 to each sentence whether it contains url or not.
We sought to remove the noise from the data as much as possible. The text has to be as clean as possible before converting it into vectors in order to avoid getting memory consumed. If any of the checks are true, an indicator will be added to the vector. If any of the checks are false, another indicator will be added.

The Preprocessing Algorithm:

Input: CSV file consisting of training documents.
Test documents.
Output: Classification model based on training data.
Class labels for test documents.

Begin

1. Separate longer text strings into smaller pieces or tokens // the list of tokens becomes input for processing//.
2. Remove numbers, strange characters, non-Arabic characters, extra spaces, punctuations, and the stop words.
3. Remove (tashkeel) the diacritics of the words Such as (fatha ‘َ’, dammah ‘ُ’, kasrah ‘ِ’).
4. Remove repeated letters.
5. Remove duplicate rows.
6. Normalization for the letters in Arabic language such as (أ and ة) done as all types of the letter "Alif" (ا إ أ) are normalized to (ا) and the letter "ta'a" (ة) is normalized into (ه).

//check for positive and negative emojis ,give initial weight of 1 and -1 to each emotion , then to keep track of number of emotions//

7. If emoji in tweet and emoji in smileys
   Then emojis_list.append(1)
8. Else If emoji in tweet and emoji in sad
   Then emojis_list.append(-1)
9. Else If emoji in tweet and emoji in neu
   Then emojis_list.append(1)
10. Else emojis_list.append(0)
11. If the number of emojis increase more than one per tweet
    Then calculate emotion weight by (sum (emotion initial weight) * count of emotions)
    Normalize it value between +1 to -1.
    // check for hashtag existence//
12. If tweet contains hashtag (#)
    Then assign weight of 1.
13. Else assign weight of 0
    // check for url existence//
14. If tweet contains url
    Then assign weight of 1.
15. Else assign weight of 0

End

C. Converting Data into TFIDF vectors

This process is called document indexing (represented as a vector). Accordingly, each dimension corresponds to a separate word. TFiDF is one of the most common methods used to calculate it. The TFiDF method is able to merge emoji check, url and hashtag into the Tfidf matrix.
.

Validation and performance evaluation were performed and cross-checking using TFIDF was used. TFIDF is a data recovery technique that measures the TF frequency of a word and the IDF frequency of its inverse text. Each word has its TF and IDF rating respective. The output of a term's TF and IDF scores is called the term's TFIDF weight .The higher the TFIDF weight, the rare the term, and vice versa. Using the TFIDF algorithm, the significance of a keyword is based on the frequency of the word in any documents and content. More importantly, it checks the web-wide relevance of the keyword, which is called the corpus. For a term t in document d, the weight $W_{t,\,d}$ of term t in document d is given by:

$$W_{t,\,d} = TF_{t,d} \log (N/DF_t)$$

where:
$TF_{t,d}$: the number of appearance of t in document d.
DF_t: the number of documents involving the term t.
N: the overall number of documents in the corpus. [1]

D. Data Categorization

In this step, ML techniques such as SVM, NB, D-Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), Logistic Regression (Log), and Stochastic Gradient Descent (Sgd) are suggested to construct the classification model. The labels are

[1] http://www.ultravioletanalytics.com/blog/tf-idf-basics-with-pandas-scikit-learn

encoded for the training process and divide the data into sets for training and testing. Then, the data were split into training and testing sets to build a well-performing ML model.

It is essential to train the model on and to test it against data that come from the same target distribution. In this paper, data was split according to specific percentage, using the higher one for training and a lower one for testing. In this study, 80% of the data was used for training; and 20% of the data was used for testing. Figure 2 shows our framework of the proposed model for SA in Arabic tweet.

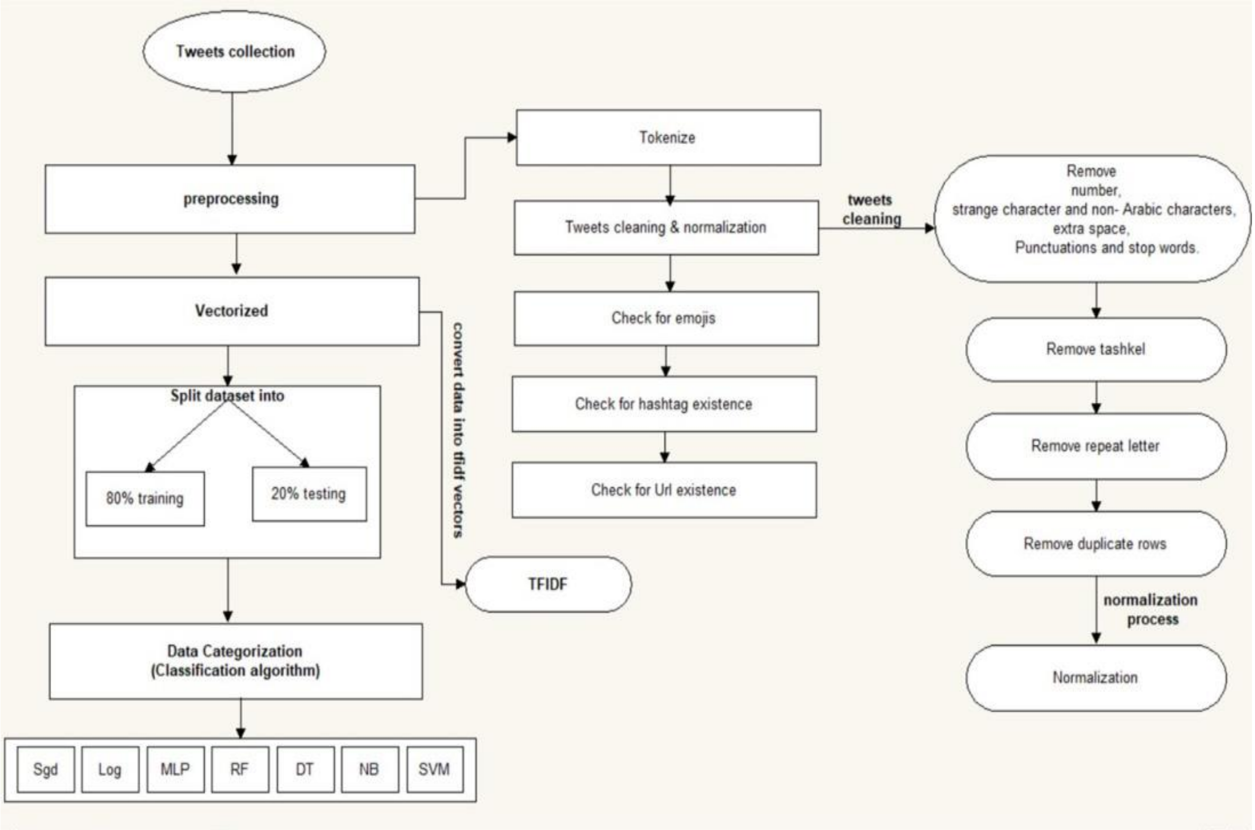

Fig 2: Framework of the proposed model for SA in Arabic Tweets

4. EXPERIMENTS

We built feature vectors by using TFIDF algorithm and encoded labels for the training process. The evaluation of the performance for classification model using Confusion Matrix was applied several mathematic rules such as recall, precision, and F-measure which are defined as follows:

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \quad (1)$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

where TP is the number of documents that are correctly assigned to the category.

ii. TN is the number of documents that are correctly assigned to the negative category.
iii. FP is the number of documents a system incorrectly assigned to the category.
iv. FN are the number of documents that belonged to the category, but are not assigned.

The success measure, namely, F1 score, a well-known F1 measure, is selected for this study which is calculated as follows:

$$\text{F1 score}=\frac{2*\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad (3)$$

*The first level (Positive and Negative Classes):
The accuracy is changed depending on the text's polarity and the classifier used. Table 6 presents the accuracy results of the different classifiers used.

Table 6 The accuracy of classifying data into two classes (Positive and Negative)

Classifier	Accuracy
MLP	88%
SVM	87%
Naïve-Bayes	86%
Sgd	85%
Log	84%
RF	79%
D-Tree	77%

Table 7 shows a comparison of the results of applying our model with [9] on the same dataset.

Table7.Comparison between the results of our study and [9]

	Algorithm						
	Svm	**Nb**	**Dt**	**Rf**	**MLP**	**LOG**	**SGD**
Model1 (our study)	87%	86%	77%	79%	88%	84%	85%
Model2 (Study [9])	87.2%	81.3%	50%	-	-	-	-

*The second level (Positive, Neutral and Negative Classes):

The accuracy is changed depending on the text's polarity. The classifier used neutral examples are involved. Table 8 presents the accuracy results of the different classifiers used.

Table 8.The accuracy when classification made on three classes (Positive, Neutral, and Negative)

Classifier	Accuracy
MLP	54%
SVM	57%
Naïve-Bayes	56%
Sgd	55%
Log	58%
RF	52%
D-Tree	45%

5. Results and Discussion

Using twitter, collected 2000 labeled tweets (1000 positive tweets and 1000 negative ones) on different subjects are used such as politics and arts, applied a model of a pre-processing and several of classifier algorithms for two levels, The first level (Positive, Negative) while the second level (positive, neutral, negative) to show different levels of accuracy when addition to neutral training examples, and experimental results showed this indeed.

- The first level: The best accuracy was achieved using MLP (88%) when the data is classified into two classes (Positive and Negative) as shown in Table 6. But, the worst accuracy was achieved by using DT (77%).
- Table 7 shows a comparison between our study and [9]. It shows that the work in [9] applied three classifiers (SVM, NB and D-Tree). Their experiments showed that SVM has better accuracy than other classifiers (87.2%). Our proposed model showed that MLP has better accurateness than other classifiers (88%).
- The second level when neutral examples are involved: The best accuracy was is 58% achieved by Logistic Regression (Log) as shown in Table 8 and the worst accuracy is 45% achieved by using DT. When the classification is based on three classes, the accuracy is changed because adding a neutral class shows different results of classification accuracy. Moreover, using neutral training examples in learning enhanced the differences between positive and negative examples.

6. CONCLUSION AND FUTUREWORKS

This paper proposes a model for opinion mining in Arabic social media networks based on machine learning using to classify tweets. It showed that the neutral examples cannot be ignored especially in Arabic language because they are reflecting more than one meaning for the same word. Furthermore, the use of neutral examples enhanced the distinction between the positive and negative examples. It showed different levels of accuracy when neutral examples are involved. The pre-processing consists of all the words found among 2000 twitters comments. They were classified according to groupings of positive and negative notes. The results of this study are linked to [9]. In [9], results showed an accuracy of 87.2 %; while our proposed work achieved 88%. This work contributes academically and practically by developing empirical evidence that focuses on the use of neutral samples in learning polarity.

For future work, fuzzy logic and larger datasets that contain pure Arabic comments could be used in the Arabic sentiment analysis. Also, work more on different Arabic vernaculars that are used in posts and other written messages online.

7. References

[1] Kumar, A., & Garg, G. (2019). Sentiment analysis of multimodal twitter data. **Multimedia Tools and Applications**, 78(17), 24103-24119.

[2] 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.

[3] Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. **Proceedings of the International Conference on Language Resources and Evaluation**,1320-1326.

[4] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011).Sentiment analysis of twitter data. **Proceedings of the Workshop on Language in Social Media** , 30-38.

[5] Mewari, R., Singh, A., & Srivastava, A. (2015).Opinion Mining Techniques on Social Media Data. **International Journal of Computer Applications**,118(6),39-44.

[6] Neethu, M. S., & Rajasree, R.(2013).Sentiment analysis in twitter using machine learning techniques. **International Conference on Computing, Communications and Networking Technologies**,1-5.

[7] Liang, P. W., & Dai, B. R. (2013). Opinion mining on social media data. **International Conference on Mobile Data Management**, 2, 91-96.

[8] Koppel, M., & Schler, J. (2006). The importance of neutral examples for learning sentiment. **Computational Intelligence**, 22(2), 100-109.

[9] Abdulla, N. A., Ahmed, N. A., Shehab, M. A., & Al-Ayyoub, M. (2013). Arabic sentiment analysis: Lexicon-based and corpus-based,**Jordan conference on applied electrical engineering and computing technologies** ,1-6.

[10] Dragoni, M., Tettamanzi, A., & da Costa Pereira, C. (2014). Using Fuzzy Logic For Multi-Domain Sentiment Analysis. **International Semantic Web Conference (Posters & Demos)**, 305-308.

[11] ElSahar, H., & El-Beltagy, S. R. (2015). Building large arabic multi-domain resources for sentiment analysis. **International Conference on Intelligent Text Processing and Computational Linguistics**, 9042¸ 23-34.

[12] Assiri, A., Emam, A., & Aldossari, H. (2015). Arabic sentiment analysis: a survey. **International Journal of Advanced Computer Science and Applications**,6(12), 75-85.

[13] Mohammad, S., Salameh, M., & Kiritchenko, S. (2016). Sentiment Lexicons for Arabic Social Media. **Proceedings of the Tenth International Conference on Language Resources and Evaluation**((LREC'16),33–37.

[14] Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2017). Arabic language sentiment analysis on health services. **1st International Workshop on Arabic Script Analysis and Recognition (ASAR**),114-118.
[15] Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. **Procedia computer science**, 120, 664-670.
[16] Badaro, G., Baly, R., Hajj, H., El-Hajj, W., Shaban, K. B., Habash, N.,& Hamdi, A. (2019). A survey of opinion mining in Arabic: a comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations. **ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP),**18(3),1-52.
[17] Alharbi, F. R., & Khan, M. B. (2019). Identifying comparative opinions in Arabic text in social media using machine learning techniques. **SN Applied Sciences**, 1(3), 213.
[18] Ghallab, A., Mohsen, A., & Ali, Y. (2020). Arabic Sentiment Analysis: A Systematic Literature Review.**Applied Computational Intelligence and Soft Computing**,2020,1-21.
[19] kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. **International Journal of Computer Applications**, 139(11),1-11.
[20] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data .**Proceedings of the Workshop on Language in Social Media**,30-38.
[21] Liang, P. W., & Dai, B. R. (2013). Opinion mining on social media data.**14th International Conference on Mobile Data Management**,2, 91-96.
[22] Al-Osaimi, S., & Badruddin, M. (2017). Sentiment Analysis Challenges of Informal Arabic Language.**International Journal of Advanced Computer Science and Applications** ,8(2) ,1-17.