MJIT 2024 Malaysian Journal of Industrial Technology

SENTIMENT ANALYSIS OF LIBYAN TWEETS USING MACHINE LEARNING ALGORITHMS

Ramadan Alsayed Alfared.
University of Zawia, Zawia, Libya.
Ramadan.Alfared@zu.edu.ly

Hanan Mohamed Aljarm.
University of Zawia, Zawia, Libya.
Pg01401068@zu.edu.ly

*Corresponding author's email: Ramadan.Alfared@zu.edu.ly

ARTICLE INFO

Handling Editor: Rahimah Mahat

Article History:
Received 25 December 2023
Received in revised form 12 January 2024
Accepted 5 February 2024
Available online 15 March 2024

Keywords:

Sentiment Analysis; Text classification; machine learning; Twitter; social media; Libyan dialect

ABSTRACT

Sentiment analysis is a highly active field of study in natural language processing, also known as opinion mining. Social media is a communication tool between internet users. Thus, these platforms become valuable data resources that can be exploited and used efficiently to support decision-making. Many researchers are still working on improving the processing of sentiment analysis in textual data (positive or negative comments). Although there are several studies in Arabic dialects using the machine learning approach, no prior work has been conducted on the Libyan dialect.
In this paper, a training dataset of Libyan tweets for sentiment analysis is used to train three machine-learning algorithms. The aim is to determine which algorithm has the best accuracy for our dataset. We use Cohen's Kappa measure to evaluate the quality and measure the reliability of the sentiment annotations which observed agreement is 89.1%. The experiments of the three algorithms showed that the decision tree algorithm achieved better results compared to the other algorithms in terms of accuracy. The results are (72%, 69% and 65%) in (Decision Tree, Support Vector Machine and Naive Bayes) respectively.

1.0 Introduction

Natural language processing (NLP) is a very active domain in Artificial Intelligence (AI) [1] that exploits the most advanced algorithms to allow understanding of human language. NLP is a branch of AI and has several applications including sentiment Analysis [2], Machine Translation [3] and Information Retrieval [4].

In recent years, many users have been interested in social networking platforms like Twitter, Facebook, and Instagram. Social media is a communication tool between internet users, it is considered a good free resource for any NLP applications. The majority use social sites to express their emotions, beliefs or opinions about products, some issues, services, and life, to gain valuable information for decision-making. A machine learning algorithm (MLA) is used to analyse the huge collection of data. There are several works on Arabic sentiment analysis (Dialects) which have gained considerable interest in the research community such as Jordanin [5], Morocan [6], Saudia [7] and Tunisian [8].

Recently Libyan dialect had the one of the attention dialects that is building a dataset for use in many NLP applications such as [9]. In the NLP field, there are three principal approaches such as machine-learning [10], Lexicon-based [11] and hybrid [12].

In this paper, we apply three ML models for the Libyan dialect sentiment analysis (SA) including *Naive Bayes*, *SVM* and *Decision Tree* to evaluate each model and compare between them in terms of accuracy of *recall* and *precision*. This comparison determines which model is appropriate for the Libyan dialect. The study includes a dataset of 2296 Libyan dialect tweets randomly retrieved over the period from the 20th of March to the 23rd of July 2023.

The remainder of this paper is structured as follows: Section 2 surveys the literature on sentiment analysis. In Section 3, we explore the different types of machine learning algorithms and sentiment analysis. We present the data gathering in Section 4. Section 5 evaluates the quality of the annotation and measures the reliability of the sentiment annotations. Section 6 is dedicated to a detailed presentation of our proposed method for sentiment analysis in the Libyan dialect by using three ML algorithms. Section 7 provides a discussion that demonstrates the efficiency and accuracy of three models for the Libyan dialect sentiment analysis. We finally draw some conclusions and future work directions in Section 8.

2.0 Previous Work

During the last decade, much research has been done in the field of analysis of feelings. Some papers studied the polarity analysis and stated that dealing with the Arabic dialect is a challenging task due to the complexity of Arabic morphology. There are several papers were conducted in this field:

The paper [13] conducts a rigorous comparison of different machine learning approaches for Arabic SA including Naive Bayes, SVM, CNN, LSTM, and a variety of the recently introduced language models. They examine the use of transformer-based language models for Arabic SA and show their superior performance compared to the existing approaches, where the best model achieves F-score scores of 0.69, 0.76, and 0.92 on the *SemEval*, *ASTD* and *ArSAS* benchmark datasets [14]. They also apply an extensive analysis of the possible reasons for failures, which show the limitations of the existing annotated Arabic SA datasets, and the challenge of sarcasm that is prominent in ADs. Finally, they highlight the main gaps in Arabic sentiment analysis research and suggest the most in-need future research directions in this area. In the paper [15] the author introduces a framework that can perform sentiment analysis on tweets written using either Modern Standard Arabic MSA or Jordanian dialectical Arabic. The core of the framework is a dialect lexicon which maps dialectical words into their corresponding MS Arabic words. The experimentation reveals that the dialect lexicon improves the accuracies of the classifiers.

The results reveal that replacing dialectical words with their corresponding MSA words improves the overall *Precision*, *Recall* and *F-Measure*. When examining the results at the class level. The authors conclude that *Precision* of the *Positive* class was slightly improved when the dialect lexicon was used with the two classifiers. The Precision of the Negative class was slightly improved when the dialect lexicon was used with the Naive Bayse classifier. The Precision of the Neutral class was not improved when the dialect lexicon was used. The Recall of the Negative class was greatly improved when the dialect lexicon was used with both classifiers.

In the paper [16], four classifiers were trained on a dataset from Twitter, namely Naïve Bayes, SVM, Multinomial Logistic Regression and K-Nearest Neighbor to conduct a comparative analysis of the performance of the classifiers. These algorithms when run against the tweets dataset the results revealed that SVM gives the highest F1-score (72.0) while the best accuracy was achieved by KNN (k=2) and it equals 92.0.

In a paper [17], two human annotators labelled tweets based on their polarity. They found that 500 of the tweets were positive and 500 were negative. Different preprocessing methods were used including removing hashtags, non-Arabic words, usernames, and URLs. They experimented with two features: unigrams and bigrams and two classifiers: SVM and Naive Bayes. Two experiments were implemented: one which included removing the stop words and one without removing them. Removing stop words led to a slight improvement in the performance. This suggests that stop words are valuable to the sentiment or that other stop words need to be removed. The results show that SVM led to a greater performance than NB giving a 4-6% increase in accuracy. The best model was SVM with unigrams, with an accuracy was 72%.

3.0 Machine learning algorithms and sentiment analysis

Machine learning algorithms use various techniques to learn patterns and make predictions from data, some common techniques include:

1. Supervised Learning: this technique involves training the algorithm on labelled data, where the input features are mapped to known output labels. The algorithm learns to generalize from this labelled data and make predictions on new, unseen data [18].

2. Unsupervised Learning: in this technique, the algorithm learns patterns and relationships in unlabelled data without any predefined output labels. It discovers hidden structures or clusters in the data [19].

3. Semi-Supervised Learning: this technique combines elements of both supervised and unsupervised learning. It uses a small amount of labelled data along with a larger amount of unlabelled data to train the algorithm [20].

4. Reinforcement Learning: this technique involves training an agent to interact with an environment and learn from feedback in the form of rewards or penalties. The agent learns through trial and error to maximize its cumulative reward [21].

5. Deep Learning: Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers (deep neural networks) to learn complex patterns from large amounts of data. It has been particularly successful in tasks such as image recognition and natural language processing [22].

6. Ensemble Methods: Ensemble methods combine multiple individual models (e.g., decision trees, neural networks) to make predictions collectively, often resulting in improved accuracy and robustness [23].

Most of studies on the sentiment analysis were conducted using first technique (Supervised Learning) and ensemble models with different machine learning algorithms, including Naive Bayes, Support Vector Machine …etc. These are just some of the techniques used in machine learning algorithms, and different algorithms may employ different combinations or variations of these techniques depending on the specific problem being solved.

4.0 Data Gathering

To collect the data from Twitter, we use the Twitter Search Application Programming Interface (API) which allows harvesting a stream of real-time tweets by querying their content. To retrieve tweets which are relevant to the Libyan Dialect, we develop a set of search queries to increase the chance of obtaining tweets that convey the Libyan dialect. To get more accurate query results, three types of queries were used to get the tweets in the Libyan dialect. The first query is a group of keywords (lexicon) that was used in the Libyan dialect, while the geocoding system was used in the second type of query. The third type of query is to combine two of the previously mentioned types of queries into one complex search. [24]. Note that for training a classifier, these query terms are replaced by placeholders. The extracted data was cleaned in a pre-processing step, e.g. by normalizing usernames and digits and eliminating Latin characters (i.e. URLs, emails). Figure (1) shows a sample of the keywords used to search for a set of related tweets.

I D	Libyan Dialect keyword	English meaning	relevant tweet which contains certain keyword	English meaning of tweet
1	باهي	Good, ok or agree	وسع بالك باهي تو تفهم الموضوع	Make your patience then you will understand the subject
2	هلبة	Many or much	قعمزي بعدين ساهل الموضوع سهل هلبة	Sit down, then everything will be easier, the matter is very easy
3	راهو	Be informed	يبصر عليك راهو تعال تفضل الحوش حوشك	Be aware, he's flirting with you. Go ahead, my house is like your home
4	كنك	What's wrong with you?	كنك راهو نبصر عليك ماتز علش مني	Be aware, he's flirting with you. do not be angry from me
5	علاش	why	علاش ديما قاعد ومكسد بروحك	Why are you always sitting on your own
6	قعمز	sit down	قعمز مكانك فيه حصة ترفيهية قنينة	Sit where you are. There is an entertaining lecture
7	توا	now	شن بندير توا و الضي هارب ليه هلبة ساعات	What am I going to do now. The electricity has been cut off for hours
8	قنين	Beautiful	متل قديم و قنين احكي هولي ثاني	An old and beautiful example tell me again
9	نبصر	I'm joking	راهو امس نبصر عليك مش جديات الحكاية	You should know that I am joking with you, and not seriously
10	نو	Very hot weather	الجو نو هلبة مش شوية	The weather is very hot and not a little

Figure 1: Some examples of the keywords (Libyan Dialect)

Like most countries, the Libyan dialect varies from one region to another, *east*, *west* and *south*. LD is characterized by the presence of letters that are not pronounced and written correctly. For example, the word "ذهب" *thahib* which means in English *gold* is pronounced as "دهب" *dahip* and the word "ثعلب" *thalip* which means in English *fox* is pronounced as "تعلب" *talip*. In fact, there is a special feature in the Libyan dialect, which is that it contains a huge amount of vocabulary. LD is also characterized by the word "هلبه" which means in English *a lot*, "توا" which means in English *now* or "باتي" meaning *my father* …etc.

5.0 Reliability of Annotations

For annotation, we recruited two teachers of Arabic. To evaluate the quality of the annotation and measure the reliability of the sentiment annotations, we conducted an inter-annotator agreement study on the annotated tweets. We use Cohen's Kappa [25] which is considered as a measure of the degree of agreement among the assigned labels, correcting for agreement by chance.

The overall observed agreement is 89.1% and resulting weighted Kappa reached k = 0.836, which indicates reliable dataset annotations and almost perfect agreement. Two annotators evaluate approximately 2296 tweets as to be *positive*, *negative* or *neutral*, but we just need only the sentences which are considering as *positive* or *negative opinion*.

Table 1: some example annotations (agreement and disagreement among annotators)

#	Tweets	English translation	Annotator1	Annotator2
1.	الاحزاب السياسية هي الي غبرت في البلاد	Political parties are what destroyed the country	Negative	Negative
2.	واحد من 10 من الليبيين يستخدموا في الايفون	One in 10 Libyans use an iPhone	Neutral (facts)	Neutral (no-clear positive evaluation)
3.	الصحراء الليبية بدات تنزار من هلبة ناس ليبين وأجانب	The Libyan Desert has become visited by many Libyans and foreigners	Positive	Neutral
4.	علمتنا الثورات العربية بأن الملك إدريس عنده حق يلغي الاحزاب السياسية	The political revolution (Arab Spring) has taught us that King Idris had the right to abolish political parties	Uncertain (unclear sentiment indicator)	Negative

Table 1 represents some example annotations from our training data. For instance, tweet #1 shows an agreement among annotators in labelling tweets with a clear negative polarity, while tweet #2 indicates a lower level of agreement: That is, when the annotators agree on the selected label neutral, but disagree on the reason for the annotation. The third example reflects a significant level of disagreement regarding the neutral and positive polarities. Furthermore, sarcastic and heterogeneous tweets have created a challenge even to human annotators. Tweet #4 represents a disagreement among annotators, whether it is a sarcastic view of very complicated and tragic circumstances, or just a negative attitude. In the context of Arabic dialect, sarcasm is difficult to detect because it uses positive indicators to express negative emotions, these expressions of sentences exists in the most of Arabic dialects.

6.0 Methodology

Our approach goes through several stages, including methods for obtaining data from Twitter, then the data filtering stage (data-preprocessing). The next step is the labeled dataset as a positive or a negative opinion. For this, we use three ML algorithms as shows in Figure 2. The dataset was obtained in textual data format is consists of 1,700 Libyan tweets. We have been divided our collected data into two datasets. The first dataset is the training data which represents around 80% of the total size while the second data constitutes of 20% as testing data.

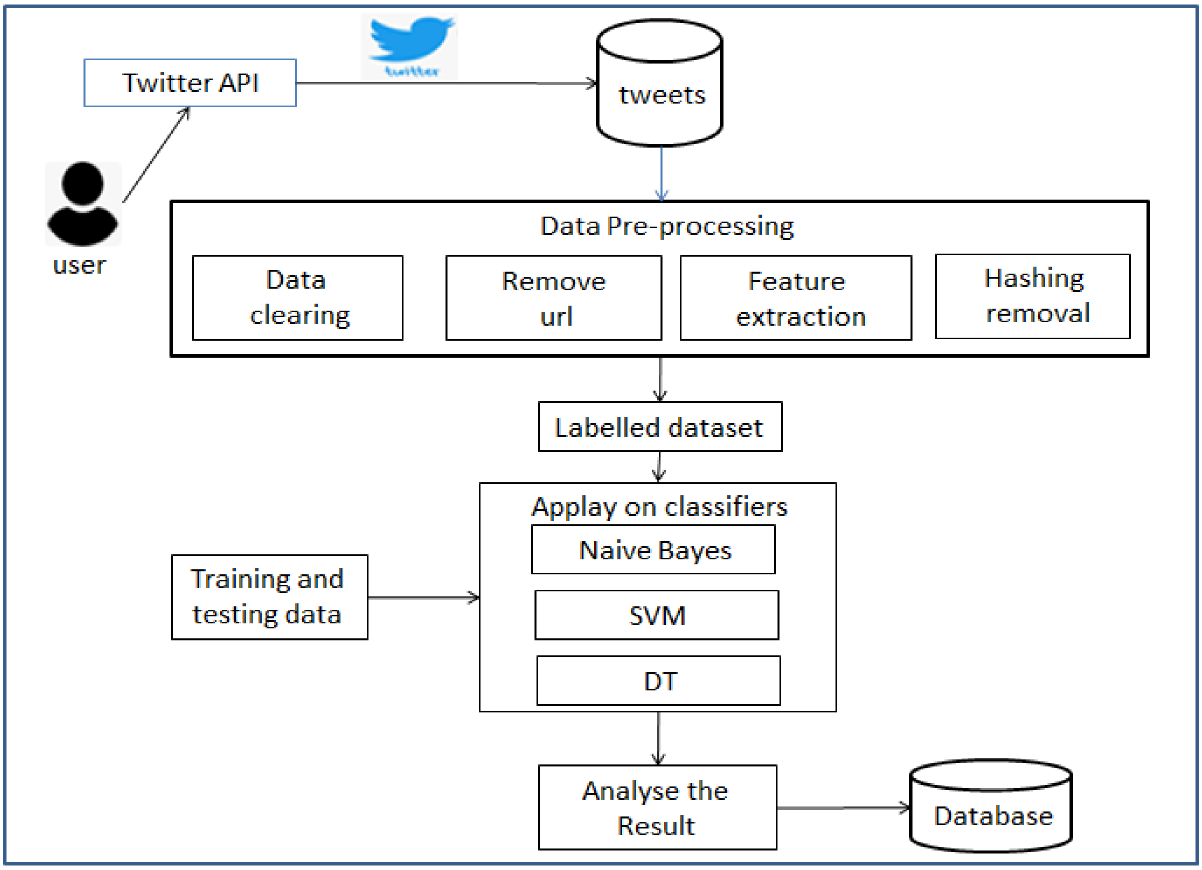

Figure 2: General form of the system

The training set will be used to train the model, while the testing set will be used to evaluate its performance (in term accuracy of the *precision* and *recall*). We then stored the dataset and the result in the database.

6.1 Definition of the three used models:

There are several data mining algorithms that can be used to implement the supervised learning such as *Naïve Bayes* (NB), *Support Vector Machine* (SVM) and *Decision Tree* (D-Tree). We use these three models of machine learning algorithms for the Libyan dialect SA, and compare them in term accuracy of *precision* and *recall*; the models are presented as following:

- **Naive Bayes** is a common machine learning algorithm used for sentiment analysis. It is based on Bayes' theorem and assumes that the features are conditionally independent of each other given the class label. Naive Bayes is one of the simplest probability classifications, which is a classification technique based on Bayes' theorem. The Naive Bayes classifier assumes that the existence of a particular feature in a class is unrelated to the existence of any other feature. It implies the assumption of independence among the predictors .Naïve Bayes is a probabilistic classifier that uses Bayes' theory to classify texts based on evidence that appears in the training data. Computes the conditional probability of each attribute in a given text based on the presence of that attribute in that category and multiplies the probabilities of all the attributes of a given text to calculate the probability of the final classification for each category (tweet).

- **Support vector machine** is a powerful tool for sentiment analysis. It is an advanced machine learning model that can accurately classify text data into positive, negative, or neutral sentiments. To use the Support victor machine for sentiment analysis, we will need a labeled dataset of text samples with their corresponding sentiment labels. As the other method, support victor machine needs to divide the labeled dataset into training and testing sets. The training set will be used to train the support victor machine algorithm, while the testing set will be used to evaluate its performance. For evaluate the model: Once trained, evaluate the performance of the Support victor machine on the testing set by comparing its predicted sentiments with their actual labels. Common evaluation metrics include *accuracy*, *precision*, *recall*, and *F1 score*.
- **Decision Tree Model** can be constructed as follows:
  1. Start with the root node, which represents the input text.
  2. Split the data based on a feature that helps in distinguishing *positive* and *negative* sentiments. This feature could be the presence of certain keywords or phrases commonly associated with *positive* or *negative* sentiments.
  3. Create child nodes for each possible outcome of the split, representing different branches of the decision tree.
  4. Repeat steps 2 and 3 recursively for each child node until a stopping criterion is met.

The stopping criterion could be reaching a maximum depth for the tree, having a minimum number of instances in a node, or achieving a certain level of accuracy.

Assign a sentiment label (positive or negative) to each leaf node based on majority voting among the instances in that node.

The decision tree can then be used to classify new input texts by traversing down the tree based on their features until reaching a leaf node with a sentiment label.

It's important to note that constructing an effective decision tree for sentiment analysis requires careful feature selection and tuning to achieve accurate results. Additionally, other techniques such as pruning and ensemble methods can be applied to improve the performance of the decision tree model.

7.0 Experimentation and result analysis

In this section, a comparative analysis has been performed in order to examine the performance of the three classification models: *DT*, *SVM* and *NB* models.
The accuracy of the models can be measured in two ways: *Confusion Matrix* and *Classification Measure*. We first provide some definitions considering ***Confusion Matrix*** which is showed in Table 2. The confusion matrix consists of the four outcomes generated by the binary classification.
-*True Positive* (TP) represents the cases where the model correctly predicted the positive class while *False Positive* (FP) describes the cases where the model incorrectly predicted the positive class.

-*True Negative* (TN) represents the cases where the model correctly predicted the negative class while *False Negative* (FN) describes the cases where the model incorrectly predicted the negative class.

Table 2: A confusion matrix

True positive TP	False positive FP	$Precection = \frac{TP}{TP+FP}\ or\ \frac{TN}{TN+FN}$
False negative FN	True negative TN	$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$
$Recall = \frac{TP}{TP+FN}\ or = \frac{TN}{TN+FP}$		

Precision measures the ability of the model to correctly classify positive or negative sentiments out of all instances it predicted as positive or negative. It is calculated as the ratio of true positives to the sum of true positives and false positives (negative or positive sentiments incorrectly classified as positive or negative, respectively). It is calculated using the formula:

P=Tp/(Tp+Fp)

Recall measures the ability of the model to correctly identify all positive or negative sentiments in the dataset. It is calculated as the ratio of true positives (TPR) (correctly classified positive or negative sentiments) to the sum of true positives and false negatives (positive or negative sentiments incorrectly classified as negative or positive, respectively). It is calculated using the formula:

R= Tp/(Tp+Fn)

F-Score is a measure of a model's accuracy on a dataset. F-measure is a harmonic mean of recall and precision, providing a single metric that balances both measures. It can be calculating by the following formula:

F=2PR/P+R

The following table illustrates the results obtained for the confusion matrices of the three models.

Table 3: results obtained for the confusion matrices of the three models.

Naïve Bayes	TP/138 **FN/89**	FP/31 **TN/88**
Support vector machine	TP/221 **FN/6**	FP/**101** **TN/18**
Decision Tree	TP/185 **FN/42**	FP/62 **TN/57**

Accuracy measures the percentage of correct predication made by the model. If we obtain the high accuracy then the model is best. Because the high accuracy means, it is close to real value while the low accuracy means then, it is far or nearby to real value. It can be calculating by the following formula:

Accuracy = (Tp+Tn)/(Tp+Fp+Fn+Tn)

Accuracy and precision are two important factors to consider when taking data measurements. Both accuracy and precision reflect how close a measurement is to an actual value, but accuracy reflects how close to measurements is to a known or accepted value. While precision reflects how reproducible measurements are, given if they are far from the accepted value. The obtained results show in Table 4.

Table 4: The calculations and results

Algorithms	Accuracy	Precision	Recall	F1-Score
NB	65%	66%	67%	64%
SVM	69%	72%	56%	53%
DT	72%	69%	67%	65%

The measurement of the performance of different algorithms is evaluated in terms of *recall*, *precision*, *f-measure*, and *accuracy*. These algorithms when ran against the tweets dataset the results revealed that DT model gives the highest F1-score (65%) while the best *precision* was achieved by SVM it equals to 72%. The DT model has gained a better level of evaluation comparing to other SVM and NB models. DT model has achieved the highest *accuracy*, *recall* and *F-score*, which were 72% in accuracy 67% in *recall* and 65% in *f-score* evaluation measure.
NB model has achieved the lowest *accuracy* which was 65%. Although, the average precision, recall and f-measure have achieved by NB model are 66%, 67% and 64% respectively.
The results show that DT model led to a greater performance than SVM giving a 2% increase in accuracy. Hence, we can consider that the DT and SVM models are the best for the Libyan dialect SA.

8.0 Conclusion and future work

We have examined three machine learning algorithm models for sentiment analysis of Libyan dialect tweets. The objective was to identify the model that could learn the Libyan dialect and opinion analysis with the highest accuracy rate. In order to do this, we have gathered roughly 2296 Libyan tweets. Each tweet has one of three labels: positive, *negative,* or *neutral.* We used Cohen's Kappa, which is considered a measure of the degree of agreement among the assigned labels, correcting for agreement by chance. The overall observed agreement is 89.1% for the annotated tweets. As an experiment, three classifiers of machine learning algorithms were used to assess our framework. We only dealt with *positive* and *negative* opinions (1727 tweets). The results obtained are promising, and this encourages us to continue working on this topic. The decision tree algorithm is mainly focused, and the result is compared with other models in terms of *recall* and *F-Score* measure. The decision tree model achieved results with an accuracy of 70%, followed by the support vector machine model with an accuracy of 69%, and then the Naive Bayes model with an accuracy of 0.65%. We can conclude that the DT model has the best evaluation in term of accuracy for Libyan dialect sentiment analysis.

As future work, there are certainly numerous ways that this work can and will be progressed and improved. On the other hand, the dataset can be extended by including new data; it could involve collecting a larger and more diverse dataset of Libyan dialect text to train machine learning algorithms. This would help improve the accuracy and generalizability of sentiment analysis models for Libyan dialect.

9.0 References

[1] Turing A."Computing Machinery and Intelligence". Mind 49. pp. 433-460, 1950.

[2] Bhumika Ga, Monika N,Kanika V, Goldi R and Priyanka B, "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python", *International Journal of Computer Applications(0975-8887),*Volume 165-No.9, May 2017.

[3] Salima H, Karima M and Kamel S, "Machine translation for Arabic dialects (survey)", Information Processing & Management, Volume 56, Issue 2, pp 262-273, March 2019.

[4] Abdelkrim A, Ahmed O and Mohammed S, "A Review on Recent Arabic Information Retrieval Techniques", EAI Endorsed Transactions on Internet of Things, Volume 8, Issue 3, 2022.

[5] Atoum, J. O and Nouman M, "Sentiment analysis of arabic jordanian dialect tweets". Int. J. Adv. Comput. Sci. Appl 10 (2), pp 256–262, 2019.

[6] Soukaina M, Brahim A, Ismail EL B, Sara A and Nabil L, "MSTD: Moroccan Sentiment Twitter Dataset", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 10 ,pp. 363-372, 2020.

[7] Alahmary R, H, Al-Dossari and Emam A, "Sentiment Analysis of Saudi Dialect Using Deep Learning Techniques," in International Conference on Electronics, Information, and Communication, Auckland, New Zealand, pp. 1–6, Jan. 2019.

[8] Salima M, Fethi Bo, Yannick E and Lamia Ha, "Sentiment Analysis of Tunisian Dialects: Linguistic Ressources and Experiments", Proceedings of the Third Arabic Natural Language Processing Workshop, Association for Computational Linguistics, pp 55–61, Valencia, Spain, Apr 2017.

[9] Husein A and Ramadan A, "A Topic-based Twitter Sentiment Analysis Training Dataset for Libyan Dialect", International Conference on Technical Sciences (ICST2019), pp 398-401, Tripoli, Libya, March 2019.

[10] Yafooz, W. M., Hizam, E.and Alromema W, "Arabic sentiment analysis on chewing khat leaves using machine learning and ensemble methods. Engineering", Technology & Applied Science Research 11 (2), (2021).

[11] Sherif M , Alamoodi A., Albahri O., Salem ,. Albahri, M, Mohammed R and Gang K, "Lexicon annotation in sentiment analysis for dialectal Arabic: Systematic review of

current trends and future directions", Information Processing & Management Volume 60, Issue 5, Sep 2023.

[12] Mustafa M and Jamal Noja, "Enhancing Arabic Sentiment Analysis Through a Hybrid Deep Learning Approach", International Journal of EducationPages: 183-189, , Culture and Society Volume 8, Issue 4, August 2023.

[13] Ibrahim F and Walid M, "A comparative study of effective approaches for Arabic sentiment analysis", *in* Information Processing & Management, Volume 58, Issue 2, March 2021

[14] Abu Bakr S, Kareem E, Samhaa R, 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, 5-6 Nov 2017.

[15] Rehab M. D, "Sentiment Analysis for Dialectical Arabic", 6th International Conference on Information and Communication Systems (ICICS), Apr. 2015.

[16] Rua I, Mawada O, Mawada T, Noor M and Izzeldein A, "Sentiment Analysis for Arabic Dialect Using Supervised Learning", In International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 12-14 August 2018.

[17] Shoukry,A.and Rafea,A, "Sentence-level Arabic sentiment analysis", in: Collaboration Technologies and Systems(CTS), International Conference on, IEEE. pp.546–550, 2012.

[18] Ismail R, Omer M, Tabir M, Mahadi N, and Amin I, "Sentiment Analysis for Arabic Dialect Using Supervised Learning," in International Conference on Computer, Control, Electrical, and Electronics Engineering, Khartoum, , pp. 1–6, Sudan, Dec. 2018.

[19] Mahyoub FHH, Siddiqui MA and Dahab MY, "Building an arabic sentiment lexicon using semi-supervised learning". Journal of King Saud University-computer and Information Sciences; 26(4): pp. 417-424, 2014.

[20] Elshakankery K., Ahmed, M. and Hilatsa, "A hybrid incremental learning approach for arabic tweets sentiment analysis". Egyptian Informatics Journal 20 (3), pp 163–171, 2019.

[21] Hammad, M and Al-awadi M, "Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning", pp. 131–139. Springer, Information technology: new generations 2016.

[22] Alotaibi A and Abul Hasanat M, "Racism Detection in Twitter Using Deep Learning and Text Mining Techniques for the Arabic Language," in First International Conference of Smart Systems and Emerging Technologies, pp. 161–164, Riyadh, Saudi Arabia, Nov. 2020.

[23] Pansy N and Rupali V, "A review on sentiment analysis and emotion detection from text", Published online: 28 August 2021.

[24] Husien A and Ramadan A, "Building a Twitter Social Media Network Corpus for Libyan Dialect". International Journal of Computer Electrical Engineering, 2017.

[25] Cohen, J, "A coefficient of agreement for nominal scales". Educational and psychological measurement, 20(1):37–46. 1960.