Opinion Mining for Arabic Dialects on Twitter

DoniaGamal, Marco Alfonse, El-Sayed M. El-Horbaty, Abdel-BadeehM.Salem

Computer Science Department, Faculty of computer and information sciences,
Ain Shams University, Cairo, Egypt

donia.gamaleldin@cis.asu.edu.eg, marco@fcis.asu.edu.eg,
shorbaty@cis.asu.edu.eg, absalem@cis.asu.edu.eg

Abstract

Opinion Mining (OM) has lately become one of the increasing areas of research identified with text mining and natural language processing. This domain is utilized to detect and extract the sentiment out of text giving valuable and beneficial information related to the author and his/her tendency for a precise topic. The fundamental task is to classify that extracted text which could be a tweet, review, blog, comment, news,etc. to a positive, negative, or neutral sentiment. Most of the instant investigations identified with this topic focus essentially on English textswith a limited and finite assets and resources accessible for miscellaneous languages like Arabic, and its different dialects like the Egyptian dialect, Gulf dialect and so on. This research focus on Arabic Dialects Opinion Mining (ADOM),different Machine Learning (ML) algorithms are applied and the experimental results showthat the Support Vector Machine (SVM) classifier gives the highestand most efficient accuracyof**93.56%** compared to other applied classifiers. Moreover, this accuracy exceeds the other Arabic related work which makes it very promising and encourages to continuein this line of researchutilizing a normalized dataset with two polarities.

Keywords: *Digital Arabic Language Preprocessing; Arabic Dialect Opinion Mining; Sentiment Classification; Twitter; Sentiment Analysis; Machine Learning, Applied Informatics.*

1. Introduction

OM, also known as Sentiment Analysis (SA), research goes back to 2003 [1].SA can be divided into several tasks which include, Sentiment Classification (SC), Sentiment Summarization (SS), Sentiment Lexicon Generation (SLG), Sentiment Quantification (SQ), Opinion Extraction (OE), Feature-Based Summary (FBS), and Opinion Spam (OS). A large portion of the SAresearchis concerned withSC, which intends to decide whether the users' opinion and attitudeare positive, neutral or negative[2].Two fundamental approaches are utilizedforthe automatic OM task. The first approach utilizes lexicons called sentiment lexicons or polarity lexicons and the second uses ML algorithms. The performance of the first approachreliesupon the lexicon scope and quality while the latter needs rich annotated datasets. However, sentiment assets and available resourcesare unequal andunbalanced in varied languages. The sentiment lexicon, opinion vocabulary or labeled data are wealthy in few languages like English and are poor in others.

Today there is a growing interest in Arabic OMforArab internet users as they are increasingly utilizing Social Media (SM) platforms.Online blogs are daily logs for their creators that include data about a specific topic that their authors areconcerned about. Generally,they utilize it to express their personal review and opinion about items, products, political view or other interests they have [3]. Twitter is one of the tremendous andgreatest online platforms that arestacked with sentiment. It is a micro-blogging site which

holdstweets.Twitter has a length constraint whichis140 characters or less per tweet. There are over 1 billion Tweets generated every 72 hours with more than 140 million online users [4]. According to [5], the total number of active users per month on Twitter in the Arab region is 11.1 million in March 2017. The Middle Eastgenerates over 27.4 million tweets dailycompared to 17.2 million tweets per day in the last two years.

Arabic Language has three distinct dialects, which are QuranicArabic (QA) also known as Classical Arabic, Modern Standard Arabic (MSA), and Colloquial Arabic. QA is the type of Arabic in which the Quran (the holy book of Islam) is written. In the sixth century A.D., the language wasmarginally not the same as the Arabic of today. MSA is the most broadly utilized version of Arabic today in Arabic speaking nations [6]. MSA is utilized as a portion of every media outlet from television to films, to daily newspapers and radio broadcasts. Thevast majority books are written in MSA in addition topoliticians'opinions in debates alongside speeches. MSA is the Arabic that is utilized in everyday life in Arabic speaking countries. Colloquial Arabic is frequently the spoken language of the most Arabs. This type of Arabic is subject to regional varieties that not only exists across nations, butalso occur in the same nation [7].

This paperfocuses on MSA and Colloquial Arabic, whichare mostly spoken in the Arab region and are used in written formsof plays, music and books. With its prominence, Egyptian Colloquial Arabic is the most famous and popular Arabic dialect.

The rest of this paper is organized as follows: Therelated work performed on Arabic datasets using the ML approachis presented in Section 2. The proposed approach of ADOM on Twitter is illustrated in detail in Section 3. Section 4 demonstrates and discusses the evaluation of the experimental results. Finally, conclusions and future workguidelines are discussed in Section 5.

2. Related Work

A large number of researches were proposed to analyze and evaluate the sentiment and obtain the opinion from the World Wide Web SM networks (Facebook, Twitter, etc.).In this section,the main focus is on theArabic OM researches.

Rushdi et al. [8] applied the SVM and Naïve Bayes(NB) classifiers on a dataset that comprises of 500 reviews of movies written inArabic. They appliedTerm Frequency- Inverse Document Frequency (TF-IDF) to weight the tokens of the opinions and they likewise stem the tokens. The SVM achieves accuracy of 90% and the accuracy of the NB classifier is 84%.

Omar et al. [9] made a comparative study on the adequacy of individual supervised MLclassifiers and ensemble algorithms for SAof Arabic Customers' Reviews. The most common text classification algorithms utilized as a base-classifiers are NB, SVM, andRocchio classifier. Theindividual classifiers'results showedthat the BernoulliNB(BNB) and SVM algorithms performed better thanvariantML algorithms.However, the results of the ensemble classifiers approach showed that it carried out robustly better than all the various individual classifiers.

Duwairi and Qarqaz[10]dealt with SA in Arabic reviews and opinions fromML perspective. They applied three different supervised classifiers on a developed datasetof tweets/comments which are NB,SVM, and K-Nearest Neighbor(KNN). The results demonstrate that SVM provides thehighest precision as well the KNN with specifying K as ten provides the highestRecall.

Shoukry and Rafea [11] illustrated the impact of the preprocessing over 1000 tweets (positive and negative) written in Egyptian dialect from Twitter to improve the accuracy.They applied the SVM classifier with two different stemmers and getalmost 76% accuracy.

Soliman et al. [12] constructed a Slang Sentimental Words, Idioms, and phrases Lexicon (SSWIL) of opinion words. In addition, they classified a collected data of Arabic news' comments which were shared on Facebook using a Gaussian kernel SVM classifier. To test the performance of the proposed classifier, several Facebook news' comments were used, where 86.86% accuracy rate was obtained with precision 88.63% and recall 78%.

Badaroet al. [13] proposed a light mobile application using lexicon-based computing for SA of Arabic tweets. The proposed technique classifies the tweet into multi-classes which are positive, negative, objective or neutral utilizingDecision Tree (DT) as the ML classifier. Experiments were performedutilizing a corpus of manually annotated 2300 Arabic Tweets and the accuracy of 67.3% has been achieved.

Aldayeland Azmi[14] endorsed a solution for the problem of tweeting in Arabic utilizing the Saudi Arabia Dialect as a basis. They recommended a hybrid approach that merges semantic orientation and ML algorithms. Through this technique, the lexical-based approach labels the training dataset. The output labeled data was utilized as training data for the SVM classifier. The experiments demonstrated that their hybrid approach progressed the F-measure of the lexical classifier by 5.76%, achieving a general accuracy of 84%.

Altawaierand Tiun[15], investigated the ML algorithms in terms of Arabic SA on Twitter. Three different algorithmswereapplied, including NB, DT, and SVM. The experimental results have shown that DT has outperformed the other techniques obtaining 78% f-measure.

Table 1 presents our comparison between different ML algorithms that are stated in this section.

Table 1: Evaluation of Arabic Sentiment Analysis

Authors	ML Classifiers	Feature Extraction	Evaluation Metric	
Rushdi et al. [7]	NB	TF-IDF	Accuracy	90.60%
	SVM			84%
Omar et al. [8]	MNB	TF-IDF	F-Macro	94.59%
	BNB			96.51%
	Rocchio with Jaccard			90.11%
	Rocchio with Cosine			92.59%
	SVM			94.61%
Duwairi and Qarqaz [9]	NB	N-gram	Macro Precision	66.21%
	SVM			75.25%
	KNN			70.97%
Shoukry and Rafea [10]	SVM	Unigram	Accuracy	79.2%
		Unigram + Bigram		80.5%

Table 1: (continued) Evaluation of Arabic Sentiment Analysis

Authors	ML Classifiers	Feature Extraction	Evaluation Metric	
		Unigram + Bigram + Trigram		80.6%
Soliman et al. [11]	SVM	N/A	Accuracy	86.86%
Badaro et al. [12]	DT	POS	Accuracy	67.3%
Aldayel and Azmi [13]	Hybrid (SVM + Lexicon)	Unigram	Accuracy	83.80%
		Bigram		84.16%
		Trigram		84.07%
Altawaier and Tiun [14]	NB	TF-IDF	Accuracy	75%
	DT			78.9%
	SVM			44%

It can benoticed from Table 1 that the most common implemented algorithms are NB and SVM. The algorithm, thatestimated to be the best,uses TF-IDF with NBwhich gives a high accuracyof 90.60%.

3. The Proposed Method of Arabic Dialects Opinion Mining

In this section, our proposed methodof ADOM is illustrated in depth. The steps performed to collect and prepare the dataset of different Arabic dialects Tweetsare clarified and the tools and techniques utilized in the research are stated. The research methodologyhas fundamental stages as shown in Fig1, whichare gathering the Arabic Dialect tweets dataset, preprocessing Tweets and annotations, feature extraction, applying and comparing different supervised ML classification methods, and demonstrating their results. These phases will be explained indetail in the following subsections.

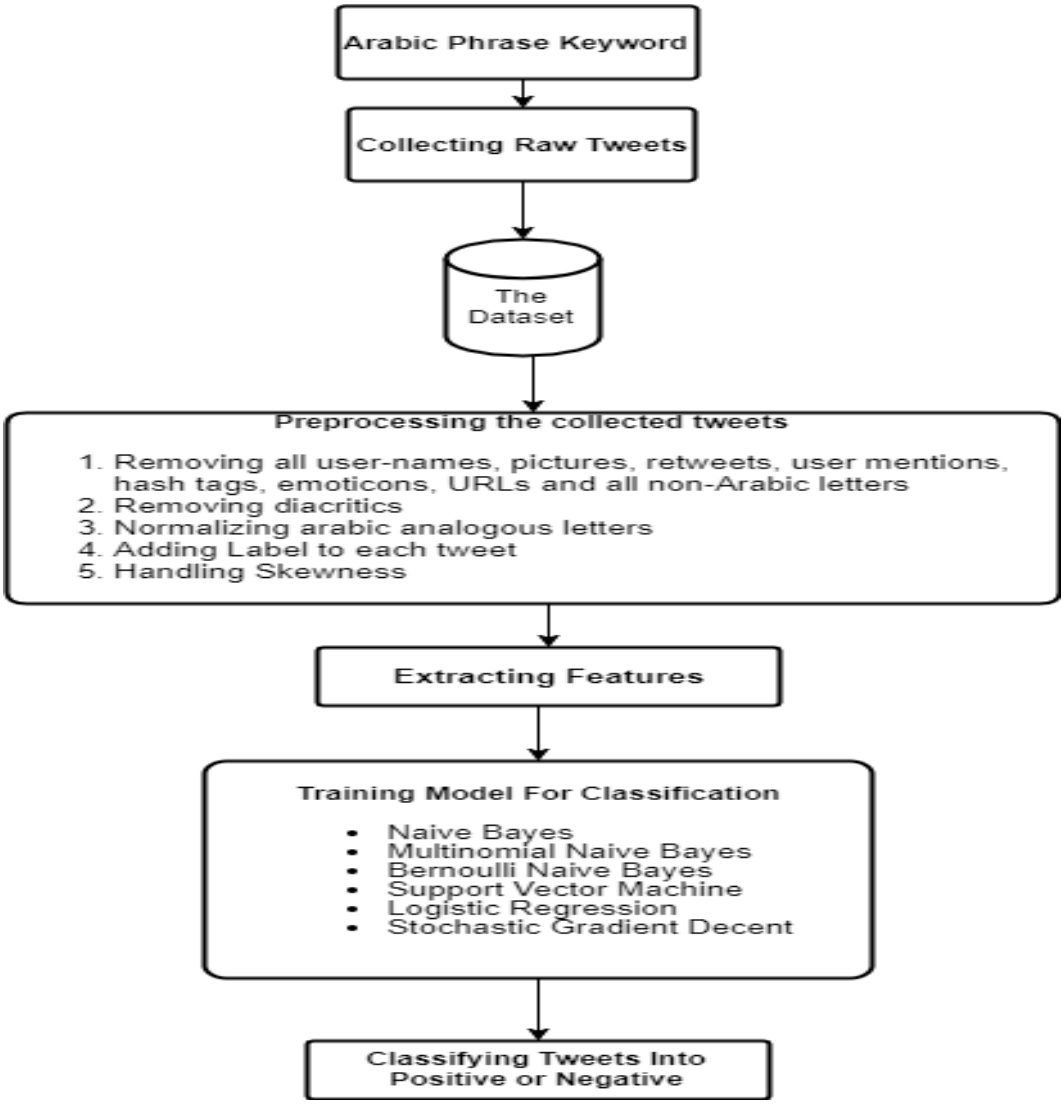

Figure 1:TheADOM Proposed Method

3.1Collecting Arabic Dialects Tweets Dataset

This step comprises of extracting tweets of7 days due to Tweepy's [16] constraint for pulling tweets. Approximately 151,500tweets, published by Arabic users, were collected. Different Arabic phrases are used as keywords for searching and collecting these tweets[17]. The tweets includemany opinions about different topics, which are expressed and written in numerous ways by individuals.

3.2Preprocessing Tweets and Annotations

The polarity in the raw data is extremely susceptible to irregularity and redundancy. The quality of the data influences the results and therefore to enhance the quality, the raw tweets are preprocessed to removeall noise from the collected tweets and improve the efficiency of the data (see Fig.1).Then comes the process of removing all user-names, profile pictures, retweets, user mentions, hash tags, emoticons, URLs and all non-Arabic letters from the

tweets to be easily manipulated and dealt with. Then the data was labeled into positive and negative classes automatically as shown in fig 2.Finally, all tweets are annotated consisting of 75,774 positive and 75,774 negative tweets to be the used dataset.

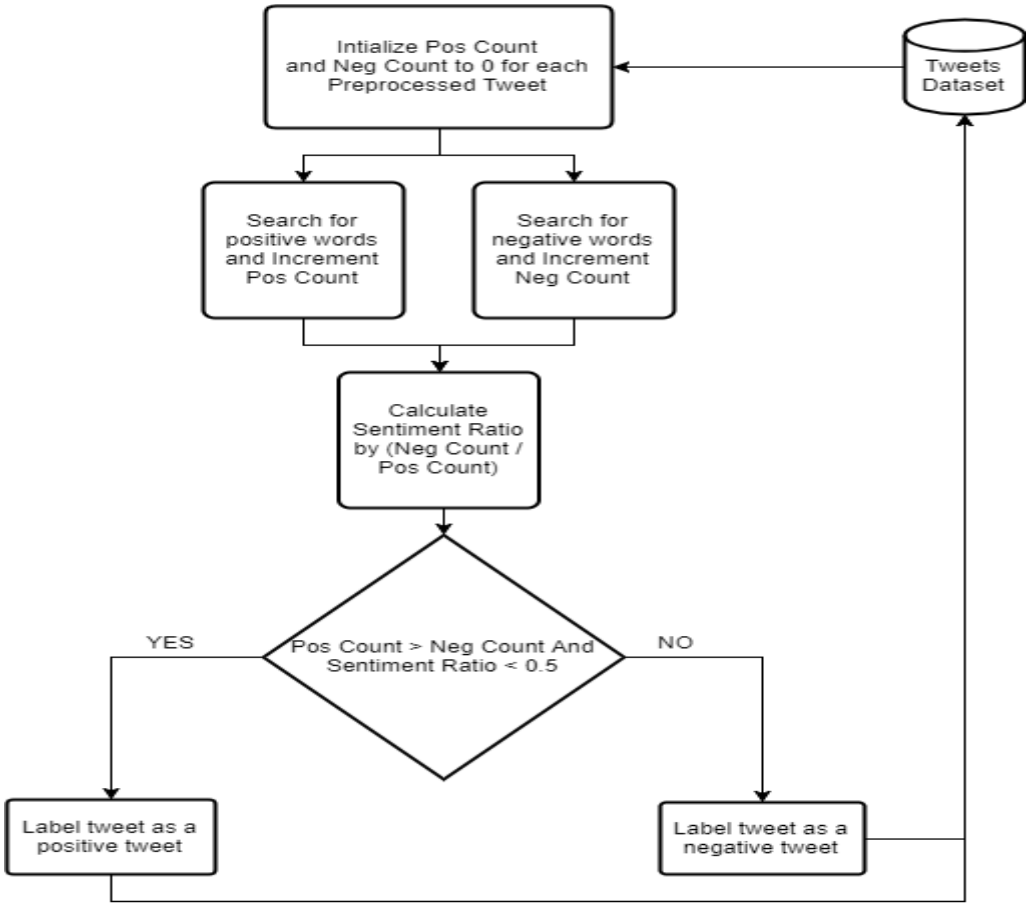

Figure 2: Labeling Tweets

3.3 Feature Extraction

The Term Frequency (TF) is used to extractthe feature set.TF is found essentially monitoring the frequency that a given phrase/expressionappearsin a given text [18].These features are individualwords and their weights are calculated to indicate the relative influence of features.

3.4 Supervised Classification

Supervised MLintends to train the data on certain pattern that one may be able to identify and distinguish it in the test part. This is a valuable and suitable method in the field of OM by training the dataabout a pattern that may indicate for ifthosesentimentsare positive or negative. In this study, six classification algorithms have been chosenincluding NB, SVM, BNB, Multinomial NB (MNB), Stochastic Gradient Decent(SGD) and Logistic Regression (LR) [19, 20, and 21].The reason behind using such algorithms lies on their effective ability to deal with text categorization where the number of features is huge [22]. The experiments have beenconducted using Python 3.6libraries. Python programming language is considered one of the most powerful languages since it is completely open source and made for integration with

external tools on cross-platforms. Moreover, ahuge numbers of python libraries are built for complex tasks in Artificial Intelligence (AI) and ML domains. Two of the widely used libraries are Tensor-flow [23]which is high-level neural network library, and Scikit-learn [24] which is used for data mining, data analysis and machine learning, etc.The two Python libraries used for the experiment are Scikit-learn and Natural Language Tool Kit (NLTK) [25]. These libraries are free that grants researchers the ability to employ ML algorithms.

4. Results and Discussion

Results can be evaluated using different methods [26]. One of the most popular method is the accuracy. The evaluation of theproposed classifiersis performed on a set of real auto-annotated Twitter posts using the same evaluation methods described in [27] (accuracy).The accuracy of the classifiers is computedaccording to Eq.1:

$$\boldsymbol{Accuracy} = \frac{\text{Number of correctly classified Tweets}}{\text{Number of all tweets}} \qquad (1)$$

The performances of the six MLalgorithms used in this research are comparedby choosing 75% of the dataset as training data and the remaining 25% as testing data.Figure 3 demonstrates these accuracies for the various classifiers under consideration when performed on the balanced dataset. The accuracies range between 85% and 94%, with BNB has the lowest accuracy while theSVMhasthe highest one. The apparent difference betweenNB and MNB is noticed also,the MNB is a more optimized modelthan NB as soon as it relates to text classification issues.

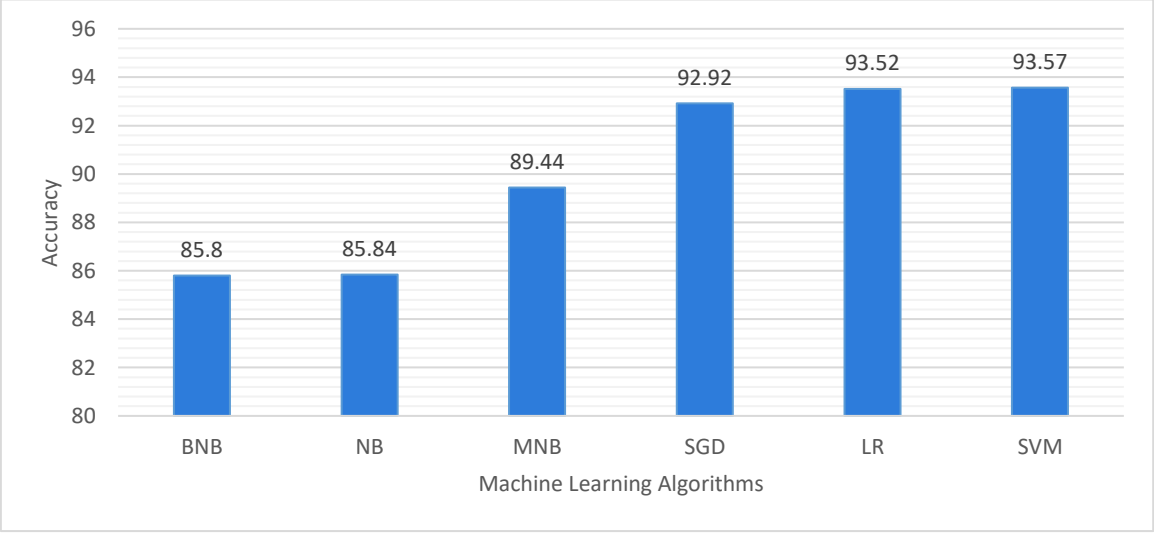

Figure 3:The Accuracies of the ML Algorithms Applied on the Balanced Arabic Twitter Dataset

From the above results, it is clear that SVM classifier outperforms the other text classification models with an accuracy above 93.5%. Fig. 4 shows a comprehensive view of the accuracies of the six supervised learning algorithms with different number of features. It shows the accuracy curves of different classification algorithms with different range of features.

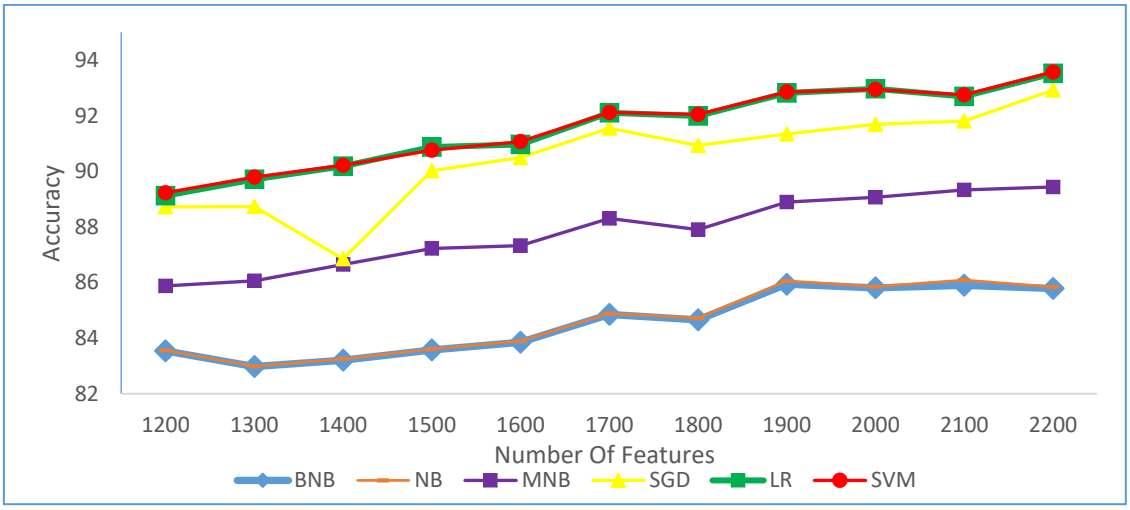

Figure 4: The Impact of Number of Features on the Overall Accuracy of MLAlgorithms

The accuraciesvary from**82**% to **94**% with an interval of two on accuracy axis and from 1200 to 2200 with an interval of one hundred on number of features axis. From figure 4,it is noticed that NB and BNB get almost similar results, the same for SVM and LR. MNB performs constantly better than NB and BNB.

5. Conclusions and Future Work

The content of SM networking such as Twitter is viewed as a significant SA resource, as usersutilize such media to generously express their views and opinions on an extensive of mixed topics. Twitter is one of the most prevalent SM networks in the Arab nations. Focusing on ADOM as a use-case, this paper documents an in-depth approach putting into consideration the challenges and difficulties facing the research community in classifying sentiments in Arabic especially in Egyptian colloquial dialect. Clarifying the data creation procedure and providing the results of different ML algorithms in this paper, researchers can be helped to better realize the nature of the dataset and to more successfully and effectively utilize this dataset in their future research work. We utilized supervised ML algorithms, namely SVM, NB, SGD, MNB, BNB and LR algorithms to determine the tweets' sentiments. The experimental resultsdemonstrate that the SVM classifier has the highest accuracyof **93.57%.**

As the future work, a multilingual Twitter content is going to be assembledand usedto construct a multilingual sentiment classifierto enhance the accuracy of SA.Negation handling, and detection of sarcasms are aimed to be considered.

References

[1]. Cambria, Erik, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. "New avenues in opinion mining and sentiment analysis."International Journal of IEEE Intelligent Systems 28, no. 2, pp: 15-21, 2015.

[2]. Gamal, Donia, Marco Alfonse, El-Sayed M. El-Horbaty, and Abdel-Badeeh M. Salem. "A comparative study on opinion mining algorithms of social media statuses.", Proceedings of Intelligent Computing and Information Systems (ICICIS), 2017 Eighth International Conference on, pp. 385-390. IEEE, 2017.

[3]. Godwin-Jones, Robert. "Blogs and wikis: Environments for online collaboration",International Journal of Language Learning & Technology, Vol 7, No. 2, pp. 12-16 ,2003.

[4]. Gantz, John, and David Reinsel. "Extracting value from chaos." IDC iview 1142, pp.1-12, 2011.

[5]. http://www.arabsocialmediareport.com/[accessed July 2018]

[6]. Al-Gedawy, Madeeh N. "Detecting Egyptian Dialect Microblogs using a Boosted PSO-based Fuzzier", International Journal of Egyptian Computer Science Journal, Vol.39, No. 1, 2015.

[7]. https://www.arabacademy.com[accessed July 2018]

[8]. Rushdi-Saleh, Mohammed, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and José M. Perea-Ortega. "OCA: Opinion corpus for Arabic." International Journal of the American Society for Information Science and Technology 62, no. 10, pp. 2045-2054, 2011.

[9]. Omar, Nazlia, Mohammed Albared, Adel Qasem Al-Shabi, and Tareq Al-Moslmi. "Ensemble of classification algorithms for subjectivity and sentiment analysis of Arabic customers' reviews." International Journal of Advancements in Computing Technology 5, no. 14:77, 2013.

[10]. Duwairi, Rehab M., and Islam Qarqaz. "Arabic sentiment analysis using supervised classification." Proceedings ofInternational Conference of Future Internet of Things and Cloud (FiCloud), pp. 579-583. IEEE, 2014.

[11]. ShoukryAmira, and Ahmed Rafea. "Preprocessing Egyptian dialect tweets for sentiment mining." In The Fourth Workshop on Computational Approaches to Arabic Script-based Languages, pp. 47- 56, 2012.

[12]. SolimanTaysir Hassan, M. A. Elmasry, A. Hedar, and M. M. Doss. "Sentiment analysis of Arabic slang comments on Facebook." International Journal of Computers & Technology 12, no. 5:3470-3478, 2014.

[13]. Badaro Gilbert, RamyBaly, RanaAkel, Linda Fayad, Jeffrey Khairallah, Hazem Hajj, KhaledShaban, and Wassim El-Hajj. "A light lexicon-based mobile application for sentiment mining of Arabic tweets." In Proceedings of the Second Workshop on Arabic Natural Language Processing, pp. 18-25. 2015.

[14]. Aldayel Haifa K., and Aqil M. Azmi. "Arabic tweets sentiment analysis–a hybrid scheme." Journal of Information Science 42, no. 6: 782-797, 2016.

[15]. AltawaierMerfatM., and Sabrina Tiun. "Comparison of machine learning approaches on Arabic twitter sentiment analysis." International Journal on Advanced Science, Engineering and Information Technology 6, no. 6: 1067-1073, 2016.

[16]. Roesslein, Joshua. "tweepy Documentation." [Online] http://tweepy. readthedocs. io/en/v3 5 ,2009.

[17]. AlyMohamed, and Amir Atiya. Labr: A large scale arabic book reviews dataset. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, vol. 2, pp. 494-498, 2013.

[18]. Joshi NehaS., and Suhasini A. Itkat. "A survey on feature level sentiment analysis." International Journal of Computer Science and Information Technologies 5, no. 4: 5422-5425, 2014.

[19]. AggarwalCharu C., and ChengXiangZhai, eds. mining text data. Springer Science and Business Media, 2012.

[20]. Srivastava Ashok N., and MehranSahami. Text mining: Classification, clustering, and applications. Chapman and Hall/CRC, 2009.

[21]. Kao Anne, and Steve R. Poteet, eds. Natural language processing and text mining. Springer Science and Business Media, 2007.

[22]. Joachims T. "Text categorization with support vector machines: Learning with many relevant features". Machine Learning: ECML-98, Lecture Notes in Computer Science, 1398, pp 137-142, 2005.

[23]. https://www.tensorflow.org/[accessed July 2018]

[24]. http://scikit-learn.org/[accessed July 2018]

[25]. https://www.nltk.org/[accessed July 2018]

[26]. El-Halees, Alaa. "A Comparative Study On Arabic Text Classification", International Journal of Egyptian Computer Science, Vol. 30, No. 2, 2008

[27]. GautamGeetika, and DivakarYadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Proceedings of seventh International Conference on Contemporary computing (IC3), pp. 437-442. IEEE, 2014.