Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
December 2022, vol. 27, no. 2, pp. 119–127
https://doi.org/10.2478/acss-2022-0013
https://content.sciendo.com

Aspect-based Sentiment Analysis and Location Detection for Arabic Language Tweets

Norah AlShammari[1], Amal AlMansour[2*]
[1,2]*King Abdulaziz University, Faculty of Computing and Information Technology, Jeddah, Saudi Arabia*

***Abstract* – The research examines the accuracy of current solution models for the Arabic text sentiment classification, including traditional machine learning and deep learning algorithms. The main aim is to detect the opinion and emotion expressed in Telecom companies' customers tweets. Three supervised machine learning algorithms, Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), and one deep learning algorithm, Convolutional Neural Network (CNN) were applied to classify the sentiment of 1098 unique Arabic textual tweets. The research results show that deep learning CNN using Word Embedding achieved higher performance in terms of accuracy with F1 score = 0.81. Furthermore, in the aspect classification task, the results reveal that applying Part of Speech (POS) features with deep learning CNN algorithm was efficient and reached 75 % accuracy using a dataset consisting of 1277 tweets. Additionally, in this study, we added an additional task of extracting the geographical location information from the tweet content. The location detection model achieved the following precision values: 0.6 and 0.89 for both Point of Interest (POI) and city (CIT).**

***Keywords* – Aspect detection, Arabic language, deep learning, location detection, sentiment analysis.**

I. Introduction

With the flow of data on social networking sites, such as Twitter, we often hear about measuring public opinion and note the great interest of researchers in this field. Sentiment analysis is the process of analysing opinions and emotions to identify the tendencies of collected data, and classifying them into positive, negative or sometimes neutral. Sentiment analysis is exceptionally important because it helps organisations and businesses measure the extent to which their customers are satisfied with a particular product and service in a way faster, more comprehensive and more reliable than traditional questionnaires. In this research paper, we propose a new framework for enhancing a social medial sentiment analysis especially in Twitter by considering the aspect and geographical location related to tweet. We studied the aspect-based sentiment analysis and added an extra step of extracting the geographical location based on the tweet content. Detecting tweet aspects and spatial location will help analyse users' reactions towards specific service aspects for each region, city or neighbourhood. This will help companies and government organisations to realise the type of service and its location and if it needs to be mended. Multiple research projects [1]–[3] have studied aspect-based sentiment classification models for the Arabic language in social network content. According to our best knowledge, none of those systems used Natural Language Processing (NLP) practices such as Part of Speech (POS) tagging and Named Entity Recognition (NER) in their proposed models. Sentiment analysis uses various NLP practices and techniques to verify whether textual data are positive, negative or neutral.

As most of existing studies of aspect-based sentiment classification models for Arabic social network content do not make use of NLP methods and tools, this study focuses on addressing this gap by proposing an aspect-based sentiment classification model that utilizes most of useful NLP practices in sentiment analysis. In addition, the proposed model identifies the geolocation of given tweets in order to analyse users' feelings by location and identify the service locations, which need more consideration. The main contributions of the research can be listed as follows:

- To develop an aspect-based sentiment classification model for Arabic text tweets using unigram features and bigram features, and compare their accuracy performance.
- To compare the aspect-based sentiment analysis accuracy of traditional machine learning algorithms with deep learning algorithms.
- To study the effect of NLP techniques such as POS tagging and Word Embeddings with deep learning algorithms.
- To identify the geolocation of Arabic text tweets, which is important especially in a large, extensive geographical area, spanning several regions such as Saudi Arabia using NLP techniques such as NER.

This paper describes the main research methods and results achieved during the project. The rest of the paper is organised as follows: Section II presents the used NLP techniques. Section III reports the related research studies. Section IV explains the experimental steps, and the used dataset. Section IV presents results and discussion. Finally, concluding remarks and future research areas are provided in Section V.

* Corresponding author's e-mail: aalmansour@kau.edu.sa

©2022 Norah AlShammari, Amal AlMansour.
This is an open access article licensed under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0).

II. Used NLP Techniques

The most basic and useful techniques in NLP are explained below:

- *Part of Speech* (POS) refers to assigning each word in the given sentence to the correct part of speech tag such as noun, verb, and adjective. POS for the most natural language is divided into eight sections, as shown in Fig. 1.
- *Word Embedding* is a type of representation that captures words with similar semantics; words with similar vectors are semantically similar. It may be one of the major advances of deep learning methods on handling NLP challenges.
- *Tokenization* is one of the most crucial pre-processing steps in NLP. Tokenization can be done using one-word sequence, which is called unigram, or into a two-word sequence, which is called bigram, or into a three-word sequence, which is called trigram.
- *Named Entity Recognition* (NER) is a prime component of NLP systems for retrieving information and extracting the relationship, etc. Numerous studies on the development of NER systems have been widespread for decades, but precise systems that use Deep Neural Network (DNN) have only been introduced in the past few years [4]. Identifying the named entities is the task of identifying entities such as the person as PER and the location as LOC and the institution as ORG. NER systems are typically used as the first stage for different NLP applications such as question answering, information retrieval, cross-reference solution, topic modelling, etc. Therefore, it is important to highlight recent developments in the identification of the named entity, in particular NER neural structures that have achieved the latest performance [4]. The location detection through NER is one of the most famous studies that has achieved high accuracy. In our case problem, we intended to identify the location in the text which is one of the most important pre-defined categories when recognising the objects, in addition to names of persons and organisations, medical symbols, etc.

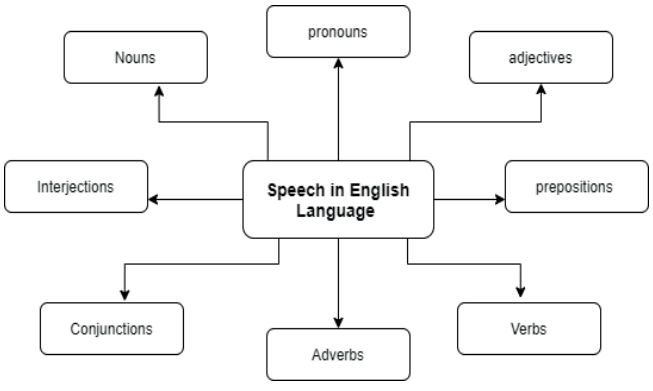


Fig. 1. Sections of speech in the English language.

III. Related Work

Machine Learning (ML) is one of the most widely used techniques for classification and prediction problems due to its higher classification accuracy and less computational time. Several studies propose the use of machine learning algorithms for sentiment classification [5]–[9], whereas other studies incorporate dictionaries, which have a list of positive and negative sentimental words to quantify the text sentiment orientation [10]–[14].

Deep Learning (DL) has recently achieved very promising results on tasks such as language understanding and image analysis, and we expect it might improve the performance in the field of sentiment classification, which is a big part of language understanding [15]. Research studies that propose the use of deep learning include [16]–[20]. Table I summarises the reviewed papers for sentiment analysis, including the highest accuracy achieved using the proposed methods.

TABLE I
Reviewed Research Studies for Sentiment Analysis

Reviewed paper	Methodology: Machine Learning (ML)/ Lexicons (Lex)/ Deep Learning (DL)	Dataset type: Tweets/ Reviews/ Other (article, stories, etc.)	Language: English/ Arabic	Highest accuracy
[5]	ML	Tweets	English	83 %
[6]	ML	Tweets	English	Not Available
[7]	ML	Tweets	Arabic	72 %
[8]	ML	Tweets	English	84.6 %
[9]	ML	Tweets	English	95.96 %
[10]	Lex	Tweets	English	72.6 %
[11]	ML, Lex	Tweets	English	71.42 %
[12]	Lex	Reviews	English	Not Available
[13]	Lex	Tweets	English	NA
[14]	Lex	Reviews	English	84.8 %
[16]	ML, DL	Other	Arabic	74.3 %
[17]	DL	Other	English	Not Available
[18]	ML, DL	Other	English	78.4 %
[19]	ML, DL	Tweets	English	80.1 %
[20]	DL	Reviews, Other	English	Not Available

IV. Methods

In this study, we detected the orientation of Arabic tweets using both traditional machine learning and deep learning approaches. For our project, only unique tweets were extracted and analysed using the functions of the Twitter API. Almost 6182 unique Arabic microblogs were collected during the period (9–19 August 2019) to identify the sentiment of customers' tweets. The primary mechanism for extracting the Arabic tweets was through certain keyword queries that mention the companies' twitter accounts. In this way, users'

thoughts or feelings through their tweets about three telecommunication companies (STC, MOBILY and ZAIN) in Saudi Arabia were collected. The collected dataset details are shown in Table II.

TABLE II
USED DATASET

Saudi telecommunication companies	Number of tweets	Number of users	Average of characters per tweet
STC	2891	2098	77.2
Mobily	1827	1127	75.2
Zain	1464	960	74.5
Total	**6182**	**4185**	**75.7**

The main processes of the proposed aspect-based sentiment classification and location detection for Arabic can be subdivided into three main phases. The objective of the first phase is to classify the text sentiments of tweets into positive, negative and neutral classes using machine learning and deep learning algorithms. The second phase is concerned with identifying the aspects of given users' tweets. Finally, in the third phase, the proposed system detects the location named entities from the textual tweets. The following subsections explain each phase in detail.

A. Sentiment Classification Phase

In the sentiment classification phase, we performed the following three main processes: pre-processing, labelling, and classification. Different machine learning, Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), and deep learning Convolutional Neural Network (CNN) algorithms were used for Arabic sentiment classification task. SVM, LR, RF, and CNN classifiers work well for text classification; hence, they are widely used in the field of sentimental analysis. Moreover, different text features and representations have been utilised to evaluate the sentiment classification and provide the highest potential performance. Firstly, both unigram and bigram features have been used by machine learning classifiers SVM, LR, and RF to detect the Arabic tweet sentiments. Then, deep learning algorithm CNN has been used with POS and Word Embedding technique. The number of annotated tweets is shown in Table III with a sample of tweets in Arabic for each sentiment class.

TABLE III
SENTIMENT-LABELLED TWEETS

Sentiment class	Number of tweets	Sample tweets in Arabic	Sample tweets translated into English
Neutral	552	للاسف واجهنا ضعف النت في منطقة الطائف وسوء تغطيه	Unfortunately, we encountered poor internet coverage in the Taif region
Negative	463	تم الارسال من جديد	Has been sent again
Positive	81	كفووو والله أرسلتها	Wow, I sent it
Total	**1096**		

B. Aspect Detection Phase

In this phase, we applied the same classifiers and NLP technologies that had been applied in sentiment classification phase. However, in this phase we investigated the performance of used classifiers in detecting the aspect rather than sentiment of tweet. The only difference from the previous sentiment analysis steps is the data labelling step. We set the topics as shown below in Table IV (Customer Service, Network, Internet, Packages, Billing, and General) for 1277 tweets manually. The number of annotated tweets is shown in Table IV with a sample of tweets in Arabic for each aspect class.

TABLE IV
ASPECT-LABELLED TWEETS

Aspect class	Number of tweets	Sample tweets in Arabic	Sample tweets translated into English
Customer Services (CS)	577	اتوقع السؤال واضح ومايحتاج ارسل لكم خاص	I expect the question is clear and does not need to be sent to you in private
Network	109	م فييه شبكه شبكه مااافيه كرهتتتت امسك جوالي	There is no network coverage, there is no network coverage, I hate to hold my phone
Internet	228	لوسمحت النت بالبيت مايشتغل وكل شهر على هالحال	Excuse me, the internet is not working at home and this is the case every month
Packages	113	ممكن اعرف كم هذا الباقه شهور	May I know how many months this package is
Billing	136	انا اسال بشكل عام ابي اعرف سعر الدقيقه في شرائ	I ask in general, I want to know the price per minute in sim….
General	114	الغاء الرقم نهائي	Delete the number permanently
Total	**1111**		

C. Location Detection Phase

NER is an important NLP task, which aims at building a machine learning model to extract the named entity from the given text. There are several NER models for the Arabic language text such as FARASA [21] and Polyglot-NER [22]. However, these NER models were trained on formal Arabic text and underperformed when testing on informal Arabic text [23]. For that reason, we applied a deep learning algorithm to build location NER that could extract location from text such as social media data. We utilized SpaCy library (spacy.io) to build location NER that could extract location from the Arabic text. We defined two location types, which are: City (CIT) and Point of Interest (POI). For the training purpose, we collected data from AQAR application. The reason is that the collected tweet dataset did not include enough location NE for training as the data from AQAR application. AQAR (https://sa.aqar.fm/ [24]) textual data have a lot of location NEs and share the similarity characteristics with Twitter data as it was mainly written using informal Arabic text. However, for the testing dataset, we utilised the data collected from Twitter PAI. The NER model was built using CNN algorithm, and the model was evaluated using precision, recall, and F1 score measures.

Annotation step was applied for three types of entity, which are:

1. Preposition (PRE): Preposition entity refers to terms that are related to location entities and came before location NE, those PRE entities will help the model to predict location NE.
2. Point of Interest (POI): Point of interest location refers to any area smaller than cities, such as residential areas, parks, malls, and centres.
3. City (CIT): This will cover all location NE refers to cities.

This study used two location NE labels, POI and CIT because of the data characteristics. Most of the collected tweets contain two or more location NE, one for the residential area, which is considered as POI label, and the second refers to the city, which is considered as CIT label. Fig. 2 shows a sample of data annotation conducted on DataTurks. The training dataset contains 798 posts collected from AQAR application by applying Beautiful-Soup python library for web scraping. However, the testing set contains 100 tweets as shown in Fig. 3, those tweets are collected from the original dataset collected in the sentiment analysis phase.



Fig. 2. Sample of labelled textual data from AQAR dataset.

Fig. 3 shows an example of the testing set labelling, and we applied the same labelling conditions used with the training set. The testing dataset represents our research problem, which is collected from Twitter and related to Saudi telecom companies' tweets.

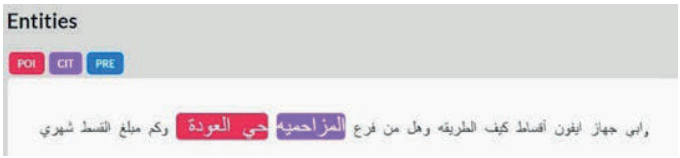


Fig. 3. Example of the testing dataset from STC dataset.

In this experiment, we utilised SpaCy library to build our location NE model. SpaCy library contains several tools for NLP, and it is designed as a modular pipeline. It contains several models to solve several NLP tasks, such as POS tagging, NER, and text classification. The NER pipeline in SpaCy is a deep learning model built by utilising CNN and Long Short-Term Memory (LSTM) architecture [25]. In this experiment, we trained the deep learning model by blank Arabic language class model, which did not contain learned entities by using annotated training dataset to extract three types of entities. In this experiment, we utilised the default execution set by SpaCy and no additional layers were added. The SpaCy implemented model applied Word Embeddings. We utilised AQAR data as training data which were split by the model to training and validation sets. We applied 50 training iterations to train and validate our model. After that we tested the performance of our model by utilising the annotated data collected from Twitter API.

V. Results and Discussion

The section shows the result of the proposed framework for the three main phases: the sentiment analysis of Arabic tweets, the aspect detection, and the location detection. Different NLP techniques and classifiers were used to report the performance results for each phase. There is limited dataset benchmark in the Arabic text especially when the study related to social media. In this research, we collected our own data from Twitter that related to Saudi telecom companies, which represented the real world dataset. As a result, it not logical to compare our results with other studies as we worked on different datasets. Therefore, the best approach to evaluate our results is to compare them with the baseline model results for each experiment phase.

A. Sentiment Classification Results

Sentiment analysis experiments were performed using a total of 220 tweets in the testing dataset that should be classified against three sentiment classes. The baseline model accuracy for sentiment analysis task was calculated by dividing the number of instances (111 tweets) into the majority class (neutral class) by the total number of instances in the testing set (220 tweets). As a result, the accuracy of baseline model was 50 %. In this experiment, we applied a set of hyper parameters for the machine learning classifiers. We utilised the default hyper parameters set by Sklearn libraries with some changes, such as n_estimators=100 for RF and gamma= 'scale' for SVM. On the other hand, to build the deep learning model, we must set the layers and activation functions. In this experiment, we applied Sequential() to build our classification model with two layers. (1) The input layer is Dense type with relu as an activation function, this layer generates 42400 parameters. (2) The output layer is also Dense type with sigmoid as an activation function, this layer generates only 66 parameters. Table V shows the obtained results for accuracy, precision, recall and F1 measures using different classifiers and language models.

TABLE V
Sentiment Classification Results

Used Classifier/ Feature		Accuracy	Precision	Recall	F1
Machine Learning (Unigram)	SVM	0.74	0.77	0.74	0.73
	RF	0.75	0.78	0.75	0.74
	LR	**0.78**	**0.80**	**0.78**	**0.77**
Machine Learning (Bigram)	SVM	0.51	0.75	0.51	0.36
	RF	0.54	0.71	0.54	0.42
	LR	0.53	0.68	0.53	0.39
Deep Learning	Neural Networks	0.75	0.75	0.75	0.74
	With Word Embeddings	**0.81**	**0.82**	**0.81**	**0.81**
	With POS	0.75	0.81	0.75	0.74

As shown in Figs. 4 and 5, all applied experiments outperformed the baseline model, and the deep neural networks with Word Embedding were found to be highly effective in the sentiment classification phase. The results additionally show that POS technique, which was used to determine the grammatical classes for each word in the tweet, did not improve the neural network accuracy. For traditional machine learning classifiers, SVM, RF and LR, the highest accuracy was achieved by LR classifier using a unigram language model, which reached 78 %. It should be also noted that employing a bigram language model negatively affected the performance for the used SVM, RF and LR classifiers by (21–25 %) in terms of accuracy respectively.

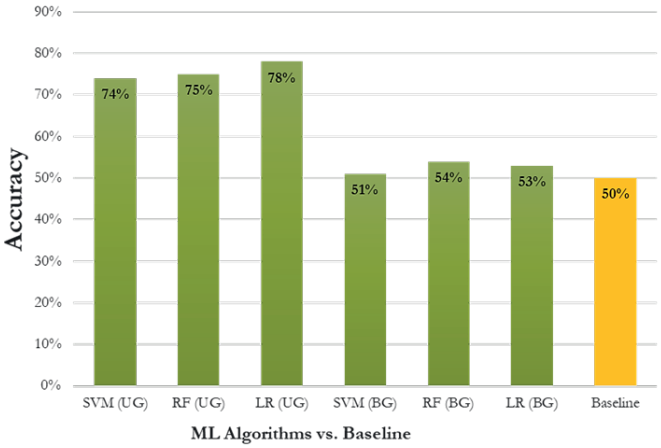


Fig. 4. Accuracy of ML models vs. baseline model for the sentiment classification phase.

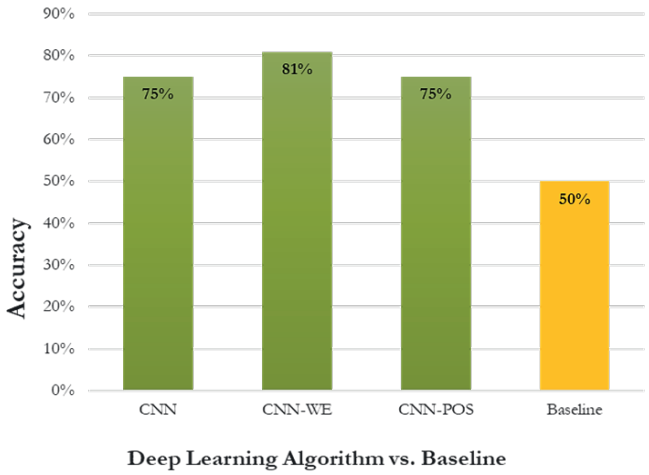


Fig. 5. Accuracy of DL models vs. baseline model for the sentiment classification phase.

A comparison of all machine learning applied models with the deep learning models is shown below in Fig. 6.

Fig. 7 shows the plot graph of training accuracy and validation over epochs. The epochs are represented as the row x and the accuracies are represented as the column y. We can see that the model in Word Embeddings algorithm was rising for the epochs and achieved accuracy of 81 %.

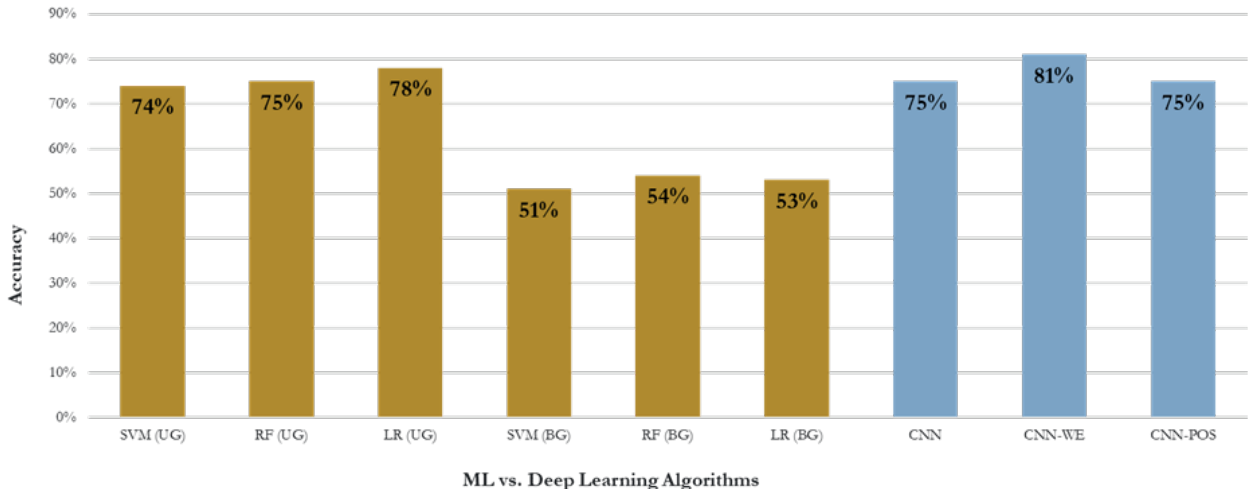


Fig. 6. Accuracy of ML models vs. DL models for the sentiment classification phase.

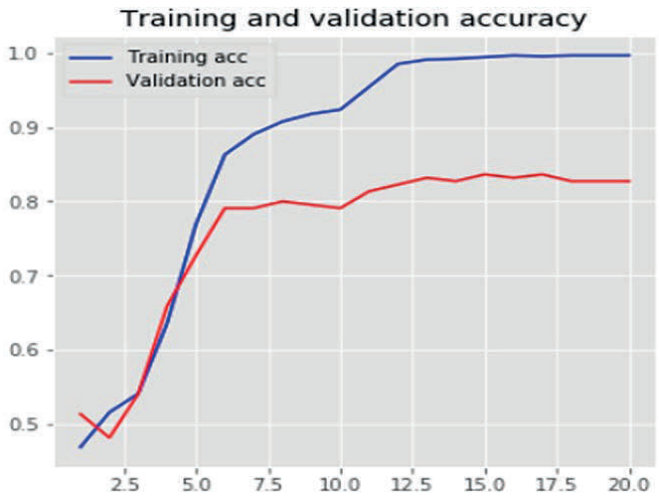


Fig. 7. Training and validation accuracy for the sentiment classification phase.

B. Aspect Detection Results

We conducted several experiments to study the impact of bigram, Word Embeddings and POS techniques on different classification algorithms. The accuracy of the baseline in the aspect detection task was calculated by dividing the number of tweets (115 tweets) in the majority class (CS class) by the total number of tweets in the testing set (256 tweets). As a result of this calculation, the accuracy of baseline model was 0.45. In this experiment, we applied the same classifiers that had been utilised in the sentiment analysis classification experiment with the same parameters for machine learning algorithms and the same layers and activation functions for deep learning classifier.

The experiment results are reported in Table VI and graphed in Figs. 8 and 9.

TABLE VI

ASPECT DETECTION RESULTS

Used classifier/ Feature		Accuracy	Precision	Recall	F1
Machine Learning (Unigram)	SVM	0.54	0.53	0.54	0.43
	RF	0.61	0.69	0.61	0.62
	LR	**0.63**	**0.65**	**0.63**	**0.59**
Machine Learning (Bigram)	SVM	0.45	0.25	0.45	0.29
	RF	0.24	0.43	0.24	0.24
	LR	0.45	0.21	0.45	0.28
Deep Learning	Neural Networks	0.59	0.63	0.59	0.59
	With Word Embeddings	0.58	0.61	0.58	0.56
	With POS	**0.75**	**0.80**	**0.75**	**0.74**

The results indicate that bigram method decreased the classifiers' performance and reported result under the baseline accuracy. We believe that the reason of this unacceptable result with the bigram method was the number of features in the bigram matrix. It should be noted that the number of features utilised to conduct the bigram experiment was three times bigger than unigram features. Most of the bigram features are considered noisy data because most of the features do not have weight. In other words, the bigram features matrix is considered a sparse matrix. On the other hand, the LR algorithm outperforms the other applied traditional algorithms using unigram with an accuracy score of 0.63. The score is considered acceptable compared to baseline model accuracy. Table VI also shows that the CNN algorithm achieved the best performance when applied with POS to extract names related to each aspect class. Word Embeddings technique underperformed the CNN algorithm compared to applying CNN algorithm without Word Embeddings technique. The achieved result is totally different from the results reported in the sentiment classification phase. Thus, the best result was achieved when applying Word Embeddings technique. It was also noted that POS achieved higher accuracy in the aspect detection task compared to the sentiment classification task.

We think the reason behind this is that in aspect detection task, the POS focuses on nouns and the nouns are expressed on the aspects more than the adjectives in the sentiment. It should be noted that the applied POS technique in aspect detection was focusing on the entities tagged as nouns; however, it was focusing on adjectives while detecting tweet sentiments.

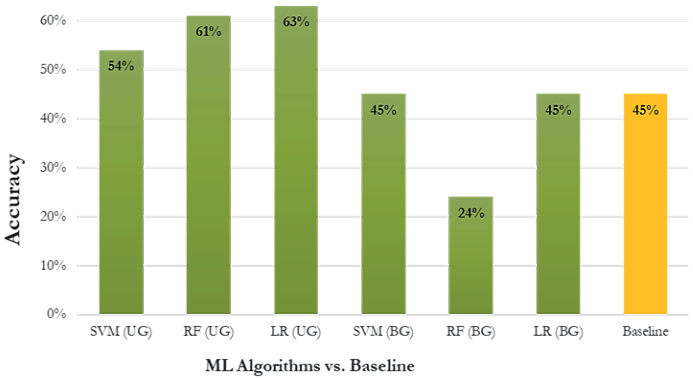


Fig. 8. Accuracy of ML models vs. baseline model for the aspect detection phase.

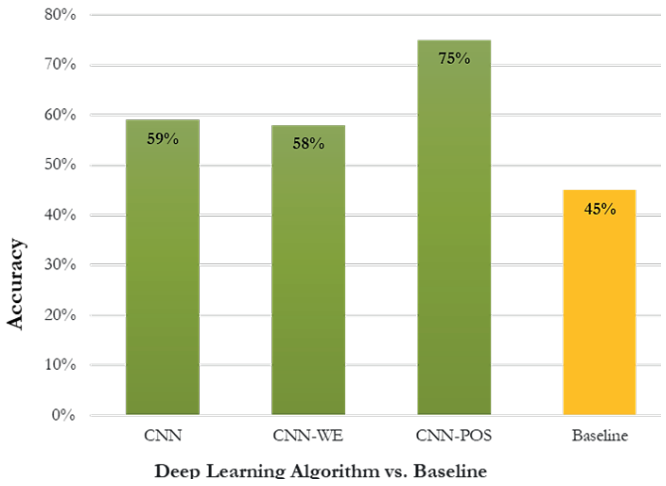


Fig. 9. Accuracy of DL models vs. baseline model for the aspect detection phase.

A comparison between all machine learning applied models with the deep learning models is shown in Fig. 10.

Fig. 11 shows the plot graph of training accuracy and validation over epochs. The epochs are represented as the row x and the accuracies are represented as the column y. We can see that the model in POS algorithm was rising for the epochs and achieved accuracy of 75 %.

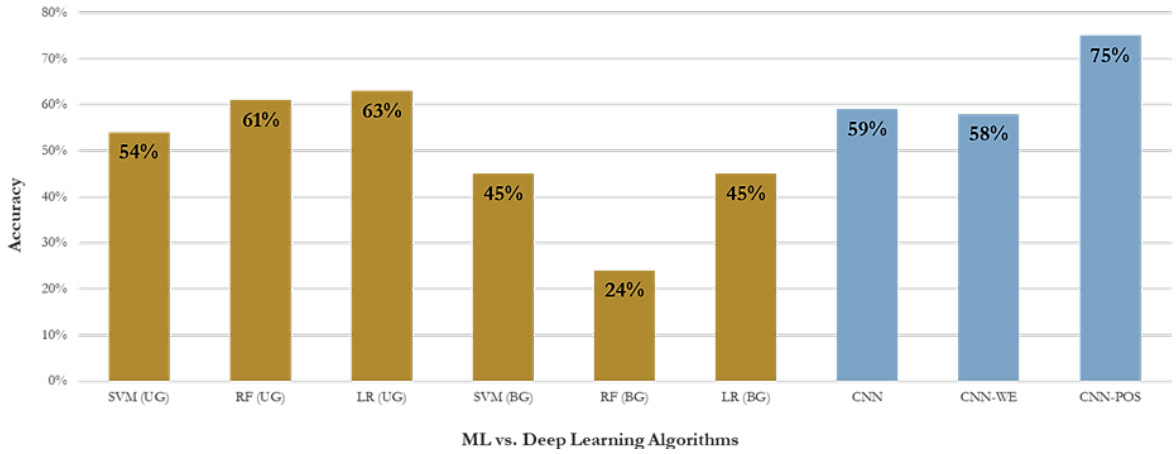


Fig. 10. Accuracy of ML models vs. DL models for the aspect detection phase.

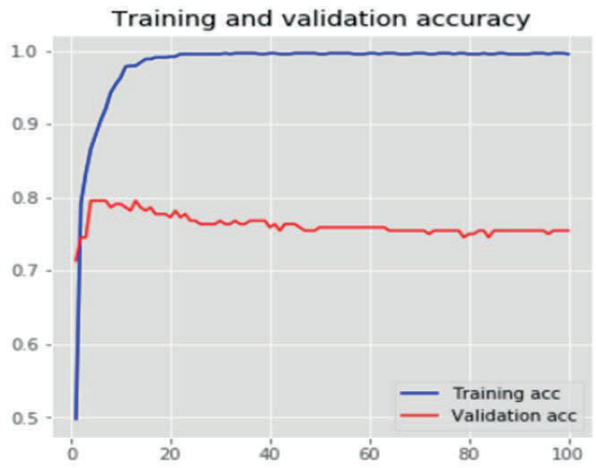


Fig. 11. Training and validation accuracy for the aspect detection phase.

C. Location Detection Results

This subsection shows the result of the proposed location NER model. Table VII shows the location NER model performance results using recall, precision, and F1 measures. In terms of precision, the trained model achieved 0.89 and 0.60 for City (CIT) and Point of Interest (POI), respectively. In our case, the precision is an important measure because it represents the percentage of the true location NE from the predicted location NE. In other words, for the predicted city NE, the percentage to be true city NE is 89 %. The recall measure represents the percentage of the true predicted location NE from all location NE in the dataset. The NER task results show that the proposed model can predict 53 % and 49 % from all POI and CIT NE, respectively. It is noteworthy that performing NER on the informal text such as social media data is a hard task compared to conducting this task on the formal text. There are several reasons why the NER model underperformed when applied to the text collected from a user generated platform such as a social media platform. Those reasons include spelling mistakes, grammar mistakes, multi-language text, and the use English words written in Arabic letters.

TABLE VII

LOCATION DETECTION RESULTS

Location named entity	Precision	Recall	F1
Point of Interest	0.60	0.53	0.56
City	0.89	0.49	0.63

VI. CONCLUSIONS

In the research, we have proposed a new framework for enhancing social medial sentiment analysis, especially in Twitter by considering the aspect and geographical location related to tweet content. The study has contributed to the field of Arabic sentiment analysis in the following way:

- The study experiment results show that the bigram has a negative impact when applied with the Arabic text to analyse user's sentiment or the related aspects. As a result of these experiments, this study recommends applying unigram features rather than bigram features with sentiment analysis on Arabic tweets.
- The conducted experiments show that applying POS or Word Embeddings techniques improve the performance of deep learning algorithm. Moreover, utilising NLP techniques with the deep learning algorithms outperforms the listed traditional machine learning algorithms.
- The result of sentiment analysis experiments shows that Word Embeddings technique improves the sentiment analysis model when applied with CNN algorithm and outperform the remaining NLP techniques. On the other hand, the aspect analysis experiment results show that

the CNN algorithms with POS technique achieve the best performance score compared to the performance of the remaining NLP techniques, including Word Embeddings technique.

- Location detection experiment results show that applying NER technique to the Arabic text by utilising deep learning algorithms (SpaCy library) is an efficient method to extract the location NE from tweets.

Future research areas are listed below:

- To study how to link the extracted location with geographical coordinates. In the present study, we have just extracted the location named entities and these extracted entities should be assigned to the correct geographical coordinates recognised during the location detection phase.
- To add other classification algorithms into our proposed model that perhaps might achieve better and higher accuracy results.
- To build a sentiment analysis dashboard that automatically monitors and evaluates the users' attitudes, hence, can help decision makers make their decision based on clear views.

REFERENCES

[1] B. Jongman, J. Wagemaker, B.R. Romero, and E.C. De Perez, "Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals," *ISPRS International Journal of Geo-Information*, vol. 4, no. 4, pp. 2246–2266, Oct. 2015. https://doi.org/10.3390/ijgi4042246

[2] D. Eilander, P. Trambauer, J. Wagemaker, and A. van Loenen, "Harvesting social media for generation of near real-time flood maps," *Procedia Engineering*, vol. 154, pp. 176–183, Aug. 2016. https://doi.org/10.1016/j.proeng.2016.07.441

[3] J.A. de Bruijn, H. de Moel, B. Jongman, J. Wagemaker, and J.C.J.H. Aerts, "TAGGS: Grouping tweets to improve global geoparsing for disaster response," *Journal of Geovisualization and Spatial Analysis*, vol. 2, Dec. 2018, Art no. 2. https://doi.org/10.1007/s41651-017-0010-6

[4] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, Aug. 2018, pp. 2145–2158. https://aclanthology.org/C18-1182/

[5] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N project report, Stanford 1.12, 2009.

[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *LREc*, vol. 10, 2010. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf

[7] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in *2012 International Conference on Collaboration Technologies and Systems (CTS)*, Denver, CO, USA, May 2012, pp. 546–550. https://doi.org/10.1109/CTS.2012.6261103

[8] M. Elgamal, "Sentiment analysis methodology of Twitter data with an application on Hajj season," *International Journal of Engineering Research & Science (IJOER)*, vol. 2, no.1, pp. 82–87, Jan. 2016. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ijoer.com/assets/articles_menuscripts/file/IJOER-JAN-2016-22.pdf

[9] Md Shoeb and A. Jawed, "Sentiment analysis and classification of tweets using data mining," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 12, Dec. 2017. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.irjet.net/archives/V4/i12/IRJET-V4I12267.pdf

[10] X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," in *Proceedings of the 22nd international conference on World Wide Web*, May 2013, pp. 607–618. https://doi.org/10.1145/2488388.2488442

[11] P. Bhoir and S. Kolte, "Sentiment analysis of movie reviews using lexicon approach," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Madurai, India, Dec. 2015, pp. 1–6. https://doi.org/10.1109/ICCIC.2015.7435796

[12] T. Hardeniya and D. A. Borikar, "An approach to sentiment analysis using lexicons with comparative analysis of different techniques," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 18, no. 3, pp. 53–57, May–Jun. 2016. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.iosrjournals.org/iosr-jce/papers/Vol18-issue3/Version-1/J1803015357.pdf

[13] H. Krishnan, M. Sudheep, and T. Santhanakrishnan, "Sentiment analysis of tweets for inferring popularity of mobile phones," *International Journal of Computer Applications*, vol. 157, no. 2, pp. 1–3, Jan. 2017. https://doi.org/10.5120/ijca2017912616

[14] L. Vu and T. Le, "A lexicon-based method for sentiment analysis using social network data," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, Las Vegas, Nevada, USA, Jul. 2017. https://www.researchgate.net/publication/319490430_A_lexicon-based_method_for_Sentiment_Analysis_using_social_network_data

[15] M.-Y. Day and C.-C. Lee, "Deep learning for financial sentiment analysis on finance news providers," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, USA, Aug. 2016, pp. 1127–1134. https://doi.org/10.1109/ASONAM.2016.7752381

[16] A. Al Sallab *et al*., "Deep learning models for sentiment analysis in Arabic," in *Proceedings of the Second Workshop on Arabic Natural Language Processing*, Beijing, China, Jul. 2015, pp. 9–17. https://doi.org/10.18653/v1/W15-3202

[17] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2015, pp. 959–962. https://doi.org/10.1145/2766462.2767830

[18] Y. Lu, et al. "Are deep learning methods better for twitter sentiment analysis?" in *Proceedings of the 23rd Annual Meeting of Natural Language Processing*, Japan, 2017. [Online]. Available: http://www.anlp.jp/proceedings/annual_meeting/2017/pdf_dir/C5-1.pdf

[19] A.M. Qamar, S.A. Alsuhibany, and S.S. Ahmed, "Sentiment classification of twitter data belonging to Saudi Arabian telecommunication companies," *Int. J. Adv. Comput. Sci. Appl*., vol. 8, no. 1, pp. 395–401, 2017. https://doi.org/10.14569/IJACSA.2017.080150

[20] A. Chakravarthy *et al*. "Hybrid architecture for sentiment analysis using deep learning," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, 2018. https://doi.org/10.26483/ijarcs.v9i1.5388

[21] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, "Farasa: A fast and furious segmenter for Arabic," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, San Diego, California, Jun. 2016, pp. 11–16. https://doi.org/10.18653/v1/N16-3003

[22] R. Al-Rfou, V. Kulkarni, B. Perozzi, and S. Skiena, "Polyglot-NER: Massive multilingual named entity recognition," in *Proceedings of the 2015 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics*, 2015, pp. 586–594. https://doi.org/10.1137/1.9781611974010.66

[23] A. Zirikly and M. Diab, "Named entity recognition for Arabic social media," in *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, Denver, Colorado, USA, Jun. 2015, pp. 176–185. https://doi.org/10.3115/v1/W15-1524

[24] AQAR. [Online]. Available: https://sa.aqar.fm/. Accessed on: March 20, 2020.

[25] A.K. Tarcar, A. Tiwari, V. Dhaimodker, P. Rebelo, R. Desai, and D. Rao, "Healthcare NER models using language model pretraining," in *The First Health Search and Data Mining Workshop (HSDM 2020), WSDM 2020 conference*, Houston, USA, Feb. 2020.

Norah AlShammari obtained M. Sc. degree in Computer Science from the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah. Her research interests include data mining and machine learning.
E-mail: n.alrakad@hotmail.com

Amal AlMansour is an Assistant Professor at the Faculty of Computing and Information Technology (FCIT), King Abdulaziz University (KAU), Jeddah, Saudi Arabia. In 2016, Amal received a PhD degree in Informatics from the Department of Computer Science, King's College London, UK. Thesis project focused on Social Computing and Mining Social Networks. She is also a certified Data Science and Advanced Analytics Associate (DECA-DS), Dell EMC. Her current research areas include artificial intelligence, data mining and machine learning, database management systems, social computing and mining social networks. Link to the research papers: https://scholar.google.com/citations?hl=en&user=yMKDTQMAAAAJ
E-mail: aalmansour@kau.edu.sa
ORCID iD: https://orcid.org/0000-0001-6689-4666