ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVED DEEP LEARNING SENTIMENT ANALYSIS FOR ARABIC

AHMED BINMAHFOUDH[1]

[1]Department of Computer Engineering, College of Computers and Information Technology, Taif University, Saudi Arabia

E-mail: [1]a.binmahfoudh@tu.edu.sa

ABSTRACT

Sentiment Analysis (SA) has recently gained great interest in Natural Language Processing (NLP). In fact, NLP consists in extracting data from texts and categorizing certain tweets as Positive, Negative, or Neutral. In this paper, we also present our participation in the Arabic Sentiment Analysis Challenge organized by King Abdullah University of Science and Technology (KAUST). Data of interest are tweets written in Arabic language, which becomes more challengeable. In this manuscript, we present the introduced system and the bi-LSTM model. Also, detail the less efficient explored solutions. Our main objective is to extract the crucial semantic data in Arabic tweets. The obtained findings about Arabic twitter corpus reveal that the performance of the developed technique is better than that proposed in the literature. Official test accuracy scores are 0.7605 with Macro-F1 score.

Keywords: *Attention, GRU, LSTM, Neural Network, Author Profiling, Gender Identification, Deep Learning*

1. INTRODUCTION

Artificial intelligence (AI) is widely used in everyday life and in many different fields such as industry, automation and expert systems, medicine and biology, education and even video games. Deep learning networks are widely employed in machine learning and, especially, in artificial intelligence.

They have shown excellent performance in various natural language processing (NLP) tasks (e.g., paraphrase identification, text summarization, machine translation, and question answering [1]. In fact, sentiment analysis (SA) has taken advantage of the development of deep neural networks (DNNs) thanks to their high efficiency and reduced need for engineered characteristics [2].

In 2021, King Abdullah University of Science and Technology (KAUST) allowed Twitter users to deal with SA. It presented to the participants 55K tweets to be trained and 20K tweets to be validated. Then, the performance of the competing teams was rated on the leaderboard, https://www.kaggle.com/c/arabic-sentiment-analysis-2021-KAUST. Finally, the winners were selected based on a set of 20K tweets with a distribution similar to that of the training and validation sets.

In the current work, we describe the model suggested to execute such sentiment-analysis for Arabic tweets. To extract important data regarding certain aspects, a bi-directional long short-term memory (bi-LSTM) model that treats the key part of the sentence is created. It was applied on a dataset containing twitter texts formed by KAUST [3]. The suggested model provided better accuracy than those given by the existing ones. We attend the rank 18 from 94 submissions registered in the competition's website, from 45 different countries.

The remaining part of this manuscript is divided into 4 sections. The literature review is presented in Section 2. Section 3 depicts the suggested attention-based proposals. The experiments conducted on a dataset including Twitter texts provided by the PAN Lab at CLEF 2018 [4] are described, in Section 4, to show the efficiency of the introduced model in terms of accuracy. In Section 5, we briefly explore the dataset used in sentiment-analysis-2021-kaust competition. Section 6 provides discussion and Section 7 concludes the paper and show our future perspectives. Techniques employed to minimize the training times and compare various neural network architectures are also defined.

2. STATE OF THE ART

The Arabic language is spoken by a large number of people around the world. According to Ethnologies, Arabic is the fifth most spoken language in the world, with over 422 million speakers. As such, there is likely to be a large amount of Arabic language content on the internet and being able to analyze the sentiment of this content could be useful for various applications.

Arabic is a complex language with many unique features that make it different from other languages. For example, it is written from right to left, and it has a large number of loanwords from other languages. This complexity means that Arabic sentiment analysis may be more challenging than sentiment analysis in other languages and developing effective machine learning models for this task could be an interesting research problem.

Sentiment analysis has a wide range of potential applications, including social media analysis, customer feedback analysis, and market research. These applications are all relevant to the Arabic-speaking world and being able to accurately analyze sentiment in Arabic could be useful for businesses and organizations operating in this region.

Sentiment analysis, also called opinion mining, represents a research field that examines opinions, feelings and attitudes about subjects, things, people, and events. These elements are generally expressed through text. It is intended to specify a predetermined sentiment classified as positive, negative, or neutral.

Sentiment analysis is performed to study the relation between the employment of language and its social aspect. Arabic is currently ranked as the fourth language used on the web, and there are approximately 168 million Arab internet users. Most sentiment analysis work focuses on the English language. The complexity of the Arabic language and the lack of resources available for Analyzing sentiments in Arabic like lexicons and datasets are the main barriers to Arabic sentiment analysis. Indeed, in 2021, a workshop on Arabic natural language processing (WANLP) organized a shared task about detecting sentiments from text written in Arabic [5]. The used dataset has 15K tweets.

In 2019, [6] introduced a hybrid system for Arabic SA, which uses approaches relying on lexicon and machine learning. The authors conducted experiments on several datasets, like ASTD and ArTwitter, utilizing different methods (e.g., machine learning and conventional learning models). The developed system provided 75.1% accuracy.

Moreover, various machine learning algorithms (e.g., SVM, NB, Ridge Regression (RR), and AdaBoost) were tested, in 2019, by [7] on a corpus of 151548 tweets. To extract features, the authors used TF-IDF and the best results (99.9%) was obtained by applying RR.

In their study, [8] employed the discrete bag-of-word (BOW) vector representations and continuous vector representations of the Doc2vec type in order to extract useful features to evaluate their corpus of 8000 posts on Facebook (4000 Arabic posts and 4000 written in Arabizi). The authors tested different classification algorithms like NB, RL, SVM and decision tree. They achieved a good performance by using logistic regression with an F1 score of 72%, for Arabic, and 78% for the Arabizi.

In 2018, [9] constructed an SA system in which CNN and LSTM were combined. The system was tested on two datasets, the ArTwitter and Arabic Health Services. In their work, the authors obtained an accuracy equal to 88.1% and 94.3% for the two used datasets, respectively.

Besides, in 2015, [10] built a large lexicon containing Arabic terms excerpts from press articles. From the extracted lexicon, an SA system was created and tested on Twitter dataset.

In 2014, [11] examined Arabic dialects. The researchers formed a lexicon of sentimental slang words and idioms (SSWIL). In the performed experiments, they utilized SVM and the created lexicon.

[12] used a dataset containing 2000 Tweets, the dataset was tested using lexicon-based systems and machine learning. The authors concluded that better findings could be obtained through the combination of the two methods.

In more recent works, the ASA big data scale was proposed in order to analyze authors' sentiment. For instance, [13] introduced the largest Book Reviews in Arabic Dataset (BRAD) and other applications for sentiment analysis. The employed dataset includes some dialectal content like Egyptian dialect.

In 2015, [14] created the Arabic Sentiment Tweets Dataset (ASTD). The dataset contains 10K Arabic tweets that were manually categorized as objective, subjective positive, subjective negative and subjective mixed.

Moreover, [15] constructed multi-domain Arabic resource including various Arabic dialects, such as Egyptian, Saudi Arabian and Emirati. It was formed by collecting more than 33k reviews. The rating showed the reviewer's general feeling toward a given object. It was extracted and normalized into positive, negative, or mixed.

3. THE INTRODUCED APPROACH

This section describes the suggested bi-LSTM model. Indeed, the bidirectional neural networks was used to classify sentiment reviews collected on Twitter. The application of the introduced model can be divided into three major steps: preprocessing, classification, and the evaluation of the system, as exposed in the diagram in Figure 1.

We created ASA relying on LSTM. In fact, the designed model showed excellent performance and was able to overcome the gradient vanishing or exploding problems. LSTM network is a recurrent neural network that has blocks of LSTM cells instead of standard neural network layers as seen in Figure 2. LSTM networks is based on a novel structure formed by memory cells; each of contains two memory blocks and an output layer.

The LSTM cell calculates its internal state following the iterative process presented below and for multiple blocks. Then, for each block, the computations were arbitrarily repeated.

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}xt + W_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{hf}h_{t-1} + W_{xf}xt + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(Woh_{t-1} + W_{xo}xt + Woc_t + b_o) \quad (3)$$

$$c_t = f_t \Theta c_{t-1} + i_t \Theta \, tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$h_t = o_t \Theta \, tanh(c_t) \quad (5)$$

where:

i_t: is the input gate that demonstrates the amount of novel data to be transferred via the memory cell.

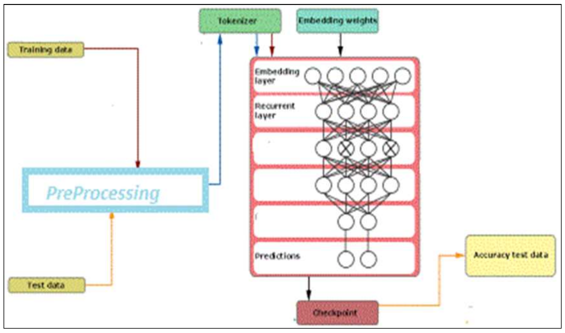


Figure 1: Representation of the bi-LSTM model

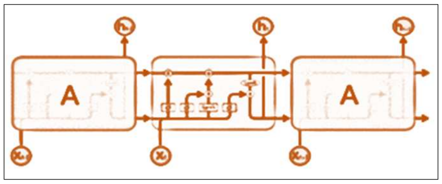


Figure 2: LSTM cells

f_t: refers to the forget gate that throws data from memory cell.

o_t: corresponds to the output gate. It shows the amount of data to be processed in the following step.

c_t: is the self-recurrent. It is generally equal to the standard RNN.

σ: is the sigmoid function.

h_t: is the final output.

Θ: designates the element-wise vector product.

W: are parameters matrices having various sub-scripts.

b: the bias vector.

3.1 Pre-Processing

Before classifying a given document, it is essential to apply the pre-processing process to clean and normalize the data. We describe below the methods used to prepare the considered data (corpus).

3.1.1 Data Analysis Levels

Three levels of data analysis can be distinguished:

- Document level: The documents analysis, in which a given document is classified as positive or negative [16], is based on the assumption that each document expresses opinions about the same entity (e.g., product).
- Sentence level: At this level, we determine whether each sentence expresses a positive, negative, or neutral opinion.
- Word level: At this more detailed analysis level, it is shown whether words express separately a positive, negative sentiment or other feeling.

This study focuses on the document level.

3.1.2 Pre-Processing Steps

The objective of this phase is to obtain a generic model for the normalization of raw texts. The different steps applied during the pre-processing phase are described below:

- Segmentation: This process allows transforming a textual document either into a sequence of sentences or of separate words depending on the analysis lev-el. It is intensively applied in a large number of automatic language processing applications [17]. Like other types of automatic language processing, segmentation has its particularities, whether at the linguistic level or at the IT level. Thus, it is based on the linguistic study and IT modelling that complement each other.
- Omitting «stop words»: Stop words are common words having less important meaning than

keywords, e.g., 'هو' ,'من' ,'له' ,'لن' ,' لم' ,' كل' ,'في' ,. In fact, this process reduces noise in textual data by using word lists or methods for dynamic stop-word identification.

- Removal of non-Arabic words and numbers: The texts written in Arabic are processed, in this work, to filter the words belonging to the other languages as well as the numbers that do not carry meaning and, therefore, do not have a polarity (positive, negative, etc.).
- Conversion of smileys: Smileys (or emoticons) are combinations of symbols; each of which represents the state, mode, or the emotion of the comment's writer towards a specific topic. We noticed that most of the previous works ignored the meaning of some special character representations by eliminating them. However, their exploration can enrich the studied corpus by transforming these characters combinations into words reflecting the ex-pressed feeling and, thus, by enriching the corpus to obtain better precision when detecting polarity. For this reason, we develop, in the present work, a small converter (symbol to word) whose functioning is presented in Table 1.
- Correction of long words: The redundancy of characters in words is a very well-known phenomenon, especially in social networks where Internet users want to show confirmation in their comments. It has been spread on the internet in forums, dating sites and any other site that offers a space for the user to express his/her opinion. Since the meaning of a word does not change if it is lengthened, a corrector that returns the word in its default form is introduced in this manuscript. For example, the word "رااااااااااااااائع" becomes “رائع".

3.2 Stemming

Arabic verbal lemmas are derived from a root. The root is a sequence of three or four letters that defines an abstract notion. The phenomenon of reducing lemmas to their roots or stemming affects the performance of a sentiment analysis system. It is the process of removing all prefixes and suffixes from a word to produce the stem or root of a word.

3.2.1 The Benefits of Stemming

This process is applied in various tasks, namely compression to reduce the size of documents, spell checking (instead of looking up a word in its complete form in a dictionary, only the root is searched), information retrieval and text analysis where stemming helps to map the grammatical variations of a word to instances of the same term.

3.2.2 Stemming Techniques

There are two methods of morphological analysis in Arabic language: "Arabic stemming" and "Arabic light stemming". In many cases, Arabic Stemmer does not seem to be effective in detecting the polarity of words from the same root but have different, sometimes opposite, meanings. For illustration, let us consider the following two words from the same root: (رائع) which means "wonderful" and (مروع) which means "terrible" in English. Despite the obvious difference in meaning, these two words come from the same root (روع) which means, "to frighten". The stemming technique would come up with reverse results in such cases.

The second technique, Arabic light stemming relies on the omission a number of prefixes and/or suffixes if the meaning is not affected. For instance, light stemming transforms the word (المسافرون) "travelers" into (مسافر) "traveler" rather than the root (سفر) "Travel". [18] developed another algorithm, called 'Arabic Khoja Stemmer'. The algorithm proposed functions by eliminating long suffixes and prefixes and comparing the rest of the word to the verb and noun patterns in order to extract the word root. This stemmer utilizes the data of a number of linguistic files including, but not limited to, diacritics, punctuation marks, articles and 168 lexical items.

The present work tested the two types of stemmers, "Arabic Khoja Stemmer" and "Arabic Light Stemmer". The results yielded are reported in the evaluation section.

Table 1: Examples of the conversion of a symbol into a word

Smiley	Comment	Translation	Comment after conversion	Translation
	جهاز ممتاز	Good smartphone	جهاز ممتاز أنا سعيد	Good smartphone, I am happy
	كاميرا سيئة	Bad camera	كاميرا سيئة أنا حزين	Bad camera, I am sad

3.3 Pre-Classification

To analyze a document, it is necessary to represent it in the right dimensions. In this paper, the dimensions are the terms (words or groups of words) that form the text. They are selected either statistically, according to their appearance in positive or negative documents, or in a finer (but also less generic) way. The finer selection most often uses a minimal lex-icon and grammatical knowledge (Part-of-speech tagging) on the words.

3.4 Text Representation

Several models of textual representation were developed in the literature:

- The "bag of words" approach: The "bag of words" is the simplest approach of text representation. It represents the text as a set of n-grams without taking into account their order of appearance nor the relationships [19] between them within the text, i.e., the text is considered as a set of words without sequentially.
- Sequentiality-based method: Techniques of sentence analysis relying on "bag of words" are often less efficient because the data derived from natural language is most often sequential.

3.5 Features Selection

The literature shows that good indicators of a sentiment differ from one work to another. In fact, some researchers proved that adjectives are more effective in determining the direction of reviews [20]. Other authors demonstrated that verbs or adverbs are more meaningful and reflect better the expressed feelings. To deal with this ambiguity, it is necessary to choose the appropriate and more efficient descriptors. In order to construct the elements of the features vector, several models of n-grams of different orders (uni-grams, big-grams and tri-grams) were developed.

- Uni-grams: They are considered the easiest features to extract. This model provides better data coverage.
- Bi-grams: This model is mainly used to detect negation.
- Trigrams: This model allows better capturing the constructs associated to expressions of sentiment.

Therefore, the process started by extracting all uni-grams, bi-grams and tri-grams in the annotated corpus. Then, the frequency of appearance of each candidate characteristics in the 250 documents was calculated. Afterwards, a dictionary containing all the candidate descriptors and their corresponding frequencies was built. Finally, for each opinion, if one of these candidate features was present in this dictionary, the frequency of this candidate will be extracted from it and placed in the feature vector representing this opinion. Then, a feature vector is constructed for each opinion based on the frequencies of the terms.

4. EXPERIMENTS

To test our method, we collected a new data using Twitter API. Fetching tweets from Twitter is a very important phase for sentiment analysis, and for this, we use the Tweepy library to access the Twitter API. Tweepy is an open-source library, which allows Python to communicate with the Twitter platform and use its API to access tweets. However, to work with the Twitter API, you must have a Twitter account, then connect to this account and create an "Application" there. Once this application has been created, you will need, in the "Keys and Access tokens" tab, to create:

- An API key (also called "Consumer Key") and its secret key.
- An "Access Token" and its secret counterpart.

Once these identifiers have been created, we put them in the 'config.ini' file, and afterwards. For our sentiment analysis classification, we fetch 1250 positive tweets and 1250 negative ones. We filter only the tweets about selling products.

Throughout this part, we will present the results we obtained using the different classification methods SVM, NB and KPPV. We will end with a 'discussion' part in which we will try to compare the performances of the different classifiers.

Positive Classification:

استثنائي. هدواء المكان وتميز الخدمات. يحتاج البوفيه الأفطار
الصباحي المزيد من الأصناف

Exceptional. The place is quiet, and the services are distinguished. The morning breakfast buffet needs more items.

Negative Classification:

اتجاه .المستثمرينلالغاء لشراء البيتكوين وباقي العملات

A tendency for investors to cancel the purchase of Bitcoin and the rest of the currencies

As results, classifiers choose positive words (exceptional استثنائي) according to Arabic sent wordnet to classify the tweet as positive and negative words (cancel الغاء) for negative tweets. We have ambiguity when we have the two in the same tweet.

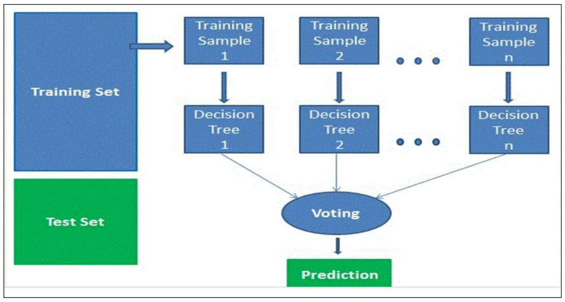


Figure 3: Architecture of Random Forest Classifier [21]

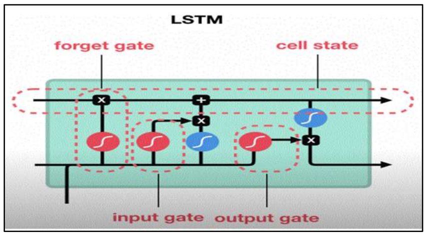


Figure 4: Graphical representation of LSTM memory cells [19]

4.1 Support Vector Machine (SVM)

We start by defining the SVM and introduce our approach for the classification task. Support Vector Machines (SVM, also known as Support Vector Networks) are the association of machine learning algorithms aimed to cope with solves classification, regression and anomaly detection problems. SVMs are credited for their solid theoretical frameworks, particular flexibility and ease. Even users with little experience in data mining can use SVMs.

We notice that with the SVM classifier we reach an accuracy of 79.27% on the test dataset, which is a good accuracy.

4.2 Random Forest Classifier

The term "Random Forest" refers to a type of supervised machine learning algorithm that rely on ensemble learning. For learning to take place in ensemble learning, algorithms of different types are combined together, or algorithms of the same type are used several times for better predictive performance. The random forest is a classification algorithm made of many decision trees. Multiple algorithms of the same model, such as multiple decision trees, make a forest of trees. Random forest algorithms are usable for regression as well as classification tasks. Figure 3 represents the architecture of random forest classifier.

With the Random Forest Classifier, we achieved an accuracy of 79.19% on the test dataset, which is an encouraging accuracy.

4.3 Naive Bayes

Naive Bayes Classifier is a popular machine-learning algorithm. It is a Supervised Learning algorithm used for classification. The naive Bayes classifier is based on Bayes' theorem. The latter is a classic of probability theory. This theorem is based on conditional probabilities (What is the probability that an event will occur knowing that another event has al-ready occurred).

With the Naive Bayes classifier, an accuracy of 78.30% is achieved on the test dataset, which is less accuracy than that achieved by the SVM and Random Forest Classifier.

4.4 Long Short-Term Memory (LSTM)

LSTM networks introduce a new structure called memory cell where each cell is composed of two memory blocks and an output layer as shown in Figure 4.

We are developing a system that will predict the sentiment of a tweet as positive or negative. This means that after the model is developed, we will have to make predictions about new tweets. This will require data preparation on these new tweets as done on the training data for the model. We will ensure that this constraint is built into the evaluation of our models by splitting the training and testing datasets before any data preparation. That means, we will use 33% of dataset data as test set and 67% of remaining data as training dataset.

These vectors are random at the start of training, but during training become meaningful to the network. We can encode the training data as sequences of integers using the Tokenizer class in the Keras API. Next, we use an LSTM network as they have been effective in solving data classification issues. Adam's efficient implementation of stochastic gradient descent is used, and we keep track of accuracy in addition to loss during training. The model is trained for 10 epochs, that means, we will pass 10 times through the training data.

With the LSTM we achieve an accuracy of 92.13% on the test data set, which is a very good accuracy (Figure 5).

4.5 Discussion

We compare the results obtained between the SVM method, Naive Bayes, Random Forest Classifier, and the LSTM neural network with the different pre-processings used. Table 2 summarizes the results obtained.

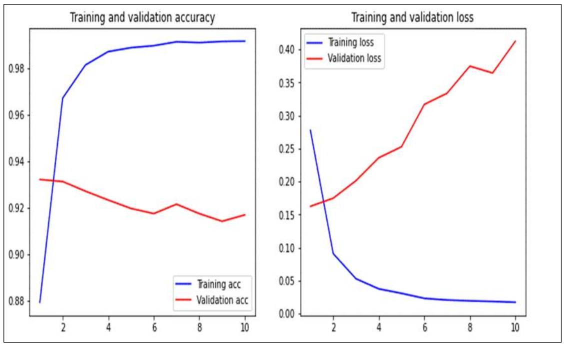


Figure 5: Training and validation accuracy and Training and validation loss

Table 2: Results comparison on Twitter data

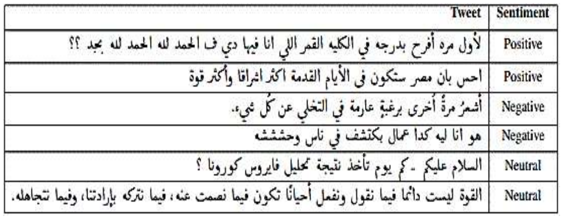
Classi fier	Accuracy
SVM	79.27%
Random Forest	79.19%
Naive Bayes	78.30%
LSTM	92.13%

We noticed that the approach that concerns LSTM Neural Networks gives a better precession (92.13%) when training the model, as long as the SVM approach had an acceptable precession (79.27%). In addition, we found that the SVM, Random Forest, and Naive bayes algorithms give similar results with an average precession of 78.92%.

Indeed, the LSTM architecture achieved good accuracy on the test data set. This allows us to say that the use of LSTM neural networks will allow us to predict the sentiments of new tweets with less error and more pre-cession, and that is an advantage for the recurrent neural networks that are used very often in the processing of sequential data to be able to use them also for sentiment analysis and also for the classification of texts.

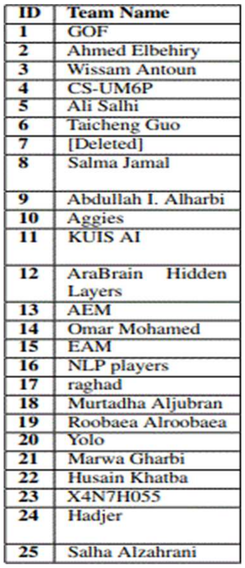
5. COMPETITIVE CONFERENCE KAUST PARTICIPATION

The dataset used in sentiment-analysis-2021-kaust competition includes 100K tweets collected from May 2012 and April 2020 and labelled positive, negative, or neutral. The tweets are written in various Arabic dialects such as Khaleeji, Hijazi, Egyptian, and modern standard Arabic. Examples of the annotated tweets used in ASAD are illustrated in Figure 6.



Tweet	Sentiment
لأول مره أفرح بدرجه في الكليه القمر اللي انا فيها دي ف الحمد لله الحمد لله بجد ؟؟	Positive
احس بان مصر ستكون فى الأيام القدمة اكثر اشراقا وأكثر قوة	Positive
أشعرُ مرةً أخرى برغبةٍ عارمة في التخلي عن كُل شيء.	Negative
هو انا ليه كدا عمال بكتشف في ناس وحششه	Negative
السلام عليكم .. كم يوم تأخذ نتيجة تحليل فايروس كورونا ؟	Neutral
القوة ليست دائما فيما نقول ونفعل أحيانًا تكون فيما نصمت عنه ؛ فيما نتركه بإرادتنا، وفيما نتجاهله.	Neutral

Figure 6: Example of tweets in ASAD



ID	Team Name
1	GOF
2	Ahmed Elbehiry
3	Wissam Antoun
4	CS-UM6P
5	Ali Salhi
6	Taicheng Guo
7	[Deleted]
8	Salma Jamal
9	Abdullah I. Alharbi
10	Aggies
11	KUIS AI
12	AraBrain Hidden Layers
13	AEM
14	Omar Mohamed
15	EAM
16	NLP players
17	raghad
18	Murtadha Aljubran
19	Roobaea Alroobaea
20	Yolo
21	Marwa Gharbi
22	Husain Khatba
23	X4N7H055
24	Hadjer
25	Salha Alzahrani

Figure 7: The top-ranked teams' result

To assess the performance of the proposed bi-LSTM model, 10-fold cross-validation was applied. The obtained accuracy rate for gender detection was equal to 0.7605 for test data, as shown in Figure 5, and 19 ranks were provided by 1,247 submissions team. The top-ranked teams are listed in Figure 7.

6. DISCUSSION

The Arabic language is spoken by a large number of people around the world. According to Ethnologies, Arabic is the fifth most spoken language in the world, with over 422 million speakers. As such, there is likely to be a large amount of Arabic language content on the internet, and being able to analyze the sentiment of this content could be useful for various applications.

Arabic is a complex language with many unique features that make it different from other languages. For example, it is written from right to left, and it has a large number of loanwords from other languages. This complexity means that Arabic sentiment analysis may be more challenging than sentiment analysis in other languages, and developing effective machine learning models for this task could be an interesting research problem.

Sentiment analysis has a wide range of potential applications, including social media analysis, customer feedback analysis, and market research. These applications are all relevant to the Arabic-speaking world, and being able to accurately analyze sentiment in Arabic could be useful for businesses and organizations operating in this region.

There are some weaknesses in our work:

- Lack of sufficient training data: LSTM models, like all neural network models, require a large amount of data to learn from. If the model is not trained on a sufficient amount of Arabic data, it may not be able to accurately capture the patterns and nuances of the language.
- Inefficient preprocessing: It is important to preprocess the data correctly before training an LSTM model. This can include things like tokenization, stemming, and stop word removal. If the data is not preprocessed effectively, the model may not be able to learn from it effectively.
- Poor model architecture: The architecture of the LSTM model, including the number of layers and the size of each layer, can also impact the model's performance. If the model is not configured properly, it may not be able to learn from the data effectively.
- Overfitting or underfitting: If the model is overfitting, it may be memorizing the training data rather than learning general patterns that can be applied to new data. On the other hand, if the model is underfitted, it may not be able to capture the complexity of the data, leading to poor performance.

It is also worth noting that there could be other, language-specific factors that contribute to the poor performance of an LSTM model on Arabic data. Without more context about the specific model and data in question, it is difficult to identify the specific cause of any issues.

However, the difference from prior work keeping in view the need statement before conclusion mentioning the achievement of this study can be presented in these 3 points:

- LSTMs are particularly good at handling sequential data, such as text. This is because they have a "memory" component that al-lows them to remember important information from earlier in the sequence and use it to inform their predictions later.
- LSTMs are able to effectively capture long-term dependencies in data. This means that they can under-stand the context and meaning of words even if they are separated by many other words in the sequence. This is especially useful for sentiment analysis, where the over-all sentiment of a piece of text may be dependent on words that are far apart from each other.
- LSTMs have been shown to perform well on a wide range of natural language processing tasks, including sentiment analysis. This means that they have been extensively tested and refined, and are likely to be a reliable choice for this task.

7. CONCLUSION

This paper introduced the system used in the Arabic Sentiment Analysis Challenge organized by KAUST university in Saudi Arabia. In fact, Arab sentiment analysis is applied to detect the individual’s opinion towards a specific event, brand, or something else. We develop Bidirectional Long Short-Term Memory Networks Language models to label a given tweet as Positive, Negative, or Neutral. The obtained results revealed the excellent performance of the suggested bi-LSTM model in Arab sentiment analysis and in dealing with natural language processing problems. Our results are comparable to the best state of the art methods.

In this study, we ameliorated the accuracy of sentiment analysis models. Research on using deep learning with LSTM for sentiment analysis could lead to the development of more accurate models for detecting sentiment in Arabic text.

As future work, we will apply LSTM to new domains. Research on using LSTM for sentiment analysis in Arabic could also explore new application areas or domains where these models could be used.

Also, we enhanced the performance of existing models. Studies that compare the performance of deep learning with LSTM models to other approaches in Arabic sentiment analysis could provide insights into how to improve the performance of existing models.

Finally, we can explore new techniques for preprocessing and representing text data: Research on using deep learning with LSTM for sentiment analysis in Arabic could also investigate new techniques for pre-processing and representing text data, which could be applied to other natural language processing tasks.

ACKNOWLEDGMENTS

The author is grateful to help from Dr. Seifeddine Mechti (University of Sfax) as one of the team. Also, to King Abdullah University of Science and Technology (KAUST) in performing this competition and providing the data.

REFERENCES:

[1] Yigitcanlar T, Kankanamge N, and Vella K, “How are smart city concepts and technologies perceived and utilized? A systematic geo-Twitter analysis of smart cities in Australia”, *Journal of Urban Technology*, Vol. 28, No. 1-2, 2021, pp. 135-154.

[2] S.V. P, and Ittamalla R, “An analysis of attitude of general public toward COVID-19 crises–sentimental analysis and a topic modeling study”, *Information Discovery and Delivery*, Vol. 49, No. 3, 2021, pp. 240-249.

[3] Alharbi B, Alamro H, Alshehri M, Khayyat Z, Kalkatawi M, Jaber II, and Zhang X, “ASAD: A Twitter-based benchmark Arabic sentiment analysis dataset, *arXiv preprint* arXiv:2011.00578, 2020.

[4] Rangel F, Rosso P, Montes-y-Gómez M, Potthast M, and Stein B, “Overview of the 6th author profiling task at pan 2018: multimodal gender identification in twitter”, *Working Notes Papers of the CLEF*, 2018, pp. 1-38.

[5] Kankanamge N, Yigitcanlar T, Goonetilleke A, and Kamruzzaman M, “Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets”, *International journal of disaster risk reduction*, Vol. 42, 2020, pp. 101360.

[6] Elshakankery K, and Ahmed MF, “HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis”, *Egyptian Informatics Journal*, Vol. 20, No. 3, 2019, pp. 163-171.

[7] Gamal D, Alfonse M, El-Horbaty ES, and Salem AB, “Twitter benchmark dataset for Arabic sentiment analysis”, *International Journal of Modern Education and Computer Science*, Vol. 11, No. 1, 2019, pp. 33-38.

[8] Ebadi A, Xi P, Tremblay S, Spencer B, Pall R, and Wong A, “Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing”, *Scientometrics*, Vol. 126, No. 1, 2021, pp. 725-739.

[9] Alayba AM, Palade V, England M, and Iqbal R, “Improving sentiment analysis in Arabic using word representation”, *In 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, 2018 pp. 13-18.

[10] Al-Ayyoub M, Essa SB, and Alsmadi I, “Lexicon-based sentiment analysis of Arabic tweets”, *International Journal of Social Network Mining (IJSNM)*, Vol. 2, No. 2, 2015, pp. 101-114.

[11] Soliman TH, Elmasry MA, Hedar A, and Doss MM, “Sentiment analysis of Arabic slang comments on facebook”, *International Journal of Computers & Technology*, Vol. 12, No. 5, 2014, pp. 3470-3478.

[12] Abdulla NA, Ahmed NA, Shehab MA, and Al-Ayyoub M, “Arabic sentiment analysis: Lexicon-based and corpus-based”, *In 2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*, 2013, pp. 1-6.

[13] Elnagar A, and Einea O, “BRAD 1.0: Book reviews in Arabic dataset”, *In 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, 2016, pp. 1-8.

[14] Nabil M, Aly M, and Atiya A, “ASTD: Arabic sentiment tweets dataset”, *In Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 2515-2519.

[15] Al-Harbi O, “Classifying sentiment of dialectal Arabic reviews: a semi-supervised approach”, *The International Arab Journal of Information Technology*, Vol. 16, No. 6, 2019, pp. 995-1002.

[16] Bashar MK, “A hybrid approach to explore public sentiments on COVID-19”, *SN Computer Science*, Vol. 3, No. 3, 2022, pp. 1-9.

[17] Chen C, Teng Z, Wang Z, and Zhang Y, “Discrete opinion tree induction for aspect-based sentiment analysis”, *In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, 2022, pp. 2051-2064.

[18] Khoja S, and Garside R, “Stemming Arabic text”, *Computer Science Department, Lancaster University, Lancaster, UK*, 1999, [Online] Available: http://zeus.cs.pacificu.edu/shereen/research.htm#stemming

[19] Chamekh A, Mahfoudh M, and Forestier G, “Sentiment analysis based on deep learning in e-commerce”, *In International Conference on Knowledge Science, Engineering and Management*, Springer, Cham, 2022, pp. 498-507.

[20] Cortis K, and Davis B, “Over a decade of social opinion mining: a systematic review”, *Artificial intelligence review*, Vol. 54, No. 7, 2021, pp. 4873-4965.

[21] Ligthart A, Catal C, and Tekinerdogan B, “Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification”, *Applied Soft Computing*, Vol. 101, 2021.