

**DEVELOPING ARABIC SENTIMENT ANALYSIS FOR SAUDI ARABIA'S
TELECOMMUNICATION COMPANIES USING DEEP AND ENSEMBLE
LEARNING**

**DESENVOLVENDO ANÁLISE DE SENTIMENTO ÁRABE PARA EMPRESAS DE
TELECOMUNICAÇÕES DA ARÁBIA SAUDITA USANDO APRENDIZAGEM
PROFUNDA E ENSEMBLE**

**DESENVOLVENDO ANÁLISE DE SENTIMENTO ÁRABE PARA EMPRESAS DE
TELECOMUNICAÇÕES DA ARÁBIA SAUDITA USANDO APRENDIZAGEM
PROFUNDA E ENSEMBLE**

How to cite:

Almutari, Sara M., & Alotaibi, Fahad M. (2023). Developing arabic sentiment analysis for Saudi Arabia's telecommunication companies using deep and ensemble learning. Revista Gestão & Tecnologia (Journal of Management & Technology). v. 23, nº 4, 2023, p: 7 - 26

Sara Manour Almutairi

Faculty of computer and information technology, King Abdulaziz Uni-versity, Saudi Arabia, Riyadh

Fahad Mazead Alotaibi

Faculty of computer and information technology, King Ab-dulaziz University, Saudi Arabia, Jeddah

Scientific Editor: José Edson Lara
Organization Scientific Committee
Double Blind Review by SEER/OJS
Received on 22/11/2023
Approved on 18/12/2023



This work is licensed under a Creative Commons Attribution – Non-Commercial 3.0 Brazil

Abstract

Sentiment analysis is a type of artificial intelligence that uses algorithms to determine whether an opinion is positive or negative. Arabic Sentiment Analysis (ASA) is responsible for assessing people's opinions, feelings, and responses to a variety of products and services on social and commercial networking sites. In this article, we develop a new Arabic sentiment analysis model for Saudi Arabia's telecommunication companies (Zain, Mobily, and STC). We create and develop a new dataset, called Sara-Dataset, for analyzing customer opinions towards Saudi Arabian communication firms. using Google Colabs libraries (e.g., Keras), our dataset consists of 50532 tweets. After processing and preparation it becomes 27294 tweets. The dataset is divided into three parts: training, validation, and testing, which each represent 70%, 10%, and 20%, respectively. The proposed model depends on deep learning models: LSTM (long-short-term memory) and CNN (conventional neural network). We evaluate our model using several parameters. The number of training epochs, the loss function, the optimizer (Adadelta, Adagrad, and Adam), batch size, and the ensemble learning approach. We evaluate our model in terms of accuracy, precision, F-measure, R-call, ROC, and loss function. When compared to other models, our results indicate significant enhancements. The best accuracy results of the LSTM model with the Adam optimizer and 32-batch size are 94.97%. The best accuracy result of the CNN model with the Adam optimizer and 32-batch size is 96.83%. Our future work is to use ensemble learning model.

Keywords: ASA; Deep learning; CoLabs; Sara-Data-set; Accuracy; Prepossessing; LSTM; CNN; ROC; Zain; Mobily; STC

Resumo

A análise de sentimento é um tipo de inteligência artificial que utiliza algoritmos para determinar se uma opinião é positiva ou negativa. A Análise de Sentimento Árabe (ASA) é responsável por avaliar as opiniões, sentimentos e respostas das pessoas a uma variedade de produtos e serviços em sites de redes sociais e comerciais. Neste artigo, desenvolvemos um novo modelo de análise de sentimento árabe para as empresas de telecomunicações da Arábia Saudita (Zain, Mobily e STC). Criamos e desenvolvemos um novo conjunto de dados, denominado Sara-Dataset, para analisar as opiniões dos clientes em relação às empresas de comunicação da Arábia Saudita. usando bibliotecas do Google Colabs (por exemplo, Keras), nosso conjunto de dados consiste em 50.532 tweets. Após processamento e preparação, são 27.294 tweets. O conjunto de dados é dividido em três partes: treinamento, validação e teste, cada uma representando 70%, 10% e 20%, respectivamente. O modelo proposto depende de modelos de aprendizagem profunda: LSTM (memória de longo-curto prazo) e CNN (rede neural convencional). Avaliamos nosso modelo usando vários parâmetros. O número de épocas de treinamento, a função de perda, o otimizador (Adadelta, Adagrad e Adam), tamanho do lote e a abordagem de aprendizagem em conjunto. Avaliamos nosso modelo em termos de exatidão, precisão, medida F, chamada R, ROC, e função de perda. Quando comparados com outros modelos, nossos resultados indicam melhorias significativas. Os melhores resultados de precisão do modelo LSTM com o otimizador Adam e tamanho de 32 lotes são 94,97%. O melhor resultado de precisão do modelo CNN com o otimizador Adam e

tamanho de 32 lotes é 96,83%. Nosso trabalho futuro é usar o modelo de aprendizagem em conjunto.

Palavras-chaves: AAS; Aprendizagem profunda; CoLabs; Conjunto de dados Sara; Precisão; Agradável; LSTM; CNN; ROC; Zain; Móvel; STC

Resumen

El análisis de sentimientos es un tipo de inteligencia artificial que utiliza algoritmos para determinar si una opinión es positiva o negativa. El Análisis de Sentimiento Árabe (ASA) es responsable de evaluar las opiniones, sentimientos y respuestas de las personas a una variedad de productos y servicios en sitios de redes sociales y comerciales. En este artículo, desarrollamos un nuevo modelo de análisis del sentimiento árabe para las empresas de telecomunicaciones de Arabia Saudita (Zain, Mobily y STC). Creamos y desarrollamos un nuevo conjunto de datos, llamado Sara-Dataset, para analizar las opiniones de los clientes sobre las empresas de comunicación de Arabia Saudita. Utilizando las bibliotecas de Google Colabs (por ejemplo, Keras), nuestro conjunto de datos consta de 50532 tweets. Después del procesamiento y preparación, se convierten en 27294 tweets. El conjunto de datos se divide en tres partes: entrenamiento, validación y pruebas, cada una de las cuales representa el 70%, 10% y 20%, respectivamente. El modelo propuesto depende de modelos de aprendizaje profundo: LSTM (memoria a corto plazo) y CNN (red neuronal convencional). Evaluamos nuestro modelo utilizando varios parámetros. El número de épocas de entrenamiento, la función de pérdida, el optimizador (Adadelata, Adagrad y Adam), el tamaño del lote y el enfoque de aprendizaje conjunto. Evaluamos nuestro modelo en términos de exactitud, precisión, medida F, llamada R, ROC, y función de pérdida. En comparación con otros modelos, nuestros resultados indican mejoras significativas. Los mejores resultados de precisión del modelo LSTM con el optimizador Adam y un tamaño de 32 lotes son del 94,97%. El mejor resultado de precisión del modelo CNN con el optimizador Adam y un tamaño de 32 lotes es del 96,83%. Nuestro trabajo futuro es utilizar el modelo de aprendizaje conjunto.

Palabras clave: ASA; Aprendizaje profundo; CoLabs; Sara-Conjunto de datos; Exactitud; Agradable; LSTM; CNN; República de China; Zaín; Movilidad; STC

1. INTRODUCTION

Sentiment analysis is an artificial intelligence technique that employs techniques to analyze whether an opinion is positive or negative. It's a powerful tool in the election process and social media to classify people's opinions towards things (e.g., products) [1–3]. Sentiment analysis is recognized as a significant technology for effectively studying customers' opinions. Preparing data, recognizing and identifying respondents, and evaluating findings are the

primary components of sentiment analysis [4–10]. There are several studies targeting sentiment analysis in e-Marketing using deep learning algorithms

The main processes of the proposed methodology are shown in Figure 1. The first step consists in specifying further the research gap through conducting a comprehensive survey on Arabic sentiment analysis using deep learning, as well as investigating the benefits and drawbacks of using such models. Then, the related works relevant to Arabic sentiment analysis using deep learning are collected and studied. After that, we create and preprocess the Sara dataset in order to construct customer opinion for Arabic sentiment analysis using LSTM and CNN. Then, we evaluate the model results in terms of accuracy, precision, recall, F-measure, and loss function. Next, we use ensemble learning techniques to improve the overall accuracy of Arabic sentiment analysis.

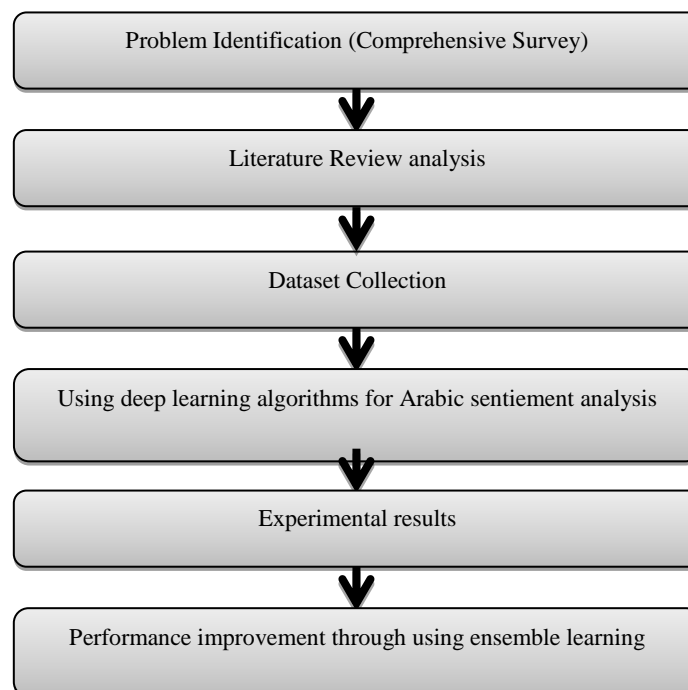


Figure 1. Developing sentiment analyses using deep learning

The rest of paper is organized as follows: Section 1 describes the introduction. Related works for Arabic sentiment analysis using deep learning is described in section 2. In Section 3, we explain our technique to develop and implement a covert channel. Experimental results

and discussion are described in Section 4. Finally, we mention our finding and conclusion in Section 5.

2. RELATED WORKS FOR ARABIC SENTIMENT ANALYSIS USING DEEP LEARNING

This section will include relevant articles on using deep learning in Arabic sentiment analysis. As depicts in Table 1 and depending on our previous works [24] we note that there is a low accuracy in using deep learning with existing sentiment analysis datasets. Also, the authors don't deal with large Arabic sentiment datasets.

Table 1.
Deep learning algorithms and its reflection for our research

Paper Number	Dataset source	Algorithms	Accuracy	Reflection	Enhancements
[18]	Arabic Sentiment Tweets Dataset (ASTD).	CNN is coupled with LSTM	65.05% using ensemble learning	Ensemble learning can enhance the accuracy of ASA	accuracy has to be improved
[19]	From Twitter. Consists of 1103 tweets (576 as positive and 527 labeled as negative)	Lexical+ SVM.	84.01%	we will use deep learning with large dataset	Hybrid algorithm enhance the accuracy of semantic analysis The data set is small
[20]	HTL and LABR which belongs to book rating	CNN+LSTM	Using HTI dataset, the accuracy is 85.38%. Using LABA dataset, the accuracy is 86.88%	Several papers using combination of CNN and LSTM	The accuracy is increased when using deep learning algorithm. And we must increase the datasets
[21]	Arabic tweets is collected that consist of 500 tweets related for education area	DT and SVM as traditional machine learning And the feed forward architecture as	Using deep learning accuracy 90%. using SVM 85%	Deep learning algorithm outperform machine learning algorithm	Deep learning with weighting characteristic need time consuming

		deep learning		
[22]	SemEval 2018) contains 4372 tweets, which are organized into three categories: training, development, and testing	Bidirectional LSTM	75.5% accuracy for validation and 49.8% for testing	SemEval is considered as large scale accuracy has to be improved
[23]	Twitter Arabic Hotels reviews	(LSTM) and Bidirectional LSTM	82.6%	When dealing with huge dataset, the RNN has the overfitting problem Solving overfitting problem

Altaher et al. [14], use deep learning algorithms. They used stop-word and stemming as pre-processing methods and they used feature and information gain as feature weighting. Then, applying a deep learning algorithm to effectively and accurately classifies Arabic tweets either as positive or negative tweets. The dataset is consisted from 500 Arabic tweets, and the tweets mainly discuss general topics about education. the feed-forward architecture as deep learning. The accuracy when using deep learning is 90%. Using SVM, DT, neural network they accuracy are 85%, 67%, and 80%, respectively

3. PROPOSED ARCHITECTURE FOR ARABIC SENTIMENT ANALYSIS USING DEEP LEARNING

Our architecture is based on deep models that are trained to detected the customer opinions towards Saudi Arabian communication firms. In this section, we go through the our solution depicted in Figure 2. Specifically, it relies on the following components:

1. Dataset collection and preprocessing: we create and develop Sara dataset for customer opinions towards Saudi Arabian communication firms

2. Model training: In this thesis, we use two deep learning models. Namely, (LSTM) Long short term memory [13, 14], and (CNN) Conventional Neural Network [15]. We intend to use Google Colabs [16] features to pre-process the dataset.
3. Colabs [16] is a Google tool that allows us to develop, implement, and run Python programs directly in the browser. On CoLab, a Jupyter notebook is also offered as a service. Colabs also provides a variety of libraries, such as Keras [17], that offers machine learning and natural language processing technologies to be applied.

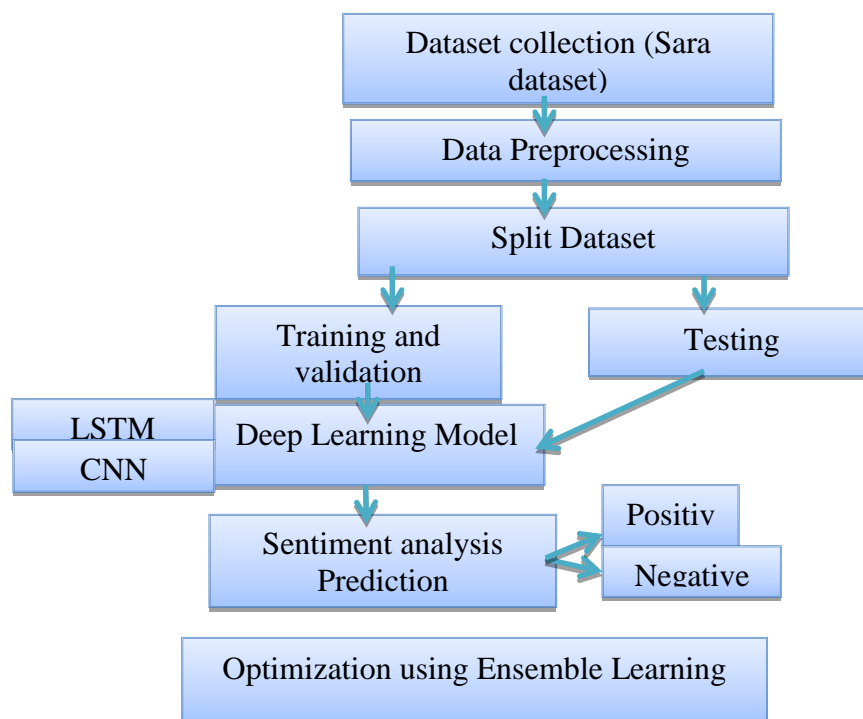


Figure 2. The architecture for detecting opinions towards Saudi Arabian communication firms.

4. Dataset: One should mention that after splitting the collected data into training, validation and testing sets. Training dataset makes up 70% of the overall Sara-dataset. Validation dataset makes up 10% of the overall Sara-dataset. Testing dataset makes up 20% of the overall Sara-dataset dataset. We divide the dataset into five parts (5 folds); four parts for training and validation and one part used testing. Following training, we utilize the testing

dataset part to see how accurate our model. During training, we will measure the performance of proposed model in terms of accuracy, precision, F-measure, R-call and ROC and error rate.

5. Evaluation: We evaluate LSTM, and CNN models using accuracy, precision, recall, F-measure, and loss function.

4. EXPERIMENTS SETTING

4.1. *Experimental environment*

The experiments of this research will be conducted using Google CoLabs [9] that is considered as a GPU hardware accelerator infrastructure. Specifically, Colabs is a Google tool that allows us to develop, implement, and run Python programs directly through a web browser. On CoLabs, the Jupyter notebook is also offered as a complementary feature. Colabs also provides a variety of libraries, such as Keras [17], that allows natural language processing techniques to be applied. We investigate the setting of:

- The number of training epochs,
- The loss function,
- The optimizer,
- Ensemble learning approach.

4.2. *Sara Dataset*

We develop a new Arabic dataset for three Saudi Arabian communication firms to improve the quality and quantity of their services by collecting customer feedback. An individual's customer is categorized into two major classes positive and negative opinions towards Saudi Arabian Communication firms. We selected three firms, which are Zain, Mobily, and Saudi Telecom (STC). We are targeting the customer tweets from Twitter from June 2022 to December 2022. As depicted in Figure 3, the targeted companies are Zain, Mobily, and STC. After pre-processing we collected around 10000 tweets belongs to STC companies. Around 9000 tweets belongs to Mobily company and the remainder tweets belongs to Zain company.

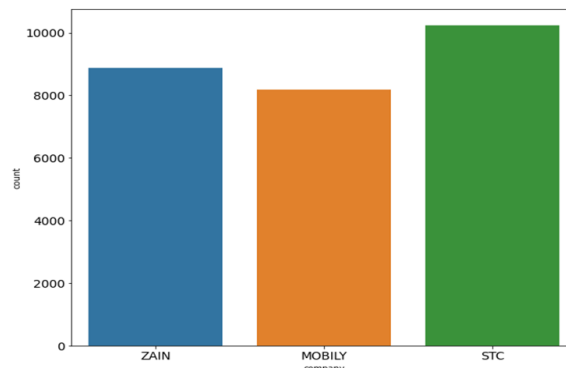


Figure 3. Companies distribution of Sara Dataset

4.3. Data set preprocessing

Several preprocessing techniques are applied to prepare Sara-dataset to be suitable for deep learning models. In this thesis we did the following pre-processing:

1. Removing re-tweets,
2. Removing diacritics,
3. Removing punctuations and removing repeating characters,
4. Normalization and removing URLs,
5. Creating word tokens
6. Removing Arabic stop words.

Our dataset consists of 50532 tweets, where after per-processing and preparation it becomes 27294 tweets as depicts in Figure 3. Table 2 shows the original tweet before pre-processing including special characters (@), numbers (3). And how pre-processing filter the tweets form diacritics, URLs and Arabic stop words.

Table 2
Original tweets and cleaning tweets during preprocessing


Original tweets	Cleaning tweet
Mobily We demand compensation for the period of service interruption for 3 days \$@	We demand compensation for the period of interruption of service for 3 days
@The internet has been weak for two days 	The internet was weak for two days

Figure 4, depicts classes distribution amongst the three companies companies which are Zain, Mobily and STC.

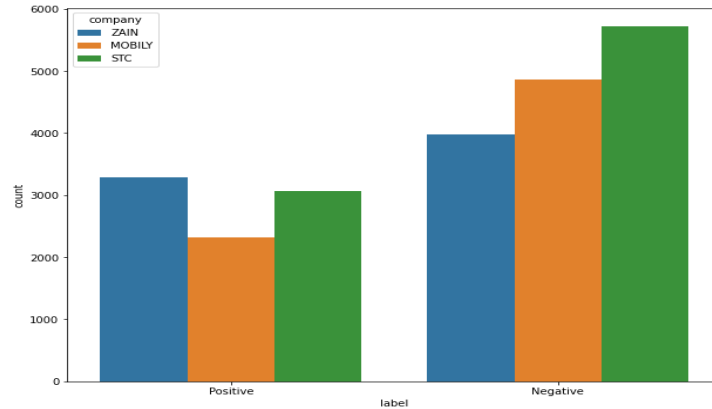


Figure 4. Classes distribution amongst companies

It should mention that after splitting the collected data into training, validation and testing sets. Training dataset makes up 70% of the overall Sara-dataset. Validation dataset makes up 10% of the overall Sara-dataset. Testing dataset makes up 20% of the overall Sara-dataset dataset as depicts in Table 3.

Table 3 Sara-Dataset splitting.

Total	Training (70%)	Validation(20 %)	Testing(10%)
23752	22927	3275	6550

As depicts in Table 4, the dataset is consist of 13 features including Id, Text, author_id, Language, City, Created data, company, tweet_count listed_count, Lable, is_retweet, followers_count and label (classes).

Table 4
Features of Sara dataset

Id	Text	Author_id	Language	City	Created data	Label
Company	Tweet_count	Listed_count	Lable	Is_retweet	Followers_count	

An example of some features including Id, Text, Language, City, company and Label. Where our target country is Saudi Arabia in Riyadh city is showed in Table 5.

Table 5
The features of Sara dataset

Id	Text	Language	City	Company	Label
1.55365E+18	“Peace be upon you, Zain’s internet is currently very weak, abundant and heavy It is stuck and there is a malfunction now on Zain Viber and the bandwidth is very bad.”	Arabic	Riyadh	Zain	Negative

4.3. Performance Measurement

Several performance metrics are typically used to evaluate the classification performance of Arabic sentiment analysis detection. Namely, they include Accuracy, Precision, Recall, F-measure and Area Under Curve. As the detection can be positive or negative. Positive classification is consisted of :

- True positive (TP): positive instance classified as positive.
- True negative (TN) : negative instance classified as negative.

Incorrect classification is consisted of:

- False positive (FP): negative instance classified as positive.
- False negative (FN): positive instance classified as negative.

This yield the confusion matrix shown in Table 6.

Table 6
Confusion Matrix in Natural langage Processing.

Classification	Positive	Negative
Positive records	TP	FN
Negative records	FP	TN

The Accuracy, Precision, Recall, and F-measure performance metrics are calculated based on the confusion matrix. Thus, the resulting performance measures are defined as follows:

- Accuracy is estimated by dividing the total correctly classified records (True positives and True negatives) by the total number of samples [1, 2]:

Accuracy =

$$\frac{\text{TruePositives} + \text{TrueNegative}}{\text{TruePositives} + \text{TrueNegatives} + \text{FalsePositives} + \text{FalseNegatives}} \quad (1)$$

Precision [1, 2] is the ratio of the number of relevant records retrieved (just True Positive) to the total number of irrelevant and relevant records retrieved (True positives and False negatives). Also, it is defined as the number of true positives over the number of true positives plus the number of false positives:

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \quad (2)$$

- Recall [1, 2] is the ratio of the number of relevant records retrieved to the total number of relevant records. it is defined as the number of true positive over the number of true positives plus the number of false negative:

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \quad (3)$$

One should mention that high precision means that an instances had been retrieved are finally more relevant results than irrelevant, while high recall means that the most of relevant results are retrieved.

- F-measure [12] is a measure of a test's accuracy for binary classification. As seen in Eq(9), The F-measure can interpret as a weighted average of the precision and recall, where an F-measure score reaches its best value at "1" and worst at "0".

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

- A Receiver Operator Characteristic (ROC) represents the true positive rate (TPR) against the false positive rate (FPR). Plotting sensitivity (true positive rate) versus specificity (false positive rate) for the different values given results in the graphical ROC curve [11, 12].

Table 7 provides a summary of the considered performance measurements such as Accuracy, Precision, F-measure, R-call, ROC, and loss value for the Arabic sentiment

analysis for Saudi Arabia communications companies. We choose three models CNN, and LSTM and the combination between them as ensemble learning.

Table 7
Models, Data set, and performance Measurements

Model	Dataset	Measurement
CNN	Sara Dataset	Accuracy, Precision, F-measure, R-call, ROC, and loss value for the positive and negative classes
LSTM		
Ensemble Learning		

4. RESULTS AND DISCUSSION

In this section, we disuse the results obtained by developing Arabic sentiment analysis using three models CNN, and LSTM and the combination between them as ensemble learning. These models applied on Sara-dataset.

4.1 Results obtained using CNN

We apply the CNN model to the Sara_Dataset to develop a Arabic sentiment analysis model. The dataset belongs to three Saudi Arabian communication firms that are improving the quality and quantity of their services by collecting customer feedback. We use different hyper-parameters, such as epoch, with different numbers, different batch sizes, and different optimizers. We evaluate our model in terms of accuracy, precision, F-measure, R-call, ROC, and loss function, as shown in Table 8. The accuracy of the best result, which was achieved using the Adam optimizer, was 96.34% using a 64-batch size and 96.83% using a 32-batch size. We use different optimizer such as Adadelta, Adagrad and Adam. Also, the best training accuracy is 97.37% when we use Adam optimizer and 128-Batch size.

Table 8
Measurement Performance for CNN Model

Model	Epochs	Batch Size	Optimizer	Training	Validation	Test			
						Accuracy	Precision	Recall	F-measure
CNN	10	32	Adadelta	62.61	62.98	47.14	22.22	47.14	30.2
			Adagrad	83.96	83.38	86.63	87.27	86.63	86.62
			Adam	96.94	95.48	96.83	96.83	96.83	96.83
		64	Adadelta	62.61	62.98	47.14	22.22	47.14	30.2
			Adagrad	67.12	68.63	61.34	77.23	61.34	56.12
			Adam	97.21	94.75	96.34	96.34	96.34	96.34
		128	Adadelta	62.6	62.98	47.14	22.22	47.14	30.2
			Adagrad	62.67	62.98	47.14	22.22	47.14	30.2
			Adam	97.37	94.75	94.38	94.58	94.38	94.36

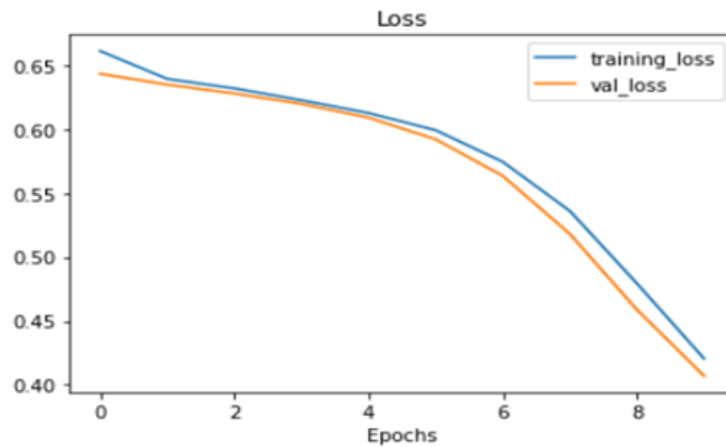


Figure 5. Trainig and validationl loss

Additionally, as shown in Figure 6 as epochs increased the the loss of training and validation are decreased. When we use the Adagrad optimizer with epoch 10 the with a 32-batch as size is the traing accuracy and validation acuurcy around 87% as depicted in Figure.

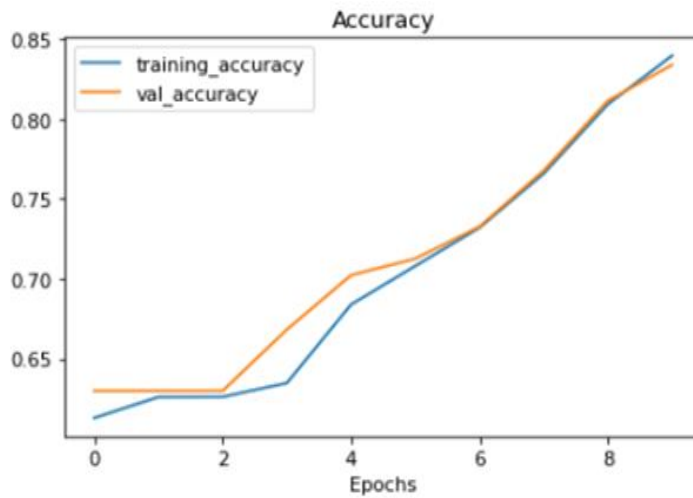
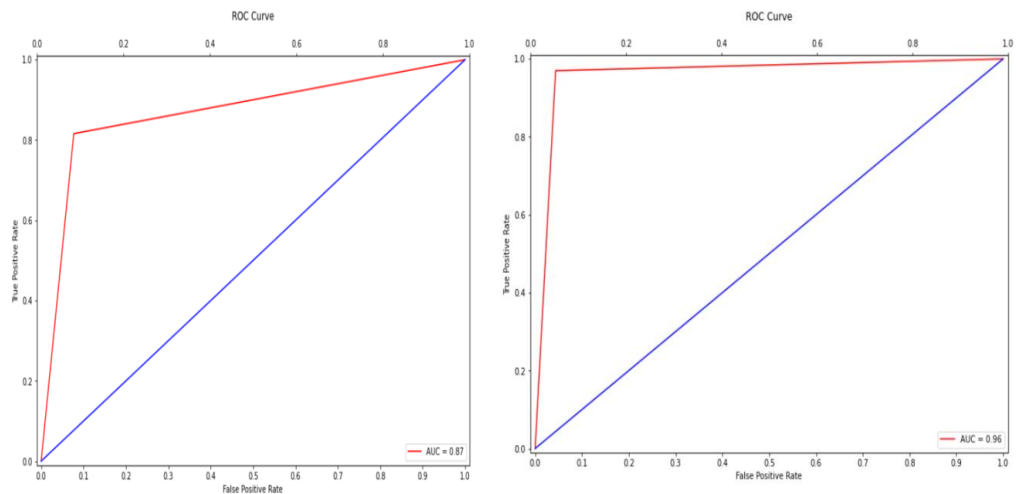


Figure 6. Training and validation accuracy.

The Roc curve of CNN model, using Adagrad and Adam optimizer, epoch=10, and with batch-size 32 and 64, are 87% and 96% respectively as depicts in Figure 7.



a. Batch size 32 Adagrad
Figure 7. ROC curve of CNN

b. Batch size 64 with Adam

5.2. Results obtained using LSTM

We apply the LSTM model to the Sara_Dataset to develop a Arabic sentiment analysis model. The dataset belongs to three Saudi Arabian communication firms that are improving the quality and quantity of their services by collecting customer feedback. We use different

hyper-parameters, such as epoch, with different numbers, different batch sizes, and different optimizers. We evaluate our model in terms of accuracy, precision, F-measure, R-call, ROC, and loss function, as shown in Table 9. Additionally, the best accuracy when we using LSTM is 94.97% with Adam optimizer, epoch 10, and batch size is 32. We use different optimizers such as Adadelata, Adagrad and Adam.

Table 9
Measurement Performance for LSTM Model

Model	Epochs	Batch-size	Optimizer	Training	Validation	Test			
						Accuracy	Precision	Recall	F-measure
LSTM	10	32	Adadelata	62.6	62.98	47.14	22.22	47.14	30.2
			Adagrad	62.64	63.14	47.14	22.22	47.14	30.2
			Adam	95.82	95.03	94.97	94.97	94.97	94.97
		64	Adadelata	62.64	63.05	47.14	22.22	47.14	30.2
			Adagrad	62.6	62.98	47.14	22.22	47.14	30.2
			Adam	95.45	94.27	93.47	93.53	93.47	93.46
		128	Adadelata	62.47	62.94	47.26	63.36	47.26	30.56
			Adagrad	62.6	62.98	47.14	22.22	47.14	30.2
			Adam	94.85	93.22	93.79	93.87	93.79	93.78

The Figure 8 shows the training and validation loss using different batch size. As epochs increase the loss function is decreased.

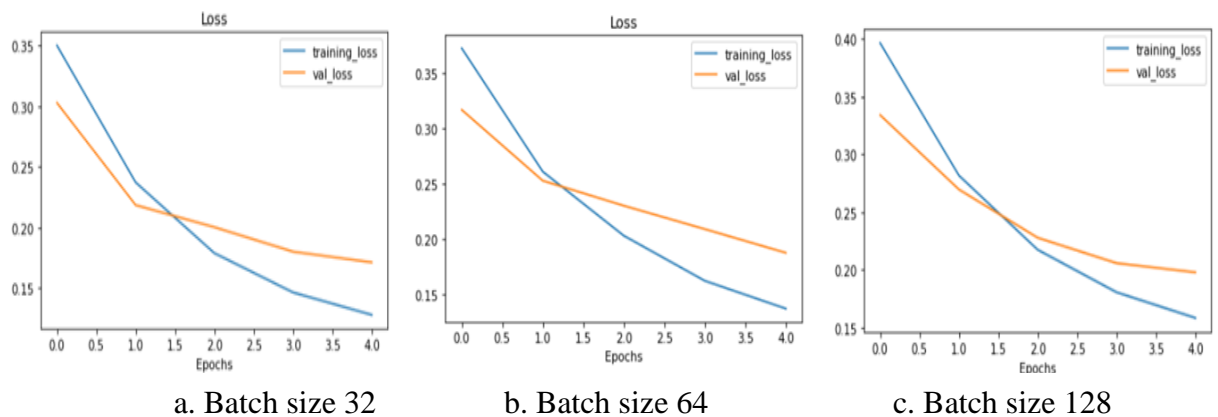


Figure 8.
The training and validation loss of LSTM

Figure 9, Depicts the confusion matrix using LSTM model with different batch size.

Where the best confusion matrix when using Adam optimizer with batch size 32

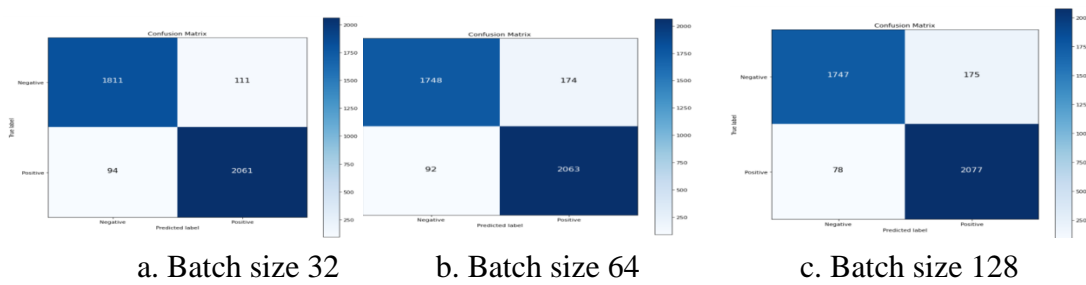


Figure 9. The Confusion matrix of LSTM Model

Figure 10, shows the relation between true and False positive which called ROC curve. Where is the curve of LSTM equal 0.95.

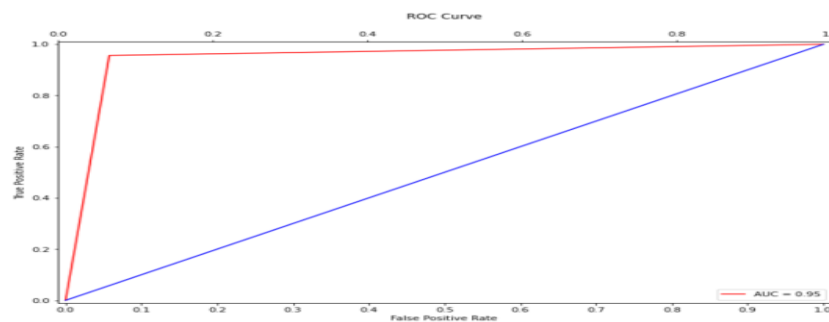


Figure 10. The Roc curve of LSTM Model

5. CONCLUSIONS

Sentiment analysis is an artificial intelligence technique that employs techniques to analyze whether an opinion is positive or negative. It's a powerful tool in the election process and social media to classify people's opinions towards things. In this article we created and developed a new dataset called, Sara-Dataset. It is considered a new Arabic dataset for three Saudi Arabian communication firms (Zain, Mobily, and STC) to improve the quality and quantity of their services by collecting customer feedback. Total records of the dataset after applying the preprocessing techniques is 27294. We developed Arabic several Arabic sentiment analysis models to detect customers opinion about communication firms. In our model, we used deep learning including LSTM and CNN to developed Arabic sentiment analysis models. We evaluate the model results in terms of validation, training and testing accuracy, precision, recall, F-measure, ROC curve and loss function. We use ensemble

learning techniques to improve the overall accuracy of Arabic sentiment analysis. Also, we used several number of hyper-parameter including training epochs, the loss function, and the optimizer. Finally, we combine LSTM with CNN model using voting and average techniques. Our results showed that the best accuracy using CNN is 96.83% with Adam optimizer, epoch=10 and batch size 32. Additionally, the best accuracy when we using LSTM is 94.97% with Adam optimizer, epoch 10, and batch size is 32. The best accuracy when we using ensemble learning is 95.81% with Adam optimizer, epoch 10, and batch size is 64.

Conflicts of Interest: “The authors declare no conflict of interest.”

REFERENCES

- Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums." *ACM transactions on information systems (TOIS)* 26.3 (2008): 1-34.
- Abdullah, Malak, and Mirsad Hadzikadic. "Sentiment analysis on arabic tweets: Challenges to dissecting the language." *International Conference on Social Computing and Social Media*. Springer, Cham, 2017.
- Al-Ayyoub, M., Khamaiseh, A. A., Jararweh, Y., & Al-Kabi, M. N. (2018). “A comprehensive survey of arabic sentiment analysis.” *Journal of Information processing management*, 56 (2019): 320-342.
- Aldayel, Haifa K., and Aqil M. Azmi. "Arabic tweets sentiment analysis—a hybrid scheme." *Journal of Information Science* 42.6 (2016): 782-797.
- Alayba, Abdulaziz M., Vasile Palade, Matthew England, and Rahat Iqbal. "Arabic language sentiment analysis on health services." In *2017 1st international workshop on arabic script analysis and recognition (asar)*, pp. 114-118. IEEE, 2017.
- Al-Smadi, Mohammad, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews." *International Journal of Machine Learning and Cybernetics* 10, no. 8 (2019): 2163-2175.
- Alayba, Abdulaziz M., Vasile Palade, Matthew England, and Rahat Iqbal. "A combined CNN and LSTM model for arabic sentiment analysis." In *International cross-domain conference for machine learning and knowledge extraction*, pp. 179-191. Springer, Cham, 2018.
- Almutairi, Sara Manour, and Fahad Mazead Alotaibi. "A Comparative Analysis for Arabic Sentiment Analysis Models In E-Marketing Using Deep Learning Techniques." *Journal of Engineering and Applied Sciences* 10.1 (2023): 19-19.
- Altaher, Altyeb. "Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting." *Int. J. Adv. Appl. Sci* 4.8 (2017): 43-49.

- Bisong, E. and Bisong, E., 2019. Google colabatory. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, pp.59-64.
- Elzayady, Hossam, Khaled M. Badran, and Gouda I. Salama. "Arabic Opinion Mining Using Combined CNN-LSTM Models." *International Journal of Intelligent Systems & Applications* 12.4 (2020).
- Guellil, Imane, Faical Azouaou, and Marcelo Mendoza. "Arabic sentiment analysis: studies, resources, and tools." *Social Network Analysis and Mining* 9.1 (2019): 1-17.
- Gulli A, Pal S. *Deep learning with Keras*. Packt Publishing Ltd; 2017 Apr 26.
- Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic tweets using deep learning." *Procedia Computer Science* 142 (2018): 114-122.
- Khalil, Enas A. Hakim, Enas MF El Houby, and Hoda Korashy Mohamed. "Deep learning for emotion analysis in Arabic tweets." *Journal of Big Data* 8.1 (2021): 1-15.
- Karthika, P., Murugeswari, R. and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5, doi: 10.1109/INCOS45849.2019.8951367.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Hernegger, M., "Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*", 22(11), pp. 6005-6022, 2019
- Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J., 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.
- Mdhaffar, S., Bougares, F., Esteve, Y., & Hadrich-Belguith, L. (2017, April). Sentiment analysis of tunisian dialects: Linguistic resources and experiments. In *Third Arabic Natural Language Processing Workshop (WANLP)* (pp. 55-61).
- Nassif, Ali Bou, Ashraf Elnagar, Ismail Shahin, and Safaa Henno "Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities." *Journal of Applied Soft Computing* 98 (2020): 106836.
- Rasool, Abdur, et al. "Twitter sentiment analysis: a case study for apparel brands." *Journal of Physics: Conference Series*. Vol. 1176. No. 2. IOP Publishing, 2019.
- Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." *Physica D: Nonlinear Phenomena* 404 (2020): 132306.
- Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation." *Australasian joint conference on artificial intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.