Emotion detection of Arabic Tweets using Arabic Transformers

Mohamed Abdel Mohsen
Faculty of Computers and Data Science, Alexandria University
cds.Mohamedabdelmohsen77929@alexu.edu.eg

Mahmoud Khalifa
Faculty of Computers and Data Science, Alexandria University
Cds.MahmoudKhalifa12770@alexu.edu.eg

Habiba Reda
Faculty of Computers and Data Science, Alexandria University
cds.Habibareda85965@alexu.edu.eg

Omnya Ramadan
Faculty of Computers and Data Science, Alexandria University
cds.OmnyaRamadan14184@alexu.edu.eg

Manar Mohamed
Faculty of Computers and Data Science, Alexandria University
CDS.manarmohamed16957@alexu.edu.eg

***Abstract*— Social media platforms have become an essential means for communicating feelings to the entire world due to rapid expansion in the Internet era and Nowadays people can express their feelings and thoughts using pictures, videos, audio, or textual content and this led to a massive amount of unstructured data generated from this Social media platforms and this data are important and could be used in different manners as we can preprocess this data and perform sentiment analysis or emotion detection that might operate on single sentences, paragraphs, or even entire articles and our goal is to know the feelings of the author such as sadness, anger, fear, sympathy, surprise, etc. The Arabic Egyptian Dialect Nile University Dataset, comprised of 10,065 tweets, The tweets of NU were categorized into 8 emotions (sadness, anger, fear, sympathy, joy, surprise, love, and none) The term "none" was used to label neutral tweets. The tweets were collected using the" Olympics" hashtags from Egypt in the period between Jul 2016 and Aug 2016. Emotion analysis could help decision makers understand and respond to the public reactions and A combination of Arabic text preprocessing techniques was tested with machine learning and deep learning models are used in this paper to achieve this goal**

I. INTRODUCTION

There is a growing interest in using emotion analysis on the Arabic text to understand public reactions and this could be useful in many fields such as monitoring people's mental health, especially after crises and pandemics such as covid-19 where Surveys by such organizations as YouGov show that, during the pandemic, workers were more stressed than ever. This included anxiety regarding health, finance, and employment. We can find numerous amounts of articles posted every day on different social media platforms so, we can use Arabic text data in different applications that will be helpful in our society.

The Arabic language is now facing many problems due to the increase in verb conjugations, the formation, and the presence of more than one word that leads to the same meaning, also, Arabic texts may include many translated and transliterated named entities whose spelling, in general, tends to be inconsistent in Arabic texts, For example, a named entity such as the city of Washington could be spelled واشنطن/وشنطن/وشنغطن/وشنطجن, The form of characters and spelling of words can vary depending on their context and the Same word may have much different meaning due to the Diacritics and many other problems that we should put under considerations when we want to perform emotion classification in Arabic text including the pre-processing techniques and the models used so that, we can get acceptable accuracy that allows for using it in real-life applications. The purpose of this study is to Compare state-of-the-art Arabic transformer models in a large Arabic dataset and improve their results using Different combinations of prepossessing techniques

The rest of this paper is structured in the following manner. The related work is reviewed in Section II. The research approach and the methods used are explained in Section III. In Section IV, Results of Arabic transformer and machine learning models are presented and the insights are discussed. Finally, the conclusion and possible future directions are discussed in Section V.

I I . RELATED WORK

In this section I will show a brief review on the previous work that have been done before to the emotion classification of Arabic Twitter data related to different topics.

The top results on the Nile University data were on Ahmed El-Sayed, Shaimaa Lazem and Mohamed Abougabal [22] work with 75.88% accuracy, 75.6% precision, 75.7% recall and 75.5% F1 score achieved by AraBert word embedding for feature extraction and the weighted voting ensemble of Logistic Regression, GRU and AraBERT.

Al-Khatib and El-Beltagy [8] best results obtained using a Complement Naive Bayes algorithm with accuracy 68.12%

Essam et al. [10] performed emotion classification in the dataset collected in Al-Khatib and El-Beltagy [8] and they used preprocessing techniques onArabic text as shown in Table I and tested them with multiple classic machine learning and deep learning models and the best accuracy achieved was 72% using LSTM and 69.41% using Logistic Regression. A comparison of the related work on different papers was shown in Table I and II.

TABLE I: The pre-processing techniques used in the literature.

Pre-processing	[8]	[9]	[3]	[10]	[11]	[13]	[22]	**Ours**
Year	2017	2018	2019	2019	2020	2021	2021	2022
Normalization	√	√	√	√	√	√	√	√
Removal of stop words	X	X	√	√	√	X	√	√
Removal of non-Arabic letters and spaces	√	X	X	√	√	√	√	√
Diacritics removal	√	√	√	√	X	X	√	√
Links, mentions, and retweet indicators removal	√	√	X	√	√	√	√	√
Remove suffixes and prefixes	X	X	X	√	√	X	√	√
Disapprobation words	X	X	X	√	√	X	√	√
Reducing words to their roots	X	X	√	√	√	X	√	√
Occurrence removal	X	√	√	√	X	X	√	√
Negation patterns	X	√	X	√	X	X	√	√
Punctuation	X	√	√	√	X	√	√	√
Replace emojis and emoticons with emotions	X	X	X	X	X	X	√	√

III. APPROACH

Our aim is to use Arabic text which represents the opinions expressed by Egyptian citizens and the emotion representing this text to analyze and Compare state-of-the-art Arabic transformer models and improve their results using Different combinations of prepossessing techniques as shown in Table I.

TABLE II: Emotion analysis related work summarized.

Features			[8]	[9]	[3]	[10]	[11]	[13]	[22]	**Ours**
Year			2017	2018	2019	2019	2020	2021	2021	2022
Dataset	Tweets		√	√	√	√	√	√	√	√
Feature Extraction	TF-IDF		X	X	X	√	√	√	√	√
Emotion Classification	Classical Machine Learning	SVM	X	X	√	√	√	X	√	√
		Logistic Regression	X	X	X	√	√	X	√	√
		Decision Tree	X	X	X	√	X	X	√	√
		Random Forest	X	X	X	√	X	X	√	√
	Deep Learning	MarBERT	X	X	X	X	X	X	X	√
		ArBERT	X	X	X	X	X	X	X	√
		AraBERT-Twitter	X	X	X	X	X	X	X	√
		AraBERT	X	X	X	X	X	X	√	√
		Qarib	X	X	X	X	X	X	X	√
Performance Evaluation	Accuracy		68.1%	64.75%	99.9%	69.4%	74%	X	75.8%	77%
	Precision		68.8%	X	99.9%	70%	73%	X	75.6%	76.8%
	Recall		68.1%	X	99.9%	69%	74%	X	75.7%	76.2%
	F-score		65.8%	62.08%	99.9%	68%	73.2%	71%	75.5%	76.2%

A. Dataset

The dataset was collected by a research group at Nile University (NU) [8]. It is a balanced dataset consists of 10,065 Arabic tweets mostly using Egyptian dialect, and was manually annotated using eight emotions (sadness, anger, joy, surprise, love, sympathy, fear and none). It was collected using the" Olympics" hashtags. The dataset was split into 80% for training and 20% for testing.

B. Experimental Design

Data pre-processing is important to minimize the noise and get a better classification accuracy. We have done 3 experiments on this dataset where each of them started by using the pre-processing techniques in Table I with some small changes, in the first experiment we removed the emojis and emoticons from the dataset and we used this data in Arabic Transformer models to measure the performance and in the second experiment we replaced emojis and emoticons with an equivalent Arabic emotion to see that if this step will give us better accuracy, and in the last experiment we replaced each emoticon with it's equivalent emoji and we kept the emojis in the dataset as it is and we passed this data in AraBERTv0.2-Twitter model as this model have had emojis added to their vocabulary and the model is expected to handle the emojis in the dataset.

The implementation of the work in [8], and [10] is not publicly available. Thus, both were replicated employing the NU dataset reported in [8]. The replication was verified by comparing its results to the results reported in both papers. In terms of feature extraction, Term Frequency-Inverse Document Frequency (TF-IDF) [16], each one was used along with the proposed pre-processing technique.

Classical machine learning techniques were tested, mainly SVM, Logistic Regression, Decision Tree and Random Forest as illustrated in Table II. As for deep learning models, we used AraBERT-Base, AraBERT-Twitter, MarBERT, ArBERT and Qarib To evaluate the performance of the different classifiers, accuracy, precision, recall, and F-score were used. For each classifier, the error metrics were calculated for each individual label (joy, sadness, sympathy, anger, fear, surprise, love, none). The final error metrics were calculated as the average of all the metrics for all the labels.

V. RESULTS & DISCUSSION

As we discussed before we have 3 experiments all the experiments have the same preprocessing in Table I but the only difference was in handling the emojis where in experiment one we removed emojis and emoticons from the dataset and in experiment two we mapped the emojis and emoticons into equivalent Arabic text and experiment three we converted each emoticon into the equivalent emoji and we kept emojis without mapping or removing it to use this data in AraBert Twitter model.

According to the classical machine learning models the best results was in experiment 2 as expected and the best performance was logistic regression model with 67% accuracy, 76.8% Precision, 76.2% Recall and 76.2% F- score.

But also note that in case of classical machine learning models' performance in experiment 1 or 2 the difference in the accuracy is not more than 1% increase as shown in Tables IV & V.

Our top results were when we used MarBERT model in experiment 2 as expected we got 77% accuracy, 76.8% Precision, 76.2% Recall and 76.2% F- score as shown in Tables VII and these results are the best among all the previous work on this dataset.

In case of the results of Arabic Transformers in experiment 1 the best performance was from MarBERT and AraBERT-Twitter with accuracy 75% so, AraBERT-Twitter performed better when in experiment 2 than experiment 3 as shown in Table VII & VIII

TABLE III: The hyper parameters tuning used in the literature.

Hyper-parameter	**AraBERT**	**AraBERT-Twitter**	**MarBERT**	**ArBERT**	**Qarib**
hidden state dimension	768	768	768	768	768
Learning rate	2e-6	2e-6	1e.6	2e-6	1e-6
Number of epochs	20	25	20	20	30
Batch size	128	128	128	128	50

V. CONCLUSION AND FUTURE WORK

The objective of the paper is to try different Arabic Transformers and try to improve the state-of-the-art accuracy of Arabic emotion classification. Our results show that the eight emotions could be detected in Arabic text with an overall accuracy of 77% in MARBERT model.

In the future, we want to make an application analyzes the feedback of the customers in products to improve the quality of it and other an application talking about analyzing the most common emojis that people use on Facebook to predict their psychological health and we want to use different transformers like ULMfits and try different techniques to improve the accuracy

TABLE IV: Classical machine learning results (Experiment 1)

Model	TF-IDF			
	AC	PR	RE	F1
Support Vector Machine	66%	70%	%64.9	64.5%
Logistic Regression	67%	68.9%	%66.25	66.1%
Decision Tree	55%	55.6%	%55.4	55.5%
Random Forest	64%	65.3%	%63.5	62.4%

TABLE V: Classical machine learning results (Experiment 2)

Model	TF-IDF			
	AC	PR	RE	F1
Support Vector Machine	65.4%	67.3%	63.4%	63.1%
Logistic Regression	66%	67%	64.5%	64%
Decision Tree	54%	54.6%	53.9%	54.1%
Random Forest	65%	64.3%	64%	63.1%

TABLE V: Classical machine learning results (Experiment 2)

TABLE VI: Arabic Transformer models (Experiment 1)

Model	Performance			
	AC	PR	RE	F1
AraBERT	73%	74%	74.9%	74.75%
AraBERT-Twitter	75%	74.8%	74.3%	74.3%
MarBERT	75%	74.9%	74.9%	74.8%
ArBERT	71%	71.4%	71%	71.1%
Qarib	74%	74.6%	74.1%	74.3%

TABLE VI: Arabic Transformer models (Experiment 2)

Model	Performance			
	AC	PR	RE	F1
MarBERT	77%	76.9%	76.3%	76.3%
Qarib	74%	-	-	-

TABLE VII: Arabic Transformer models (Experiment 2)

TABLE VI: Arabic Transformer models (Experiment 3)

Model	Performance			
	AC	PR	RE	F1
Arabert Twitter	%74	%74	73.8%	73.8%

REFERENCES

[1] M. Baali, N. Ghneim,” Emotion analysis of Arabic tweets using deep learning approach”, Journal of Big Data, 2019.

[2] V. Tripathi, A. Joshi, P. Bhattacharyya,” Emotion analysis from text: A survey”, Center for Indian Language Technology Surveys, 2016.

[3] M. El-Masri, N. Altrabsheh, H. Mansour,” Successes and challenges of Arabic sentiment analysis research: a literature review”, Social Network Analysis and Mining, 2017.

[4] H. Alhazmi, M. Alharbi,” Emotion Analysis of Arabic Tweets during COVID-19 Pandemic in Saudi Arabia”, International Journal of Ad- vanced Computer Science and Applications (IJACSA), 11(10), 2020.

[5] A. Al-Laith, M. Alenezi,” Monitoring People’s Emotions and Symptoms from Arabic Tweets during the COVID-19 Pandemic”. Information, 12, 86, 2021.

[6] A. Al-Khatib, S. El-Beltagy,” Emotional Tone Detection in Arabic Tweets”, 18th International Conference on Computational Linguistics and Intelligent Text Processing, 2017.

[7] M. Heikal, M. Torki, N. El-Makky,” Sentiment Analysis of Arabic Tweets using Deep Learning”. Procedia Computer Science, 2018.

[8] A. Essam, H. Raafat, N. Hazem, S. Samer,” Towards The Early Detection of Depression using Sentiment Analysis of Arabic Tweets”, Graduation Project Report, Computer and Systems Engineering Depart- ment, Alexandria University, 2019.

[9] M. A. Al-Hagery, M. A. Al-Assaf, F. M. Al-Kharboush,” Exploration of the best performance method of emotions classification for arabic tweets”, Indonesian Journal of Electrical Engineering and Computer Science, vol. 19, no. 2, pp. 1010–1020, 2020.

[10] A. Gruzd, (2016). ”Netlytic: Software for Automated Text and Social Network Analysis” [Online]. Available: https://netlytic.org/.

[11] N. Al-Twairesh,” The Evolution of Language Models Applied to Emo- tion Analysis of Arabic Tweets”, Information, 12, p. 84, 2021.

[12] A. Safaya, M. Abdullatif, D. Yuret,” KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media”, In Proceedings of the Fourteenth Workshop on Semantic Evaluation; International Committee for Computational Linguistics, 2020.

[13] B. Talafha, M. Ali, M. E. Za’ter, H. Seelawi, I. Tuffaha, M. Samir, W. Farhan, H. T. Al-Natsheh,” Multi-Dialect Arabic BERT for Country- Level Dialect Identification”. In Proceedings of the Fifth Arabic Natural Language Processing Workshop (WANLP), 2020.

[14] C. Sammut, G. I. Webb,” TF–IDF”, In Encyclopedia of Machine Learning; Springer: Boston, MA, USA, 2010.

[15] M. Tomas, G. Corrado, K. Chen, J. Dean,” Efficient Estimation of Word Representations in Vector Space”, 2013.

[16] P. Jeffrey, R. Socher, C. Manning,” Glove: Global Vectors for Word Representation”, Empirical Methods in Natural Language Processing (EMNLP), 2014.

[17] A. Wissam, B. Fady, H. Hazem,” AraBERT: Transformer-based Model for Arabic Language Understanding”, Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection 2020.

[18] S. Hochreiter, J. Schmidhuber,” Long Short-Term Memory”, Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] Sh. Apeksha, D. Nyavanandi, L. Simone,” Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU”, Journal of Artificial Intelligence and Soft Computing Re- search, 2019.

[20] S. Albawi, T. A. Mohammed, S. Al-Zawi,” Understanding of a convoy- lutional neural network”, International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017.

[21] M. Abdul-Mageed, A. Elmadany, E.M.B. Nagoudi,” ARBERT & MAR- BERT: Deep bidirectional transformers for Arabic”, In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021.

[22] Ahmed El-Sayed, Shaimaa Lazem and Mohamed Abougabal, An Improved Emotion-based Analysis of Arabic Twitter Data using Deep Learning