Emotion Detection Model in Different Saudi Dialects

by

Ruba Alsulami - 2110692
Rahaf Alfudhayl - 2111658
Ghadi Aljuhani - 2111982
Shada Basudan -2114824
Joud Samkari - 2113282

Artificial Intelligence Department

College of Computer Science and Engineering
JeddahUniversity, SAUDI ARABIA

Supervisor
D.Rehab Qarout

(6 May 2024)

Abstract

Detecting emotions is essential in human communication and can help us with various applications. Our project focused on discovering emotions in the Saudi dialect using the latest natural language processing techniques. Our main goal was to identify emotional text written in the Saudi dialect, including feelings of anger, sadness, fear, surprise, neutrality, disgust, and happiness, sarcasm, and create a simple platform that displayed the text along with its emotional classification. This method helped us comprehend and interpret a given text's emotional content. By analyzing emotions from text, we could determine how people felt about a product, company, or restaurant and their opinions on a particular subject. The data we used was from the X platform so that we could use and analyze people's posts. Most lemmatization and stemming tools were based on Modern Standard Arabic (MSA), but posts on social media apps in Arabic were often in Arabic dialects, not MSA. This made preprocessing and noise filtering more challenging. The emotion analysis model was applied to classify emotions using the MARBERT model, which was a deep learning model pre-trained based on both MSA and Arabic dialects and had high performance compared to other models. One of the challenges in this field was that people used different dialects in social media, which made analyzing the text difficult because no fixed rules and vocabulary could be followed. We applied MARABERT to Saudi dialects data, giving us high accuracy, with 81% for detecting emotion.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Emotions are an essential part of human life. They are often defined as a complex pattern of reactions[1]. Emotions are an important section of language and are complex to detect. The Arabic language is one of the languages classified as complex because it has many dialects and because of its complex script. Analyzing people's feelings and opinions on a particular topic in this country is essential. Emotion detection involves identifying and classifying individuals' emotions through social media posts(previously tweets)[2]. It helps take customer opinions about the products or services provided by entities and companies seeking to improve their offerings[3]. Many people use social media to express their opinions on specific topics. Emotion detection will help gain insight into public opinion, saving time and effort for many people. Detecting emotions in Arabic posts on X Platform[1] is difficult because these posts are written with linguistic and spelling errors, even with grammar, especially since the Arabic dialects do not follow the Modern Standard Arabic morphology. Many researchers in natural language processing have studied emotion detection in English. However, few of them have studied it in Arabic because of the obstacles it poses as a different language. Most feeling analyses of Arabic posts are focused on classifying a sentiment into a positive or negative feeling but do not go deeper[4]. Until recently, the Arabic language still suffers from a lack of studies despite the recent growing interest in revealing emotions and processing the Arabic language. There is an urgent need to develop models to detect emotions on Arabic social media, especially in the dialects of this region. As we mentioned

[1]https://X.com

before, social media is a way for people to express their feelings, which makes it full of helpful information that may enable us to use it as a data source[5]. People's feelings on social media can also be analyzed and classified, which helps to know society's prevailing opinion on a particular topic[5][4]. Furthermore, we can use sentiment analysis or emotion detection to analyze and classify Arabic texts on social media, specifically those in Saudi dialects. This may help to know the general opinion of Saudi society on Saudi matters or affairs that have prevailed recently, such as the recent developments in the education sector or the health sector, the Riyadh Season, and the significant developments in tourism.

1.2 Problem Definition

This study aims to detect emotions in Arabic texts on the X platform[2], specifically the Saudi dialect. It is known that the Arabic language is one of the most dicult languages, and it may be challenging to analyze. Even more dicult that many dialects may not have xed rules or terminology that can be followed. Therefore, this may make it dicult to detect or analyze the feelings contained in the text. Given the need for published research on detecting emotions in texts in the Saudi dialect, we must work on this. The study aims to build a model capable of classifying texts in the Saudi dialect based on the emotions contained in the text. This proposed model may benefit decision-makers in the Kingdom of Saudi Arabia by knowing the community's opinions on some prevailing topics.

1.3 Aim of the Project

Our aim is to create a system that can identify different emotions in posts written in the Saudi Arabian dialect on the social media X platform[3]. We hope to understand better the emotions conveyed through social media content, specifically on the platform[4]. This project will contribute to the field of natural language processing for the Arabic language and its various dialects, especially the Saudi dialect.

[2]Ibid
[3]Ibid
[4]Ibid

1.4 Objectives

- **Objective 1:** Review studies on emotion detection in Arabic texts and understand the recent updates and techniques.
- **Objective 2:** Working on data sets consisting of Saudi dialects on X platform[5], such as preprocessing, classifying to labels, etc.
- **Objective 3:** Detecting emotions in Saudi dialect text using the MARABERT model.
- **Objective 4:** Evaluating the performance of an emotion detection model.

1.5 Plan (Gantt chart)

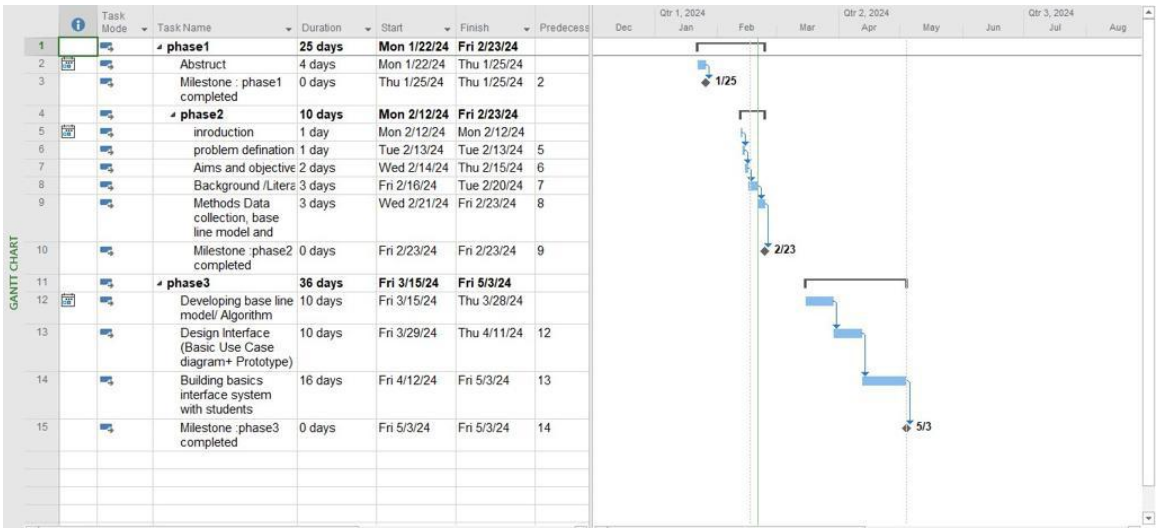

Figure 1.1: Gantt chart

[5]Ibid

Chapter 2

Background/Literature Review

2.1 Background

Historically, natural language processing for Arabic has gone through three waves. The first wave was in the 1980s, with Microsoft MS-DOS 3.30 and Arabic support, and this period was primarily focused on MSA. The second wave was in the early 2000s until 2010, when NLP of the Arabic language became essential to the Western world, especially after September 11. The USA funded large projects for companies and research centers to develop NLP tools for Arabic and its dialects. The third wave overlapped with the rise of deep-learning neural models and social media. This period witnessed a proliferation of Arab researchers interested in Arabic NLP and increased publications at top conferences from the Arab world[6].

Language is complicated, and processing it computationally is not direct. The principal building block of language is words; in natural language processing (NLP), we must turn words into a numerical format to create the right representation that can assist machines in recognizing language. All these dissimilar knowledge blocks have to be considered when we want to turn words into a numerical format to create a suitable representation that can help ML models recognize language and implement greater on the various NLP tasks and applications. Thus, to represent language, the vector space model is used, so words are represented as vectors of numbers. The dissimilar approaches for raising these vectors are language description models or language models (LM). The previous recurrent neural networks (RNN) had the restriction of considering the last words of a series. Nevertheless, transformers control this limitation by looking at all the words surrounding the

target word and giving more weight to essential wordsthis is known as self-attention. As for contextual language models, a multilingual BERT was released for non-English languages, which has been trained on Wikipedia dumps of 100+ languages, Arabic one of them. However, pre-training monolingual BERT for non-English languages has performed better than the multilingual BERT[7]. Therefore, Wissam Antoun proposed the AraBert model, a monolingual version of BERT for Arabic[8].

AraBERT was the initial Arabic-particular transformer-based language model. The start of AraBERT will help improve the performance of many Arabic NLP tasks. Recently, a large set of transformer-based Arabic language models has been developed. Also, there are MARBERT Arabic variants of other models that were released. While most of these models were trained on modern standard Arabic (MSA) data, some, such as MARBERT, included dialectal Arabic in their training data. Most of these models were evaluated on a small set of Arabic NLP tasks, and the models trained on dialectal Arabic are the most effective ones in the study, where the most effective model, MARBERT, reached an FPN of 0.724 on the SA task and an F1-sarcastic of 0.584 on the sarcasm detection[6].

In another paper [9], They pre-trained the BERT Model specifically for the Arabic language. Furthermore, they evaluated AraBERT on three Arabic NLU downstream tasks: Sentiment Analysis (SA), Named Entity Recognition (NER), and Question Answering (QA). In the pretraining process, they employ the Masked Language Modeling (MLM) task. This could improve the pre-training task by forcing the model to predict the whole word instead of getting hints from parts of the word. They also employ the Next Sentence Prediction (NSP) task that helps the model understand the relationship between two sentences. The dataset that has been used was manually scraped Arabic news, also two publicly available large Arabic corpora: the 1.5 billion words Arabic Corpus and OSIAN: the Open-Source International Arabic News Corpus. The size of the whole pretraining dataset is 24GB of text. In the Arabic language, words can have different forms but with the same meaning, this could lead to a large amount of redundancy. To solve this problem, they performed Sub-Word Units Segmentation. In the Sentiment Analysis task, both versions of AraBERT outperform multilingual BERT. In the Named Entity Recognition task, AraBERTv0.1 improved results by 2.53 points in F1 score, scoring 84.2 compared with the Bi-LSTM-CRF model, making AraBERT the new state-of-the-art for NER on AN-ERcorp. Finally, in the question-answering task, They report a 2% absolute increase in the sentence match score over multilingual BERT, which is the previous state-of-the-art[9].

The article[4] provides an overview of studies on emotion detection of Arabic posts. As mentioned before Several advanced architectures have been proposed since the release of initial transformer models: BERT, XLNet, GPT, RoBERTa, and ALBERT. Typically, transformers and their variants are pre-trained on large unlabeled datasets. Furthermore, pre-trained language models (PLMs) can be fine-tuned and used on a downstream task. The article discovered that most studies utilized either traditional ML with features, emotion lexicons, or deep learning. The study[4] mentioned some pre-trained models. First is the AraBERT model, which is a BERT-based model pre-trained in the Arabic language. AraBERTs applicability to tasks involving dialects is limited because it is pre-trained using MSA data. Also, the ArabicBERT model increases the amount of corpus used in the earlier AraBERT. Another model discussed in this study is ELECTRA, which contains two modules, a generator and a discriminator. In addition, AraGPT2 which is based on the original GPT2. Finally, the MARBERT model was trained to increase its capacity to handle dialectal Arabic[4].

2.2 Literature Review

In tracking the latest studies that discussed natural language processing, we noticed the confusion between sentiment analysis(SA) and emotion detection(ED) during the research process. Although they are two different fields of natural language processing, they overlap in some way, which creates this confusion.

Here we discuss the highlights of these studies: Developing System-based Machine Learning for Predicting and Analyzing Arabic Sentiment[3]. sentiment analysis is a specialized application within the domain of natural language processing (NLP) utilized to analyze textual input to identify people's feelings and emotions being expressed. In recent years, multiple methodologies based on Machine Learning (ML) Systems have been implemented for Arabic sentiment analysis (ASA), demonstrating the potential to accurately detect sentiment across various datasets. The objective of this research is to present a comparative investigation into the most effective approaches for conducting sentiment analysis of posts in the Arabic language by using five ML techniques(namely Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and XGBoost) by using ASTC dataset and the Artwitter dataset after the experiment in the two datasets it was noted that the Random Forest (RF) model had the highest level of accuracy 87.7% when applied

to the ASTC dataset. On the other hand, the SVM model obtained 90.3% accuracy using the Artwitter dataset.

Unveiling Sentiments: A Comprehensive Analysis of Arabic Hajj-Related Tweets from 20172022 Utilizing Advanced AI Models[10]. the objective of this paper is to present a comprehensive sentiment analysis of posts discussing the annual Hajj pilgrimage over six years by using deep learning (CNN, LSTM, BERT-MINI)and Machine learning (LR, RF, Naive Bayes, SVM, KNN, XGBoost) in manually collected Arabic posts after the experiment in Ml the random forest have the best accuracy = 89.77% and in deep learning the BERT-MINI have the best accuracy = 93.88%.

Exploring the Role of Emotions in Arabic Rumor Detection in Social Media[11]. this research detects rumors in Arabic by analyzing emotions using the latest natural language processing techniques. No data was ready for their research, so they collected it from X platform[1].[11] this paper used machine learning models to distinguish between rumors and non-rumors. The AraNet model performed excellently for news sentiment, which can be used in modern languages like the BERT model. On the other hand, the SenticNet model did not perform well. The RF classifier performs well when the number of features is increased in most cases. The results showed that combining emotions in news and comments improved performance. The emotional and textual features of comments also contribute to better detecting rumors.

Detecting emotions has a deeper classification by defining them more precisely and classifying them into more than one or three labels. For example, in emotion analysis, they are classified into (positive, negative, and normal), but in emotion detection, they are classified into (sadness, anger, fear, and joy). After deep learning and machine learning techniques, most recent research has come with combined models that perform better than individual models.

Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model[12]. In this research, the CNN-BiLSTM model is proposed. It combines two models to obtain better accuracy and extract information effectively from social media. The CNN component can capture local patterns in the text, such as word associations, while the BiLSTM component can capture long-term dependencies between words[12]. They created a dataset of 9 million posts in Arabic through the X API. They classified Arabic text based

[1]https://X.com

on sarcasm, emotion, feelings, and topic. The model achieved high accuracy where (f-measure, accuracy) were (97,97.58) % (84,86) % (95,97) % and (82,81.6) % respectively.

We also noticed that much of the research that discusses our topic relies on the SemEval-2018-Ar-Ec data set, including these two studies:

An Ensemble Deep Learning Approach for Emotion Detection in Arabic Tweets[5]. this research proposes a solution to the problem of emotion detection in Arabic text by proposing an ensemble deep-learning approach to analyze user-generated text from X. The research also talks about the challenges associated with the Arabic language, such as the complexity of the language, the lack of Arabic sources, and the differences in Arabic dialects. Also mentioned some surveys and studies that used deep learning and machine learning or pre-trained language models (PLMs) and ensemble techniques. Most studies used the SemEval-2018-Ar-Ec dataset. The use of PLMs has shown some advances in accuracy compared to other models in these studies. The proposed model is based on three advanced deep-learning models: BI-LSTM, BI-GRU, and MARABERT. The model for multi-label emotion has six layers: A preprocessing layer, a Word embedding layer, a Processing Layer, a Testing Layer, an Ensemble layer, and a Classification layer. The model showed superior performance over individual models (Bi-LSTM, Bi-GRU, and MARBERT) With Jaccard Score = 0.540, Precision = 0.634, F1 Score Macro = 0.701, which also used SemEval-2018-Ar-Ec dataset.

Deep learning for emotion analysis in Arabic tweets[13]. the research aimed to develop an emotion analysis model for categorizing emotions in Arabic social media text. Despite the increase in Internet users of the Arabic language, the study indicates that we urgently need to develop more accurate models of detecting emotions in Arabic texts in MSA and dialectal Arabic using a large-scale emotion lexicon. They discuss some studies, such as what model was used, how the data was processed, and how it was classified. From these readings, the model developed by the researchers was proposed. The model implemented a novel multilayer bidirectional long short-term memory (BiLSTM) trained on top of pre-trained word embedding vectors. The approach they followed to develop a framework for predicting users emotions from their posts is data preprocessing, which contains Initial preprocessing, stop word removal, creation of an emoji lexicon, and stemming; also, feature extraction improves the performance of different machine learning models. In addition, network architecture builds a deep learning model of recurrent neural network (RNN) and BiLSTM layers. A BiLSTM is a sequence model with two LSTMs: forwarding and backward

direction input. The data that they used in this research is SemEval-2018 Task1. The proposed model achieved about a 9% enhancement in validation accuracy compared to other models and achieved the best performance when using the emoji lexicon.

Chapter 3

Methodology

3.1 Dataset

3.1.1 Data Collection

We used 207,451 posts collected by "Musharraf Al-Ruwaili" in the research "Issues of Dialectal Saudi Twitter Corpus"[14]. We extracted 1,081 posts from them to use in our research. This data was not labeled; we solved this problem by manually voting for each post, and the label with the highest rating was assigned to the Post. We cleaned all noises that could affect the accuracy of the results. Then, we added 120 posts that we collected manually by searching on the X platform[1] using keywords that are related to each label and added them to our dataset to balance the data. There are many problems with posts on the platform[2] due to poor writing by many users and factors related to technology, culture, and social media. The original data size was so huge that we had problems cleaning it, and we reduced this data to what we saw fit to train the model on it. Many reasons can affect the quality of the results, the most important of which are punctuation marks such as (commas, full stops,...), as they represent an essential part of writing to express many things and clarify meanings, but they may cause problems when entered into the form. There is also the problem of abbreviation. Many users shorten words and prepositions to save space in the post, and some believe that using abbreviations gives a post a modern feel[14]. There is also the problem of repeating letters. Saudi users on the platform[3] express themselves by

[1] https://X.com
[2] Ibid
[3] Ibid

repeating letters and words to express feelings or emphasize a specific meaning. Some believe that repeating letters increases the intensity or enthusiasm of expression. Alternatively, it could be part of the activity on social media to attract attention and increase engagement with their posts[14].

3.1.2 Data Preprocessing

We have taken steps to improve the quality of the dataset by implementing a preprocessing technique that enhances its clarity. The primary aim of this technique is to ensure that the data is accurate, consistent, and valid for further analysis.

Preprocessing	Text
Original text	@Ahmadovih أحب اسبانيا!!! جداً، احس كلها حياه 😍🇪🇸 Mar 06, 2017
Remove diacritics.	@Ahmadovih أحب اسبانيا!!! جداً، احس كلها حياه 😍🇪🇸 Mar 06, 2017
Remove English Characters.	@ أحب اسبانيا!!! جداً، احس كلها حياه 😍🇪🇸
Remove punctuation.	أحب اسبانيا!!! جدا احس كلها حياه 😍🇪🇸
Translate the emojis to words.	أحب اسبانيا!!! جدا احس كلها حياه es وجه مبتسم مع عينين على شكل قلب
Replace flag.	أحب اسبانيا!!! جدا احس كلها حياه إسبانيا وجه مبتسم مع عينين على شكل قلب
Remove duplicate characters.	أحب اسبانيا جدا احس كلها حياه إسبانيا وجه مبتسم مع عينين على شكل قلب
Remove Stop Words.	أحب اسبانيا جدا احس حياه إسبانيا وجه مبتسم مع عينين على شكل قلب

Figure 3.1: Example to clarify preprocessing

- **Remove English Characters:** By eliminating any English characters, including numbers, from an Arabic sentence, we enhance its clarity and ensure more accurate processing.
- **Remove Diacritics:** Stripping a sentence of all diacritics not only simplifies it but also enhances its readability, a crucial step in our text processing guidelines.

- **Remove Punctuation:** All punctuation will be removed from the sentence in order to make it more transparent, for instance
  `[#@&%$!(){}[]:;.,<>?[]=+/|\''ŕ~^*_-].`
- **Translate the Emojis to Words:** Translate the emojis to words in order to clarify sentences. For example:

Emoji	الترجمة
😘	وجه يرسل قبلة
❤️	قلب احمر
😳	مندهش
😉	غمزة

Figure 3.2: Emoji lexicon

- **Replace Flag:** The flag will appear as (SA - SN - RS - SC - SC...) and will be replaced with (السعودية - سيشيل - سيراليون - صربيا - السنغال...).
- **Remove Duplicate Characters:** will remove the duplicate that the characters will appear more than two such as "اللللله" convert to "الله" or "اللللهم" will convert to "اللهم".
- **Remove Stop Words:** Remove the words that do not change the meaning of the sentence, such as in Saudi delicate (عشانهم -معايا - مااحد - توهم...) and in Modern Standard Arabic such as(أنت - إلى - منذ - نحن...).

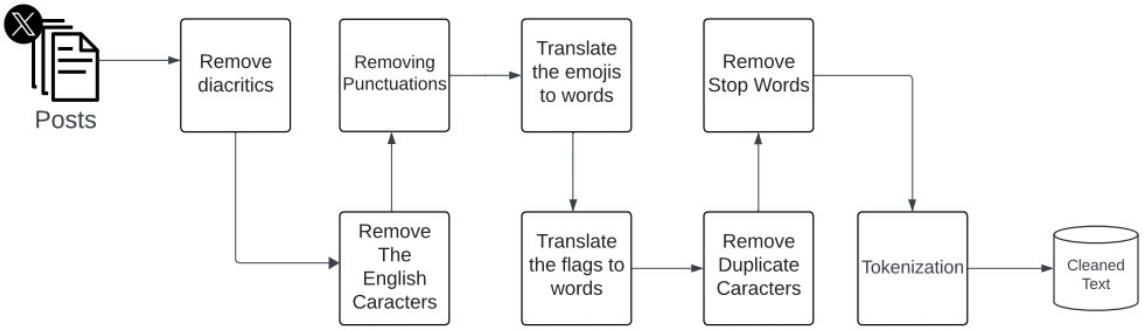

Figure 3.3: Preprocessing steps

3.2 Baseline Model and Algorithm

Our team used a pre-trained model to train on our data. Specifically, we decided to use the MarBERT Model, a pre-trained Bidirectional Encoder Representation from the Transformers model. It was trained on data from X in Arabic dialects and comprises 15.6B tokens[15]. The MarBERT Model was a great choice for our project because it was specifically trained on data from Arabic dialects[15]. Given that the data we collected consisted of posts written in Saudi dialects, we felt this model would be a very appropriate fit for our needs.

A fine-tuning approach was applied to make the model suitable for our data. During the fine-tuning process, the number of layers and weights remained unchanged while we modified other parameters. For example, we set the batch size to 22, the epoch to 10, and used the AdamW optimizer with a learning rate 3e-5. Overall, we are confident that these adjustments helped us train the model for our specific application.

The MarBERT transformer model was developed based on Google's BERT architecture and was subsequently retrained on Arabic posts using the same architecture as the BERT transformer. The model comprises 12 attention layers, 12 attention heads, 768 hidden dimensions, and a maximum sequence length 512.

The framework describes our model 3.4. First, input posts are tokenized by splitting words into subword tokens. Second, multiple transformer layers are used to produce representations of represent the words in the posts. Thirdly, BERT's pre-trained model uses only the [CLS] token for classification tasks, fed into a Softmax function to predict the probability of output classes.

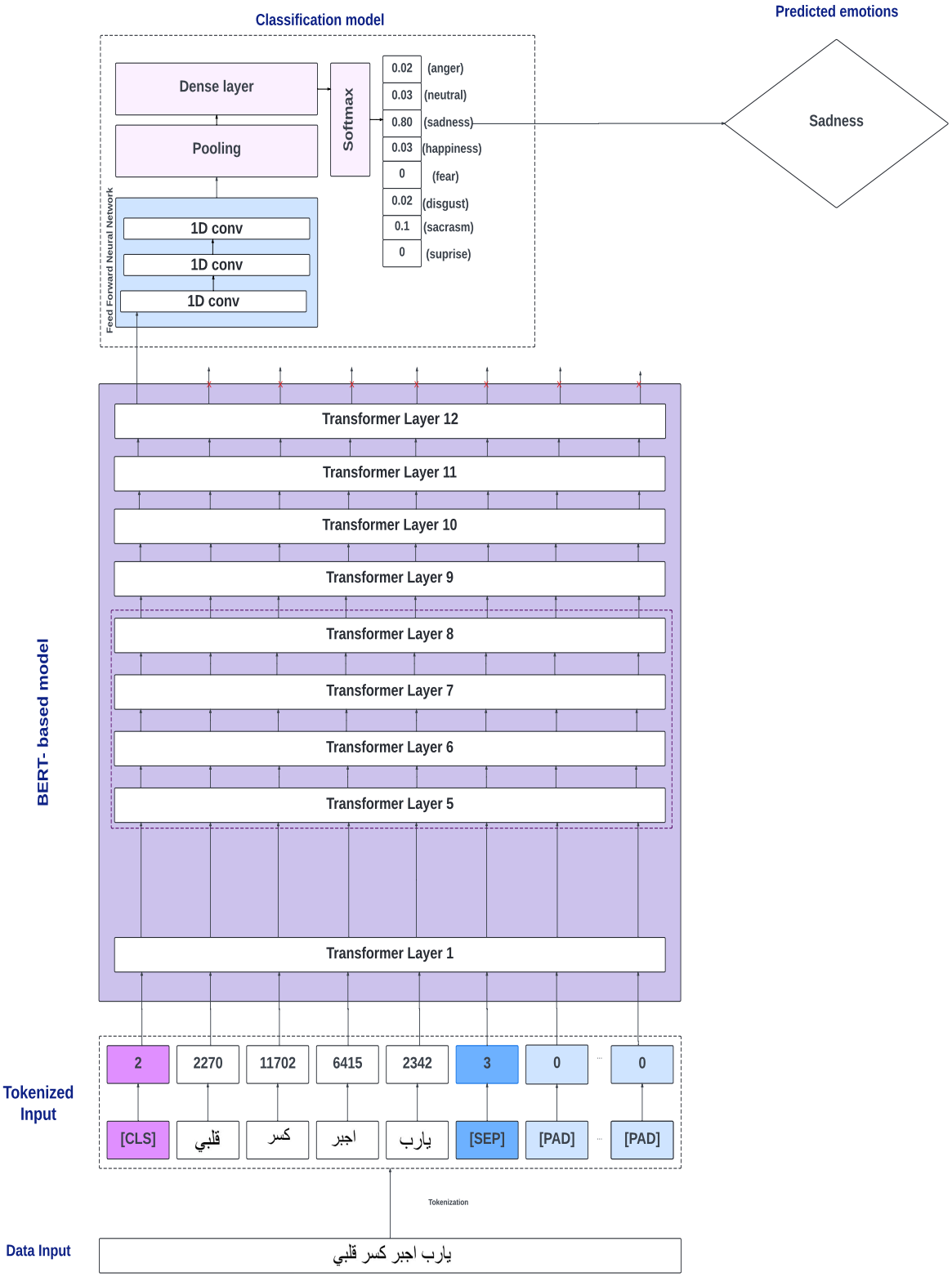

Figure 3.4: MarBERT Model Architecture

3.3 Design Basic Use Case Diagram

Use case diagram: The attached use case diagram represents a Text Emotion Detection System. The system's primary goal is to detect the emotions conveyed in a given text.

1. **User:** The user interacts with the system and provides a text for emotion detection.
2. **Provide the Text:** The user provides input text for emotion detection.
3. **Verify the text:** Check that the text will proceed to the next step if it is in Arabic. However, if all the text is non-Arabic, an error message will be displayed: "Invalid text."
4. **Implement Algorithm:** The text provided by the user will be preprocessed. This includes several steps, including removing the stop word, duplicate characters, punctuation, and numbers, to clean and normalize the text for further analysis, and then predict the entering text.
5. **Percentage of each Label:** If the user enters more than one text, it will calculate the probabilities of each label.
6. **Display the Detected Emotion:** it will appear emotions classification of the text has 8 classes(anger, happiness, neutral, sadness,fear, disgust, surprise, sarcasm)

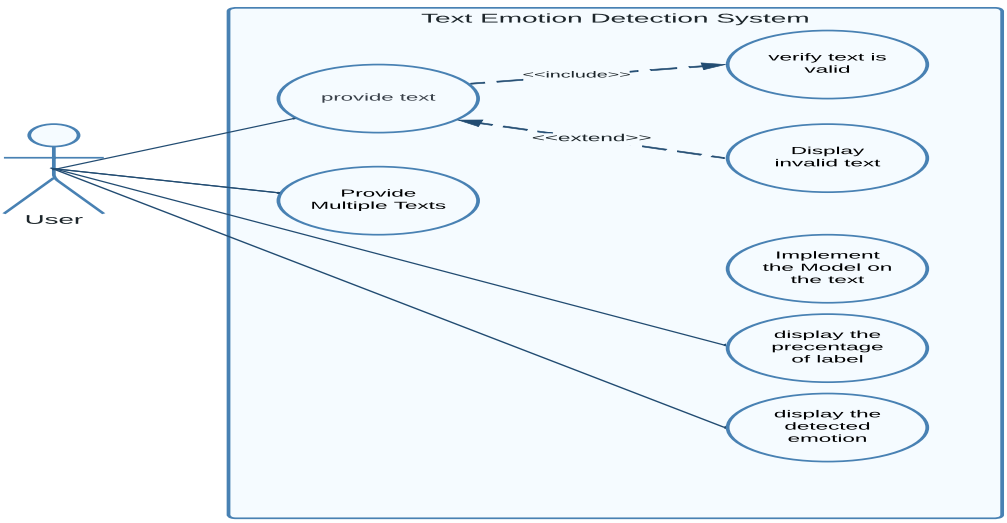

Figure 3.5: Use Case diagram

3.4 Design Prototype

We created a website for emotion detection that anyone can use for individuals, companies, or institutions. The website prototype is designed to explain the system. Initially, the user writes text, which is checked for validity. If the text contains invalid text, an invalid message is displayed. Then, the user writes the text again. The system successfully processes the Arabic text and displays the emotion. In addition, the system can handle multiple texts. When multiple texts are provided, the emotion in each text will be analyzed, and the emotion will be displayed with a percentage label.

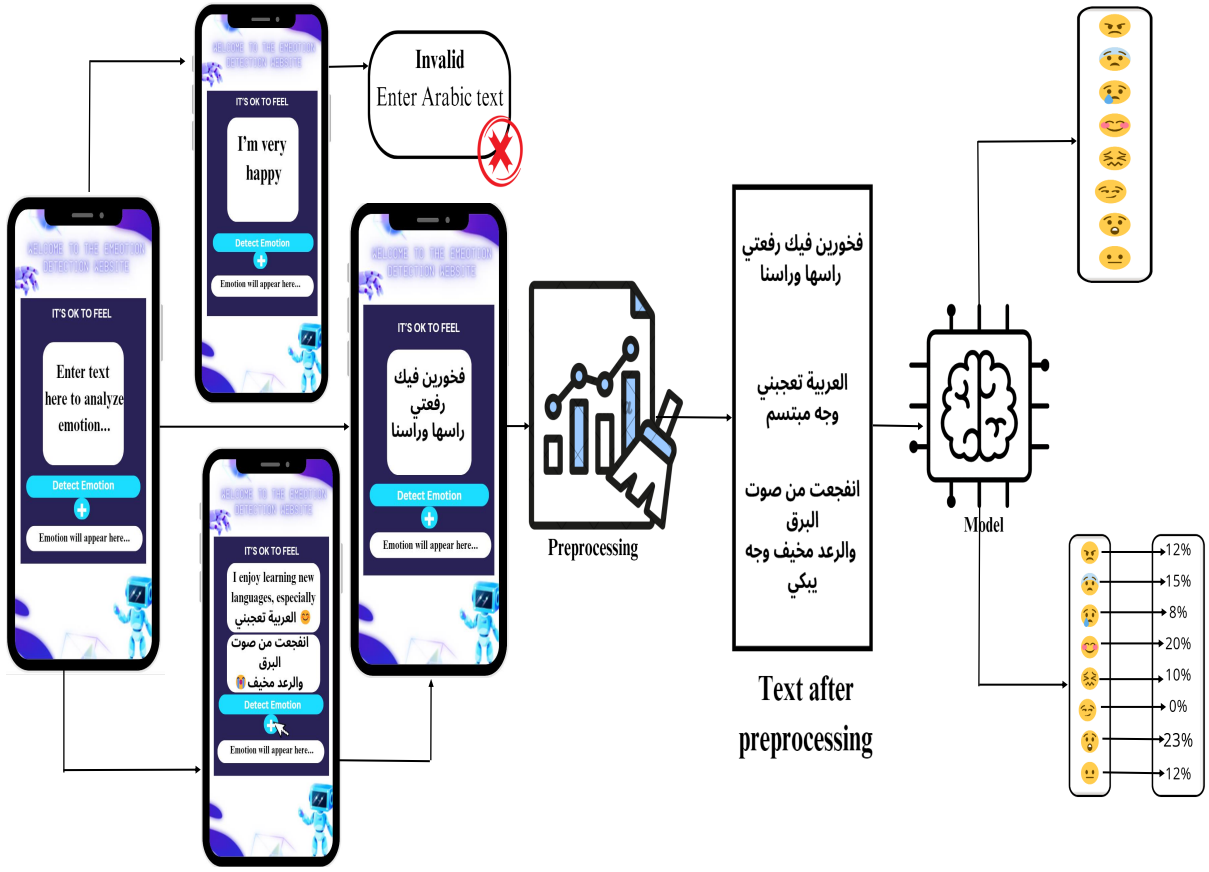

Figure 3.6: Prototype interfaces

Chapter 4

Implementation and Outcome

The implementation has been completed on a minimally used Intel i7 12th Gen CPU 2.7 GHz, 16 GB GPU NVIDIA Tesla T4. Libraries of Scikit learn 1.2.2, re 2.2.1, libraries under the transformers 4.40.0 and torch 2.2.1+cu121 platform in python 3.10.12. The AutoModelForSequenceClassification class from the Hugging Face[1] Transformers Library in Python loads the "Marabert" pre-trained model. As Mentioned in Chapter 3 Section 1.1 we utilized some of Meshrif Alruily's data published in the Issues of Dialectal Saudi Twitter Corpus[14]. In addition, we collected some of the data ourselves from X platform[2]. The combined result was 1201 posts, of which 840 were for training, and 361 were for testing. We assessed the model's performance using various evaluation metrics.

4.1 Evaluation Metrics

- **Accuracy:** Measure correct predictions for all predictions.

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{Total number of predictions}}$$

  Suitable for balanced data, but for unbalanced data, it may reflect something other than the model's actual performance.

[1] https://huggingface.co/
[2] https://X.com

- **Precision:** Measures the true positive predictions among all positive predictions made by the model.

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Precision is essential to know the model's ability to avoid false positives.

- **Recall:** Measures positive predictions for all positive cases, including ones mispredicted by a wrong label.

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Recall is essential to ensure the model can detect all positive cases.

- **F1 score:** Harmonic mean of precision and recall.

$$F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Suitable for unbalanced data and reflects the model performance. While both F1-score and accuracy are essential metrics for evaluating classification models, they serve different purposes. They should be used depending on the specific characteristics of the dataset and the task at hand.

- **Macro-average:** The metric is calculated for each class independently, and then the average is calculated for all classes. Macro is suitable for the multi-class problem because it gives a fair assessment of the model's performance across all classes, ensuring that all classes contribute to calculating the final metrics ratios.

$$\text{Macro-precision} = \frac{\text{Precision}_A + \text{Precision}_B + \text{Precision}_C}{3}$$

$$\text{Macro-recall} = \frac{\text{Recall}_A + \text{Recall}_B + \text{Recall}_C}{3}$$

Macro-F1-score is the harmonic mean of macro-precision and macro-recall.

4.2 Experimental Setup

The model underwent testing through a series of experiments that involved altering certain variables to gauge the model's performance. The data was split into 30% test and 70% training data. After conducting several experiments, the following parameter values were deemed reliable: The Learning Rate is equal to 3e-5, and the AdamW Optimizer was the most effective for the model.

4.3 Preprocessing Variations

As mentioned earlier in Chapter 3, section 1.2, we have taken several steps to preprocess the texts. We have observed that the model's performance has improved slightly even after applying numerous changes to the data; this indicates the model's efficacy in handling noise in the data. The table 4.1 shows that our attempts to feed the data without the emoji lexicon produced lower accuracy than when the lexicon was included. We have utilized the Adam optimizer to optimize the weight and bias for each epoch. We also noticed the difference in performance when the data was balanced and when it was not; as we mentioned previously, we manually made the data balance. During this stage, we noticed the difference.

Preprocessing is essential, no matter how powerful the model is, and it affects its performance.

Optimizer	Adam		
	Metrics		
	precision	recall	f1-score
With Emoji Lexicon	**0.74**	**0.75**	**0.73**
Without Emoji Lexicon	0.73	0.73	0.71

Table 4.1: Results of MarBERT Model Accuracy with and without Emoji Lexicon

4.4 Hyperparameter Tuning

Finding the optimal hyperparameters for a model was manually done by testing various values and comparing them to choose the best ones. We tested different batch sizes and epoch values, and the results of these experiments were recorded in the table 4.3. As we mentioned earlier, the Learning Rate is set to 3e-5, and the Optimizer used was AdmW

Optimizer. All experiments were done on preprocessed data. Our best results were obtained when the batch size was 22 and the epoch was 10, as shown in Table 4.2.

	Epoch = 5			Epoch = 10		
Batch Size	Metrics					
	precision	recall	f1-score	precision	recall	f1-score
22	0.79	0.79	0.78	**0.82**	**0.80**	**0.80**
32	0.73	0.72	0.71	0.78	0.79	0.78
42	0.73	0.71	0.70	0.78	0.77	0.77
52	0.73	0.73	0.72	0.77	0.77	0.77

Table 4.2: Results of MarBERT Model Accuracy on different epochs and batch size

Parameters	
Optimization	AdamW
Epochs	10
Batch Size	22
Learning Rate	3e-5

Table 4.3: Best Parameters of Fine-Tuning Process

4.5 Results

Confusion matrix: a table that compares predicted labels to actual labels, summarizing the performance of a classification model.

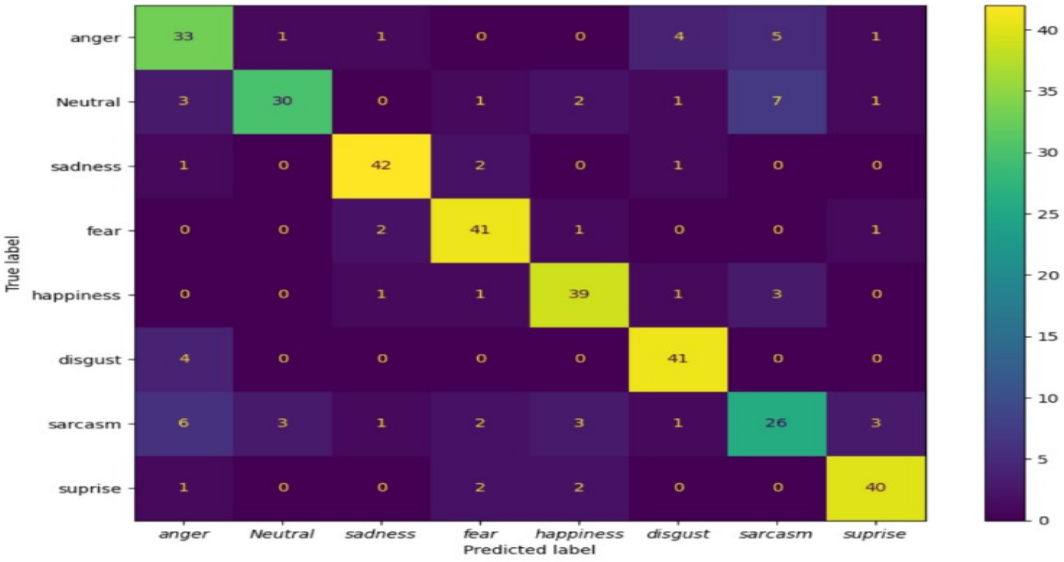

Figure 4.1: Confusion matrix

We discover the emotions mentioned in the previous sections (anger, fear, sadness, disgust, happiness, surprise, neutral, sarcasm) and notice that the model discovers most of them correctly. Sometimes, we need clarification on anger and disgust because the model is confused between these two classes and between anger and sarcasm; the problem is not with the model. The model is confusing because many people express anger with sarcasm, which is common in the Kingdom of Saudi Arabia. There is confusion between anger and disgust because, when expressing feelings, it is possible for a person to feel disgust and anger simultaneously. For this reason, the model is sometimes confused by these feelings. However, the effectiveness and accuracy of the MARBERT model in detecting these emotions have achieved high accuracy. In interpreting the results, the initial percentage was 81%.

```
Text 2 is not valid: Hello world
Note: there is some texts that are invalid so it was not predicted
مرحبا SA كيف الحال {'label': 'Neutral', 'score': 0.9120412468910217}
السلام عليكم ورحمة الله وبركاته {'label': 'Neutral', 'score': 0.8685177564620972}
تحياتي لكم جميعًا EG في مصر الجميلة {'label': 'happiness', 'score': 0.9384638071060181}
anger  : 0.0
Neutral  : 66.66666666666666
sadness  : 0.0
fear  : 0.0
happiness  : 33.33333333333333
disgust  : 0.0
sarcasm  : 0.0
suprise  : 0.0
```

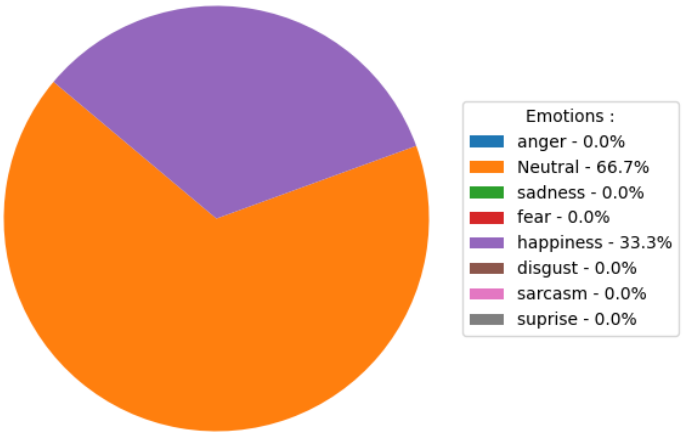

Figure 4.2: Output Example

Output: The user enters the text, which is checked to see if it is valid. Preprocess the entered text and send it to the model to detect the emotions in the text and show them to the user. If the text is invalid because it does not contain any Arabic letters, an error

message is shown to the user, and he enters a new text. If the user enters more than one text, we calculate the percentage of each emotion and show it to the user.

4.6 Discussion

A few researchers have studied emotion detection in Arabic, focusing on emotion classification in English and the Latin alphabet. Our research focuses on Arabic, especially Saudi dialects, and we accurately detect emotions. The initial results were reasonable compared to other researches, as the accuracy in our model reached 81%. In other research, the percentage ranged between 70.01% and 72.5% [2], and in another research, the performance compared to other models alone had an accuracy of 0.540 and an overall F1 score of 0.701[5]. The main problem is that the data we found needed to be labeled, which was a big problem. Moreover, it was 200000 posts, which is a problem since it requires more time and effort to label. We tried to create a small labeled data to train the rest of the data. However, it gave us the wrong emotion classication. Therefore, we reduced and classied the data. Manually, which gave high accuracy in detecting emotions; other limitations gave low accuracy, such as the emojis present in a data set, so we translated each emoji into the text to make the model with high accuracy. Duplicate characters and English characters were also a problem. As shown in Figure 4.1, the model confuses sarcasm with the rest of the classifications, especially with anger and disgust; if the data increases, we will get better results because the limitation is that it is challenging to classify Arabic text, in particular, into different dialects. In future research, we suggest exploring additional emotional categories or more complex emotional states and giving explicit Arabic data to make the model detect emotions optimally. We recommend automatically collecting data on a specific topic to detect emotion without the user entering text, and we hope that future studies will focus on discovering sarcasm in Saudi posts to know the difference between anger and sarcasm, given its importance in improving knowledge of opinions in many fields. Researchers can provide models of natural language processing and develop algorithms that can help detect the difference between sarcasm and anger. In addition, revealing implicit emotions will help in many things, especially in public decisions, to know peoples emotions towards this decision or belonging to a particular party, and to know customers emotions towards this company or party. Therefore, this will help to understand the general opinion of the Saudi community on current issues or topics of interest, such as significant developments in

tourism, the Riyadh Season, and the latest improvements in the health or education sectors, and it is easy to apply because it will take the average of emotion.

Chapter 5

Conclusion

Emotion detection in Arabic text on the X platform[3]presents a significant challenge due to the wide variety of dialects, delicate nuances, and frequent errors in spelling and grammar. Preprocessing the text and ensuring its appropriate representation with accurate emotional context required considerable effort and time. However, our perseverance paid off as we adopted the MARBERT model, which has been trained on a wide range of Arabic dialects. This enabled us to achieve remarkable results in accurately identifying and detecting emotions in Arabic text.

[3]https://X.com

References

[1] "Frontiers | Detection of emotion by text analysis using machine learning — frontiersin.org," https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.119032 [Accessed 02-05-2024].

[2] W. Alshehri, N. Al-Twairesh, and A. Alothaim, "Affect analysis in arabic text: Further pre-training language models for sentiment and emotion," *Applied Sciences*, vol. 13, no. 9, p. 5609, 2023.

[3] A. A Aladeemy, A. Alzahrani, T. HH Aldhyani, O. Ibrahim Khalaf, S. Nagi Alsubari, S. Algburi, and S. N Deshmukh, "Developing system based machine learning for predicting and analyzing arabic sentiment," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1--11, 2024.

[4] G. Alqahtani and A. Alothaim, "Emotion analysis of arabic tweets: Language models and available resources," *Frontiers in Artificial Intelligence*, vol. 5, p. 843038, 2022.

[5] A. Mansy, S. Rady, and T. Gharib, "An ensemble deep learning approach for emotion detection in arabic tweets," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[6] I. A. Farha and W. Magdy, "Benchmarking transformer-based language models for arabic sentiment and sarcasm detection," in *Proceedings of the sixth Arabic natural language processing workshop*, 2021, pp. 21--31.

[7] N. Al-Twairesh, "The evolution of language models applied to emotion analysis of arabic tweets," *Information*, vol. 12, no. 2, p. 84, 2021.

[8] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint*, 2020.

[9] ——, "Arabert: Transformer-based model for arabic language understanding. arxiv 2020," *arXiv preprint arXiv:2003.00104*, 2020.

[10] H. M. Alghamdi, "Unveiling sentiments: A comprehensive analysis of arabic hajj-related tweets from 2017--2022 utilizing advanced ai models," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 5, 2024.

[11] H. F. Al-Saif and H. Z. Al-Dossari, "Exploring the role of emotions in arabic rumor detection in social media," *Applied Sciences*, vol. 13, no. 15, p. 8815, 2023.

[12] A. A. Jalil and A. H. Aliwy, "Classification of arabic social media texts based on a deep learning multi-tasks model," *Al-Bahir Journal for Engineering and Pure Sciences*, vol. 2, no. 2, p. 12, 2023.

[13] E. A. H. Khalil, E. M. E. Houby, and H. K. Mohamed, "Deep learning for emotion analysis in arabic tweets," *Journal of Big Data*, vol. 8, no. 1, p. 136, 2021.

[14] M. Alruily, "Issues of dialectal saudi twitter corpus." *Int. Arab J. Inf. Technol.*, vol. 17, no. 3, pp. 367--374, 2020.

[15] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "Arbert & marbert: deep bidirectional transformers for arabic," *arXiv preprint arXiv:2101.01785*, 2020.