NLP Report 1
Emotion Detection in Arabic Tweets

Mohamed Zaki 46-1939
Ahmed Attawia 46-4982

May 8, 2023

1 Introduction

The Arabic language, owing to its complex nature, offers many challenges in NLP tasks in comparison to English. Arabic grammar is extremely complex when compared to other languages, and many Arabic words consist of multiple parts each having its own meaning. Not to mention words that are inferred from the context such as hidden subjects or omitted objects. These characteristics provide us with a great challenge in tackling NLP tasks in Arabic and is an area of research that is not heavily tackled. In this project, we attempt to tackle the under-researched problem of emotion detection in Arabic tweets.

2 The Dataset

The dataset we will use in this project is the SemEval-2018 dataset[2]. This dataset is a compilation of more than 22,000 tweets in three languages: English, Arabic, and Spanish. This dataset is further divided into 5 sections tackling different NLP tasks. In this project, we are mainly concerned with the Arabic section and its fifth partition specifically, which tackles emotion detection. This section, from here on referred to as the dataset, is a collection of 4,000+ tweets written in Arabic. Each tweet has a label for the following emotions: Anger, Anticipation, Disgust, Fear, Joy, Love, Optimism, Pessimism, Sadness, Surprise, and Trust.

2.1 Data Preprocessing

To be able to effectively use the dataset in our project, some preprocessing had to be done first. The data preprocessing pipleine is as follows:

1. Removing English characters.
2. Removing digits.
3. Removing stop words.
4. Removing diacritics.
5. Normalizing Arabic text.
6. Removing Emojis.
7. Removing non Arabic characters.

2.2 Data Analysis

In this section, we attempt to do some simple analysis on the dataset to gain some insights and find some limitations.

2.2.1 How many tweets are in each emotion label?

In this figure, we can see the distribution of emotions in our dataset. We can see that the emotions Anticipation, Trust, and Surprise are not well represented in this dataset, and this could become a limitation later when trying to predict them.

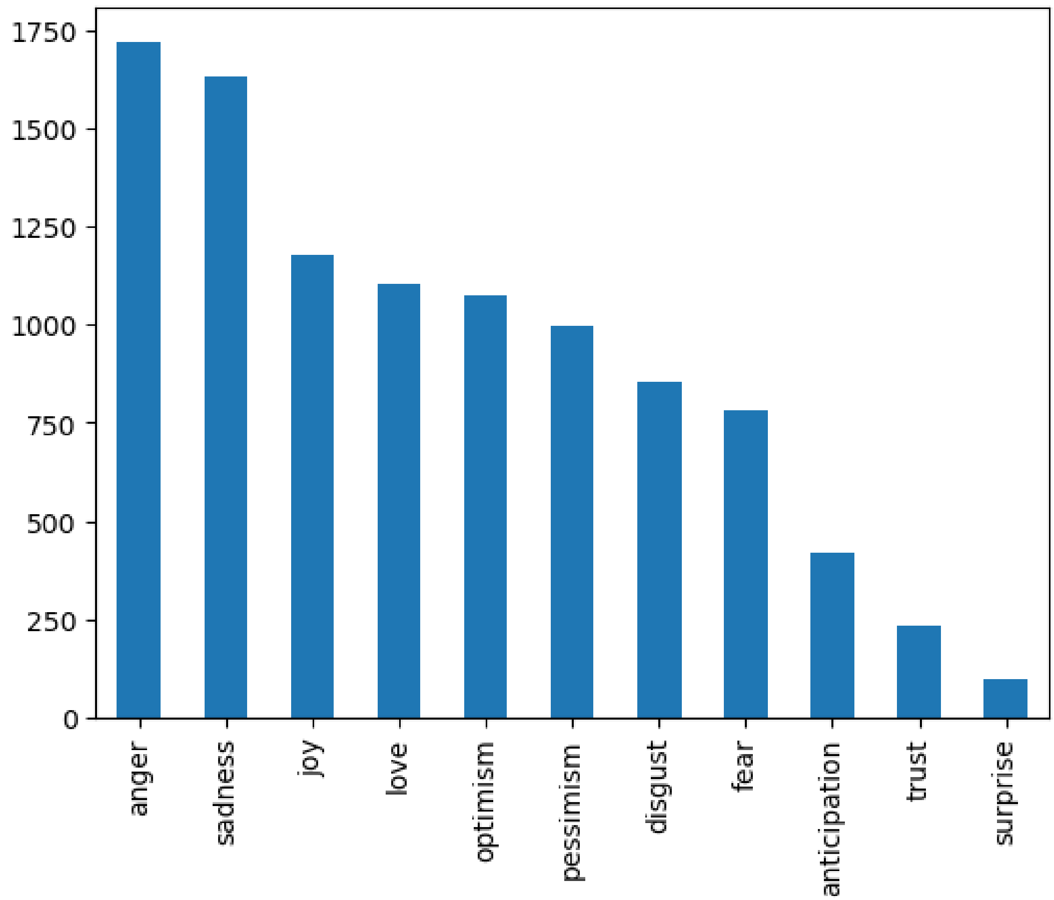

Figure 1: Tweet count for each emotion

2.2.2 The number of emotions per tweet

In this figure, we can see that the majority of tweets have more that one emotion detected. This can be beneficial as this way emotions are more accurately represented and more samples for each emotion is present.

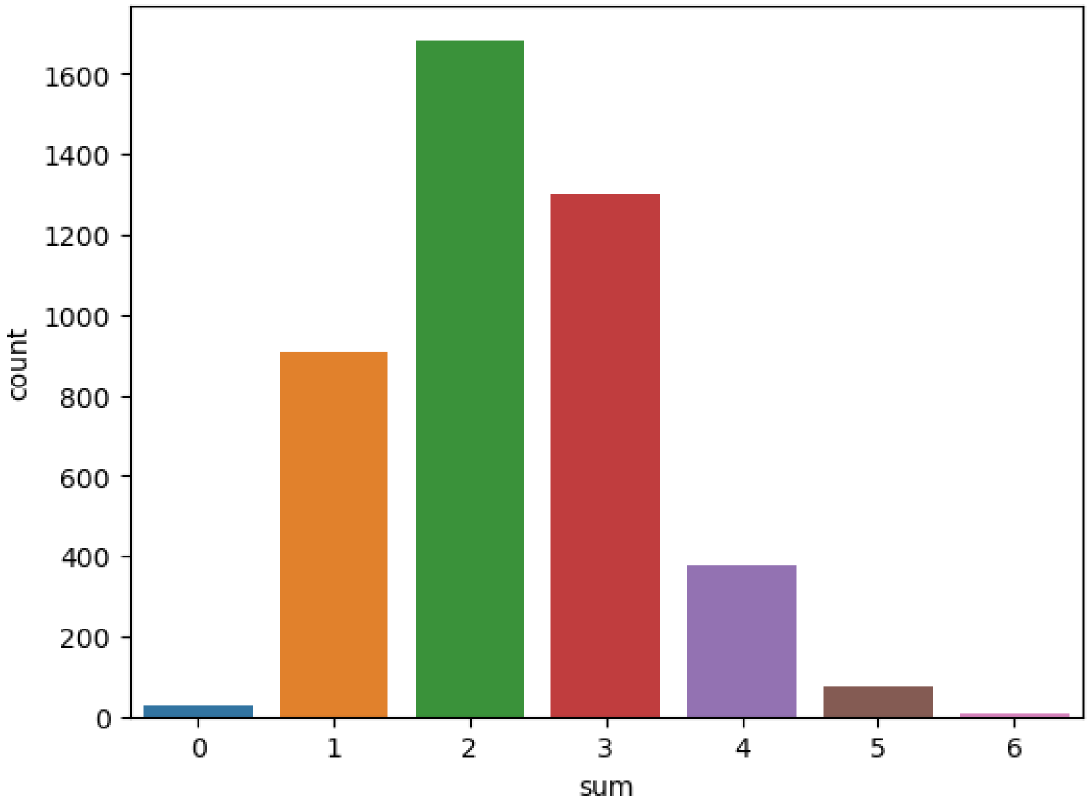

Figure 2: Count of emotion labels in tweets

2.2.3 Does tweet length correlate to number of emotions?

This figure was an attempt to find a relation between the length of the tweet and the number of emotions represented. The theory behind this was that as the tweet has more words in it, it could be describing more emotions. As we can see from the figure, this was disproven by the data.

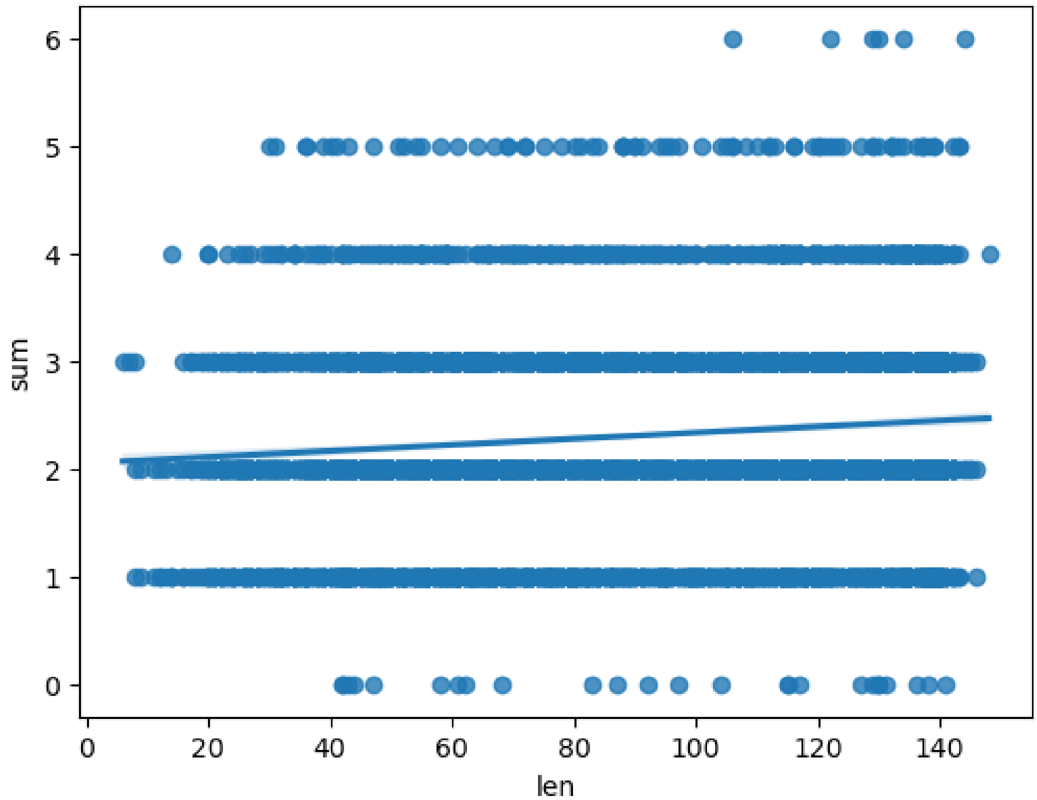

Figure 3: Relation between length of tweet and emotion count

3 System Architecture

Based on work from Mansy et al.[1] we will attempt to tackle this project by training three different models on the dataset to classify emotions, and then combining all 3 outputs using either an ensemble layer or a stacking approach where we feed the outputs into a fourth model and use that prediction as our final output. The models that will be used are as follows:

- **Bi-LSTM model:**Bi-directional LSTMs (Long Short-Term Memory) are a type of recurrent neural network that processes input sequences in both forward and backward directions using two separate hidden layers. This allows the network to capture both past and future context when making predictions, making it well-suited for tasks such as speech recognition and natural language processing.
- **Bi-GRU model:**Bi-directional GRUs (Gated Recurrent Units) are similar to bi-directional LSTMs in that they also process input sequences in both forward and backward directions. However, GRUs use a simpler architecture that replaces the LSTM's memory cell and output gate with a single update gate, making them quicker to train and more computationally efficient. Despite their simpler design, bi-directional GRUs can still be effective for tasks such as machine translation and speech recognition.
- **Pretrained language model (MARBERT Transformer):**A pretrained language model based on BERT has been introduced which is called MARBERT. It was pretrained based on both MSA and Arabic dialects, unlike ARABERT which was pre-trained only on Arabic MSA. Because MARBERT transformer was used before in the emotion detection task, it has an emotion-related contextual understanding experience.

References

[1] Mansy, A., Rady, S., and Gharib, T. An ensemble deep learning approach for emotion detection in arabic tweets. *International Journal of Advanced Computer Science and Applications 13* (01 2022).

[2] Mohammad, S. M., Bravo-Marquez, F., Salameh, M., and Kiritchenko, S. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)* (New Orleans, LA, USA, 2018).