# Dealing with Emojis in Arabic Sentiment Analysis

## 1 Approach 1: Replacing Emojis

Initially, we replaced emojis in the text with specific words to simplify the text data and reduce noise. Despite improving the text's readability, this approach inadvertently introduced a bias in the dataset. As a result, labels 2, 3, and 6 were not effectively represented in the model's training process.

Using (GRU) model, we achieved an accuracy of 72 after implementing the initial approach. While this is a notable improvement, we found that the model's performance was hindered by the biased dataset, as certain labels were underrepresented.

## 2 Approach 2: Leveraging MARBERT with uncleaned data

To address the bias and improve model performance, we adopted a different strategy. We leveraged the MARBERT model, which is pre-trained on Arabic sentiment analysis tasks. This model demonstrated a better understanding of emojis and the nuanced emotions they convey.

Using the MARBERT model, we generated weighted embeddings that captured the emotions expressed by emojis. This approach significantly improved our model's performance, achieving an accuracy of 77 with the GRU model and 75 with logistic regression. The model's ability to capture and distinguish emojis' sentiments played a crucial role in achieving these results.

## 3 Conclusion

In sentiment analysis tasks, emojis offer valuable context and sentiment indicators. However, dealing with emojis requires careful consideration to avoid introducing biases or losing valuable information. Our experiments demonstrate the effectiveness of leveraging BERT-based models like MARBERT, which not only handle emojis effectively but also provide a more nuanced understanding of the emotional context within text.

Our findings suggest that by properly addressing emojis' impact on sentiment analysis, we can achieve more accurate and representative results, enhancing the quality of sentiment classification models.