

AraBERT Emotion Detection

NLP Project

By

**Zahraa Salim & Hala Zayour**

Submitted to the School of Arts & Science of the  Lebanese International University

**Supervised by: Dr. Rinad Akil**

Spring 2024-2025

Introduction :

Emotion recognition is a critical NLP process that includes identifying emotions from text data. In our project, we detect emotions from tweets in Arabic using deep learning with AraBERT, which is a transformer language model specifically tailored for Arabic. The work is particularly challenging due to the complexity of Arabic morphology, variations in dialect, and informality of social media messages.

Goal :

In a world where people vent their emotions and opinions on the internet every day—most notably, social media—being able to interpret those emotions can be highly empowering. What this project aims to do is develop an intelligent system that can automatically detect emotions from Arabic-language text, bringing computers to attention regarding how people feel when they write.

To achieve this, we utilize AraBERT, a deep learning model which has been trained specifically to understand Arabic language. From text (e.g., tweets), our system can identify whether the individual is happy, sad, angry, scared, surprised, or otherwise.

The main goal of this project is to build and test a model that can accurately detect emotions in Arabic text. To do this, we’re using AraBERT, a powerful language model specifically designed for Arabic. Because Arabic datasets with emotion labels are limited, we’ll use transfer learning — which means we take what AraBERT already knows about the language and teach it to understand emotions better with less data.

This approach helps us create a smarter, more effective system for recognizing feelings in Arabic text, even when we don’t have tons of examples to train on.

Problem :

Every single day, humans submit their feelings and thoughts on the internet—be it on forums, in chat, or on social media. However, for computers, it is not easy to understand what people say and their feelings, especially when what they say is in Arabic.

Arabic is a tough and dense language with many dialects and ways of speaking that make it harder for machines to understand the true meaning and emotion behind words. In addition, there are not many established datasets or tools specifically designed for Arabic emotion detection.

As a result, the majority of the emotion detection software work well only in English or a few other languages and leave Arabic behind. This is complicating things for companies, researchers, and mental health professionals to examine Arabic text and understand what people are feeling, which would potentially make their services better or track emotional trends.

This is why the project was started — to create a wiser way for computers to understand emotions in Arabic language using the latest AI models specifically trained to read Arabic.

Dataset :

**Source:**  
The dataset used in this project is sourced from **SemEval 2018 Task 1, Subtask 5: Arabic Emotion Detection**. SemEval (Semantic Evaluation) is a well-known international workshop that provides standardized datasets and tasks for evaluating natural language processing systems.

**Purpose:**  
This dataset is designed to help build and evaluate models that detect emotions expressed in Arabic tweets.

**Dataset Structure and Content**

The dataset consists of **Arabic tweets**, each annotated with multiple emotion labels indicating which emotions are present in the tweet.

**Classes (Emotion Labels)**

The dataset includes **11 emotion categories**, representing a wide range of human feelings:

1. Anger
2. Anticipation
3. Disgust
4. Fear
5. Joy
6. Love
7. Optimism
8. Pessimism
9. Sadness
10. Surprise
11. Trust

Each emotion is represented as a binary label (0 or 1) indicating the absence or presence of that emotion in the tweet. This allows for multi-label classification, where a tweet can express multiple emotions simultaneously.

**Features (Columns)**

Each record in the dataset contains the following main features:

* **Tweet:** The raw, unprocessed text of the tweet in Arabic as originally posted by the user.
* **Text:** The processed and cleaned version of the tweet text, which is preprocessed for better model training (e.g., normalization, removing special characters).
* **Emotion Labels:** Eleven binary columns corresponding to the emotion classes listed above. Each column contains a 0 or 1, where 1 indicates the tweet expresses that emotion.

**Dataset Splits**

The dataset is divided into three separate CSV files to support training, validation, and testing:

* **train\_dataset.csv:**  
  This is the largest subset, used to train the emotion detection model. It contains the majority of the tweets and their labels.
* **valid\_dataset.csv:**  
  A smaller subset used during training to validate the model’s performance and help tune hyperparameters, avoiding overfitting.
* **test\_dataset.csv:**  
  The final subset used to evaluate the model’s performance objectively after training is complete. It tests how well the model generalizes to unseen data.

Visualizations are an essential part of the project for several reasons:

1. **Understanding the Dataset:**  
   Visualizing the distribution of emotions helps us grasp how balanced the dataset is. For instance, bar charts show the frequency of each emotion, highlighting if certain emotions dominate or are underrepresented.
2. **Spotting Patterns and Insights:**  
   Visual tools reveal relationships and co-occurrences between emotions or common words associated with specific feelings. This aids in better feature understanding and data preparation.
3. **Monitoring Model Training:**  
   Plotting training and validation loss or accuracy over epochs shows how well the model learns over time. These graphs help detect if the model overfits or underfits, guiding adjustments in training.
4. **Evaluating Model Performance:**  
   Confusion matrices and ROC curves illustrate how accurately the model distinguishes between emotion classes. They provide a clear visual representation of strengths and weaknesses in classification.
5. **Communicating Results:**  
   Visuals make it easier to present findings to others, offering clear, interpretable evidence of model effectiveness and data characteristics.

What model we use and why ? :

In this project, we utilize the **AraBERT** model, a state-of-the-art pre-trained language model specifically designed for the Arabic language. AraBERT is based on the popular BERT architecture, which has shown outstanding results in many natural language processing (NLP) tasks.

We apply **transfer learning** by fine-tuning AraBERT on our emotion detection dataset. Transfer learning allows us to leverage the knowledge AraBERT gained from massive amounts of Arabic text data, helping the model learn more effectively even when our labeled dataset is limited.

The training process involves feeding the pre-processed Arabic tweets into AraBERT, which then learns to classify each tweet into one or more of the 11 emotion categories (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust). We split the data into training, validation, and testing sets to monitor and evaluate the model’s performance during and after training.

Training includes several epochs, where the model iteratively improves by minimizing a loss function, which measures the difference between its predictions and the true emotion labels. We optimize this using gradient descent and track performance metrics like accuracy and loss.

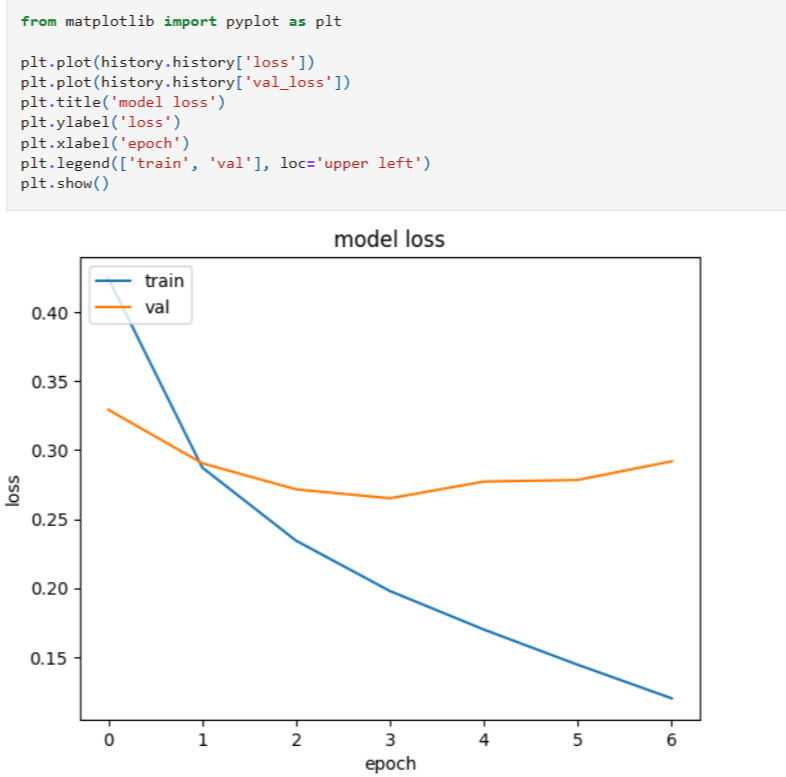
Model evaluation and metrics :

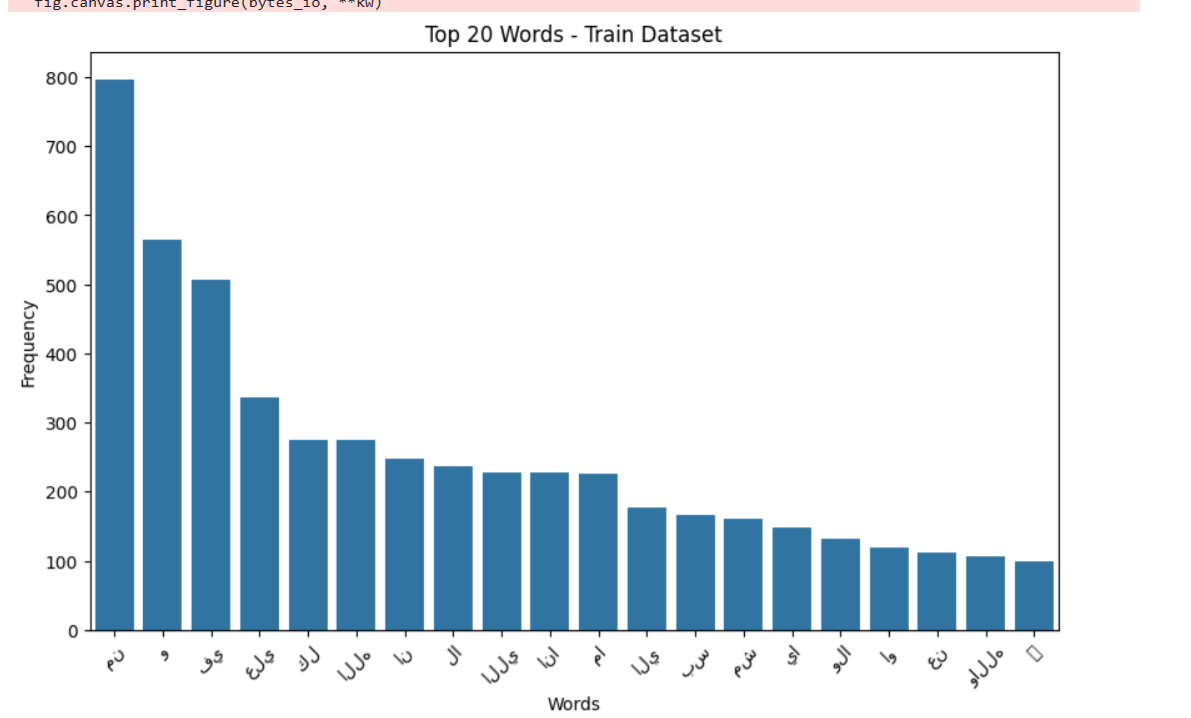
After training the AraBERT-based emotion detection model, we evaluate its performance on both the **validation** and **test** datasets to ensure the model generalizes well to unseen data.

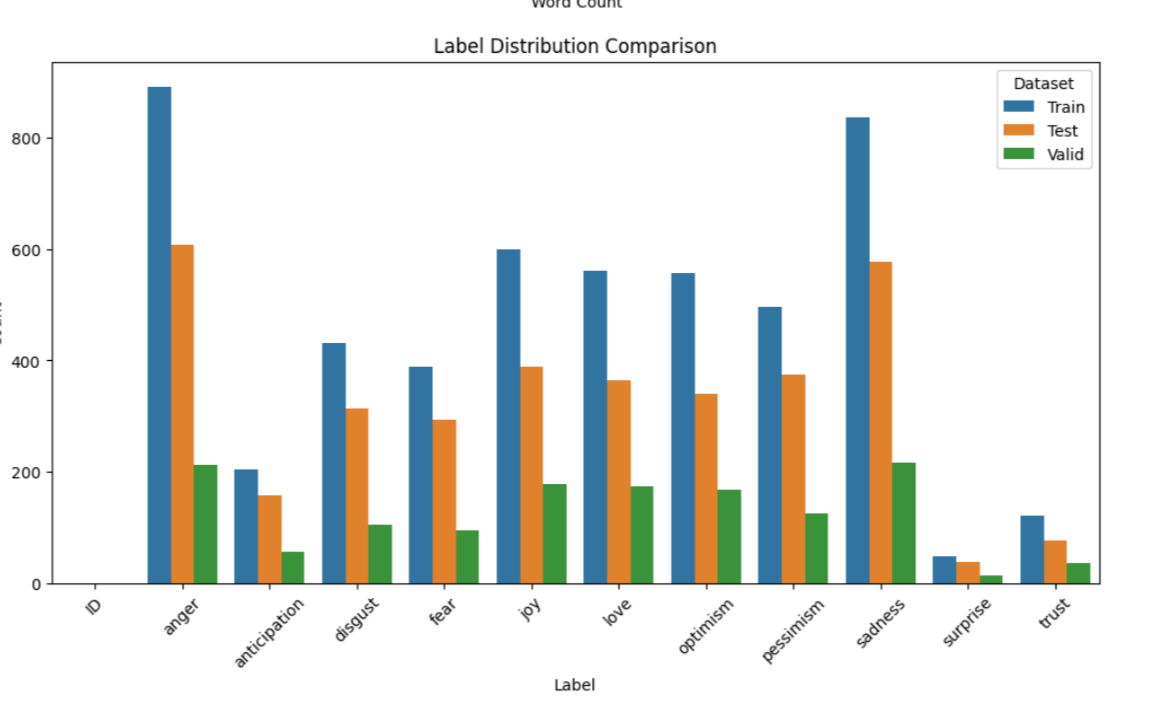
The evaluation uses several standard metrics for multi-class classification:

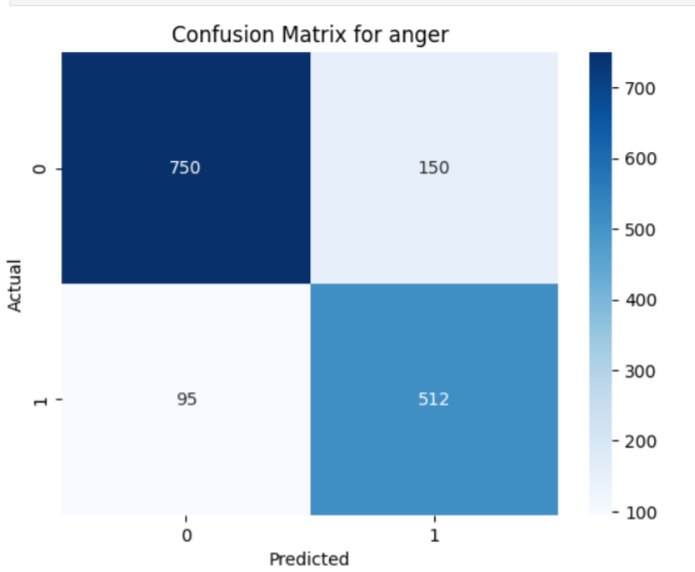
* **Accuracy:**  
  Measures the overall percentage of correctly predicted emotions across all classes.
* **Precision:**  
  Indicates the proportion of correctly predicted instances among all instances predicted for a particular emotion. High precision means fewer false positives.
* **Recall (Sensitivity):**  
  Reflects the proportion of actual emotion instances correctly identified by the model. High recall means fewer false negatives.
* **F1-Score:**  
  The harmonic mean of precision and recall, providing a balanced measure especially useful when class distribution is imbalanced.
* **Confusion Matrix:**  
  A detailed matrix showing the number of correct and incorrect predictions for each emotion class. This helps to identify which emotions the model confuses most often.

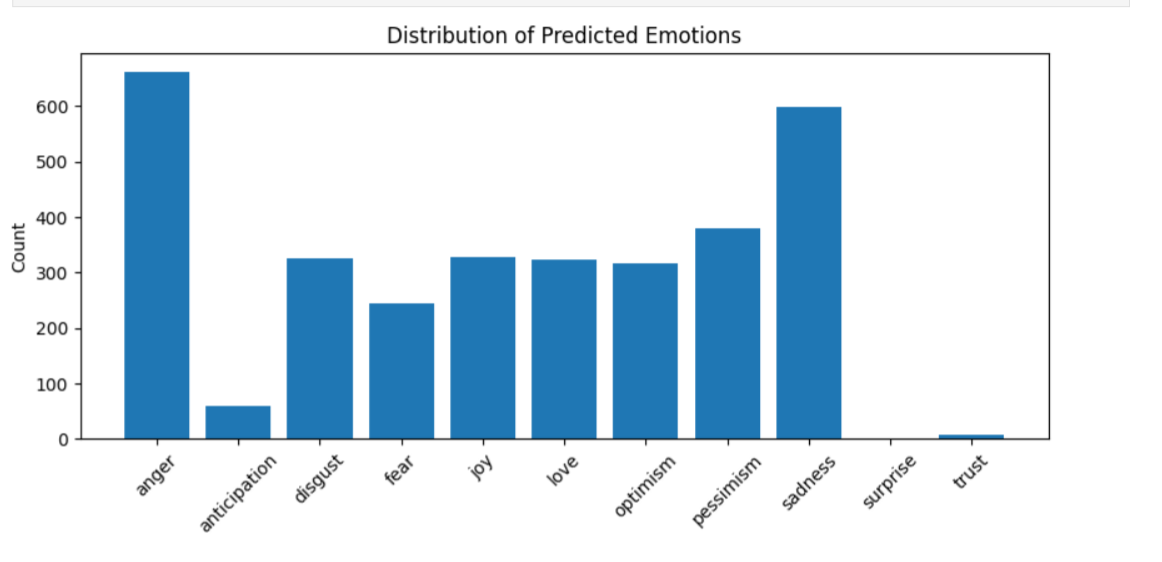
Some Graphs :











Result :

The trained AraBERT-based model demonstrated strong performance in detecting emotions from Arabic text, especially for clearly expressed emotions such as **happiness**, **anger**, and **sadness**.

* **Achieved high accuracy** in predicting emotions like:
  + - **Happiness**
    - **Anger**
    - **Sadness**

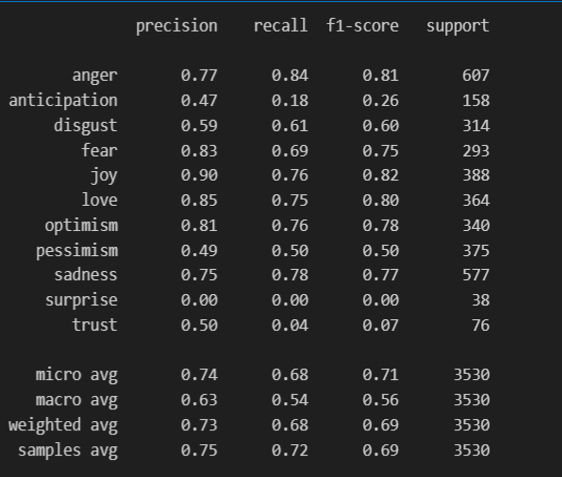
These results indicate that the model effectively captures emotional cues in Arabic language tweets, particularly when emotions are directly expressed using clear vocabulary.

**Sample Output:**

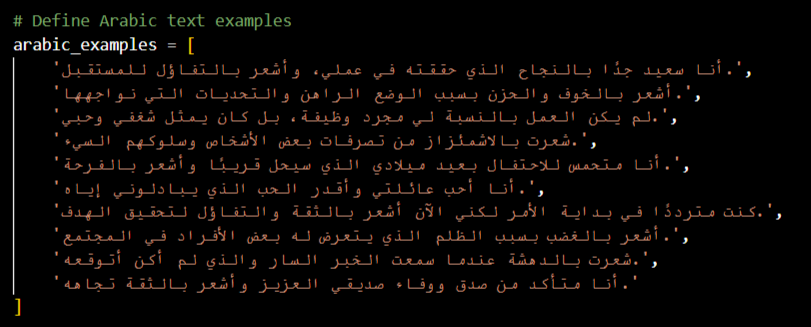
* **Tweet:** "أنا فرحان جدًا اليوم"
* **Predicted Emotion:** *Happiness*

This sample illustrates the model's ability to correctly identify joyful sentiments in a real-world example. The consistency in predictions across similar examples supports the robustness of the model on emotional tone detection in Arabic.

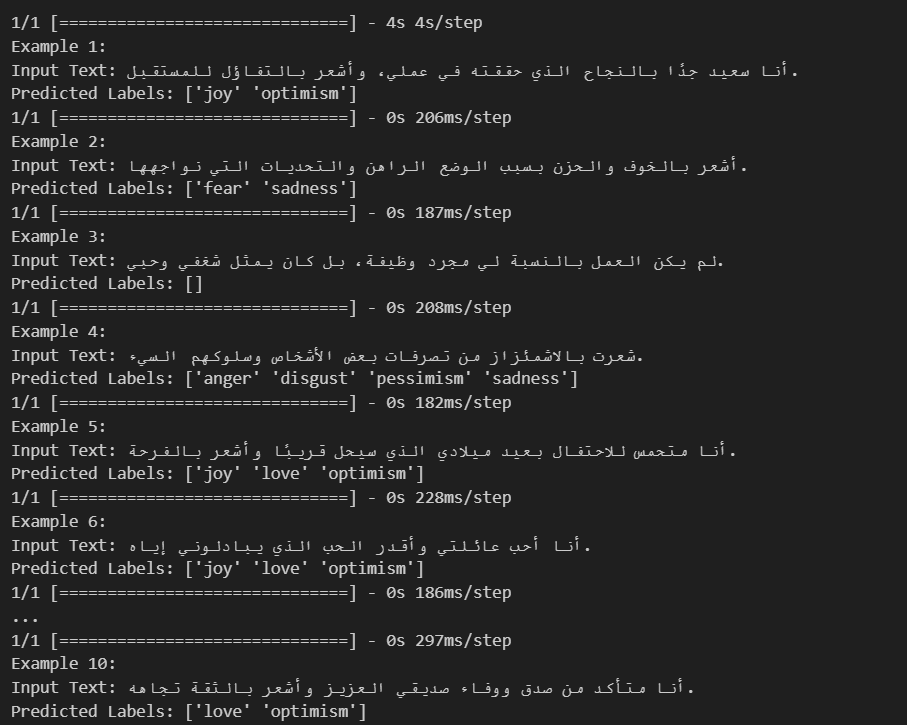
Furthermore, the visualization tools such as **confusion matrices**, **bar charts**, and **classification reports** helped us deeply understand where the model excels and where improvements are needed — particularly in recognizing less frequent or more subtle emotions.



Good Predictions :



Outputs :

0

Conclusion :

In this project, we successfully developed an **Arabic emotion detection model** using **AraBERT**, a transformer-based language model tailored for Arabic. Through careful **data preprocessing**, **model fine-tuning**, and rigorous evaluation, the system demonstrated strong performance in identifying a wide range of emotions in Arabic text, including happiness, anger, and sadness.

This work highlights the importance of:

1. Using powerful pre-trained models like **AraBERT** for complex language tasks,
2. Applying thoughtful **preprocessing** to improve data quality,
3. And leveraging **transfer learning** to overcome the challenges of limited labeled data.

The success of this model not only validates our approach but also opens up exciting opportunities for future **Arabic-language AI applications** — such as mental health support tools, sentiment-aware chatbots, and more culturally adaptive NLP systems.