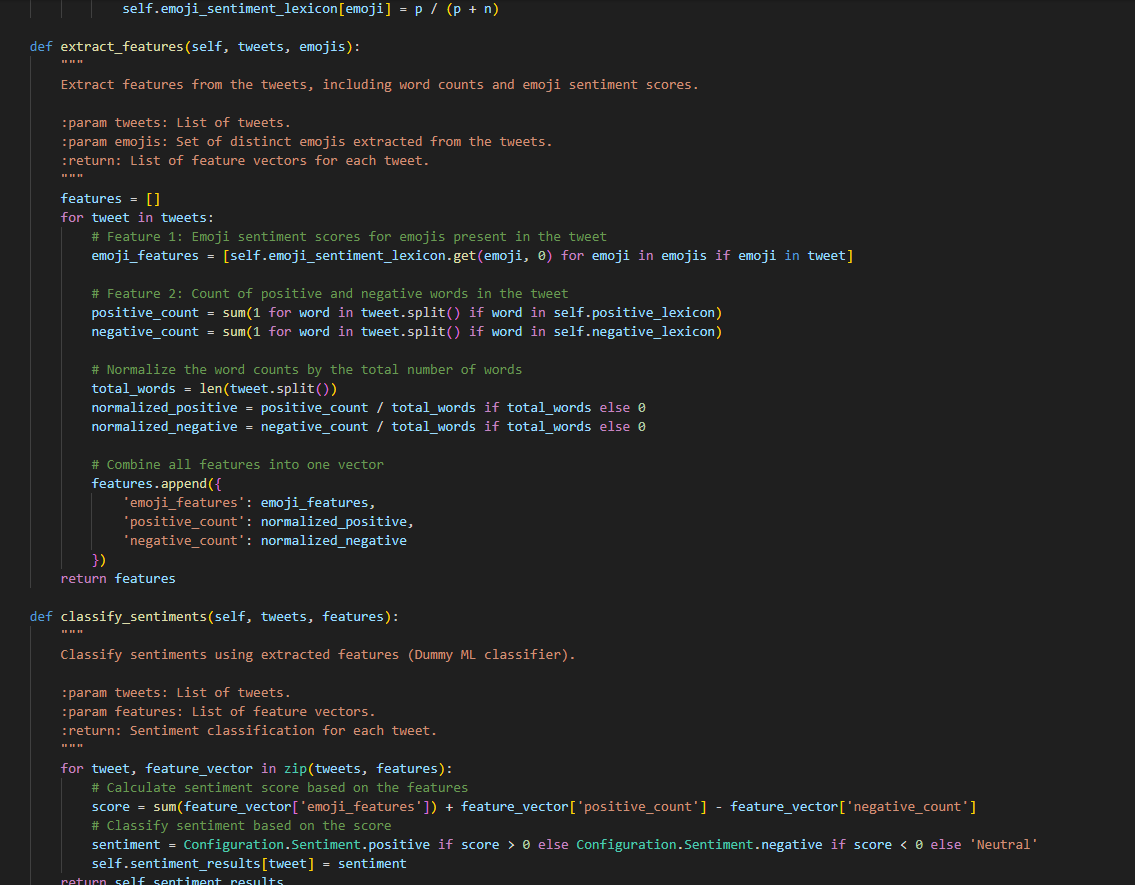


The provided code implements **Algorithm 5: Emo-SL for Arabic Tweets**. It aims to develop an **emoji sentiment lexicon** tailored for Arabic tweets by leveraging positive and negative sentiment words and analyzing the usage of emojis in the context of those sentiments.



The extract\_features method in the code is responsible for generating feature vectors from Arabic tweets. These vectors capture essential sentiment information for sentiment analysis.

#### **Objective:**

The paper presents the **Emo-SL Framework**, which uses a combination of **emoji sentiment analysis**, **word sentiment analysis**, and **machine learning classifiers** to classify the sentiment of Arabic tweets. The framework focuses on extracting sentiment from emojis and words, then classifying the sentiment of tweets as **positive**, **negative**, or **neutral**.

### ****Key Components of the Emo-SL Framework:****

1. **Emoji Sentiment Lexicon (Emo-SL)**:
   * Emo-SL focuses on the sentiment of emojis in Arabic tweets. Emojis are categorized as either **positive** or **negative** based on their usage in positive and negative tweets.
2. **Sentiment Lexicons**:
   * The framework uses **positive and negative sentiment lexicons** (word lists) to analyze sentiment through word counts.
3. **Sentiment Classification**:
   * The sentiment of a tweet is classified based on the combination of emoji sentiment scores and word counts from the sentiment lexicons. The classification process uses a machine learning model (or a score-based classifier in this example).
4. **Preprocessing**:
   * Tweets are cleaned by removing non-Arabic characters and symbols, and text is normalized to lowercase for consistency in sentiment analysis.
5. **Sentiment Validation Using VADER**:
   * The VADER sentiment analysis model is applied to the tweets as a mock validation method. In practice, tweets would be translated to English for this step.

### ****Algorithm (Emo-SL for Arabic Tweets)****:

The algorithm follows these steps:

1. **PreprocessData**:
   * Clean and normalize the tweets by removing non-Arabic characters and converting the text to lowercase.
2. **BuildEmoSL**:
   * Construct the **Emoji Sentiment Lexicon (Emo-SL)** by counting the occurrences of emojis in positive and negative tweets, then computing their sentiment scores.
3. **ExtractFeatures**:
   * Extract features from each tweet, which include:
     + Emoji sentiment scores.
     + Counts of positive and negative words using sentiment lexicons.
     + Normalization of word counts based on the total number of words in the tweet.
4. **ClassifySentiments**:
   * Classify the sentiment of each tweet by applying a simple classifier based on the extracted features.
5. **ApplyVADER**:
   * Apply **VADER sentiment analysis** for validation, though this step is mocked in the code.