**SENTIMENTAL ANALYSIS – EMOTIONS IN TEXT**

**DATASET:**

The given dataset consists of two attributes, namely text and label. Text is the sentences which show the emotion and Label is the emotion. Label is the Target Variable. The size of Dataset is (3613,2).

**CHECKING FOR NULL VALUES AND BALANCED OR IMBALANCED DATA:**

The dataset has no null or missing values. The given Dataset is Balanced.

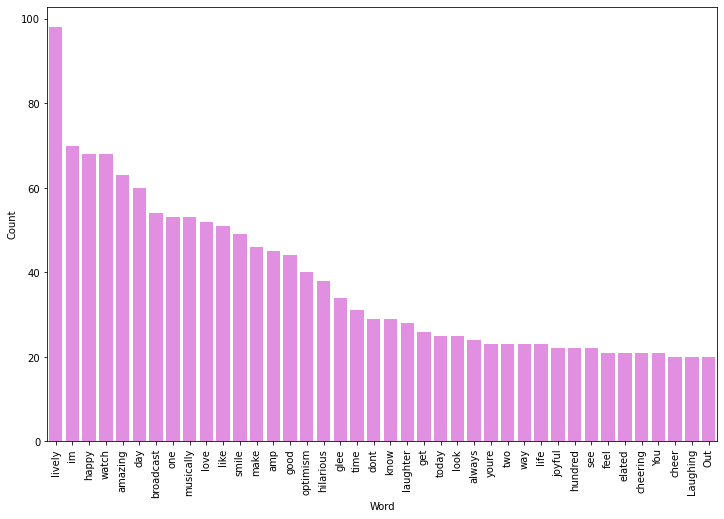
DATA PRE-PROCESSING TECHNIQUES:

1. Removing Punctuations
2. Converting Number to Words.
3. Converting Accented Characters to Ascii Characters
4. Converting SMS language into Exact words.
5. Removing Emoji’s.
6. Removing stopwords.
7. Applying Lemmatization Technique.

EXPLORATORY DATA ANALYSIS:

Plotting the Top 40 Frequent words in each emotion label for more understanding of data.

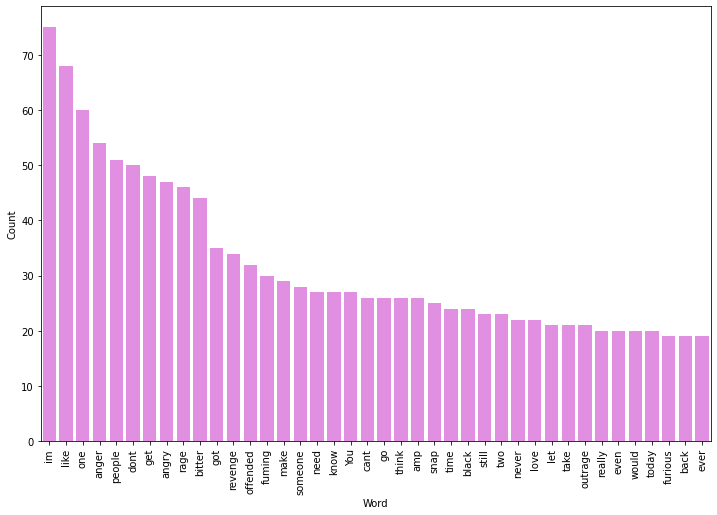
Words in JOY Vs Frequency of word:



##### Words in FEAR Vs Frequency of word:

##### d2.png

Words in ANGER Vs Frequency of word:

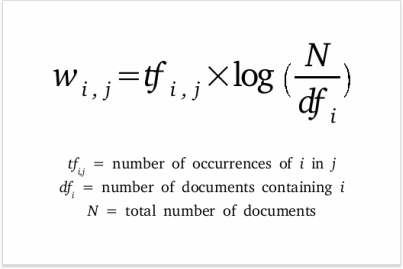


##### Words in SADNESS Vs Frequency of word:

##### d4.png

**CONVERTING TEXT INTO VECTOR:**

Converting the text into tfidf-vector for model building using scikit learn library.



**MODEL BUILDING:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.NO** | **MODEL NAME** | **ACCURACY%** |  |
| **1** | **Logistic regression** | **82** |  |
| **2** | **Naïve Bayes** | **73** |  |
| **3** | **Decision Tree** | **80** |  |
| **4** | **Random Forest** | **85** |  |
| **5** | **SVM** | **85** |  |
| **6** | **Passive Aggresive** | **85** |  |