

# **Day 70**

#### DIY

### **Q1. Problem Statement: Sentiment Analysis**

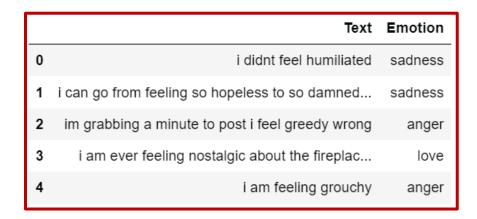
Write a Python program that reads the *mood\_data.txt* (provided on LMS) file into a DataFrame. Then the following the given tasks, that has to be taken into consideration while constructing the solution.

Here dataset contains two columns where one is our target ("emotion" has 6 different categories) and another is the independent variable ("Text" contains data in form of sentences).

- 1. Load the mobile mood\_data.txt data into a DataFrame
- 2. Generate tokens and remove punctuations, stop words and lower all rows
- Join all the tokens as they were before and store them in a new column named "cleaned text"
- 4. Now remove all single characters, extra space, and special characters and store processed data in a new column named "processed\_text"
- Create a final DataFrame containing dependent variable(emotion) and processed text
- 6. Extract independent variables (Xs) and dependent variables (Ys) into separate data objects
- 7. Generate tokens and do vectorization
- 8. Build a model with Multinomial Naive Bayes, Random Forest, Random Forest (Entropy), SVM and compare their accuracy

#### **Dataset:**





## **Sample Output:**

3. Join all the tokens as they were before and store them in a new column named "cleaned\_text"





4. Now remove all single characters, extra space, and special characters and store processed data in a new column named "processed\_text"



8. Build a model with Multinomial Naive Bayes, Random Forest, Random Forest (Entropy), SVM and compare their accuracy

Model Accuracy

Multinomial Naive Bayes 0.740625

Random Forest(Gini) 0.828958

Random Forest(Entropy) 0.811458

SVC by SVM 0.810833

