

## **Experiment No: 01**

### **Group Members:**

<b>Name</b>	<b>Roll No</b>
Abbasali Agharia	01
Yash Chand	05
Devang Mate	32

**Title:** Twitter Sentiment Analysis

### **Abstract:**

In an age where social media has emerged as a primary channel for individuals and businesses to express themselves and engage with their audiences, understanding the sentiments conveyed in Twitter posts has become increasingly crucial. Twitter Sentiment Analysis, a subfield of Natural Language Processing (NLP), provides the means to computationally identify and categorize tweets, enabling us to discern the emotional tone and stance of the authors. The project involves several key steps, including data preprocessing, feature extraction, and the training of a Naive Bayes classifier for sentiment analysis. By leveraging NLP techniques, this project contributes to understanding public sentiment on Twitter.

## **Methodology:**

### **1. Data Collection and Preprocessing:**

The project begins by importing the necessary libraries, including NumPy and Pandas, and reading a CSV file containing tweet data.

The data is filtered to retain only the 'text' and 'sentiment' columns.

The dataset is split into training and test sets, with neutral sentiments excluded from the training set.

### **2. Word Cloud Visualization:**

Two word clouds are generated to visualize positive and negative words. These visualizations provide insights into the most frequently occurring words in tweets associated with each sentiment.

### **3. Text Tokenization and Filtering:**

Tweets in the training set are tokenized, and words are filtered to remove URLs, mentions, hashtags, and the "RT" (retweet) tag.

Stop words from the NLTK library are removed to clean the text further.

### **4. Feature Extraction:**

Word features are extracted by analyzing the frequency distribution of words in the cleaned tweets.

These features are used to create a feature set for training the sentiment analysis model.

### **5. Sentiment Analysis Model (Naive Bayes):**

A Naive Bayes classifier is trained using the feature set and associated sentiments.

The classifier is capable of categorizing tweets as either 'Positive' or 'Negative' based on the extracted word features.

### **6. Sentiment Classification and Evaluation:**

The trained Naive Bayes classifier is used to classify tweets in the test set as 'Positive' or 'Negative.'

The number of correctly classified 'Negative' and 'Positive' tweets is recorded and compared to the total number of tweets in each category.

### **7. Results:**

The project provides an assessment of the Naive Bayes classifier's performance in classifying sentiments and presents the accuracy of the sentiment analysis on the test dataset.