# Twitter Sentiment Analysis

## Project Overview:

Twitter Sentiment Analysis is a data analytics project that involves analyzing a dataset of tweets to determine the sentiment expressed in each tweet—whether it is positive or negative. The project aims to gain insights into public opinions, trends, and sentiments shared on Twitter, utilizing data analytics techniques.

## Project Objectives:

➢Data Exploration:

- Explored the Sentiment dataset to understand its structure, features, and size.
- Identified key variables such as tweet content, timestamp, and sentiment labels.

- Data Cleaning:
  - Performed data cleaning tasks to handle missing values, duplicate entries, and irrelevant information.
  - Ensured data quality by addressing any anomalies or inconsistencies in the dataset.

- Exploratory Data Analysis (EDA):
  - Conducted exploratory data analysis to gain initial insights into tweet patterns, sentiment distributions, and temporal trends.
  - Utilized visualizations (e.g., histograms, word clouds) to represent key aspects of the dataset.

- Sentiment Distribution:
  - Visualized the distribution of sentiment labels (positive, negative) in the dataset.
  - Analyzed the balance of sentiment classes to understand potential biases.

- Word Frequency Analysis:
  - Analyzed the frequency of words in tweets to identify common terms and themes.
  - Created word clouds to visualize the most frequent words in positive and negative sentiments.

- Text Preprocessing:
  - Preprocessed tweet text by removing stop words, special characters, and URLs.
  - Tokenized and lemmatized words to prepare the text for sentiment analysis.

- Sentiment Prediction Model:
  - Implemented a sentiment prediction model using machine learning or natural language processing techniques.
  - - Trained the model on a subset of the dataset and evaluated its performance using metrics like accuracy and F1 score.

**Insights gained:**

- Identified the length of positive and negative tweets.
- Analyzed the most frequent time stamps of tweets.
- Found out the most common words in tweets using counter module and wordcloud.

- Performed feature engineering using CountVectorizer for bagofwords and used pipeline to combine all preprocessing and feature extraction.
- Implemented Bernouli naive bayes to train the model for binary classification.
- Utilized performance metrics like accuracy, confusion matrix, precision, recall, f1-score to check the performance of the model.
- Deployment: dumped the model using pickle module
- Created a web application using a framework called streamlit.