Social Media Sentiment Analysis using VADER

# 1. Objective

The objective of this project is to analyze public sentiment on Twitter using Natural Language Processing (NLP) techniques.   
The analysis uses a labeled dataset to predict whether a tweet expresses a positive, negative, or neutral opinion.   
This enables us to understand public mood and opinions about events, brands, or social issues in real-time.

# 2. Dataset

We use the Sentiment140 dataset from Kaggle, which contains 1.6 million tweets labeled as either positive or negative.   
It includes tweets, their text, and sentiment labels (0 for negative, 4 for positive).   
For the sake of balance and analysis, we use a representative sample of these tweets.

# 3. Tools & Libraries Used

- Python  
- Pandas  
- NLTK  
- Matplotlib  
- Seaborn  
- WordCloud  
- VaderSentiment

# 4. Data Preprocessing

Text data is cleaned by removing URLs, mentions, hashtags, numbers, and special characters.   
We also tokenize the text and remove stopwords. Finally, text is converted to lowercase.   
This preprocessing ensures better accuracy for sentiment prediction models.

# 5. Sentiment Analysis using VADER

We use VADER (Valence Aware Dictionary for sEntiment Reasoning), a rule-based sentiment analysis tool designed for social media text.   
VADER returns a compound score for each tweet, which is then classified as Positive (score ≥ 0.05),   
Negative (score ≤ -0.05), or Neutral (score in between).

# 6. Results & Visualizations

The sentiment analysis shows distribution of tweets into positive, negative, and neutral categories.   
We also use word clouds to visualize frequently used terms in the tweets.   
Positive tweets dominate the dataset with a substantial number of neutral and negative tweets as well.

# 7. Conclusion

This project shows how social media sentiment analysis can be performed using simple NLP tools.   
It can be extended to real-time applications using Twitter APIs, and improved using deep learning-based sentiment classifiers.   
This analysis is beneficial for businesses, researchers, and government agencies to understand public sentiment trends.