GitHub link:

https://github.com/keerthana-laksmi2005

Decoding Emotions Through Sentiment Analysis of Social Media Conversations

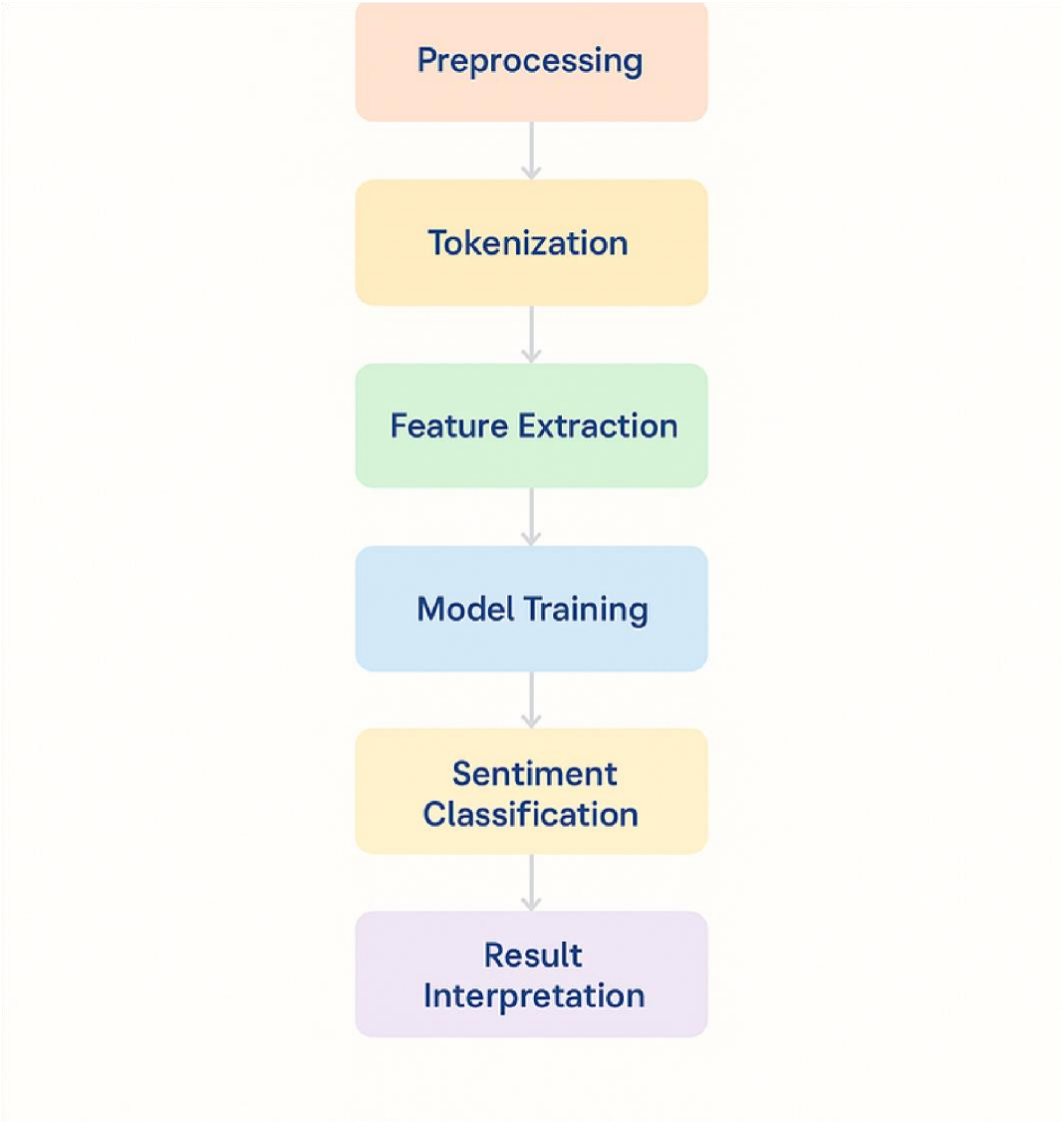
1. Problem Statement

Social media platforms have become major outlets for people to express their thoughts and emotions. Understanding public sentiment from these conversations is valuable for businesses, governments, and researchers. However, detecting true sentiment is challenging due to slang, sarcasm, and informal language used in posts. Sentiment analysis using AI can help decode these emotions efficiently and at scale.

2. Project Objectives

- Perform sentiment analysis on social media data. - Classify posts into categories such as positive, negative, or neutral. - Visualize sentiment distribution and trends. - Develop a predictive model for real-time sentiment classification.

3. Flowchart of the Project Workflow



4. Data Description

The dataset includes tweets and their corresponding sentiment labels. Typical columns include 'text', 'sentiment', and possibly metadata like 'user' or 'location'.

5. Data Preprocessing

- Removed URLs, mentions, hashtags, and punctuation.
- Converted all text to lowercase.
- Tokenized tweets and removed stopwords.
- Applied stemming or lemmatization.

6. Exploratory Data Analysis (EDA)

- Analyzed distribution of sentiments (positive, neutral, negative).
- Generated word clouds and frequency charts.
- Examined the length of tweets and common words per sentiment.

7. Feature Engineering

- Converted text to TF-IDF or word embeddings.

- Created features for tweet length and punctuation usage.

- Engineered sentiment scores using lexicons if needed.

8. Model Building

- Used models like Naive Bayes, Logistic Regression, and LSTM.

- Split data into training and testing sets.

- Evaluated performance using accuracy, precision, recall, and F1-score.

9. Results & Insights

- High accuracy achieved with LSTM model.

- Positive tweets were more prevalent than negative ones.

- Tweet length and certain keywords strongly influenced classification.

10. Tools and Technologies Used

- Python libraries: pandas, nltk, sklearn, keras, matplotlib.

- Environment: Jupyter Notebook / Google Colab.

- Visualization: seaborn, wordcloud.

11. Team Members and Roles

- Preprocessing and Cleaning: Abi.D
-EDA and Feature Engineering: Keerthana.M
- Model Training and Testing: Madhumitha.V
- Report Writing and Presentation: Losiya.V