

Decoding Emotions through Sentimental Analysis of Social Media Conversation

Phase:2

STUDENT NAME: SIVA JENANI.S

REGISTER NUMBER: 422623104002

INSTITUTION: UNIVERSITY COLLEGE OF ENGINEERING-PANRUTI

DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING

DATE OF SUBMISSION: 10.05.2025

GITHUB REPOSITORYLINK: https://github.com/SJenani/JenaniS_sentiment_analysis.git

1.Problem Statement:

- Social media platforms have become primary sources for individuals to express emotions, opinions, and reactions. Understanding these emotional expressions through sentimental analysis is critical for various industries, such as marketing, politics, public health, and customer service.
- This project focuses on decoding emotions embedded in social media conversations using sentiment analysis techniques.
- The problem is a **classification problem** (positive, negative, neutral, or emotion categories like happiness, sadness, anger, etc.).
- By solving this problem, organizations can better understand public opinion, enhance user engagement, and make data-driven decisions.

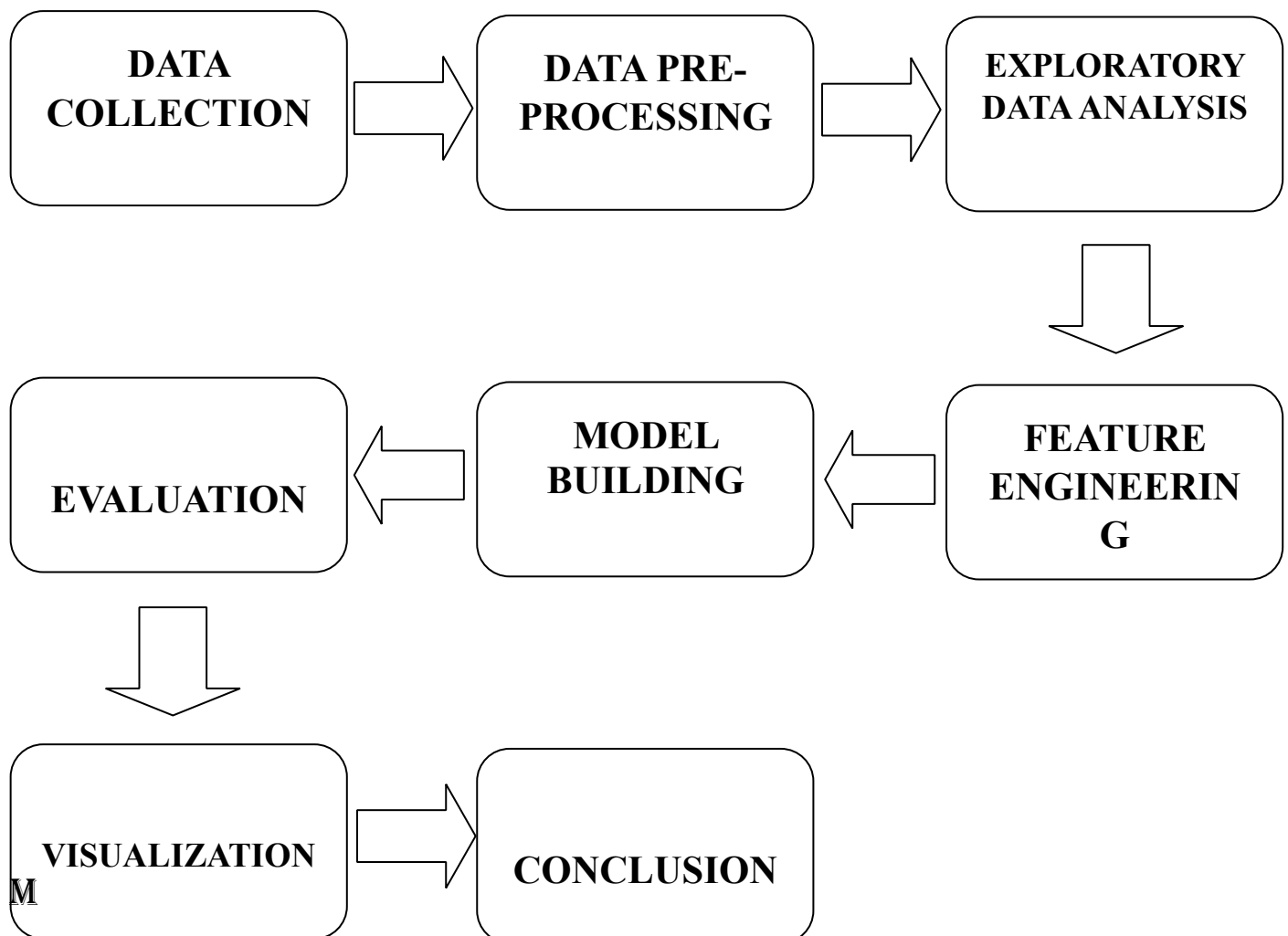
2. Project Objectives:

- Build a machine learning model to classify sentiments in social media conversations.

- Achieve high accuracy and generalization on unseen social media data.
- Analyze the emotional trends over time and across different topics
- Develop a solution that can be integrated into real-world applications like brand monitoring, customer feedback analysis, and crisis detection.

3. Flowchart of the Project Workflow:

Data Collection → Data Preprocessing → Exploratory Data Analysis → Feature Engineering → Model Building → Evaluation → Visualization → Conclusion



4. Data Description:

- **Dataset Name:** Twitter Sentiment140 / Kaggle Datasets (or similar dataset from open APIs).
- **Type of Data:** Textual (Unstructured).
- **Records and Features:** Approximately 1.6 million tweets; features include text, sentiment label, user info, etc
- **Static or Dynamic:** Static dataset for this phase.
- **Target Variable:** Sentiment label (e.g., Positive, Negative, Neutral).

5. Data Preprocessing:

- Removal of unwanted characters (hashtags, mentions, URLs).
- Lowercasing of text.
- Handling missing values.
- Tokenization and lemmatization.
- Encoding target labels (e.g., Positive = 1, Negative = 0).
- Splitting data into training and testing sets.

6. Exploratory Data Analysis (EDA):

- Univariate analysis: distribution of sentiment labels using countplots.

- Word clouds for positive and negative sentiments.
- Bivariate analysis: frequent word pairs using bigrams.
- Sentiment trends over time (if timestamp available).

Insights:

- Positive sentiments are more dominant.
- Specific keywords correlate strongly with particular emotions.

7. Feature Engineering: • Text vectorization using TF-IDF / Word Embeddings (Word2Vec, GloVe).

- Creating new features like text length, number of hashtags, etc.
- Using n-grams (bigrams, trigrams) for better text representation.
- Optional dimensionality reduction using PCA or TruncatedSVD for visualization.

8. Model Building:

- Models implemented: Logistic Regression, Random Forest Classifier, and LSTM (optional advanced).
- Justification:
 - Logistic Regression for baseline performance.
 - Random Forest for non-linear text classification.
 - LSTM for capturing sequential patterns in conversations (if deep learning is used).
- Performance Metrics: Accuracy, Precision, Recall, F1-Score.

9. Visualization of Results & Model Insights:

- Confusion Matrix to understand classification performance.
 - ROC-AUC Curves for model comparison.
 - Feature importance plots (for traditional ML models).
 - Word Importance using SHAP values (optional advanced).
-

10. Tools and Technologies Used: • Programming Language: Python

- **IDE/Notebook:** Google Colab / Jupyter Notebook
- **Libraries:** pandas, numpy, scikit-learn, NLTK, seaborn, matplotlib, TensorFlow (optional for deep learning).
- **Visualization Tools:** seaborn, matplotlib, WordCloud.

11. Team Members and Contributions:

Team members	Roles
MADHUMATHI S	Data Collection & Preprocessing
KAVIYA P	EDA & Feature Engineering
SIVA JENANI S	Model Building & Evaluation
DHANUSH M	Visualization & Reporting
RAGAVAN R	Brainstorming & Reviews

