Project Statement Document

Problem Statement

With the rapid growth of social media platforms such as Twitter, Facebook, and Instagram, millions of users share their opinions, emotions, and feedback daily. Manually analyzing such large volumes of unstructured text to understand public sentiment is not feasible. Organizations need automated systems capable of extracting sentiment patterns, detecting user opinions, and identifying emotional tone in real time. This project addresses the need for a scalable Machine Learning-based sentiment analysis system to classify text into positive, negative, or neutral sentiments.

Scope of the Project

The scope of this project includes the complete development of an end-to-end sentiment analysis application using Machine Learning techniques. The project covers dataset collection, text cleaning, preprocessing, model training, evaluation, deployment, and testing. The system will allow users to input any social media text and receive real-time sentiment classification. It will also include visualizations and performance metrics for better understanding. The scope excludes advanced deep learning techniques and multilingual sentiment analysis unless added in future versions.

Target Users

This project is designed for the following target users: 1. **Businesses and Organizations** – To analyze customer opinions and monitor brand sentiment. 2. **Researchers and Data Scientists** – To study sentiment trends and improve NLP models. 3. **Students and Educators** – For academic projects, learning NLP, and understanding sentiment analysis workflows. 4. **Social Media Analysts** – To track audience engagement and public reactions. 5. **Developers** – To integrate sentiment analysis into applications or APIs.

High-Level Features

1. **Automated Text Preprocessing** Includes tokenization, stopword removal, punctuation cleaning, and lemmatization to prepare text for modeling. 2. **Machine Learning-Based Sentiment Classification** Supports multiple ML algorithms such as Logistic Regression, SVM, and Naive Bayes for accurate sentiment prediction. 3. **Sentiment Visualization Dashboard** Generates charts and graphical insights showing sentiment distribution and trends. 4. **Real-Time Sentiment Prediction** Allows users to input custom text and receive instant sentiment analysis results. 5. **Model Evaluation and Reporting** Provides performance metrics including accuracy, precision, recall, F1-score, and confusion matrix. 6. **Modular and Scalable Architecture** Designed so new datasets, ML models, or APIs can be easily integrated in the future.