Project Report

On

SOCIAL MEDIA TIMELINE ANALYSIS

Submitted in partial fulfillment for the award of

Diploma in Big Data Analytics (PG-DBDA)

from C-DAC, ACTS (Hyderabad)

Guided by:

Mr. Ganga Prasad R

Presented by:

Mr. Sravani Surasani	**PRN Number 240850325035**
Mr. Mohsin Mirza	**PRN Number 240850325022**
Mr. Yash Sonawane	**PRN Number 240850325031**
Mr. Bopalkar Pranav	**PRN Number 240850325008**
Mr. Balarami Reddy	**PRN Number 240850325007**

Centre for Development of Advanced Computing (C-DAC), Hyderabad

ACKNOWLEDGEMENT

This project "**SOCIAL MEDIA SENTIMENTAL ANALYSIS**" was a great learning experience for us and we are submitting this work to Advanced Computing Training School (CDAC ACTS). We are very glad to mention the name of *Mr. Ganga Prasad R* for his valuable guidance to work on this project. His guidance and support helped me to overcome various obstacles and intricacies during the course of project work.

We are highly grateful to Manager (ACTS training Centre), C-DAC, for her guidance and support whenever necessary while doing this course Diploma in *Big Data Analytics (PG-DBDA)* through C-DAC ACTS, Hyderabad.

Our heartfelt thanks goes to *Mr.Sadhu Srinivas* (Course Coordinator, PG-DBDA) who gave all the required support and kind coordination to provide all the necessities and extra hours to complete the project and throughout the course up to the last day here in C-DAC ACTS, Hyderabad.

From:

Mr. Sravani Surasani	**PRN Number 240850325035**
Mr. Mohsin Mirza	**PRN Number 240850325022**
Mr. Yash Sonawane	**PRN Number 240850325031**
Mr. Bopalkar Pranav	**PRN Number 240850325008**
Mr. Balarami Reddy	**PRN Number 240850325007**

TABLE OF CONTENTS

1. Introduction of Project:

In today's digital age, social media platforms such as Twitter, Facebook, and Instagram have become major sources of public opinion and user-generated content. People express their thoughts, opinions, and emotions on various topics, making social media a valuable source of data for sentiment analysis. Sentiment analysis, also known as opinion mining, is a technique used in Natural Language Processing (NLP) to determine the emotional tone behind a text. It helps in understanding whether the sentiment expressed in a post, comment, or tweet is **positive, negative, or neutral**. Businesses, governments, and researchers use sentiment analysis to gauge public opinion, track trends, and make data-driven decisions.

This project focuses on implementing a **Social Media Sentiment Analysis** system that extracts, processes, and analyzes textual data from social media platforms. The system utilizes machine learning and NLP techniques to classify sentiments, providing insights into public perception on various topics such as brand reputation, political opinions, or customer feedback.

The main objectives of this project include:

- Collecting social media data through APIs or web scraping.
- Preprocessing and cleaning the text data for accurate analysis.
- Applying sentiment classification models such as **ARIMA,FACEBOOK PROPHET**
- Visualizing sentiment trends through graphs and dashboards for better interpretation.

By implementing sentiment analysis, this project aims to showcase how businesses and organizations can leverage social media insights for decision-making, crisis management, and enhancing customer experiences.

2. Product Overview and Summary

2.1 Purpose:

The purpose of this project is to analyze sentiments expressed on social media platforms and classify them into positive, negative, or neutral categories. As social media serves as a powerful tool for communication and opinion sharing, businesses, governments, and researchers can leverage sentiment analysis to gain insights into public perception, market trends, and brand reputation.

2.2 Scope and Overview:

This project will cover multiple aspects of sentiment analysis, focusing on the following areas:

Data Collection: Gathering social media data through APIs (e.g., Twitter API) or web scraping.

Preprocessing: Cleaning and normalizing text data, removing stop words, emojis, special characters, and handling misspellings.

Sentiment Classification: Implementing machine learning or deep learning algorithms like Naïve Bayes, SVM, or LSTMs to classify text.

Visualization & Insights: Presenting findings through graphs, dashboards, and trend analysis to identify patterns over time.

Real-world Applications: Analysing customer reviews, brand mentions, and public sentiment on trending topics for marketing, political, and business intelligence.

2.3 Feasibility Study

- **Technical Feasibility:**
  The project requires expertise in Natural Language Processing (NLP), machine learning, and data visualization. It will utilize programming languages such as Python along with libraries like NLTK, TextBlob, VADER, Scikit-learn and Matplotlib. Cloud services or local databases may be used for storing and processing large datasets.

- **Operational Feasibility:**

Organizations, marketers, and researchers can easily integrate sentiment analysis into their decision-making processes. The system can be automated to continuously monitor and analyze real-time social media trends, improving operational efficiency.

3. Overall Description:

Overall Description

Social media sentiment analysis is a data-driven approach used to determine the sentiment (positive, negative, or neutral) behind text-based content shared on social media platforms. This project aims to extract, preprocess, analyse, and visualize sentiment trends from platforms like Twitter, Facebook, and Instagram.

The key objectives of this project include:

- Automated Sentiment Classification: Using NLP and machine learning techniques to classify social media posts.
- Trend Analysis & Visualization: Representing sentiment trends over time using graphical analysis.
- Business & Social Impact: Helping businesses understand customer feedback, tracking public sentiment on social issues, and analysing brand reputation.
-

3.1 Features of the System:

- Data Collection: Extracting real-time data from social media using APIs or web scraping techniques.
- Text Preprocessing: Cleaning and normalizing data (removing stop words, special characters, handling misspellings, etc.).
- Sentiment Detection: Using machine learning and deep learning models for sentiment classification.
- Data Visualization: Displaying sentiment trends using graphs, charts, and dashboards.

3.2 Technologies Used:

To develop the Social Media Sentiment Analysis system, the following technologies will be utilized:
Programming Languages & Frameworks

- Python: Primary programming language for data processing, machine learning, and visualization.
- Jupyter Notebook/PyCharm: Development environments for writing and testing code.

Data Collection & Processing

- BeautifulSoup: For web scraping if API access is restricted.
- Pandas & NumPy: For data manipulation and handling large datasets.

Natural Language Processing (NLP) Libraries

- NLTK (Natural Language Toolkit): For text preprocessing, tokenization, and sentiment analysis.
- TextBlob & VADER: For lexicon-based sentiment analysis of short text and social media comments.

Machine Learning & Deep Learning

- Scikit-learn: For implementing ML models like Naïve Bayes, SVM, and Random Forest.
- TensorFlow: For deep learning models such as LSTMs for sentiment classification.

Data Visualization & Dashboarding

- Matplotlib & Seaborn: For generating visual representations of sentiment trends.
- Plotly & Dash: For interactive dashboards to display sentiment insights.

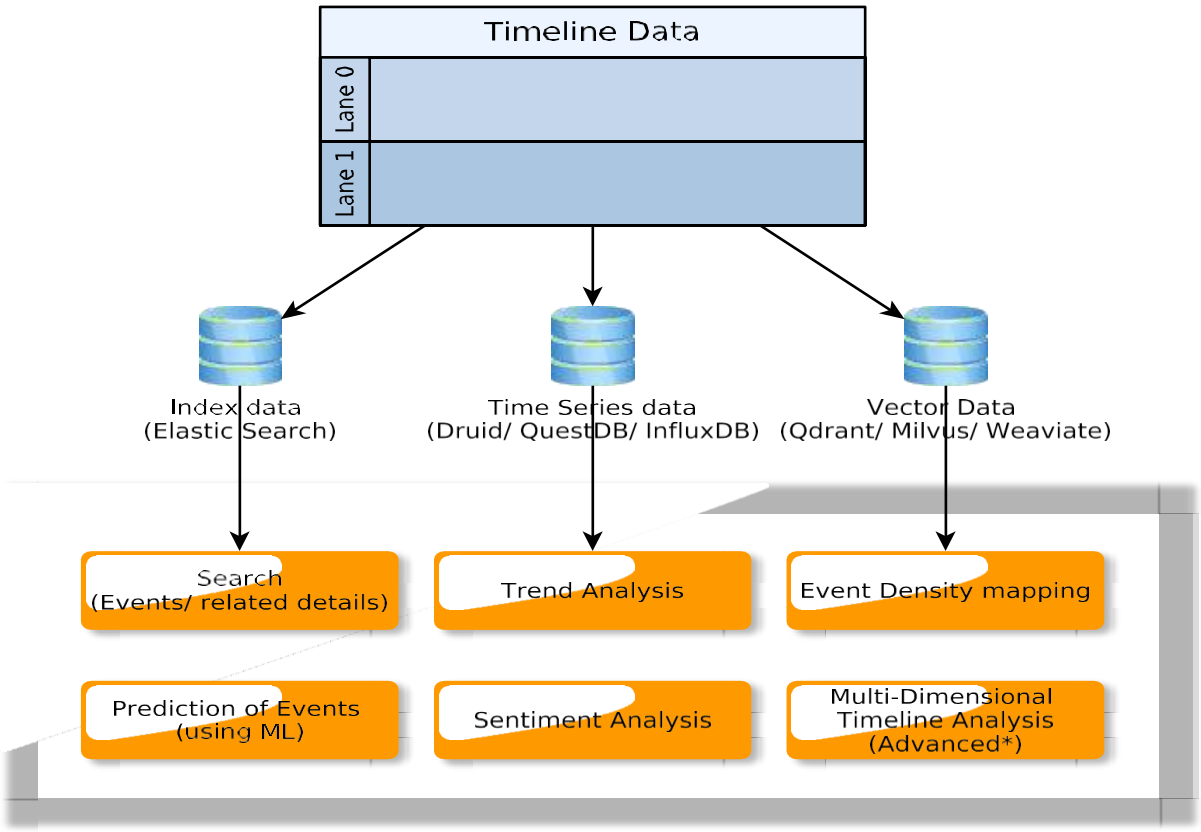
Database & Storage

- MySQL / SQL lite: For structured data storage.
- ChromaDB: For handling large amounts of social media data.
- Apache Druid: For scalability and handling large datasets.

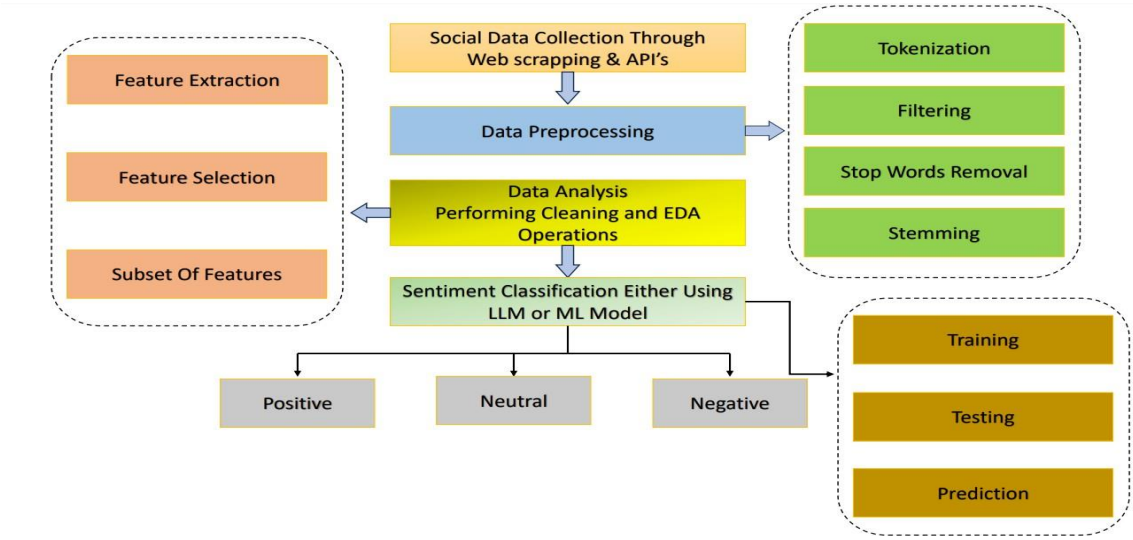
4. REQUIREMENTS

4.1 FUNCTIONAL REQUIREMENTS

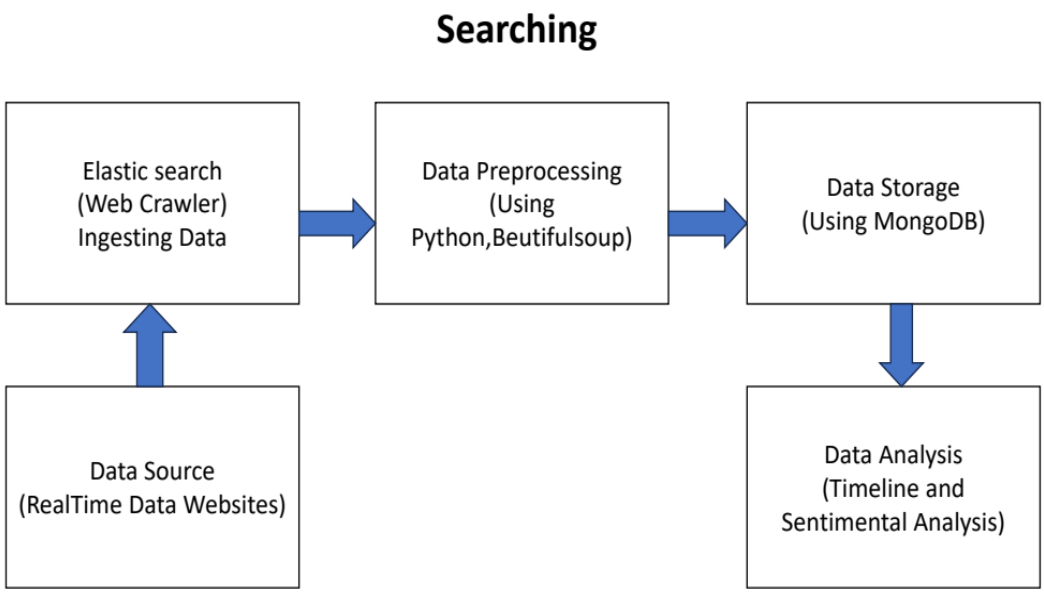
4.1.1 Complete System:



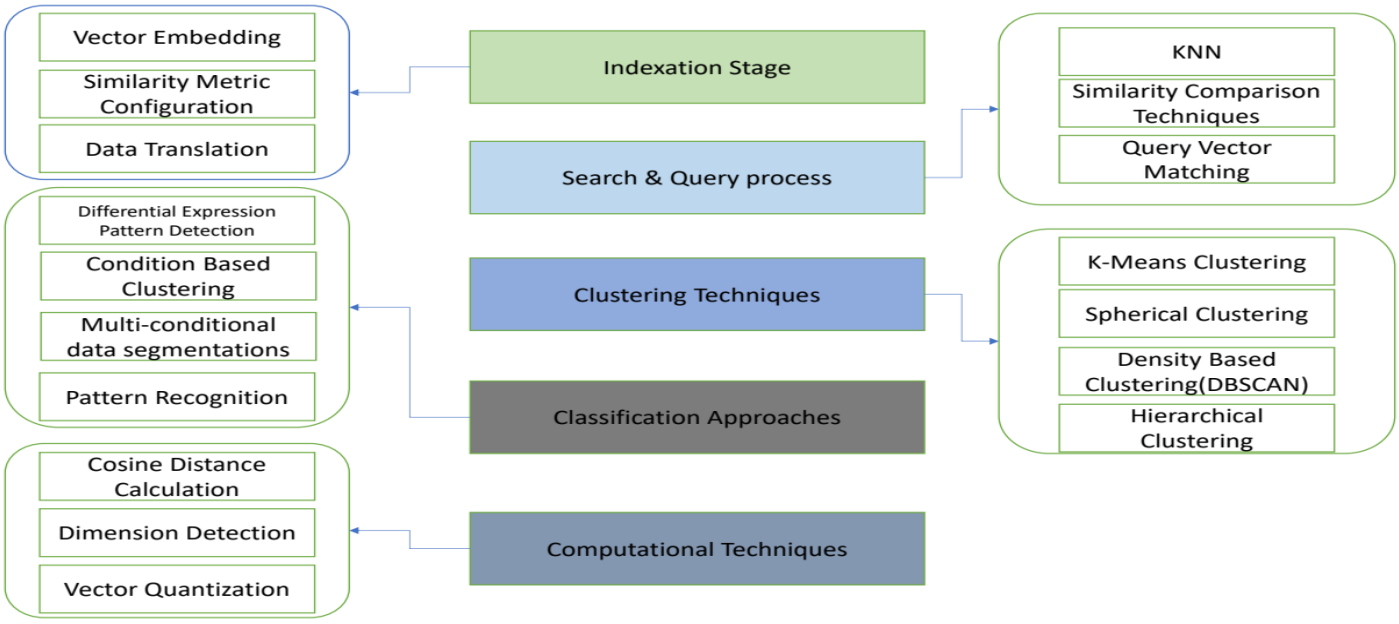
4.1.2 Detailed Data Pipeline:



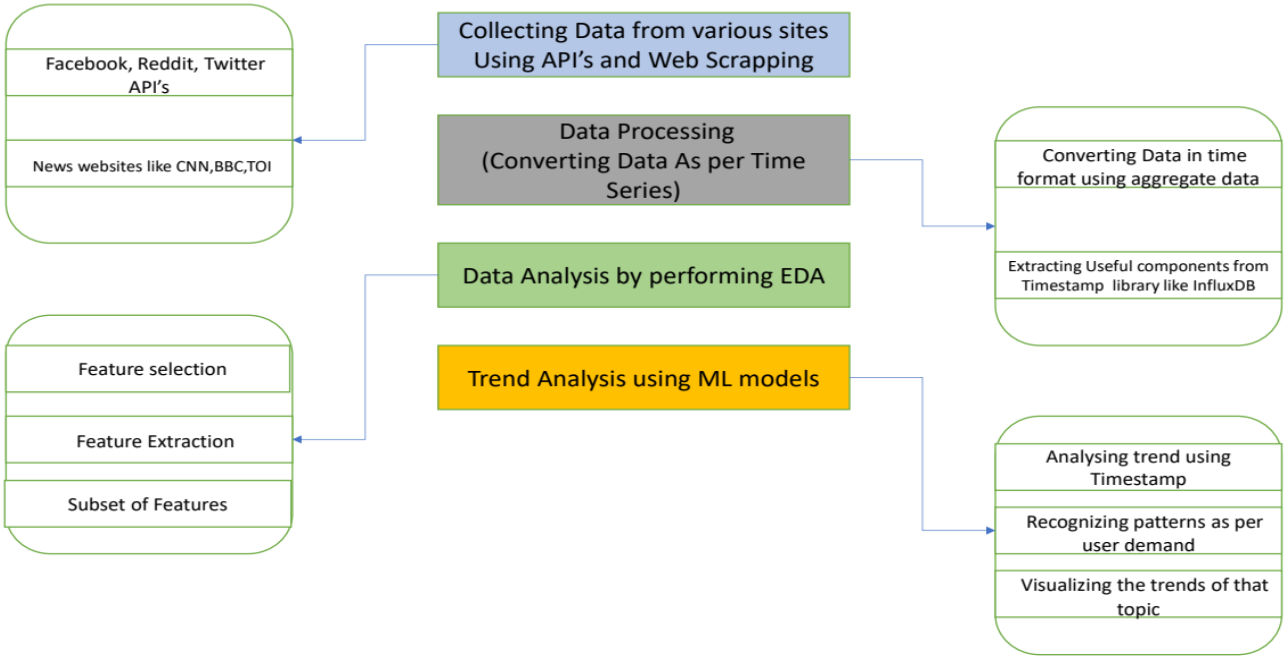
4.1.3 Initial Stage (searching)



4.1.4 Event Density Mapping

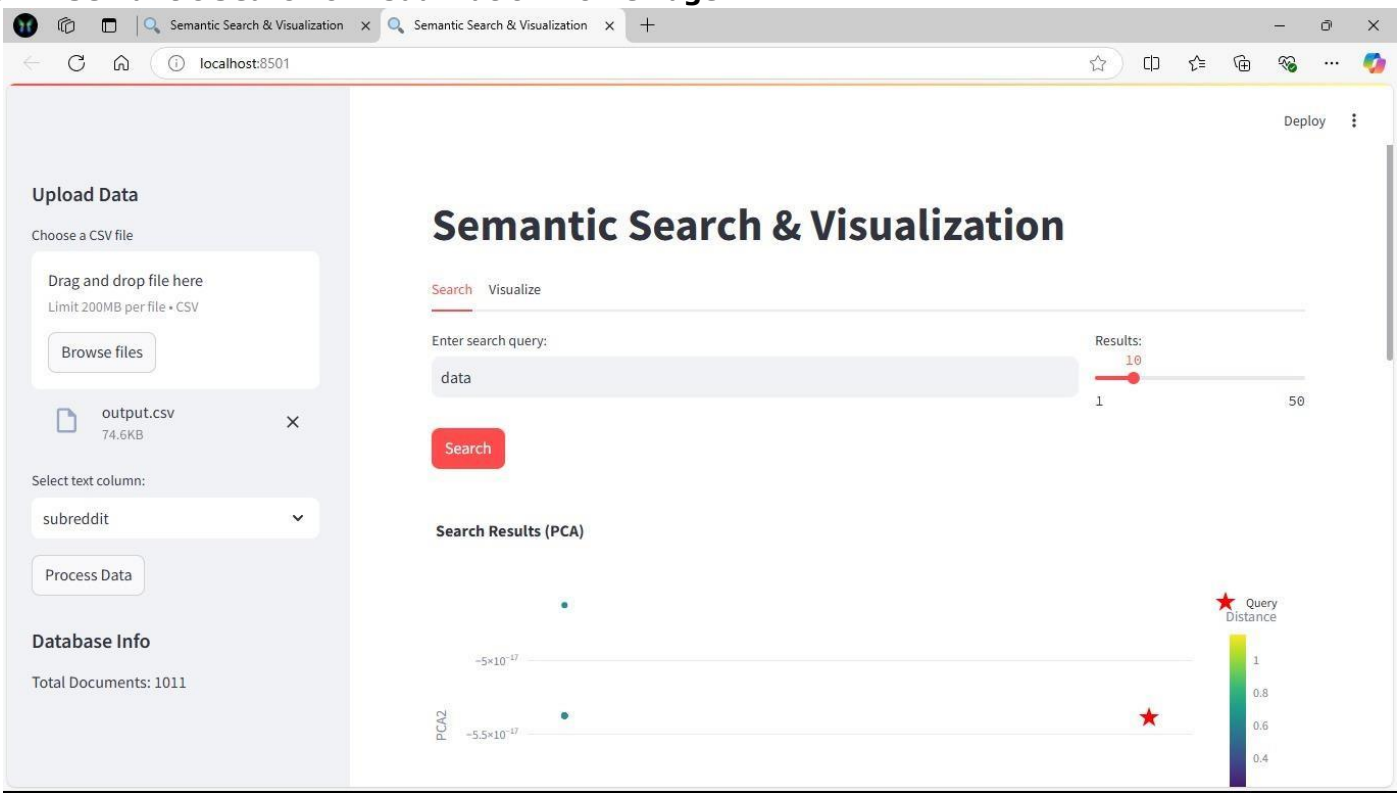


4.1.5 Data Processing

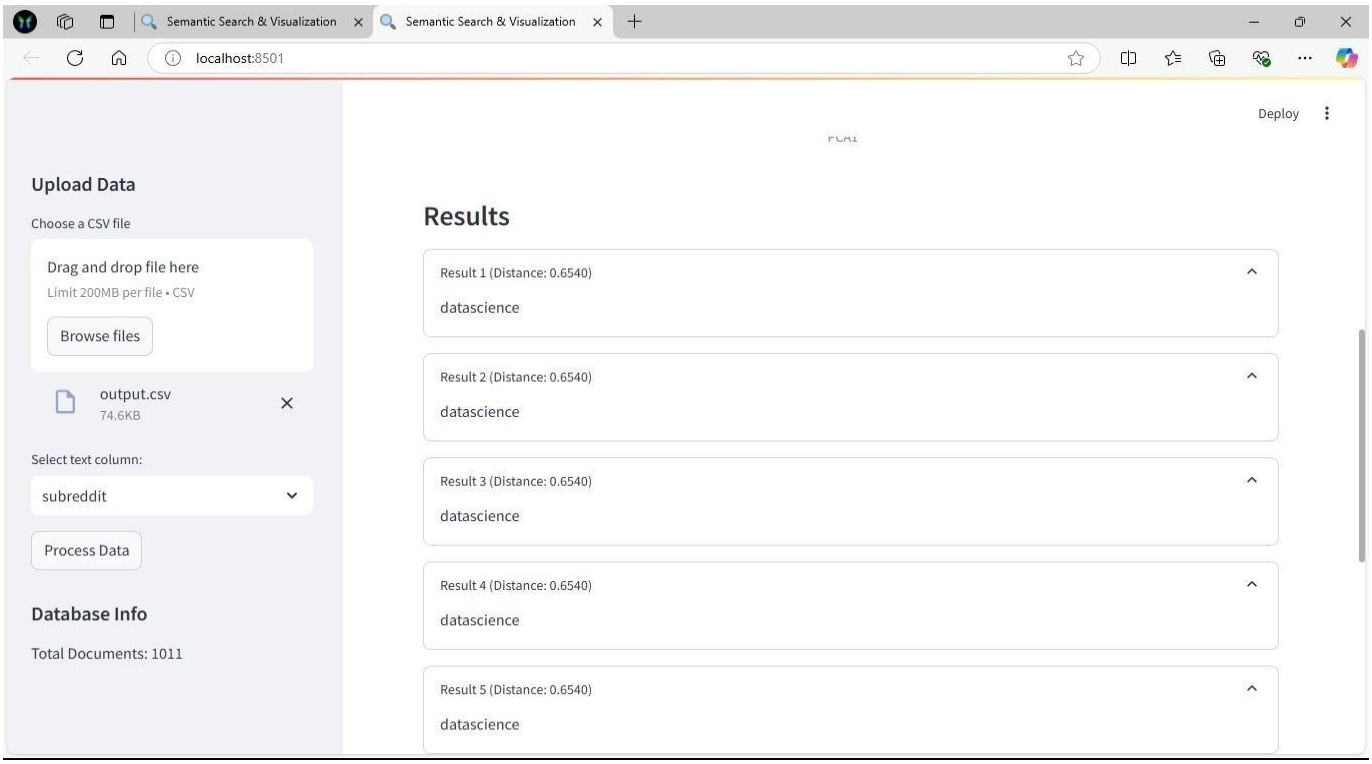


5.1 User Interface:

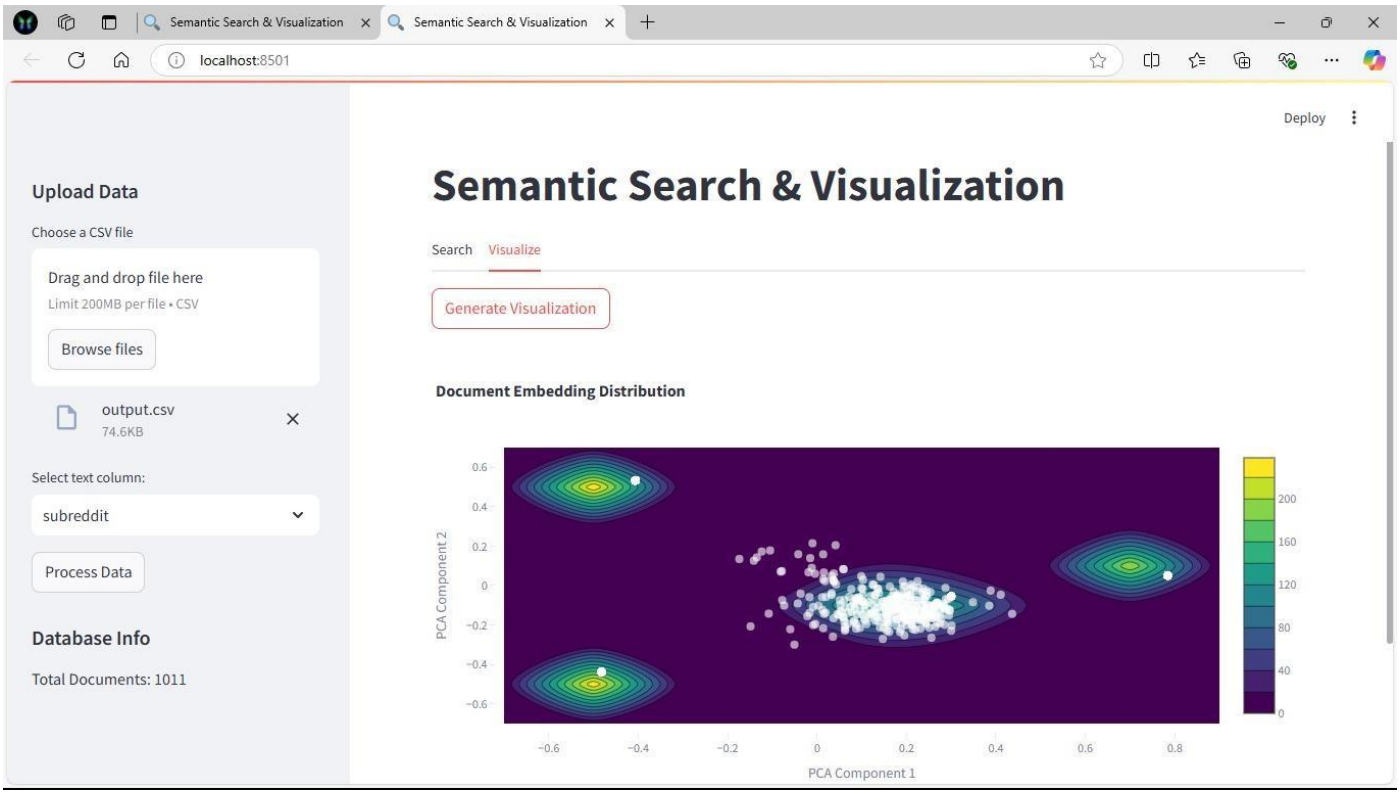
5.1.1Semantic Search & Visualization Home Page:



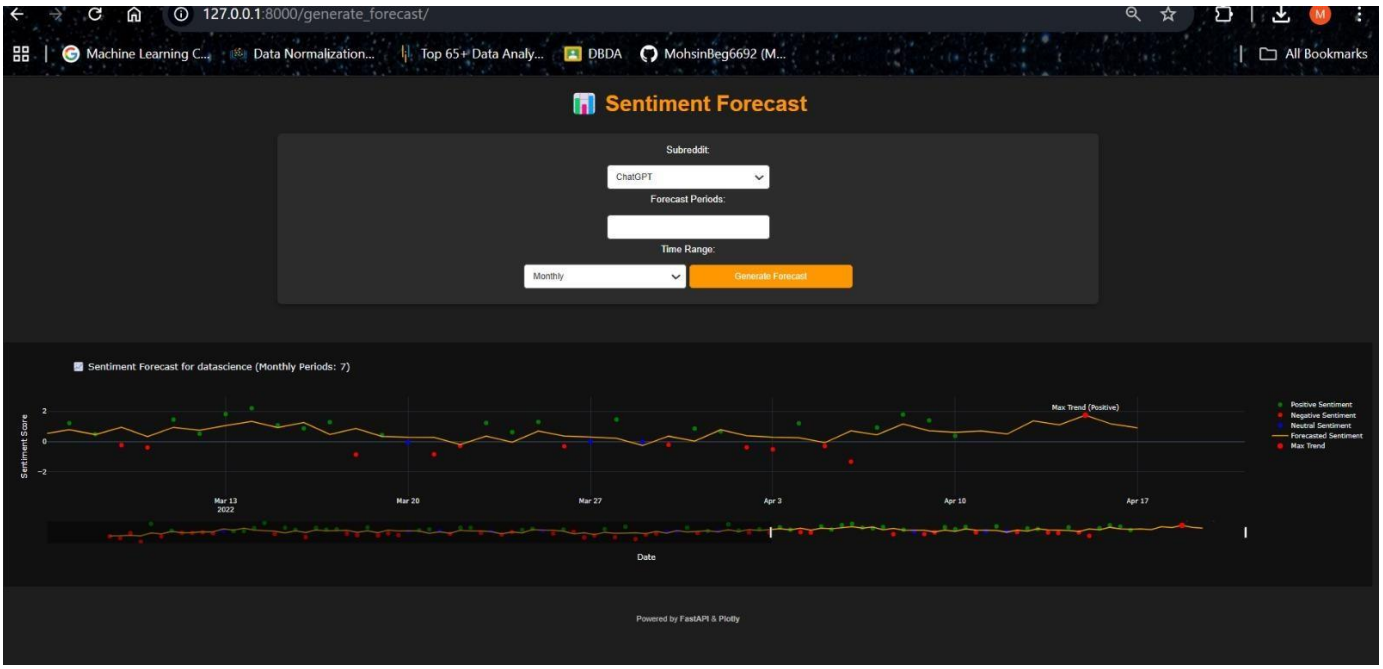
5.1.2 Results:

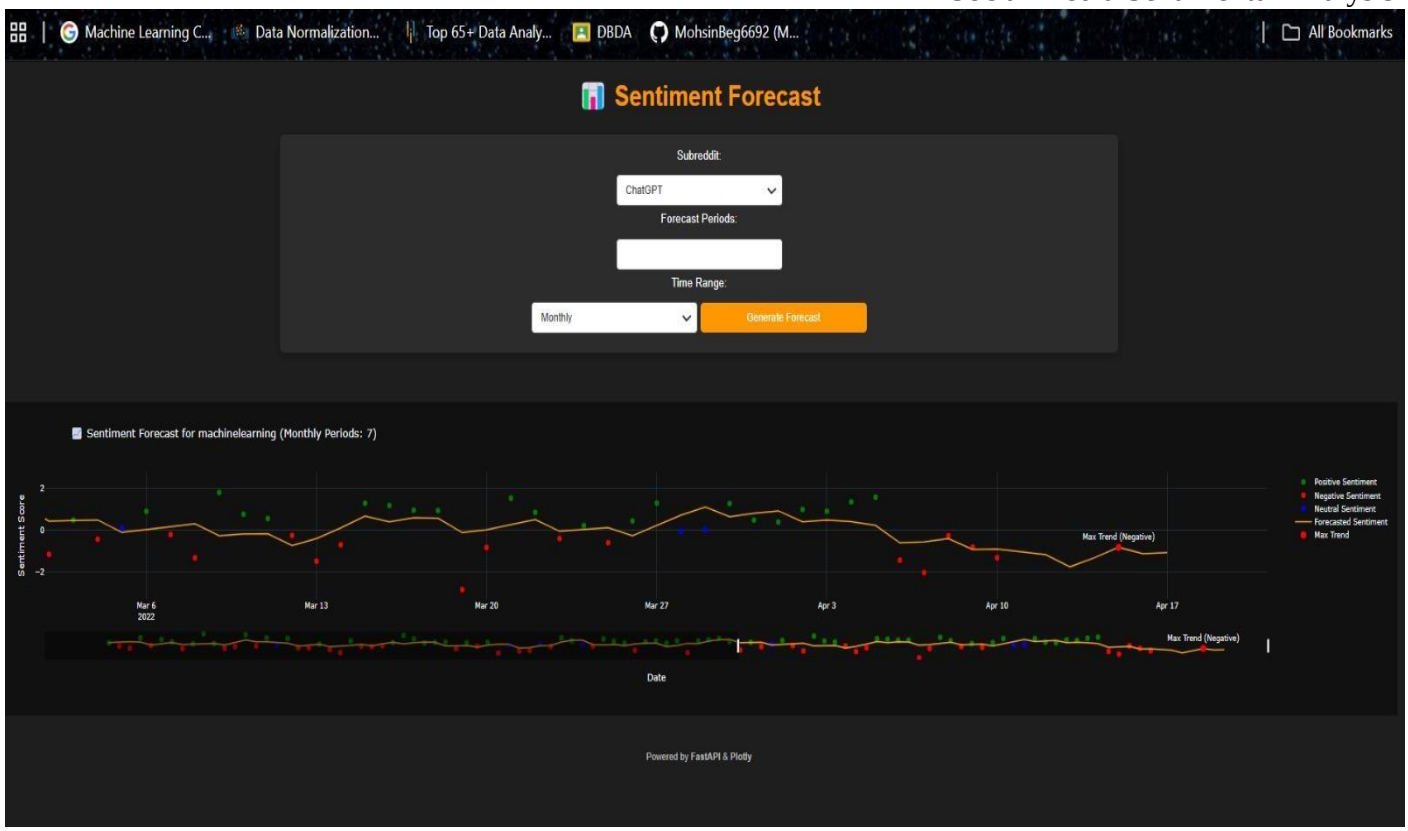


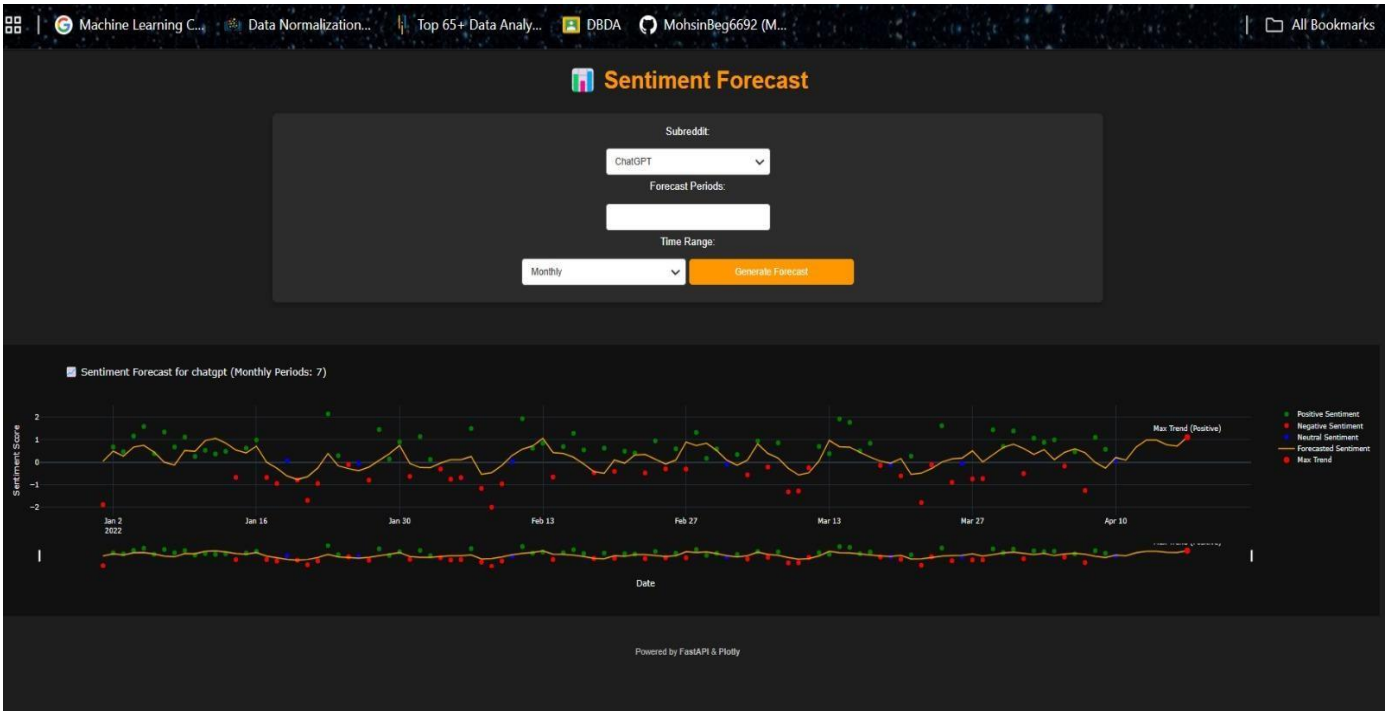
5.1.3 Visualization (Document Embedding Visualization):



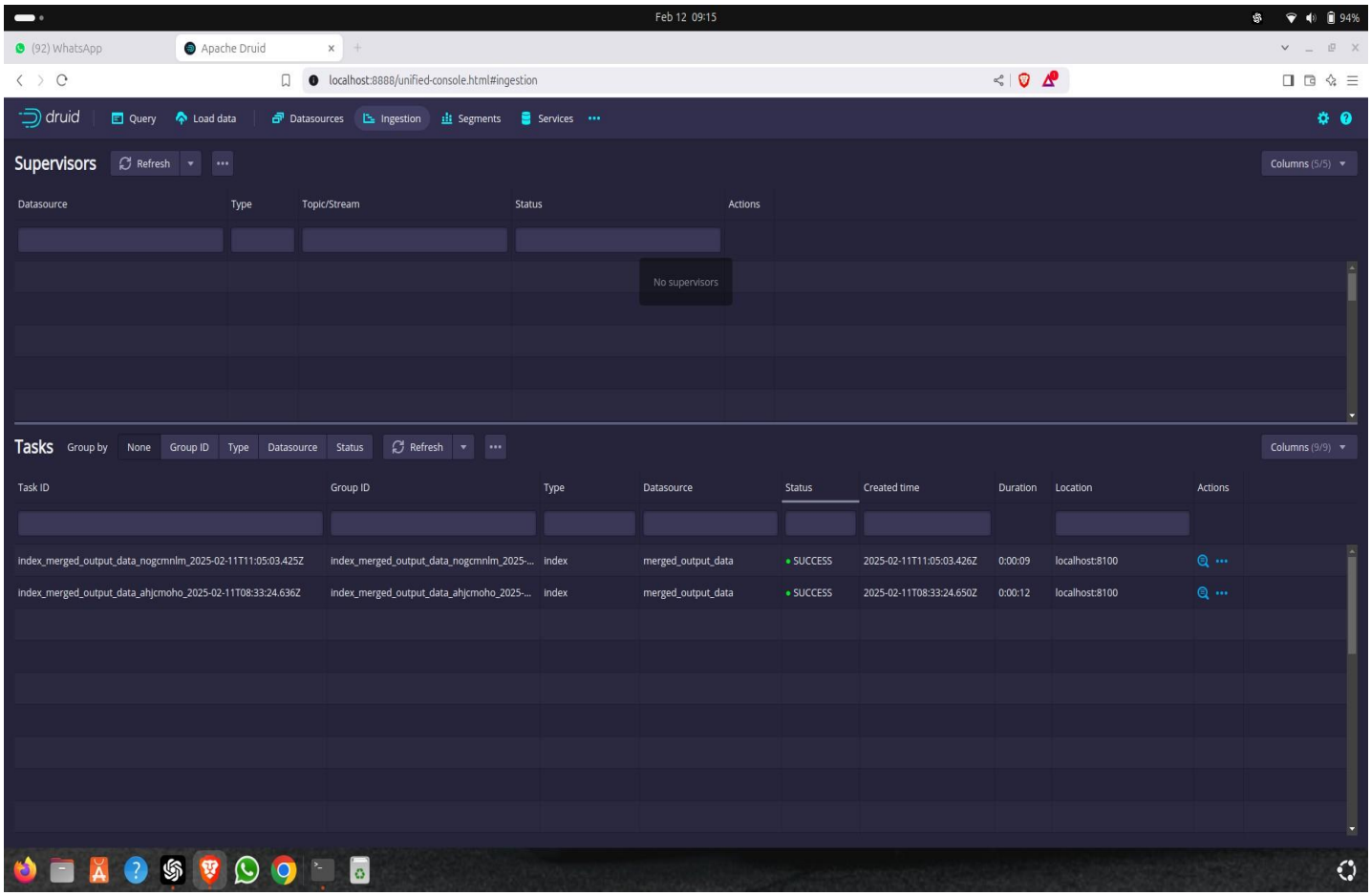
5.1.4 Sentimental forecast:




Machine Learning C...
Data Normalization...
Top 65+ Data Analy...
DBDA
MohsinBeg6692 (M...
All Bookmarks
Sentiment Forecast
Subreddit:
ChatGPT
Forecast Periods:
Time Range:
Monthly
Generate Forecast
Sentiment Forecast for machinelearning (Monthly Periods: 7)
Sentiment Score
2
0
-2
Mar 6
2022
Mar 13
Mar 20
Mar 27
Apr 3
Apr 10
Apr 17
Max Trend (Negative)
Positive Sentiment
Negative Sentiment
Neutral Sentiment
Forecasted Sentiment
Max Trend
Max Trend (Negative)
Date
Powered by FastAPI & Plotly


Machine Learning C...
Data Normalization...
Top 65+ Data Analy...
DBDA
MohsinBeg6692 (M...
All Bookmarks
Sentiment Forecast
Subreddit:
ChatGPT
Forecast Periods:
Time Range:
Monthly
Generate Forecast
Sentiment Forecast for chatgpt (Monthly Periods: 7)
Sentiment Score
Max Trend (Positive)
Positive Sentiment
Negative Sentiment
Neutral Sentiment
Forecasted Sentiment
Max Trend
Jan 2
2022
Jan 16
Jan 30
Feb 13
Feb 27
Mar 13
Mar 27
Apr 10
Date
Powered by FastAPI & Plotly

6.Database (Apache Druid):



7.Project Management Methodology:

Scrum Agile Methodology Used.

8.Future Scope:

The field of social media sentiment analysis is continuously evolving, with several potential advancements and applications in the future:

Enhanced Real-time Analysis

- Implementing real-time streaming sentiment analysis to monitor trends instantly.
- Integration with big data processing frameworks like Apache Spark for handling large-scale data efficiently.

Multilingual Sentiment Analysis

- Expanding sentiment analysis to multiple languages using multilingual NLP models such as BERT, mT5, and GPT-based models.
- Addressing challenges in slang, regional dialects, and cultural variations in sentiment expression.

Emotion Detection & Advanced Analytics

- Moving beyond positive, negative, and neutral classifications to detect specific emotions like anger, joy, sadness, and sarcasm.
- Utilizing deep learning techniques for better contextual understanding.

Integration with Business Intelligence

- Providing real-time brand monitoring and customer sentiment insights for businesses.
- Enhancing marketing strategies through AI-driven predictive sentiment analysis.

Ethical and Privacy Considerations

- Implementing privacy-preserving sentiment analysis using federated learning.
- Ensuring compliance with data privacy regulations (GDPR, CCPA) while analszing public social media

Conclusion:

Social media sentiment analysis is a powerful tool that helps organizations, businesses, and policymakers understand public opinion, track trends, and make data-driven decisions. This project successfully demonstrates the ability to extract, preprocess, analyze, and visualize sentiment trends from social media platforms using machine learning and NLP techniques.

By leveraging advanced AI models, real-time data processing, and multilingual support, sentiment analysis can continue to evolve, providing more accurate and insightful results. The future holds promising advancements in deep learning, emotion detection, and ethical AI, making sentiment analysis an indispensable tool for industries such as marketing, politics, finance, and customer service.

This project serves as a foundation for further research and development, aiming to create more intelligent, real-time, and human-like sentiment analysis systems in the coming years.