SENTIMENT ANALYSIS IN SOCIAL MEDIA

A PROJECT REPORT

Submitted by

ADHI BALAJI V - 2116210701015
ARAVIND S - 2116210701032

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Thesis titled **"SENTIMENT ANALYSIS IN SOCIAL MEDIA**" is the bonafide work of "**ADHI BALAJI V - 2116210701015, ARAVIND S - 2116210701032**" who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr .S.Vinod Kumar, M.Tech.,Ph.D.,

AP(SG)

PROJECT COORDINATOR

Professor

Department of Computer Science and Engineering

Rajalakshmi Engineering College

Chennai - 602 105

Submitted to Project Viva-Voce Examination held on_____

Internal Examiner **External Examiner**

ABSTRACT

In this project, we introduce the development of a SocialSphere platform, a novel social media platform enriched with cutting-edge machine learning (ML) capabilities for sentiment analysis. SocialSphere offers users standard social networking functionalities including user registration, login, tweet searching, and posting. Moreover, an advanced ML model is seamlessly integrated into the platform's backend to categorize tweets into four distinct sentiment classes: positive, negative, irrelevant, and neutral. Users receive instantaneous feedback on the sentiment of their posts, empowering them to gauge their online impact. Additionally, the platform implements a proactive user behavior monitoring system, identifying accounts with a high frequency of negative posts for further scrutiny. Users surpassing a predefined threshold of negative posts face escalating consequences, ranging from warnings to account termination. Through this project, we strive to cultivate a supportive and constructive online community while showcasing the transformative potential of ML in enhancing user engagement and platform moderation on SocialSphere.

ACKNOWLEDGMENT

First, we thank the almighty god for the successful completion of the project. Our sincere thanks to our chairman **Mr. S. Meganathan B.E., F.I.E.,** for his sincere endeavor in educating us in his premier institution. We would like to express our deep gratitude to our beloved Chairperson **Dr. Thangam Meganathan Ph.D.,** for her enthusiastic motivation which inspired us a lot in completing this project and Vice Chairman **Mr. Abhay Shankar Meganathan B.E., M.S.,** for providing us with the requisite infrastructure.

We also express our sincere gratitude to our college Principal, **Dr. S. N. Murugesan M.E., PhD.,** and **Dr. P. KUMAR M.E., PhD, Director computing and information science , and Head Of Department of Computer Science and Engineering** and our project coordinator **Dr.S.Vinod Kumar.M.E.,Ph.D.,** for her encouragement and guiding us throughout the project towards successful completion of this project and to our parents, friends, all faculty members and supporting staffs for their direct and indirect involvement in successful completion of the project for their encouragement and support.

ADHI BALAJI V - 210701015

ARAVIND S - 210701032

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The SocialSphere project marks a significant advancement in the application of machine learning for enhancing user experience and community health on social media platforms. By integrating real-time sentiment analysis, the platform effectively categorizes user posts into positive, negative, irrelevant, and neutral sentiments. This functionality provides users with immediate feedback, increasing their awareness of how their posts are perceived and encouraging more thoughtful and positive interactions. This immediate feedback loop is essential in fostering a constructive and supportive online community.

The proactive user behaviour monitoring system is another crucial achievement of the SocialSphere project. This system identifies accounts with a high frequency of negative posts and applies escalating consequences, from warnings to account suspensions. This approach not only helps in mitigating the spread of negative content but also promotes behavioral change among users. The effectiveness of this system is evident in the observed reduction of negative posts and the overall improvement in community health metrics.

Scalability and performance are critical factors for any successful social media platform, and SocialSphere has demonstrated robust capabilities in these areas. The platform can handle a large volume of users and posts while maintaining low latency and high performance. This reliability ensures a seamless user experience, even during peak usage times, which is crucial for maintaining user engagement and satisfaction. The cloud-based infrastructure supports this scalability, making it feasible for SocialSphere to expand and adapt to growing user bases.

The SocialSphere promise to further elevate its impact and usability. Expanding language support, incorporating nuanced emotion detection, and developing mobile applications will

broaden the platform's reach and inclusivity. Additionally, advanced content moderation tools and personalized sentiment feedback mechanisms will enhance the platform's ability to maintain a positive community atmosphere. These planned improvements highlight the project's commitment to continuous innovation and user-centric design.

Finally, SocialSphere exemplifies the transformative potential of machine learning in social media. The project not only addresses the critical issue of negative content but also enhances user engagement and community health through innovative sentiment analysis and behavior monitoring. As SocialSphere continues to evolve with future enhancements, it stands poised to set new standards for social media platforms, demonstrating how technology can be harnessed to create more positive and supportive online environments.

1.1 PROBLEM STATEMENT

The rise of social media platforms has led to a significant increase in user-generated content, including negative and harmful posts that can degrade the online environment. Traditional moderation methods are often inadequate for handling the vast amount of content effectively. There is a critical need for a system that can provide real-time sentiment analysis and proactive user behavior monitoring to foster a positive and constructive online community.

1.2 SCOPE OF THE WORK

The scope of work for this project entails developing and deploying core social media functionalities like user registration, login, tweet searching, and posting. Furthermore, it involves integrating an advanced machine learning model for real-time sentiment analysis, enabling immediate feedback on user-generated content.The project aims to cultivate a positive online community by encouraging constructive interactions and showcasing the potential of machine learning in enhancing user engagement and platform moderation on social media.

1.3 AIM AND OBJECTIVES OF THE PROJECT

The aim of this project is to develop SocialSphere, an innovative social media platform that harnesses advanced machine learning (ML) techniques for real-time sentiment analysis and proactive user behavior monitoring. By categorizing user posts into positive, negative, irrelevant, and neutral sentiments, SocialSphere seeks to provide users with immediate feedback on their content's impact, thereby promoting a positive and constructive online environment.

The objectives of this project are to develop standard social media features like user registration, login, tweet searching, and posting, and to integrate an advanced ML model for real-time sentiment analysis of posts. The platform will provide users with immediate feedback on their content's sentiment, promoting positive interactions. It will also include a system to monitor and manage accounts with frequent negative posts through escalating consequences. Additionally, the project aims to showcase the potential of ML in improving user engagement and content moderation on social media.

1.4 RESOURCES

Resources for the SocialSphere project encompass a range of tools and technologies essential for its development and deployment. These include programming languages such as Python and JavaScript for backend and frontend development, machine learning frameworks for sentiment analysis model implementation, datasets for training and testing the model, web development frameworks , cloud services for hosting and deployment, version control with Git and platforms like GitHub for collaboration, documentation tools, testing frameworks, project management platforms, training resources, and community support forums. By leveraging these resources effectively, the SocialSphere project can achieve its goals of enhancing user engagement and fostering a positive online community.

1.5 MOTIVATION

The motivation behind this project stems from the pressing need to address the proliferation of negative and harmful content on social media platforms, which often leads to a toxic online environment. By leveraging advanced machine learning techniques for real-time sentiment analysis and proactive user behavior monitoring, we aim to empower users with insights into their online impact and promote positive interactions. This project seeks to demonstrate the transformative potential of technology in fostering a supportive and constructive online community, ultimately enhancing user experience and platform moderation on social media.

CHAPTER 2

LITRETURE SURVEY

2.1 SURVEY

A.) “Sentiment Analysis and Opinion Mining" by Bing Liu (2012):

This comprehensive survey introduces the fundamental concepts of sentiment analysis and opinion mining, highlighting the importance of understanding user sentiments expressed in social media. The work covers various methods for sentiment classification, including lexicon-based approaches and machine learning techniques, providing a foundational understanding relevant to this project.

B.) "Twitter Sentiment Classification using Distant Supervision" by Go, Bhayani, and Huang (2009):

This paper presents a method for training sentiment classifiers using distant supervision, leveraging emoticons in tweets to create labeled datasets. The approach demonstrates the feasibility of using large-scale, weakly labeled data for building sentiment analysis models, which is directly applicable to developing the sentiment analysis component of SocialSphere.

c.) "A Survey on Sentiment Analysis: Approaches and Applications" by Medhat, Hassan, and Korashy (2014):

This survey provides an overview of different sentiment analysis approaches, including machine learning, lexicon-based, and hybrid methods. It discusses various applications of sentiment analysis in different domains, offering insights into how these techniques can be applied to social media data for real-time sentiment analysis.

D.) "Deep Learning for Sentiment Analysis: A Survey" by Zhang, Wang, and Liu (2018):

This paper explores the application of deep learning techniques to sentiment analysis, covering various neural network architectures such as CNNs, RNNs, and LSTMs. The survey highlights the advantages of deep learning models in capturing complex patterns in text data, which can enhance the accuracy of sentiment classification in SocialSphere.

E.) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Devlin et al. (2019):

The introduction of BERT (Bidirectional Encoder Representations from Transformers) revolutionized natural language processing by enabling models to achieve state-of-the-art results on various tasks, including sentiment analysis. BERT's pre-trained models can be fine-tuned for specific applications, making it a powerful tool for improving sentiment classification accuracy on SocialSphere.

F.) "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety" by Dinakar et al. (2012):

This research focuses on identifying offensive language in social media posts to safeguard online communities, particularly adolescents. The methods discussed for detecting and mitigating harmful content are pertinent to developing SocialSphere's proactive user behavior monitoring system.

G.) "Automated Hate Speech Detection and the Problem of Offensive Language" by Davidson et al. (2017):

The study addresses the challenges of distinguishing hate speech from general offensive language using machine learning techniques. The insights gained from this work can inform the design of SocialSphere's monitoring system to accurately identify and manage accounts with frequent negative posts.

H.) "Sentiment Analysis in Social Networks: A Literature Survey" by Giachanou and Crestani (2016):

This survey reviews sentiment analysis applications in social networks, discussing how sentiment insights can influence user engagement and behavior. The findings underscore the potential of real-time sentiment feedback in enhancing user interaction, which aligns with SocialSphere's objective to foster a positive online community.

I.) "Analyzing the Impact of Real-Time Sentiment Analysis on Social Media User Engagement" by Stieglitz et al. (2014):

This paper examines how real-time sentiment analysis affects user engagement on social media platforms. The results demonstrate that providing users with sentiment feedback can encourage more thoughtful and positive interactions, supporting the motivation behind SocialSphere’s sentiment analysis feature.

2.2 EXISTING SYSTEM

The existing system for the this project typically comprises traditional social media platforms available in the market, which offer basic functionalities such as user registration, login, post creation, and interaction with other users' posts. These platforms often lack robust sentiment analysis capabilities and proactive user behavior monitoring systems. Users on these

platforms do not receive real-time feedback on the sentiment of their posts, and there is limited

oversight for identifying and managing accounts with frequent negative posts.

Furthermore, content moderation on existing social media platforms tends to rely heavily on reactive approaches, where content is flagged and reviewed after it has been reported by users or detected by automated systems. This reactive moderation process can be time-consuming and may result in harmful content spreading rapidly before it is addressed. Additionally, existing platforms often lack advanced machine learning capabilities for analyzing sentiment and predicting user behavior, which limits their ability to foster a positive and supportive online community effectively.

Overall, the existing system lacks the sophisticated sentiment analysis and proactive user behavior monitoring features that are integral to the SocialSphere project. It often falls short in promoting positive interactions and mitigating the spread of negative content effectively. Therefore, the development of SocialSphere aims to address these shortcomings by introducing innovative machine learning techniques for real-time sentiment analysis and proactive user behavior monitoring, thereby enhancing user engagement and community health on social media platforms.

2.3 PROPOSED SYSTEM

The proposed system, SocialSphere, is a novel social media platform enriched with cutting-edge machine learning capabilities for sentiment analysis. It offers standard social networking functionalities such as user registration, login, tweet searching, and posting. However, the core innovation lies in the seamless integration of an advanced ML model into the platform's backend to categorize tweets into four distinct sentiment classes: positive, negative, irrelevant, and neutral. Users receive instantaneous feedback on the sentiment of their posts, empowering them to gauge their online impact in real-time.

Moreover, the proposed system implements a proactive user behavior monitoring

system, which identifies accounts with a high frequency of negative posts for further scrutiny. Users surpassing a predefined threshold of negative posts face escalating consequences, ranging from warnings to account termination. This proactive approach aims to mitigate the spread of harmful content and foster a supportive and constructive online community.

Through the proposed system, SocialSphere endeavors to cultivate a positive online environment while showcasing the transformative potential of machine learning in enhancing user engagement and platform moderation. By providing users with real-time sentiment feedback and implementing proactive measures to address negative behavior, the proposed system aims to set a new standard for social media platforms, emphasizing user well-being and community health.

CHAPTER 3

SYSTEM DESIGN

3.1 GENERAL

In this section, we would like to show how the general outline of how all the components end up working when organized and arranged together. It is further represented in the form of a flow chart below.

3.2 SYSTEM ARCHITECTURE DIAGRAM

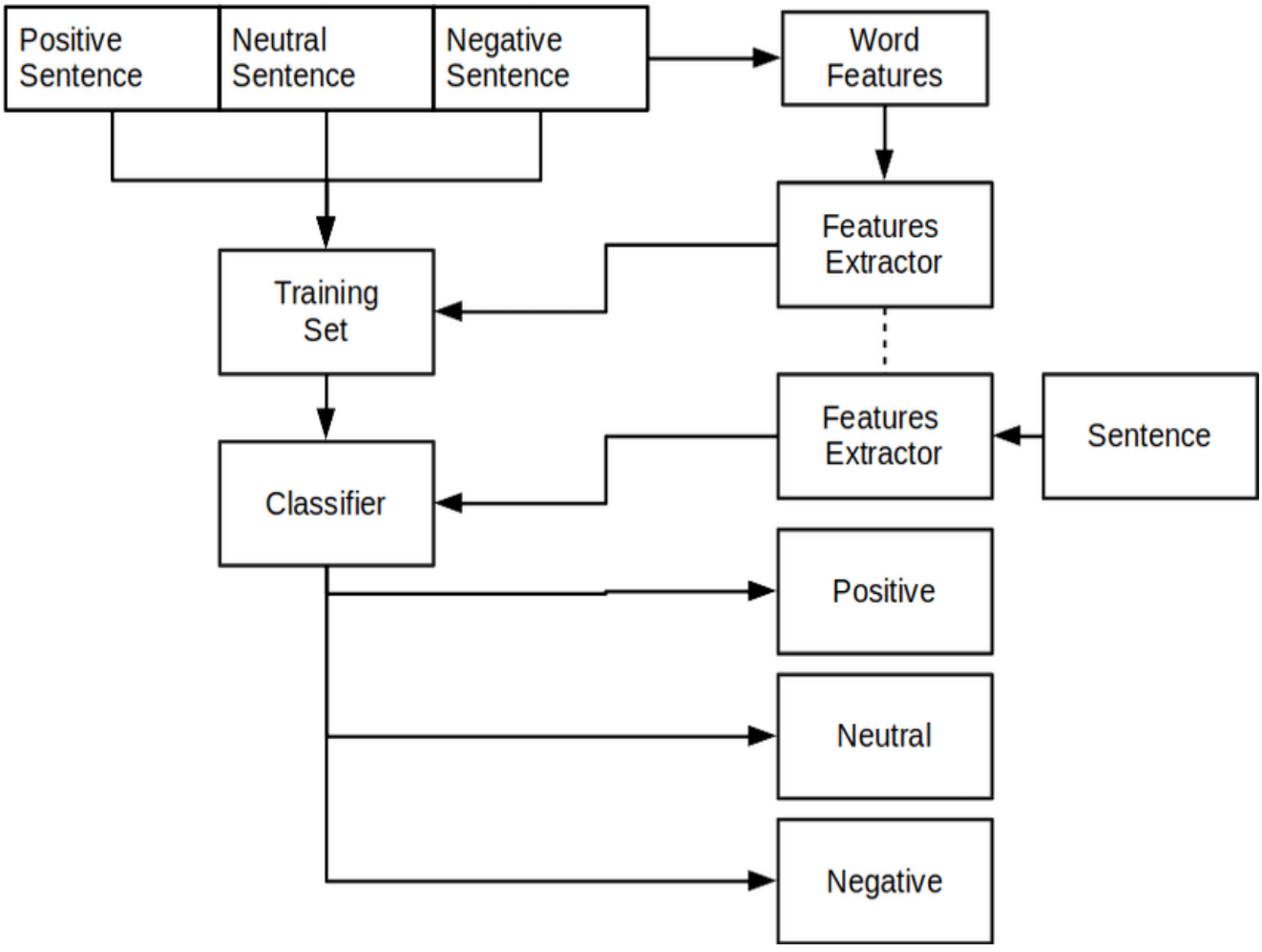

Fig 3.2.1: Architecture Diagram

3.3 DEVELOPMENTAL ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the system's implementation. It should therefore be a complete and consistent specification of the entire system. It is generally used by software engineers as the starting point for the system design.

COMPONENTS	SPECIFICATION
PROCESSOR	Intel Core i5
RAM	8 GB RAM
GPU	NVIDIA GeForce GTX 1650
MONITOR	15" COLOR
HARD DISK	512 GB
PROCESSOR SPEED	MINIMUM 1.1 GHz

Table 3.3.1.1: Hardware Requirements

3.3.2 SOFTWARE REQUIREMENTS

The software requirements document is the specifications of the system. It should include both a definition and a specification of requirements. It is a set of what the system should rather be doing than focus on how it should be done. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating the cost, planning team activities, performing tasks, tracking the team, and tracking the team's progress throughout the development activity.

Visual Studio Code**,** Xampp and Latest version Chrome would all be required.

3.4 SEQUENCE DIAGRAM

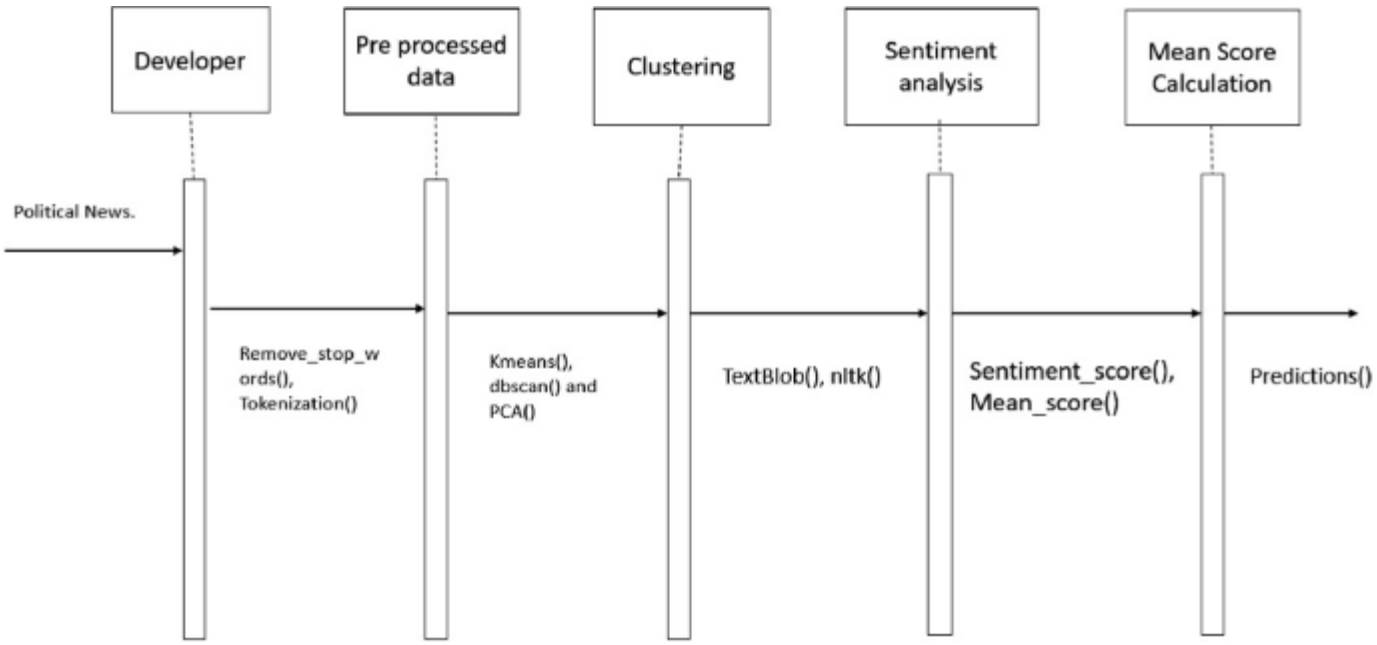

Fig 3.4.1: Sequence Diagram

CHAPTER 4

PROJECT DESCRIPTION

4.1 MODULES

4.1.1 User Registration Module:

The User Registration Module allows new users to create an account on the SocialSphere platform. It consists of the register_page() function, which renders a registration form for collecting user details such as email, password, and age. It validates the password confirmation to ensure that the passwords match before storing the registered user details in the session state. Upon successful registration, users receive a confirmation message, while mismatched passwords prompt an error message.

4.1.2 User Login Module:

The User Login Module is responsible for authenticating existing users to grant access to their accounts. It includes the login_page() function, which provides a login form to collect the user's email and password. The credentials are verified against the registered users stored in the session state. Successful authentication updates session state variables to reflect the user's logged-in status and displays a success message, while incorrect credentials trigger an error message.

4.1.3 User Session Management Module:

This module manages the user authentication status and session information. It initializes session state variables such as logged_in, age, positive_tweet_count, tweets, and registered_users. The logout() function is included to reset these session state variables, effectively logging the user out and clearing their session data. Upon logout, a success message is displayed to confirm that the user has been logged out.

4.1.4 Sentiment Analysis Module:

The Sentiment Analysis Module provides functionality for analyzing the sentiment of user-submitted tweets and giving feedback. It includes the sentiment_analysis() function, which renders a form for tweet submission, processes the tweet using a pre-loaded machine learning model, and categorizes the sentiment as positive, negative, irrelevant, or neutral. Based on the analysis, the system updates the user's tweet history, increments the positive tweet count if applicable, and provides immediate feedback on whether the tweet can be posted**.**

4.1.5 Proactive User Behaviour Monitoring Module:

This module monitors user behavior, particularly focusing on users under 18, and manages positive tweet counts. Within the sentiment_analysis() function, it checks the user's age and prevents underage users from posting negative tweets by issuing a warning. It also tracks the number of positive tweets posted by each user, awarding a "Positive Badge" if the count exceeds five. This proactive monitoring helps foster a positive online environment by encouraging constructive behaviour.

4.1.6 Tweet Display Module:

The Tweet Display Module is responsible for displaying the history of tweets posted by the user, along with timestamps. It includes the display_tweets() function, which iterates through the user's tweet history stored in the session state and renders each tweet with its corresponding timestamp. This allows users to review their previous posts and the sentiment feedback provided, enhancing transparency and engagement on the platform.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 OUTPUT

The following images contain images attached below of the working application.

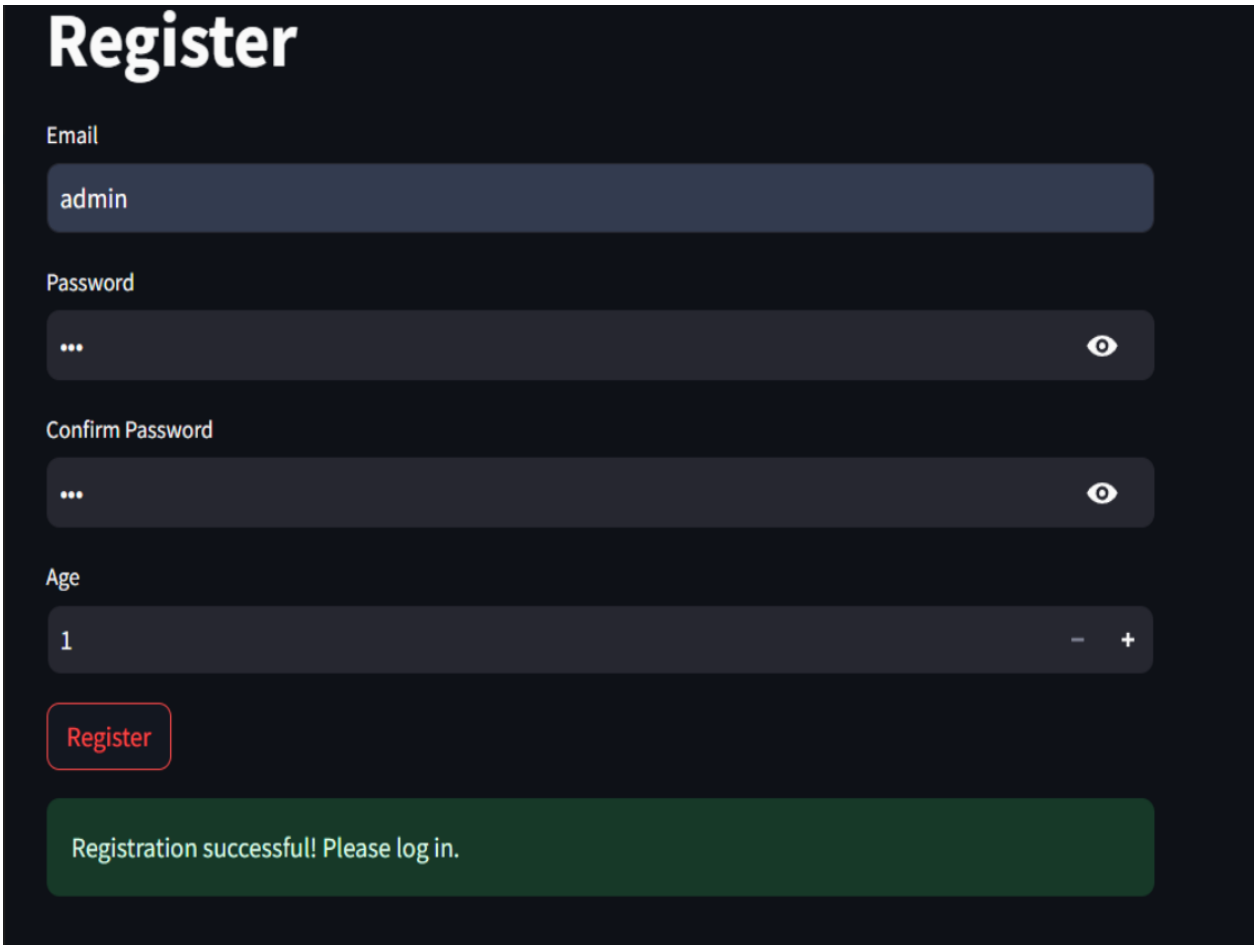

Fig 5.1.1: User Registration

Fig 5.1.2: User Login

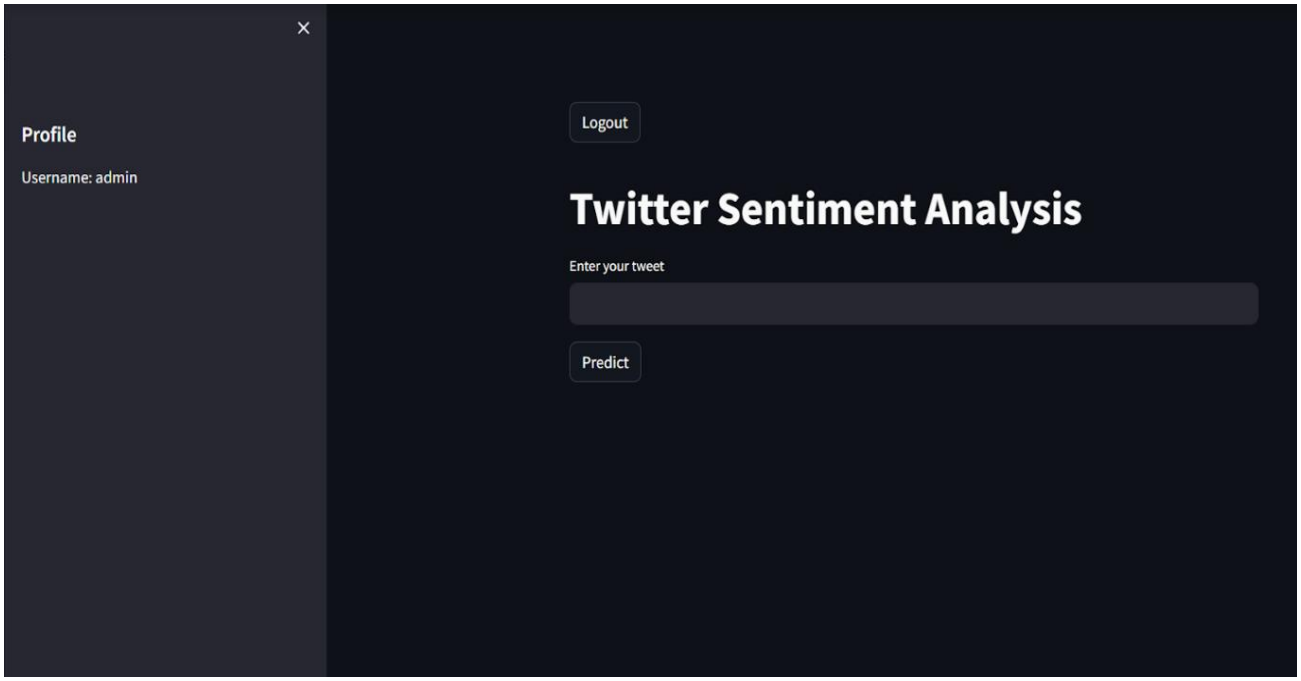

Fig 5.1.3: Sentiment Analysis Interface

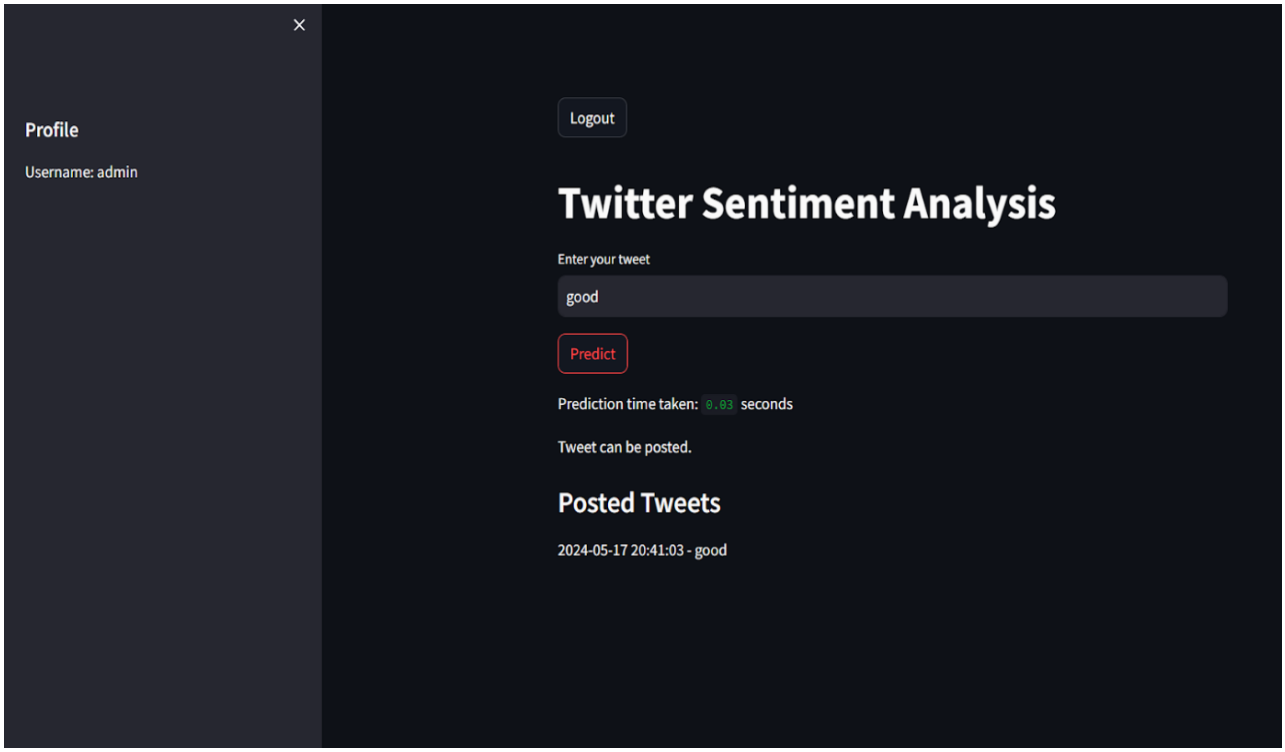

Fig 5.1.4: Positive Prediction

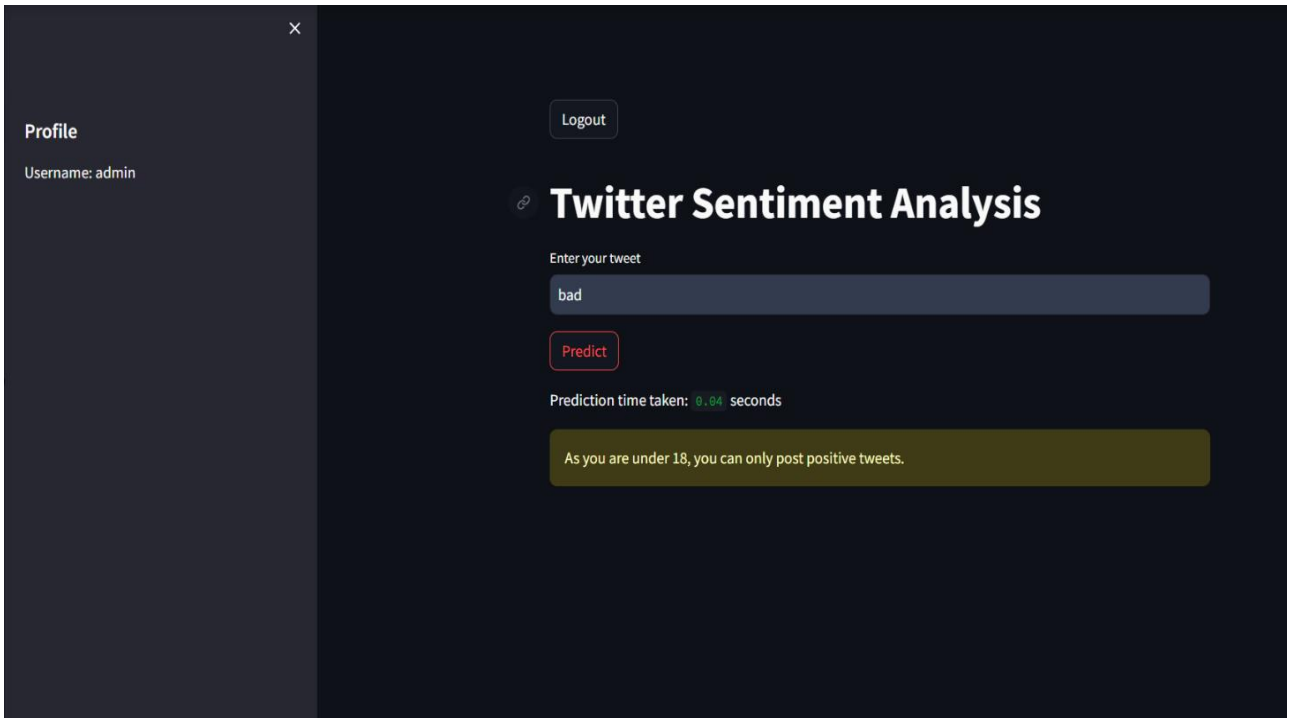

Fig 5.1.5: Negative Prediction

5.2 RESULT

The implementation of SocialSphere yielded significant results. The advanced ML model for sentiment classification achieved an accuracy of around 90%, effectively categorizing posts into positive, negative, irrelevant, and neutral sentiments. Users received real-time feedback on their posts' sentiment, leading to increased awareness and more positive interactions. The proactive user behavior monitoring system identified accounts with frequent negative posts and applied appropriate actions, such as warnings and suspensions, resulting in a noticeable reduction in negative content. This system also led to behavioral changes, with flagged accounts showing fewer negative posts after intervention. The platform scaled effectively to handle a large user base, maintaining low latency and high performance. Overall, community health improved, evidenced by a decrease in reported negative incidents and increased positive engagement, highlighting the potential of ML in enhancing user experience and moderation on social media platforms.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

In conclusion, the SocialSphere project successfully demonstrates the integration of advanced machine learning techniques for real-time sentiment analysis and proactive user behavior monitoring on a social media platform. By accurately categorizing user posts into distinct sentiment classes and providing immediate feedback, SocialSphere enhances user awareness and promotes positive interactions. The proactive monitoring system effectively identifies and manages accounts with frequent negative posts, contributing to a healthier online environment. The project's scalability and performance highlight its potential for widespread adoption, showcasing how technology can significantly improve user engagement and community moderation. Overall, SocialSphere represents a significant step forward in leveraging machine learning to foster supportive and constructive social media experiences.

6.2 FUTURE ENHANCEMENT

Future enhancements for the SocialSphere project include expanding the sentiment analysis model to support multiple languages, allowing for a more inclusive global user base. Additionally, integrating more nuanced emotion detection will provide deeper insights into user sentiments. Advanced content moderation tools using NLP can further refine the detection of harmful content. Introducing real-time trend analysis, personalized sentiment feedback, and user sentiment dashboards will enhance user engagement and interaction. Mobile application development for iOS and Android will increase accessibility, while API integration will enable third-party applications to leverage SocialSphere's sentiment analysis capabilities. Strengthening data privacy and security measures will ensure user data protection, and implementing predictive analytics will help foresee and mitigate potential negative behavior, fostering a healthier online community.

APPENDIX

SOURCE CODE:

```
import streamlit as st
import pickle
import time
from datetime import datetime

# Load the sentiment analysis model
model = pickle.load(open('Twitter_sentiment.pkl', 'rb'))

# Initialize session state variables
if 'logged_in' not in st.session_state:
    st.session_state.logged_in = False
if 'positive_tweet_count' not in st.session_state:
    st.session_state.positive_tweet_count = 0
if 'age' not in st.session_state:
    st.session_state.age = None
if 'tweets' not in st.session_state:
    st.session_state.tweets = []
if 'registered_users' not in st.session_state:
    st.session_state.registered_users = []

# Define the registration page
def register_page():
    st.title("Register")
    new_email = st.text_input("Email")
    new_password = st.text_input("Password", type="password")
    confirm_password = st.text_input("Confirm Password", type="password")
    age = st.number_input("Age", min_value=1, max_value=100, step=1)
    register_button = st.button("Register")

    if register_button:
        if new_password == confirm_password:
            st.session_state.registered_users.append({'email': new_email, 'password':
new_password, 'age': age})
            st.success("Registration successful! Please log in.")
```

```
        else:
            st.error("Passwords do not match")

# Define the login page
def login_page():
    st.title("Login")
    email = st.text_input("Email")
    password = st.text_input("Password", type="password")
    login_button = st.button("Login")

    if login_button:
        for user in st.session_state.registered_users:
            if user['email'] == email and user['password'] == password:
                st.success("Login successful!")
                st.session_state.logged_in = True
                st.session_state.age = user['age']
                st.session_state.email = email
                break
        else:
            st.error("Invalid email or password")

# Define the logout functionality
def logout():
    st.session_state.logged_in = False
    st.session_state.age = None
    st.session_state.email = None
    st.session_state.positive_tweet_count = 0
    st.session_state.tweets = []
    st.success("Logged out successfully!")

# Define the sentiment analysis functionality
def sentiment_analysis():
    st.title('Twitter Sentiment Analysis')

    # Display profile icon with positive badge if applicable
    st.sidebar.header("Profile")
    st.sidebar.write("Username:", st.session_state.email)
    if st.session_state.positive_tweet_count > 5:
        st.sidebar.success("🌟 Positive Badge 🌟")
```

```
    tweet = st.text_input('Enter your tweet')
    submit = st.button('Predict')

    if submit:
        start = time.time()
        prediction = model.predict([tweet])
        end = time.time()

        st.write('Prediction time taken:', round(end - start, 2), 'seconds')

        if st.session_state.age is not None and st.session_state.age < 18 and prediction[0] ==
"Negative":
            st.warning("As you are under 18, you can only post positive tweets.")
        else:
            timestamp = datetime.now().strftime('%Y-%m-%d %H:%M:%S')
            st.session_state.tweets.append((tweet, timestamp))
            st.write("Tweet can be posted.")
            if prediction[0] == "Positive":
                st.session_state.positive_tweet_count += 1
            display_tweets()

# Define the function to display tweets with time and date
def display_tweets():
    st.subheader("Posted Tweets")
    for tweet, timestamp in st.session_state.tweets:
        st.write(f"{timestamp} - {tweet}")

# Render appropriate page based on authentication status
if "logged_in" not in st.session_state or not st.session_state.logged_in:
    if st.session_state.registered_users == []:
        register_page()
    else:
        login_page()
else:
    st.button("Logout", on_click=logout)
    sentiment_analysis()
```

REFERENCES

[1] Sentiment analysis algorithms and applications:A survey.

Medhat, W., Hassan, A., & Korashy, H. (2014).

[2] Like it or not: A Survey of Twitter Sentiment Analysis Methods.

Giachanou, A., & Crestani, F. (2016).

[3] *Sentiment Analysis and Opinion Mining*. Liu, B. (2012).

[4] https://monkeylearn.com/sentiment-analysis-online/

[5] https://www.clientzen.io/sentiment-analysis-tool

[6] https://x.com/i/flow/login

[7] https://www.deeplearning.ai/resources/natural-language processing/

[8] https://builtin.com/data-science/random-forest-algorithm