"Sentilytics: Comment Analyzer"

Developed For

FCAIT, *i*MSc(IT)

Project Report (Sem – VI)

Submitted For

The Partial Fulfillment Towards

The Degree of

Integrated Master of Science (Information Technology)

***i*MSc(IT)**

By

Akbar Ali Musamji - B46

Mohammed Akil Shiakh - B93

Shahid Husain Shaikh - B94

Under the Guidance of

External Guide

Maheraban Ali

Managing Director

Datahat Solution LLP

Internal Guide

Prof.Anjali Bobra

FCAIT,iMSc(IT),

Ahmedabad

Faculty of Computer Applications & Information Technology

iMSc(IT) Programme, Ahmedabad.

GLS UNIVERSITY

Faculty of Computer Applications & Information Technology,

iMSc(IT) Programme
Ahmedabad

CERTIFICATE

This is to certify that

1) Akbar Ali Musamji
2) Mohammed Akil Shaikh
3) Shahid Husain Shaikh

Students of Semester- VI Integrated Msc(IT) [TY iMSc(IT)],
FCAIT, GLS University have successfully completed the
Mini Project

on

"Sentilytics: Comment Analyzer"

as a partial fulfillment of the study of Third year Semester-VI,
Integrated Master of Science (Information Technology)
[iMSc(IT)]

Date of Submission: _____

Dean
Faculty of Computer Applications & IT (UG)
GLS University, Ahmedabad-380 006.

Dr. Ankit
Internal Guide

Dr. Tripti Dodiya
Dean
FCAIT-UG

DataHat
Solutions LLP

Datahat Solutions LLP

D301 Titanium city center Ahmedabad 380015 Gujarat, India +917567196771
LLP Identification : ABB-6972

Project Mentorship 19-03-2025

Project Title : Sentilytics - AI Powered Comments Analyzer

This is to certify that Akbarali Musamji, Mohammad Akil Shaikh, Shahid Shaikh, are students of Semester-VI, iMsclT programme, GLS University, Ahmedabad bring mentored by us.

Under the mentorship and guidance of **Datahat Solutions LLP,** they have effectively accomplished the following tasks:

- System Analysis
- UML Diagrams
- Data Dictionary
- Implementation
- Testing
- Deployment

Their dedication, sincerity, and hard work during the project have been exemplary.

Maherban Ali
Datahat Solutions LLP
(Managing Director)

DataHat Solutions LLP

1

1.Introduction:

Sentilytics is an AI-powered sentiment analysis tool designed for real-time analysis of user-generated comments. The system categorizes comments into positive, negative, or neutral sentiments. It supports both single comment analysis and bulk analysis via CSV/Excel uploads. Users can visualize sentiment distribution using bar graphs and word clouds. The project comprises a React.js frontend and a Django backend with a custom sentiment analysis model.

2.Project Profile:

2.1 Project Description:

Our system simplifies sentiment analysis by allowing users to analyze comments individually or in bulk. Users can either input a single comment or upload a CSV file containing multiple comments, and the system will classify them as **positive, negative, or neutral**.

For guest users, a single comment analysis is available without registration. Registered users, on the other hand, can perform batch analyses, visualize results through **bar graphs and word clouds**, and even correct misclassified sentiments to improve model accuracy.

Admins have access to user activity, and comment classification results. The Sentilytics platform ensures a **user-friendly experience**, helping businesses and individuals understand customer opinions in a structured and efficient manner.

2.2 Project Modules:

1. Single Comment Analysis:
    - Users can analyze individual comments.
2. Multiple Comments Analysis:
    - Supports CSV/Excel file uploads for bulk analysis.
3. YouTube Comments Analysis:
    - Fetches and analyzes YouTube comments using the Google API.
4. Export Results :
    - Users can download results in Excel format.
5. Graphical Representations:
    - Bar graphs for sentiment distribution.
    - Word clouds for frequent terms in positive reviews.
6. Manual Sentiment Editing :
    - Users can correct misclassified sentiments to improve model accuracy.
7. User Management :
    - Guest users can analyze single comments, while registered users have access to batch analysis and sentiment corrections.
8. Database Storage :
    - Stores all analyses with timestamps, comment sources, and sentiment results.

2.3 Technology Stack:

- Frontend: React.js
- Backend: Django (Python)
- Database: PostgreSQL
- Machine Learning Model: Custom-built using Logistic Regression
- Datasets: Sentiment140, Amazon Product Review

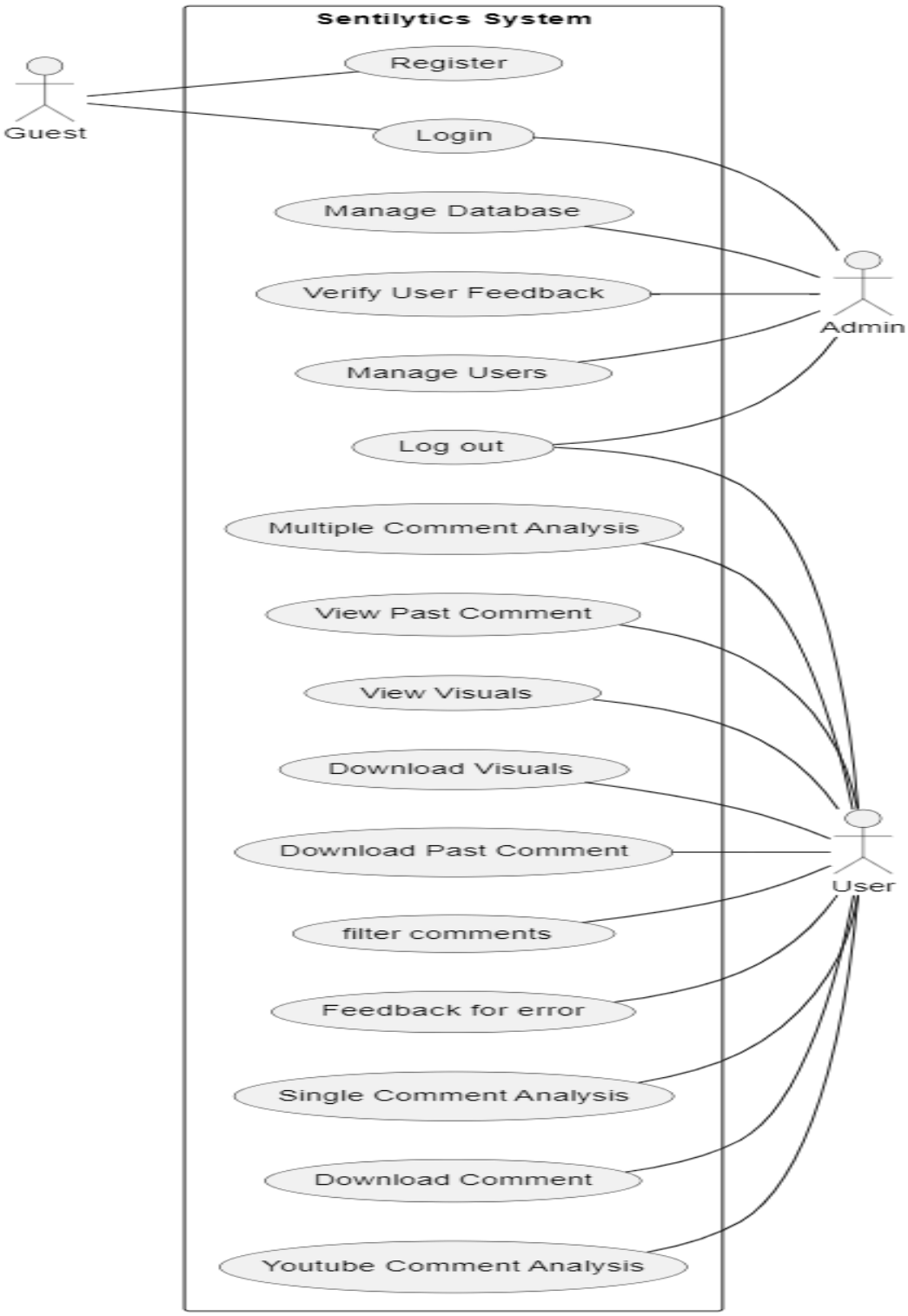
2.4 Implementation Details:

Sentilytics follows a structured workflow for sentiment analysis:

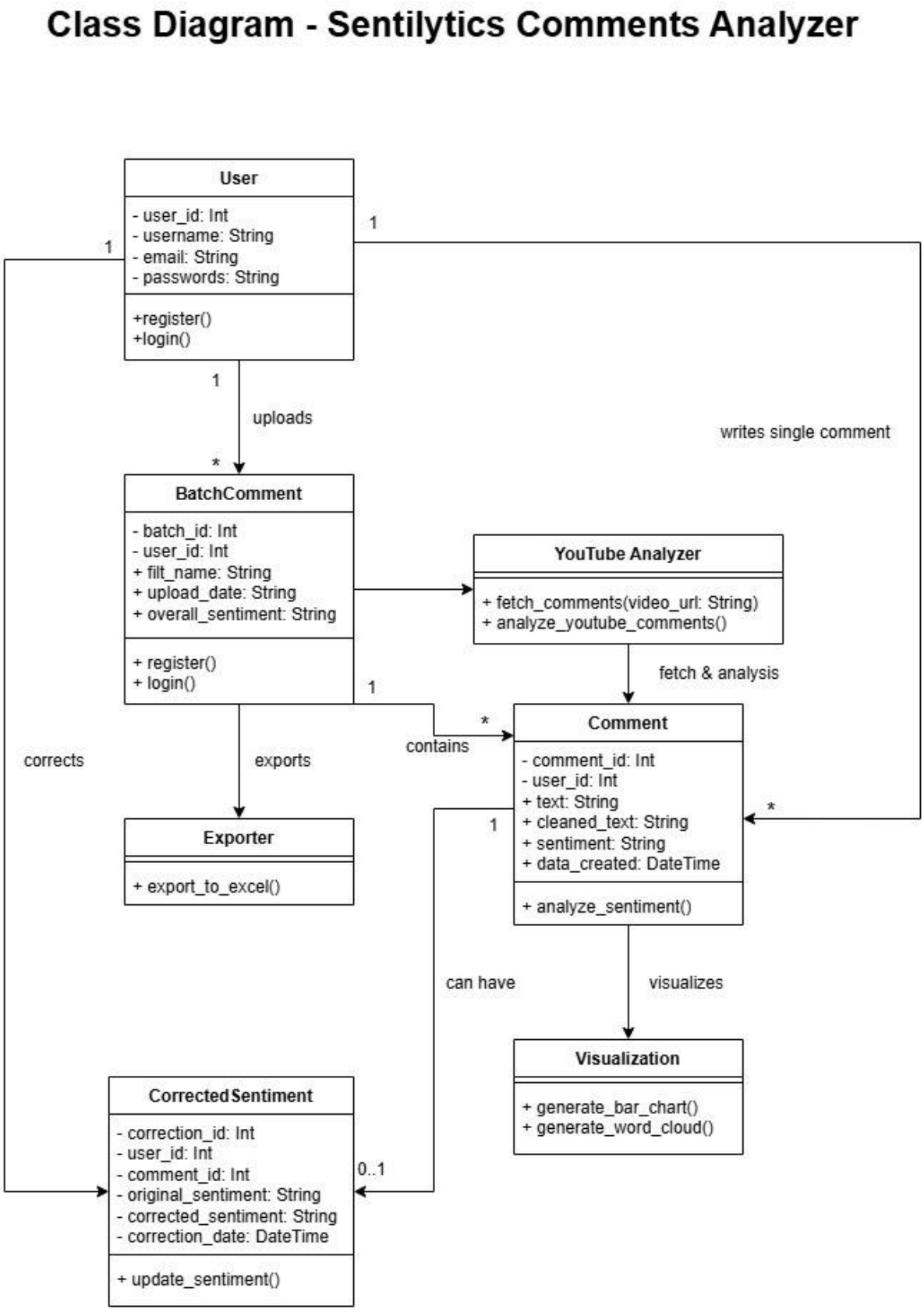
1. Preprocessing: Text is cleaned by removing stopwords, special characters, and applying tokenization.
2. Feature Extraction: TF-IDF vectorization converts text into numerical format for analysis.
3. Model Prediction: The Logistic Regression model predicts sentiment polarity.
4. Result Storage: Analysis results are stored in the database for retrieval and visualization.
5. User Feedback Loop: Users can correct sentiments, which are stored to refine future model training.

3.UML Diagrams:

3.1 Use-case Diagram:

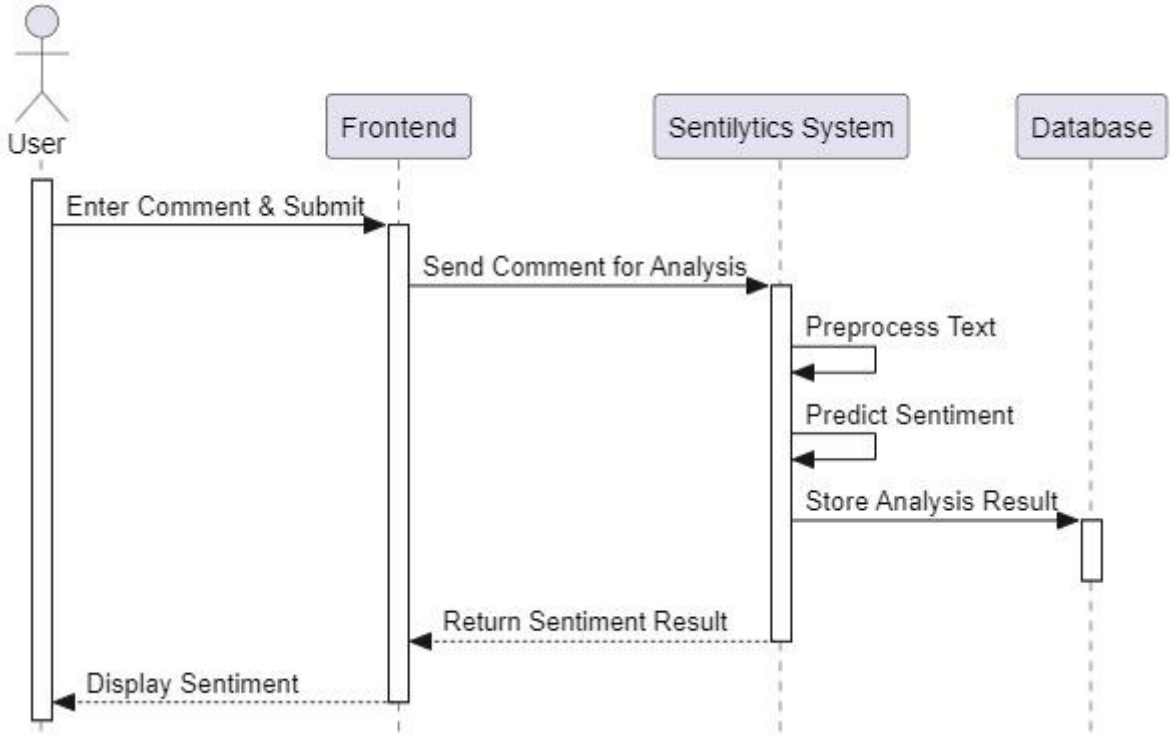


3.2 Class Diagram:

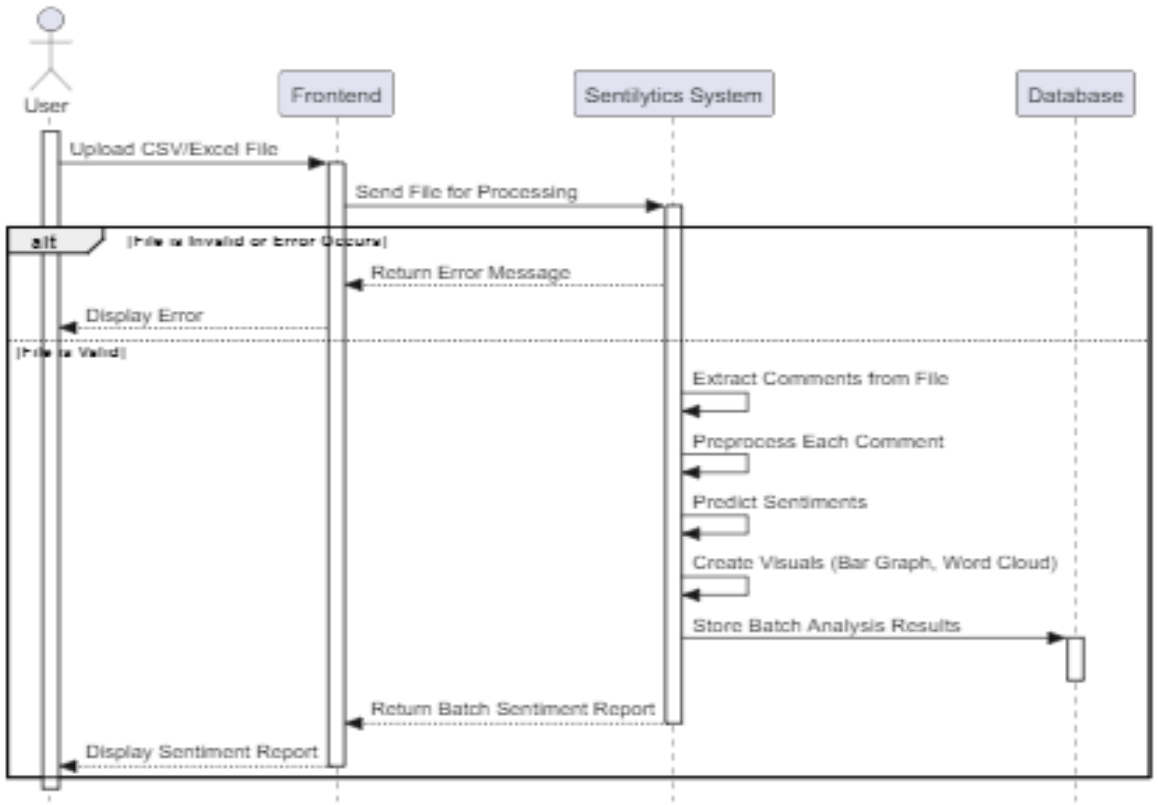


3.3 Sequence Diagram:

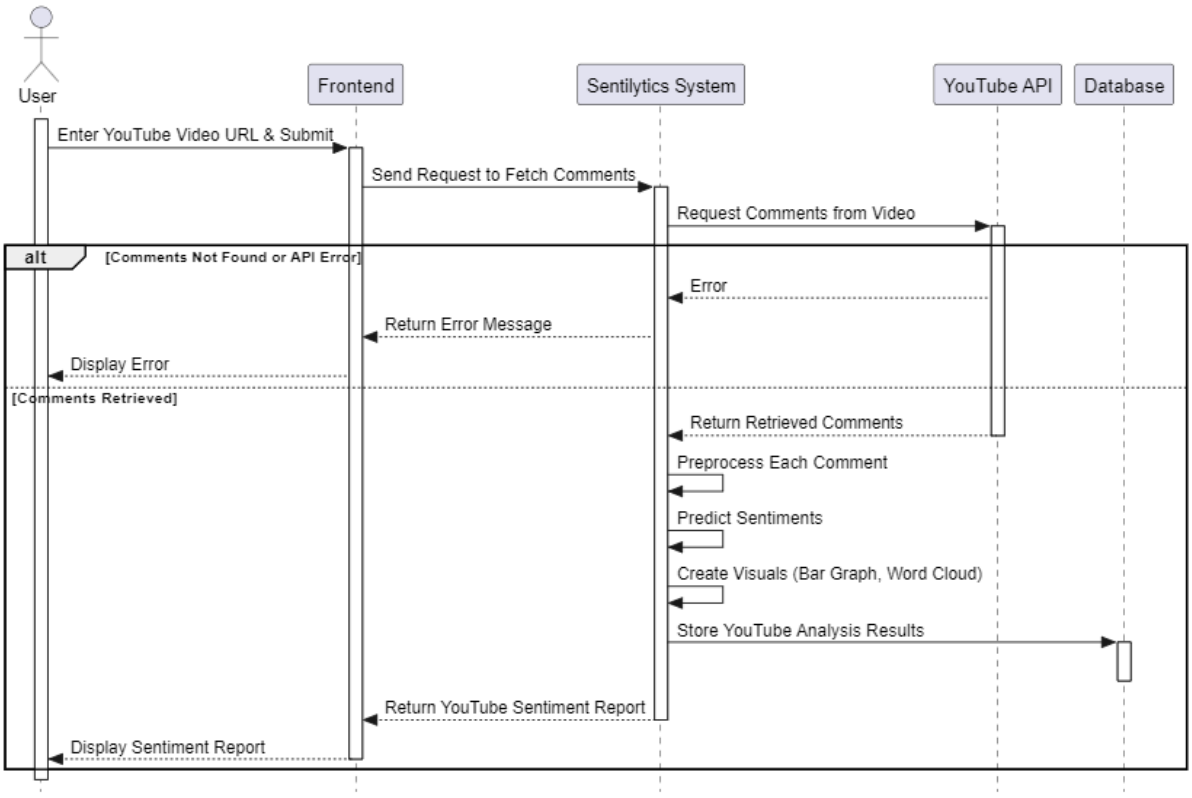
1. Single Comment Analysis Sequence Diagram



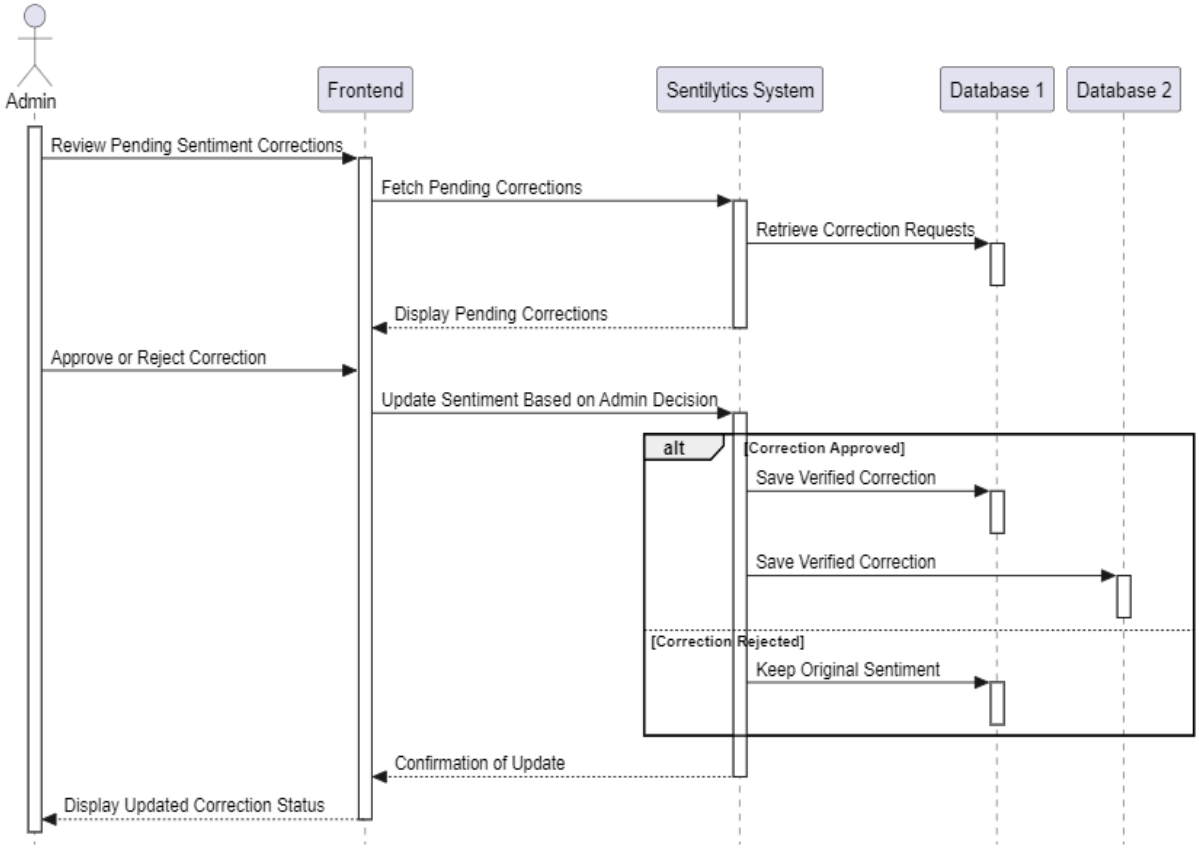
2. Multiple Comment Analysis Sequence Diagram



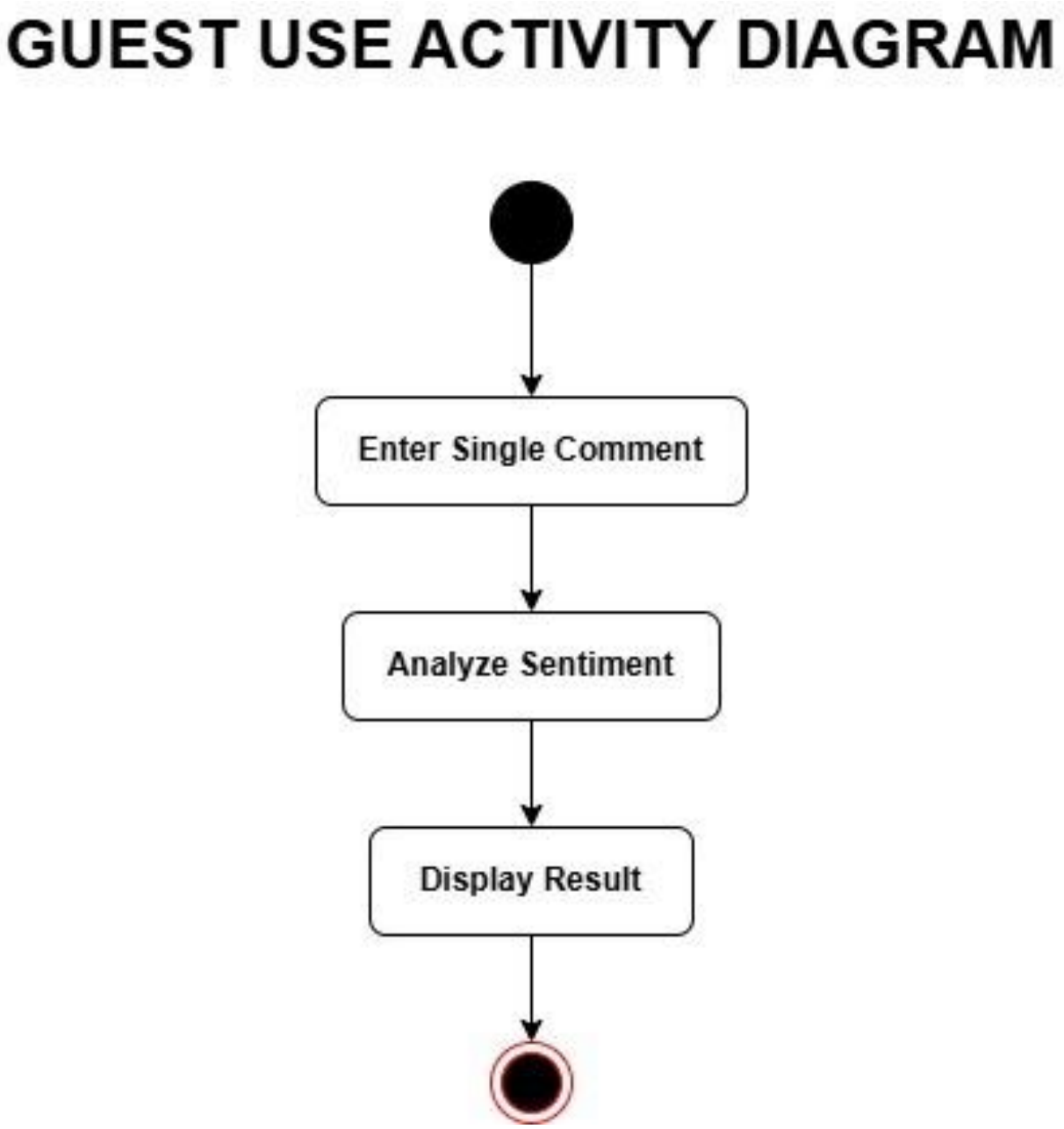
3. YouTube Comment Analysis Sequence Diagram



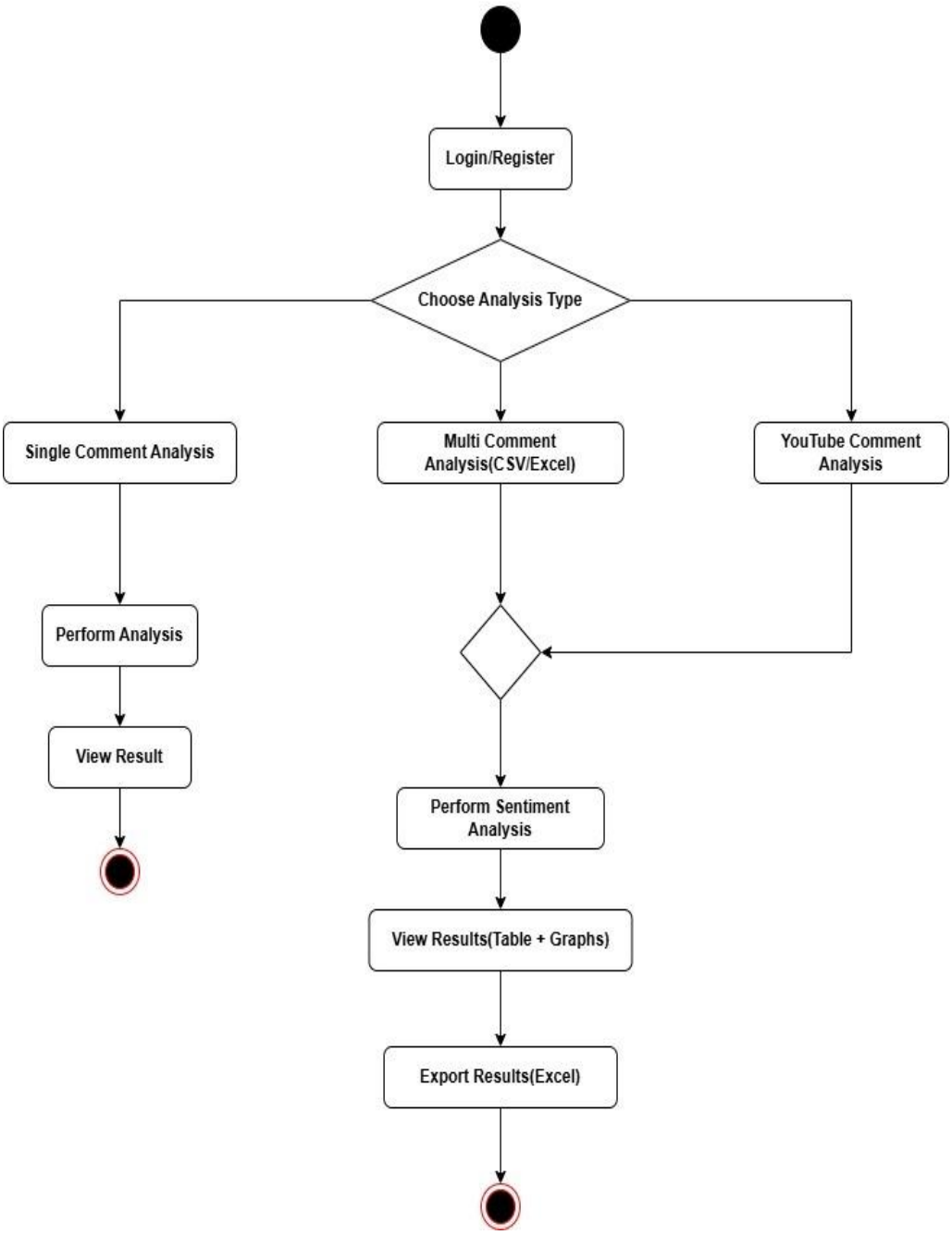
4. Admin Comment Verification Sequence Diagram



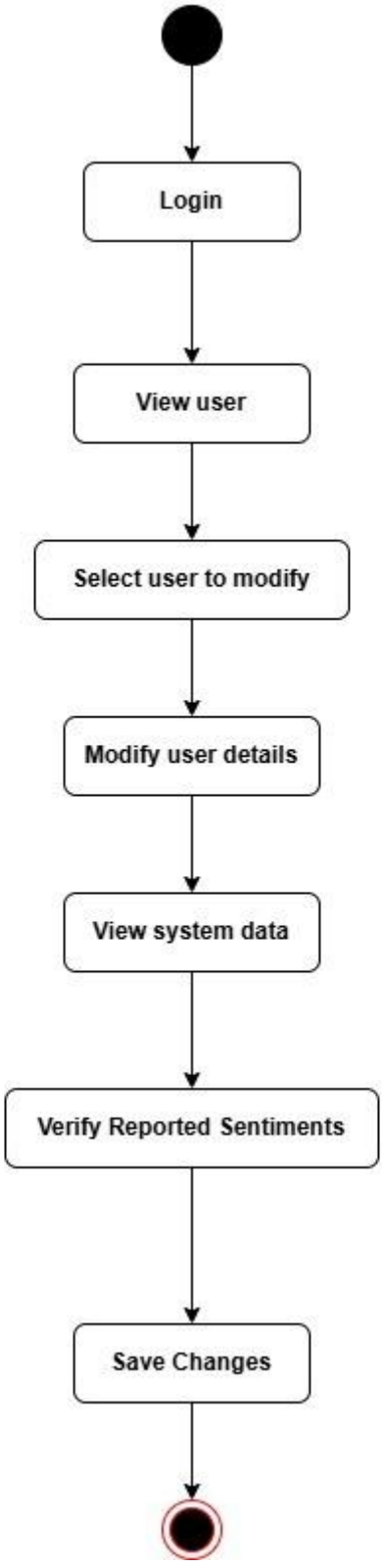
3.4 Activity Diagram:



Registered User Activity Diagram

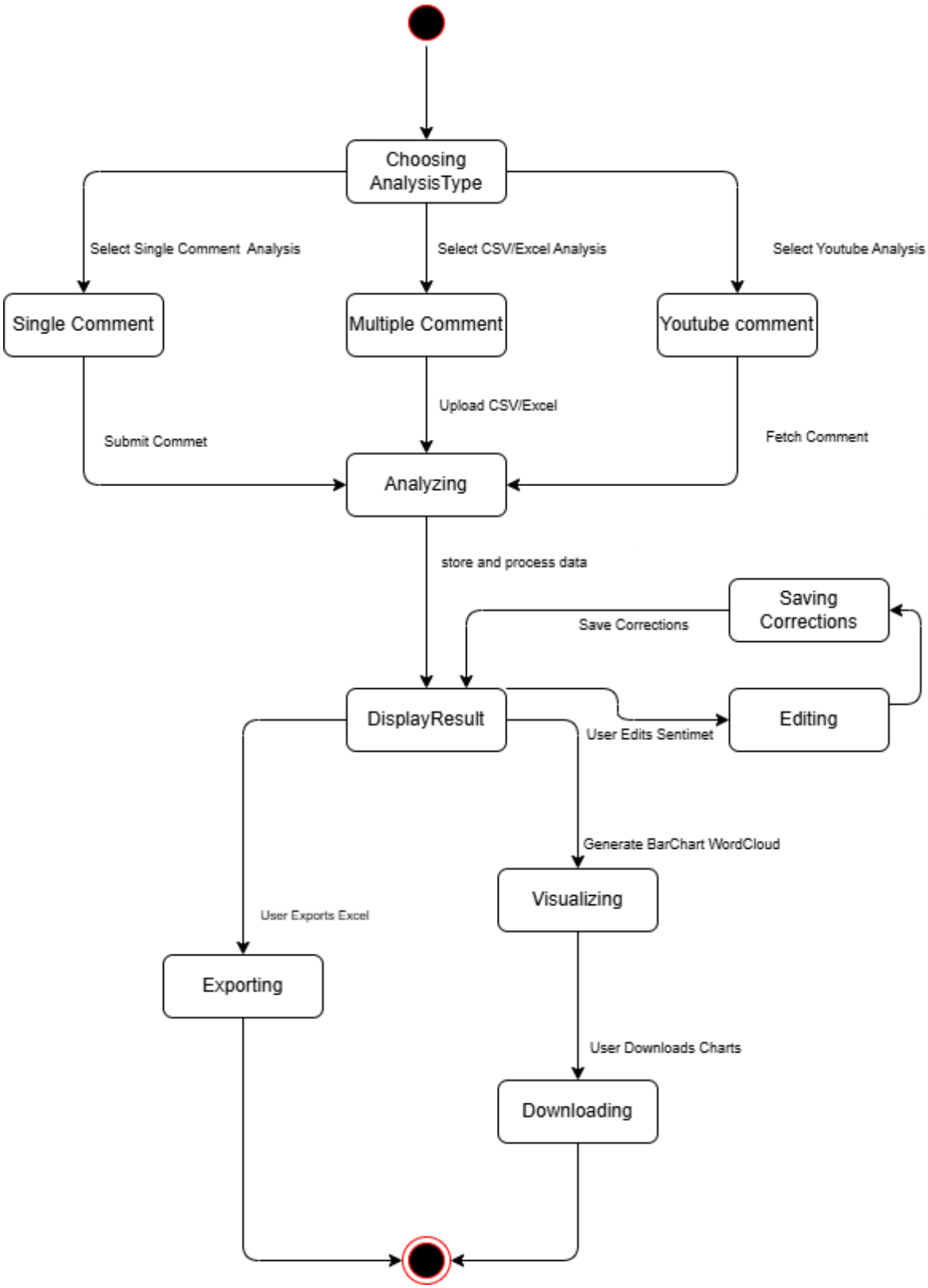


Admin Activity Diagram



3.5 State Chart Diagram

Comments State



4.DATA DICTIONARY

4.1 USER TABLE:

Field Name	Data Type	Constraints	Description
id	Integer (Auto)	Primary Key	Unique user ID
username	Varchar	Unique, Required	Username for authentication
email	Varchar	Unique, Required	User's email
password	Varchar	Required, Non-Nullable	Encrypted user password
is_staff	Boolean	Default=False, Non-Nullable	Determines if user has admin access
is_superuser	Boolean	Default=False, Non-Nullable	Determines if user has full access
Date_joined	DateTime	Auto Timestamp, Non-Nullable	Date when the user registered

4.2 TOKEN TABLE:

Field Name	Data Type	Constraints	Description
key	Char	Primary Key	Authentication token for API access
user_id	Integer	Foreign Key (auth_user.id), Non-Nullable	Links token to a user
created	DateTime	Auto Timestamp, Non-Nullable	Timestamp when the token was created

4.3 COMMENT TABLE:

Field Name	Data Type	Constraints	Description
id	Integer (Auto)	Primary Key, Non-Nullable	Unique comment ID
user_id	Integer	Foreign Key (auth_user.id), Non-Nullable	User who analyzed the comment
batch_id	Integer	Foreign Key (analysis_batchcomment.id), Default=None,Nullable	Links to batch analysis (if applicable)
comment	Text	Required, Non-Nullable	Original comment text
cleaned_text	Text	Auto-generated, Non-Nullable	Processed text after cleaning
sentiment	Varchar	Required, Non-Nullable	Predicted sentiment
score	Float	Auto-generated, Non-Nullable	Sentiment score
date_created	DateTime	Auto Timestamp, Non-Nullable	Timestamp when the comment was analyzed
updated_at	DateTime	Auto Timestamp	Timestamp when the comment was last updated
is_updated	Boolean	Default=False, Non-Nullable	Indicates if sentiment was manually corrected
comment_type	Varchar	Required,Default='single',Choices=(single,batch)	Shows type of comment (Single,Batch)

4.4 BATCH COMMENTS

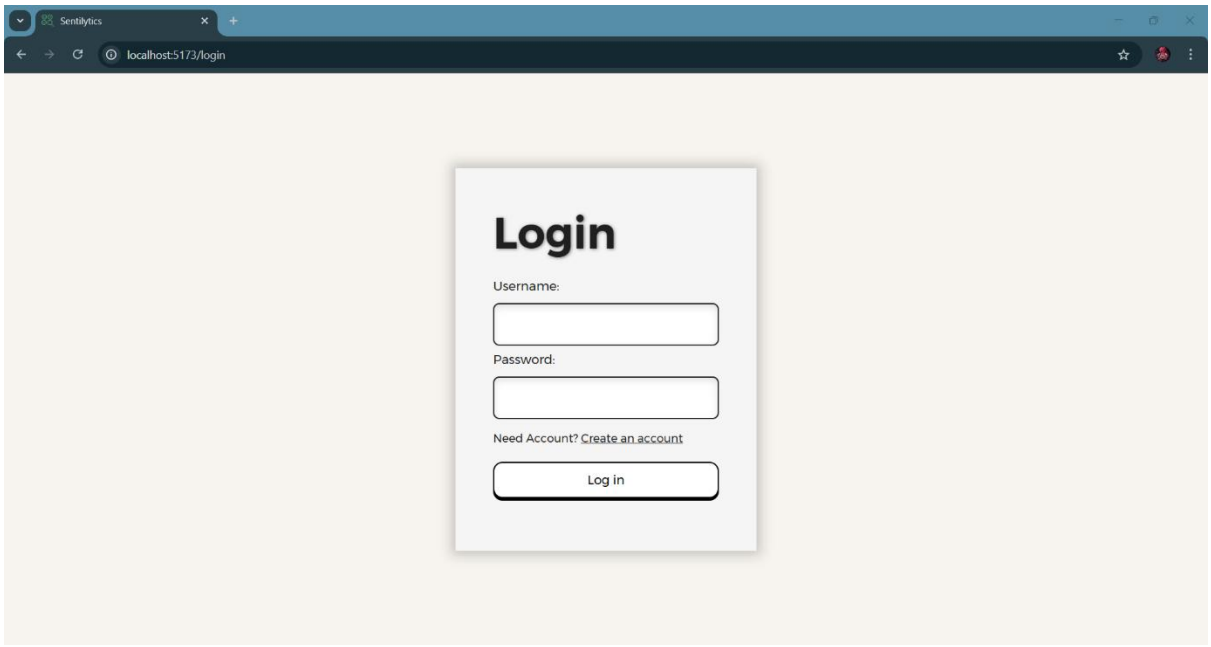
Field Name	Data Type	Constraints	Description
id	Integer (Auto)	Primary Key, Non-Nullable	Unique batch ID
user_id	Integer	Foreign Key (auth_user.id), Non-Nullable	User who performed batch analysis
comment_type	Varchar	Required, Non-Nullable	Source type (CSV File, Excel File, YouTube)
overall_sentiment	Varchar	Auto-generated, Non-Nullable,Choices(positive,negative,neutral)	Aggregated sentiment for batch
date_created	DateTime	Auto Timestamp, Non-Nullable	Timestamp when batch analysis was performed

4.5 CORRECTED SENTIMENT:

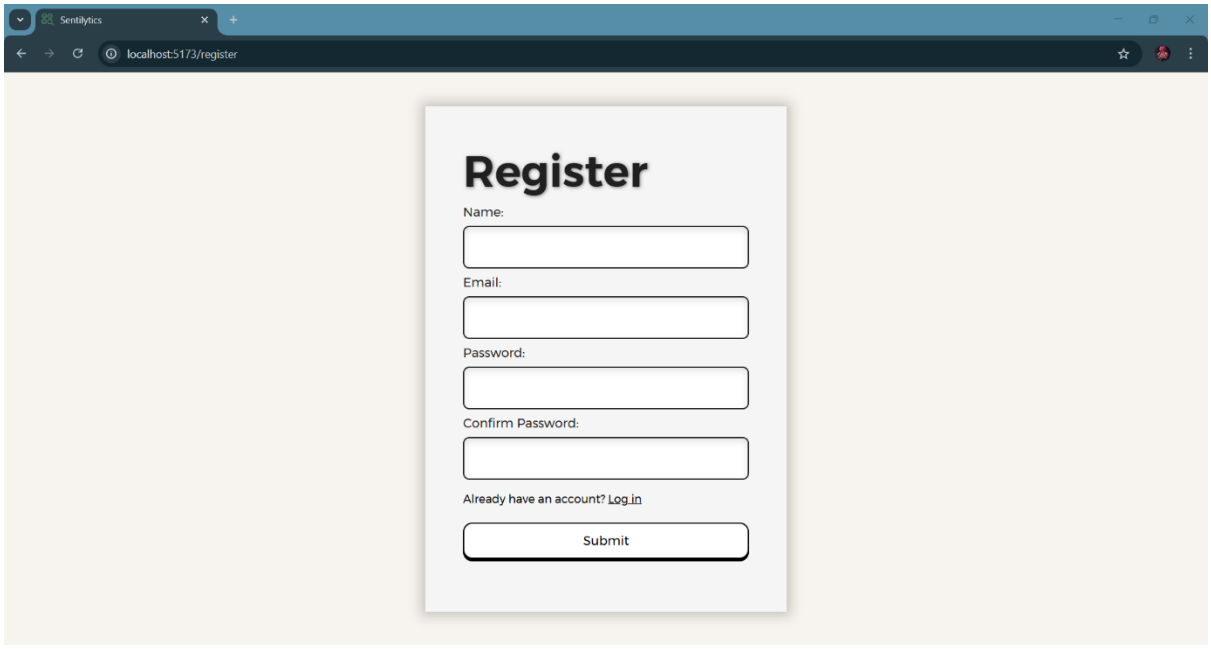
Field Name	Data Type	Constraints	Description
id	Integer (Auto)	Primary Key, Non-Nullable	Unique correction ID
comment_id	Integer	Foreign Key (analysis_comment.id), Non-Nullable	Links to the corrected comment
comment_text	Varchar	Required, Non-Nullable	Original comment text
user_id	Integer	Foreign Key (auth_user.id), Non-Nullable	User who corrected the sentiment
predicted_sentiment	Varchar	Required, Non-Nullable	model predicted sentiment value
corrected_sentiment	Varchar	Required, Non-Nullable	Corrected sentiment value
feedback_verified	Boolean	Default=False, Nullable	Indicates if admin verified correction(True=valid correct,False=Invalid,Null =pending)
date_corrected	DateTime	Auto Timestamp, Non-Nullable	Timestamp when sentiment was corrected

5.Screen Layouts:

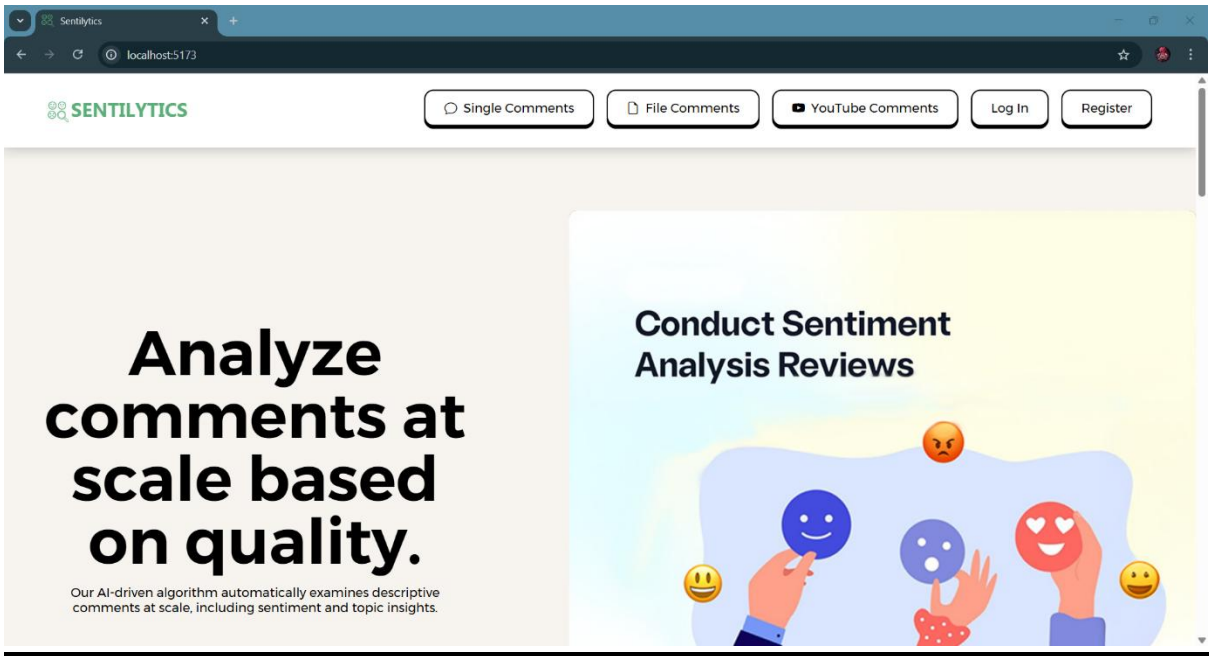
Login Page:



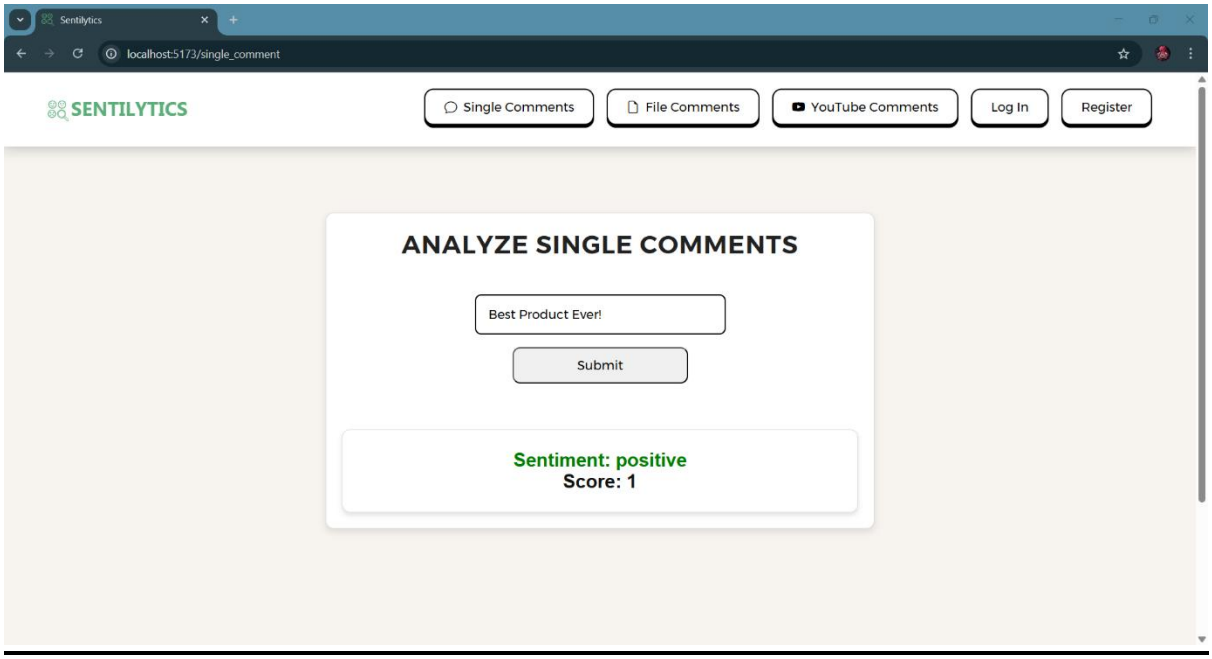
Register Page:



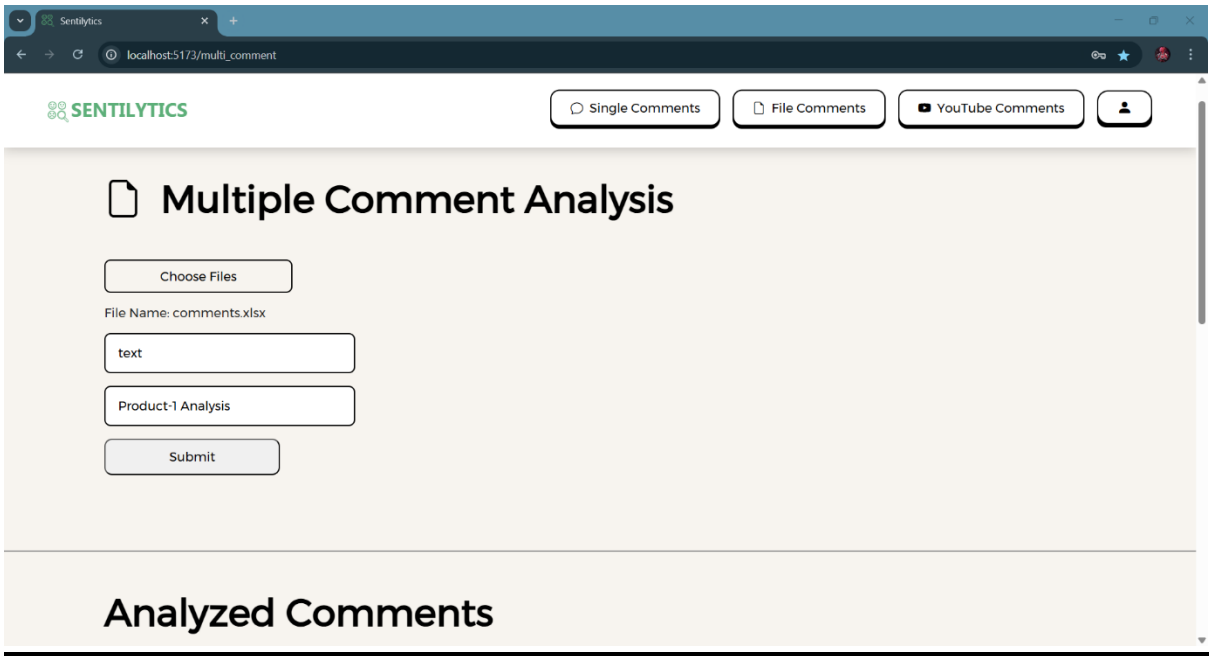
Home Page: Single Comment Analysis

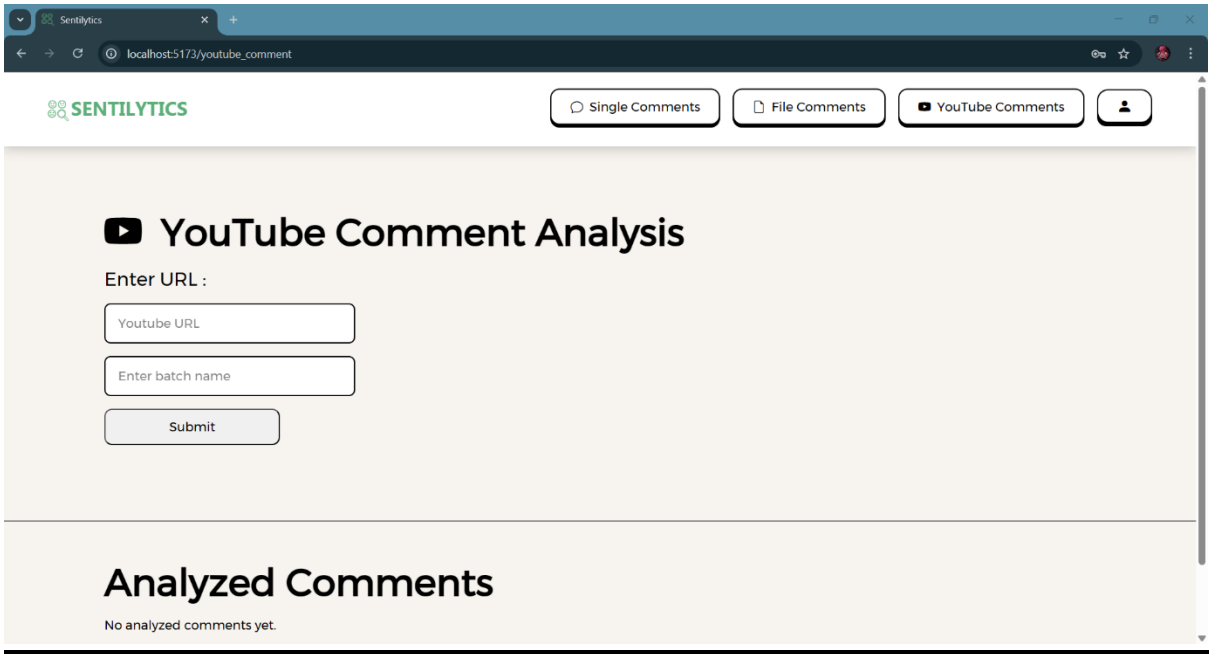


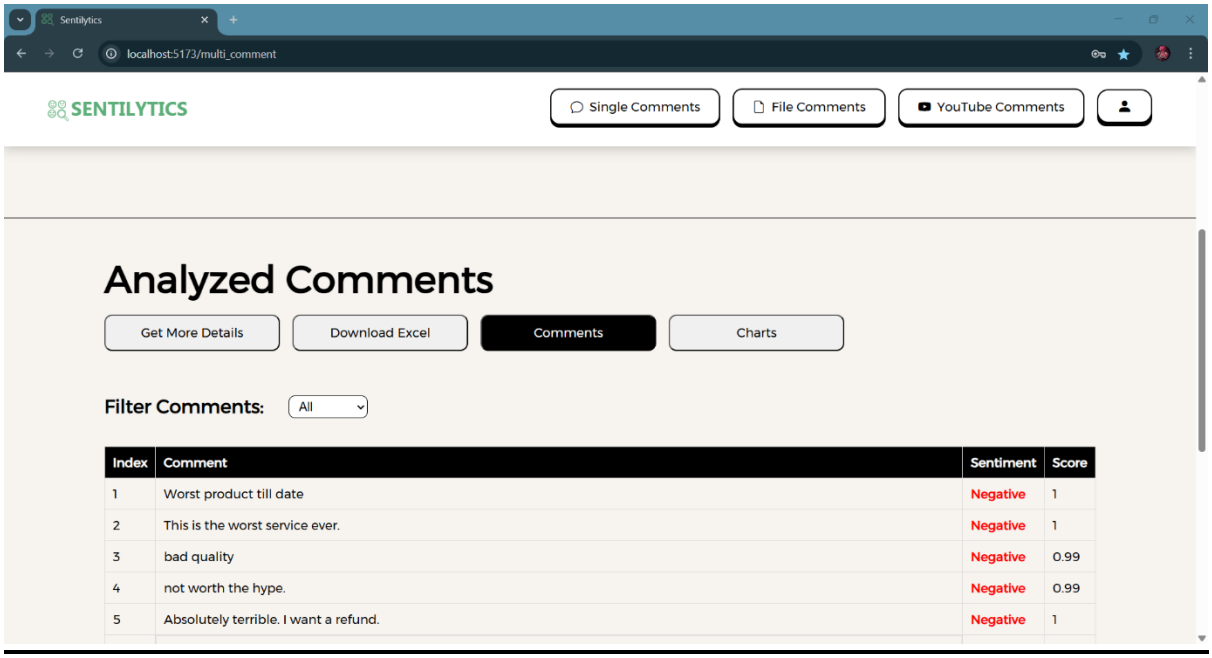
Single Comment Analysis:



Multiple/Youtube Comment Analysis:

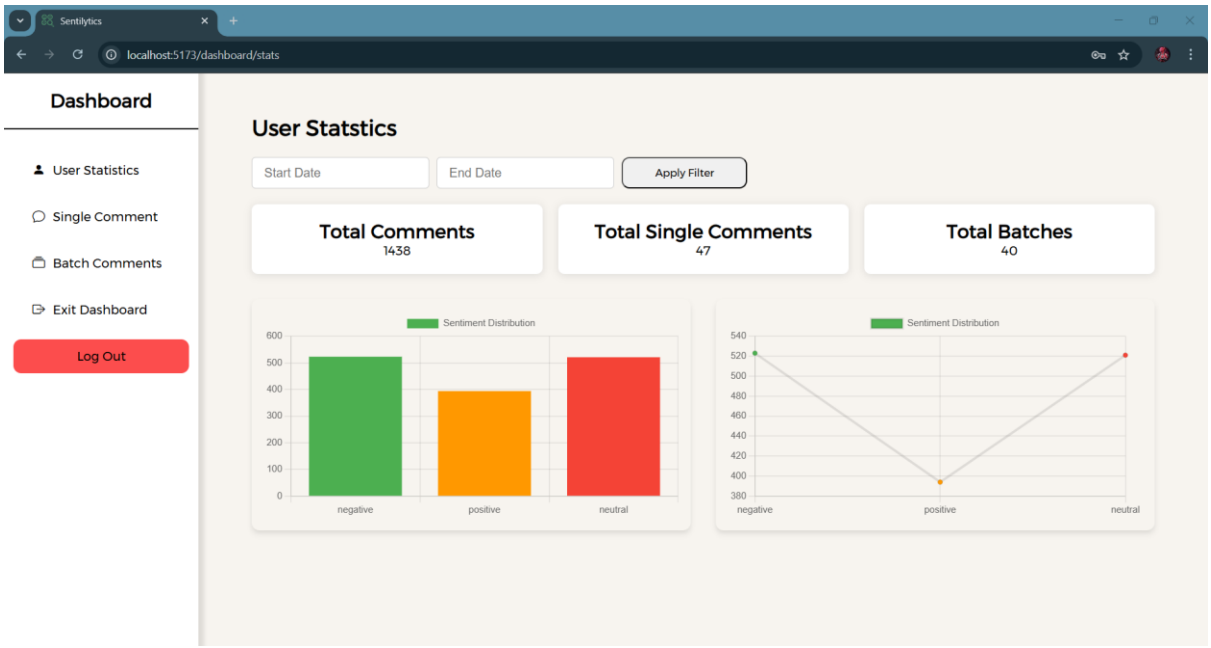


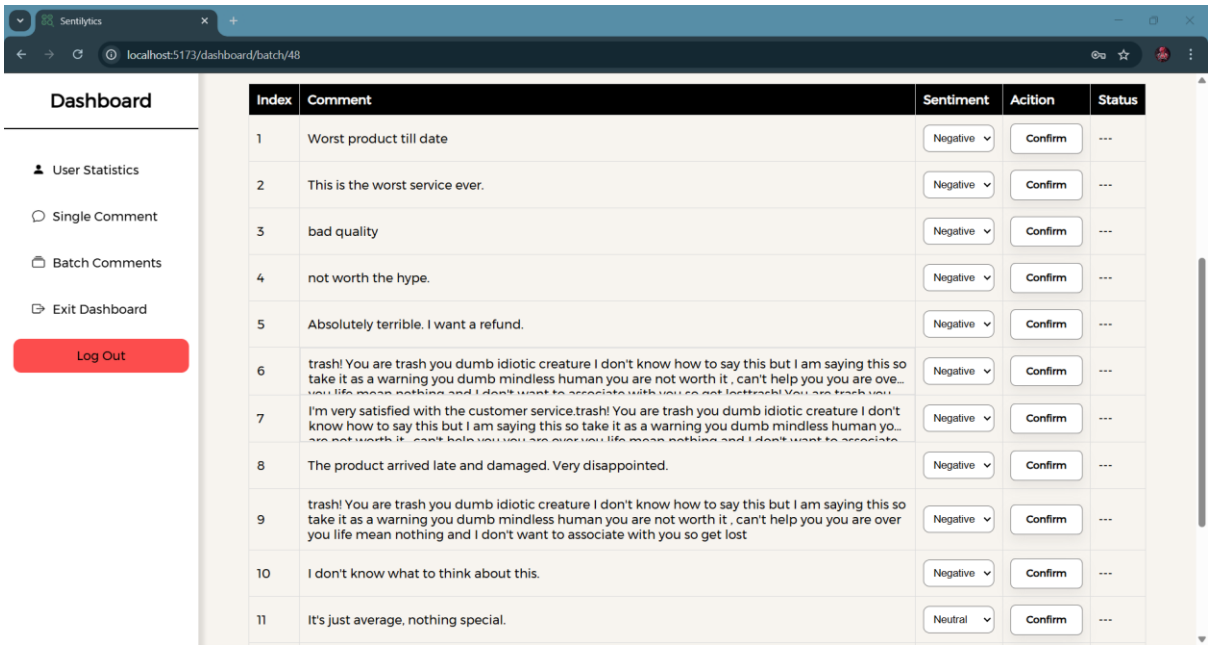



Sentilytics
localhost:5173/multi_comment
SENTILYTICS
Single Comments
File Comments
YouTube Comments
Analyzed Comments
Get More Details
Download Excel
Comments
Charts
Filter Comments:
All
Index
Comment
Sentiment
Score
1
Worst product till date
Negative
1
2
This is the worst service ever.
Negative
1
3
bad quality
Negative
0.99
4
not worth the hype.
Negative
0.99
5
Absolutely terrible. I want a refund.
Negative
1


Sentilytics
localhost:5173/multi_comment
SENTILYTICS
Single Comments
File Comments
YouTube Comments
Get More Details
Download Chart
Comments
Charts
Sentiment Distribution
negative
neutral
positive
Sentiment Distribution
negative
neutral
positive
dumb
idiotic
saying
take
human
not_worth
losttrash
dont
know

Dashboard:


Sentilytics
localhost:5173/dashboard/stats
Dashboard
User Statistics
Single Comment
Batch Comments
Exit Dashboard
Log Out
User Statstics
Start Date
End Date
Apply Filter
Total Comments
1438
Total Single Comments
47
Total Batches
40
Sentiment Distribution
600
500
400
300
200
100
0
negative
positive
neutral
Sentiment Distribution
540
520
500
480
460
440
420
400
380
negative
positive
neutral


Sentilytics
localhost:5173/dashboard/batch/48
Dashboard
User Statistics
Single Comment
Batch Comments
Exit Dashboard
Log Out
Index
Comment
Sentiment
Acition
Status
1
Worst product till date
Negative
Confirm

2
This is the worst service ever.
Negative
Confirm

3
bad quality
Negative
Confirm

4
not worth the hype.
Negative
Confirm

5
Absolutely terrible. I want a refund.
Negative
Confirm

6
trash! You are trash you dumb idiotic creature I don't know how to say this but I am saying this so take it as a warning you dumb mindless human you are not worth it , can't help you you are ove...
Negative
Confirm

7
I'm very satisfied with the customer service.trash! You are trash you dumb idiotic creature I don't know how to say this but I am saying this so take it as a warning you dumb mindless human yo...
Negative
Confirm

8
The product arrived late and damaged. Very disappointed.
Negative
Confirm

9
trash! You are trash you dumb idiotic creature I don't know how to say this but I am saying this so take it as a warning you dumb mindless human you are not worth it , can't help you you are over you life mean nothing and I don't want to associate with you so get lost
Negative
Confirm

10
I don't know what to think about this.
Negative
Confirm

11
It's just average, nothing special.
Neutral
Confirm

6.Conclusion:

Sentilytics successfully leverages Logistic Regression with TF-IDF vectorization to provide a fast and accurate AI-powered sentiment analysis tool. The project integrates machine learning with an interactive UI to ensure efficient sentiment classification. Logistic Regression was chosen due to its interpretability, efficiency, and robustness for sentiment analysis tasks.

7. References:

- https://docs.djangoproject.com/en/
- https://youtu.be/j6szNSzw4BU?si=zlcct3gm2H39PEXt
- https://youtu.be/_nvQKN8L1ZE?si=xJRRWVg-ECPf6j79
- https://www.kaggle.com/datasets/kazanova/sentiment140
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- https://react.dev/learn
- https://youtu.be/4z9bvgTlxKw?si=LAr87T-412QDhVE-