

**A
Project Report
On
"EDA APP AND TWITTER
SENTIMENT ANALYSIS DASHBOARD"**

Prepared by

**Kenee Patel-19DCS089
Parth Patel-19DCS098
Shruti Patel-19DCS102**

Under the guidance of

Krishna Patel

Assistant Professor

A Report Submitted to

**Charotar University of Science and Technology
for Partial Fulfillment of the Requirements for the
6th Semester [CS357] Software Group Project-IV**

Submitted at



**CSE
DEPSTAR
APRIL-2022**

CERTIFICATE

This is to certify that the report entitled **“EDA APP AND TWITTER SENTIMENT ANALYSIS DASHBOARD”** is a bonafied work carried out by **Shruti Patel(19DCS102)** under the guidance and supervision of **Assistant Prof. Krishna Patel** for the subject **[CS357] Software Group Project-IV (CSE)** of 6th Semester of Bachelor of Technology in **DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Krishna Patel

Assistant Professor Cum
Research Fellow (CSE)
DEPSTAR, Changa, Gujarat.

Dr. Parth Goel

Head Of Department –
Computer Science &
Engineering, DEPSTAR,
CHARUSAT, Changa,
Gujarat

Dr. Amit Ganatra

DEPSTAR Dean, FTE
CHARUSAT, Changa,
Gujarat.

DECLARATION BY THE CANDIDATES

We hereby declare that the project report entitled “**EDA App and Twitter Sentiment Analysis Dashboard**” submitted by us to Devang Patel Institute of Technology and Advanced Research, CHARUSAT, Changa in partial fulfilment of the requirement for the award of the degree of **B.Tech** in Computer Engineering, from Department of Computer Science and Engineering , DEPSTAR, is a record of bonafide CS357 Software Group Project carried out by us under the guidance of **Prof. Krishna Patel**. We further declare that the work carried out and documented in this project report has not been submitted anywhere else either in part or in full and it is the original work, for the award of any other degree or diploma in this institute or any other institute or university.

Parth Patel(19DCS098)

Prof. Krishna Patel

Assistant Professor Cum Research Fellow

Computer Science & Engineering (CSE)

DEPSTAR, Changa, Gujarat.

ABSTRACT

In today's scenario, where we are generating data at an unprecedented rate and we are living in an era where future oil of the world is going to be the data. We are witnessing a paradigm shift where more and more companies and organizations are relying on the data driven decision making. We this, comes the advancement and dominance of the fields of data analytics and data science. These two domains are heavily data dependent and thus the importance of the data is increased. EDA stands for Exploratory Data Analysis [11] which is an important is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. [8] EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. With this, the need of sentiment analysis is also coming into demand as the use of social media is firing to a record.

Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information. [11]

Sentiment analysis systems help organizations gather insights from unorganized and unstructured text that comes from online sources such as emails, blog posts, support tickets, web chats, social media channels, forums and comments. [8] Algorithms replace manual data processing by implementing rule-based, automatic or hybrid methods. Rule-based systems perform sentiment analysis based on predefined, lexicon-based rules while automatic systems learn from data with machine learning techniques. A hybrid sentiment analysis combines both approaches.

In addition to identifying sentiment, opinion mining can extract the polarity (or the amount of positivity and negativity), subject and opinion holder within the text. Furthermore, sentiment analysis can be applied to varying scopes such as document, paragraph, sentence and sub-sentence level

ACKNOWLEDGEMENT

We, the developers of “EDA APP AND TWITTER SENTIMENT ANALYSIS DASHBOARD”, with immense pleasure and commitment would like to present the project assignment. An EDA is thorough examination meant to uncover underlying structure of dataset to expose the trends and patterns beneath it. Sentiment analysis refers to identifying as well as classifying the sentiments that are expressed in the text source.

We hereby avail this opportunity to express our gratitude to number of people who extended their valuable time, full support and cooperation in developing the project. Every work that one completes successfully stands on the constant encouragement, good will and support of the people around. I hereby avail this opportunity to express my gratitude to number of people who extended their valuable time, full support and cooperation in developing the project.

We express deep sense of gratitude towards our Head of the CSE Department, Dr. Parth Goel and project guide Prof. Krishna Patel for the support during the whole session of study and development. It is because of them, that we were prompted to do hard work, adopting new technologies.

They altogether provided us favourable environment, and without them it would not have been possible to achieve my goal.

Thanks,

Parth Patel(19DCS098)

TABLE OF CONTENTS

DECLARATION.....	i
ABSTRACT.....	ii
ACKNOWLEDGEMENT.....	iii
CHAPTER 1: INTRODUCTION.....	1
1.1 PROJECT OVERVIEW.....	2
1.2 PROJECT OBJECTIVE.....	3
1.3 SCOPE.....	3
1.4 SYSTEM PHASES.....	4
1.5 TOOLS AND TECHNOLOGIES USED.....	4
CHAPTER 2: PROJECT MANAGEMENT.....	06
2.1 PROJECT PLANNING.....	07
CHAPTER 3: SYSTEM REQUIREMENTS STUDY.....	09
3.1 REQUIREMENTS.....	10
3.2 LIMITATIONS	10
3.3 EXPECTED OUTCOMES.....	11
3.4 FUTURE OUTCOMES.....	11
CHAPTER 4: SYSTEM ANALYSIS.....	12
4.1 USER CHARACTERISTICS.....	13
4.2 USE CASE DIAGRAM.....	13
4.3 ACTIVITY DIAGRAM.....	14
4.4 SEQUENCE DIAGRAM.....	15
4.5 FLOW CHART FOR EDA APP.....	16
4.6 FLOW CHART FOR SENTIMENT ANALYSIS DASHBOARD.....	17
CHAPTER 5: SYSTEM IMPLEMENTATIONS AND STANDARDS.....	18
5.1 IMPLEMENTATION SCREENSHOTS.....	19
5.2 CODING STANDARDS.....	24
5.3 OPERATING ENVIRONMENT.....	25
5.4 RISK ANALYSIS AND MANAGEMENT.....	26
5.5 TEST SUITE DESIGN.....	29
5.6 TEST CASES.....	30
CHAPTER 6: CONCLUSION.....	32
6.1 CONCLUSION.....	33
6.2 DIFFICULTIES FACED.....	33
WORK CITED/ BIBLIOGRAPHY.....	34
LIST OF FIGURES	
FIGURE 2.1.....	07
FIGURE 4.1.....	13
FIGURE 4.2.....	14
FIGURE 4.3.....	15
FIGURE 4.4.....	16
FIGURE 4.5.....	17
FIGURE 5.1.....	19
FIGURE 5.2.....	20
FIGURE 5.3.....	20

FIGURE 5.4.....	21
FIGURE 5.5.....	21
FIGURE 5.6.....	22
FIGURE 5.7.....	22
FIGURE 5.8.....	23
FIGURE 5.9.....	23

LIST OF TABLES

TABLE 5.1.....	24
TABLE 5.2.....	26
TABLE 5.3.....	28
TABLE 5.4.....	30
TABLE 5.5.....	31
TABLE 5.6.....	31

CHAPTER 1: INTRODUCTION

1.1 PROJECT OVERVIEW

The overall project is divided into two phases.

The two phases are as follows

1. Development of exploratory data analysis application
2. Development of Twitter sentiment analysis dashboard

So talking about the phase one of the project the project is all about development and analysis of the exploratory data analysis application which is also commonly known as EDA app.

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. [8]

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset.

The second phase of project that is the development and the analysis of Twitter sentiment analysis dashboard. This is regarding the analysis of sentiments or which is commonly called as sentiment analysis. this kind of analysis is very famous and very common and in demand nowadays because lots and lots of companies are trying to influence their customers with the help of social media. social media is a place millions of users swipe and oh content switching now and then 24 by 7. Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information. [11]

1.2 PROJECT OBJECTIVE

Main objective behind the development of this project was to facilitate and is of working for the Aspiring and current data analysts and data scientists. We know that in each and every project of data science and Data Analytics EDA is an important part. In order to decide whether which data set is correct for the project and which data set is not correct for the project each and every analyst and data scientist has to code extra 60 to 120 lines extra for each and every data set. So we have tried to automate this process. The exploratory data analysis application that we have built tries to automate the process and also aims at saving time of the analyst.

The second phase of our project was to develop Twitter sentiment analysis dashboard. We need to clarify that dashboard means feeding on live data. Now this tool will help sentiment analysts to analyse the sentiment for the current trend of emotions of the people towards certain tweets. Twitter is a Battleground of tech companies, its Battleground of the world where people battle with hashtags and their words. In order to determine the sentiment of an overall population towards a particular tweet many companies and Organisation have started building up sentiment analysis teams which performs analysis on the sentiments of people on the social media. This is particularly helpful for the company in order to develop their future publicity campaign in order to make changes in the environment so that the productivity is increased and also their profit is increase which is the ultimate goal of each and every organisation.

1.3 SCOPE

Never-before-explored setting: The simple technology used in the development of the application, provides a lucid experience specially for the EDA app as it is very simple to use.

Native Support: The project was started with the domain of Data analytics and the project is leaned towards making the work of data analysts easier by automating certain tasks and to provide a ready-made platform through which they can get insights from the data and thus can dedicate their time towards the analysis phase.

1.4 SYSTEM PHASES

Phase: 1

- EDA APP.
- Upon discussion with analysts, we found out a problem.
- We started working on them.
- We used streamlit and python for it.

Phase: 2

- Twitter sentiment analysis dashboard.
- We used flask, vuejs frameworks for it.
- Used NAIVE BAYES CLASSIFIER in backend.
- Created pipeline of the model.
- Developed front end and back end separately.
- Waiting for the twitter tokens.

1.5 TOOLS AND TECHNOLOGIES USED

- **Python (Version = 3.10.0)**

Python is a high-level, general-purpose programming language. [1] We used python to develop our machine learning model. We made use of many libraries also.

- **Node js**

Node.js is an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser. [2] We used node js in our backend development.

- **Vue js**

Vue.js is an open-source model–view–viewmodel front end JavaScript framework for building user interfaces and single-page applications. [3] We used Vue js in our frontend development.

- **Flask**

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. [4] We used flask as our framework for backend development.

- **VS Code**

Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. [5] We used vs code as our primary IDE.

- **Jupyter Notebook**

Project Jupyter is a project and community whose goal is to "develop open-source software, open-standards, and services for interactive computing across dozens of programming languages". [6] We used jupyter notebook for developing our machine learning model.

- **Nltk**

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language. [7]

- **Venv**

The module used to create and manage virtual environments is called venv [8]

- **Bootstrap**

Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains HTML, CSS and JavaScript-based design templates for typography, forms, buttons, navigation, and other interface components. [9]

- **Streamlit**

Streamlit is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time [10]

CHAPTER 2:

PROJECT MANAGEMENT

2.1 PROJECT PLANNING

The SPIRAL model is a risk-driven process model generator for software projects. Based on the unique risk patterns of a given project, the spiral model guides a team to adopt elements of one or more process models, such as incremental, waterfall, or evolutionary prototyping.

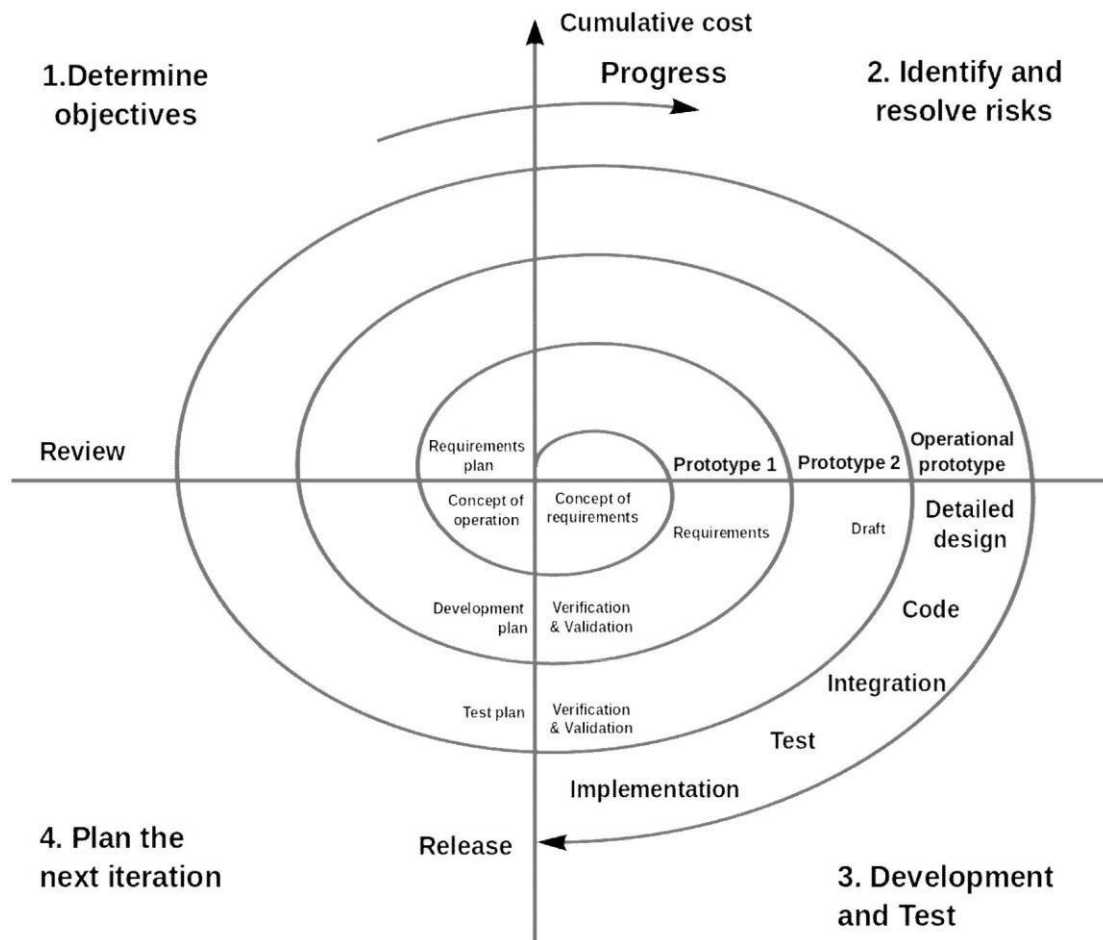


FIGURE-2.1: SPIRAL CHART OF EDA

1. The requirements are known in advance of implementation.
2. The requirements have no unresolved, high-risk implications, such as risks due to cost, schedule, performance, safety, security, user interfaces, organizational impacts, etc.
3. The nature of the requirements will not change very much during development or evolution.

4. The requirements are compatible with all the key system stakeholders' expectations, including users, customer, developers, maintainers, and investors.
5. The right architecture for implementing the requirements is well understood.
6. There is enough calendar time to proceed sequentially.

Perform four basic activities in every cycle:

This invariant identifies the four activities that must occur in each cycle of the spiral model:

1. Consider the win conditions of all success-critical stakeholders.
2. Identify and evaluate alternative approaches for satisfying the win conditions.
3. Identify and resolve risks that stem from the selected approach(es).
4. Obtain approval from all success-critical stakeholders, plus commitment to pursue the next cycle.

Key idea: On each iteration identify and solve the sub-problems with the highest risk.

Advantages:

1. Realism: the model accurately reflects the iterative nature of software development on projects with unclear requirement.
2. Flexible: incorporates the advantages of the waterfall and evolutionary methods.
3. Comprehensive model decreases risk.
4. Good project visibility.

CHAPTER 3:

SYSTEM REQUIREMENT STUDY

3.1 REQUIREMENTS

3.1.1 HARDWARE REQUIREMENTS:

- Recommended 1 GB of RAM.
- Recommended 14 inch of display.
- A touchscreen for lucid experience (optional)

3.1.2 SOFTWARE REQUIREMENTS:

- Web Browser (Chrome Recommended).
- Consistent Internet Connection.

3.2 LIMITATIONS

- Gives general EDA not Specific EDA
- Dirty Data might affect the EDA
- Dashboard usage limited to twitter's scope
- Incorrectly Targeted Sentiment.
- Inability to perform well in different domains, inadequate accuracy and performance.
- Insufficient labeled data, incapability to deal with complex sentences that require more than sentiment words and simple analyzing.

3.3 EXPECTED OUTCOMES

- This project will save initial cost and time taken by data analyst to finalize the data set or to judge whether the data set is appropriate or not.
- This project will give a glimpse of data and it's underlying impurities.
- The Project provides a detailed information about the data and their types
- The project provides correlation between the data from the data set.
- The project provides analysis dashboard for the tweets.
- The project provides a platform to interact with the sentiment analysis.

3.4 FUTURE ENHANCEMENT

- The team will try hard to make it more efficient than the existing application.
- Optimization of the ML algorithm.
- Connection on frontend with backend.
- Testing of the project.
- Integration of facebook with twitter

CHAPTER 4:

SYSTEM ANALYSIS

4.1 USER CHARACTERISTICS

- The end user is the soul user of the application.
- The admin can make the updates

4.2 USE CASE DIAGRAM

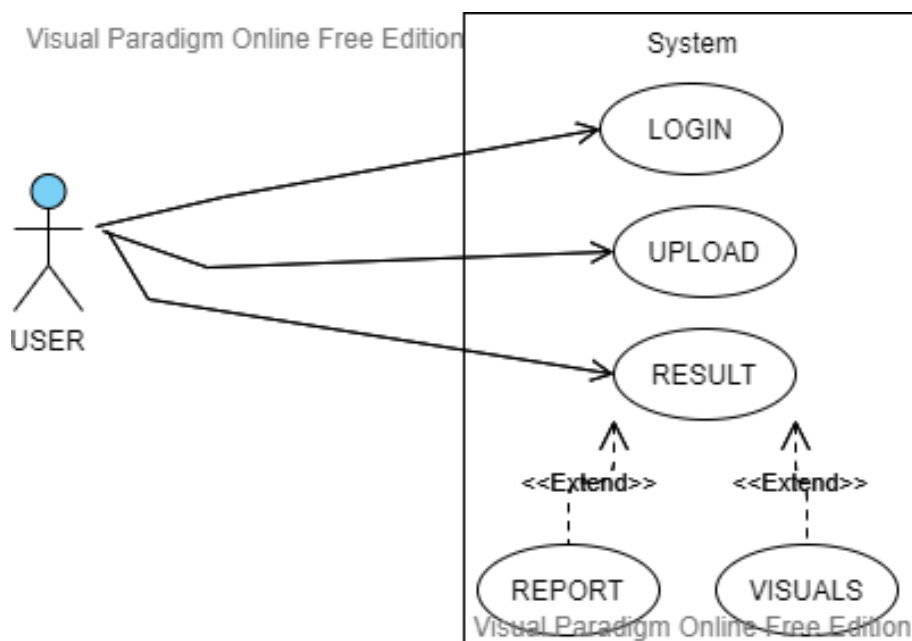


FIGURE-4.1 USE CASE DIAGRAM OF EDA

The figure shows the use case diagram of the phase

4.3 ACTIVITY DIAGRAM

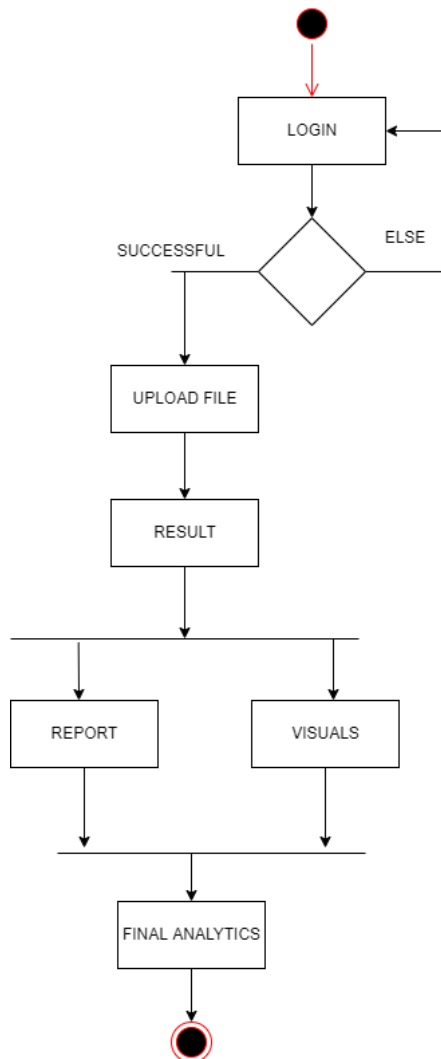


FIGURE-4.2 ACTIVITY DIAGRAM OF EDA

The figure shows the activity diagram of the phase

4.4 SEQUENCE DIAGRAM

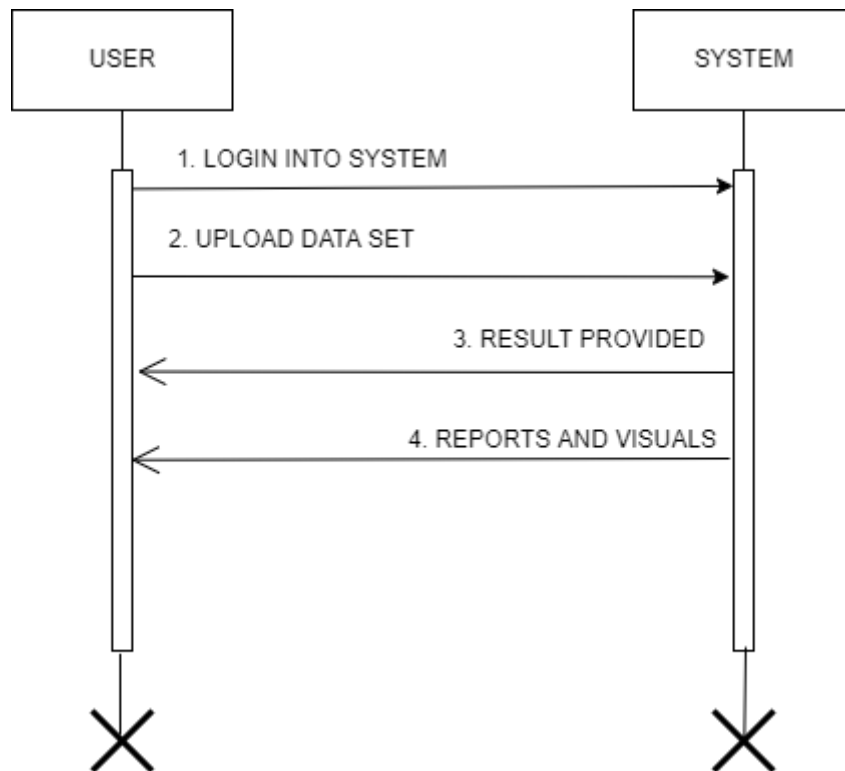


FIGURE-4.3 SEQUENCE DIAGRAM OF EDA

The figure shows the sequence diagram of the phase

4.5 FLOW CHART FOR EDA APP

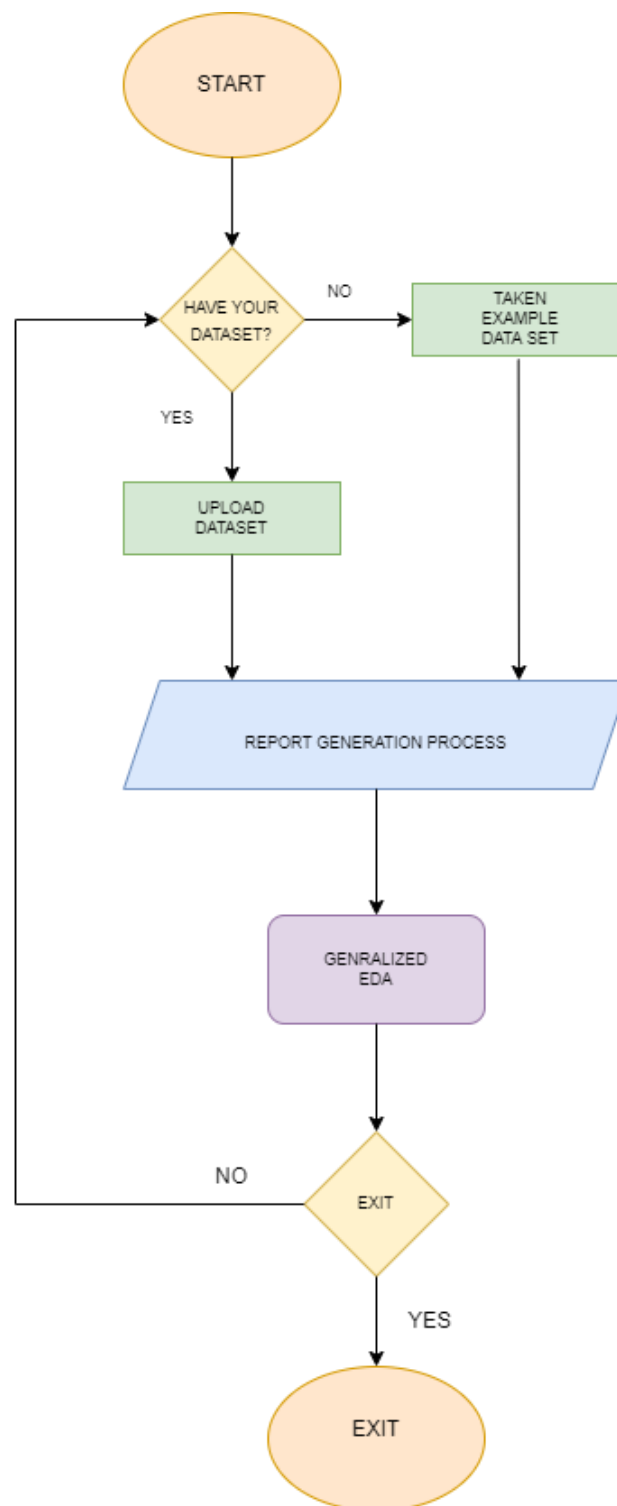


FIGURE-4.4 FLOWCHART FOR EDA

The figure depicts the flow chart of EDA

4.6 FLOW CHART FOR SENTIMENT ANALYSIS DASHBOARD

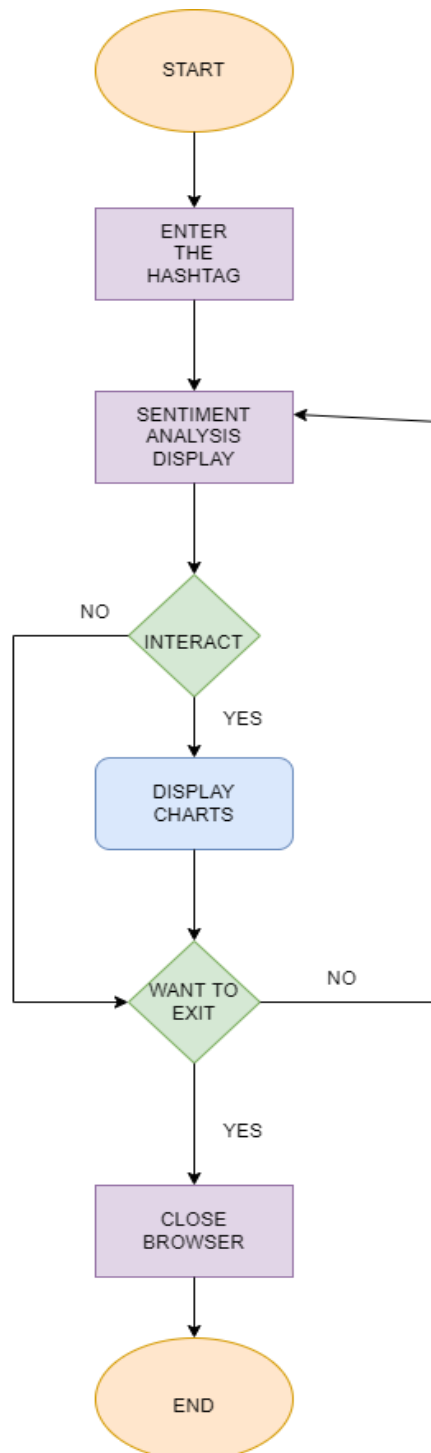


FIGURE-4.5 FLOWCHART FOR SENTIMENTAL ANALYSIS

The figure depicts the flow chart of sentiment analysis dashboard

CHAPTER 5:

SYSTEM IMPLEMENTATION AND

STANDARDS

5.1 IMPLEMENTATION SCREENSHOTS

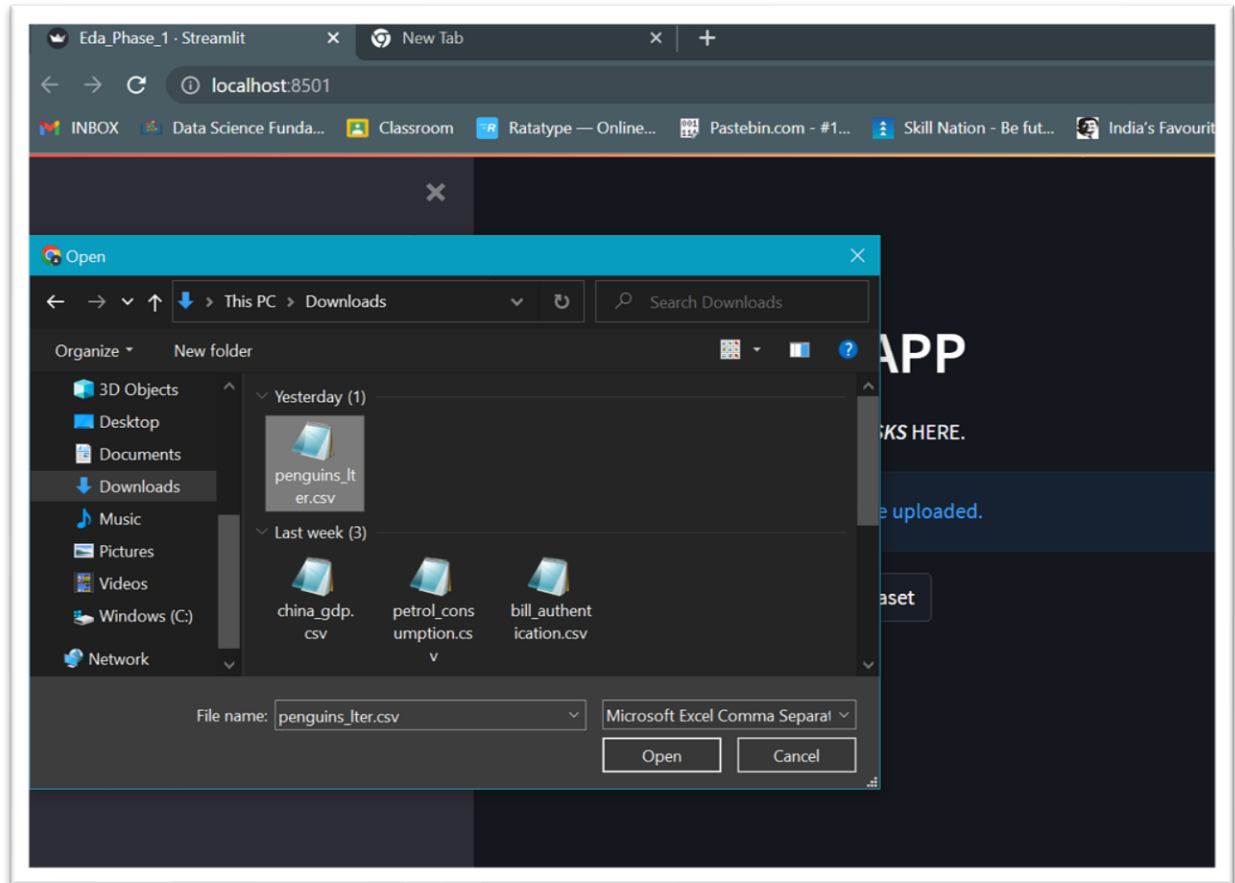
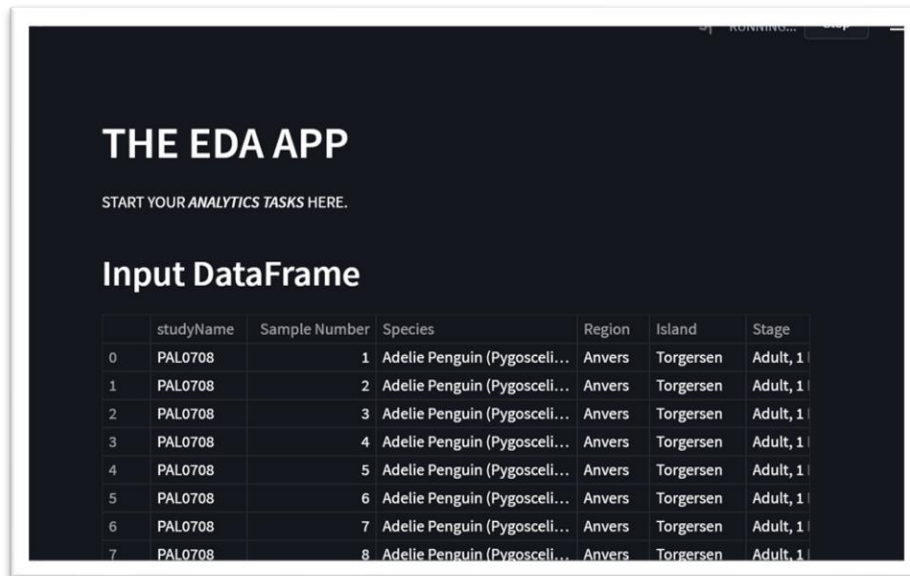


FIGURE-5.1: UPLOAD SECTION OF EDA APP

This is the upload section of the EDA app. The user can upload the data set from this section.



THE EDA APP

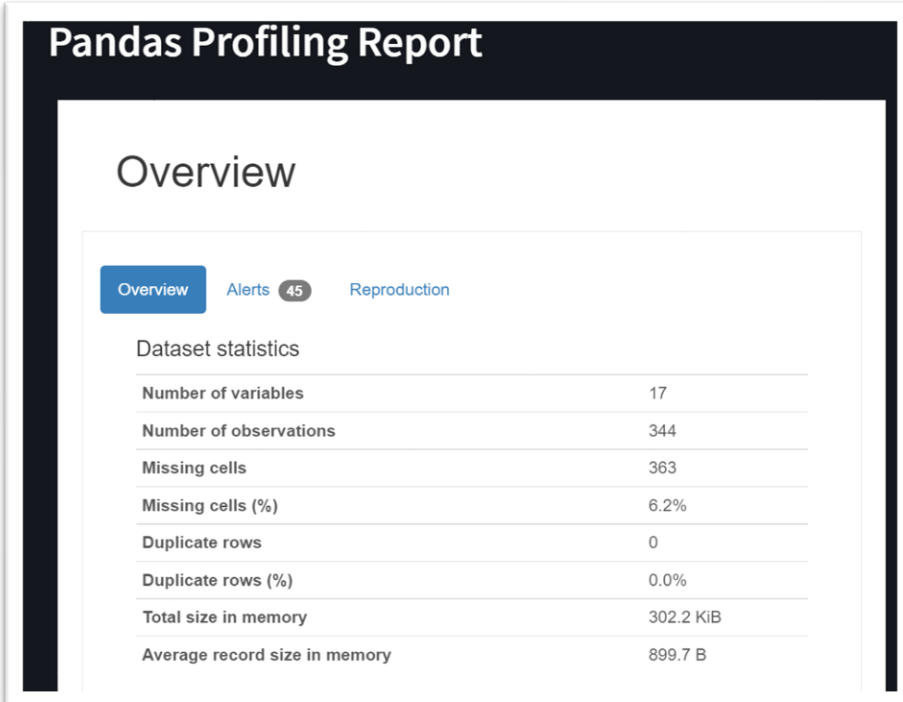
START YOUR ANALYTICS TASKS HERE.

Input DataFrame

	studyName	Sample Number	Species	Region	Island	Stage
0	PAL0708	1	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
1	PAL0708	2	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
2	PAL0708	3	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
3	PAL0708	4	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
4	PAL0708	5	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
5	PAL0708	6	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
6	PAL0708	7	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1
7	PAL0708	8	Adelie Penguin (Pygosceli...	Anvers	Torgersen	Adult, 1

FIGURE-5.2: THE DATASET UPLOADED DISPLAYED IN THE PANEL

Once the data set is uploaded. The application will start to look at the data set. Firstly, the glimpses of the data set is shown. The dataset is converted into dataframe.



Pandas Profiling Report

Overview

Overview Alerts 45 Reproduction

Dataset statistics

Number of variables	17
Number of observations	344
Missing cells	363
Missing cells (%)	6.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	302.2 KiB
Average record size in memory	899.7 B

FIGURE-5.3: THE STATISTICS OF THE EDA

The report generated consists of the overview section, which gives even the minute details of the dataset.

itemID	Sentiment	Sentiment text
0	1	0 is so sad for my APL frie...
1	2	0 I missed the New Moon trail...
2	3	1 omg its already 7:30 :O
3	4	0 .. Omgaga. Im sooo im gunna CRy. I'...
4	5	0 i think mi bf is cheating on me!!! ...

FIGURE-5.4: THE GLIMPSES OF THE DATASET USED

The above figures shows the head section of the dataset which is used to train the model.

	precision	recall	f1-score	support
0	0.72	0.76	0.74	8203
1	0.78	0.73	0.76	9210
accuracy			0.75	17413
macro avg	0.75	0.75	0.75	17413
weighted avg	0.75	0.75	0.75	17413

FIGURE-5.5: ACCURACY REPORT

The above figure shows the accuracy statistics of the model used. The accuracy is not great but sufficient to deploy the pipeline.

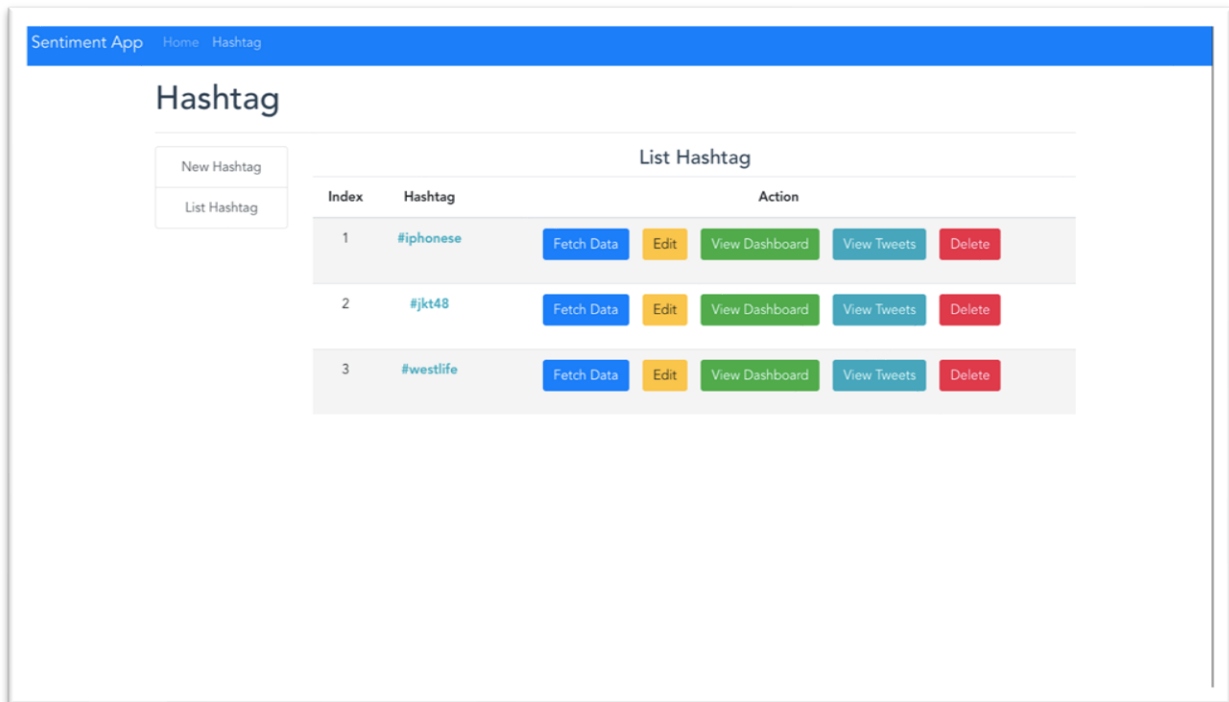


FIGURE-5.6: DASHBOARD PANEL

The dashboard panel of the sentiment website. The panel shows the hashtags used before.

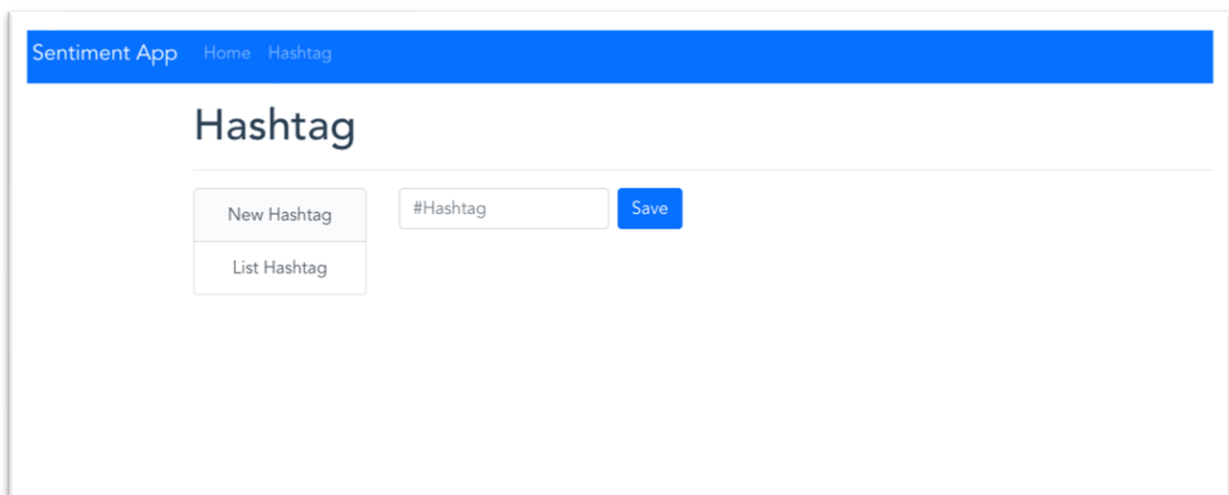


FIGURE-5.7: HASHTAG SECTION

This section allows the user to add or remove the hashtags.

Sentiment App Home Hashtag			
Tweets contain #iphonese			
Index	Username	Tweet	Sentiment
1	@hmednasri14	RT @imMadhu1: iPhone SE 2020 (\$399) + #AppleWatch Series 5 (\$399) + #iPad 10.2 inch (\$329) + #AirPods 2nd Gen (\$199) = \$1326. #iPhone 11 P...	negative
2	@Popcornsmile	RT @phanssss: อากาศวันนี้ร้อนไทยใช้ iMessages มากๆ ทุกวันนี้คุยกับเพื่อนเรื่อง sensitive สำหรับประเทศไทย จะคุยในนี้ และ sticker คือน่ารัก...	positive
3	@nnuch12550326	RT @phanssss: อากาศวันนี้ร้อนไทยใช้ iMessages มากๆ ทุกวันนี้คุยกับเพื่อนเรื่อง sensitive สำหรับประเทศไทย จะคุยในนี้ และ sticker คือน่ารัก...	positive
4	@Ratul_Deoun	RT @reneritchie: More #iPhoneSE reviews: @iJustine: https://t.co/sEvUbRdtZC @MKBHD: https://t.co/9qH84JcfKI @UrAvgConsumer https://t.co/...	positive
5	@WChayklang	RT @phanssss: อากาศวันนี้ร้อนไทยใช้ iMessages มากๆ ทุกวันนี้คุยกับเพื่อนเรื่อง sensitive สำหรับประเทศไทย จะคุยในนี้ และ sticker คือน่ารัก...	positive
6	@4582kilometers	RT @phanssss: อากาศวันนี้ร้อนไทยใช้ iMessages มากๆ ทุกวันนี้คุยกับเพื่อนเรื่อง sensitive สำหรับประเทศไทย จะคุยในนี้ และ sticker คือน่ารัก...	positive
7	@nearacharger	RT @Vyyper: ??????????NEW VIDEO????????? Should you upgrade to the iPhone SE 2020? https://t.co/d0s67rZLEK RETWEETS PLEASE? ?????????? #Apple #techno...	positive
8	@atothepleng	RT @phanssss: อากาศวันนี้ร้อนไทยใช้ iMessages มากๆ ทุกวันนี้คุยกับเพื่อนเรื่อง sensitive สำหรับประเทศไทย จะคุยในนี้ และ sticker คือน่ารัก...	positive

FIGURE-5.8: RESULTS FOR #IPHONSE

Sample result for the tweet #iphonese

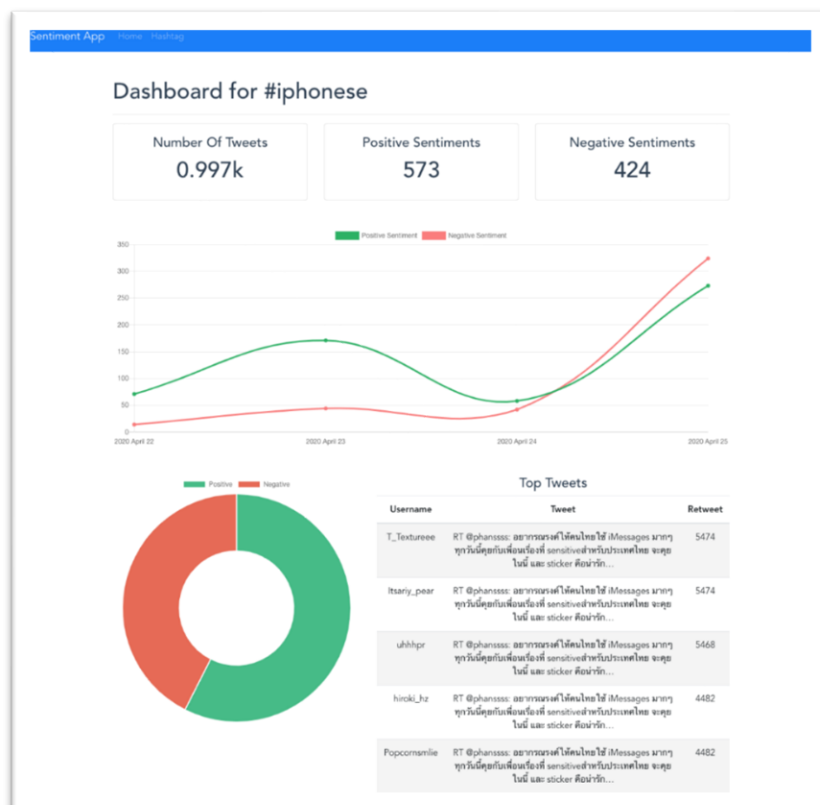


FIGURE-5.9: STATISTICS FOR THE TWEET

Stats. Panel for the user to interact

5.2 CODING STANDARDS

- Use of CamelCase for naming of functions.
- Assets Management Folders are separately maintained for Prefabs, Scripts, 3D models, Scenes etc.
- Reusable codes The scripting in the modules and functions are done in such a way that it can be REUSED in different scripts also.
- Comments are added wherever required, for easy understanding

	Part of Class/Interface Declaration Notes	Notes
1	Class (static) variables	First the public class variables if any, then the protected, then package level (no access modifier), and then the private.
2	Instance variables	First public, then protected, then package level (no access modifier), and then private
3	Class (static) methods	First the public, then the protected, then package level (no access modifier), and then the privat
4	Class Methods	First public, then protected, then package level (no access modifier), and then private.

TABLE-5.1

In this project, we have used the white box testing because the source code is completely developed by us and available to us. We have chosen this techniques because of following reasons:

- Easy to find out errors

- Make it efficient for use Optimize the code

5.2.2 Advantages:

- As the tester has knowledge of the source code, it becomes very easy to find out which type of data can help in testing the application effectively.
- It helps in optimizing the code.
- Extra lines of code can be removed which can bring in hidden defects.
- Due to the tester's knowledge about the code, maximum coverage is attained during test scenario writing.

5.3 OPERATING ENVIRONMENT

- The basic hardware requirement is to have a processor of 2.1 GHz and recommended RAM of 4 GB for smooth processing. RAM size less than 4 GB will also do the job but patience can be tested. For storage options, less than 1 GB of hard disk space will be required in case if one opts to download the report. The app is compatible with all the major operating systems like Windows, Linus, Chrome OS, MAC OS etc.
- For windows, it is recommended to have at least windows 8.1 for enhanced experience.
- For Linux, ubuntu linux or kali linux is recommended.
- Web browser like Chrome or Edge will be required.
- The app is compatible with all the web browsers except the internet explorer

5.4 RISK ANALYSIS AND MANAGEMENT

A risk management process contains 4 major steps.

- **Risk identification** - Identify project, product and business risks
- **Risk analysis** - Assess the likelihood and consequences of these risks
- **Risk planning** - Draw up plans to avoid or minimise the effects of the risk
- **Risk monitoring** - Monitor the risks throughout the project.

5.4.1 RISK INFORMATION:

Project	EDA Application
Risk Type	Operational Risk
Priority	3
Risk Factor	Further Project Execution will depend on the functioning of the application. The application may not work for certain datasets
Probability	30%
Impact	Project completion will be delayed if certain datasets that needs to be processed will cause the stoppage error.
Monitoring Approach	To check the population set of the data sets to be used in the project in the testing phase in order to find and rectify the problem.
Contingency Plan	Modification in the testing strategy and to change the deadlines of the project in order to get sufficient time for the damage control
Estimated Resources	2 additional data analysts to be added by next week for further guidance in both development and testing.

TABLE-5.2

5.4.2 RISK IDENTIFICATION:

According to my point of view, in the risk identification process, certain risk that our application has are as follows:

- **Technical Disadvantage:**

The technology that is being used/provided by the application may become obsolete before launching it or in a short time after the launch.

- **Development Environment:**

There is high probability that, change in development flow or method due to some miscellaneous reasons can affect the application.

For example, if a developer leaves the project and back up is not ready then, the development of the application might get delayed.

The situation may get worse if lead developer leaves the project.

- **Business Impact:**

The market already have the similar applications which can give tremendous competition and impact the brand value.

- **Application not meeting the expectation**

This is the most common risk that every project possesses, that whether their application will satisfy the needs of the end user or will it fail once launched.

5.4.3 Risk Projection

RISK	CATEGORY	PROBABILITY	IMPACT
Technical Disadvantage	TR	30%	4
Development Environment	PR	40%	8
Business Impact	BR	30%	5
Application not meeting the expectation	BR	35%	9

TABLE-5.3

TR- Technical Risk

PR- Project Risk

BR- Business Risk

5.4.4 Risk Planning:

- An effective strategy for dealing with risk must consider three issues
 - Risk mitigation (i.e., avoidance)
 - Risk monitoring
 - Risk management and contingency planning

Risk mitigation (avoidance) is the primary strategy and is achieved through a plan. As we have identified the risks, so, we will be planning to either avoid, monitor, or will try to eliminate the risk. Here, we can divide the risk into multiple sub lists. For example, for the risk of technical disadvantage, we can further more go into details and can try to avoid the risk by segregating the technologies that can cause the risk. For Development risk, we can have a back up plan so risk can be managed and the impact can be reduced. Similarly, for the business impact, we can avoid the risk by adding some unique features in our application which the rivals are not offering.

5.4.5 Risk monitoring

We can monitor the development risk by following a specific SDLC model like agile. For the technical disadvantage, we can hire some versatile developers who can adapt very quickly to the changing tech and can help in reducing the risk.

5.5 TEST SUITE DESIGN

If each test case represents a piece of a scenario, such as the elements that simulate a completing a transaction, use a test suite.

Test suites can identify gaps in a testing effort where the successful completion of one test case must occur before you begin the next test case. When you run a test suite in sequential mode, you can choose to stop the suite execution if a single test case does not pass. Stopping the execution is useful if running a test case in a test suite depends on the success of previous test cases.

Test suites are also useful for the following types of tests –

Build verification tests: A collection of test cases that perform a basic validation of most the functional areas in the product. The tests are executed after each product build and before the build is promoted for use by a larger audience.

Smoke tests: A collection of test cases that ensure basic product functionality. Typically, smoke tests are the first level of testing that is performed after changes are made to the system under test.

End-to-End integration tests: A collection of test cases that cross product boundaries and ensure that the integration points between products are exercised and validated.

Functional verification tests: A collection of test cases that focus on a specific product function. Executing this type of test with a test suite ensures that several aspects of a specific feature are tested.

5.6 TEST CASES

A test case, in software engineering, is a set of conditions or variables under which a tester will determine whether an application, software system or one of its features is working as it was originally established for it to do. The mechanism for determining whether a software program or system has passed or failed such a test is known as a test oracle. In some settings, an oracle could be a requirement or use case, while in others it could be a heuristic. It may take many test cases to determine that a software program or system is considered sufficiently scrutinized to be released. Test cases are often referred to as test scripts, particularly when written - when they are usually collected into test suites.

TEST CASE ID : T01

Module Name : Working with the browsers

Pre-condition : The website of EDA should open in different browsers

Test id	Test case	Expected outcome	Actual outcome	Pass/Fail
1	EDA APP opening in different browser	The app should be opening in chrome and edge	The app opened in chrome and edge	pass

TABLE-5.4

Test Case ID: T02

Module Name : EDA should generate report for uncleaned data

Test id	Test case	Expected outcome	Actual outcome	Pass/Fail
1	EDA app giving results in chrome	EDA app should give results	EDA app gives result	PASS
2	EDA app giving results in edge	EDA app should give results	EDA app gives result	PASS

TABLE-5.5

TEST CASE ID : T03

Module Name : Working with the browsers

Pre-condition : The website of sentiment analysis should open in different browsers

Test id	Test case	Expected outcome	Actual outcome	Pass/Fail
1	website opening in different browser	The website should be opening in chrome and edge	The website opened in chrome and edge	pass

TABLE-5.6

CHAPTER 6:

CONCLUSION

6.1 CONCLUSION

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information. we have completed the project work using software engineering and system analysis and design approach. This project is completed with the primary functionalities. Due to lack of skilled knowledge and time constraints, the project cannot be fully completed so far. we have created an experience which performs the basic functionalities but yet there is a scope of improvement and advancement which can be taken care of in the near future.

6.2 DIFFICULTIES FACED

Lack of full knowledge about the domain or technology:

This domain was new for all of us and neither of us had a past experience for this domain. So, the problem of where to start and what to start was pretty evident and also, and as we went deep down in the project, it became complex and difficult to manage.

Lack of proper coordination:

When the group was formed and the project was finalized, none of us were having knowledge of the domain, so a chaotic scenario was observed where everyone started to wait for other team member to take the charge, and ultimately the issue was resolved but it was difficult.

WORK CITED/ BIBLIOGRAPHY:

1. <https://www.python.org/>
2. <https://www.analyticsvidhya.com/blog/2021/05/exploratory-data-analysis-eda-a-step-by-step-guide/>
3. <https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/>
4. <https://hypefactors.com/blog/limitations-of-sentiment-analysis-for-reputation-management/>
5. <https://v2.vuejs.org/v2/guide/?redirect=true>
6. <https://code.visualstudio.com/docs>
7. <https://flask.palletsprojects.com/en/2.1.x/>
8. https://en.wikipedia.org/wiki/Exploratory_data_analysis
9. <https://www.journaldev.com/53190/exploratory-data-analysis-python#:~:text=Exploratory%20Data%20Analysis%20%E2%80%93%20EDA,or%20through%20some%20python%20functions.>
10. <https://monkeylearn.com/sentiment-analysis/>
11. <https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>