

SENTIMENT ANALYSIS & DATA VISUALIZATION

Submitted in partial fulfillment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

Somiya Panikar (20102095)

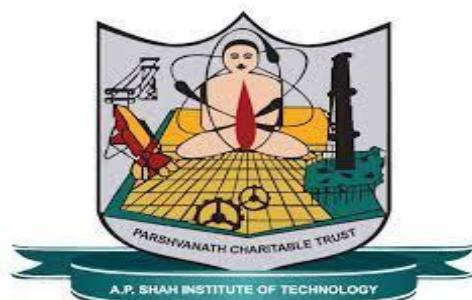
Mitali Mohite (20102119)

Parth Mishra (20102162)

Shubham Mojdra (20102108)

Guide

Prof. Suchita Dange



**Department of Computer Engineering
A.P. Shah Institute of Technology, Thane**

Academic Year 2022-2023



Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

CERTIFICATE

This is to certify that the Mini Project 1B entitled "**Sentiment Analysis & Data Visualization**" is a Bonafede work of "**Somiya Panikar (20102095)**", "**Mitali Mohite (20102119)**", "**Shubham Mojidra (20102108)**", "**Parth Mishra (20102162)**" submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering in Computer Engineering**

Guide

Prof. Suchita Dange

Project Coordinator

Prof. Deepak Khachane

Head Of Department

Prof. Sachin Malave

Date:



Parshvanath Charitable Trust's
A. P. SHAH INSTITUTE OF TECHNOLOGY
(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

Project Report Approval for Mini Project-2A

This project report entitled “*Sentiment Analysis & Data Visualization*” by “*Somiya Panikar*”, “*Mitali Mohite*”, “*Shubham Mojidra*”, “*Parth Mishra*” is approved for the degree of *Bachelor of Engineering in Computer Engineering, 2022-23.*

Examiner Name

1. _____

2. _____

Signature

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Somiya Panikar (20102095)

Mitali Mohite (20102119)

Shubham Mojdra (20102108)

Parth Mishra (20102162)

Date:

Abstract

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities. Whenever we need to make a decision, we want to hear others' opinions. Second, it presents many challenges in research problems. Sentiment analysis is the procedure by which information is extracted from the opinions and emotions of people in regards to entities, events. In decision making, the opinions of others have a significant effect on the company. Our major task was to take the audit report from the user in pdf format, do segmentation of the pdf, convert it into a desirable csv format for the model, do sentiment analysis on it and then visualize the result using various charts. Our project's goal was to make data analysis easier and better visualized. Our project model's accuracy is quite good. As a result, we were able to identify the polarities, forecast the audit's mood, and create a stunning visualization of it.

Keywords - *sentiment analysis, computational, predict*

Contents

1.	Introduction	1
2.	Literature Survey	2
3.	Problem Statement.....	4
4.	Objective & scope.....	5
5.	Experimental Setup.....	6
	 5.1 Hardware requirement	
	 5.2 Software requirement	
6.	System Design.....	7
	 6.1 UML Diagram	
	 6.2 Algorithm/Process (with Expected input and outputs)	
7.	Implementation.....	12
	 7.1 Code	
8.	Result.....	16
	 8.1 Design and Screenshots	
9.	Conclusion.....	24
10.	References.....	25
11.	Annexure 1- Project Planning (Using Gantt chart)	26
12.	Acknowledgement	27

List of figures

6.1.1 UML Diagram

Fig 6.1.1 The UML/ Use Case diagram of our website..... 15

6.1.2 Architectural Diagram

Fig 6.1.2 The architectural diagram of our website..... 16

6.1.3 DFD

Fig 6.1.3.1 The Zero Level DFD of our website..... 16

Fig 6.1.3.2 The First Level DFD our website..... 17

Fig 6.1.3.3 The Second Level DFD of our website..... 17

7.0 Gantt Chart

Fig 7.0 Gantt Chart..... 27

Chapter 1

Introduction

Large banks can use AI-based sentiment analysis software to gauge customer opinions about their brand or their products in an attempt to improve customer experiences. Dataset we get is in raw format, there's no such eye pleasing insights for data scientist. As a result, we hope that our project (Sentiment Analysis & Data Visualization) would assist our users (customers) by giving them a thorough analysis of the data they have provided in the form of audit reports. The system's interface will display the analyzed data as graphs and charts, which will make it simpler to interpret the conclusions. The previously examined reports that will be kept in the database of our system will allow our user to view their progress. The product is intended for companies who require an audit of the company's financial statements for analysis purposes, such as banks, creditors, and regulators. The product is supposedly open source. It is a web-based application that employs a data analytics methodology. The Sentiment analysis and Data visualization system provide customers with an easy way to examine their documents. Users (companies) will also be able to examine all previous analysis and visualization on a single platform.

Chapter 2

Literature Survey

[1] Sentiment analysis is currently the most popular study area under the Natural Language Processing umbrella (NLP). The main goal of this research issue is to identify the feelings and opinions of the clients or users through text. Despite the fact that a lot of research has been done in this area using a variety of models, sentiment analysis is still regarded as a difficult subject with a lot of conflicts that need to be resolved. The use of slang terms, new accents, grammar and spelling errors, etc. are some of the current difficulties. With diverse types of data and various machine learning algorithms, this work aims to conduct a review of the literature. Sentiment analysis is a technique that document analysts have used for summarizing or modifying certain knowledge texts. Incorporating NLP, statistical methodologies, and machine learning approaches, sentiment analysis is frequently used to mine subjective data from internet material. Sentiment analysis has a wide range of applications but its major focus is on large-scale text analysis rather than manual analysis. In order to harvest social data and make the business process profitable, opinion mining technology is commonly used. In order to demonstrate how problems have been solved up to this point and the many algorithms employed in sentiment analysis, this overview highlights the research that has been done in the field over the past ten years.

[2] Social media is a developing phenomenon that is having a significant impact on society. Sites like Facebook, Twitter, LinkedIn, and Instagram each have unique features and applications. For instance, Facebook is regarded as a social network where each member has a reciprocal relationship with every other member of the network. In this instance, the relationship is undirected. On the other hand, no one in the Twitter network is required to have a reciprocal relationship with anyone else. The relationship is either directed or undirected in this situation.

For data analysis in this review of the literature, we concentrate on Twitter, an online networking site that lets users send and read 140-character messages known as "tweets". In contrast to Facebook, where users may regulate the privacy of their profiles, Twitter allows unregistered users to read and monitor the majority of tweets in addition to providing publicity. Another significant social networking and microblogging platform is Twitter. The wealth of data offered by Twitter, including tweets, user profiles, and the number of followers and followings in the network, is crucial for data analysis. As a result, most studies look into and compare different data analysis methods to understand the most recent technologies. Similar concept can be applied to analyze business intelligence.

[3] The dissemination of information to potential readers has always been a problem for technical and professional communicators. Additional difficulties are brought on by complex information from technological fields like science, medicine, and the environment. Because we are constantly exposed to new information in our daily lives, pictures are frequently the most effective form of communication. But how can we use data visualizations to explain complicated concepts to audiences who aren't experts? Data visualization is the use of visuals to display huge amounts of data in accordance with specific criteria or groups, such as the aggregation of data into graphs, charts, scatter plots, or other common visualization styles. As such, it becomes important for the field to understand the landscape of current research. We have made an effort to apply visualization in the financial industry in light of this literature assessment.

Chapter 3

Problem Statement

“It becomes difficult for data analyst to go through large amount of data” :

In an effort to enhance customer experiences, large banks can use AI-based sentiment analysis tools to measure client opinions about their brand or their products. The dataset we receive is in raw format; data scientists cannot access such aesthetically pleasing insights. The main objectives of any business are to increase revenue and keep clients.

“Understanding customer’s views and emotions becomes difficult which can affect the company’s progress” :

You can use sentiment analysis to learn how many people are feeling about your brand or product. Simply said, there is usually too much data for this to be done manually. For enterprises to acquire deeper insights, specialized SaaS technologies have made it easier. Everything from employee surveys and social media posts to consumer feedback could be included in this. These sources' sentiment data can be utilized to guide important business decisions.

“Lack of proper analysis and estimation leads to customer turnovers” :

Use a corporation that just introduced a new product as an example. The business can feed the information into a feedback management tool rather than sifting through hundreds of evaluations. It will categorize incoming feedback using a sentiment analysis algorithm. The business can react appropriately and learn more quickly what customers think of their new offering. Customers' favorite aspects and potential improvement areas can be found.

Companies can estimate how many customers feel a particular way about their product using this kind of study. A company's priorities can be influenced by the quantity of respondents and the general polarity of the opinion about, say, "online documentation." To prevent client turnover and maintain their competitiveness, businesses should, for instance, concentrate on producing better documentation.

Chapter 4

Objective and Scope

- The goal is to provide a platform that simplifies the process of analyzing and visualizing a bank's audit report.
- To develop a user-friendly but efficient sentiment analysis and data visualization platform.
- The developed system's main goal is to take bank audit reports as input and analyze them using a ML model.
- Following analysis, the system will visualize the results in graphs and charts using tableau or Py Plotly package.

When it comes to measuring attitudes, feelings, and emotions towards their brand, organizations can measure them using sentiment analysis, which is a particularly potent tool. As of now, practically all sentiment analysis initiatives have been carried out by businesses and brands using information from social media, survey replies, and other sources of user-generated material. But in our system, sentiment analysis will be done on the bank's auditor report. The auditor's report expresses the auditor's judgment regarding the accuracy and conformity with GAAP of a company's financial statements. Because banks, creditors, and regulators need an audit of a company's financial statements, the audit report is crucial. To make data easier for the human brain to grasp and draw conclusions from, data visualization is the practice of putting information into a visual context, like a map or graph. Data visualization's major objective is to make it simpler to spot patterns, trends, and outliers in big data sets. We will use graphs and charts to visually represent the output from our ML model. Banks may easily examine and visualize their financial audit results utilizing our solution.

Chapter 5

Experimental Setup

5.1 Hardware Requirement

- Processor- Intel Core i5
- RAM – 8GB
- Hard Disk – 1TB
- Display Card –GTX 1660 TI
- Mouse – Logitech Serial Mouse
- Keyboard – Standard 104 Enhanced Keyboard

5.2 Software Requirement

- Web Server – Heroku
- Browser – Microsoft Edge, Google Chrome
- Server-side scripting – Python Django
- Libraries –
- Database – SQLite
- Language –HTML, CSS, JS(BOOTSTRAP)

Chapter 6

System Design

6.1.1 UML Diagram

It is a structural diagram that depicts a static view or structure of a system. It presents an outline for the system. It stresses the elements to be present that are to be modeled. Through this we can also show the flow between modules, and how the data flows between them.

What can we achieve through this?

- Improved integration between structural models like class diagrams and behavior models like activity diagrams.
- Added the ability to define a hierarchy and decompose a software system into components and subcomponents.

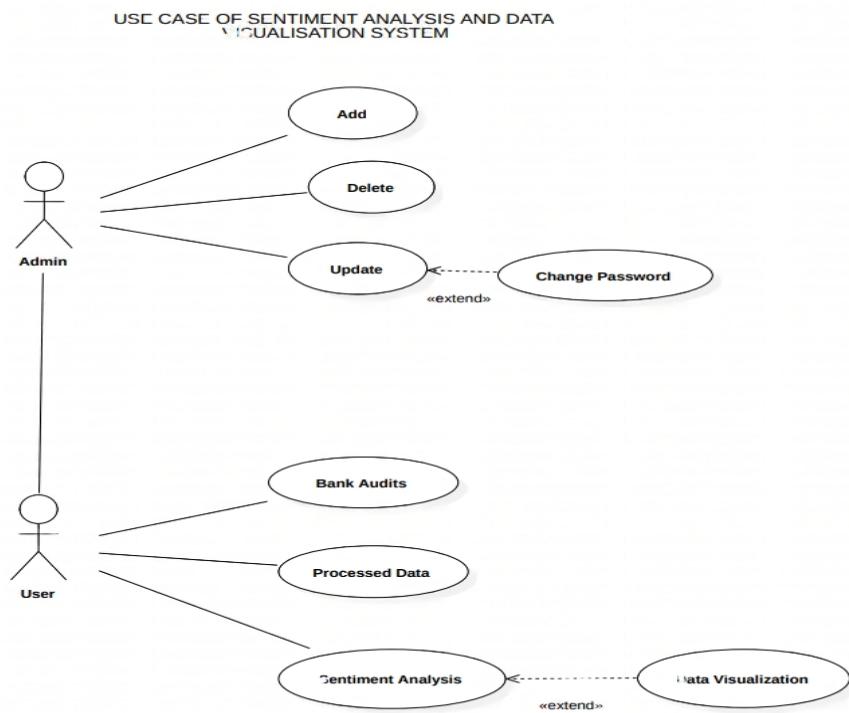


Fig 6.1.1 The UML/ Use Case diagram of our website

6.1.2 Architectural Diagram

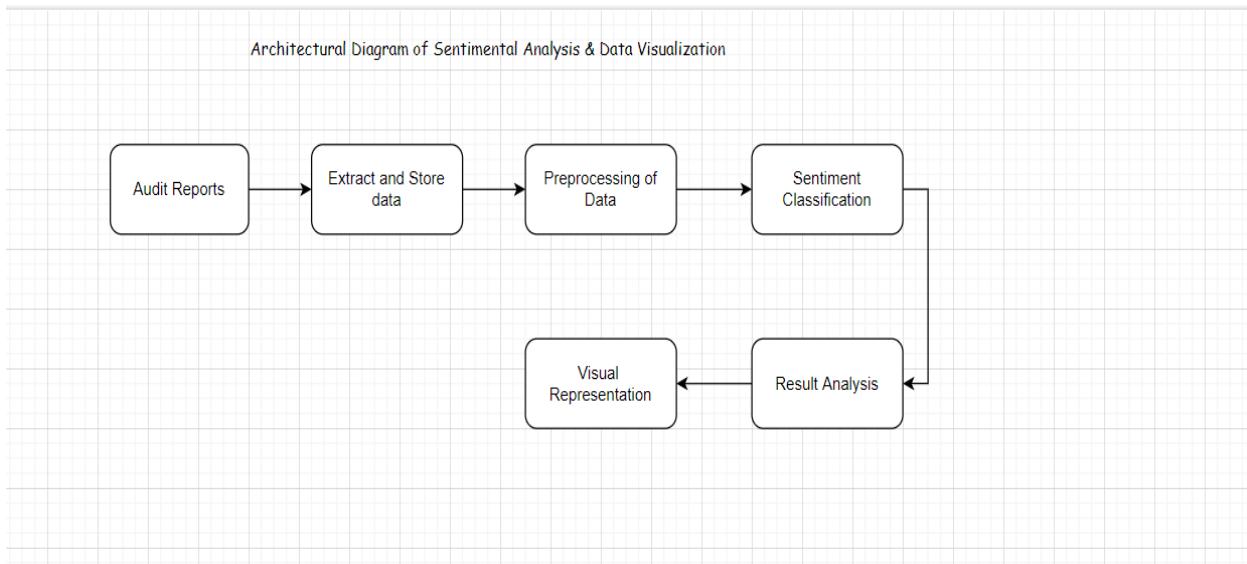


Fig 6.1.2 The architectural diagram of our website

6.1.3 DFD

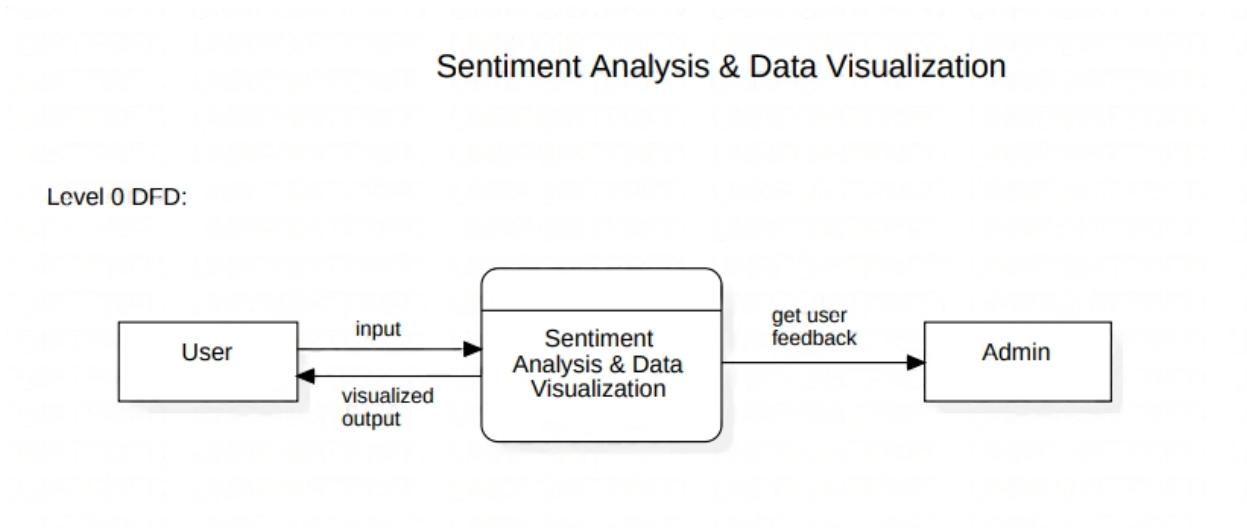


Fig 6.1.3.1 The Zero Level DFD of our website

System Architecture and Information Engineering

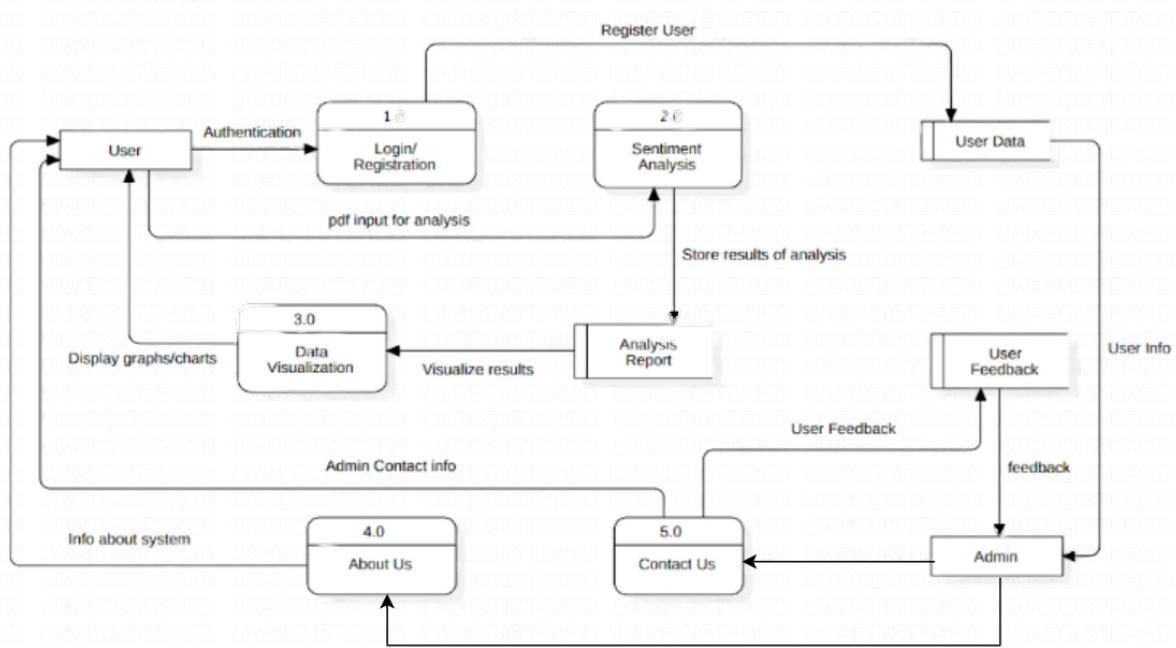


Fig 6.1.3.2 The First Level DFD our website

Level 2 DFD of Sentiment Analysis:

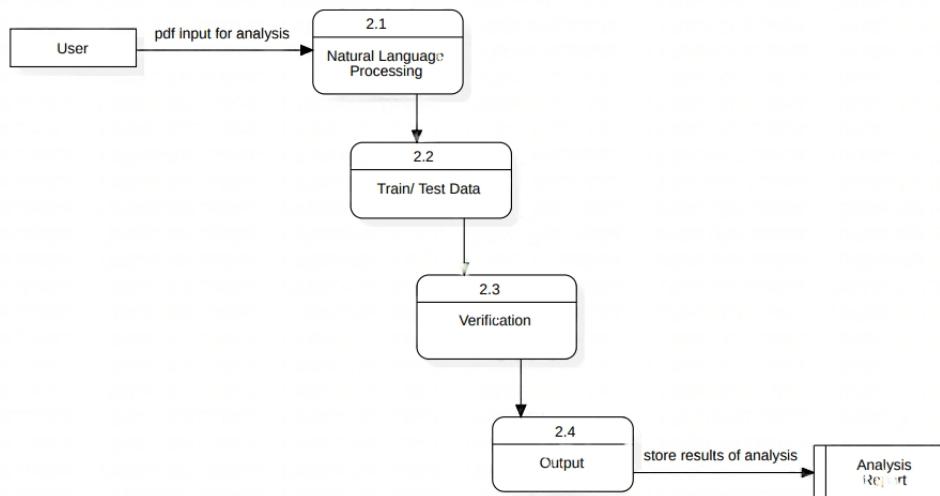


Fig 6.1.3.3 The Second Level DFD our website

6.2 Algorithm/Process (with Expected input and outputs)

Modules of System

The project is divided into various modules to help the work divide into small yet efficient modules. Each module's output acts as an input to the next module. Each module requires a specific set of knowledge and skill which takes time hence they have been divided accordingly.

6.2.1 Home Page

It contains basic homepage tabs like contact us, Sign Up etc. That are easy to use for user. It will also contain the single sentence sentiment analysis as a demo for the users. For visualization part the user will have to Log In.

6.2.2 Contact Module

In this module, it will allow the users to contact the site developers through email. So it will take the name, email address and the message from the users.

6.2.3 Login/ Register Module

Through this module, the user will be able to sign in and create an account. When users access the system through Portal Direct Entry, they're considered guests until they log in. The Login Module is a portal module that allows users to type a username and password to log in. The visualize feature is visible only if the user is logged in or authenticated.

6.2.4 About Us Module

In this module, it will contain all the contact information related to our website. So it will include email id, social networking site details, developers name and details, etc.

6.2.5 Visualize Module

In this module, it will allow the users to input an audit report in pdf format, do sentiment analysis on it and visualize it. This module will visualize the result using various charts like Bar chart, Pie Chart, WordCloud.

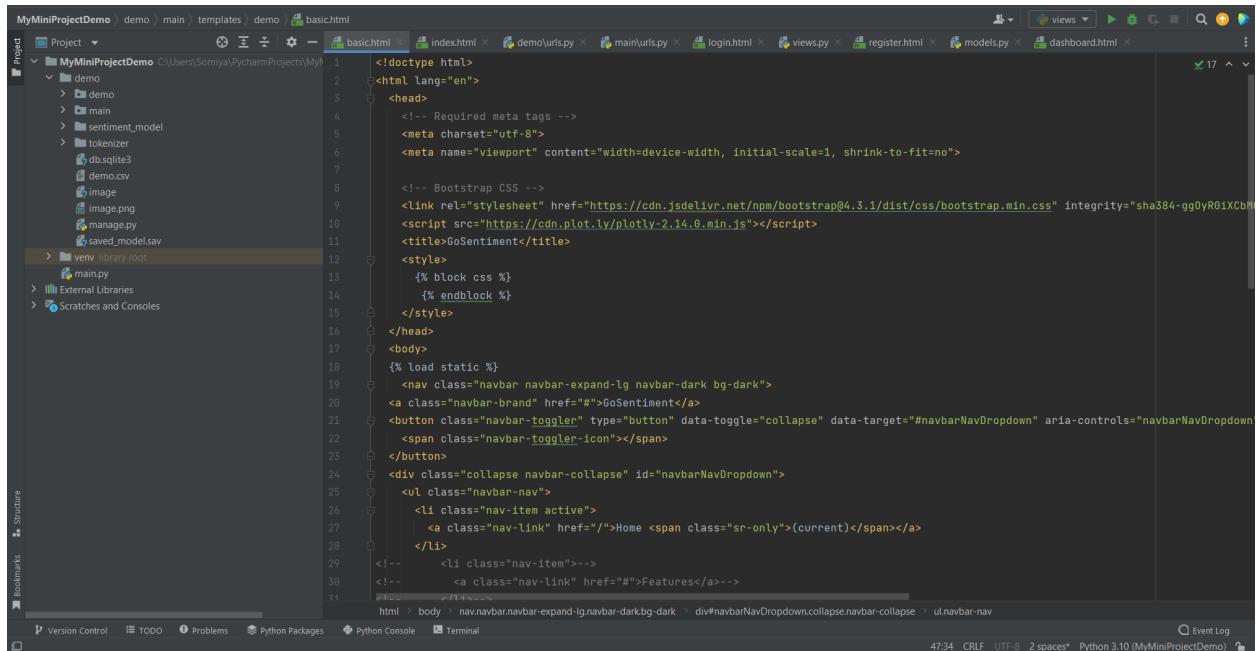
6.2.6 Gunning Fog Module

This module provides users with a button to submit audit reports in pdf format. The analysis and various relevant scores/ratings are displayed along with different types of graphs for easy comprehension. This module gives us an overview of the report. Using this we conclude the difficulty level of apprehending the audit and make changes accordingly.

Chapter 7

Implementation

7.1 Code



The screenshot shows the PyCharm IDE interface with the 'basic.html' file open in the editor. The code is an HTML template for a navigation bar. It includes Bootstrap CSS imports, a title, and a navbar with active links for 'Home' and 'Features'. The PyCharm interface includes a project tree on the left, toolbars at the top, and various status indicators at the bottom.

```
<!doctype html>
<html lang="en">
  <head>
    <!-- Required meta tags -->
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

    <!-- Bootstrap CSS -->
    <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@4.3.1/dist/css/bootstrap.min.css" integrity="sha384-gg0yR8iXcbM...
    <script src="https://cdn.plot.ly/plotly-2.14.0.min.js"></script>
    <title>GoSentiment</title>
    <style>
      {% block css %}
      {% endblock %}
    </style>
  </head>
  <body>
    {% load static %}
    <nav class="navbar navbar-expand-lg navbar-dark bg-dark">
      <a class="navbar-brand" href="#">GoSentiment</a>
      <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarNavDropdown" aria-controls="navbarNavDropdown" aria-expanded="false" aria-label="Toggle navigation">
        <span class="navbar-toggler-icon"></span>
      </button>
      <div class="collapse navbar-collapse" id="navbarNavDropdown">
        <ul class="navbar-nav">
          <li class="nav-item active">
            <a class="nav-link" href="/">Home <span class="sr-only">(current)</span></a>
          </li>
          <!-->
          <li class="nav-item">
            <a class="nav-link" href="#">Features</a>
          </li>
        </ul>
      </div>
    </nav>
  </body>
</html>
```

The screenshot shows the PyCharm IDE interface with the project 'MyMiniProjectDemo' open. The current file is `urls.py` located in the `demo` directory. The code defines URL patterns for various views including admin, predict, submit, dashboard, and contact. A PEP 8 warning 'E231 missing whitespace after ','' is visible in the status bar.

```
1. Import the include() function: from django.urls import include, path
2. Add a URL to urlpatterns: path('blog/', include('blog.urls'))



from django.contrib import admin
from django.urls import path, include
from . import views

urlpatterns = [
    path('admin/', admin.site.urls),
    path('', include('main.urls')),
    path('predict', include('main.urls')),
    path('submit', include('main.urls')),
    path('dashboard', include('main.urls')),
    # path('inputcsv', include('main.urls')),
    path('load_data', include('main.urls')),
    path('loginpage', include('main.urls')),
    path('logoutpage', include('main.urls')),
    path('register', include('main.urls')),
    # path('dataset', include('main.urls')),
    path('contact', include('main.urls')),
    path('aboutus', include('main.urls')),
    #path('dashboard', views.dashboard, name='dashboard'),
]
```

The screenshot shows the PyCharm IDE interface with the project 'MyMiniProjectDemo' open. The current file is `views.py` located in the `main` directory. The code imports various Python libraries and defines a `load_dataset` function that reads a pickle file named `saved_classifier`.

```
import json
import pickle

import fitz
import plotly
from django.http import HttpResponseRedirect
from django.shortcuts import render, redirect
from textblob import TextBlob
from google.transliteration import transliterate_text
from deep_translator import GoogleTranslator
from django.conf import settings
from django.core.mail import send_mail
import plotly.express as px
import plotly.graph_objs as go
import pandas as pd
import re
from wordcloud import WordCloud, STOPWORDS
from collections import Counter
import matplotlib.pyplot as plt
from django.contrib.auth.models import User
from django.contrib.auth import authenticate, login, logout

wc_data = []
input_df = None
fname = None
filename = 'saved_model.sav'
saved_classifier = pickle.load(open(filename, 'rb'))

def load_dataset(dataset):
    n_lines = 1
    sentiment_by_words()
```

```

MyMiniProjectDemo > demo > main > views.py
Project  basic.html  index.html  demourls.py  mainurls.py  login.html  views.py  register.html  models.py  dashboard.html
MyMiniProjectDemo > demo > main > sentiment_model > tokenizer > db.sqlite3 > demo.csv > image > image.png > manage.py > saved_models.sav > venv library root > main.py > External Libraries > Scratches and Consoles
46     return user_df
47
48
49     def clean_text(unformatted_text):
50         unformatted_text = str(unformatted_text)
51         unformatted_text = re.sub(r'@[A-Za-z0-9]*', ' ', unformatted_text)
52         unformatted_text = re.sub(r'#[A-Za-z0-9]*', ' ', unformatted_text)
53         unformatted_text = re.sub(r'&', ' ', unformatted_text)
54         unformatted_text = re.sub(r'\n', ' ', unformatted_text)
55         unformatted_text = re.sub(r'\r', ' ', unformatted_text)
56         unformatted_text = re.sub(r'\t', ' ', unformatted_text)
57         unformatted_text = re.sub(r'\?', ' ', unformatted_text)
58         unformatted_text = re.sub(r'\!', ' ', unformatted_text)
59         unformatted_text = re.sub(r'\.', ' ', unformatted_text)
60         unformatted_text = re.sub(r'\+', ' ', unformatted_text)
61         unformatted_text = re.sub(r'\https?:\/\/[\S]+', ' ', unformatted_text)
62
63     return unformatted_text.lower()
64
65     def count_plot(x, y):
66         fig = go.Figure()
67         layout = go.Layout(
68             title='Multiple Reviews Analysis',
69             xaxis=dict(title='Category'),
70             yaxis=dict(title='Count')
71         )
72         fig.update_layout(dict=layout, overwrite=True)
73         fig.layout.template = "plotly_dark"
74         fig.add_trace(go.Bar(name='Multi Reviews', x=x, y=y))
75         graph1 = json.dumps(fig, cls=plotly.utils.PlotlyJSONEncoder)
76
77         return graph1
78
79     sentiment_by_words()

```



```

MyMiniProjectDemo > demo > main > views.py
Project  basic.html  index.html  demourls.py  mainurls.py  login.html  views.py  register.html  models.py  dashboard.html
MyMiniProjectDemo > demo > main > sentiment_model > tokenizer > db.sqlite3 > demo.csv > image > image.png > manage.py > saved_models.sav > venv library root > main.py > External Libraries > Scratches and Consoles
190     return HttpResponse('Mail Sent Successfully')
191
192
193     def load_data(request):
194         global wc_data, input_df, fname
195         count_positive, count_negative, count_neutral = 0, 0, 0
196
197         if request.method == "POST":
198             try:
199                 csvget = request.FILES["csv_files"]
200                 fname = csvget.name
201                 user_dataframe = load_dataset(csvget)
202                 input_df = user_dataframe.copy()
203                 wc_data = []
204
205             except:
206                 return render(request, "demo/dashboard.html")
207             opt = "Sentences"
208             if input_df is not None:
209                 input_df1 = input_df.head(5)
210                 data_html = input_df1.to_html()
211                 list_of_columns = list(input_df.columns)
212
213                 if opt:
214                     temp_df = input_df[opt]
215                     input_df['sentiment'] = ''
216                     input_df[opt] = input_df[opt].apply(clean_text)
217                     for i in range(temp_df.shape[0]):
218                         wc_data += str(temp_df.iloc[i]).split()
219                         text = TextBlob(str(temp_df.iloc[i]))
220                         result = text.sentiment.polarity
221                         if result >= 0.1:
222                             count_positive += 1
223                             text.sentiment = 'Positive'
224
225             sentiment_by_words()

```

The screenshot shows the PyCharm IDE interface with the following details:

- Project Tree:** The project is named "MyMiniProjectDemo" located at "C:\Users\Somya\PycharmProjects\MyMiniProjectDemo". It contains several packages and files: demo, main, tokenizer, db.sqlite3, demo.csv, image, image.png, manage.py, saved_model.sav, venv (library root), main.py, and External Libraries.
- Code Editor:** The file "views.py" is open. The code implements two views: `loginpage` and `register`.
 - `loginpage` handles POST requests to log in users by authenticating them and redirecting to the home page if successful, or rendering the login template if unsuccessful.
 - `register` handles POST requests to create new users by extracting form data, creating a User object, and saving it to the database.
- Status Bar:** The status bar at the bottom shows the current file is "views.py", the line count is 276, and the Python version is 3.10 (MyMiniProjectDemo).

Chapter 8

Result

The system acts as an effective tool to analyze the financial reports and give us analytics and insights into the businesses. It helps us to understand whether the company is in loss or in profit based on the auditor's opinions in the financial reports.

With inputs from various existing systems, and with addition of our own datasets, a functional ML model is implemented to give us usable results.

Understanding is further deepened and made convenient with the help of various plots. Interactive graphs can help people draw meaningful and fruitful conclusions.

Various scores are calculated, giving users an indication of how difficult the audits are to read and grasp. Scores also suggest the educational qualification one may require to understand the audits with ease.

8.1 Design and Screenshots

8.1.0 Home Page

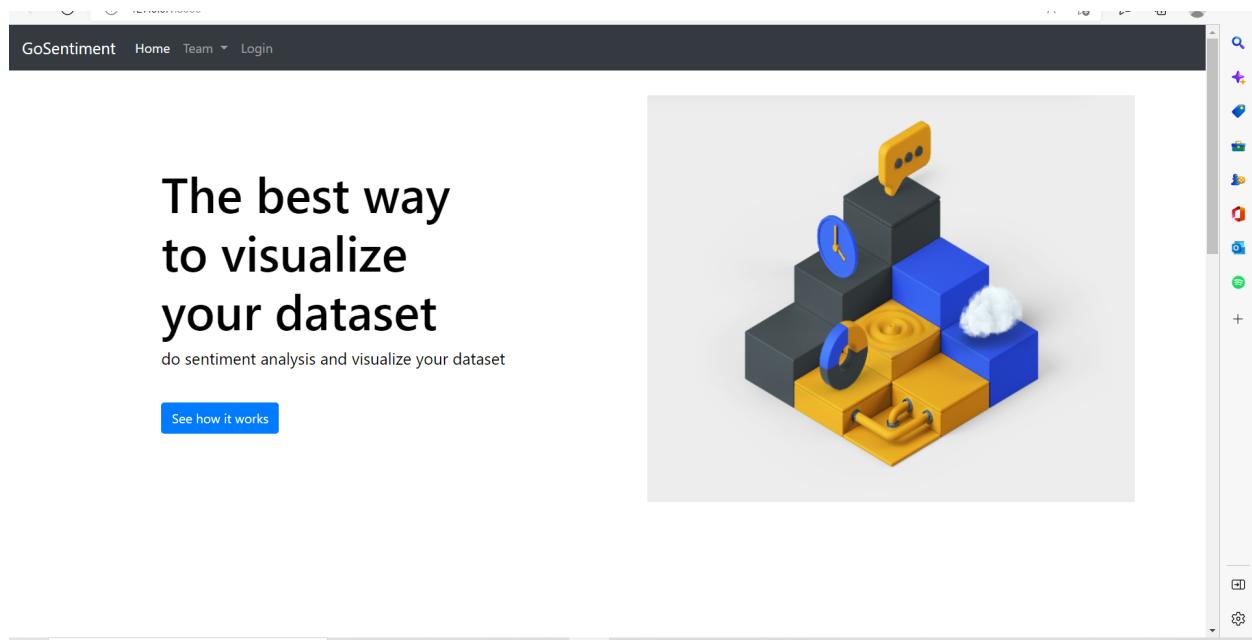


Fig 8.1.0 The Main or Home page of our website,i.e, the page that will be visible to users as soon as they enter our site

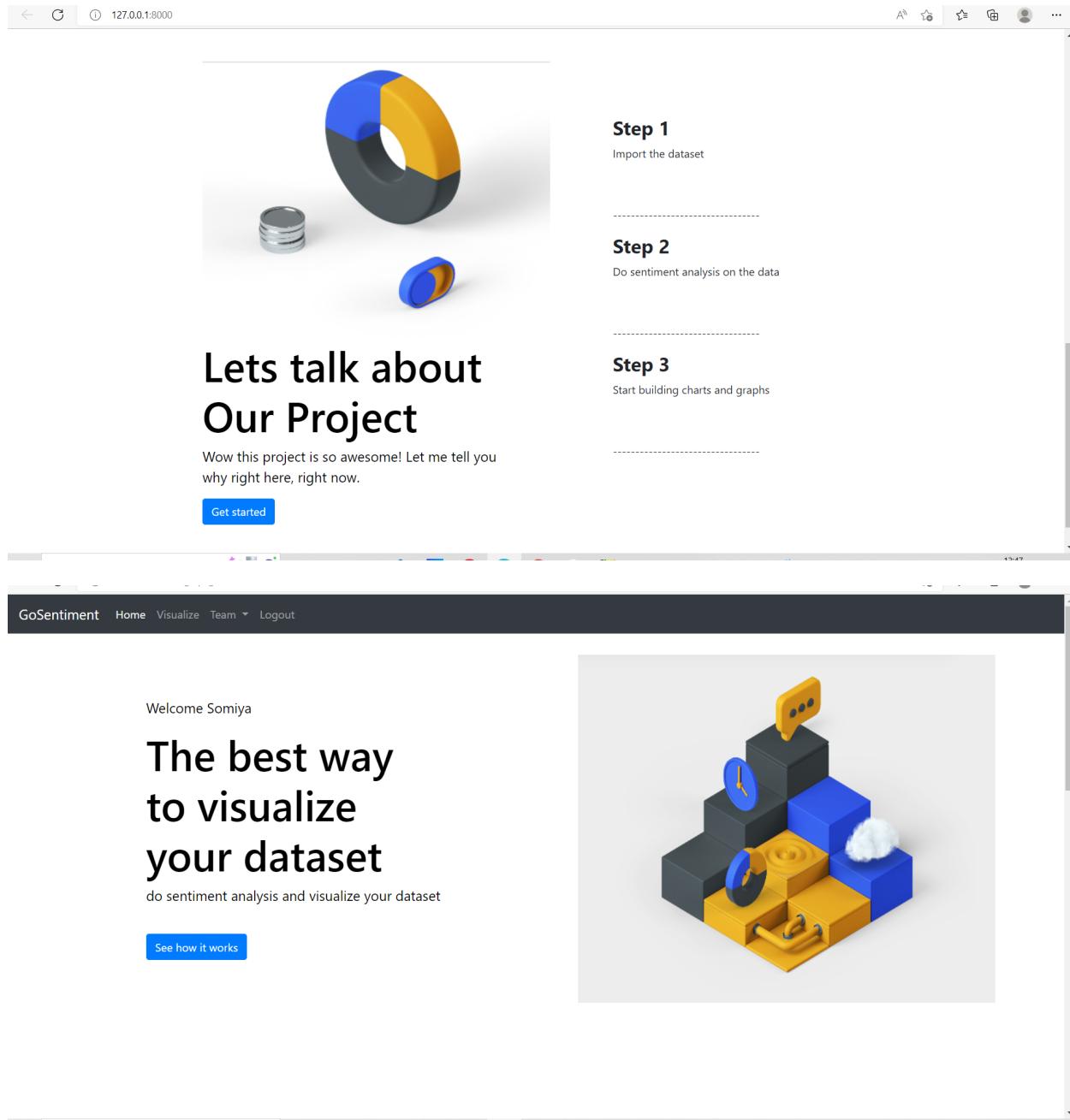


Fig 8.1.0 The Main or Home page of our website,i.e, the page that will be visible to users as soon as they login

8.1.1 Contact Module

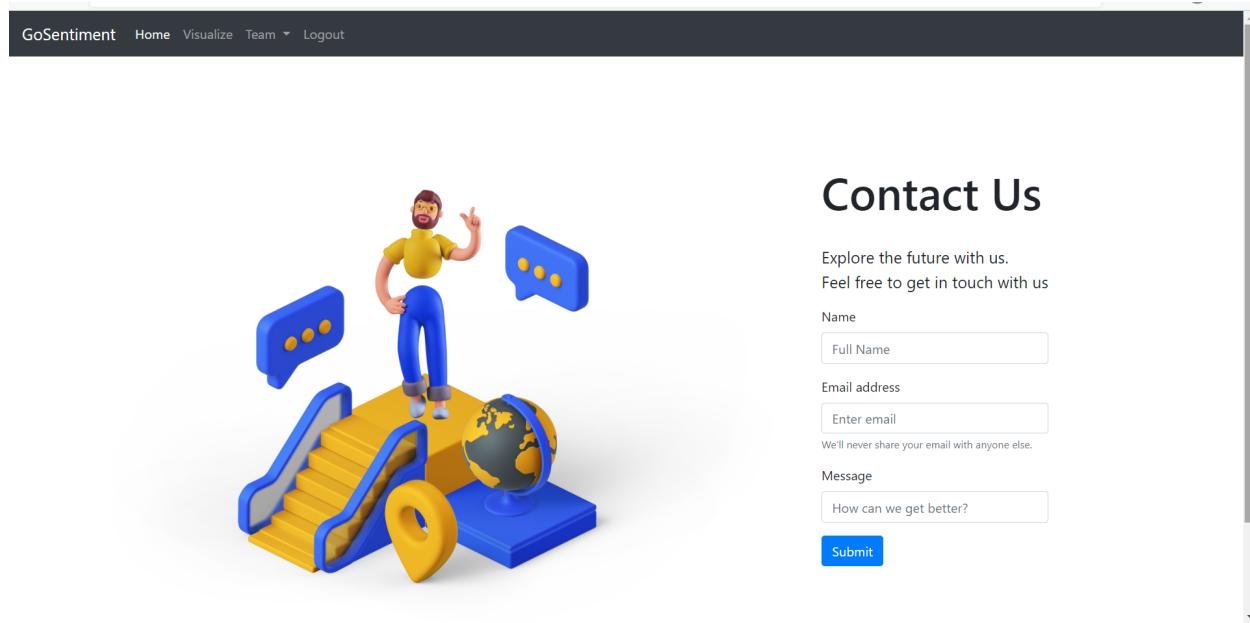
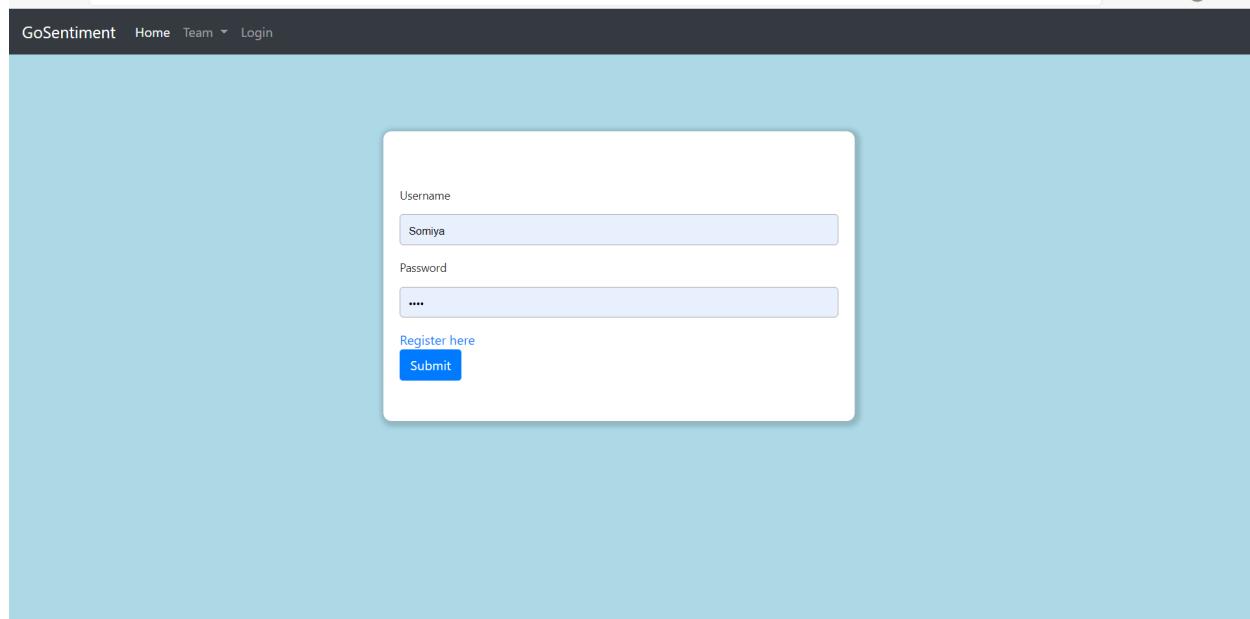


Fig 8.1.1 The Contact us page of our website

8.1.2 Login/Register Module



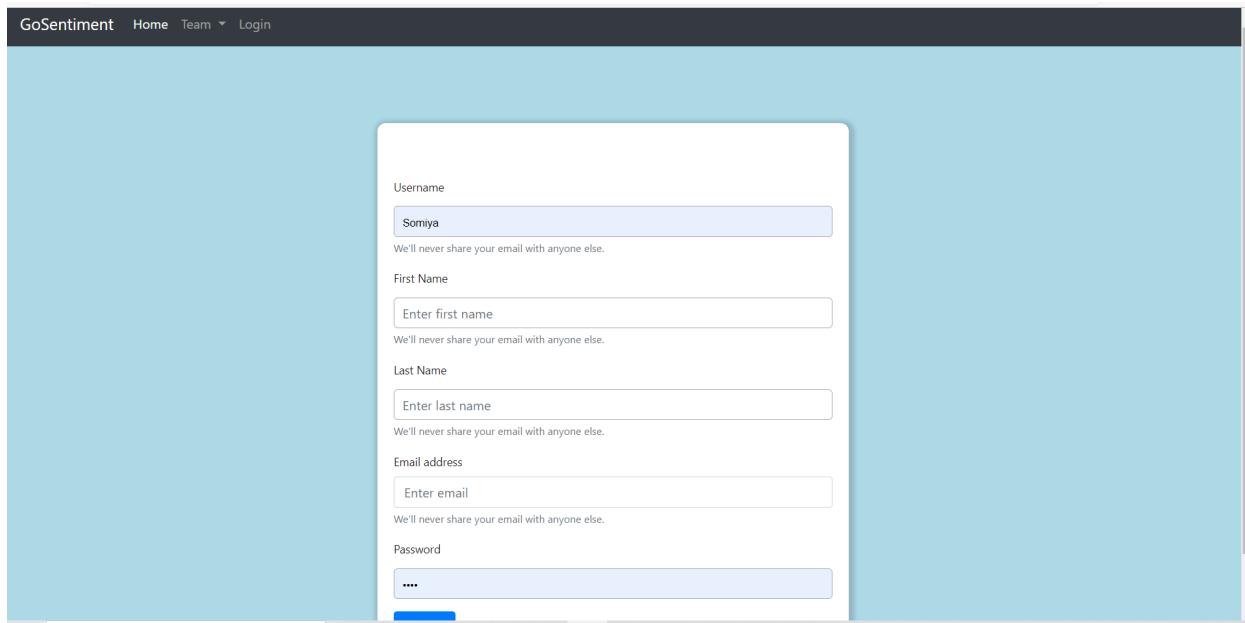


Fig 8.1.2 The Login and Register page of our website

8.1.3 About Us Module

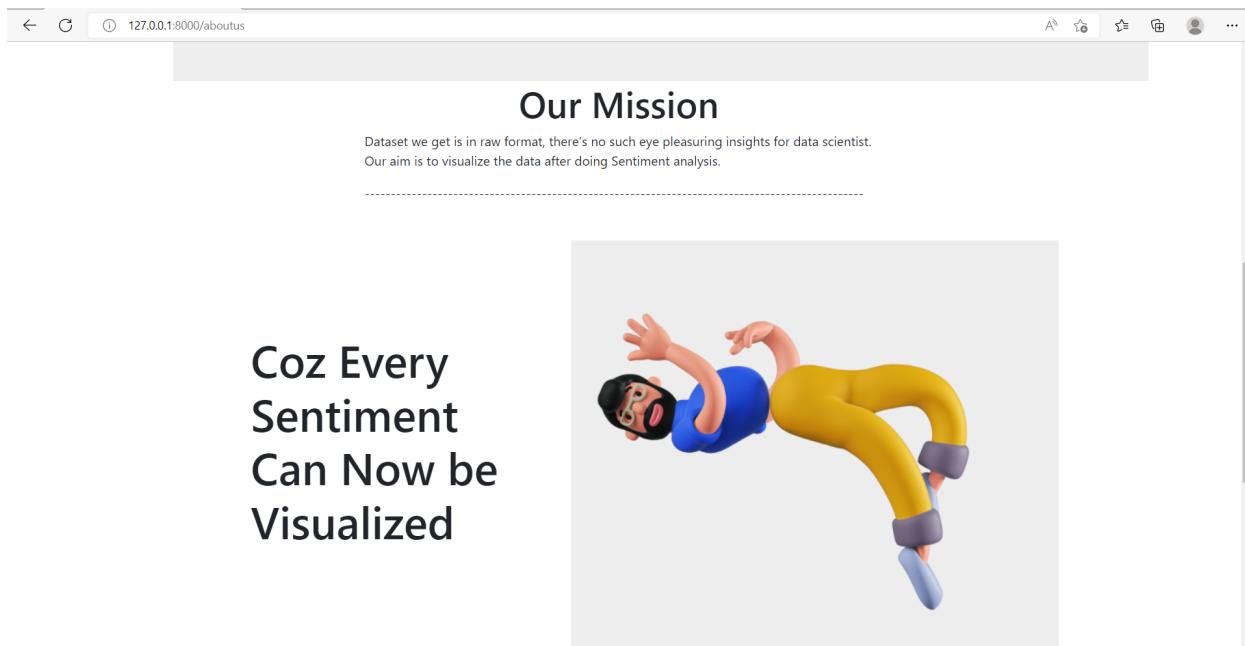
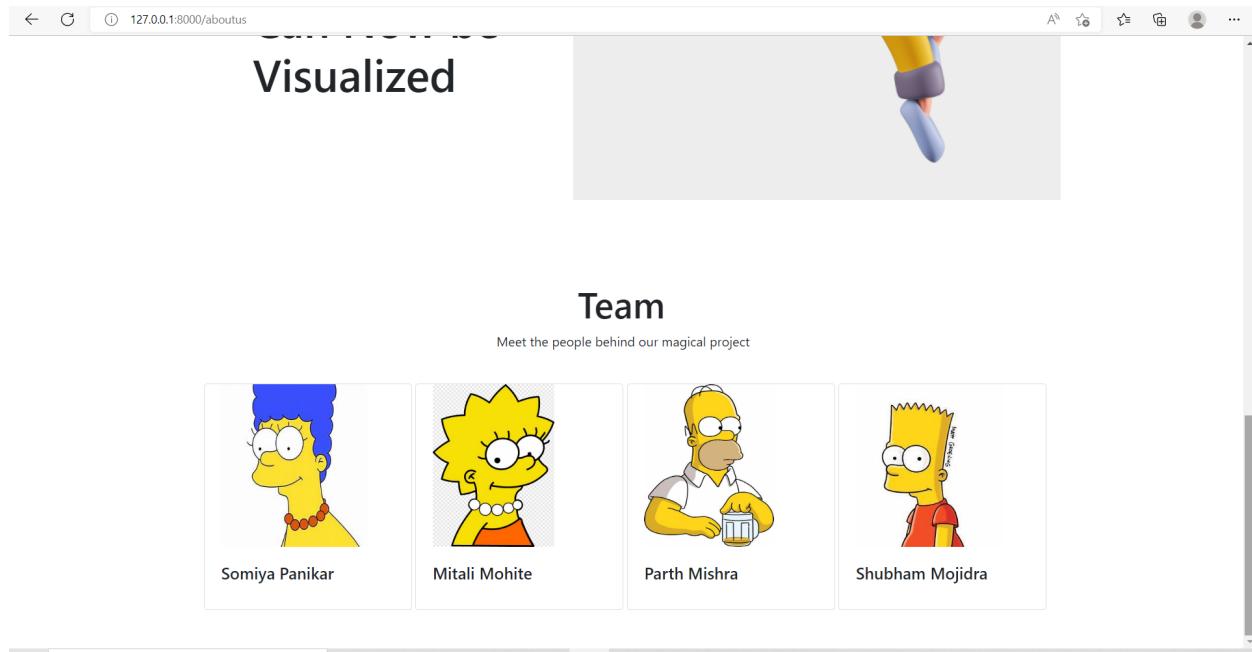


Fig 8.1.3 The About Us page of our website



8.1.4 Visualize Module

The screenshot shows a web browser window with the URL 127.0.0.1:8000/load_data. The navigation bar includes links for GoSentiment, Home, Visualize, Team, and Logout. A file upload input field shows "Choose File No file chosen". A blue "Analyze" button is located below the input field. A table displays the results of the analysis:

	Unnamed: 0	Sentences
0	0	Following is the draft auditor's report for Adverse opinion. Please note that following changes must be made
1	1	where applicable
2	2	Where the Company is requirement to prepare consolidated financial statements under the Companies
3	3	Act 2013, in the auditor's report, the term "financial statements", wherever appearing, would be
4	4	replaced by the term "Standalone financial statements".

Your Results

The Sentiment is Neutral



Fig 8.1.4 The Visualize page of our website

8.1.5 Gunningfog Module

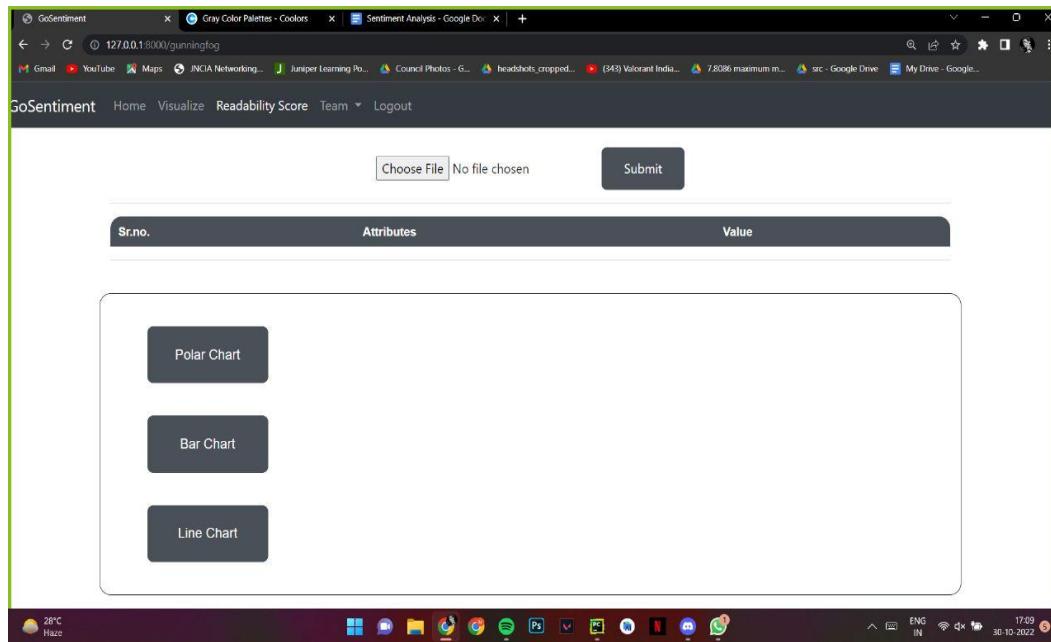


Fig 8.1.5 Landing page for gunningfog module

Sr.no.	Attributes	Value
1	Word Count	2387
2	Sentence Count	232
3	Avg Sentence Count	10.288793103448276
4	Syllable count	462
5	Poly syllable count	0
6	Syllable per word	0.2
7	Difficult Words	0
8	Gunning Fog Index	6.115517241379311
9	Flesch Reading Ease	179.47
10	Smog Index	3.1
11	Dale-Chall index	0.51

Fig 8.1.6 Various analytics for the audit submitted

Your Results

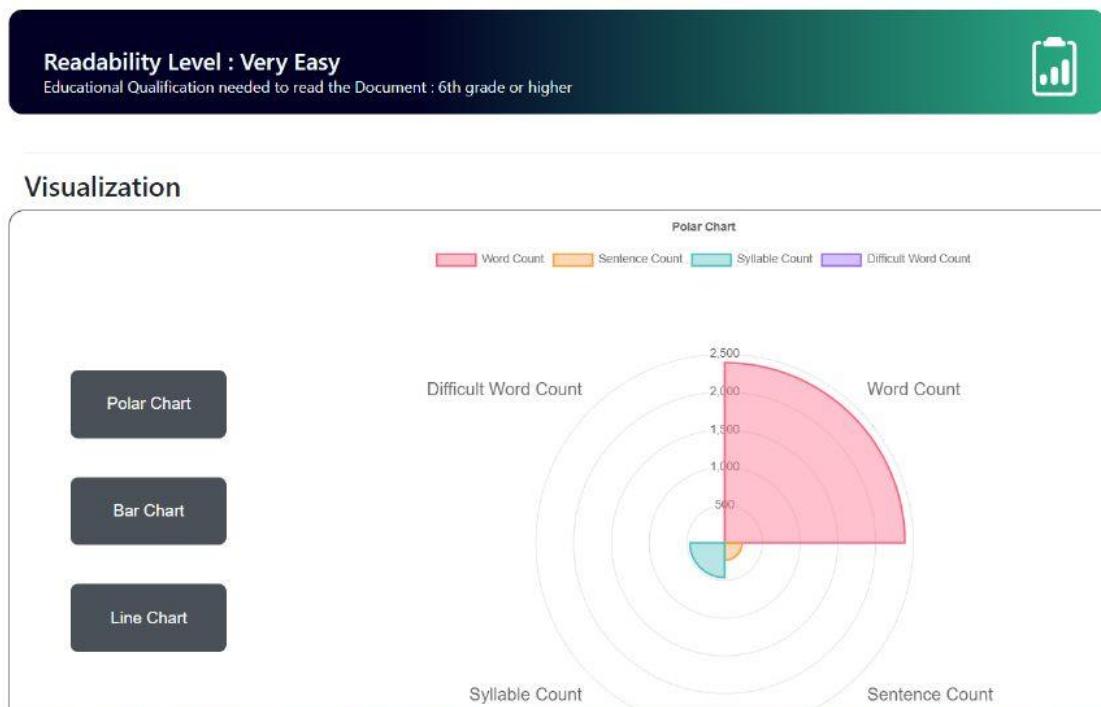


Fig 8.1.7 Analysis of audit reports

Chapter 9

Conclusion

We considered two methods for building our model. First we tried implementing using pre pre-defined Text Blob library and then finally we built our own model using Naive Bayes. Training is very easy and fast using this approach. Testing is straightforward, calculating the probabilities from the data provided. Our major task was to take the audit report from the user in pdf format, do segmentation of the pdf, convert it into a desirable csv format for the model , do sentiment analysis on it and then visualize the result using various charts. Sentiment analysis is or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. We have also made use of Gunning Fog Function to find the readability of the text in the audit reports. Using this User can get a glimpse of how easily readable their document is and make changes accordingly. The output of this function is then visualized using eye-pleasing graphs. The aim of our project was to ease the process of data analysis and visualize it properly. The accuracy of our project model is pretty nice. Hence we were able to successfully find the polarities and eventually predict the sentiment of the audit and also visualize it beautifully.

References

- [1] Gutha Jaya, Vadlamani Ravi, Bheemidi Vikram Reddy, Mohammad Zaheeruddin ,Harshal Jaiswal , “Sentiment Classification of Indian Banks’ Customer Complaints”.
- [2] Melik Masarifoglu; Umit Tigrak; Sefa Hakyemez; Guven Gul; Erdal Bozan; Ali Hakan Buyuklu; Arzucan Özgür, “Sentiment Analysis of Customer Comments in Banking using BERT-based Approaches”.
- [3] Gim Hoy Soong; Chye Cheah Tan, “Sentiment Analysis on 10-K Financial Reports using Machine Learning Approaches “.
- [4] Boyd-Graber, J. and P. Resnik. Holistic “sentiment analysis across languages: multilingual supervised latent Dirichlet allocation”. In Proceedings of EMNLP-2010, 2010.
- [5] Breck, E., Y. Choi, and C. Cardie. “Identifying expressions of opinion in context”. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2007), 2007.
- [6] Das, S. and M. Chen. “Yahoo! for Amazon: Extracting market sentiment from stock message boards”. In Proceedings of APFA-2001, 2001.

Annexure 1- Project Planning (Using Gantt chart)

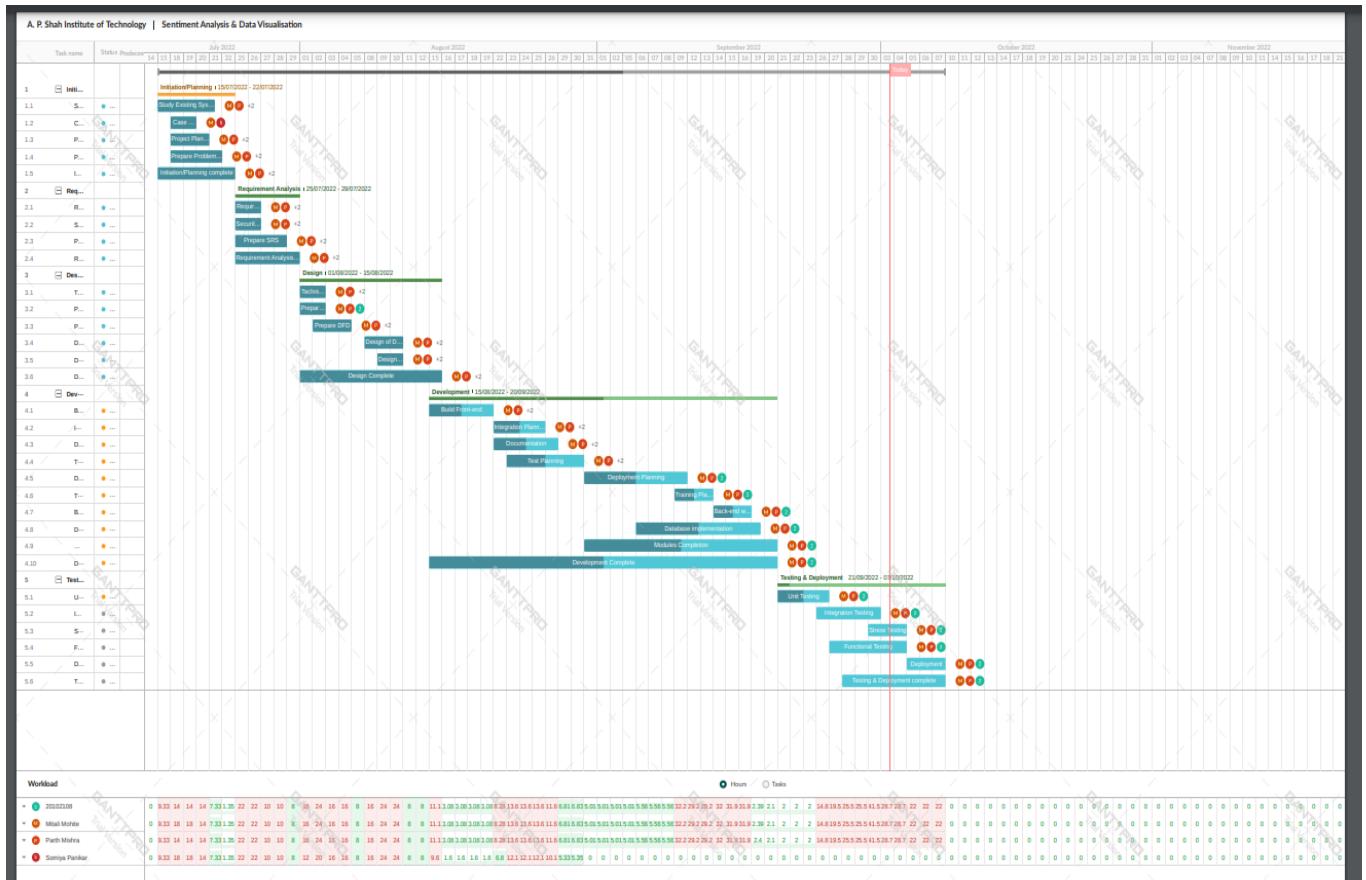


Fig 7.0 Gantt Chart

Acknowledgement

We have great pleasure in presenting the report on Sentiment Analysis & Data Visualization. We take this opportunity to express our sincere thanks towards our guide Prof. Suchita Dande, Department of Computer Engineering, APSIT, Thane and Prof. Sachin Malave, the HOD of Computer Engineering department for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards her constant encouragement, support and guidance through the development of the project. We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Somiya Panikar
20102095

Parth Mishra
20102162

Mitali Mohite
20102119

Shubham Mojidra
20102108