



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

**Rajiv Gandhi University of Knowledge Technologies – Nuzvid**

**Nuzvid, Eluru, Andhra Pradesh – 521202.**

**SENTIMENT ANALYSIS ON COMPANY REVIEWS BY  
USING MACHINE LEARNING**

A Project Progress Report

Submitted in partial fulfillment for the degree of

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

Submitted by

SK.KHASIM SAHEB (N180727)

G.SRAVANI (N180827)

CH.LAKSHMI PRASANNA (N180798)

CH.KEERTHANA (N180137)

S.V..KEERTHI (N180822)

*Under the Esteem Guidance of*

**Mr. KALAPALA SRAVAN KUMAR**



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

**Rajiv Gandhi University of Knowledge Technologies – Nuzvid**

**Nuzvid, Eluru, Andhra Pradesh – 521202.**

**CERTIFICATE OF COMPLETION**

This is to certify that the work entitled, “**Sentiment Analysis On Company Reviews By Using Machine Learning**” is the bonafied work of **SK.KHASIM SAHEB (N180727)** , **G.SRAVANI (N180827)**, **CH.LAKSHMI PRASANNA (N180798)** , **CH.KEERTHANA (N180137)**, **S.V.KEERTHI (N180822)** carried out under my guidance and supervision for 3rd year mini project of **Bachelor of Technology** in the department of Computer Science and Engineering under RGUKT IIIT, Nuzvid. This work is done during the academic session March 2022 – September 2022, under our guidance.

---

**Mr. KALAPALA SRAVAN KUMAR**

Assistant professor,

Department of CSE,

RGUKT, Nuzvid

---

**Mr.Chiranjeevi Sadu**

Assistant Professor,

Head of the Department,

Department of CSE,

RGUKT, Nuzvid.



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**  
**Rajiv Gandhi University of Knowledge Technologies – Nuzvid Nuzvid, Krishna,**  
**Andhra Pradesh – 521202.**

**CERTIFICATE OF EXAMINATION**

This is to certify that the work entitled, “**Sentiment Analysis On Company Reviews By Using Machine Learning**” is the bonafied work of **SK.KASIM SAHEB (ID No:N180727), G.SRAVANI (ID No:N180827) , CH.LAKSHMI PRASANNA (ID No:180798) , CH.KEERTHANA(ID No:180137),S.V.KEERTHI(ID No:N180822)**and here by accord our approval of it as a study carried out and presented in a manner required for its acceptance in 3rd year of **Bachelor of Technology** for which it has been submitted. This approval does not necessarily endorse or accept every statement made, opinion expressed or conclusion drawn, as recorded in this thesis. It only signifies the acceptance of this thesis for the purpose for which it has been submitted.

---

**Mr.Kalapala Sravan Kumar**

Assistant Professor,

Department of CSE,

RGUKT-Nuzvid.

---

**Project Examiner**

RGUKT-Nuzvid.



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**  
**Rajiv Gandhi University of Knowledge Technologies – Nuzvid Nuzvid, Krishna,**  
**Andhra Pradesh – 521202.**

**DECLARATION**

We “SK.KASIM SAHEB (ID No:N180727), G.SRAVANI (ID No:N180827) , CH.LAKSHMI PRASANNA (ID No:180798) , CH . KEERTHANA (ID No:180137) , S. V. KEERTHI (ID No:N180822)” hereby declare that the project report entitled “Sentiment analysis on company Reviews ” done by us under the guidance of **Mr. Sravan Kumar, Assistant Professor**, is submitted for the fulfillment of mini project during the academic session November 2021-February 2022 at RGUKT-Nuzvid.

We also declare that this project is a result of our own effort and has not been copied or imitated from any source. Citations from any websites are mentioned in the references. The results embodied in this project report have not been submitted to any other university or institute for the award of any degree or diploma.

**Date:10-07-2023**

**Place: Nuzvid**

SK.KASIM SAHEB (N180727)

G.SRAVANI (N180827)

CH.L.PRASANNA (N180798)

CH.KEERTHANA (N180137)

S.V.KEERTHI (N180822)

## ACKNOWLEDGEMENT

We would like to express our profound gratitude and deep regards to our guide **Mr. Kalapala Sravan Kumar** for his exemplary guidance, monitoring and constant encouragement to us throughout the B.Tech course. We shall always cherish the time spent with him during the course of this work due to the invaluable knowledge gained in the field of reliability engineering.

We are extremely grateful for the confidence bestowed in us and entrusting our project entitled **“Sentiment analysis on company Reviews ”**.

We express our gratitude to Mr. Chiranjeevi Sadu (HOD of CSE) and other faculty members for being a source of inspiration and constant encouragement which helped us in completing the project successfully.

Finally, yet importantly, we would like to express our heartfelt thanks to our beloved God and parents for their blessings, our friends for their help and wishes for the successful completion of this project.

## **ABSTRACT**

This project aims to create a college website to help students choose the best companies. NLP, FullStack, and ML techniques are used, including sentiment analysis. The project will begin by gathering company reviews, and training a sentiment analysis model using ML algorithms. Additional features such as packages, Basic description about the company, Recruitment process, related images, especially genuine reviews will be provided on web page. The Aim is to provide students with valuable information about each company. The project acknowledges the limitations of sentiment analysis algorithms and the lack of context, which can result in inaccurate results. The long-term goal of the project is to help students make better career choices through this Project.

## Table of Contents

<b>CHAPTER 1</b>	8
<b>INTRODUCTION</b>	8
1.1 Overview	8
1.2 Natural Language Toolkit (NLTK)	9
1.3 Machine learning (ML)	10
1.4 Data preprocessing	10
1.5 Importance of this project	11
<b>CHAPTER 2</b>	12
<b>REQUIREMENTS AND ANALYSIS</b>	12
2.1 Hardware components	12
2.2 Software components	13
<b>CHAPTER 3</b>	14
<b>PROPOSED MODEL AND FLOW OF THE PROJECT</b>	14
3.1 Proposed model	14
3.2 Flow of the project	14
3.3 Advantages	15
3.4 Applications	16
<b>CHAPTER 4</b>	18
<b>IMPLEMENTATION</b>	18
4.1 Implementation of the model	18
4.2 XAMPP Installation	18
4.3 Import all necessary libraries and Define actions	19
4.4 Model training	20
4.5 Web Implementation	20
<b>CHAPTER 5</b>	23
<b>RESULT</b>	23
5.1 Performance of the model	23
5.2 Limitations	25
<b>CHAPTER 6</b>	25
<b>CONCLUSION</b>	26
6.1 Summary	26
6.2 Future Work	26
6.3 References	26

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview:

The College Website for Company Reviews and Career Guidance project aims to develop a web-based platform to assist college students in selecting the best companies for their career paths. The website leverages Natural Language Processing (NLP), FullStack, and Machine Learning (ML) techniques, with a focus on sentiment analysis, to gather and analyze company reviews. By providing students with valuable information about each company, including packages, basic company descriptions, recruitment processes, related images, and genuine reviews, the project seeks to empower students in making informed career decisions.

The project acknowledges the limitations inherent in sentiment analysis algorithms and the potential lack of context, which can affect the accuracy of results. However, the primary objective is to empower students in their career decision-making process by presenting them with authentic reviews, company descriptions, recruitment processes, related images, and additional features that facilitate a better understanding of each company.

Key features of the project include data collection from company reviews, ML algorithm-based sentiment analysis model training, and the development of a user-friendly web interface. The NLP techniques employed enable the project to analyze the sentiment expressed in reviews, providing students with an overall perception of each company to aid their decision-making process.

The project report will encompass various chapters, starting with an introduction that highlights the project's significance and objectives. It will delve into the technologies utilized, including NLTK and ML, and emphasize the importance of data preprocessing. The proposed model and project flow will be outlined, focusing on the advantages and potential applications of the system.

Implementation details will cover the model implementation, installation of XAMPP, necessary library imports, model training, and the web implementation. The performance of the model will be evaluated, considering its accuracy, limitations, and the challenges associated with sentiment analysis.

The conclusion will summarize the project's achievements, emphasizing its aim to provide students with valuable information to facilitate better career choices. Future work will be suggested to address the limitations and enhance the website's functionality.



## 1.2 Natural Language Processing Tool Kit (NLTK):

**NLTK** (Natural Language Toolkit) is a fundamental component of our project, enabling us to incorporate advanced Natural Language Processing (NLP) techniques. As an open-source library widely used in the field of NLP, NLTK offers a comprehensive set of tools and resources that are instrumental in analyzing and processing human language data.

**Tokenization:** NLTK provides robust tokenization capabilities, allowing us to break down text into individual tokens, such as words or sentences. By utilizing NLTK's tokenization functionality, we can effectively segment and structure the textual data we gather, facilitating subsequent analysis.

**Stemming and Lemmatization:** NLTK includes powerful stemmers and lemmatizers, which are essential for reducing words to their base or root forms. This process helps in standardizing the vocabulary used in company reviews, minimizing variations, and enabling more accurate analysis. By applying NLTK's stemming and lemmatization techniques, we can enhance the quality and consistency of our textual data.

**Part-of-speech (POS) Tagging:** NLTK provides pre-trained models and tools for part-of-speech tagging. With NLTK's POS tagging functionality, we can assign grammatical tags to words in company reviews, such as nouns, verbs, adjectives, and more. This information offers valuable insights into the syntactic structure of the text, aiding in more nuanced analysis.

**Sentiment Analysis:** NLTK offers pre-trained models and utilities for sentiment analysis, a key component of our project. Leveraging NLTK's sentiment analysis capabilities, we can assess the sentiment expressed in company reviews. This allows us to categorize reviews as positive, negative, or neutral, providing valuable information to students about the overall sentiment towards different companies.

By harnessing the power of NLTK, we can preprocess textual data, perform sentiment analysis on company reviews, and extract meaningful information for presentation on our website. Additionally, NLTK's extensive range of functionalities equips us with the necessary tools to handle the linguistic complexities inherent in our project.

## 1.3 Machine Learning:

The ML component of our project utilizes the Naive Bayes classification algorithm for sentiment analysis. Naive Bayes is a widely-used algorithm for text classification tasks, including sentiment analysis. It leverages labeled data to train a model that can predict the sentiment (positive, negative, or neutral) of company reviews.

During the training phase, the model learns the relationships between textual features and sentiments from the labeled dataset. It estimates the probabilities of specific words or features occurring in each sentiment class. These probabilities are then used to classify new, unlabeled company reviews based on their textual features.

In our project report, we discuss the implementation details of the Naive Bayes algorithm, including dataset preparation, feature extraction, and probability estimation. We evaluate the performance of the sentiment analysis model using standard metrics like accuracy.

## 1.4 Data Preprocessing:

Data preprocessing is a crucial step in our project, as it ensures that the collected company reviews are in a suitable format for further analysis and model training. In our case, the reviews were manually gathered from popular platforms such as Glassdoor and AmbitionBox for each company.

**The data preprocessing phase involves several steps:**

**1. Data Collection:** We collected the company reviews manually from Glassdoor and AmbitionBox. This process involved searching for each company and gathering a substantial number of reviews to ensure a representative dataset. By manually collecting the reviews, we aimed to gather genuine and diverse opinions expressed by employees or candidates.

**2. Text Cleaning:** Raw text data often contains noise and irrelevant information that can hinder analysis. In the data preprocessing phase, we cleaned the reviews by removing unnecessary characters, such as special symbols or punctuation marks. We also eliminated any HTML tags or formatting artifacts that might be present in the collected data.

**3. Tokenization:** To analyze the textual data effectively, we employed tokenization. This process involves breaking down the reviews into individual tokens, typically words or sentences.

**4. Stopword Removal:** Stopwords are common words that do not carry significant meaning in the context of sentiment analysis, such as "the," "is," and "and."

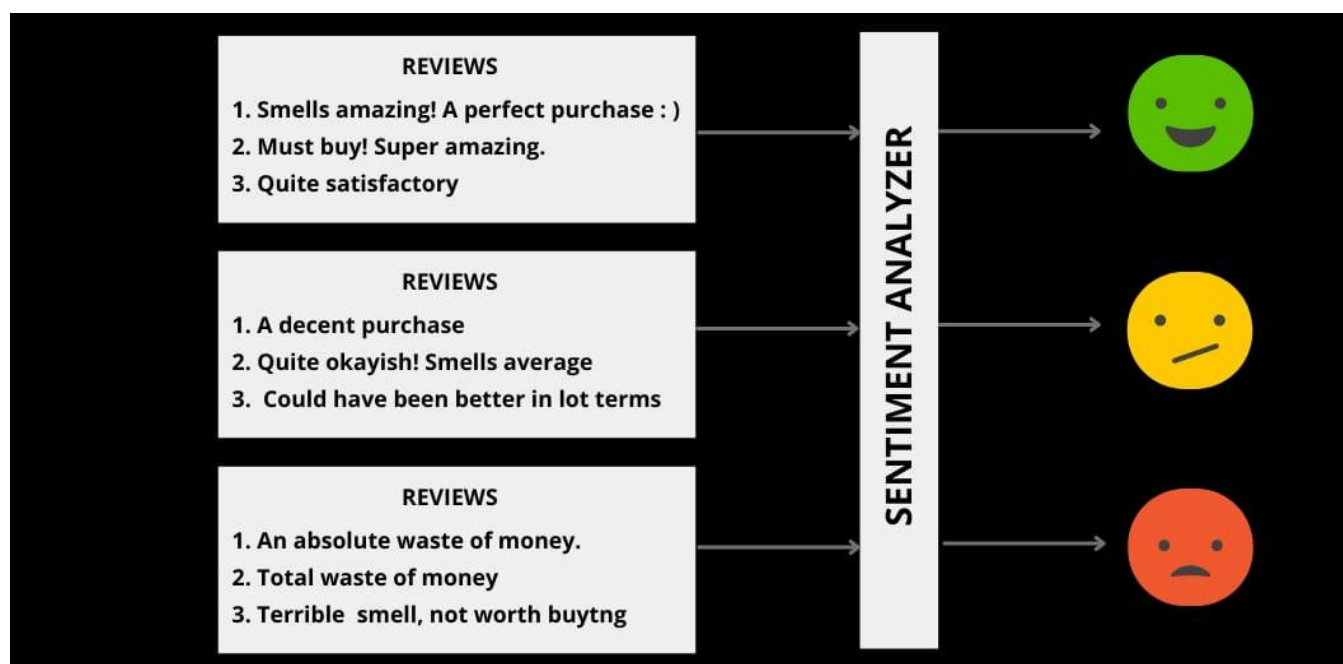
**5. Lowercasing:** To achieve consistency and avoid duplications, we converted all the text to lowercase. This step ensures that words with different cases are treated as the same, improving the accuracy of subsequent analysis.

**6. Data Labeling:** For sentiment analysis, we labeled the collected reviews with the appropriate sentiment category (positive, negative, or neutral). This step involved manually assigning sentiment labels based on the overall tone and sentiment expressed in each review.

By performing these data preprocessing steps, we obtained a clean, structured dataset ready for further analysis and model training. The preprocessed data serves as the foundation for training our sentiment analysis model using the Naive Bayes algorithm.

### 1.5 Importance of this project:

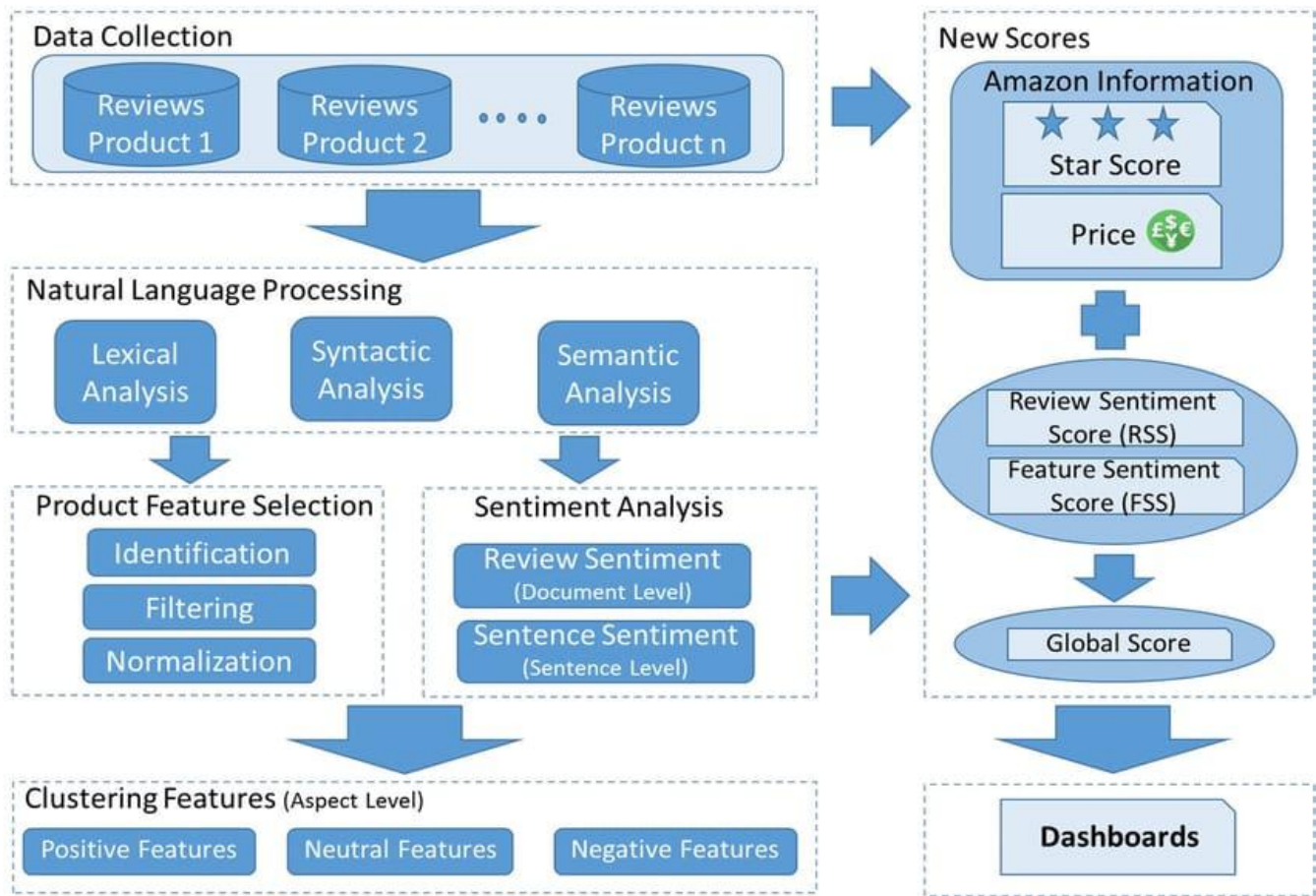
1. Empowering students in making informed career choices.
2. Bridging the information gap about companies for students.
3. Utilizing NLP and ML techniques for accurate sentiment analysis.
4. Enhancing career decision-making through valuable information.
5. Long-term goal of continuous improvement for student guidance.



*figure.1 overview of this project*

## CHAPTER 2

### ARCHITECTURE:



*figure.2 Achitecture of sentiment analysis*

### REQUIREMENTS AND ANALYSIS

#### 2.1 Hardware components:

- Processor: 64-bit, quad-core, 2.5 GHz minimum per core
- RAM: 4 GB or more.
- HDD: 20 GB of available space or more.
- Display: Dual XGA (1024 x 768) or higher resolution monitors.
- Keyboard: A standard keyboard

## 2.2 Software components:

**1. XAMPP:** XAMPP is a software package that provides a local development environment for web applications. It includes Apache as the web server, MySQL as the database management system, and PHP as the programming language. XAMPP simplifies the setup and configuration of the development environment, allowing you to develop college websites .

**2. HTML, CSS, and JavaScript:** These are the fundamental building blocks of web development. HTML (Hypertext Markup Language) is used for structuring the content of web pages, CSS (Cascading Style Sheets) is used for styling and layout, and JavaScript is used for adding interactivity and dynamic functionality to the website.

**3. PHP:** PHP (Hypertext Preprocessor) is a popular server-side scripting language used for dynamic web development. In your project, PHP can be utilized to handle server-side processing, database interactions, and rendering dynamic content on the web pages.

**4. MySQL:** MySQL is a widely used open-source relational database management system. It allows you to store, manage, and retrieve the collected company reviews, related information, and user data. MySQL provides a robust and efficient way to handle data persistence in your project.

**VADER Sentiment Analysis:** VADER is a rule-based sentiment analysis tool specifically designed for sentiment analysis of social media texts. It uses a combination of lexicons and grammatical rules to assess the sentiment expressed in text. VADER assigns sentiment scores to individual words and calculates an overall sentiment polarity score for a given text.

**NLP Libraries (e.g., NLTK):** NLP libraries are crucial software components for text processing and analysis tasks. Libraries like NLTK (Natural Language Toolkit) or SpaCy provide functionalities such as tokenization, stemming, lemmatization, part-of-speech tagging, and sentiment analysis. These libraries enable you to process and analyze the company reviews and extract meaningful insights.

**Machine Learning Libraries :** ML libraries are utilized for training the sentiment analysis model. Libraries like Scikit-learn or TensorFlow provide implementations of ML algorithms, including Naive Bayes, which is employed for sentiment analysis in your project. These libraries facilitate the training, evaluation, and integration of the sentiment analysis model into the web application

**Pandas:** a powerful Python library, plays a vital role in data preprocessing for sentiment analysis. It simplifies tasks such as data loading, exploration, cleaning, text preprocessing, transformation, filtering, integration, and export. With its extensive set of functions, Pandas facilitates efficient data manipulation and prepares the dataset for sentiment analysis tasks.

## CHAPTER 3

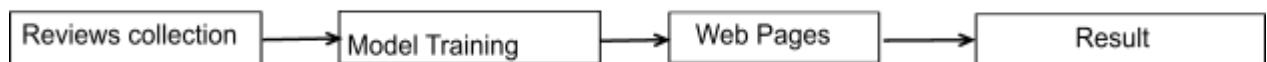
### PROPOSED MODEL AND FLOW OF THE PROJECT

#### 3.1 Proposed model

This project creates a college website using NLP, FullStack, and ML techniques, including sentiment analysis. It gathers company reviews and trains a sentiment analysis model using ML algorithms. The website provides features like company packages, descriptions, recruitment processes, related images, and genuine reviews to help students make informed career choices. The project acknowledges the limitations of sentiment analysis algorithms and aims to improve students' decision-making over the long term.

#### 3.2 Flow of the project:

**Broadly project flow is classified as:**



Based on our project abstract, here's a proposed flow for our project:

##### 1. Data Collection:

- Gather company reviews from sources like Glassdoor and AmbitionBox.
- Manually collect a significant number of reviews for each company.
- Extract relevant information such as review text, date, rating, and metadata.

##### 2. Data Preprocessing:

- Clean and preprocess the collected reviews by removing irrelevant characters.
- Perform text normalization, removing stopwords, and handling noise in the data.

##### 3. Sentiment Analysis Model Training:

- Label the reviews with sentiment categories (positive, negative, neutral).
- Train a sentiment analysis model using ML algorithms, such as Naive Bayes.
- Utilize NLP techniques and libraries (e.g., NLTK) for sentiment analysis and polarity scoring.

##### 4. Web Development:

- Use XAMPP as the web development framework for building the college website.
- Design an intuitive and user-friendly interface for the website.
- Implement features for user registration and login to access website functionalities.

## **5. Company Profile Pages:**

- Create individual profile pages for each company.
- Display basic company information (description, industry, location, employee count).
- Provide additional details about the company (packages, recruitment process)

## **6. Reviews and Sentiment Analysis Display:**

- Showcase the collected company reviews on their respective profile pages.
- Integrate sentiment analysis results, showing the overall sentiment of the reviews.
- Use visual indicators or labels to represent sentiment scores for each company.

## **7. Additional Features and Content:**

- Include related images or media to enhance the company profile pages.
- Offer options for users to like or provide feedback on company profiles or specific reviews.

## **8. Search and Filtering Functionality:**

- Implement search options for users to find specific companies based on location, name.
- Enable users to narrow down their search and focus on companies that match their preferences.

## **9. Deployment and Testing:**

- Deploy the website on a web server using XAMPP.
- Conduct thorough testing to ensure proper functionality and user experience.

## **10. Long-term Goal and Future Work:**

- limitations of sentiment analysis algorithms and strive for continuous improvement.
- Gather feedback from seniors and incorporate enhancements based on their needs.
- Explore additional features or modules to expand the capabilities of the college website.

### **3.3 Advantages:**

**1. Accessible Information:** The project provides students with a centralized platform to access valuable information about different companies. Students can easily explore company profiles.

**2. Informed Decision-Making:** By utilizing sentiment analysis techniques, the project enables students to make informed career choices.

**3. Genuine Reviews:** The project incorporates genuine company reviews from sources like Glassdoor and AmbitionBox. This ensures that students have access to authentic opinions and experiences shared by current or former employees, enhancing the credibility of the information provided.

**4. User-Friendly Interface:** The website's user interface is designed to be intuitive and user-friendly. Students can easily navigate through company profiles, search for specific companies, and access the desired information. The interface promotes a seamless and efficient user experience.

**5. Additional Features:** The inclusion of additional features such as company packages, related images, and user feedback options enhances the overall functionality and interactivity of the website.

**6. Continuous Improvement:** The project acknowledges the limitations of sentiment analysis algorithms and the lack of contextual understanding. This recognition encourages continuous improvement and refinement of the sentiment analysis model to provide more accurate and reliable results over time.

### **3.4 Applications:**

**1. Career Decision Making:** The project assists students in making informed career choices by providing valuable information and insights about different companies.

**2. College Placement Programs:** The project can be utilized by colleges and universities during placement programs to offer students comprehensive information about participating companies.

**3. Internship and Job Search:** Students can use the project to research companies, read reviews, and gather insights when searching for internships or job opportunities.

**4. Company Research and Analysis:** The project provides a platform for researchers, HR professionals, and recruiters to gather data and analyze company reviews for market research or talent acquisition purposes.

**5. Educational Institutions:** The project can be integrated into career counseling services at educational institutions to provide students with a reliable resource for career guidance.



**6. Alumni Networking:** The project can facilitate alumni networking by providing a platform for students to access information about companies where alumni are employed.

**7. Industry Insights:** The sentiment analysis results and aggregated reviews can offer industry insights, helping professionals and organizations understand trends and perceptions within specific sectors.

**8. Continuous Improvement of Companies:** Companies can utilize the project to gain insights into employee experiences, identify areas for improvement, and enhance their employer brand.

**9. User Feedback and Engagement:** The project's feedback and engagement features allow users to provide input, share experiences, and contribute to the community, fostering active user participation.

**10. Research and Development:** The project can serve as a foundation for further research and development in the field of sentiment analysis, NLP, and ML techniques applied to company reviews.

## CHAPTER 4

### IMPLEMENTATION

#### 4.1 Implementation of the model

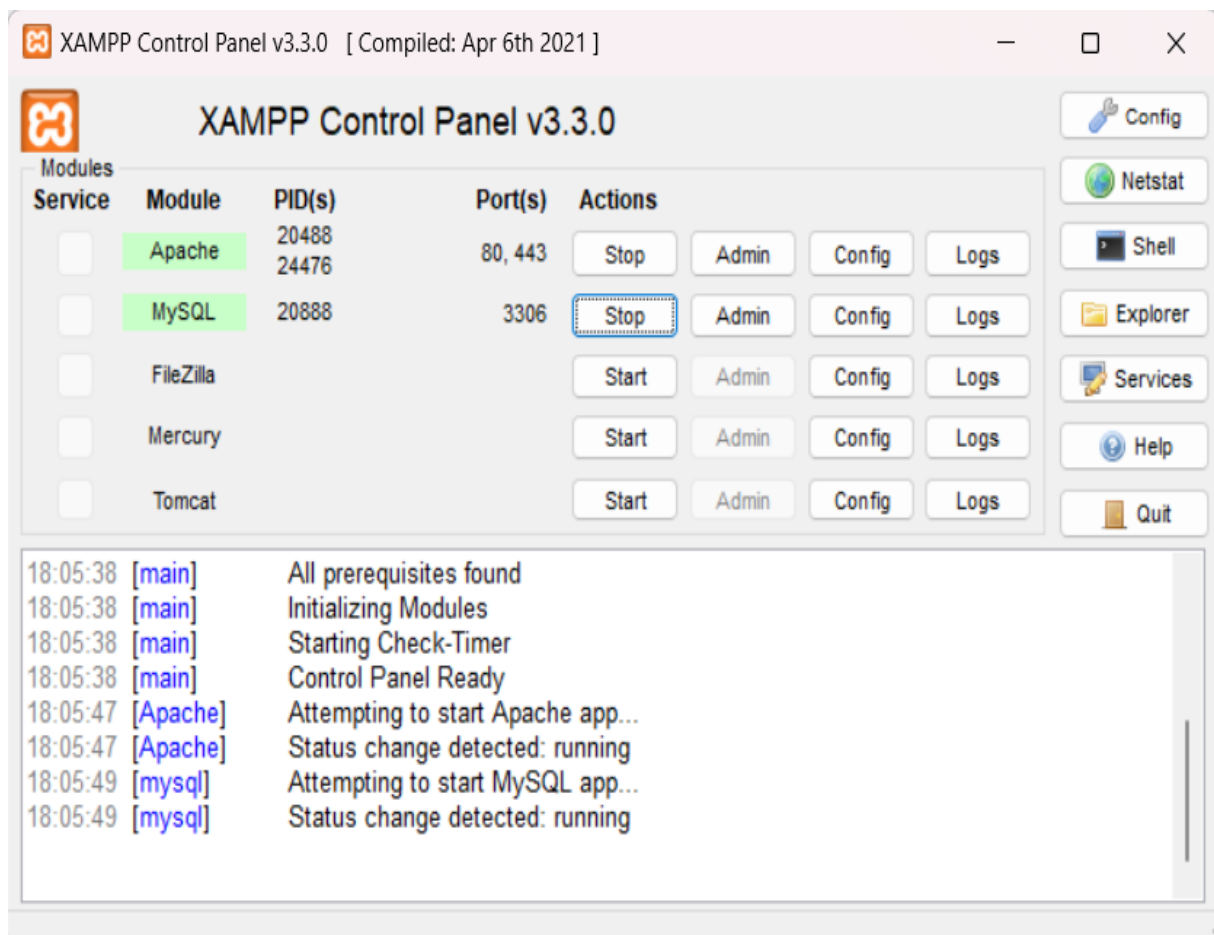
Various models are implemented to get the efficient results.

##### Experimented models:

Naive bayes classification : Accuracy(70%)  
NLTK Libraries : Accuracy(70%)  
SQLite :Accuracy(73%)

#### 4.2 XAMPP Installation:

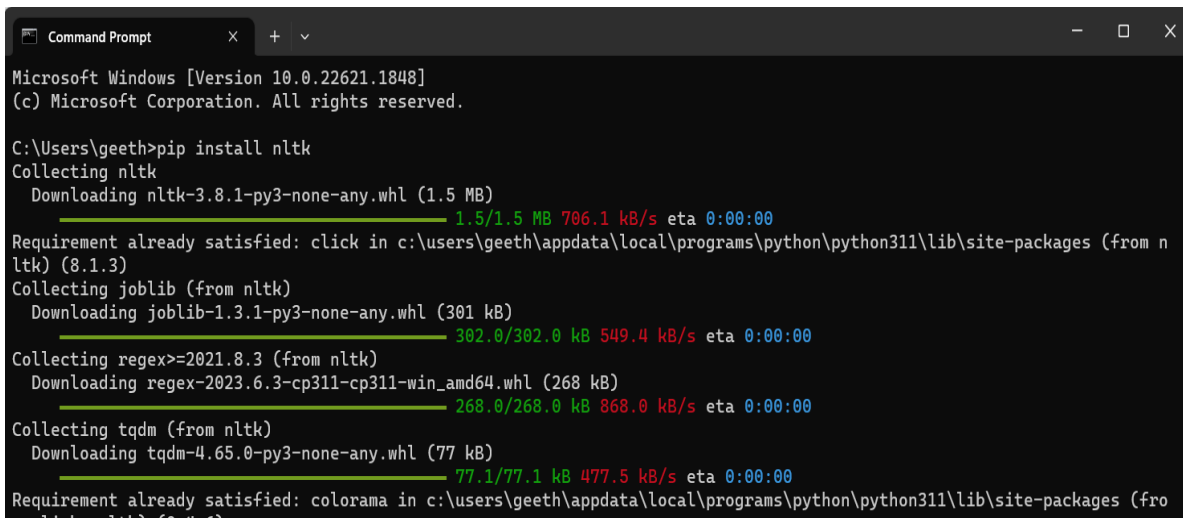
- XAMPP installation involves setting up the XAMPP software on your local machine.
- XAMPP provides a bundled package that includes Apache , MySQL , PHP , and other necessary components.
- Installing XAMPP is a straightforward process where we follow the instructions provided by the XAMPP installation wizard for our specific operating system.
- Once installed, we can configure and run the local web server to host our web application.



*figure.3 XAMPP installation*

### 4.3 Import all necessary libraries and Define actions:

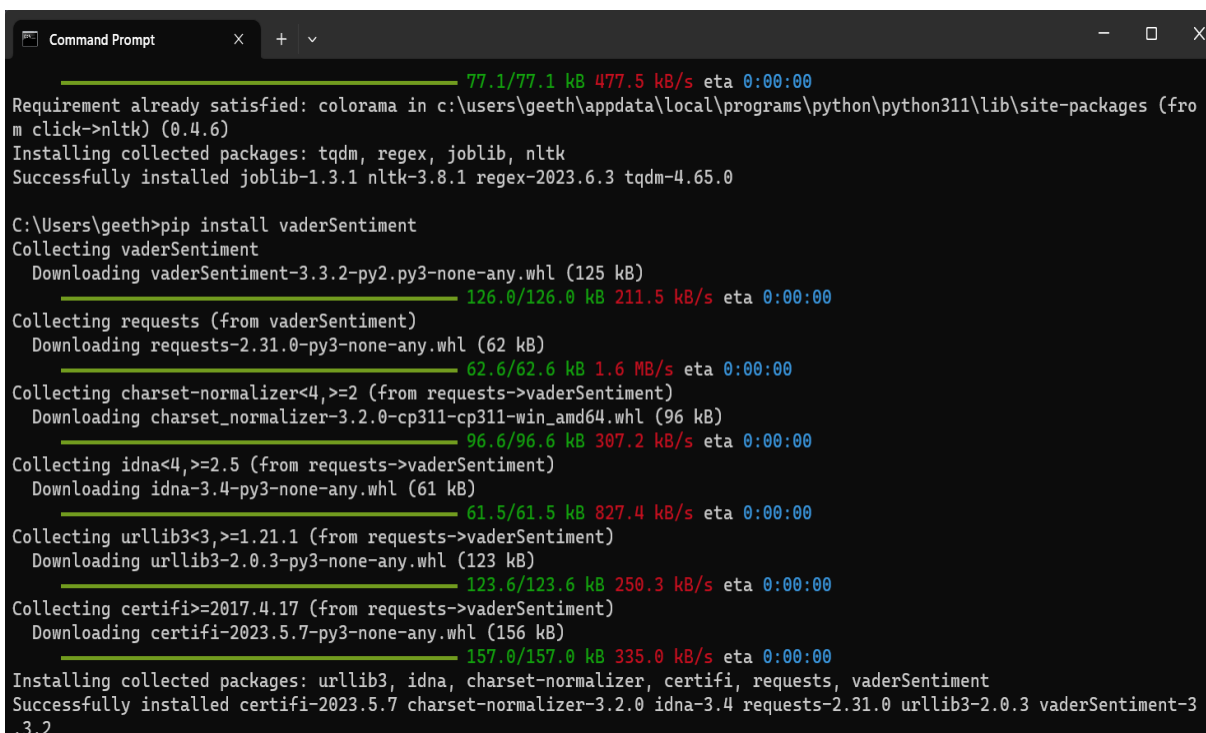
- In this step, we import the required libraries and packages into our project environment.
- Common libraries used for web development with Python include Flask, Django, or other frameworks suitable for our project.
- You define the necessary actions and functionalities that your web application will perform, such as handling user authentication, processing search queries, displaying company profiles, and managing user interactions.
- These actions define the behavior and functionality of your web application.



```
Microsoft Windows [Version 10.0.22621.1848]
(c) Microsoft Corporation. All rights reserved.

C:\Users\geeth>pip install nltk
Collecting nltk
  Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)
    1.5/1.5 MB 706.1 kB/s eta 0:00:00
Requirement already satisfied: click in c:\users\geeth\appdata\local\programs\python\python311\lib\site-packages (from nltk) (8.1.3)
Collecting joblib (from nltk)
  Downloading joblib-1.3.1-py3-none-any.whl (301 kB)
    302.0/302.0 kB 549.4 kB/s eta 0:00:00
Collecting regex<=2021.8.3 (from nltk)
  Downloading regex-2023.6.3-cp311-win_amd64.whl (268 kB)
    268.0/268.0 kB 868.0 kB/s eta 0:00:00
Collecting tqdm (from nltk)
  Downloading tqdm-4.65.0-py3-none-any.whl (77 kB)
    77.1/77.1 kB 477.5 kB/s eta 0:00:00
Requirement already satisfied: colorama in c:\users\geeth\appdata\local\programs\python\python311\lib\site-packages (from click==8.1.3) (0.4.6)
```

*figure 4:NLTK installation*



```
Requirement already satisfied: colorama in c:\users\geeth\appdata\local\programs\python\python311\lib\site-packages (from click->nltk) (0.4.6)
Installing collected packages: tqdm, regex, joblib, nltk
Successfully installed joblib-1.3.1 nltk-3.8.1 regex-2023.6.3 tqdm-4.65.0

C:\Users\geeth>pip install vaderSentiment
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
    126.0/126.0 kB 211.5 kB/s eta 0:00:00
Collecting requests (from vaderSentiment)
  Downloading requests-2.31.0-py3-none-any.whl (62 kB)
    62.6/62.6 kB 1.6 MB/s eta 0:00:00
Collecting charset-normalizer<4,>=2 (from requests->vaderSentiment)
  Downloading charset_normalizer-3.2.0-cp311-cp311-win_amd64.whl (96 kB)
    96.6/96.6 kB 307.2 kB/s eta 0:00:00
Collecting idna<4,>=2.5 (from requests->vaderSentiment)
  Downloading idna-3.4-py3-none-any.whl (61 kB)
    61.5/61.5 kB 827.4 kB/s eta 0:00:00
Collecting urllib3<3,>=1.21.1 (from requests->vaderSentiment)
  Downloading urllib3-2.0.3-py3-none-any.whl (123 kB)
    123.6/123.6 kB 250.3 kB/s eta 0:00:00
Collecting certifi>=2017.4.17 (from requests->vaderSentiment)
  Downloading certifi-2023.5.7-py3-none-any.whl (156 kB)
    157.0/157.0 kB 335.0 kB/s eta 0:00:00
Installing collected packages: urllib3, idna, charset-normalizer, certifi, requests, vaderSentiment
Successfully installed certifi-2023.5.7 charset-normalizer-3.2.0 idna-3.4 requests-2.31.0 urllib3-2.0.3 vaderSentiment-3.3.2
```

*figure 5:Vader Sentiment installation*

## 4.4 Model Training:

- Model training involves utilizing the collected and preprocessed data to train a sentiment analysis model.
- You select the appropriate ML algorithm, such as Naive Bayes, SVM, or Neural Networks, for sentiment analysis.
- Using the labeled data, you train the model to learn patterns and relationships between the text inputs and their sentiment labels.
- This training process enables the model to classify the sentiment of new text inputs accurately.

```
data['sentiment'] = data['rating'].apply(label_sentiment)

# Splitting the Dataset
X_train, X_test, y_train, y_test = train_test_split(data['Reviews'], data['sentiment'], test_size=0.2, random_state=42)

# Vectorizing the Text
cv = CountVectorizer()
X_train_cv = cv.fit_transform(X_train)
X_test_cv = cv.transform(X_test)

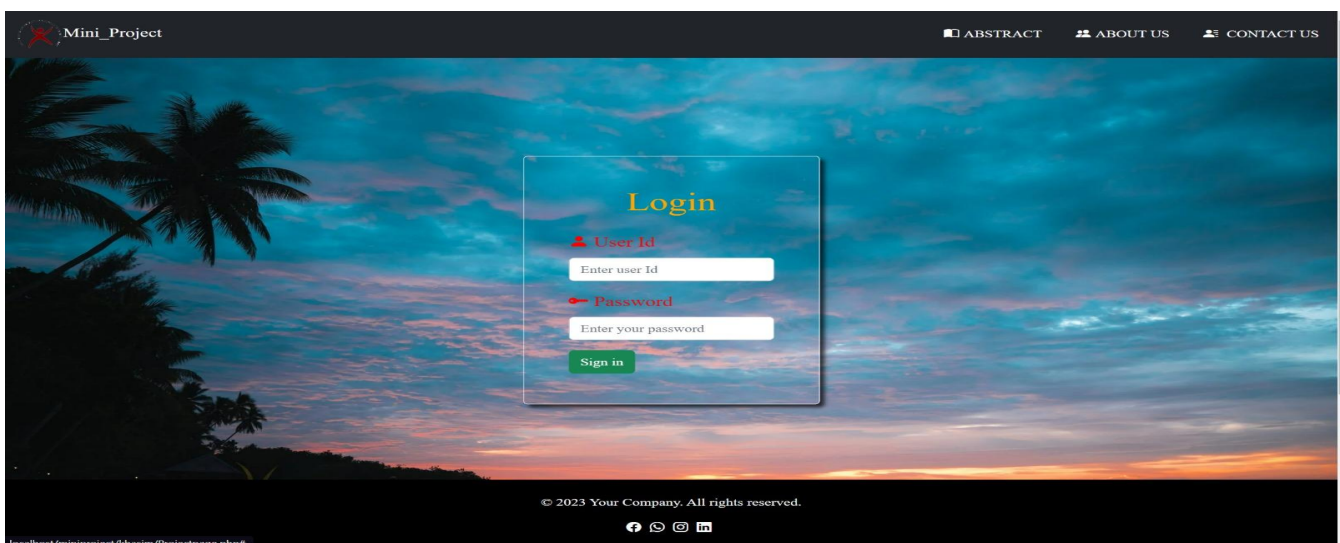
# Training the Model
nb = MultinomialNB()
nb.fit(X_train_cv, y_train)

# Testing and Evaluating the Model
y_pred = nb.predict(X_test_cv)
print('Accuracy:', accuracy_score(y_test, y_pred))
print('Confusion Matrix:', confusion_matrix(y_test, y_pred))
```


## 4.5 Web Implementation:


- Web implementation refers to the process of integrating the trained sentiment analysis model into your web application.
- You develop the different components of your web application, including the login page, search functionality, company profile pages, and any additional features.


### Login Page:





## Search Page:

Mini\_Project

HOME





ABOUT US

CONTACT US

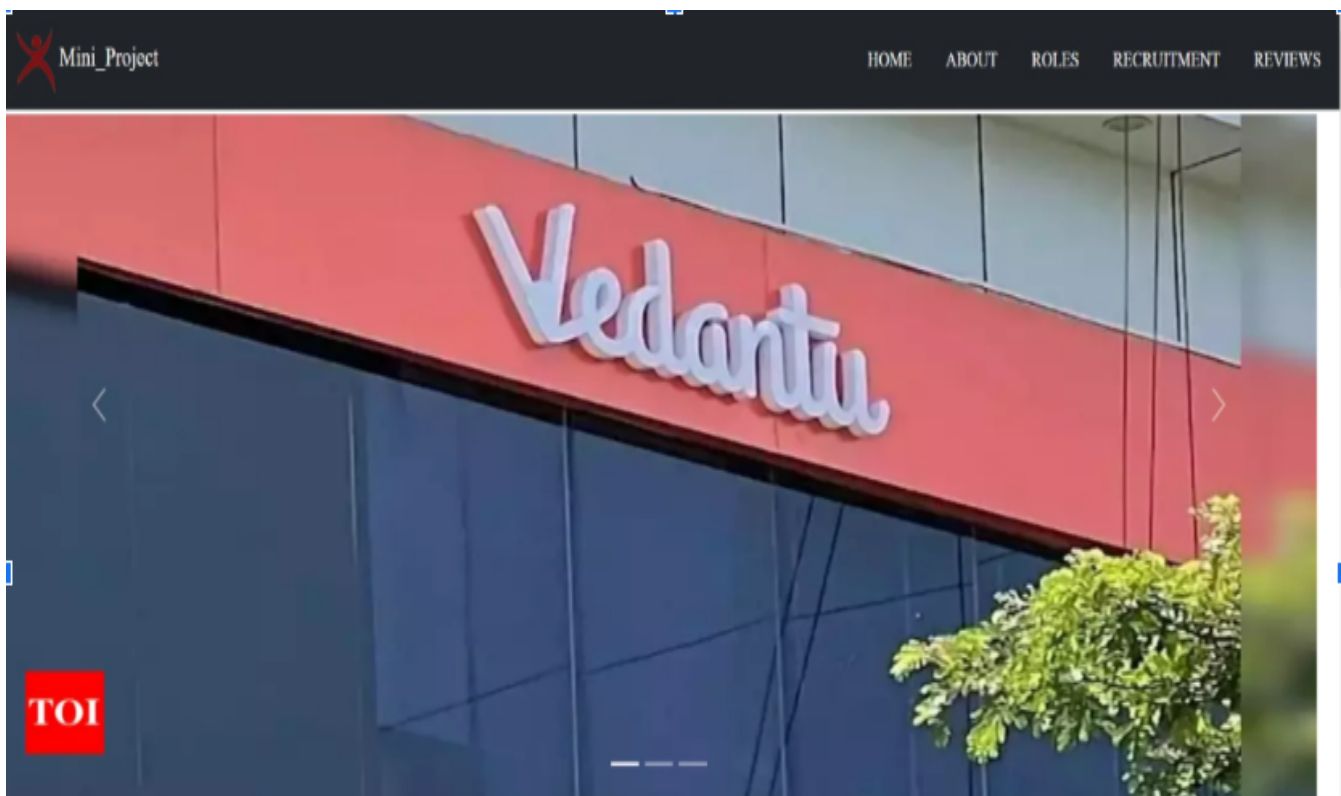


Company Name	Location
Analog Devices	Bengaluru
Vedanthu	Maharastra
Thought Works	Bengaluru
Byjus	Chennai
Gridlex	Haryana
HCL	Mumbai
Synopsys	Bengaluru
Test Sigma	Bengaluru
PrimeSoft	Hyderabad
Purple.com	Mumbai
PromptCloud	Bengaluru
Wipro	Bengaluru

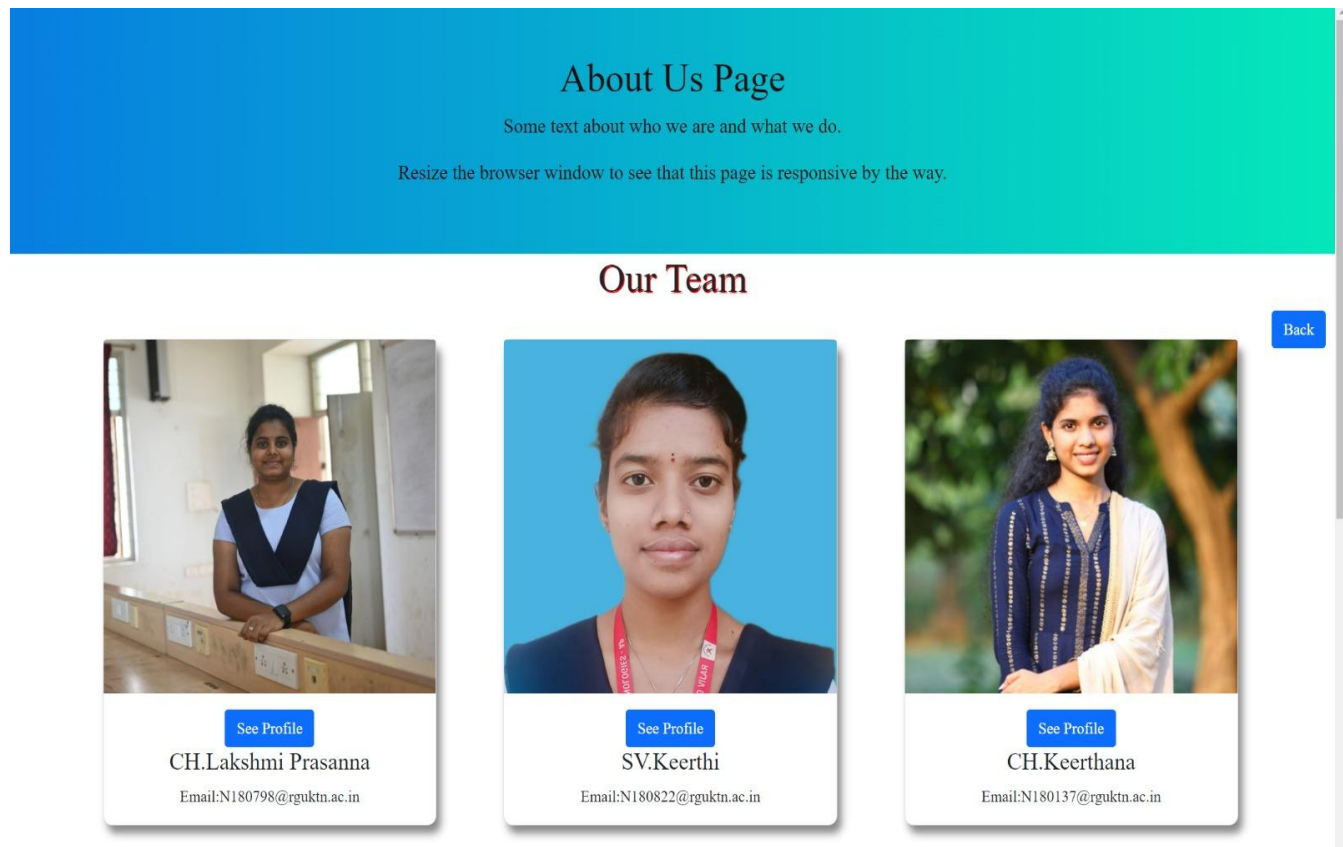
© 2023 Your Company. All rights reserved.



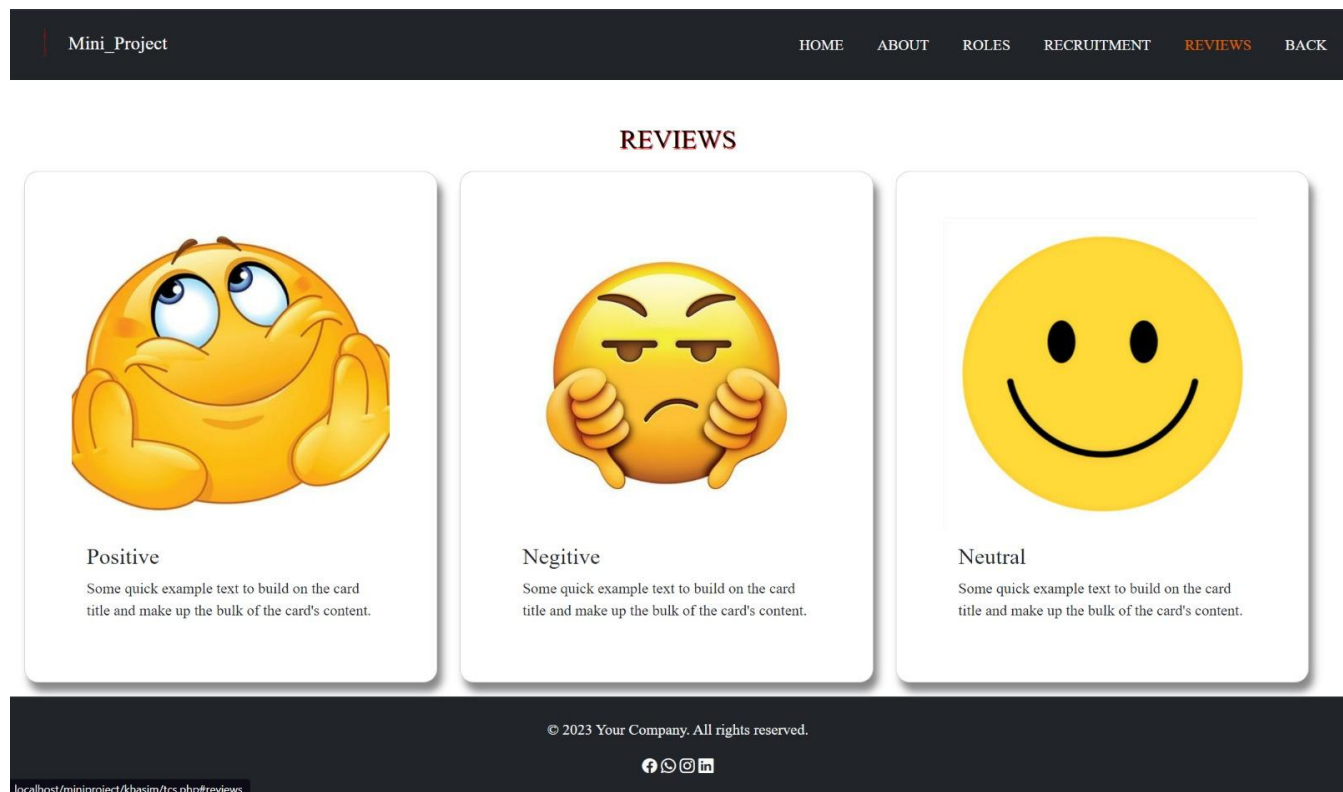
## Example company page:



## About us Page:



## Reviews Classification:



These are the steps we followed during the implementation of our projet

## CHAPTER 5

### RESULT

#### 5.1 Performance of the model:

In our project, Naive Bayes classification is employed for sentiment analysis of company reviews. The Naive Bayes algorithm is chosen due to its simplicity, efficiency, and effectiveness in text classification tasks. The first step involves data preparation, where the collected company reviews are preprocessed to remove noise, perform text normalization, and handle other preprocessing tasks. This includes tokenization, removing stopwords, and potentially applying stemming or lemmatization techniques to standardize the text data.

Next, relevant features are extracted from the preprocessed reviews. This may involve techniques like Bag-of-Words, TF-IDF, or word embeddings to represent the textual data as numerical features that can be processed by the Naive Bayes classifier. Once the features are extracted, the Naive Bayes classifier is trained using the preprocessed and feature-engineered data. During the training process, the classifier estimates the probabilities of different features given each sentiment category (positive, negative, neutral), leveraging the Naive Bayes assumption of feature independence.

After training, the performance of the Naive Bayes classifier is evaluated using a validation set. Performance metrics such as accuracy, precision, recall, or F1 score are calculated to assess the classifier's effectiveness in predicting the sentiment of unseen reviews. Once the classifier is trained and evaluated, it is used for sentiment analysis on new and unseen company reviews. The classifier predicts the sentiment category (positive, negative, or neutral) of each review based on the extracted features and the learned probabilities from the training phase.

The sentiment analysis results are integrated into the web application. The predicted sentiment for each company review is displayed on the company profile page, providing users with insights into the overall sentiment expressed by employees or candidates about the company. By leveraging Naive Bayes classification, we can effectively categorize the sentiment of company reviews, allowing users of our website to make informed decisions based on the sentiment analysis results. The simplicity and efficiency of Naive Bayes make it a suitable choice for sentiment analysis in our project.




## Reviews classification by using ML,NLTK:

MINI\_PROJECT


HOMEABOUTRECRUITMENTREVIEWS

REVIEWS




Positive

- > nice place manage personal life work
- > great place work
- > please provide better hikes
- > good start working freshers
- > good
- > best company



Negative

- > dont expect much increments terms percentage
- > much put
- > see con
- > advice management
- > proper hikes salary
- > rt manager obev evrthing anvthing client



Neutral

- > challenging work freshers learn
- > slow career growth scope learning
- > good company pressure work life balance
- > offer good pay need get initial
- > management administration site opportunities projects
- > worth work

## Additional features like Salaries:

Designation	Average Salary
Software Engineer Experience: 0 - 6 yrs	₹ 5,23,817/year (₹2.2L/yr - ₹12.0L/yr)
Technical Lead Experience: 4 - 13 yrs	₹ 14,06,349/year (₹6.3L/yr - ₹29.5L/yr)
Lead Engineer Experience: 2 - 11 yrs	₹ 8,50,210/year (₹4.5L/yr - ₹16.0L/yr)
Senior Analyst Experience: 1 - 8 yrs	₹ 4,74,106/year (₹2.1L/yr - ₹10.0L/yr)
Analyst Experience: 0 - 8 yrs	₹ 3,91,462/year (₹2.0L/yr - ₹8.0L/yr)
Senior Software Engineer Experience: 2 - 9 yrs	₹ 9,66,725/year (₹4.0L/yr - ₹21.0L/yr)



## 5.2 Limitations:

**Certainly! Here are the limitations of our project in a concise format:**

- 1. Subjectivity and Context:** Difficulty capturing nuances and context in sentiment analysis.
- 2. Incomplete Dataset:** Reliance on limited or biased data may impact accuracy.
- 3. Lack of Real-Time Updates:** Dataset may not reflect the most current sentiments.
- 4. Dependency on User-Generated Content:** Authenticity and bias of user reviews can affect reliability.
- 5. Limited Scope of Sentiment Analysis:** Other important factors beyond sentiment may influence career choices.
- 6. Scalability and Maintenance:** Managing a growing dataset and ensuring long-term accuracy.
- 7. Language and Cultural Bias:** Potential bias towards certain languages or cultures in sentiment analysis.

## CHAPTER 6

### CONCLUSION

#### 6.1 Summary:

The project aims to create a college website that provides valuable information about companies to help students make informed career choices. It incorporates NLP, FullStack, and ML techniques, including sentiment analysis. The project collects company reviews, trains a sentiment analysis model using ML algorithms, and presents additional features like company packages and genuine reviews on the website. The project acknowledges the limitations of sentiment analysis algorithms and aims to continuously improve accuracy. The long-term goal is to assist students in making better career decisions through the comprehensive information provided on the platform.

#### 6.2 Future Work:

##### 1. Making the Project Dynamic:

- Introduction of real-time updates for current and up-to-date information.
- Regular fetching and updating of company reviews from various sources.
- Ensuring that sentiment analysis and company information reflect the latest data.

##### 2. Collecting Reviews from Seniors and Storing in the Database:

- Incorporating firsthand experiences from senior students or professionals.
- Storing reviews in a database for easy retrieval and integration.
- Providing users with valuable insights from seniors related to each company.

##### 3. Adding Additional and More Accurate Features about Each Company:

- Enhancing the project with comprehensive information for decision-making.
- Including details on career growth, benefits, work culture, and achievements.
- Incorporating updated recruitment processes and employee testimonials.

#### 6.3 References:

- Each company reviews are manually collected from Glass door and ambition box websites
- visit:[https://www.glassdoor.co.in/Reviews/index.htm?overall\\_rating\\_low=3.5&page=1&locId=2940587&locType=C&occ=Software%20Engineer](https://www.glassdoor.co.in/Reviews/index.htm?overall_rating_low=3.5&page=1&locId=2940587&locType=C&occ=Software%20Engineer) (Glass door)
- visit:<https://www.ambitionbox.com/reviews> (Ambition box)
- To understand sentiment analysis we watched some youtube videos
- visit:<https://youtu.be/SMecj9AIDF8>





