

**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**

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*

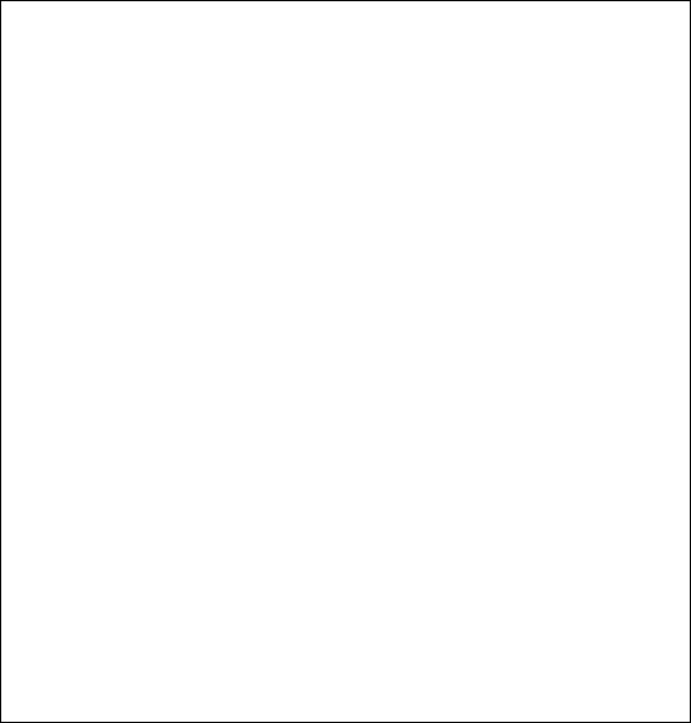
**Mr.**

**KALAPALA**

**SRAVAN**

**KUMAR**

1



**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

rd

3

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**

**Mr.Chiranjeevi**

**Sadu**

Assistant

professor,

Assistant

Professor,

Department

of

CSE,

Head

of

the

Department,

RGUKT,

Nuzvid

Department

of

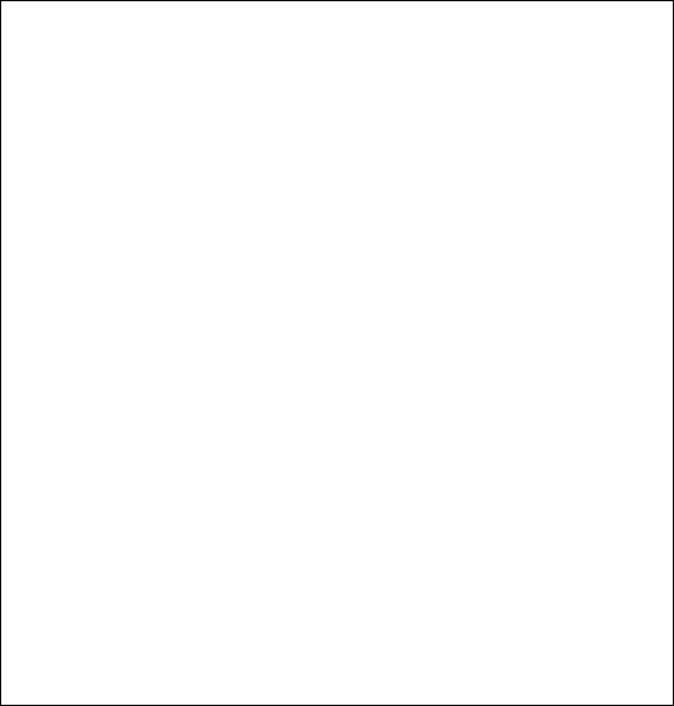
CSE,

RGUKT,

Nuzvid.



2



**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

rd

3

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**

**Project**

**Examiner**

Assistant

Professor,

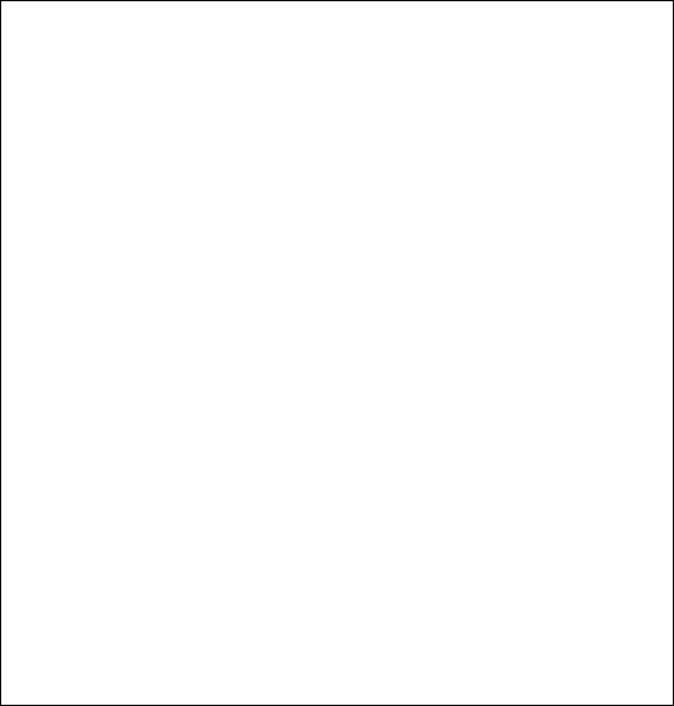
RGUKT-Nuzvid.

Department

of

CSE,

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

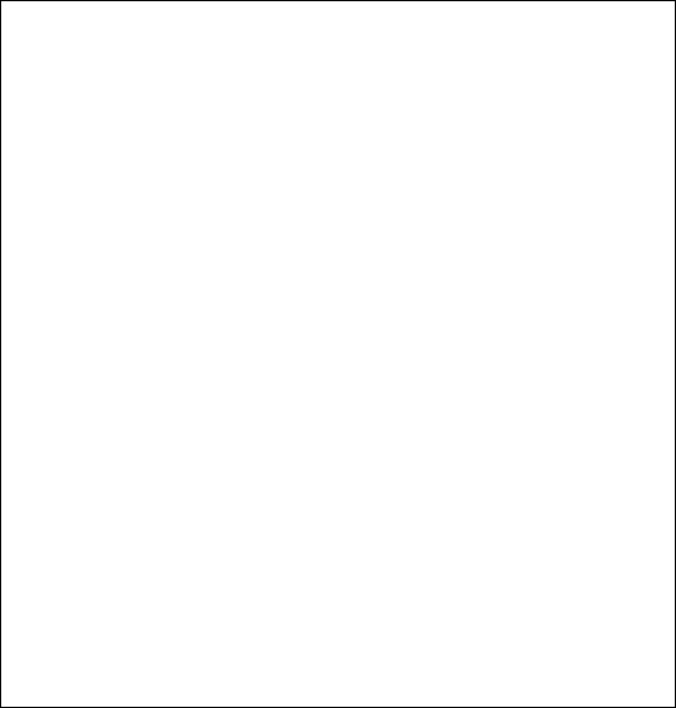
(N180798)

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

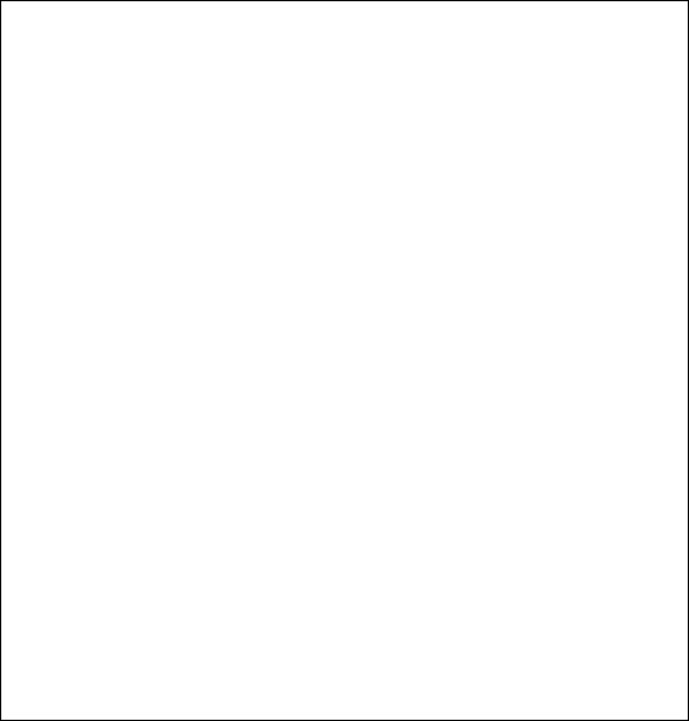
successful

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

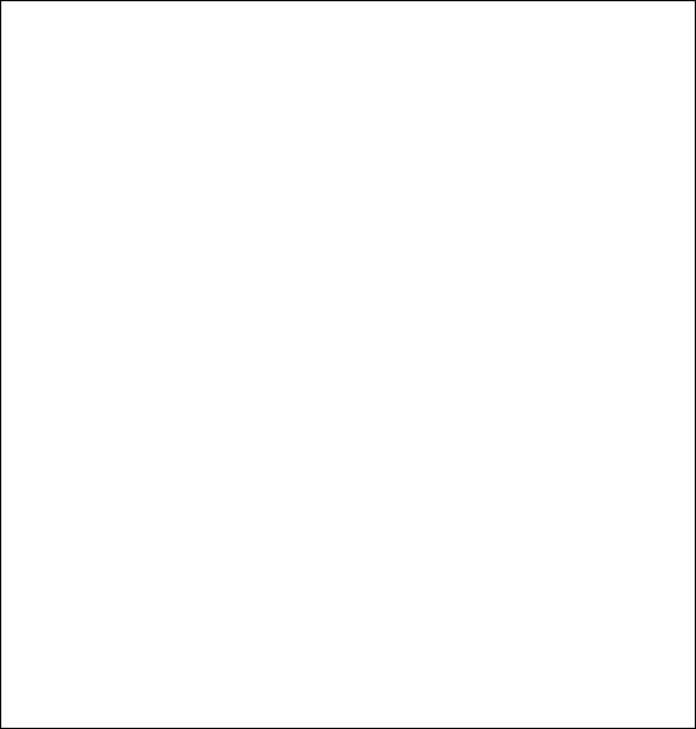
career

choices

through

this

Project.



**Table**

**of**

**Contents**

**CHAPTER**

**1**

8

**INTRODUCTION**

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

**CHAPTER**

**2**

12

**REQUIREMENTS**

**AND**

**ANALYSIS**

12

2.1

Hardware

components

12

2.2

Software

components

13

**CHAPTER**

**3**

14

**PROPOSED**

**MODEL**

**AND**

**FLOW**

**OF**

**THE**

**PROJECT**

14

3.1

Proposed

model

14

3.2

Flow

of

the

project

14

3.3

Advantages

15

3.4

Applications

16

**CHAPTER**

**4**

18

**IMPLEMENTATION**

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

**CHAPTER**

**5**

23

**RESULT**

5.1

Performance

of

the

model

5.2

Limitations

**CHAPTER**

**6**

**CONCLUSION**

6.1

Summary

6.2

Future

Work

**6.3**

References

23

23

25

25

26

26

26

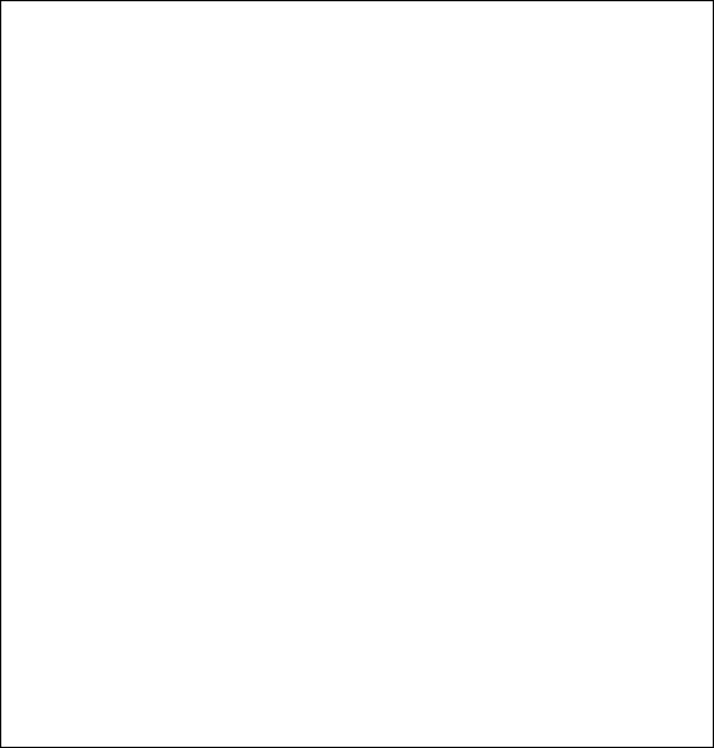
26

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.

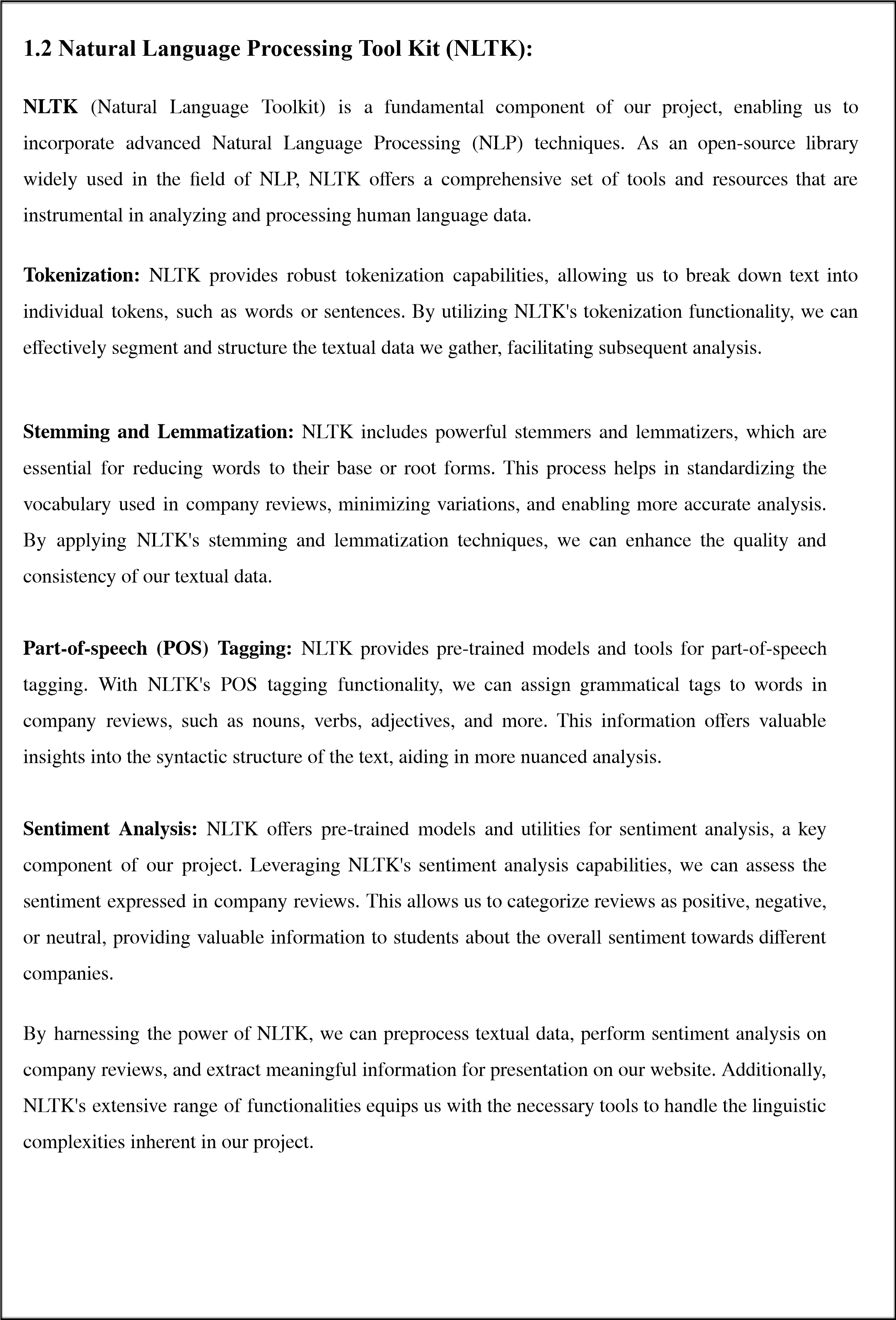
The project acknowledges the limitations inherent in sentiment analysis algorithms and the potential lack of context, which can affect the accuracy of results. However, the primary objective is to empower students in their career decision-making process by presenting them with authentic reviews, company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

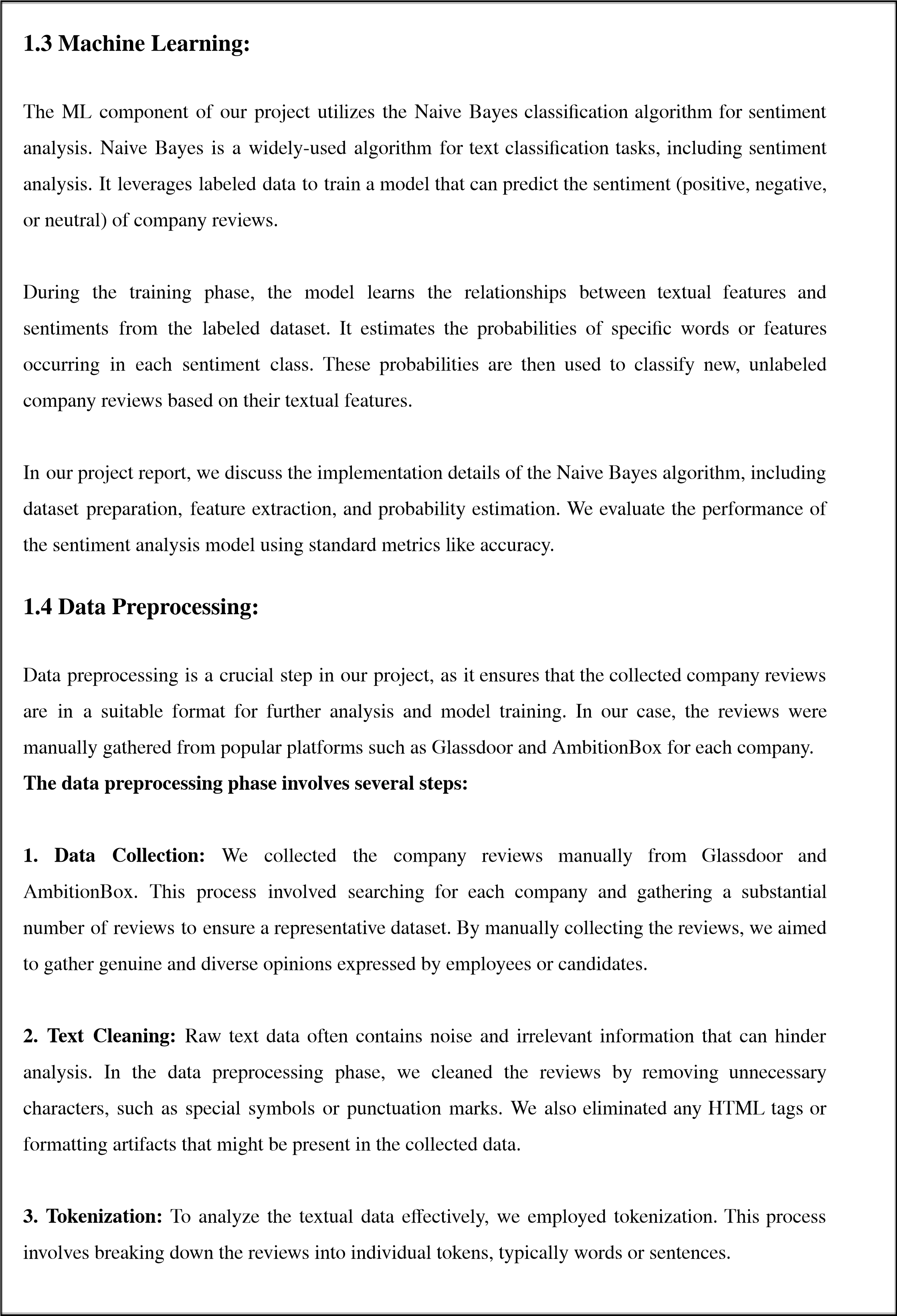
Key features of the project include data collection from company reviews, ML algorithm-based sentiment analysis model training, and the development of a user-friendly web interface. The NLP techniques employed enable the project to analyze the sentiment expressed in reviews, providing students with an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

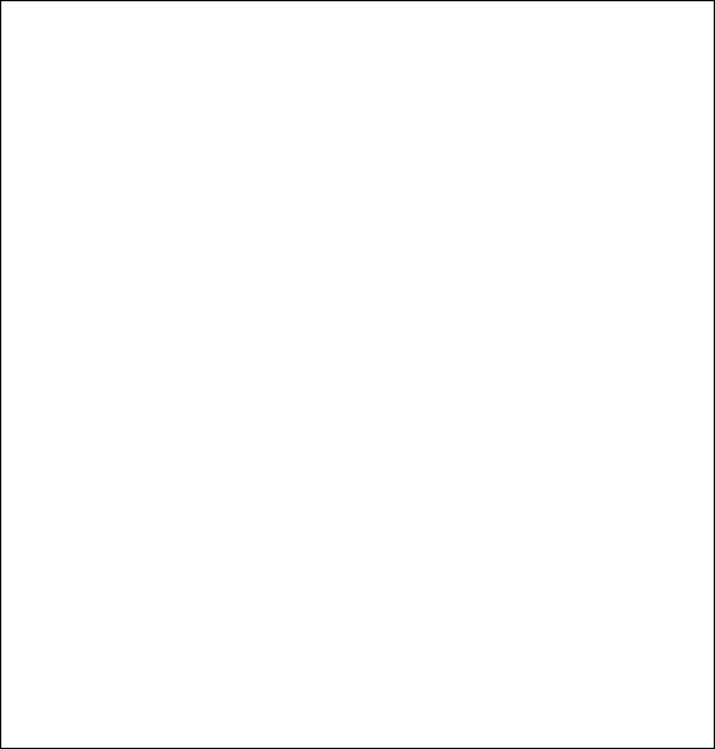
The project report will encompass various chapters, starting with an introduction that highlights the project's significance and objectives. It will delve into the technologies utilized, including NLTK and ML, and emphasize the importance of data preprocessing. The proposed model and project flow will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







**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.

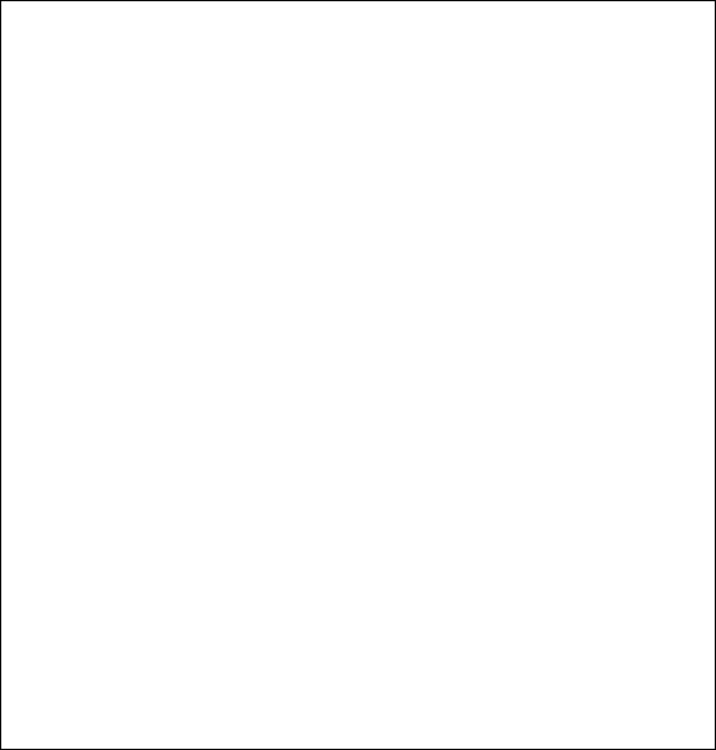
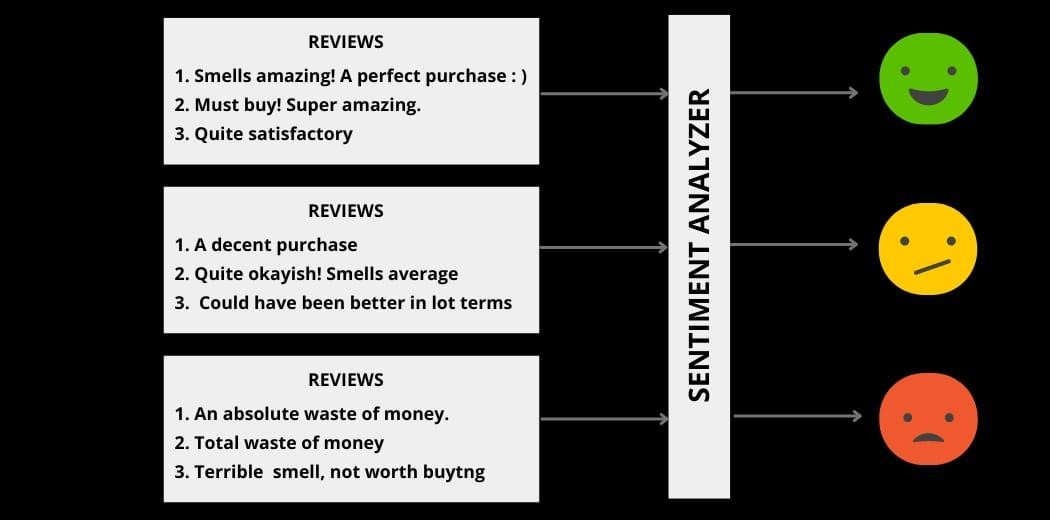
*figure.1*

*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.

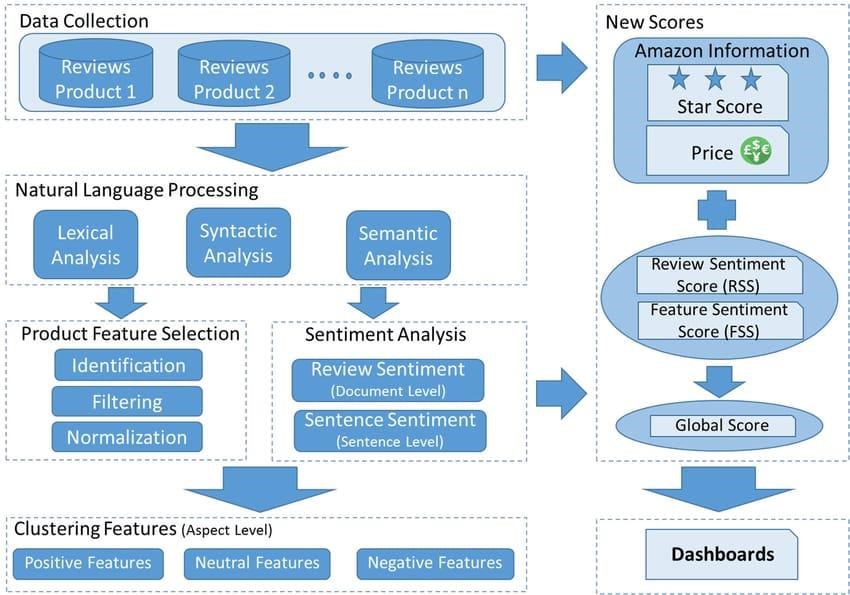
•

Keyboard:

A

standard

keyboard



**2.2 Software components:**

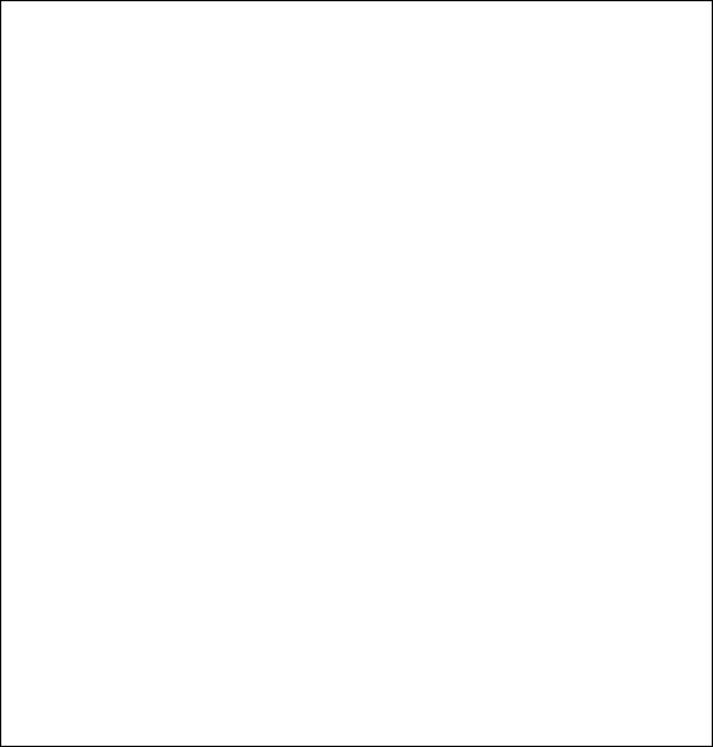
1. **XAMPP:** XAMPP is a software package that provides a local development environment for web applications. It includes Apache as the web server, MySQL as the database management system, and PHP as the programming language. XAMPP simplifies the setup and configuration of the development environment,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

1. **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.

1. **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.

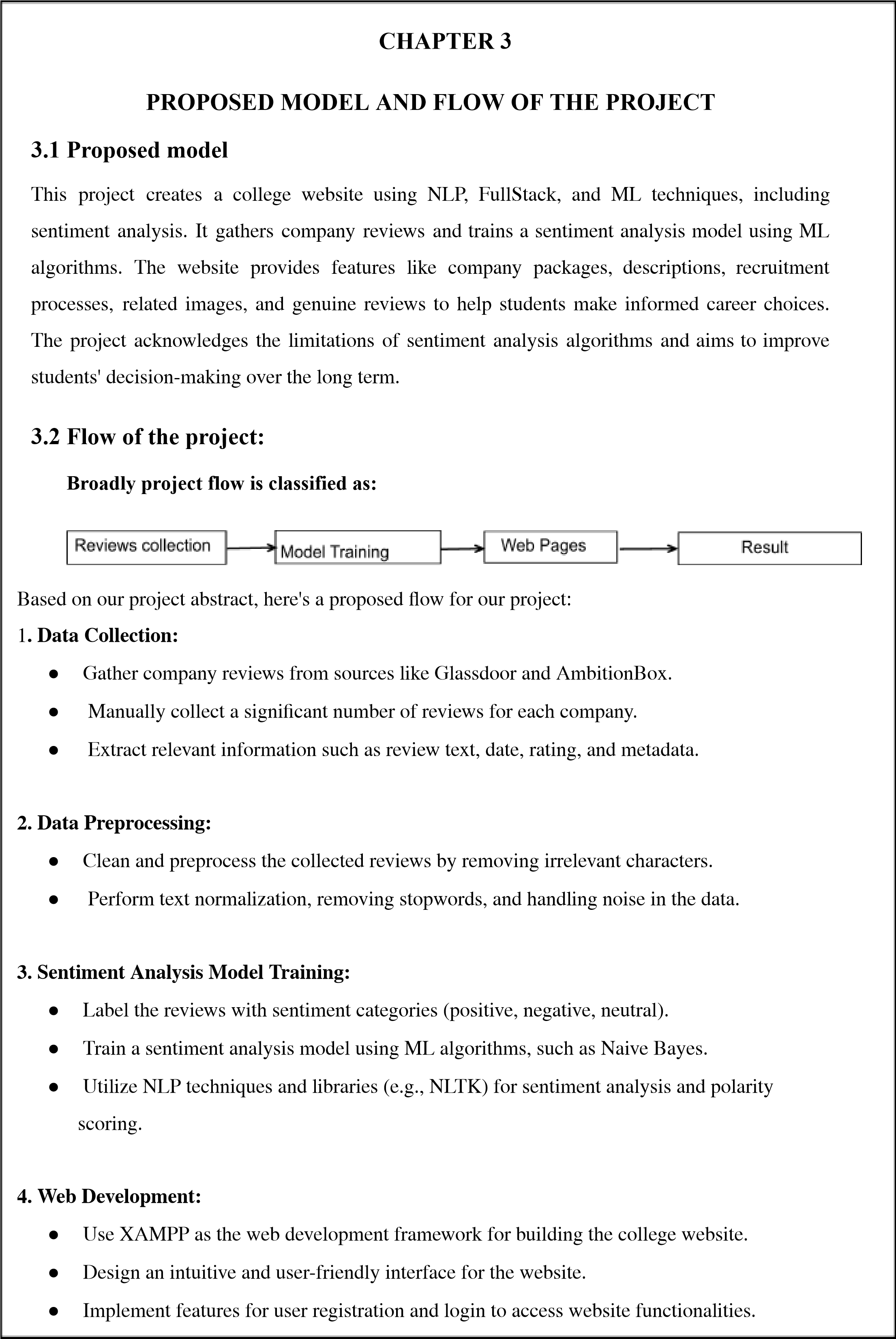
1. **MySQL:** MySQL is a widely used open-source relational database management system. It allows you to store, manage, and retrieve the collected company reviews, related information, and user data. MySQL provides a robust and efficient way to handle data persistence in your project.

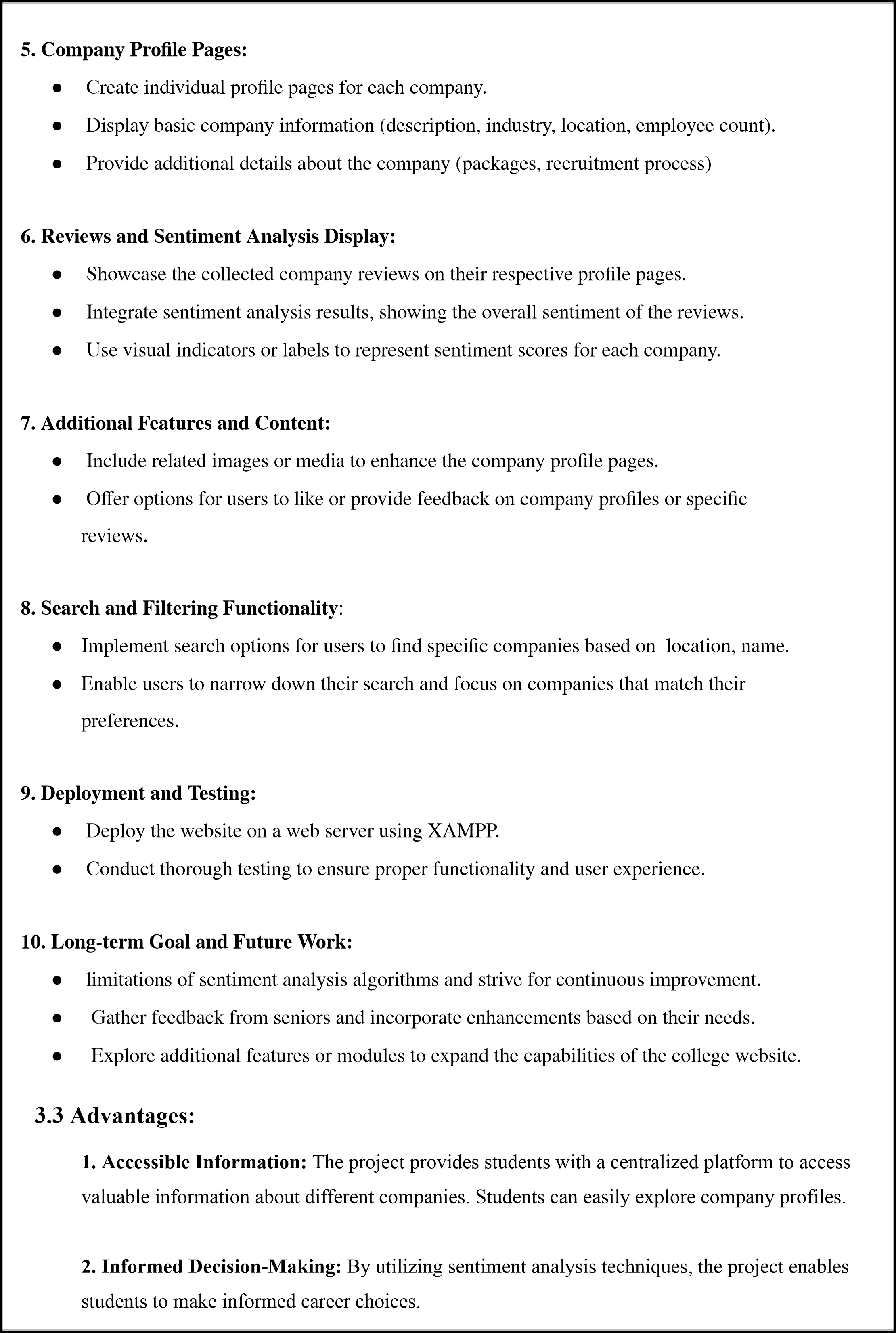
**VADER Sentiment Analysis:** VADER is a rule-based sentiment analysis tool specifically designed for sentiment analysis of social media texts. It uses a combination of lexicons and grammatical rules to assess the sentiment expressed in text. VADER assigns sentiment scores to individual words and calculates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

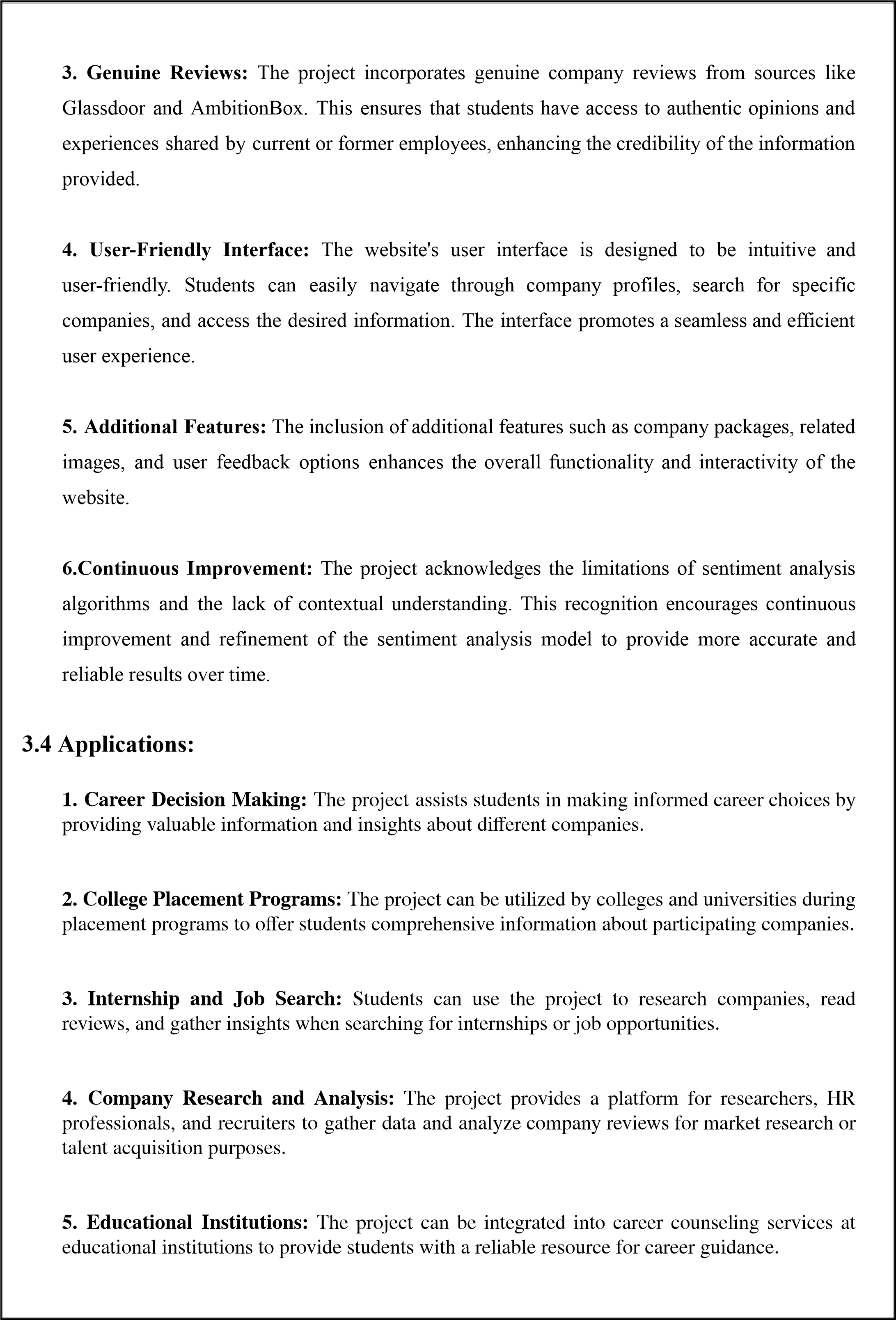
 **NLP Libraries (e.g., NLTK):** NLP libraries are crucial software components for text processing and analysis tasks. Libraries like NLTK (Natural Language Toolkit) or SpaCy provide functionalities such as tokenization, stemming, lemmatization, part-of-speech tagging, and sentiment analysis. These libraries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

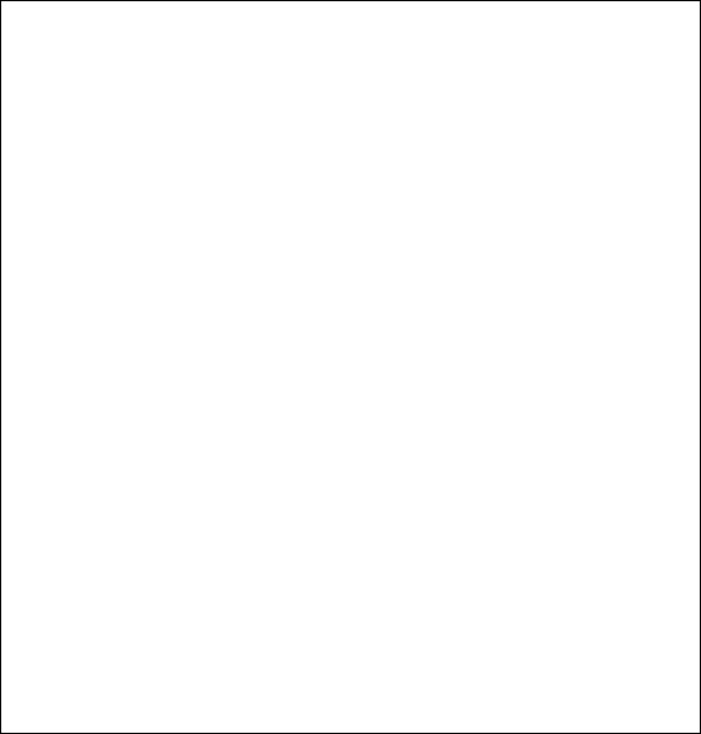
**Machine Learning Libraries :** ML libraries are utilized for training the sentiment analysis model. Libraries like Scikit-learn or TensorFlow provide implementations of ML algorithms, including Naive Bayes, which is employed for sentiment analysis in your project. These libraries facilitate the training,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









**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

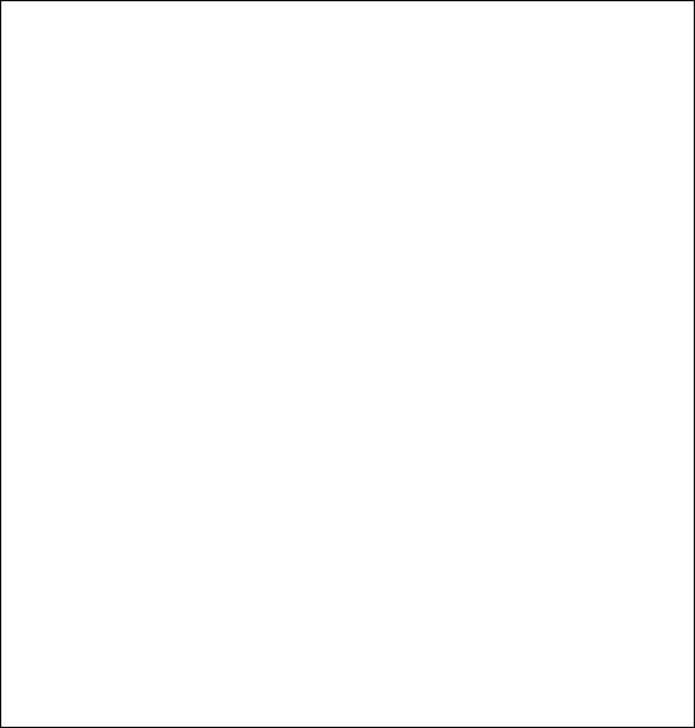
techniques

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

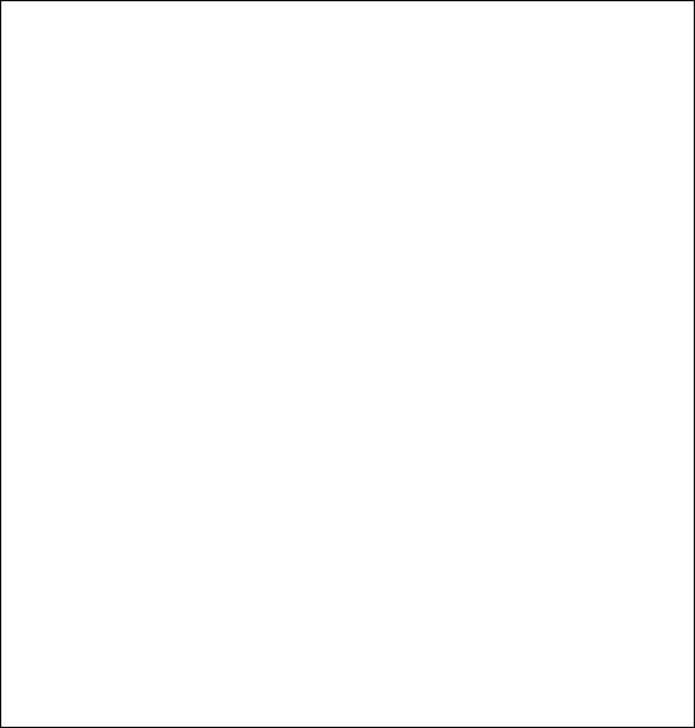
web

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.

*figure*

*4:*

*NLTK*

*installation*

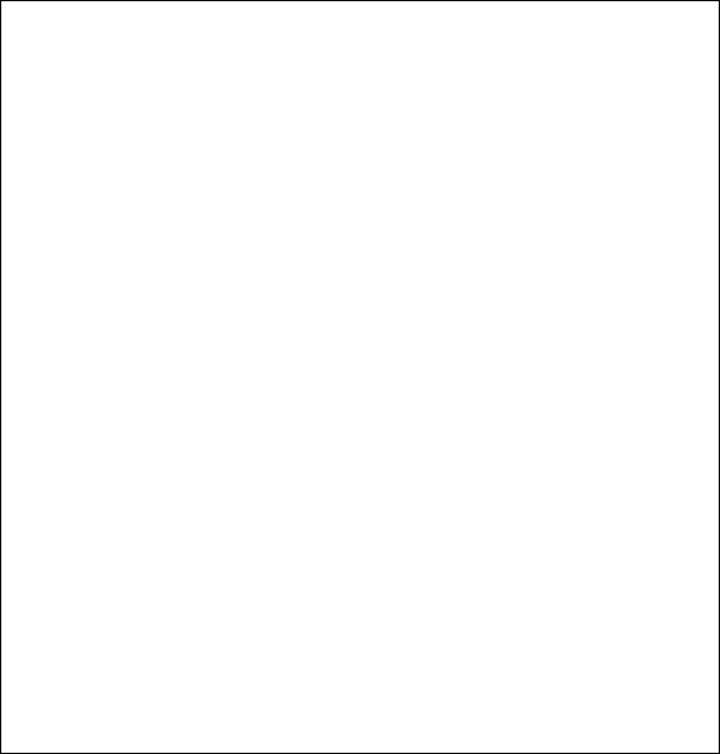
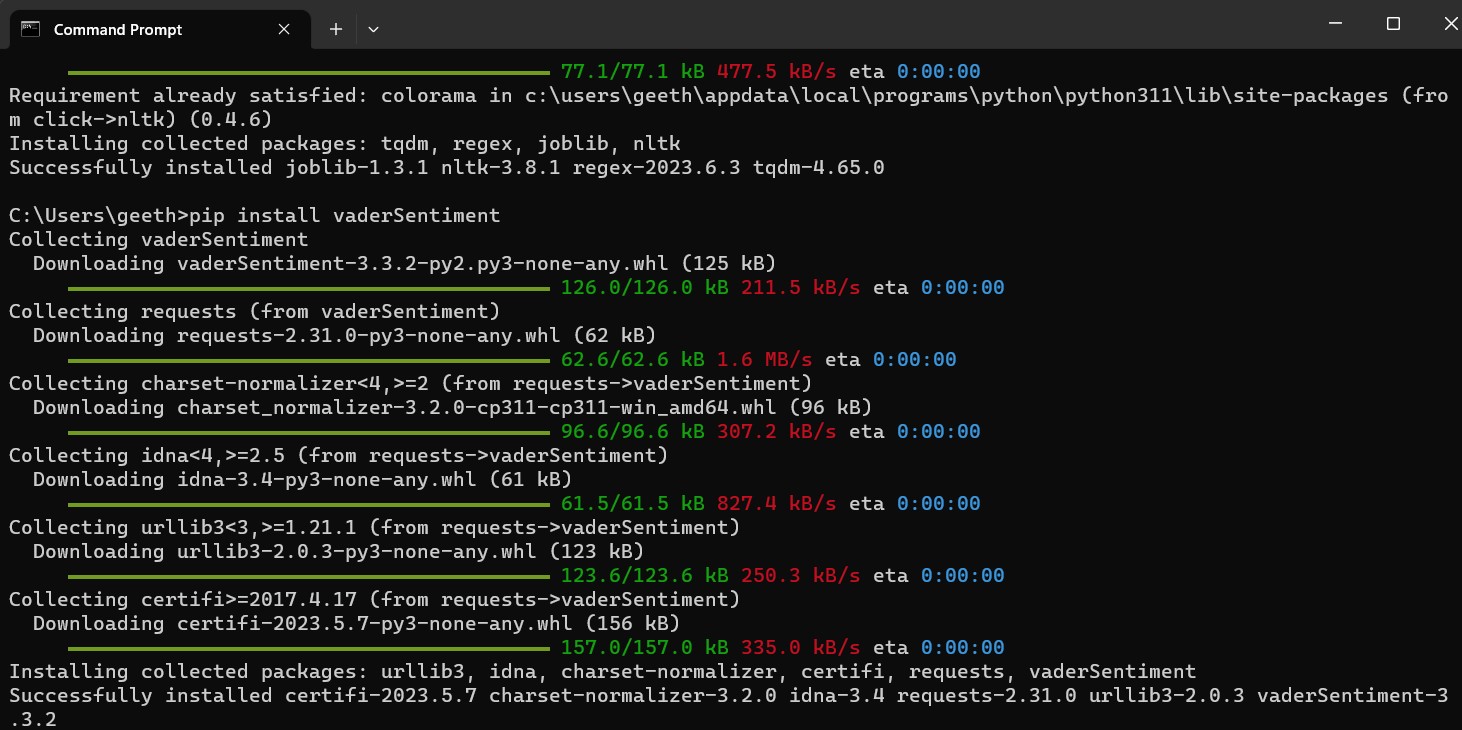
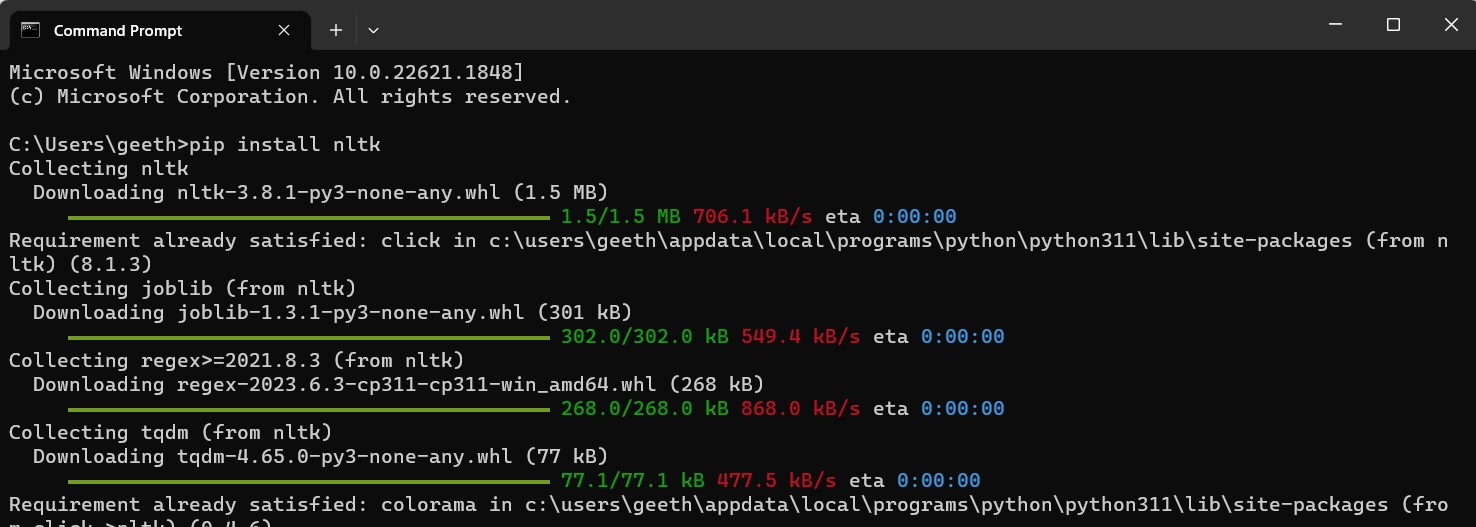
*figure*

*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.

**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

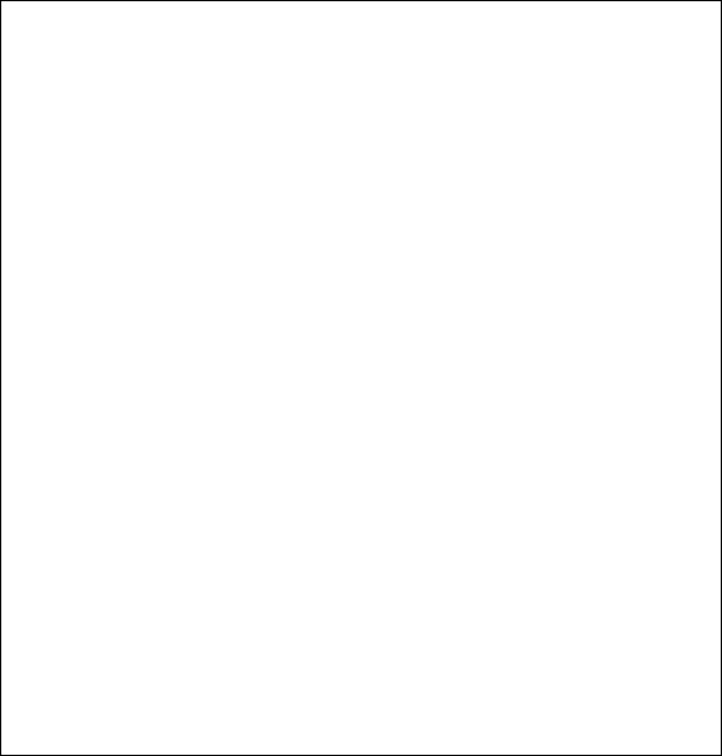
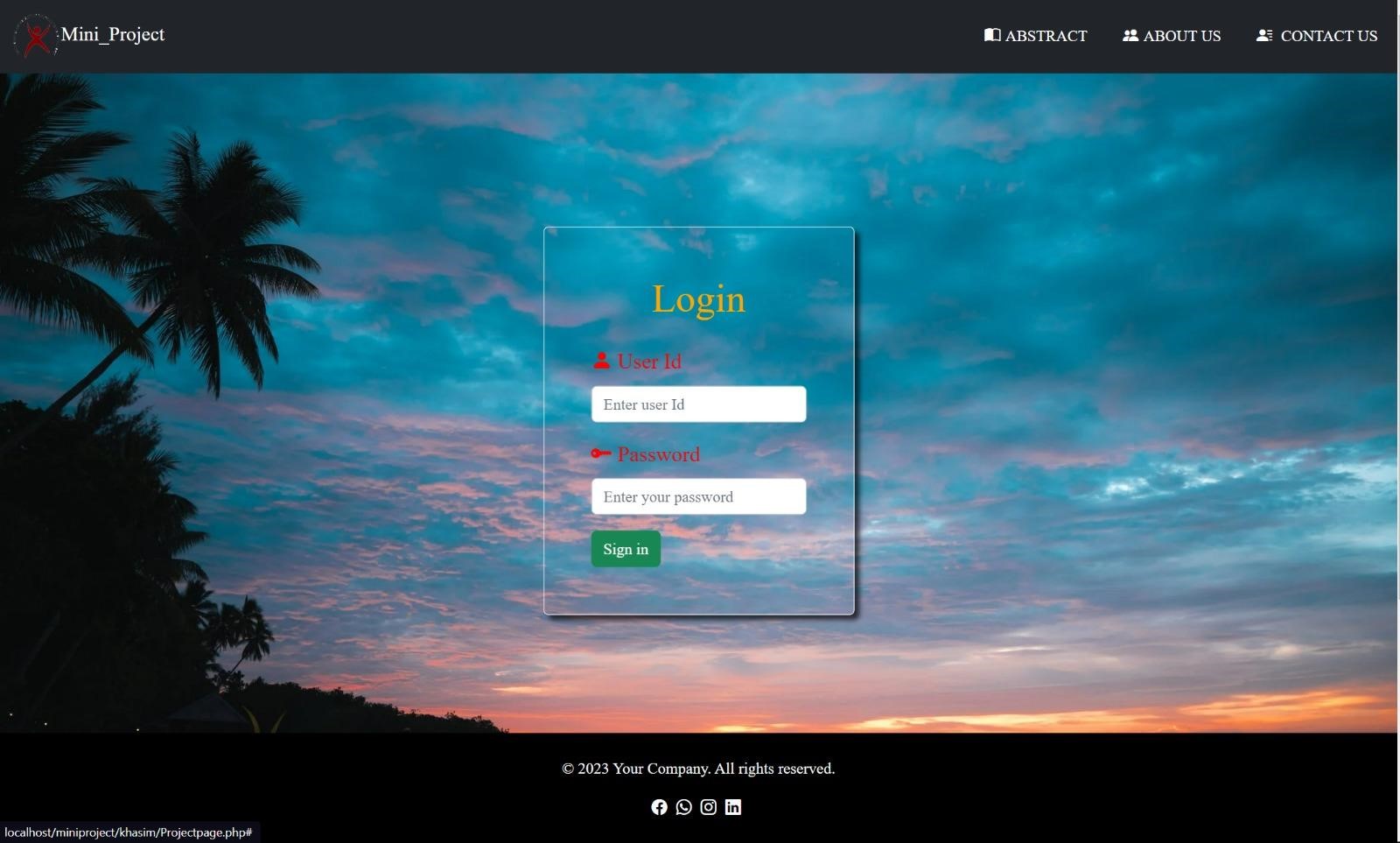
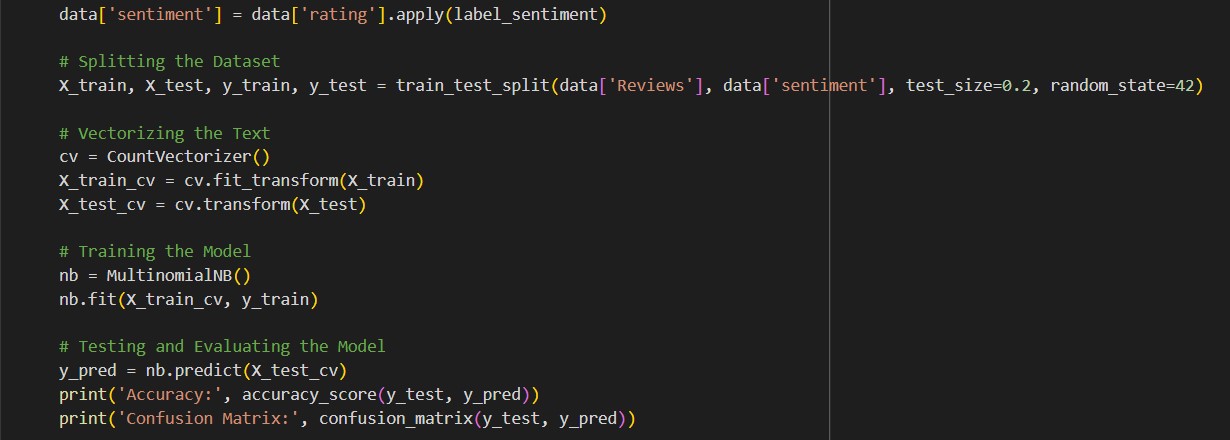
any

additional

features.

**Login**

**Page:**



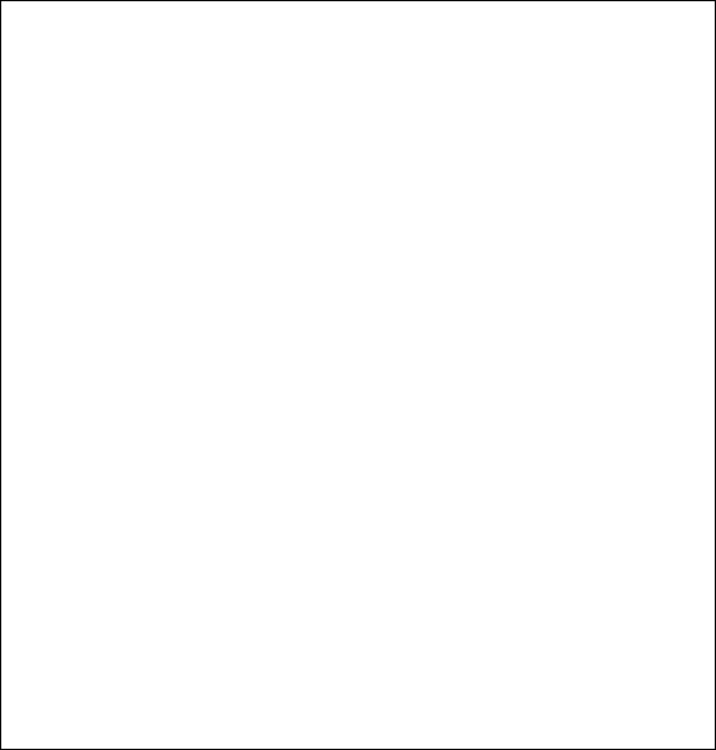
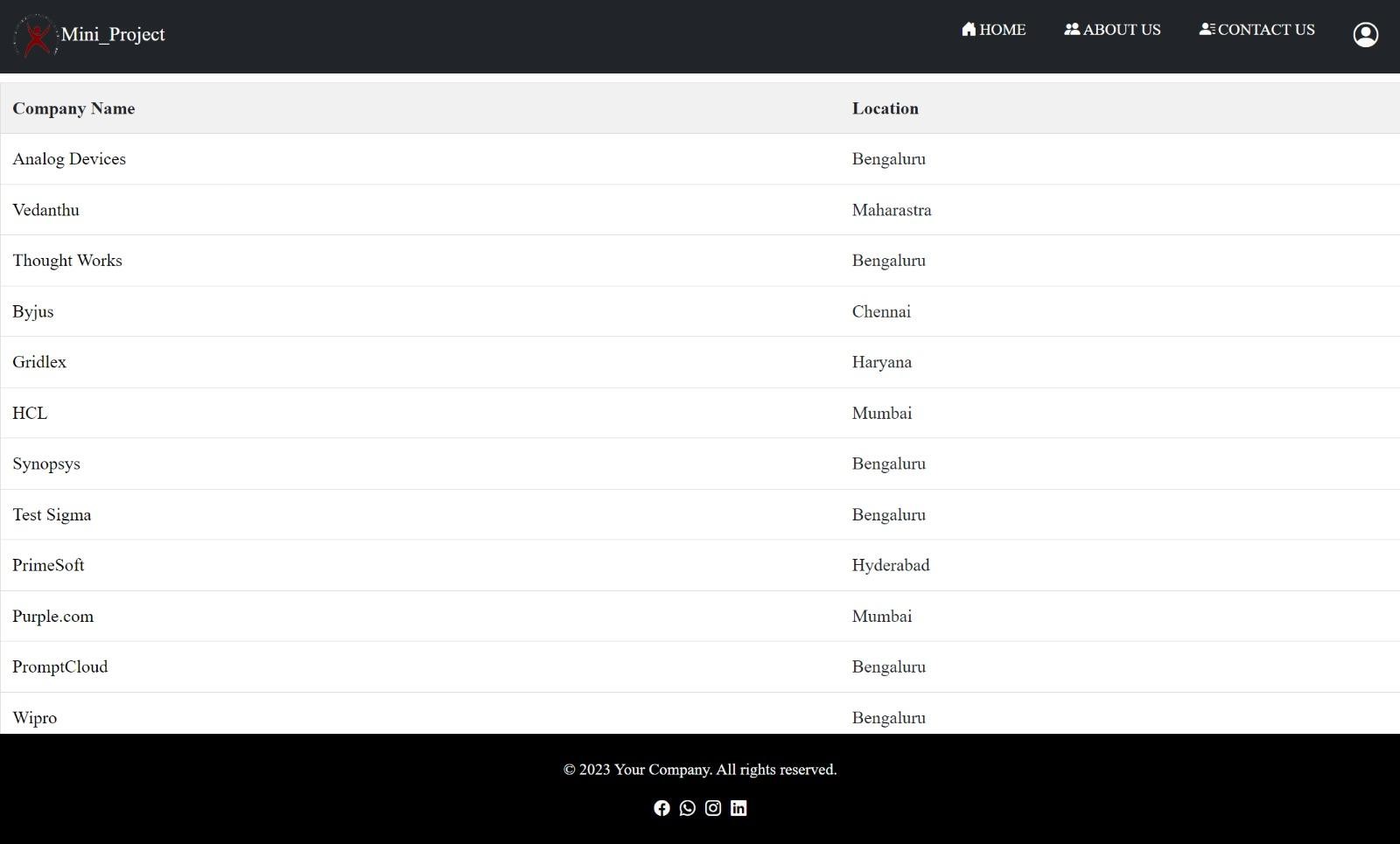
**Search**

**Page:**

**Example**

**company**

**page:**



**About**

**us**

**Page:**

**Reviews**

**Classification:**

These

are

the

steps

we

followed

during

the

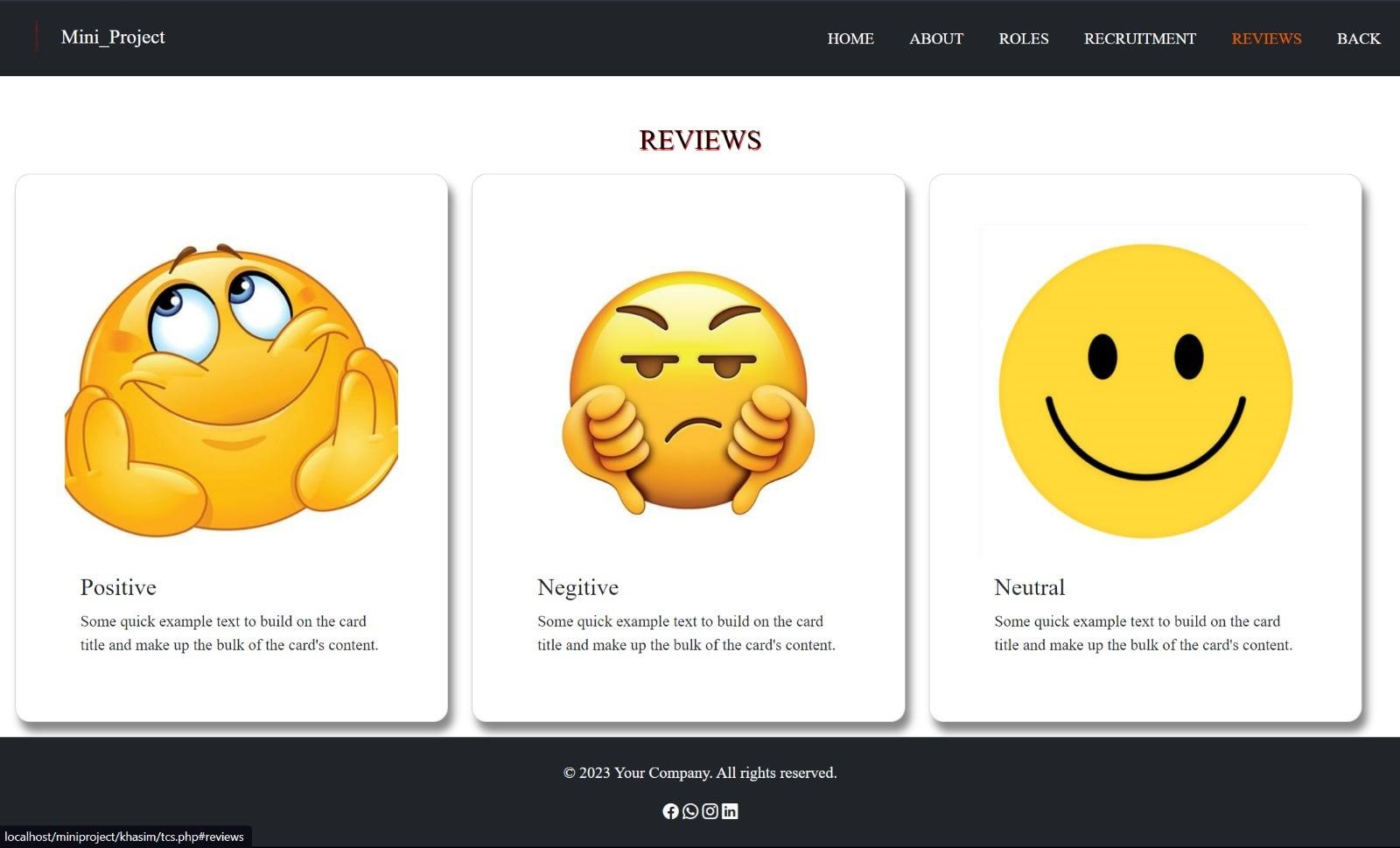
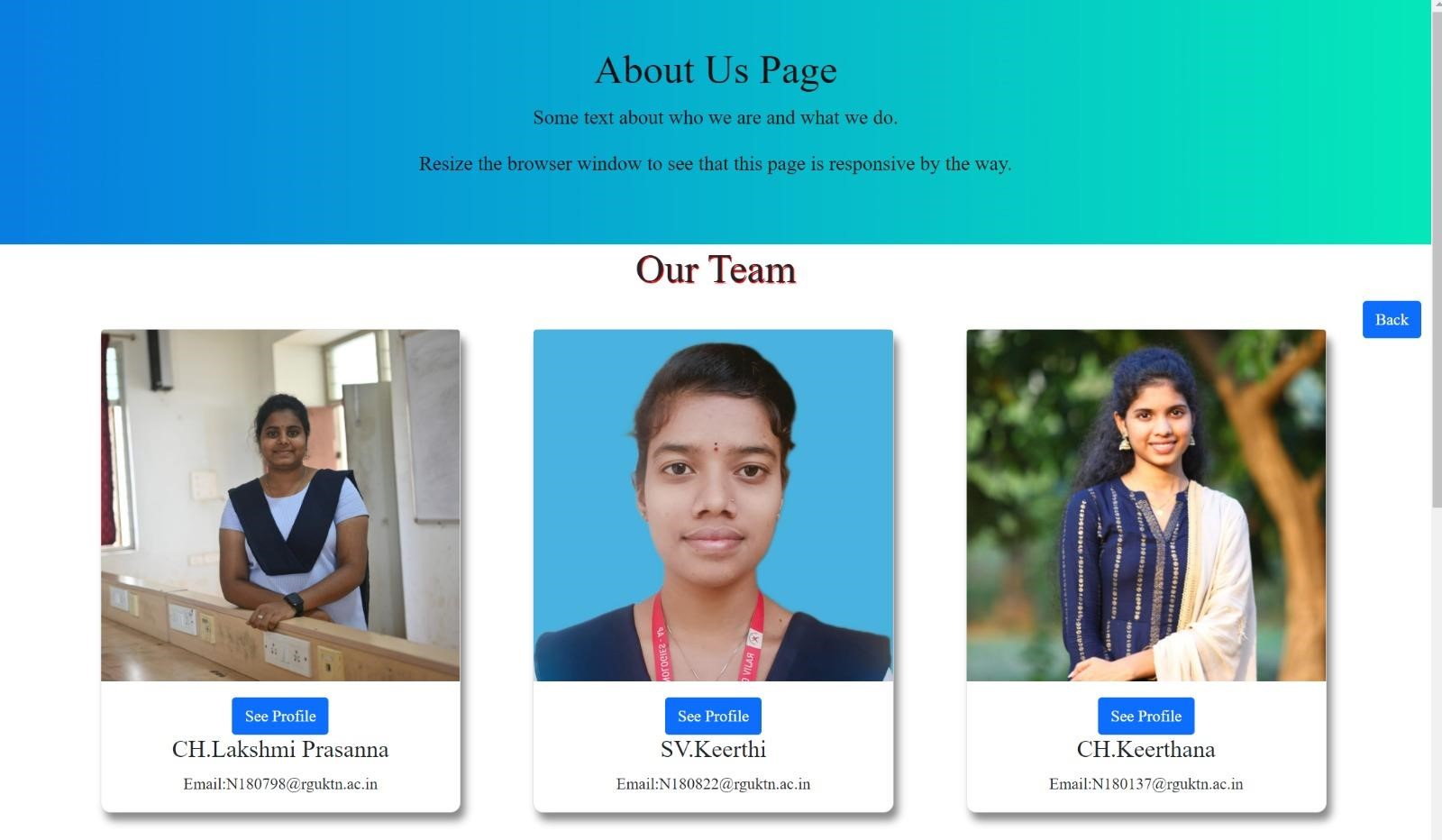
implementation

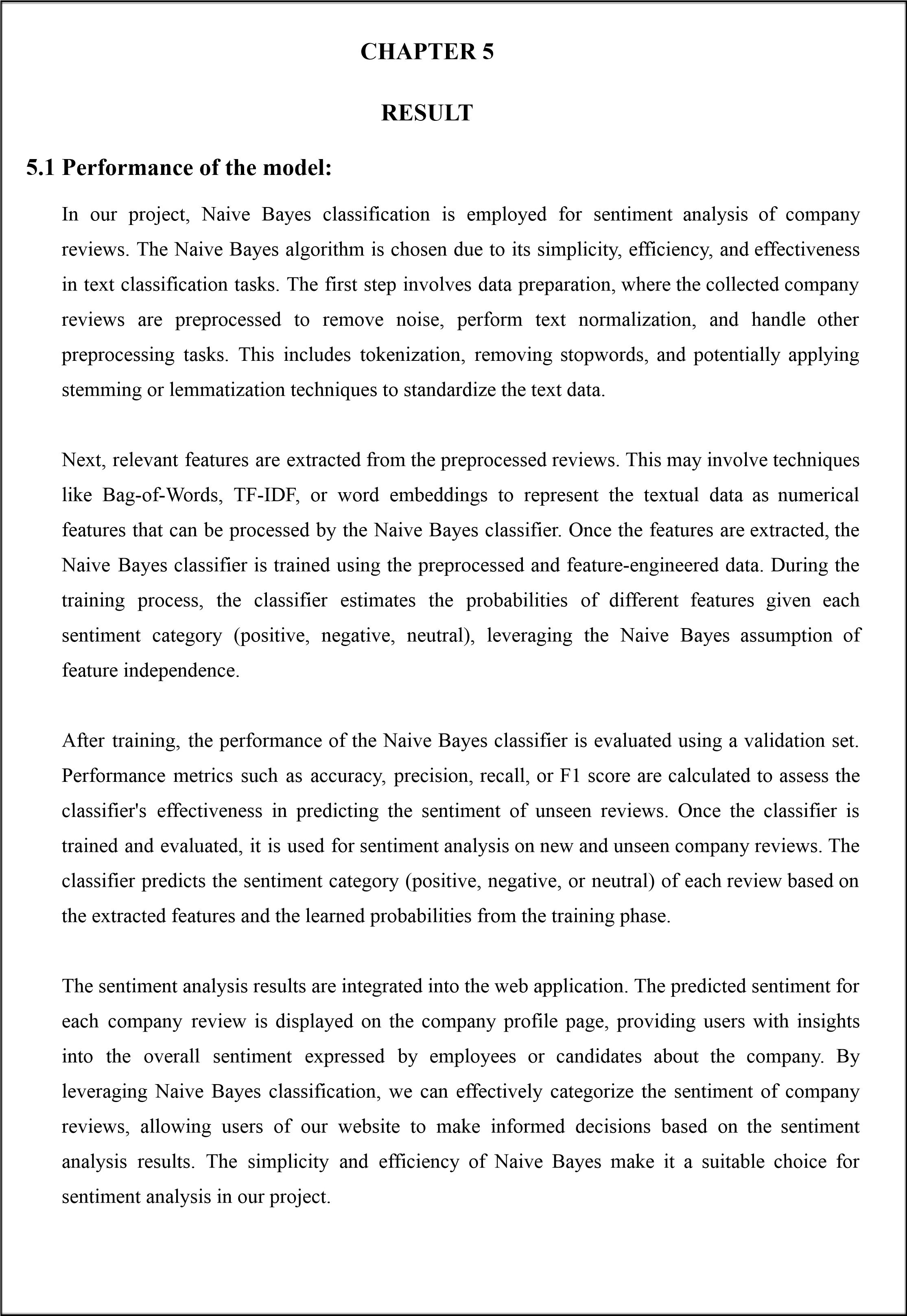
of

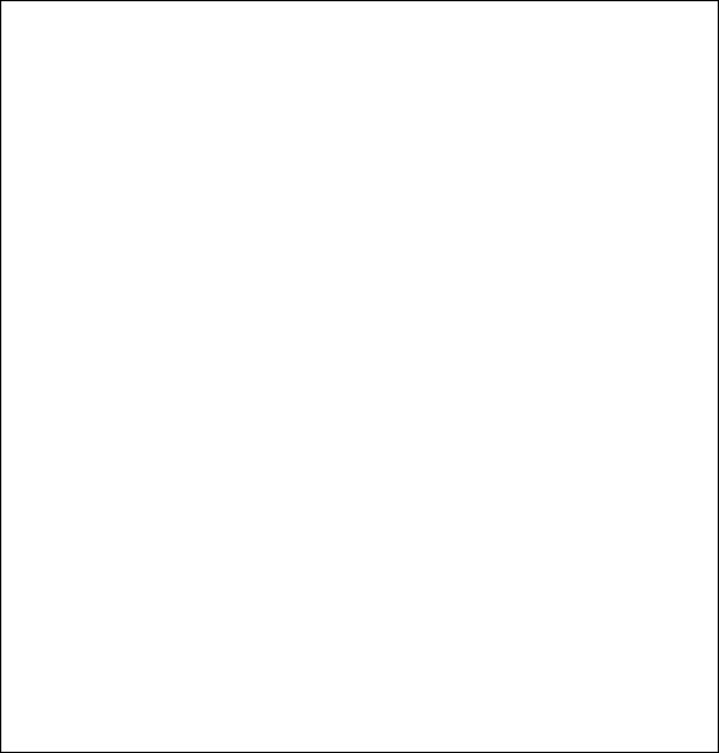
our

proje

t







**Reviews**

**classification**

**by**

**using**

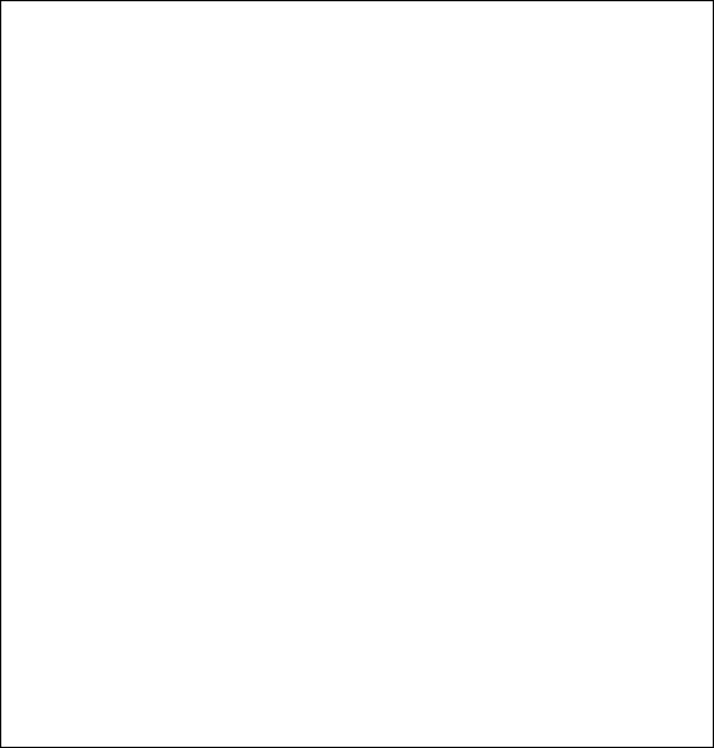
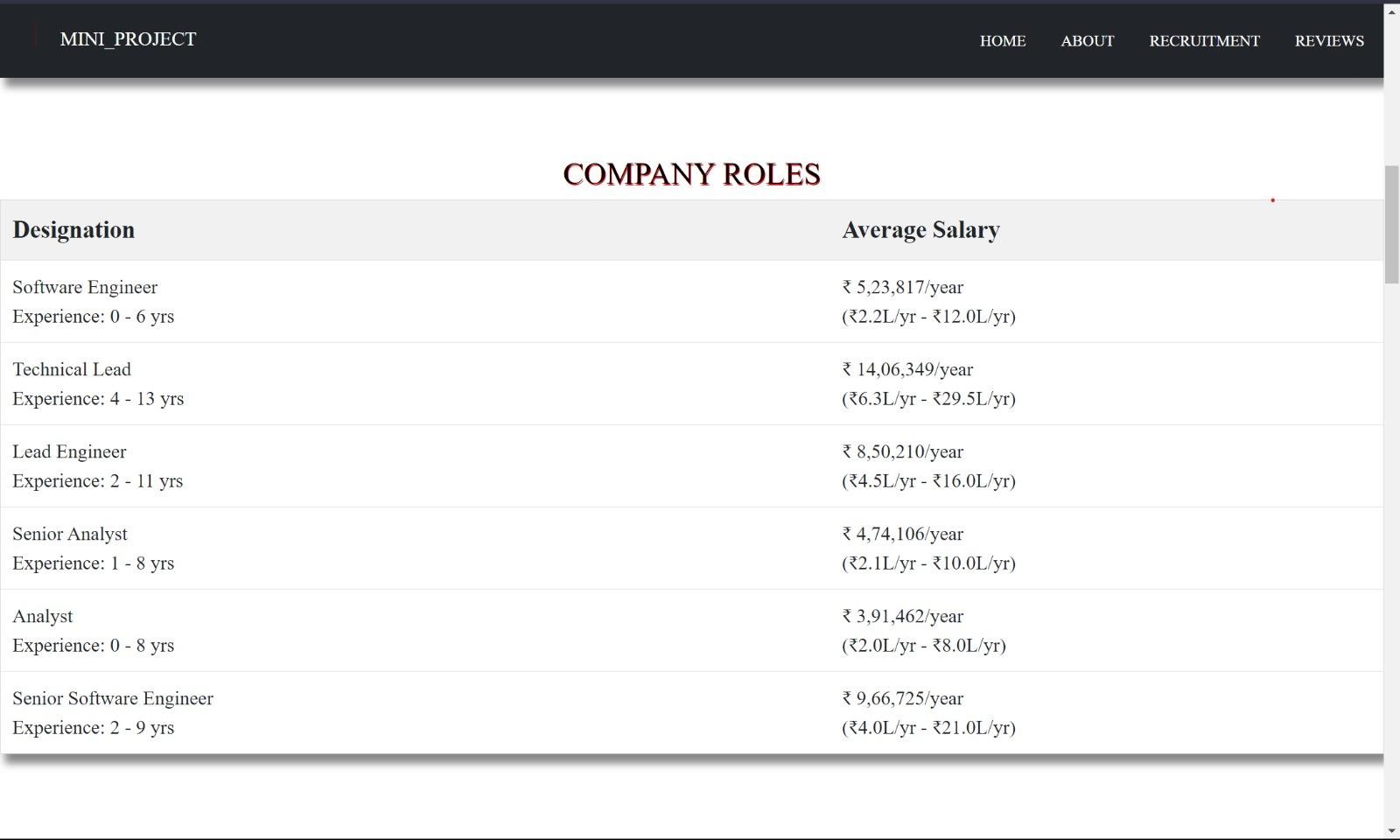
**ML,NLTK:**

**Additional**

**features**

**like**

**Salaries:**



**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

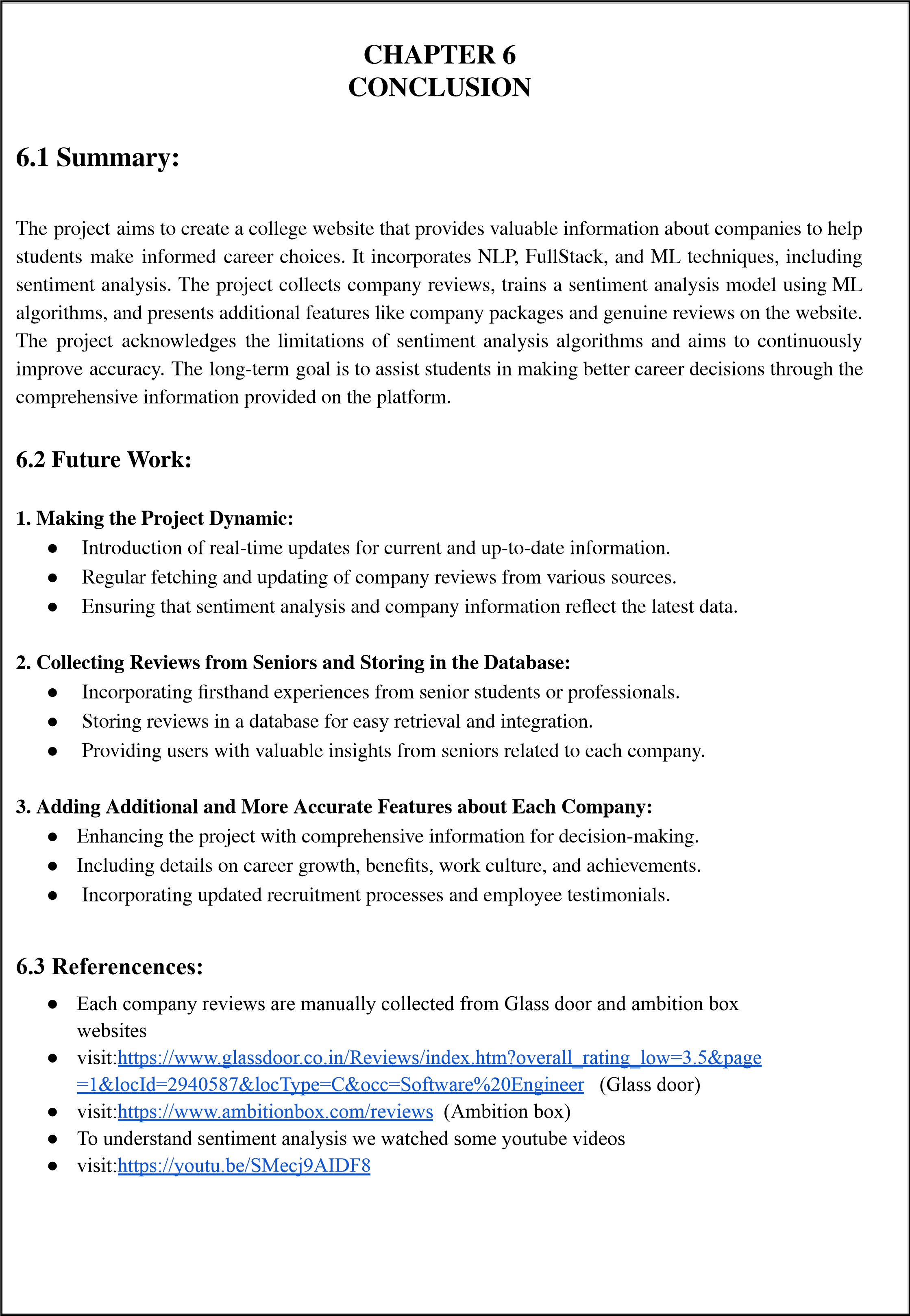
or

cultures

in

sentiment

analysis.



27 28 29