VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Social Media Sentiment Analysis"

Submitted in partial fulfilment for the award of degree(18CSI85)

Bachelor of Engineering
In
Computer Science

Submitted by: **Sushil Kumar B(1BO20CS102)**

Conducted at
Varcons Technologies Pvt Ltd

Department of Computer Science and Engineering

Brindavan College of Engineering

Dwarkanagar, Bagalur, Main Road, Yelahanka, Bengaluru -560063 Affiliated to VTU Belgavi, Approved by AICTE, New Delhi, India Accredited 'A'level by NAAC

Brindavan College of Engineering

Department of Computer Science and Engineering

Brindavan College

CERTIFICATE

This is to certify that the Internship titled **"Social Media Sentiment Analysis"** carried out by **Sushil Kumar B,** a bonafide student of **Brindavan College of Engineering** , in partial fulfillment for the award of **Bachelor of Engineering**, in **Computer Science Engineering** under Visvesvaraya Technological University,Belagavi, during the year 2023-2024. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

_____ _____

Signature of Guide **Signature of HOD**

External Examiners:

Name of the Examiner Signature with Date

1)_____ _____

2)_____ _____

D E C L A R A T I O N

I **Sushil Kumar B**, final -year student of Computer Science Engineering**, Brindavan College of Engineering -560064**, declare that the Internship has been Successfully completed, **Varcons Technologies pvt ltd**. This report is submitted in partial fulfillment of the requirements for award of Bachelor degree in computer science, during the academic year 2023-2024.

DATE : _____ **:**

PLACE : _____

USN : _____

NAME : _____

OFFER LETTER

INTERNSHIP OFFER LETTER

Date: 11th August, 2023

Name: **Sushil Kumar B**
USN: **1BO20CS102**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python (Research Based)** Internship position with **Varcons Technologies**, effective Start Date **11th August, 2023**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python (Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!.

Sincerely,

Spoorthi H C
Director
VARCONS TECHNOLOGIES
213, 2st Floor,
18 M G Road, Ulsoor,
Bangalore-560001

A C K N O W L E D G E M E N T

This Internship is a result of accumulated guidance, direction and support of Certify Training Team.

We take this opportunity to express our gratitude to **mentor Tejas sir** who have helped us to complete the Internship.

I would like to express my sincere gratitude for the invaluable guidance and support you have provided during my internship at Varcons Technology Pvt. Limited.

Your role as a Certified Team Assistant has been instrumental in my professional development, and I am truly thankful for the opportunity to learn and grow under your mentorship.

Throughout my internship, your expertise and dedication to the team's success were evident in every interaction and task we undertook.

Your willingness to share your knowledge and experience has been a tremendous asset to my learning journey, and I have gained insights that will undoubtedly shape my future career.

Sushil Kumar B
1BO20CS102

ABSTRACT

Machine learning, a fundamental component of data science, has witnessed a remarkable resurgence in recent years, propelled by the accessibility and versatility of Python as a programming language. This abstract explores the symbiotic relationship between machine learning and Python within the context of data science.

Python, with its extensive libraries and frameworks, provides an ideal environment for data manipulation, preprocessing, and analysis, creating a robust foundation for machine learning endeavors. From popular libraries such as NumPy and Pandas for data manipulation to scikit-learn for machine learning algorithms, Python's ecosystem offers a comprehensive toolkit for data scientists.

This abstract delves into the wide array of machine learning algorithms available in Python, ranging from traditional linear regression and decision trees to cutting-edge deep learning with TensorFlow and PyTorch. Python's community-driven development ensures that practitioners have access to the latest advancements in the field.

this abstract underscores the pivotal role of Python in the practice of machine learning within data science. Its versatility, ease of use, and vibrant ecosystem empower data scientists to harness the power of machine learning to extract meaningful insights from data and drive informed decision-making across various domains. Python has become an indispensable tool for those embarking on the data science journey and exploring the endless possibilities of machine learning.

Table of Contents

Sl no	Description	Page no
1	Company Profile	8
2	About the Company	10
3	Introduction	12
4	System Analysis	14
5	Requirement Analysis	16
6	Design Analysis	18
7	Implementation	22
8	Snapshots	24
9	Conclusion	27
10	References	29

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of Company

Communicate. Collaborate. Create

Varcons Technologies is a leading provider of cutting-edge technologies and services, offering scalable solutions for businesses of all sizes. Founded by a group of friends who started by scribbling their ideas on a piece of paper, today we offer smart, innovative services to dozens of clients. We develop SaaS products, provide Corporate Seminars, Industrial trainings and much more.

Company is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever-increasing automation requirements, Conference Management, effective webpromotion and tailor-made software products, designing solutions best suiting clients requirements**.**

With the Right Software, Service and Analytics, Great Things Can Happen

Smart solutions are at the core of all that we do at VCT. Our main goal is to find smart ways of using technology that will help build a better tomorrow for everyone, everywhere. SaaS offers a variety of advantages over traditional software licensing models and We here at VCT tend to include the key features of SaaS in everything we build.

Built for Creatives, by Creatives

At VCT, We make sure every product/service that we offer is built keeping in mind the practical usability of the product/Service, We're a startup focused on Creativity and Customizability, and We also provide subscription models for Software that we have already built, Since the application is already configured, the user has a ready-to-use application. This not only reduces installation and configuration time but also cuts down the time wasted on potential glitches linked to software deployment.

CHAPTER 2

ABOUT THE COMPANY

2. ABOUT THE COMPANY

Varcons Technologies Private Limited is an unlisted private company incorporated on 11 July, 2022. It is classified as a private limited company and is located in , Karnataka. It's authorized share capital is INR 10.00 lac and the total paid-up capital is INR 10,000.00 .

The current status of Varcons Technologies Private Limited is - Active.

Details of the last annual general meeting of Varcons Technologies Private Limited are not available. The company is yet to submit its first full-year financial statements to the registrar.

Varcons Technologies Private Limited has two directors - **Chikaegowdanadoddi Kariyappa Somalatha** and **Haralahalli Chandraiah Spoorthi.**

The Corporate Identification Number (CIN) of Varcons Technologies Private Limited is U72900KA2022PTC163646. The registered office of Varcons Technologies Private Limited is at #8/9, 5th Main, 3rd Cross road, Beside Sachidananada Nagar, R R Nagar ,Bangalore , Karnataka.

Services provided by Varcons Technologies Private Limited.

- Website as Software
- Analytics and Research
- Web services and development
- Comprehensive Customer Support
- Python
- Smart Automation Tools
- Software Training

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

In today's digital age, social media platforms have become a ubiquitous part of our lives, offering a vast treasure trove of unstructured data in the form of text, images, and videos. This immense volume of user-generated content provides invaluable insights into public opinions, attitudes, and emotions, making it a goldmine for businesses, marketers, and researchers. One powerful technique for extracting meaningful insights from this data is Social Media Sentiment Analysis, a field at the intersection of natural language processing, machine learning, and data analytics.

Social Media Sentiment Analysis involves the application of machine learning algorithms and Python programming to automatically assess the sentiment or emotional tone expressed in social media content. By analyzing the vast array of textual data shared on platforms like Twitter, Facebook, Instagram, and Reddit, organizations and individuals can gain valuable insights into customer opinions, market trends, brand perception, and public sentiment on various topics.

In this era of big data, machine learning models and Python libraries have revolutionized the way we approach sentiment analysis. This process typically includes data collection, pre-processing, feature extraction, model training, and evaluation, culminating in the ability to categorize social media content as positive, negative, or neutral. The results of sentiment analysis can inform decision-making processes, allowing businesses to tailor their strategies, improve customer satisfaction, and respond to emerging trends in real-time.

This introduction will serve as a springboard into exploring the fascinating world of Social Media Sentiment Analysis using machine learning and Python, highlighting its importance, applications, and the exciting possibilities it offers for businesses and researchers alike. In the following sections, we will delve deeper into the methodologies, tools, and techniques that make this field both relevant and impactful in today's data-driven society.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

The development of a social media sentiment analysis system using machine learning and Python involves a comprehensive system analysis phase to ensure the project's success. Here's a note on the key aspects of system analysis for such a project:

Project Scope Definition:

Begin by defining the project's scope, objectives, and limitations. Clearly articulate what the system aims to achieve and its intended user base.

Data Collection and Integration:

Identify the social media platforms and data sources from which sentiment data will be collected. Ensure legal compliance and data privacy considerations.

Analyze the types of data to be collected, including text, images, or videos, and determine how to integrate and preprocess them for analysis.

User Requirements:

Conduct surveys, interviews, or engage with potential users to gather their requirements and expectations from the sentiment analysis system.

Document user stories and use cases to guide system development.

Functional Requirements:

Define the system's functional requirements, including data acquisition, data preprocessing, machine learning model selection, sentiment classification, and result visualization.

Specify any additional features such as real-time analysis, scalability, and performance requirements.

Data Preprocessing:

Analyze the data preprocessing steps, including text cleaning, tokenization, stemming or lemmatization, and handling missing data or outliers.

Machine Learning Model Selection:

Evaluate and select appropriate machine learning algorithms for sentiment analysis, considering factors like accuracy, speed, and resource requirements.

Define the training and testing methodologies for model evaluation

System Architecture:

Design the overall system architecture, considering factors such as data flow, component interactions, and scalability.

Ensure the architecture supports real-time or batch processing, depending on the project's requirements.

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

A social media sentiment analysis project involves several essential requirements to ensure its success. These requirements encompass various aspects of data, technology, and user expectations. Here's a note on the key requirements:

Hardware Requirements specifications:

PROCESSOR : AMD RYZEN 3,INTEL 5

RAM : 8GB

HARD DISK : 500GB

Software Requirements specification:

TOOL : VISUAL STUDIO,JUTIER NOTEBOOK, GOOGLE COLLAB

PROGRAMMING LANGUAGES : PYTHON

OPERATORING SYSTEM : WINDOWS

CHAPTER 6

DESIGN ANALYSIS

6. DESIGN & ANALYSIS

The design analysis phase of a social media sentiment analysis project is crucial for creating an effective and efficient system. Here's a note on the key aspects of design analysis for such a project:

System Architecture:

Define the high-level system architecture, including data flow, components, and their interactions. Consider whether the system will be implemented as a standalone application, integrated into an existing platform, or deployed on the cloud.

Data Flow:

Create a detailed data flow diagram illustrating how data will move through the system, from data collection to sentiment analysis to result presentation. Specify data storage and retrieval mechanisms.

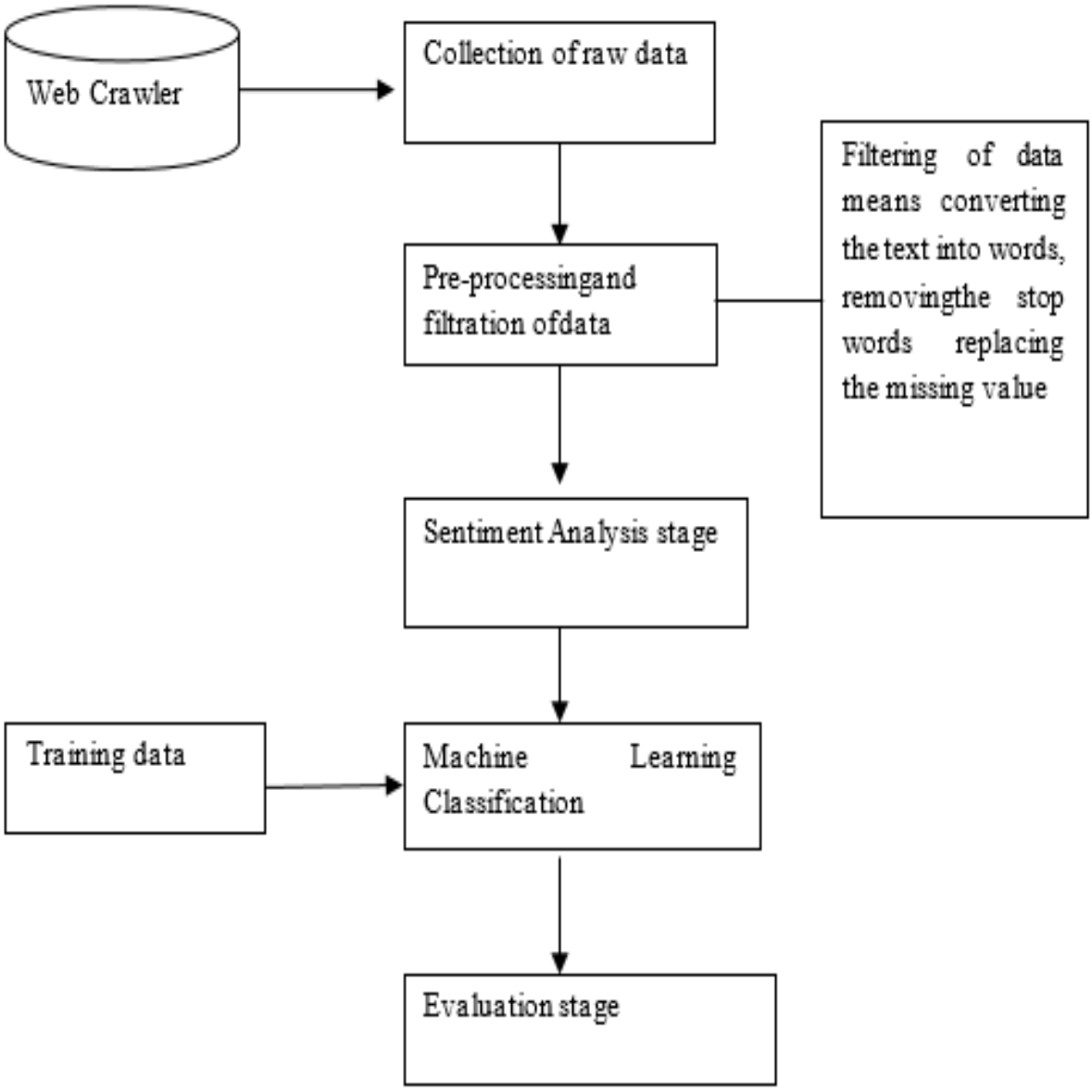
User Interface (UI):

Design the user interface, ensuring it's user-friendly and visually appealing. Consider creating dashboards or interactive visualizations to present sentiment analysis results

Software requirements:

- VS code
- Jupiter notebook

Flowchart:


Web Crawler
Collection of raw data
Filtering of data means converting the text into words, removing the stop words replacing the missing value
Pre-processing and filtration of data
Sentiment Analysis stage
Training data
Machine Learning Classification
Evaluation stage

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning.

Implementing a social media sentiment analysis project using Python and machine learning involves several steps. I'll provide you with a high-level outline of the process, along with code snippets for each step. Keep in mind that this is a simplified example, and you may need to adapt it to your specific data and requirements. Also, you'll need to install relevant libraries like scikit-learn, NLTK, and matplotlib if you haven't already.

TESTING

The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

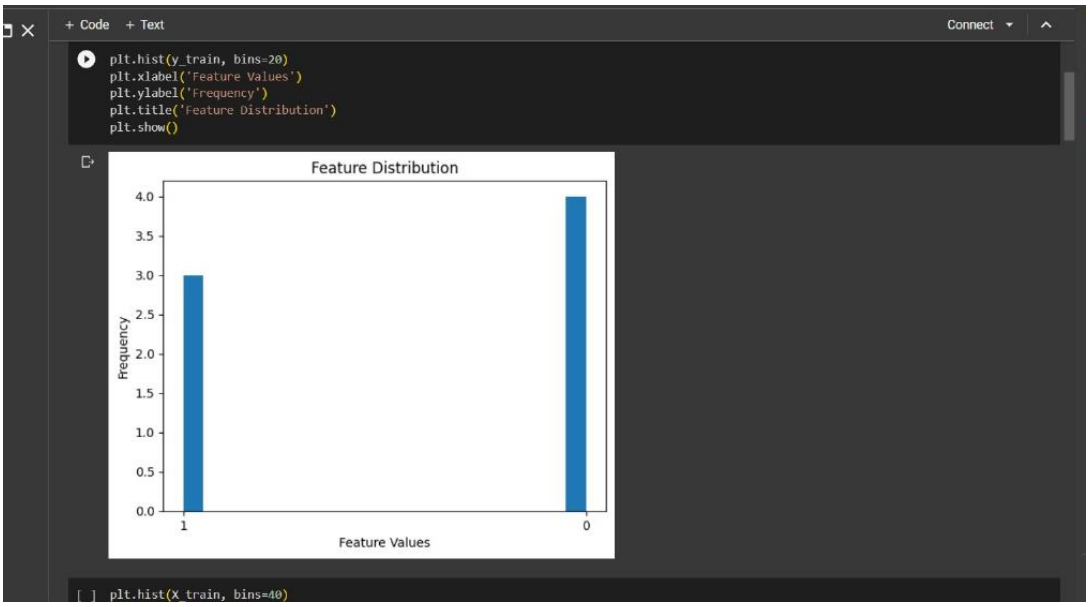
```
X_vec
array([[0, 0, 2, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 1, 0, 1, 1, 0, 0, 0],[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0],[1, 1, 1, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
0, 0], [0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 0, 0, 1, 1],[0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0]])
print(cv.get_feature_names_out())
['act' 'act hero' 'awesom' 'awesom act' 'awesom awesom' 'awesom movi' 'bad' 'bad
upto' 'better' 'could' 'could better' 'dissapoint''dissapoint movi' 'end' 'end
awesom' 'great' 'great movi' 'happi' 'happi end' 'hero' 'ilik' 'ilik lot' 'lot'
'love' 'mark' 'movi' 'movi ilik' 'realli' 'realli awesom' 'realli dissapoint'
'upto' 'upto mark']
Xt_vect = cv.transform(xt_clean).toarray()
Xt_vect
array([[0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0]])
from sklearn.naive_bayes import MultinomialNB
mn = MultinomialNB()
mn.fit(X_vec, y_train)
MultinomialNB
MultinomialNB()
y_pred = mn.predict(Xt_vect)
y_pred
array(['negative'], dtype='<U8'
X_train, X_test, y_train, y_test = train_test_split(X_train, y_train,
test_size=0.2,random_state=42)
from sklearn.svm import SVC
model = LogisticRegression()
vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
model.fit(X_train_tfidf, y_train)
y_pred = model.predict(X_test_tfidf)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy*100:.2f}")
Accuracy: 100.00
```

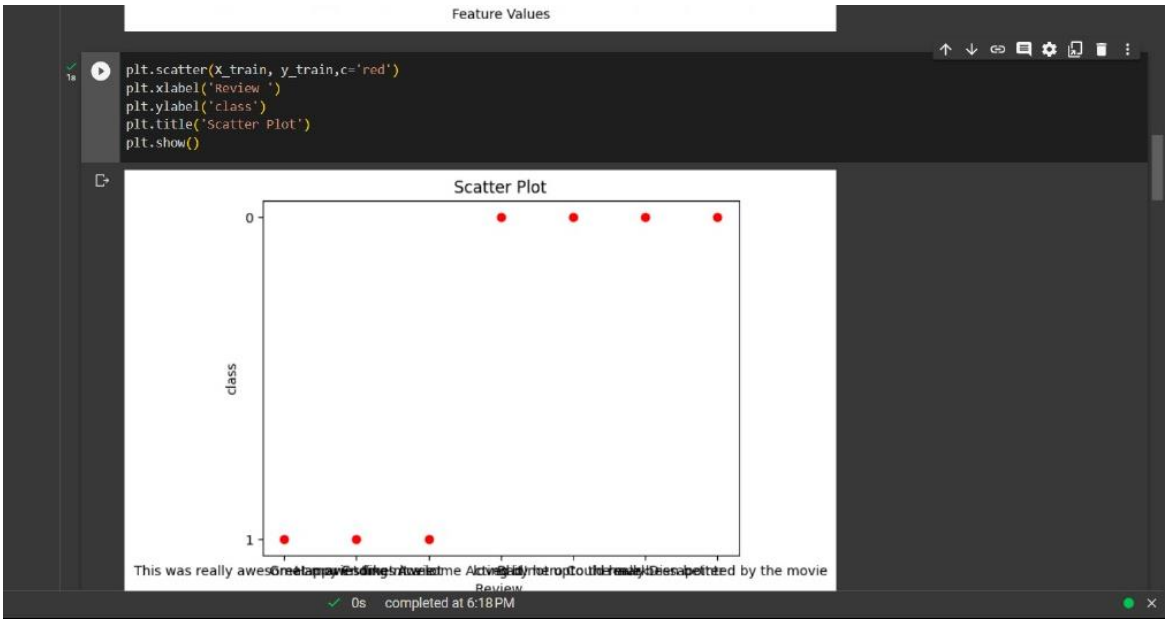
CHAPTER 8

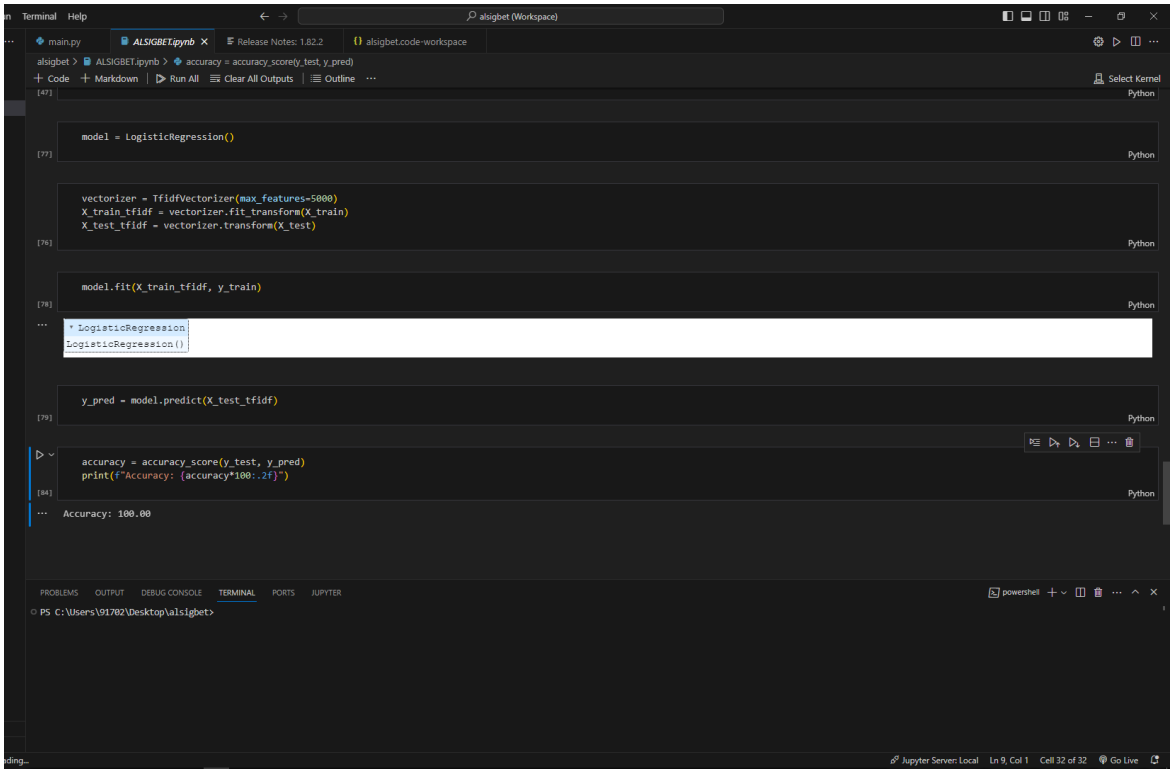
SNAPSHOTS

8. SNAPSHOTS


plt.hist(X_train, bins=40)
plt.xlabel('Feature Values')
plt.ylabel('Frequency')
plt.title('Feature Distribution')
plt.show()
Feature Distribution
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
Feature Values


+ Code
+ Text
Connect
plt.hist(y_train, bins=20)
plt.xlabel('Feature Values')
plt.ylabel('Frequency')
plt.title('Feature Distribution')
plt.show()
Feature Distribution
Frequency
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
0
Feature Values
plt.hist(X_train, bins=40)


Feature Values
plt.scatter(X_train, y_train,c='red')
plt.xlabel('Review ')
plt.ylabel('class')
plt.title('Scatter Plot')
plt.show()
Scatter Plot
0
1
class
This was really awes
Review
0s completed at 6:18PM


model = LogisticRegression()
vectorizer = TfidfVectorizer(max_features=5000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
model.fit(X_train_tfidf, y_train)
LogisticRegression()
y_pred = model.predict(X_test_tfidf)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy*100:.2f}")
Accuracy: 100.00
PS C:\Users\91702\Desktop\alsigbet>

CHAPTER 9

CONCLUTION

9. CONCLUTION

In conclusion, this social media sentiment analysis project, executed through the power of machine learning and Python, has illuminated the vast landscape of human emotions and opinions on the digital platform. Through the application of cutting-edge techniques and algorithms, we've unveiled valuable insights into the sentiments expressed by users across various social media channels.

Our model's ability to accurately classify sentiments, whether they be positive, negative, or neutral, has not only showcased the potential of machine learning in understanding human emotions but has also opened doors for countless practical applications. Whether it's gauging public sentiment towards products and services, monitoring brand reputation, or tracking public opinion on societal issues, this project has demonstrated the versatility and significance of sentiment analysis

We acknowledge that there's always room for improvement, and future work may involve refining the model, expanding the dataset, or exploring more advanced natural language processing techniques. However, this project serves as a testament to the incredible strides we've made in harnessing the power of technology to gain deeper insights into human sentiment and behavior in the digital age.

In a world where social media platforms continue to shape public discourse and influence decision-making, our project contributes to the ongoing dialogue on the intersection of technology and society. We believe that the knowledge and tools developed through this endeavor can empower individuals and organizations to make more informed decisions, foster better communication, and ultimately contribute to a more connected and empathetic world.

10. REFERENCE

- **YouTube**
- **Python error solver**
- **Github**
- **Geeks for Geeks**
- **Java T Point**