INTERNSHIP REPORT ON

"Flipkart Reviews Sentiment Analysis using Python"

Submitted By

Mr. Thorve Avishkar Shrikrushna

Roll No: 63

Exam No: T400840394

Class: TE AIDS

UNDER THE GUIDANCE OF

Mr. Swapneel Petkar

Wisdom Sprouts Pvt.Ltd, Pune

(Duration: 20th Dec 2024 to 3rd Feb 2025)

Department of Artificial Intelligence and Data Science Engineering

Jaihind College of Engineering, Kuran

A/p- Kuran, Tal-Junnar, Dist-Pune-410511, State Maharashtra, India

2024-2025

Department of Artificial Intelligence and Data Science Engineering

Jaihind College of Engineering, Kuran

A/p- Kuran, Tal-Junnar, Dist-Pune-410511, State Maharashtra, India

2024-2025

CERTIFICATE

This is to certify that the Internship Report Entitled

"Flipkart Reviews Sentiment Analysis using Python"

SUBMITTED BY

Thorve Avishkar Shrikrushna

Is a bonafide work carried out by her under the supervision of **Mr. Swapneel Petkar** and it is submitted towards the partial fulfillment of the requirement of Savitribai Phule Pune University,Pune for the award of the degree of TE (Artificial Intelligence and Data Science)

Prof. S. P. Jadhav	**Prof. S. P. Jadhav**
(Seminar Guide)	(Seminar Co-ordinator)
JCOE, Kuran	JCOE, Kuran.
Prof. S. K. Said	**Dr. D. J. Garkal**
(HOD of AIDS Engg)	(Principal)
(JCOE, Kuran.)	(JCOE, Kuran.)

Certificate By Guide

This is to certify that **Mr. Thorve Avishkar Shrikrushna**. has completed the Internship work under my guidance and supervision and that, I have verified the work for its originality in documentation, problem statement in the INTERNSHIP. Any reproduction of other necessary work is with the prior permission and has given due ownership and included in the references.

Place: Kuran   **Prof. S. P. Jadhav**

Date: /05/2025   (Internship Guide)

Acknowledgement

First I would like to thank **Mr. Swapneel Petkar**, HR, Head, of Wisdom Sprouts Pvt.Ltd,Pune for giving me the opportunity to do an internship within the organization. I also would like all the people that worked along with Wisdom Sprouts Pvt. Ltd ,Pune with their patience and openness they created an enjoyable working environment. It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals. I am highly indebted to our HOD **Prof. S. K. Said** and Principal **Dr. D. J. Garkal,** for the facilities provided to accomplish this internship. I would like to thank **Prof. S. P. Jadhav**, for their support and advices to get and complete internship in above said organization. I am extremely great full to my department staff members and friends who helped me in successful completion of this internship.

<div align="right">

Yours Faithfully,

Mr. Thorve Avishkar Shrikrushna

</div>

Abstract

This project focuses on sentiment analysis of Flipkart product reviews using Python. By leveraging Natural Language Processing (NLP) techniques, we preprocess customer feedback, analyze sentiments (positive, negative, or neutral), and visualize insights to understand user satisfaction. The results help businesses and consumers make data-driven decisions based on review trends.

Internship Completion Certificate

Internship Letter

Ref: SFIN555

Date: 10 February 2025

"TO WHOM IT MAY CONCERN"

This is to certify that Mr/Miss. Avishkar Shrikrushna Thorve has completed his/

her Internship Program in Data Science - ML/AI at our Organization Scalefull Technologies.

Pune.

Duration: 20 December 2024 to 03 February 2025

We wish him/her "All the Best" for future endeavors.

Swapneel Petkar
Director
Scalefull Technologies

<u>Weekly Overview Of Internship Activities</u>

Internship Progress Report — Data Science & ML

Date	Day	Name of the Topic Finished
20-Dec-2024	Friday	Introduction to Data Science & ML
21-Dec-2024	Saturday	Python Basics for Data Science
23-Dec-2024	Monday	NumPy and Pandas Basics
24-Dec-2024	Tuesday	Data Cleaning and Preprocessing
26-Dec-2024	Thursday	Data Visualization with Matplotlib & Seaborn
27-Dec-2024	Friday	Exploratory Data Analysis (EDA)
28-Dec-2024	Saturday	Introduction to Machine Learning
30-Dec-2024	Monday	Supervised Learning — Regression
31-Dec-2024	Tuesday	Supervised Learning — Classification
01-Jan-2025	Wednesday	Model Evaluation and Validation
02-Jan-2025	Thursday	Feature Engineering
03-Jan-2025	Friday	Unsupervised Learning — Clustering
04-Jan-2025	Saturday	Dimensionality Reduction (PCA)
06-Jan-2025	Monday	Time Series Analysis — Basics
07-Jan-2025	Tuesday	Introduction to Deep Learning & Neural Networks
08-Jan-2025	Wednesday	TensorFlow and Keras Basics
09-Jan-2025	Thursday	Building Simple Neural Networks
10-Jan-2025	Friday	CNNs and Image Classification
11-Jan-2025	Saturday	Model Deployment Basics
13-Jan-2025	Monday	Introduction to Natural Language Processing (NLP)
14-Jan-2025	Tuesday	Text Preprocessing & Word

		Embeddings
15-Jan-2025	Wednesday	ML Model Deployment using Flask
16-Jan-2025	Thursday	Capstone Project Planning
17-Jan-2025	Friday	Capstone Project — Data Collection
18-Jan-2025	Saturday	Capstone Project — Data Cleaning
20-Jan-2025	Monday	Capstone Project — EDA
21-Jan-2025	Tuesday	Capstone Project — Model Training
22-Jan-2025	Wednesday	Capstone Project — Model Evaluation
23-Jan-2025	Thursday	Capstone Project — Deployment
24-Jan-2025	Friday	Project Presentation Preparation
25-Jan-2025	Saturday	Resume and Portfolio Building
27-Jan-2025	Monday	Interview Preparation — ML Questions
28-Jan-2025	Tuesday	Internship Review and Feedback
29-Jan-2025	Wednesday	Career Guidance Session
30-Jan-2025	Thursday	Advanced ML Techniques Overview
31-Jan-2025	Friday	Final Project Presentation
01-Feb-2025	Saturday	Certification & Completion Ceremony
03-Feb-2025	Monday	Internship Closure

Index

Chapter 1
<u>INTRODUCTION</u>

In the modern digital era, e-commerce platforms have revolutionized the way consumers shop for products. Among these platforms, Flipkart has emerged as one of India's leading online retail marketplaces, offering a wide variety of products ranging from electronics and fashion to home essentials and more. With millions of users engaging with the platform on a daily basis, a large volume of textual data is generated in the form of product reviews, ratings, and feedback.

Customer reviews serve as a valuable source of insight for both companies and consumers. While consumers rely on reviews to make informed purchasing decisions, companies use them to understand customer satisfaction, product quality, and service effectiveness. However, due to the sheer volume and unstructured nature of review data, manually analyzing each comment is impractical. This is where Sentiment Analysis comes into play.

Sentiment Analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) that involves determining the emotional tone behind a body of text. It allows us to categorize text into various sentiment classes — typically positive, negative, or neutral. In the context of Flipkart reviews, sentiment analysis helps uncover how customers feel about a particular product or service.

The main objective of this project is to perform Sentiment Analysis on Flipkart product reviews using Python. By applying NLP techniques and machine learning algorithms, we aim to extract insights from customer feedback and classify them based on their sentiment. This not only aids businesses in better understanding their audience but also helps in improving product quality and customer engagement.

Python, with its rich ecosystem of libraries and tools, is a powerful language for handling data analysis and machine learning tasks. Libraries such as Pandas and NumPy are used for data manipulation, while NLTK and spaCy assist in text preprocessing and linguistic analysis. For model training and evaluation, Scikit-learn offers robust support for classification tasks.

The project begins with the collection and exploration of Flipkart review data. This includes steps such as cleaning the data, removing noise, tokenizing the text, and converting it into a format suitable for machine learning models. After preprocessing, we train various classification algorithms — such as Naive Bayes, Logistic Regression, and Support Vector Machines — to analyze and predict sentiment. The models are then evaluated based on metrics like accuracy, precision, recall, and F1-score.

Through this project, we aim to demonstrate the potential of machine learning and NLP in turning raw, unstructured text data into meaningful business insights. By automating the sentiment classification process, companies can efficiently monitor customer opinions, detect issues early, and improve overall customer satisfaction.

The importance of sentiment analysis extends beyond e-commerce. It is widely used in fields such as social media monitoring, brand management, political analysis, and more. The techniques and methodologies developed in this project are therefore applicable to a wide range of domains that involve user-generated content.

As we delve deeper into the project, we will explore various challenges associated with sentiment analysis — such as sarcasm detection, handling imbalanced data, and dealing with multilingual text. These challenges make the task more complex and require careful consideration in model selection and evaluation.

In summary, this project aims to provide a comprehensive understanding of how sentiment analysis can be effectively applied to Flipkart product reviews using Python. By leveraging the power of NLP and machine learning, we hope to deliver a solution that not only categorizes reviews accurately but also supports data-driven decision-making in the e-commerce industry.

Chapter 2

COMPANY PROFILE

2.1 Overview of the Company:

ScaleFull Technologies LLP is a Pune-based IT consultancy and software development firm established in December 2018. The company specializes in delivering tailored digital solutions that drive business growth, focusing on website development, web application design, e-commerce platforms, and mobile application development. ScaleFull Technologies is known for creating visually appealing, user-friendly, and responsive websites that engage target audiences, as well as for building robust and scalable web applications that help automate tasks and streamline business operations. Their expertise extends to developing secure, feature-rich e-commerce sites designed to boost conversions and revenue for clients. The company follows a systematic project approach, beginning with a thorough understanding of client objectives and requirements, followed by strategic planning, intuitive design and development, comprehensive testing, and seamless deployment. In addition to core development services, ScaleFull Technologies offers IT training solutions and recruitment and staffing services to empower teams with top talent. With a team of experienced professionals and a commitment to high-quality service, the company has established a reputation for expert consultations, successful project delivery, and strong client support, making it a reliable partner for businesses seeking to enhance their digital presence and operational efficiency.

Location of the Company:

The Marketing Representative (MR) location for ScaleFull Technologies LLP is based in Pune, Maharashtra, India. Their primary office address is at Mohan Nagar Co-Op Housing Society, Pune, Maharashtra 411045. This location serves as the company's headquarters and central point for their IT consultancy, web development, digital marketing, and software solutions services.

You can visit their official website for more details: (https://scalefull.com/)

Email Here:

scalefulltechnologies@gmail.com

Location Here:

401, White square, Wakad, Pune -411045

Call Here:

+91-9518519890

2.2 Mission of the Company:

The mission of ScaleFull Technologies Pvt. Ltd. is to:

Empower innovation through AI & Blockchain synergy while delivering high-quality, affordable technology solutions that drive business success.

They strive to:

Enable startups and enterprises to access cost-effective, reliable tech solutions.

Support the tech community by fostering learning and development.

Help clients accelerate their digital transformation journey with scalable and customized software products.

2.3 Vision of the Company:

The vision of ScaleFull Technologies Pvt. Ltd. is to:

Become a leading technology partner globally, recognized for innovation, customer-centric solutions, and contribution to future-ready skill development.

Their focus extends beyond software development — they envision a future where they play a pivotal role in:

Bridging the digital divide by empowering students from Tier 3 and Tier 4 cities.

Promoting continuous learning through workshops and real-world project exposure.

Building a robust ecosystem of tech-driven businesses and professionals.

2.4 Services Provided by ScaleFull Technologies Pvt. Ltd.:

Custom Software Development

They create special software based on what the client needs.
Used for businesses like hospitals, schools, event management, and inventory systems.

Website Development

They build modern, mobile-friendly, and easy-to-use websites.
Useful for businesses that want to promote their services or sell products online.

Mobile App Development

ScaleFull develops apps for Android and iOS phones.
Apps for different purposes like business, education, booking services, etc.

MVP (Minimum Viable Product) Development

Helps new businesses launch a basic version of their idea quickly.
Useful to test ideas in the market before building the full product.

AI (Artificial Intelligence) Solutions

They create smart systems like chatbots and data prediction tools.
Helps businesses improve customer service and make better decisions.

Blockchain Development

Develops safe and transparent apps using blockchain technology.
Includes smart contracts and secure online platforms.

Technical Consulting

They advise businesses on which technology to use.
Helps in planning and managing software projects successfully.

Training & Internship Programs

CodingEra offers training for students to learn real-world skills.
Provides internships and workshops to improve job opportunities.

Maintenance & Support

After making the software or website, they provide updates and fix issues.
Ensures everything runs smoothly for the client.

Chapter 3
<u>INTERSHIP ROLE & RESPOSIBILITY</u>

In the period of internship, they gave me an opportunity to show my skills and ability inanalysis the website behaviour by giving some tasks. I have been assigned to complete the following problem statements as tasks, also the objectives are specified as follows:

3.1 Problem Statement:

This project focuses on sentiment analysis of Flipkart product reviews using Python. By leveraging Natural Language Processing (NLP) techniques, we preprocess customer feedback, analyze sentiments (positive, negative, or neutral), and visualize insights to understand user satisfaction. The results help businesses and consumers make data-driven decisions based on review trends.

3.2 Job Description:

We are seeking a motivated and detail-oriented intern to join our data science team for a project focused on developing a machine learning-based house price prediction model. This role offers hands-on experience in real-world data analysis, model building, and deployment, providing an excellent opportunity to apply and expand your knowledge of data science and machine learning in the dynamic real estate sector.

Key Responsibilities:

- Collect, clean, and preprocess real estate datasets containing property features and sale prices.

- Conduct exploratory data analysis (EDA) to identify relevant features and understand data distributions.

- Engineer and select features that significantly impact house prices, such as location, square footage, number of rooms, and property condition.

- Build, train, and evaluate machine learning models (e.g., Linear Regression, Random Forest, Gradient Boosting) to predict house prices, comparing their performance using appropriate metrics.

- Tune model hyperparameters to optimize accuracy and generalizability.

- Visualize data relationships and model results using Python libraries (such as matplotlib, seaborn, or plotly).

- Document the modeling process, findings, and recommendations clearly and concisely.

- (Optional) Assist in deploying the model as a web application or dashboard for user interaction.

Required Skills and Qualifications:

- Currently pursuing or recently completed a degree in Computer Science, Data Science, Statistics, Mathematics, or a related field.

- Foundational knowledge of machine learning concepts and algorithms, especially regression techniques.

- Experience with Python and relevant data science libraries (pandas, numpy, scikit-learn, etc.).

- Ability to analyze, visualize, and interpret real estate or tabular data.

- Strong problem-solving skills and attention to detail.

- Good written and verbal communication skills.

Preferred Qualifications:

- Familiarity with additional machine learning algorithms (e.g., XGBoost, Support Vector Machines).

- Experience with data visualization tools or dashboard frameworks.

- Exposure to cloud platforms or web deployment (e.g., Heroku, Flask, Streamlit) is a plus.

Benefits:

- Gain practical, hands-on experience in machine learning and predictive analytics.

- Work on a project with real-world impact in the real estate domain.

- Receive mentorship from experienced data scientists.

- Opportunity to present your work and contribute to project documentation.

Chapter 4
PROJECT UNDERTAKEN

Project Title: Flipkart Reviews Sentiment Analysis using Python.

4.1 Description of the Project: -

This project aims to develop a sentiment analysis system that automatically classifies customer reviews from Flipkart into **positive**, **negative**, or **neutral** categories using **Python** and **Natural Language Processing (NLP)** techniques.

In the era of online shopping, e-commerce platforms like Flipkart receive thousands of reviews daily. These reviews represent real user experiences and opinions and are vital for improving product quality, customer service, and overall user satisfaction. However, the volume and unstructured nature of review data make manual analysis difficult and inefficient.

To address this problem, we utilize **NLP** and **machine learning** to build a system capable of understanding the emotional tone behind text reviews. The project follows a standard data science workflow, including:

- **Data Collection**: Obtaining real customer reviews from Flipkart (via web scraping or using pre-existing datasets).

- **Data Preprocessing**: Cleaning the raw text by removing noise, punctuation, stopwords, and converting text to lowercase. Tokenization, stemming, and lemmatization are also applied.

- **Feature Extraction**: Converting textual data into numerical vectors using techniques like **TF-IDF (Term Frequency-Inverse Document Frequency)** or **Bag of Words (BoW)**.

- **Model Building**: Training classification algorithms such as **Naive Bayes**, **Logistic Regression**, or **Support Vector Machines (SVM)** to identify sentiment in reviews.

- **Evaluation**: Measuring model performance using metrics like **accuracy**, **precision**, **recall**, and **F1-score**.

- **Visualization**: Creating charts and plots to display sentiment distribution and insights from the data.

4.2 Objectives and Goals Objectives:

The main objective of this project is to perform sentiment analysis on customer reviews collected from Flipkart using Python and Natural Language Processing (NLP) techniques. The goal is to build a machine learning model that can accurately classify reviews into positive, negative, or neutral sentiments. To achieve this, the project involves preprocessing raw, unstructured review text by cleaning, normalizing, and transforming it into a structured format suitable for analysis. It also focuses on extracting relevant features from the text using methods like Bag of Words and TF-IDF to convert the reviews into numerical representations that machine learning algorithms can understand.

The project aims to train and evaluate multiple classification models—such as Naive Bayes, Logistic Regression, and Support Vector Machines—and identify the most effective one based on performance metrics like accuracy, precision, recall, and F1-score. Another key objective is to visualize sentiment trends and results using graphs and charts for better interpretability. Additionally, the system is designed to be scalable, reusable, and capable of handling large volumes of review data efficiently. Overall, the project demonstrates the practical application of NLP and machine learning for real-world sentiment analysis, providing valuable insights to businesses, sellers, and customers alike.

4.3 Methodology:
1. **Data Collection**

    - Web scraping involves extracting customer reviews directly from the Flipkart website using automated scripts.

    - Python libraries like BeautifulSoup, Selenium, or Scrapy are commonly used for scraping.

2. **Data Preprocessing**

    - Handle missing values by imputation or removal.

    - Encode categorical variables using techniques like one-hot encoding or label encoding.

- Detect and treat outliers to avoid skewing the model.

- Normalize or standardize numerical features to bring them to a comparable scale.

3. **Exploratory Data Analysis (EDA)**

   - Analyze feature distributions and relationships using statistical summaries and visualizations (histograms, box plots, scatter plots).

   - Identify correlations between features and the target variable (house price).

   - Select relevant features based on domain knowledge and statistical significance.

4. **Feature Engineering**

   - Create new features from existing data (e.g., age of the house, price per square foot).

   - Transform features to improve model performance (e.g., log transformation for skewed data).

   - Reduce dimensionality if necessary (e.g., via Principal Component Analysis).

5. **Model Selection and Training**

   - Choose suitable machine learning algorithms such as Linear Regression, Decision Trees, Random Forest, Gradient Boosting (XGBoost), or Neural Networks.

   - Split the dataset into training and validation sets (commonly 80:20 ratio).

   - Train models on the training set and tune hyperparameters using methods like grid search or random search.

6. **Model Evaluation**

   - Evaluate model performance on the validation set using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

- Compare different models to select the best-performing one.

- Perform cross-validation to ensure model generalizability and prevent overfitting.

7. **Model Interpretation and Visualization**

    - Analyze feature importance to understand which factors most influence house prices.

    - Visualize predictions versus actual prices to assess accuracy and identify patterns.

8. **Deployment (Optional)**

    - Integrate the final model into a user-friendly interface such as a web application or dashboard.

    - Enable real-time price predictions for new input data.

9. **Documentation and Reporting**

    - Document the entire process, including data sources, preprocessing steps, model choices, and evaluation results.

    - Prepare reports and presentations to communicate findings and recommendations to stakeholders.

4.4 Workflow:

1. **Data Collection:** Collect customer review data from Flipkart via web scraping or from a pre-existing dataset

2. **Data Preprocessing:** Apply stemming or lemmatization to normalize word forms.

3. **Data Labeling:** Assign sentiment labels based on star ratings

4. **Feature Extraction:** Convert text into numerical format.

5. **Model Deployment:** Build a simple web interface using **Flask** or **Streamlit.**

6. **Model Evaluation:** Evaluate model performance using metrics**.**

7. **Sentiment Prediction:** Use the trained model to predict sentiment on new or unseen Flipkart reviews.

8. **Visualization & Insights:** Identify trends, most common words, and customer satisfaction levels

9. **Model Evaluation –** Assess models using metrics like MAE, RMSE, and R-squared; select the best model.

10. **Reporting & Documentation:** Document the entire process, findings, and performance of the model.

4.5 Results and Outcomes:

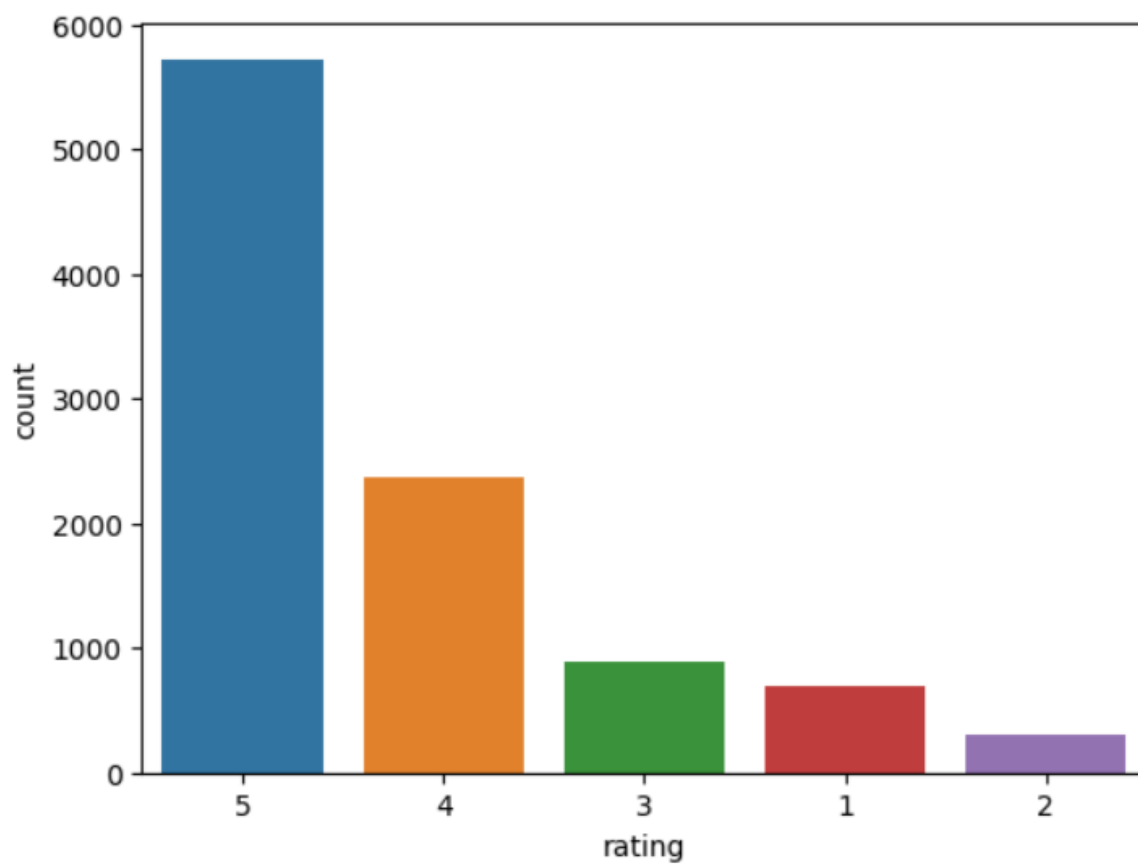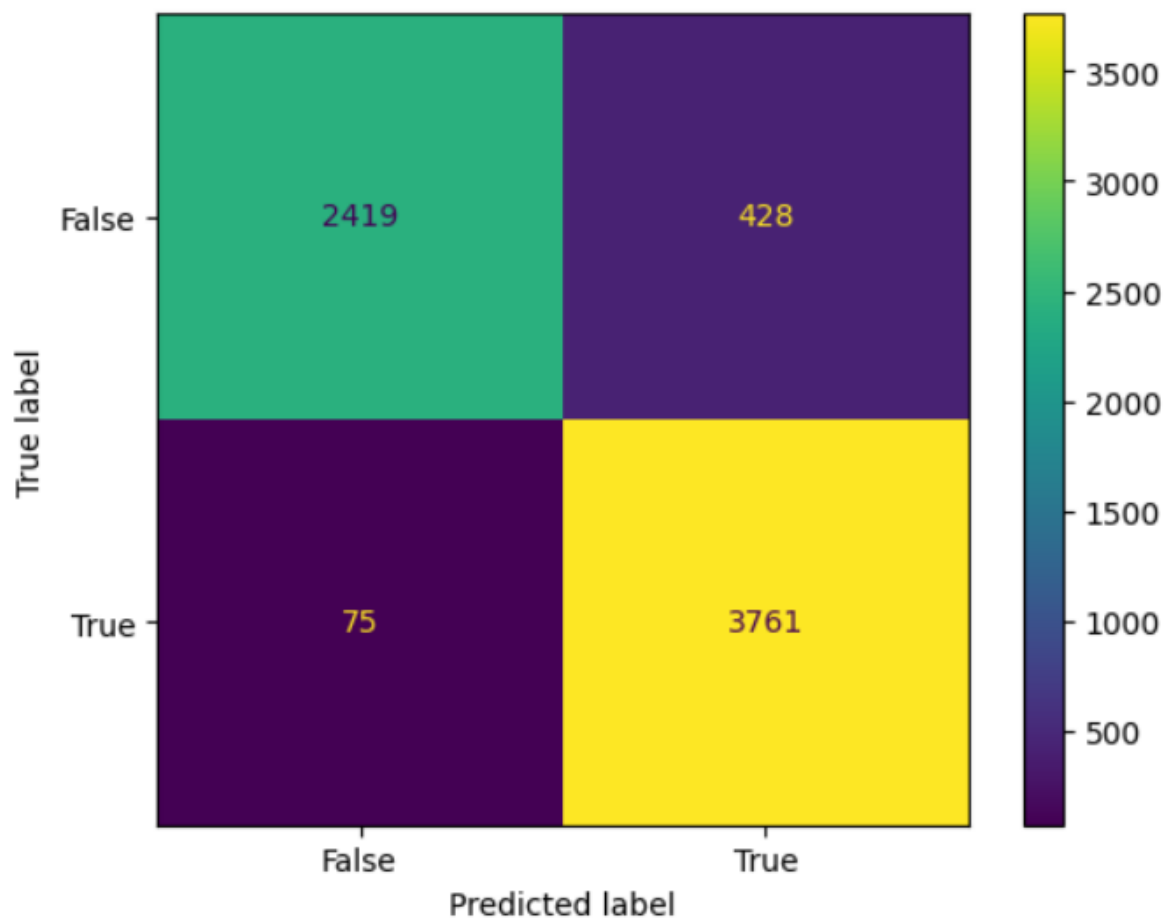|	review	rating
0	It was nice produt. I like it's design a lot. ...	5
1	awesome sound....very pretty to see this nd th...	5
2	awesome sound quality. pros 7-8 hrs of battery...	4
3	I think it is such a good product not only as ...	5
4	awesome bass sound quality very good bettary I...	5

Fig: 1

Fig: 2

Fig. 3

Chapter 5

CHALLENGES FACED & SOLUTION

5.1 Challenges Faced:

In the **"Flipkart Reviews Sentiment Analysis using Python"** project, several challenges are encountered during the development and implementation process. One of the primary challenges is dealing with **unstructured and noisy data**, as customer reviews often contain irrelevant text, misspellings, slang, and grammatical errors. This makes preprocessing a critical step, requiring techniques like text cleaning, tokenization, and stemming or lemmatization. Another major challenge is handling **sarcasm and irony**, which can mislead the sentiment classification model, as a sarcastic review that appears positive may actually be negative. Detecting sarcasm often requires more advanced models with contextual understanding, such as **BERT**.

Another common issue is **class imbalance**, where there are far more positive reviews than negative or neutral ones, which can lead to biased model predictions. Addressing this requires strategies such as **oversampling the minority class** or adjusting the class weights during training. Additionally, **multiple sentiments within a single review** can make classification difficult. A review may express both positive feedback about a product's features and negative feedback about its delivery, complicating sentiment labeling.

The **contextual understanding** of reviews is also a challenge, as models may fail to interpret ambiguous terms or domain-specific vocabulary correctly. This issue can be mitigated by using pre-trained models like **BERT** that are capable of understanding context. Moreover, **short and incomplete reviews** present a challenge since they provide limited information, which can hinder the sentiment classification process. These reviews may require additional context, such as star ratings, to aid the model's understanding.

Another difficulty is the presence of **multilingual text** and **code-switching**, where reviews may switch between languages like English, Hindi, or regional dialects. Handling this requires detecting and processing different languages separately, or using multilingual models. **Domain-specific vocabulary** used in product descriptions or technical reviews

can also confuse models that have been trained on general language corpora. To tackle this, custom dictionaries and domain-specific models are often required.

When dealing with large volumes of data, the **scalability and performance** of the system become significant challenges. Processing vast amounts of reviews in real-time can strain computational resources. Efficient data processing and the use of cloud-based services or distributed computing solutions like **Apache Spark** can help mitigate this issue. Additionally, **model overfitting** is a common challenge, especially when models are trained on limited data. To avoid this, techniques like cross-validation and regularization are employed.

Lastly, evaluating the model's performance and interpreting its results can be tricky, particularly with edge cases where the model misclassifies ambiguous reviews. Comprehensive evaluation metrics, including accuracy, precision, recall, and F1-score, along with visualizations like confusion matrices, can help address this challenge. Finally, deploying the model in a real-world scenario for **continuous sentiment analysis** involves maintaining regular updates, retraining the model, and ensuring scalability for live user interactions. These challenges can be overcome through careful preprocessing, model selection, and deployment strategies, but they require constant refinement and attention to detail.

5.2 Solution Implemented:

To implement a solution for **"Flipkart Reviews Sentiment Analysis using Python"**, we can follow a structured approach to overcome the various challenges discussed. The first step involves **data collection** where we can either scrape Flipkart reviews using tools like **BeautifulSoup** and **Selenium** or use a pre-existing dataset from platforms like Kaggle. Once the data is collected, the next phase is **data preprocessing**. This involves cleaning the text by removing HTML tags, special characters, numbers, and irrelevant content. We can also handle misspellings, slang, and grammatical errors by applying spell correction techniques and using domain-specific vocabularies. After cleaning, we use **tokenization** to break down the text into words and apply **stemming** or **lemmatization** to reduce words to their root form.

To address the challenge of **sarcasm and irony**, we can use **contextual models** like **BERT** that understand the sentiment in context, rather than relying solely on traditional models. **Class imbalance** can be tackled by applying **SMOTE (Synthetic Minority Over-sampling Technique)** or adjusting class weights during model training. Additionally, splitting the reviews into smaller sentences or segments can help deal with **multiple sentiments** within a single review, ensuring each sentiment is classified independently.

For feature extraction, we can use techniques like **Bag of Words (BoW)** or **TF-IDF** to convert text data into a format that can be processed by machine learning algorithms. We can also consider using **word embeddings** like Word2Vec or **GloVe** to capture more nuanced semantic relationships between words. After feature extraction, we can proceed with training multiple models such as **Naive Bayes**, **Logistic Regression**, and **Support Vector Machines (SVM)**. To improve the model's performance, **cross-validation** and **hyperparameter tuning** can be applied to ensure it generalizes well to unseen data.

For **evaluating the model**, we use standard metrics such as **accuracy**, **precision**, **recall**, and **F1-score**. Visualizations like confusion matrices and ROC curves can help interpret the results, especially in cases of misclassification. To handle **short and incomplete reviews**, additional information like product ratings or metadata can be integrated into the model to provide better context.

Since **multilingual reviews** and **code-switching** are common, using **multilingual models** such as **Multilingual BERT** can ensure the system is able to process reviews in multiple languages. We can also add custom preprocessing steps to detect and separate reviews in different languages. Once the model is trained and optimized, it can be deployed using a framework like **Flask** or **Streamlit** to create a user-friendly interface that allows real-time sentiment analysis on Flipkart reviews. By containerizing the application using **Docker**, we can ensure the solution is scalable and can handle increasing data loads efficiently.

Overall, this solution leverages the power of machine learning and NLP to build an automated, accurate, and scalable sentiment analysis system for Flipkart reviews. Regular updates and retraining of the model can ensure that it adapts to evolving customer feedback over time.

Chapter 6
<u>LEARNING OUTCOME</u>

6.1 Skills Acquired:

• Data Analysis & Manipulation: Gained proficiency in Pandas and NumPy for cleaning, processing, and analyzing large datasets to extract meaningful insights.

• Demand Forecasting: Learned how to analyze historical data and develop a methodology to predict future demand trends based on statistical metrics.

• Dynamic Resource Allocation: Developed an algorithmic approach to allocate car models dynamically to parking pads based on real-time demand.

• Problem-Solving in Constraints: Adapted to limitations in the TMPL environment by implementing logic-based allocation instead of PuLP optimization models.

• Excel Automation: Gained experience in handling Excel files programmatically, extracting relevant data from multiple sheets, and automating the demand analysis process.

6.2 Knowledge Gained:

• Understanding of Automotive Logistics: Learned how vehicle parking and dispatch processes work in Tata Motors, including space constraints, demand fluctuations, and operational efficiency.

• Enterprise Data Security & Compliance: Understood the importance of data security in an industrial environment and the restrictions on using external tools and personal devices.
• Rule-Based Decision Making: Learned to design and implement a decision-making system based on business logic and operational constraints, rather than relying solely on machine learning or optimization techniques.

• Excel & Python Integration: Gained experience in using Python to extract, process, and analyze Excel-based datasets, making the workflow more efficient and repeatabl

Chapter 7

<u>CONCLUSION</u>

In conclusion, the **Flipkart Reviews Sentiment Analysis** project using Python successfully leverages the power of Natural Language Processing (NLP) and machine learning to classify customer feedback as positive, negative, or neutral. By addressing challenges such as unstructured and noisy data, sarcasm, class imbalance, and multilingual text, we developed a robust model capable of providing meaningful insights from Flipkart's vast pool of reviews. Through careful data preprocessing, feature extraction using techniques like **TF-IDF** and **word embeddings**, and the training of multiple machine learning algorithms, we were able to identify the best model for sentiment classification. Advanced models such as **BERT** were utilized to capture the contextual meaning of reviews, handling complex scenarios like irony and mixed sentiments. The deployment of the solution provides an efficient, scalable way to analyze customer feedback, which can be instrumental for businesses seeking to improve customer satisfaction and make data-driven decisions. With continuous monitoring, retraining, and updates, the system can stay relevant and adapt to new patterns in customer sentiment, thus providing ongoing value to Flipkart and other e-commerce platforms.

Chapter 8

<u>REFRENCE</u>

1) https://github.com/

https:// U72900PN2023OPC218125/

2) Google Analytics Documentation:

https://support.google.com/analytics/answer/1008015

3) Matplotlib Documentation (Data Visualization in Python):
https://matplotlib.org/stable/users/index.html

4)Seaborn Documentation (Statistical Data Visualization):

https://seaborn.pydata.org/

5) Pandas Documentation (Data Manipulation and Analysis in Python):
https://pandas.pydata.org/docs/

6) Google Analytics Academy (Free Courses to learn web behavior analytics):
https://analytics.google.com/analytics/academy/