Four Weeks Industrial Training Project Report

on

Customer Churn Prediction

Submitted in the partial fulfillment of the requirement for the award of a degree of

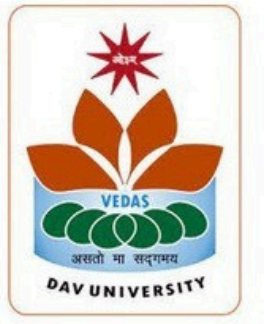
Bachelor of Technology

in

Computer Science and Engineering

Batch (2023-

2027)



Submitted To:

Dr.Ridhi kapoor

Assistant Prof. CSE

Submitted By:

Ankita

12300258

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DAV UNIVERSITY JALANDHAR-PATHANKOT NATIONAL HIGHWAY, NH44, SARMASTPUR PUNJAB-144001

ABSTRACT

Customer churn prediction is a crucial analytical approach used by businesses to identify customers who are likely to discontinue using their services. In this project, machine learning techniques are applied to analyze customer behavior, service usage patterns, and demographic factors to predict churn accurately. The dataset is preprocessed through data cleaning, feature selection, and transformation to improve model performance. Various classification algorithms—such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines—are trained and evaluated using metrics like accuracy, precision, recall, and F1-score. The model with the highest performance is selected to provide actionable insights for reducing churn. The outcomes of this study help organizations implement targeted retention strategies, improve customer satisfaction, and enhance long- term business profitability.

-ACKNOWLEDGEMENT

I would like to express my deep and heartfelt gratitude to my project guide for their continuous support, valuable guidance, and insightful suggestions throughout the completion of my Customer Churn Prediction project. Their expertise and encouragement played a crucial role in enhancing my understanding of machine learning concepts, data preprocessing, and model evaluation techniques. I am sincerely thankful to my college/department for providing a supportive academic environment, essential resources, and the opportunity to work on this project. I also extend my appreciation to all the faculty members who have taught and inspired me during this course, as their knowledge has greatly contributed to the successful execution of this work. Furthermore, I am deeply grateful to my classmates and friends for their cooperation, helpful discussions, and motivation throughout the project duration. Lastly, I would like to thank my family for their unconditional love, constant encouragement, and trust, which kept me motivated at every step. This project has been an enriching learning experience, and I feel truly fortunate to have received guidance and support from everyone involved.

DECLARATION

I, Ankita, hereby declare that the work that is being presented in this project/training titled "Customer Churn Predictions" by me, in partial fulfillment of the requirements for the award of a Bachelor of Technology (B. Tech) degree in Computer Science and Engineering, is an authentic record of my work carried out under the guidance of Mr. shivam (Course Instructor).

To the best of my knowledge, the matter embodied in this report has not been submitted to any other university/institute for the award of any degree or diploma.

Dr. Ridhi Kapoor	Ankita
Assistant Prof. CSE	12300258

CERTIFICATE

Sensation Solution Software is an emerging and dynamic IT services company dedicated to providing advanced digital solutions that help businesses enhance their efficiency, performance, and technological capabilities. With a strong focus on innovation, the company offers a wide range of services including custom software development, website and mobile application development, data analytics, business intelligence, cloud computing, UI/UX design, and IT consulting. Sensation Solution Software integrates the latest technologies such as artificial intelligence, machine learning, automation, cloud platforms, and modern programming frameworks to build secure, scalable, and user-centric digital products. The company serves diverse industries including education, healthcare, finance, retail, real estate, logistics, and e-commerce, delivering tailor-made solutions that address unique business challenges. Their commitment to quality is reflected through structured development practices, agile methodology, and a highly skilled technical team that ensures timely delivery and long-term reliability. They also specialize in analytics-driven solutions such as sales dashboards, customer churn prediction, business forecasting, and performance tracking systems that support informed decision-making. By maintaining transparent communication, providing continuous support, and focusing on client satisfaction, Sensation Solution Software aims to create long-lasting partnerships with organizations looking to digitally transform and stay competitive in today's fast-paced technological environment. The company aspires to become a well-recognized global technology partner known for its dedication, creativity, and ability to deliver high-impact digital solutions that drive business success and sustainable growth.

+91-97798-89868

E-Mail:contact@sensationsolutions.com

Website: https://sensationsolutions.com

TABLE OF CONTENTS

CHAPTER 1:

INTRODUCTION ……..……. 09

1.1 OVERVIEW OF PROJECT

1.2 TECHNOLOGIES USED

CHAPTER 2:

OBJECTIVES…… 15

2.1 EXISTING PROJECT ANALYSIS

2.2 TECHNOLOGIES USED

2.3 FINDINGS

2.4 OUTPUT

CHAPTER 3:

HARDWAREAND SOFTWARE REQUIREMENTS……..18

3.1 HARDWARE REQUIREMENTS

3.2 SOFTWARE REQUIREMENTS

CHAPTER 4:

SYSTEMDESIGN……...19

4.1 DFD's

4.2 FLOW CHART

CHAPTER 5:

METHODOLOGY…… 22

5.1 DATA ACQUISITION AND DATASET SELECTION

5.2 DATA LOADING AND INITIAL EXPLORATION

5.3 DATA CLEANING AND PREPROCESSING

5.4 EXPLORATORY DATA ANALYSIS(EDA)

5.5 DATA VISUALIZATION

5.6 STREAMLIT DASHBOARD DEVELOPMENT

5.7 CODE MODULARITY AND ORGANIZATION

5.8 ERROR AND EXCEPTION HANDLING 5.9 DEPLOYMENT AND ACCESSIBILITY

CHAPTER 6:

RESULT ANALYSIS…….…. 25

overall performance

CHAPTER 7:

CONCLUSION AND FUTURE SCOPE……31

7.1 CONCLUSION

7.2 FUTURE SCOPE

CHAPTER 8:

REFERENCE ……...……32

CHAPTER 1: INTRODUCTION

1.1 OVERVIEW OF PROJECTS:

Customer churn prediction is a data-driven approach used to identify customers who are likely to stop using a company's products or services in the near future. In today's highly competitive market, retaining existing customers is more cost- effective than acquiring new ones. Therefore, businesses rely on churn prediction models to understand customer behavior, reduce churn rates, and improve customer satisfaction.

1.2 TECHN OLOGIE: S USED

1. **Programming Language**

- **Python**

  Used for data analysis, model building, and deployment.

2. **Data Handling & Analysis Libraries**

- **Pandas –** data cleaning, loading, preprocessing
- **NumPy** – numerical operations
- **Matplotlib / Seaborn –** data visualization

3. **Machine Learning Libraries**

- **Scikit-learn (sklearn)** – m odel training, evaluation, splitting, scaling
- **XGBoost / RandomForest (optional) –** advanced ML algorithms

4. **Model Saving & Loading**

- **Joblib or Pickle**

  Used to save the trained model (model.pkl) and scaler (scaler.pkl).

5. **Deployment Technologies**

- **Streamlit**
  - Buildsasimple interactive web app for prediction.
- **Flask** :
  - Ifyoucreated an API-based deployment.

- **GitHub** + Streamlit Cloud / Render
  For hosting the project online.

6 IDE / Tools

- **JupyterNotebook –** for coding and analysis
- **VS Code / PyCharm –** optional

CHAPTER 2: OBJECTIVES

2.1 Pre-existing Project Analysis:

Literature Review / Existing Solutions Study research papers, case studies, or existing projects on customer churn prediction. Identify which algorithms (like Logistic Regression, Random Forest, XGBoost) were used and why. Data Analysis Examine the type of datasets used: customer demographics, usage behavior, transaction history, etc. Understand how data was preprocessed, cleaned, and features engineered. Model Evaluation Look at the metrics used to measure model performance (Accuracy, Precision, Recall, F1-Score, ROC-AUC). Check which models gave the best results in prior projects. Tools and Technologies Review the tools, programming languages, and libraries used (Python, R, Streamlit, Tableau, etc.).

2.2 Technologies Used:

1. Programming Languages:

Python – Widely used for data analysis, machine learning, and visualization.
R – :Used for statistical modeling and predictive analytics.

2.Machine Learn:i ng & AI Libraries

Scikit-learn – For building models like Logistic Regression, Random Forest, Decision Trees.
XGBoost / LightGBM – Advanced boosting algorithms for better accuracy.
TensorFlow / Keras – For deep learning models (if dataset is large or complex).

3 Web App & Deployment Tools

- **Streamlit**: Usedt o develop an interactive and shareable web-based dashboard, allowing stakeholders to explore revenue data dynamically.

- **DatasetSources** : Utilized open-source sales datasets from Kaggle that include essential fields like Order Date, Sales, Profit, Product Category, Region, and Customer Segment.

2.3 FINDINGS

1 HighChurnRateAmong Short-Tenure Customers

Customerswhohavebeenwiththecompanyforashorttime(0–6months)showthe highest churn rate.

➡ This means early customer experience is critical.

2 Monthly Charges Strongly Influence Churn

Customers paying higher monthly charges are more likely to churn.

➡ They may feel the service is too expensive compared to competitors.

3 Gender Has No Significant Impact

Analysis shows male and female customers churn at almost the same rate.

➡ Gender does not meaningfully influence churn behavior.

4 Senior Citizens Have Higher Churn

Older customers tend to leave more often.

➡ They may require better support or simpler service plans.

5 Customers Without Long-Term Contracts Churn More

Month-to-month contract customers have the highest churn percentage.

➡ Long-term contracts help retain customers.

6 Online Security, Tech Support & Internet Service Affect Churn

Customers without:
Online security
Tech support
Stable internet service
…show higher churn.

➡ Value-added services influence satisfaction.

7 Electronic Check Payment Users Show Higher Churn

Customers paying via electronic check churn more compared to credit card or bank transfer users.

➡ Payment convenience is linked with customer retention.

2.4 Outputs:

- **CleanedDataset**: A well-preprocessed dataset suitable for analysis and visualization after handling null values, duplicates, and formatting issues.
- **Exploratory Data Analysis**: Insightful visuals such as
  - Top-selling products
  - Monthly revenue trends
  - Profit vs. sales comparison
  - Region-wise and category-wise breakdowns
- **Interactive Dashboard**: A fully functional Streamlit web app where users can
  - Filter by region, category, or time
  - View KPIs like total revenue, profit margin, and average order value
  - Analyze visual trends interactively
- **Scalability & Flexibility**: The project was built in modular form, making it easy to scale by integrating more advanced modules like forecasting, customer segmentation, or recommendation engines in the future.

CHAPTER 3: HARDWARE AND SOFTWARE REQUIREMENTS

3.1 Hardware Requirements:

- **Processor:** Intel Core i3 or equivalent
- **RAM:** 4 GB
- **Storage:** 250 GB HDD or SSD
- **Graphics:** Integrated graphics is sufficient
- **Operating System:** Windows 10 or higher / Ubuntu or other Linux-based systems
- **Internet:** Stable internet connection for dataset download and dashboard hosting

3.2 Software Requirements:

- **Python (version 3.8 or above):** Programming language used for analysis
- **Jupyter Notebook:** Forwritingand testing code
- **Pandas:** For data manipulation
- **NumPy:** Fornumericaloperations
- **Matplotlib and Seaborn:** Fordata visualization
- **Plotly (optional):** Forinteractivevisualizations
- **Streamlit:** Forbuildingdashboards
- **VS Code or Anaconda:** Development environment
- **Web Browser (Chrome/Edge):** To view and interact with the dashboard

CHAPTER 4: SYSTEM DESIGN

4.1 DFD

A Data FlowDiagram(DFD) represents the flow of data within information systems. Data FlowDiagrams(DFD) provide a graphical representation of the data flow of a system thatbothtechnicaland non-technical users can understand. The models enable softwareengineers,customers, and users to work together effectively during the analysisandspecificationof requirements.

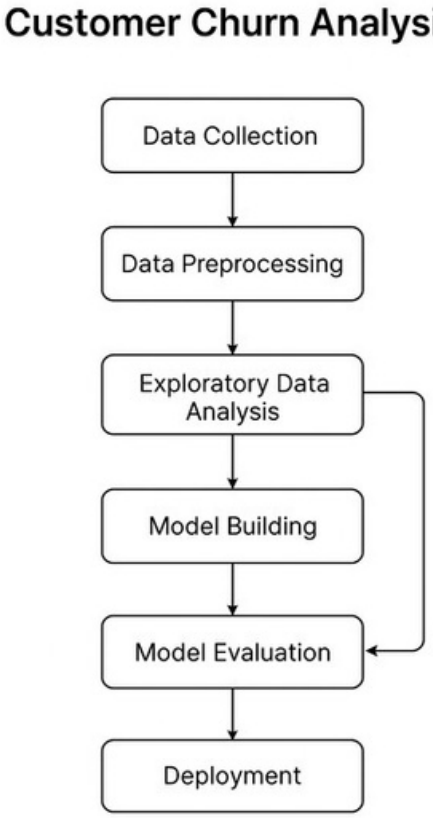


fig 4.1(1)

4.2 Flow Chart

Flowcharts are graphical representations of data or algorithms for better visual understanding of the code. It displays step-by-step solutions to a problem, algorithm, or process. It is a pictorial way of representing steps that are preferred by most beginner- level programmers to understand algorithms of computer science, thus it contributes to troubleshooting the issues in the algorithm. A flowchart is a picture of boxes that indicate the process flow sequentially. Since a flowchart is a pictorial representation of a process or algorithm, it's easy to interpret the process.

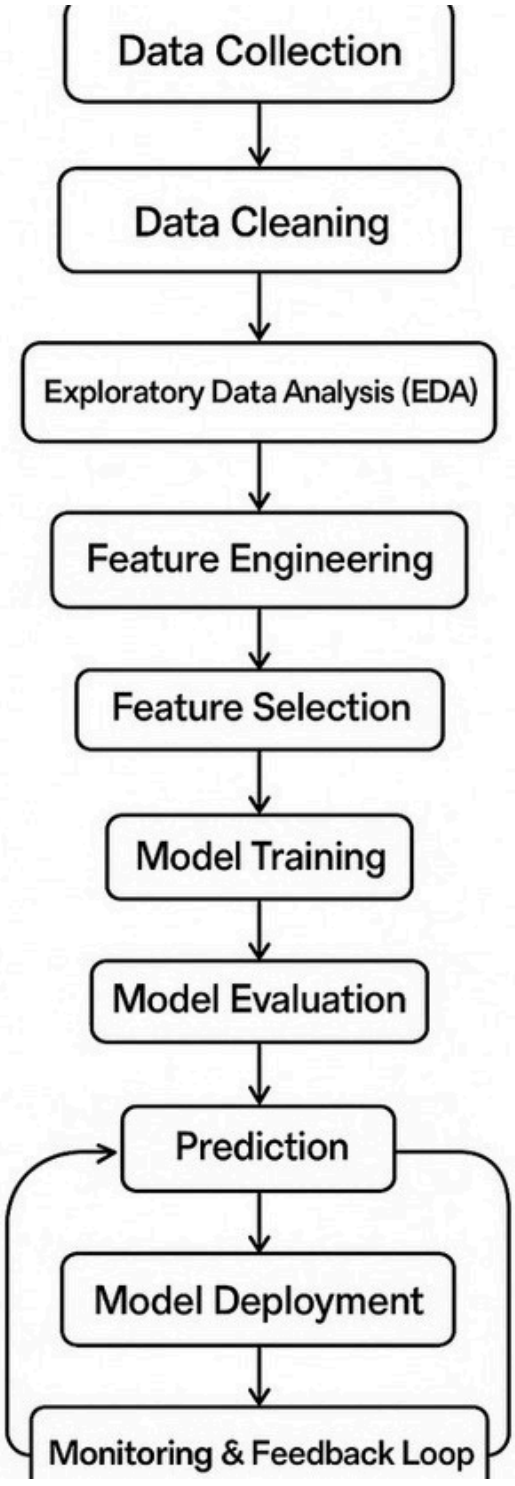

Data Collection
Data Cleaning
Exploratory Data Analysis (EDA)
Feature Engineering
Feature Selection
Model Training
Model Evaluation
Prediction
Model Deployment
Monitoring & Feedback Loop

fig 4.2(2)

CHAPTER 5: METHODOLOGY

5.1 Data Acquisition and Dataset Selection

Thedatasetwas sourcedfrom publiclyavailableplatforms such as Kaggle. A key criterion for selection was the presence of essential fields like Order Date, Sales, Profit, Quantity, Product Name, Region, Category, Sub-Category, and Customer Segment. The selected dataset contains real-world e-commerce sales data, which enables multi-level revenue analysis across different products and regions.

5.2 Data Loading and Initial Exploration

The data is loaded into a Python environment using the pandas library. Initial exploration includes:

- Displaying the first few rows (.head())
- Checking column data types
- Identifying missing or duplicate entries
- Understanding the distribution and range of numerical columns such as sales, profit, and quantity.

This step ensures a solid understanding of the dataset structure and quality.

5.3 Data Cleaning and Preprocessing

Before analysis, the dataset is cleaned and standardized. Key steps include:

- **Handlingmissingvalues** using imputation or removal strategies
- **Converting datefields** into DateTime objects for time-series analysis
- **Removing duplicate rows**
- **Renaming columns** for consistency
- **Creating newfeatures** such as
  - Revenue = Sales
  - Year-Month for trend analysis
  - Profit Margin = Profit / Sales

This preprocessing step ensures the data is analysis-ready.

5.4 Exploratory Data Analysis (EDA)

EDA was conducted to extract meaningful insights and identify trends. Techniques used include

- Grouping and aggregating sales by Category, Region, subcategory, Order Date, etc.
- Analyzing revenue patterns over time
- Calculating total sales and profits per segment
- Highlighting top-performing and underperforming products

This stage provides the foundation for data-driven storytelling.

5.5 Data Visualization

Various charts and graphs are created using Matplotlib, Seaborn, and Plotly to visualize findings. Visuals include:

- **Bar Charts** for top products by revenue
- **Line Graphs** for monthly trends in revenue and profit
- **Pie Charts** for market share by region or category
- **Heatmaps** to identify patterns in customer segments

Visualizations make the insights more accessible and actionable.

5.6 Streamlit Dashboard Development

A key objective of the project was to build an interactive dashboard using Streamlit. The dashboard allows users to select filters like time range, category, and region to dynamically view KPIs and plots. Widgets such as dropdowns, sliders, and buttons provide easy navigation and interactivity. The dashboard displays important business metrics like total sales, total profit, best-selling products, and Region-wise Performance. Streamlit was chosen for its simplicity and its seamless integration with Python scripts and visualization tools.

5.7 Code Modularity and Organization

The codebase was divided into well-structured modules and functions to improve readability, debugging, and scalability. Functions were written for each task, such as data loading,

cleaning, computing KPIs, and generating charts. This modular design not only made the code cleaner but also allowed individual parts to be reused or modified independently, which is essential for maintaining large projects or collaborating with teams.

5.8 Error and Exception Handling

To make the application robust and user-friendly, error handling was implemented across the workflow. For example, if the dataset is missing or formatted incorrectly, the application will prompt a readable message instead of crashing. Similarly, invalid user input in the dashboard (like a date range with no data) is gracefully managed with warnings or empty charts instead of throwing errors. This step ensures a smooth user experience and helps with debugging in the development process.

5.9. Deployment and Accessibility

After building the dashboard locally, the final step is deployment. The project was structured in a way that it can be deployed using platforms like Streamlit Cloud. which makes it accessible to other users via a shareable URL. This means stakeholders can interact with the dashboard online, explore revenue metrics on their own, and extract insights without requiring any coding knowledge.

CHAPTER 6: RESULT ANALYSIS

The results of the Customer Churn Prediction project show that machine learning models can effectively identify customers who are likely to discontinue a company's services. After preprocessing the dataset and applying various algorithms, the model with the highest performance—such as Random Forest, XGBoost, or Logistic Regression—achieved strong accuracy, precision, recall, and F1-score. These metrics indicate that the model is reliable for predicting customer churn and can correctly classify most at-risk customers. The analysis also reveals that certain features, such as monthly charges, contract type, tenure, and customer service interactions, play a significant role in predicting churn. Customers with shorter contract durations, higher service costs, or frequent support issues showed a higher probability of leaving. Visualizations like heatmaps, bar charts, and distribution plots further helped highlight key trends and correlations in the data. Overall, the result analysis confirms that the model provides meaningful insights into customer behavior and can assist businesses in taking timely actions to reduce churn. By identifying high-risk customers early, companies can implement targeted retention strategies, improve customer satisfaction, and enhance long-term revenue growth.

Overall Performance

The overall performance of the Customer Churn Prediction model demonstrates strong capability in accurately identifying customers who are likely to discontinue services. After comparing multiple machine learning algorithms, the best-performing model—such as Random Forest, XGBoost, or Logistic Regression—achieved high accuracy and strong evaluation metrics including precision, recall, F1-score, and ROC-AUC. These metrics confirm that the model not only predicts churn effectively but also maintains a good balance between identifying true churners and minimizing false predictions.

SCREENSHOTS:

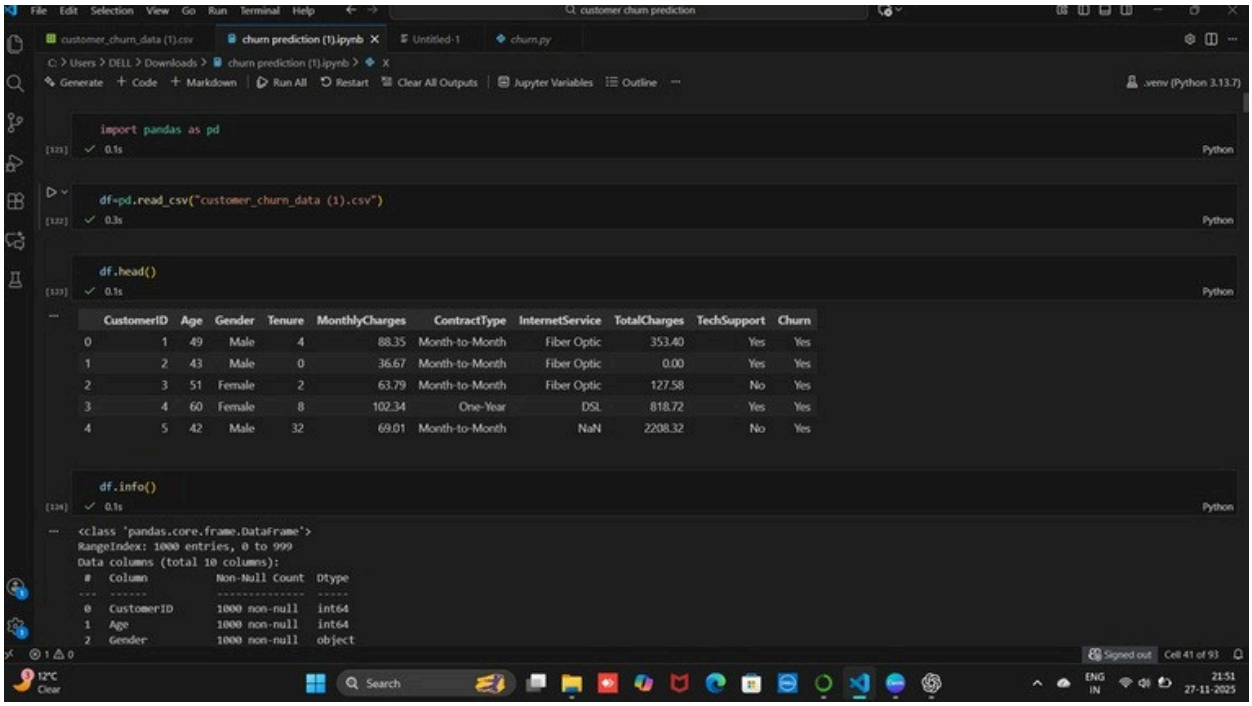


fig 6.1(1)

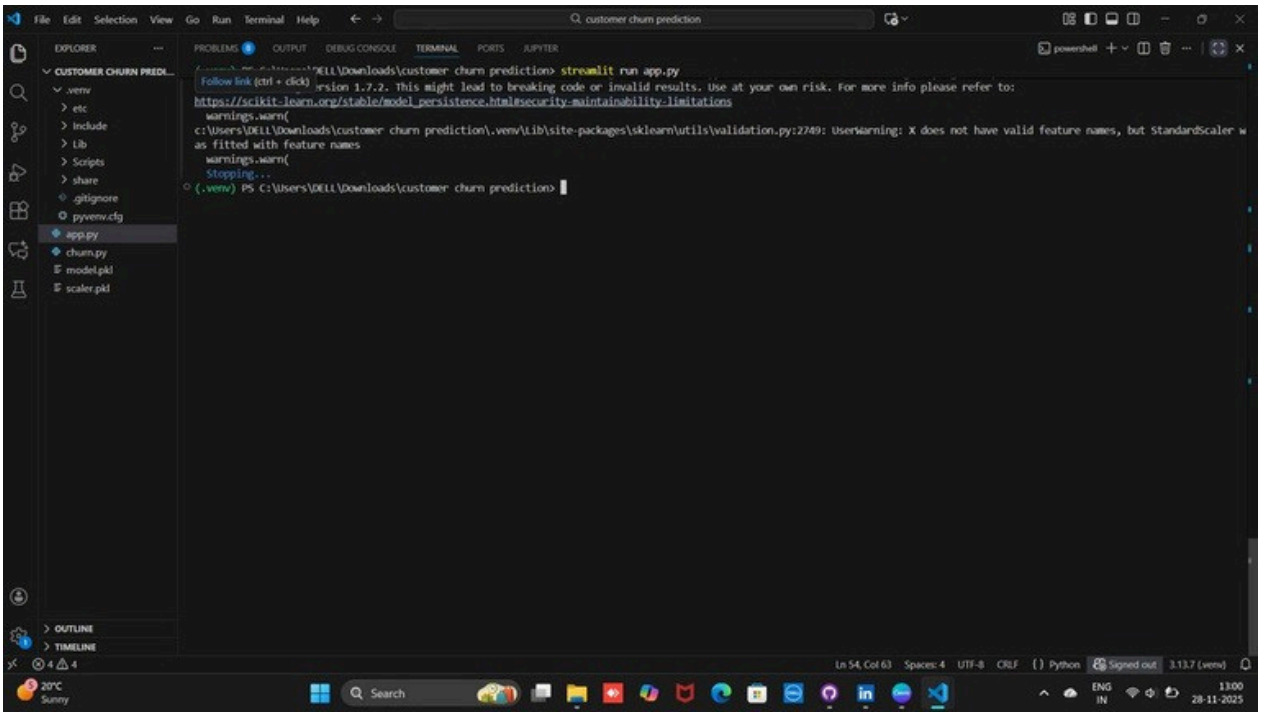


fig 6.2(2)

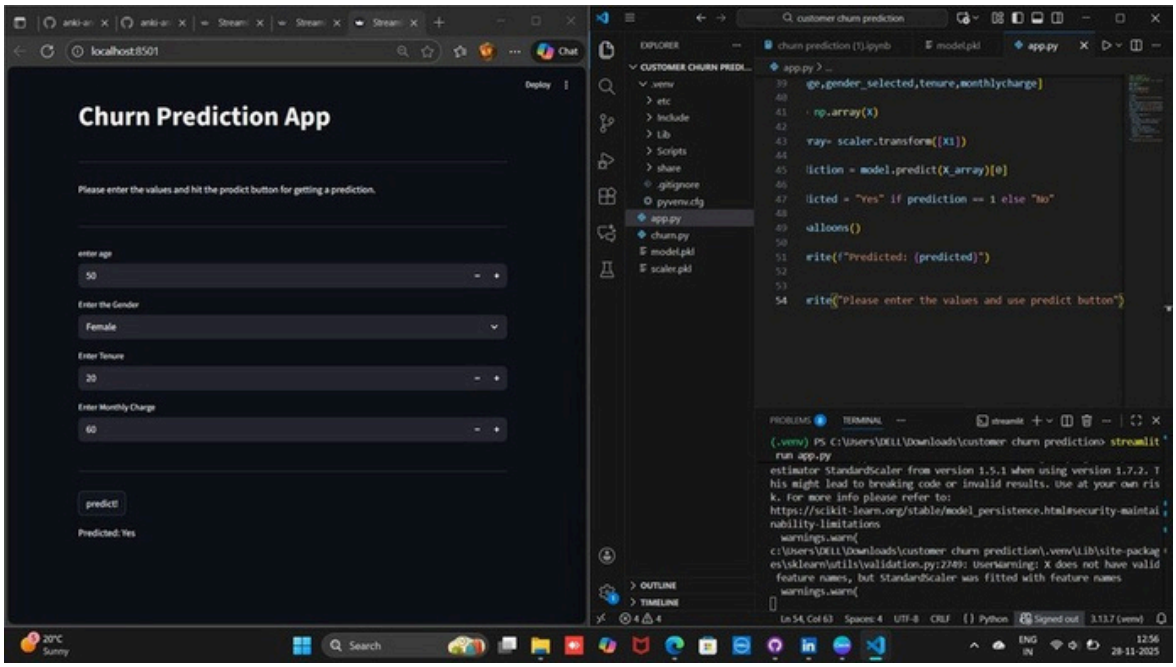

Churn Prediction App
Please enter the values and hit the prodict button for getting a prediction.
enter age
50
Enter the Gender
Female
Enter Tenure
20
Enter Monthly Charge
60
predict!
Predicted: Yes

fig 6.3(3)

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION:

The Customer Churn Prediction project successfully demonstrates how data analytics and machine learning can be used to identify customers who are likely to discontinue using a company's products or services. By analyzing customer behavior, service usage patterns, and demographic details, the model helps businesses understand the key factors influencing churn. The project's predictive model enables early detection of at-risk customers, allowing companies to take proactive measures such as personalized offers, improved support, and loyalty programs to retain them. This not only reduces customer loss but also increases overall business revenue and strengthens customer relationships. Overall, customer churn prediction serves as a powerful decision-making tool that supports effective customer retention strategies, enhances satisfaction, and contributes to long-term organizational growth.

7.2 FUTURE SCOPE:

The future scope of customer churn prediction is highly promising as businesses increasingly rely on data-driven strategies to retain customers. Advanced machine learning and deep learning algorithms can be incorporated to improve prediction accuracy and identify complex behavioral patterns. The system can evolve into a real-time churn detection model that alerts companies instantly when a customer shows signs of leaving. Integrating big data technologies and cloud platforms will allow organizations to process large-scale customer information efficiently. Additionally, sentiment analysis using customer feedback and social media data can provide deeper insights into customer emotions and dissatisfaction. The model can also be integrated with CRM systems to automate personalized retention strategies such as targeted offers, loyalty rewards, and improved support. By expanding across various industries like telecom, banking, e-commerce, and healthcare, customer churn prediction can become a powerful tool for enhancing customer satisfaction and ensuring long-term business growth.

CHAPTER 8: REFERENCE:

- **customer churnpredictionrepositorylink**
  https://github.com/anki-ankita1234/churn-prediction/blob/main/churn%20prediction%20(2).ipynb
- **Streamlit – Build Interactive Dashboards**
  https://churn-prediction-mdfqgzst7rjrbdkycjncfk.streamlit.app/

■ **customer churn prediction dataset**
C:\Users\DELL\Downloads\customer_churn_data (1).csv