**DEVELOPMENT OF ONLINE SHOPPER’S INTENSION**

A PROJECT REPORT

Submitted by

**BHUVANESH S,**

**THIYANESHWAR K,**

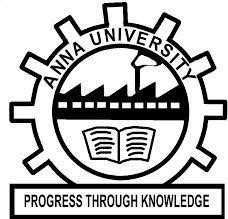
in partial fulfillment for the award of the degree of

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND BUSINESS SYSTEMS**

**CHENNAI INSTITUTE OF TECHNOLOGY CHENNAI-600 069**



**ANNA UNIVERSITY: CHENNAI 600 025**

# ANNA UNIVERSITY: CHENNAI 600 025

**BONAFIDE CERTIFICATE**

Certified that this report titled **Online shopper’s intension** is the bonafide work of **S Bhuvanesh (2104020244014)** and **Thiyaneshwar K(210420244057)** who carried out the work under my supervision.

|  |  |
| --- | --- |
| **SIGNATURE** | **SIGNATURE** |
| Dr. B.Sundarambal. | Dr. B.Sundarambal. |
| **HEAD OF THE DEPARTMENT** | **SUPERVISOR** |
| Associate Professor | Associate Professor |
| Department of Computer Science | Department of Computer Science |
| and Business Systems. | and Business Systems. |
| Chennai Institute of Technology, | Chennai Institute of Technology, |
| Pudhupedu, Kundrathur, | Pudhupedu, Kundrathur, |
| Chennai -600069 | Chennai -600069 |

**Submitted for viva-voce held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ at Chennai Institute of Technology,kundrathur.**

INTERNAL EXAMINER EXTERNAL EXAMINER

# ACKNOWLEDGEMENT

We convey our profound thanks and gratitude to our honorable Chairman Mr. P. Sriram, Chennai Institute of Technology, Chennai for providing an excellent academic climate, which made this endeavor possible.

We also express our gratitude to our beloved Principal Dr. A. Ramesh, who constantly nurtured our standard of education and devoted their precious time to our needs.

We are deeply indebted to pay our sincere thanks to our respectable Head of the Department and Supervisor Dr. B.Sundarambal, Department of Computer Science and Business Systems for showing some exuberant consent for our project and providing us with all the facilities in the department to complete the project.

We thank all our department teaching and non-teaching staff, who have helped directly and indirectly to complete this project on time.

Finally, we express our heartfelt and deep sense of gratitude to all faculty members in our division and to our friends for their helping hands, valuable support and encouragement during the project work.

# ABSTRACT

In the fast-paced world of e-commerce, understanding customer behaviour and purchase intentions is of utmost importance for businesses seeking to optimize their marketing strategies and increase revenue. This project aims to leverage the "Online Shoppers Intention dataset," a comprehensive collection of visitor attributes and website interaction metrics, to build a predictive model that can forecast whether a visitor is likely to make a purchase or not.

Through a binary classification approach, we will employ various machine learning algorithms to analyse the dataset and identify patterns and relationships between user attributes and their purchase decisions. The attributes, including administrative, informational, and product-related page visits, session duration, bounce rate, exit rate, and other factors, will serve as valuable predictors for the purchase intent.

**Table of contents**

|  |  |  |
| --- | --- | --- |
| **CHAPTER** | **TITLE** | **PAGE NO** |
|  | **ACKNOWLEDGEMENT** | **[3](#Acknowledgement)** |
|  | **ABSTRACT** | **[4](#Abstract)** |
| **1.** | **INTRODUCTION**  SUPERVISED LEARNING | **[6](#introduction)**  **[7](#SL)** |
| **2.** | **LITERATURE SURVEY**  EXISTING PROBLEM  PROPOSED SOLUTION | **[8](#EXISTING_PROBLEM)**  **[9](#PROPOSED_SOLUTION)** |
| **3.** | **THEORETICAL ANALYSIS**  BLOCK DIAGRAM  HARDWARE/SOFTWARE | **[11](#Theoretical_analysis)**  **[11](#Theoretical_analysis)** |
| **4.** | **EXPERIMENTAL INVESTIGATION** | [**12**](#Experi_investigation) |
| **5.** | **FLOWCHART** | [**14**](#Flowchart) |
| **6.** | **DATASET INFORMATION** | [**15**](#Dataset_information) |
| **7.** | **DATA VISUALIZTION** | [**17**](#Data_visualiztion) |
| **8.** | **MODEL BUILDING** | [**24**](#Model_Building) |
| **9.** | **DJANGO DEPLOYMENT** | [**28**](#Deployment) |
| **10.** | **ADVANTAGES AND DISADVANTAGES**  APPLICATION | [**31**](#advan)  [**32**](#appli) |
| **11.** | **CONCLUSION** | [**33**](#conclu) |

**CHAPTER 1**

**INTRODUCTION**

The rapid growth of e-commerce has revolutionized the way consumers shop, bringing convenience and a vast array of products right to their fingertips. In this digital era, understanding online shoppers' behaviour and purchase intentions has become paramount for businesses seeking to thrive in the competitive market. Analyzing user interactions on websites can provide valuable insights into customer preferences and help tailor marketing strategies to boost revenue.

This report presents a data-driven approach to predict online shoppers' purchase intentions using the "Online Shoppers Intention dataset." The dataset comprises a rich collection of attributes and features, including administrative, informational, and product-related page visits, session durations, bounce rates, exit rates, and more. Leveraging this diverse set of metrics, we aim to build an accurate and robust predictive model that determines whether a website visitor is likely to make a purchase or not.

The primary objective of this project is to develop a binary classification model that can efficiently identify potential customers who are more likely to convert into buyers. This will enable businesses to focus their marketing efforts on the right audience, optimize their website's user experience, and tailor their product offerings to meet customer demands.

## **1.1 Supervised Learnin****g**

Supervised machine learning is a fundamental subfield of artificial intelligence and machine learning. It involves training a model on labelled data, where each data point has corresponding input features and output labels. The primary goal of supervised learning is to build a predictive model that can make accurate predictions or decisions on new, unseen data.

The process of supervised learning begins with data collection, where a dataset with labelled examples is gathered. The data is then pre-processed to handle missing values, clean noisy data, and scale features if required. Next, a suitable model or algorithm is chosen based on the nature of the problem - classification or regression. During model training, the labelled data is fed into the selected algorithm, allowing the model to learn patterns and relationships between the input features and output labels. The model iteratively adjusts its internal parameters to minimize the prediction error and improve its ability to generalize to new data. After the model is trained, it is evaluated on a separate dataset called the test set. This evaluation helps determine how well the model generalizes to unseen data and assesses its performance using various evaluation metrics, such as accuracy, precision, recall, F1-score, and confusion matrix. Supervised learning has widespread applications in various domains. In image classification, models can learn to identify objects in images based on labelled examples. In natural language processing, supervised learning can be used to build sentiment analysis systems or language translation models. Credit risk assessment models in finance use supervised learning to predict a customer's creditworthiness based on historical data. Supervised learning's success lies in its ability to learn from labelled data and make informed predictions on new, unseen data. However, the quality and size of the labelled dataset play a vital role in the model's performance. In cases where obtaining labelled data is costly or time-consuming, techniques like semi-supervised and transfer learning can be employed to leverage both labelled and unlabelled data for better model performance.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 EXISTING PROBLEM**: Data analysis and machine learning experts have shown significant interest in addressing the challenge of predicting customer behaviour during online purchases. Previous research has explored various approaches to tackle this issue, considering factors such as browsing habits, past purchase history, demographics, and user preferences. These studies aimed to categorize customers into different groups based on their likelihood of making a purchase, utilizing either traditional statistical techniques or machine learning algorithms.

Surfing Habits: One approach considered in previous research is analyzing the browsing habits of online shoppers. This involves studying the duration of time spent on different product pages, the sequence of product views, and the intensity of interaction with various website elements. By understanding browsing patterns, it becomes possible to identify potential customers who exhibit behaviours indicating a higher propensity to make a purchase.

Previous Purchasing Behaviour: Another aspect examined is the analysis of a customer's past purchase history. This includes factors such as the frequency of purchases, average order value, and preferred product categories. By leveraging this information, it becomes possible to discern patterns and trends that can help predict future purchase behaviour.

Demographics: Demographic information, such as age, gender, location, and income level, has also been considered in predicting customer behaviour. Analyzing demographic data helps identify customer segments that are more likely to engage in online shopping and make purchases. It allows businesses to tailor their marketing strategies and product offerings to specific customer groups.

User Preferences: Understanding user preferences is crucial in predicting customer behaviour. By analysing data on customer preferences, such as product ratings, reviews, and wish lists, it becomes possible to identify the types of products that resonate with individual customers. This information can be leveraged to personalize the online shopping experience and make targeted product recommendations.

Classic Statistical Techniques: Some previous solutions employed traditional statistical techniques to categorize customers into distinct groups based on their likelihood of making a purchase. These techniques include regression analysis, clustering algorithms, and decision trees. Statistical models allow for interpretability and understanding of the factors influencing customer behaviour. Machine Learning Algorithms: Other studies focused on utilizing machine learning algorithms for customer behaviour prediction. Algorithms such as logistic regression, random forest, support vector machines, and neural networks have been employed to train models that can accurately classify customers into different categories. Machine learning models can capture complex patterns and interactions among various features, leading to more accurate predictions.

By considering these factors and employing various techniques, researchers have made significant strides in predicting customer behaviour during online shopping. These studies have contributed to the development of models and algorithms that enable businesses to anticipate customer actions, optimize marketing strategies, and enhance the overall online shopping experience.

**2.2 PROPOSED SOLUTION:**

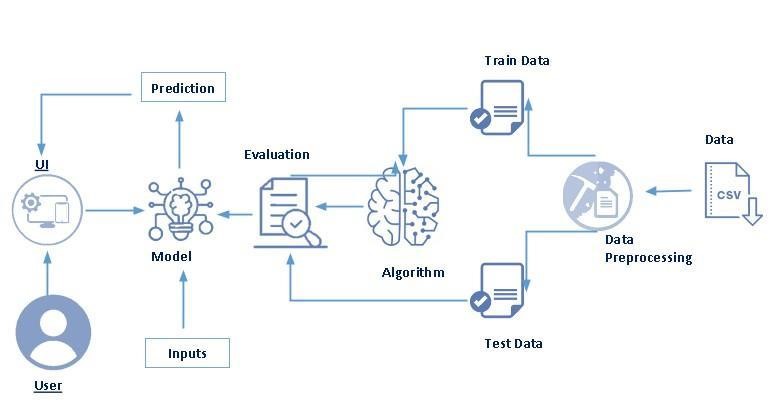
In this project, we propose to utilize classification algorithms, specifically Logistic Regression, Random Forest, KNN ,and Ada Boost Algorithms to develop a predictive model for online shopping behaviour. The proposed solution involves training the model using a dataset that consists of relevant features such as browsing time, product categories viewed, previous purchase history, and demographic information. The trained model will then be used to predict whether a customer is more likely to make a purchase or engage in window shopping. By leveraging these classification algorithms, we aim to achieve accurate predictions that can assist businesses in understanding and targeting their customer base effectively. The insights gained from this predictive model can help businesses tailor their marketing strategies, optimize their product offerings, and enhance the overall online shopping experience to improve conversion rates and customer satisfaction.

Once the best-performing model is identified, it will be saved in the joblib format, allowing for easy integration into existing systems and future use. This will enable businesses to make real-time predictions on customer behaviour during online shopping, thereby optimizing their decision-making processes and maximizing their chances of success.

**CHAPTER** **3**

### **THEORETICAL ANALYSIS**

**3.1 BLOCK DIAGRAM:**



**3.2 Hardware / Software designing:**

|  |  |
| --- | --- |
| HARWARE | 1. COMPUTER SYSTEM      1. INTERNET CONECTIVITY |
| SOFTWARE | 1. VS CODE      1. DJANGO      1. WORD      1. DATASET MANAGEMENT      1. PYTHON LANGUAGE AND LIBRARIES |

**CHAPTER 4**

**EXPERIMENTAL INVESTIGATIONS**

**Data Collection:** Gather online shopping data from various sources, such as e-commerce websites, APIs, or web scraping techniques. Collect data on browsing patterns, product categories viewed, previous purchase history, and demographic information of users.

**Data Pre-processing:** Clean the collected data by removing duplicates, handling missing values, and correcting inconsistencies. Encode categorical variables using techniques like one-hot encoding or label encoding. Normalize numerical features to ensure they are on a similar scale.

**Feature Selection:** Conduct exploratory data analysis to gain insights into the collected data. Use statistical techniques or feature importance methods (e.g., correlation analysis, information gain, or L1 regularization) to identify the most significant features. Select a subset of features that are highly correlated with the target variable (customer behaviour) and remove irrelevant or redundant features.

**Data Splitting:** Split the pre-processed data into training and testing datasets . Allocate a certain percentage of the data for training the models and the remaining portion for evaluating their performance.

**Model Training:** Apply classification algorithms such as Logistic Regression, Random Forest, and KMeans clustering to train predictive models. Configure the models with appropriate parameters and hyperparameters. Train each model on the training dataset using the selected features.

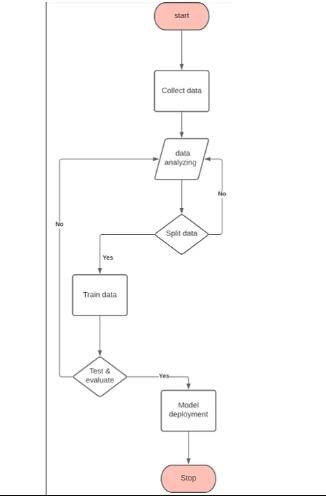
**Model Evaluation:** Evaluate the performance of each model using suitable evaluation metrics such as accuracy, precision, recall, and F1-score. Compare the performance of different models to identify the best-performing one. Assess the models' ability to predict customer behaviour during online shopping.

**Model Selection and Saving**: Select the best-performing model based on the evaluation results. Saving and selecting the model in the Joblib format for future use.

The above investigation provides a foundation for your project, laying the groundwork for subsequent steps such as setting up a Flask application, creating a user interface, handling user requests, and making predictions using the trained model.

**CHAPTER 5**

**FLOWCHART**



**CHAPTER 6**

**DATASET INFORMATION**

**Attributes/Features:**

Administrative: Represents the number of pages visited by the user on administrative pages of the website.

Administrative Duration: The total time spent by the user on administrative pages. Informational: Represents the number of pages visited by the user on informational pages of the website.

Informational Duration: The total time spent by the user on informational pages. Product Related: Represents the number of pages visited by the user on product-related pages of the website.

Product Related Duration: The total time spent by the user on product-related pages

.

Bounce Rate: The bounce rate of the website, which is the percentage of visitors who enter and leave the site without interacting further with it.

Exit Rate: The percentage of pageviews on the website that were the last in the session.

Page Value: Represents the average value of the pages viewed by the user before completing an e-commerce transaction.

Special Day: Indicates the closeness of the site visiting time to a special day (e.g., Mother's Day, Valentine's Day, etc.).

Month: The month of the year when the visit took place.

Operating System: The operating system used by the visitor.

Browser: The browser used by the visitor

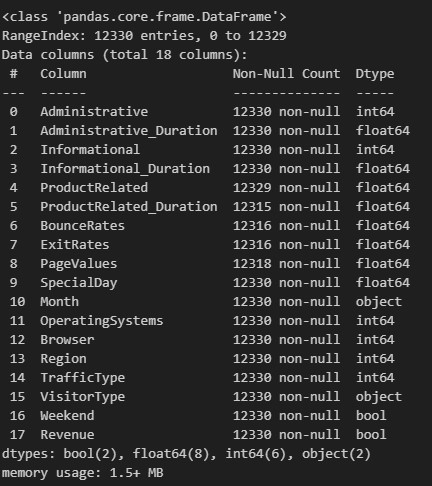
Region: The geographical region from which the visitor is browsing the website.

Traffic Type: The type of web traffic, such as direct, referral, search engine, etc.

Visitor Type: Indicates whether the visitor is a new visitor, returning visitor, or other.

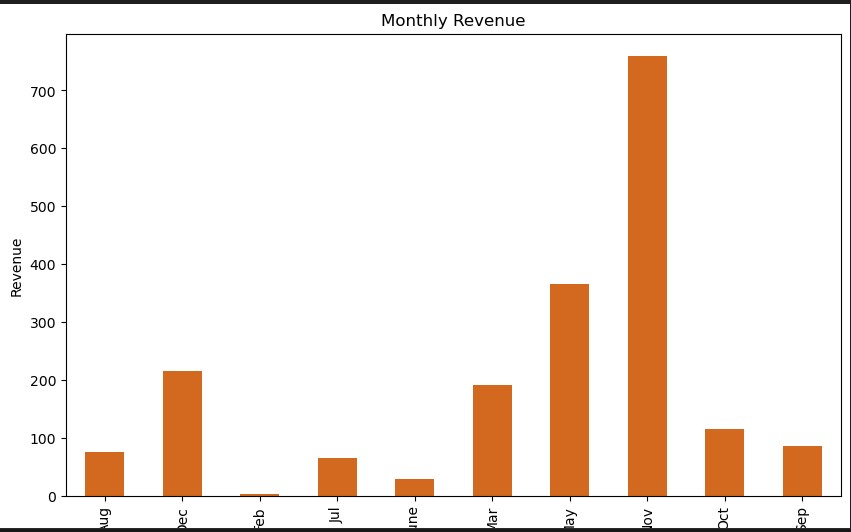
Weekend: A binary variable indicating whether the visit occurred on a weekend or not

Revenue: The target variable indicating whether the visitor made a purchase (1) or not (0).



**CHAPTER 7**

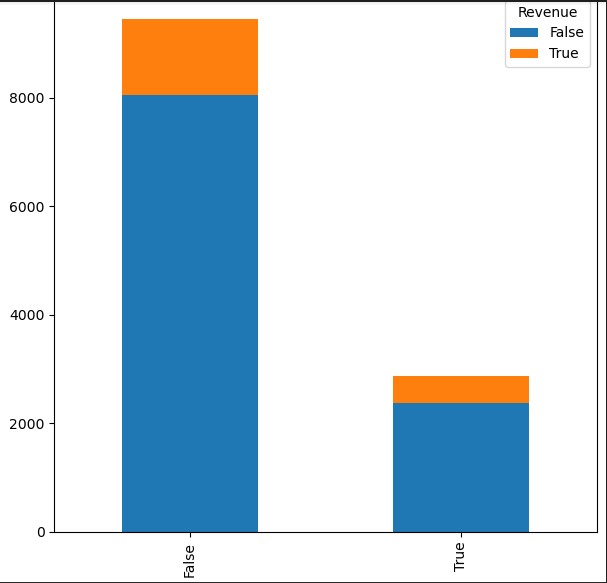
**DATA VISUALIZATION**





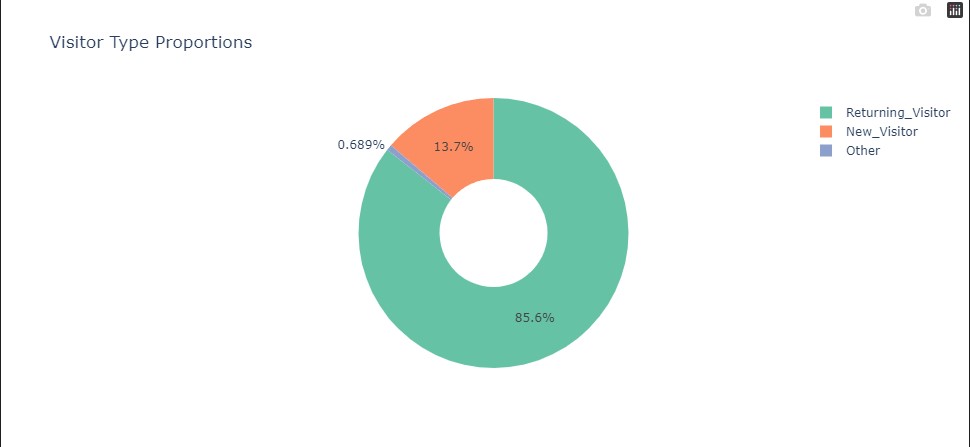
**Inference:**

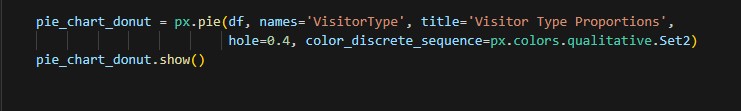
This bar chart visualizes the monthly revenue generated by the e-commerce website. It can help identify seasonal patterns or months with higher or lower revenue, allowing businesses to plan marketing campaigns or allocate resources accordingly.



**Inference:**

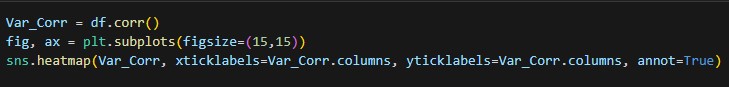
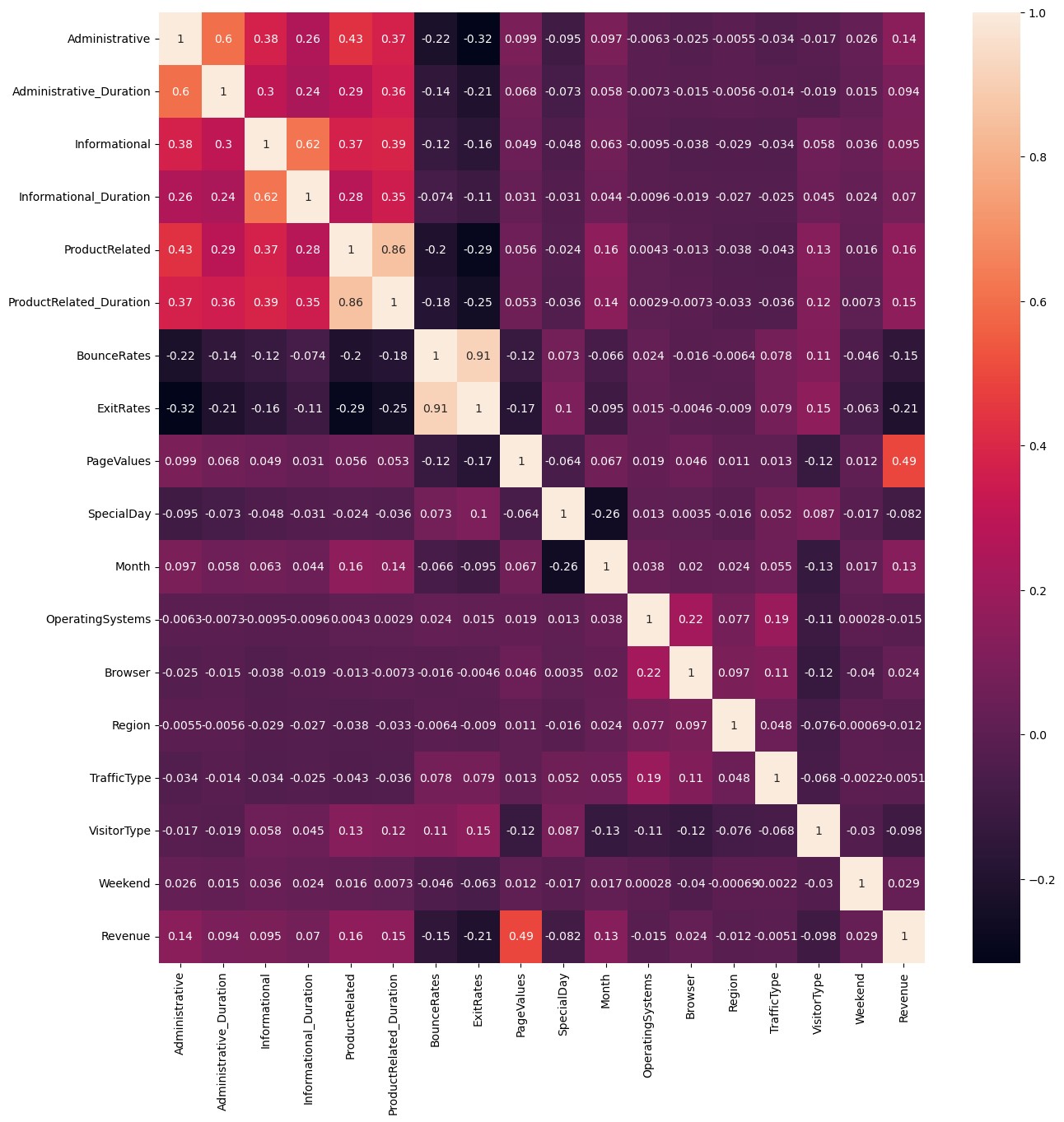
There is less product sold on weekdays and comparably more on weekends**.**

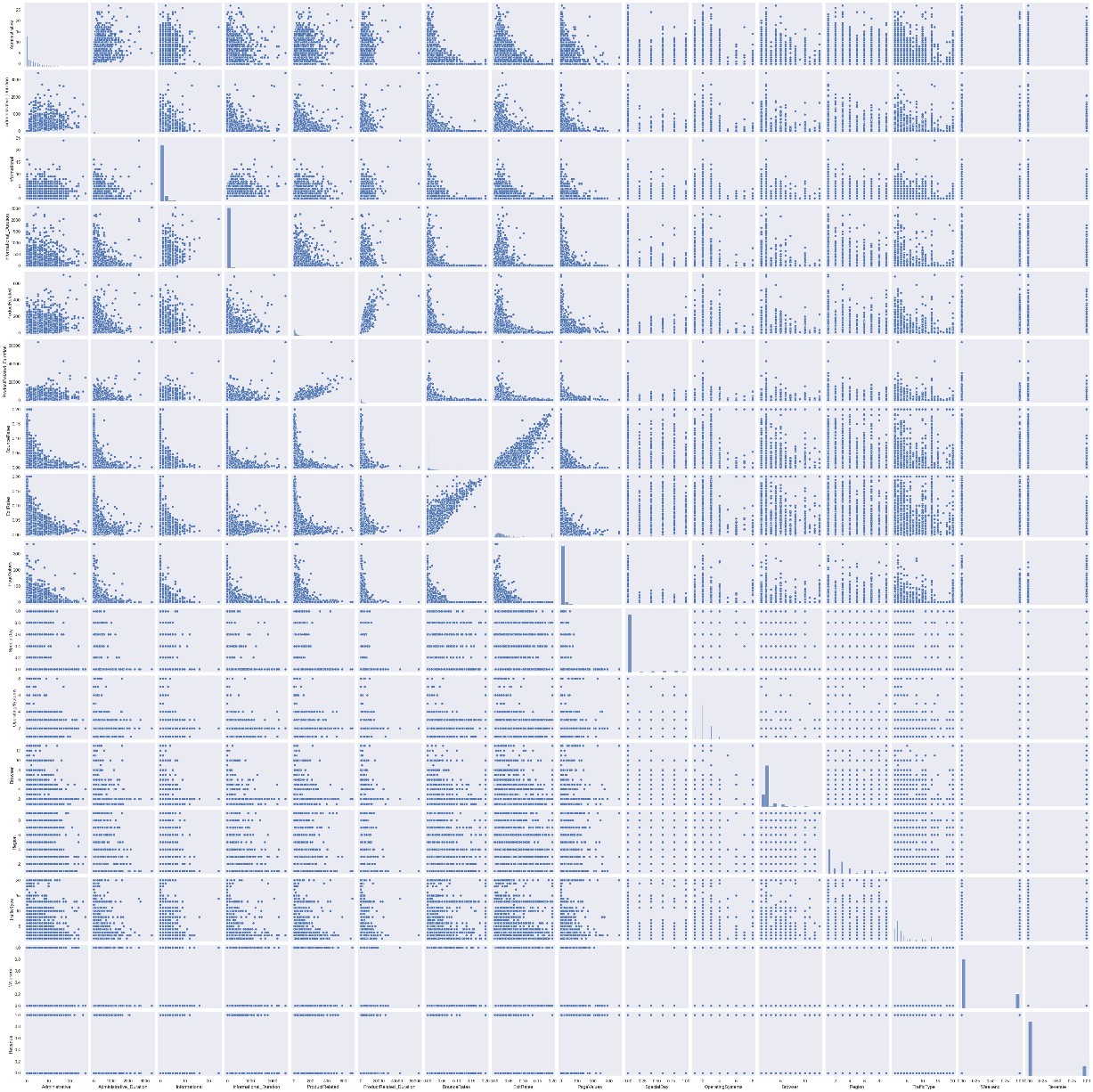




**Inference :**

This donut-style pie chart shows the proportion of different visitor types (new visitors, returning visitors, and others). The central hole (hole=0.4) in the donut chart style creates a visually appealing representation. Analyzing this chart can provide insights into customer loyalty and retention. A larger "donut" for returning visitors suggests a strong customer base, while a higher proportion of new visitors might indicate effective marketing strategies in attracting new customers.





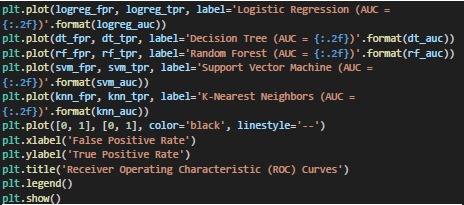


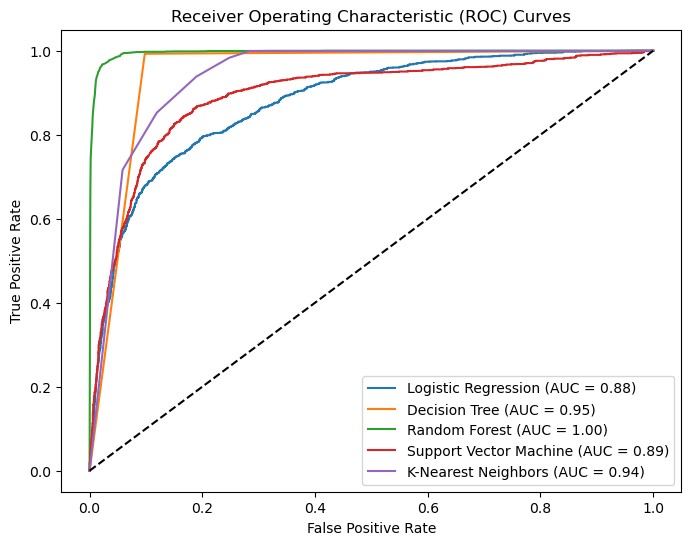
**Inference:**

This visualization show about the correlation between attributes.

**ROC CURVE:**



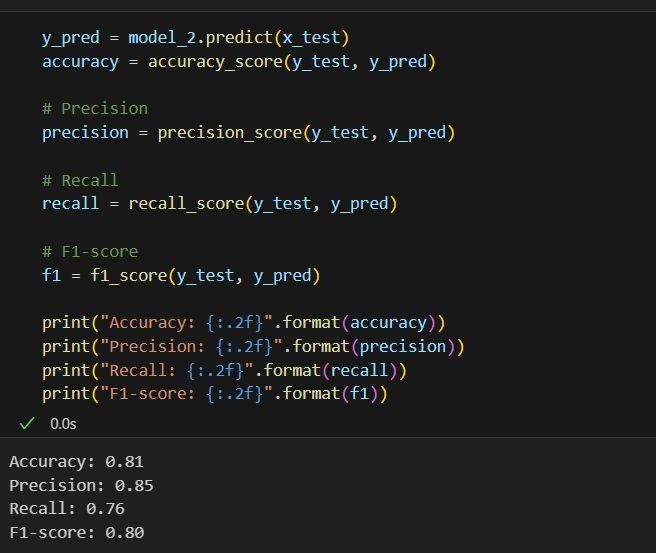
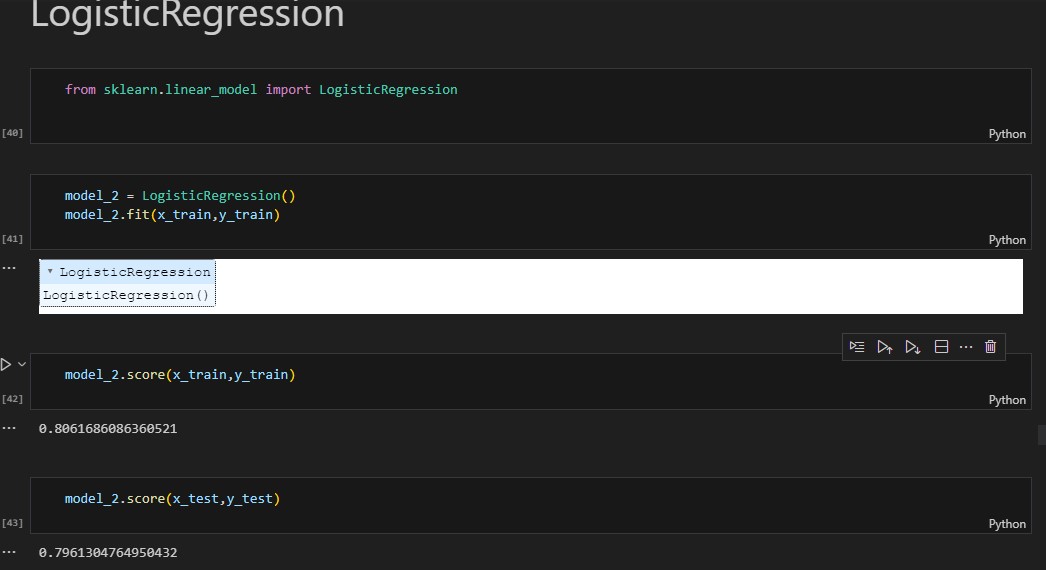




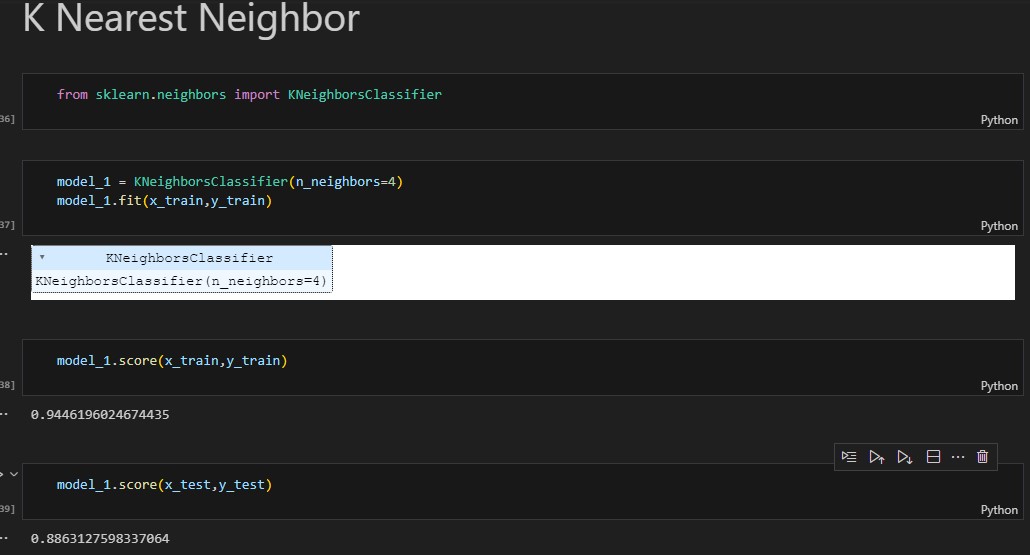
**CHAPTER 8**

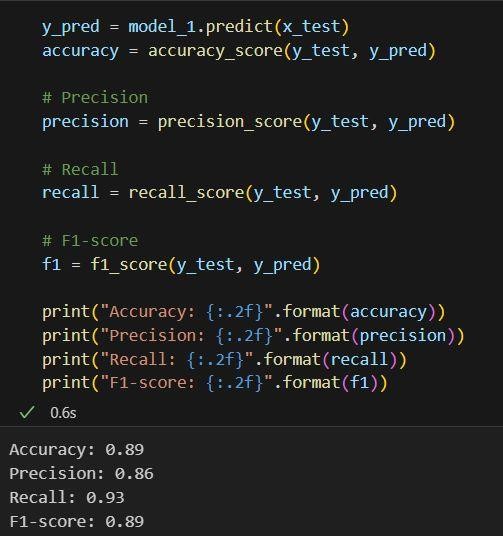
**MODEL BUILDING**

**8.1 LOGISTIC REGRESSION:**

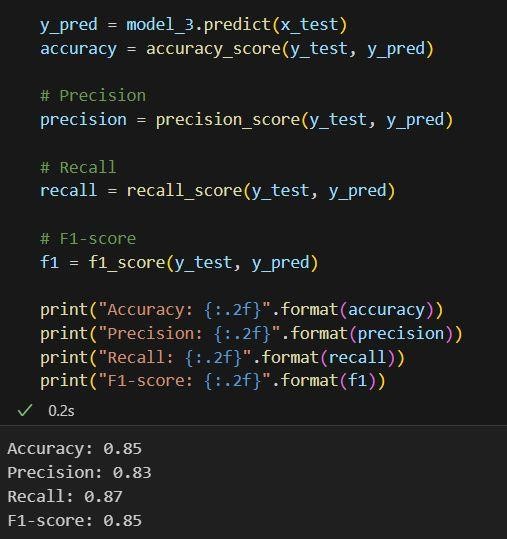
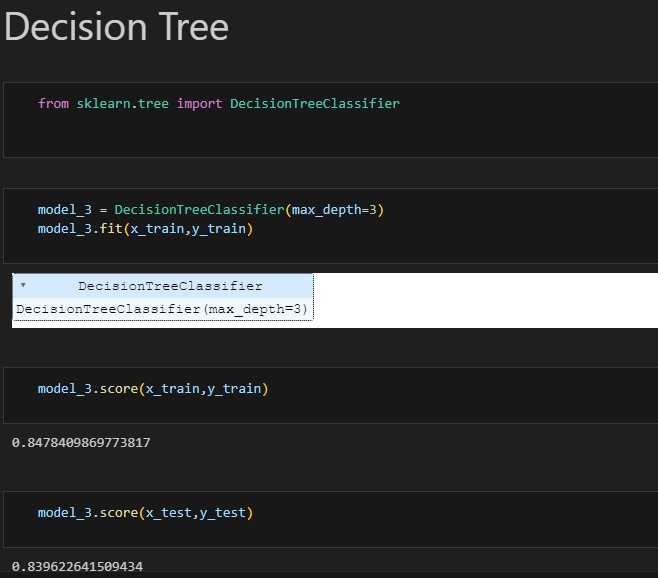


### **8.2 KNN :**

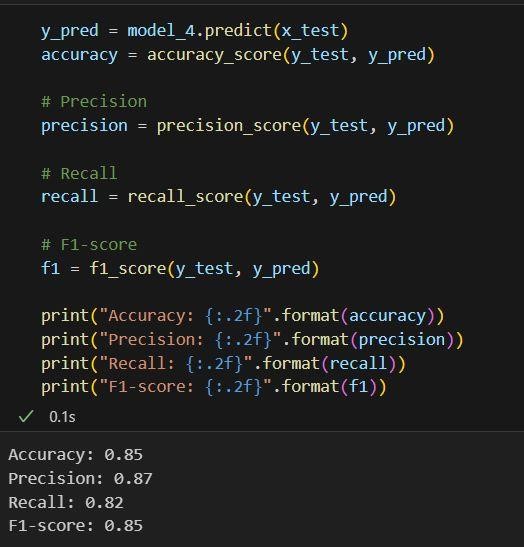
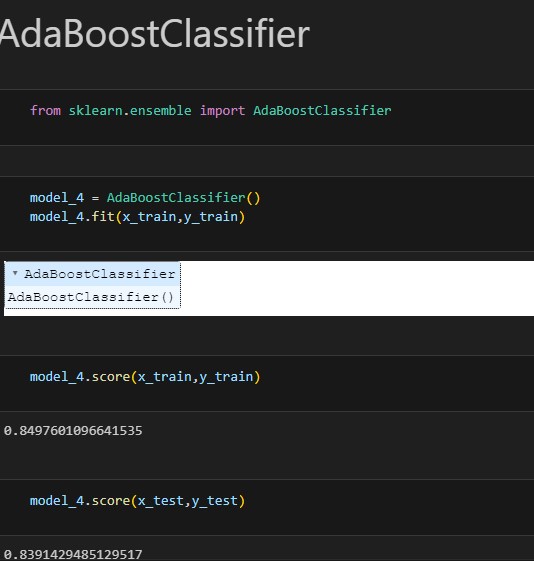




**8.3 DECISION TREE:**



**8.4 AdaBoostClassifier:**



**CHAPTER 9**

**DJANGO DEPLOYMENT AND WEB PAGE UI AND**

**OUTCOME:**

**What is Django?**

Django is a high-level, open-source web framework written in Python that follows the "batteries-included" philosophy. It is designed to make web development faster and easier by providing a robust set of tools and libraries for common tasks, allowing developers to focus on building web applications rather than dealing with low-level details. Django was created in 2005 by Adrian Holovaty and Simon Willison and has since become one of the most popular web frameworks in the Python ecosystem.

**Key features and reasons for using Django**:

1. Rapid Development: Django's "batteries-included" approach provides readyto-use components for common web development tasks, such as URL routing, database ORM, user authentication, and more. This enables developers to build web applications quickly and efficiently.

2. Scalability: Django is designed to handle high-traffic and large-scale applications. It follows best practices to ensure that applications can scale up to meet increasing demands.

1. ORM (Object-Relational Mapping): Django's built-in ORM allows developers to interact with the database using Python classes and objects, rather than writing SQL queries directly. This abstraction simplifies database operations and enhances code readability.

1. Admin Interface: Django automatically generates an admin interface for managing database records. It allows developers to perform CRUD (Create, Read, Update, Delete) operations on database models without writing custom admin interfaces.

1. Security: Django emphasizes security and comes with built-in features to prevent common web vulnerabilities, such as SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF).

1. URL Routing and View Handling: Django's URL dispatcher maps URLs to views, making it easy to organize different parts of the web application. Views handle HTTP requests and return appropriate responses.

1. Template Engine: Django's template engine allows developers to create dynamic HTML templates, separating design from logic. This promotes code reusability and enhances front-end development.

1. Form Handling: Django provides a form handling system that simplifies form creation, validation, and data cleaning. It streamlines user input and data processing.

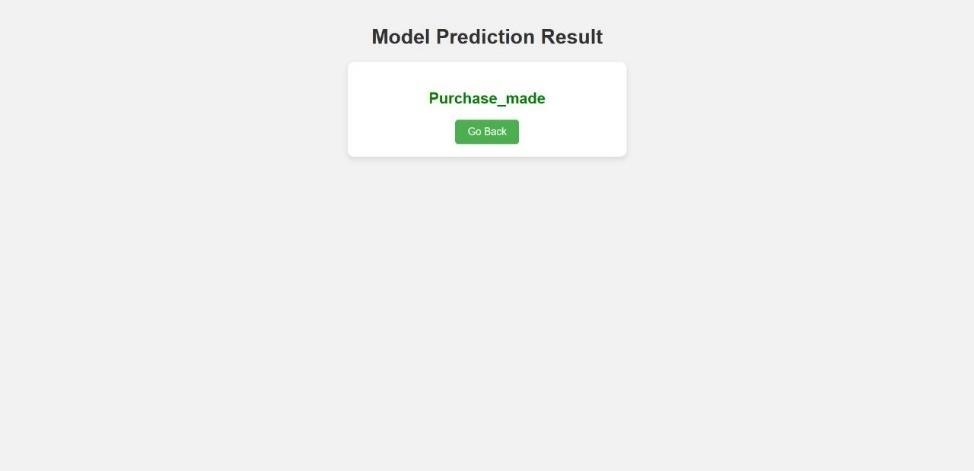
1. Internationalization and Localization: Django supports multiple languages and provides tools for translating web applications into different languages, making it suitable for creating multilingual applications.

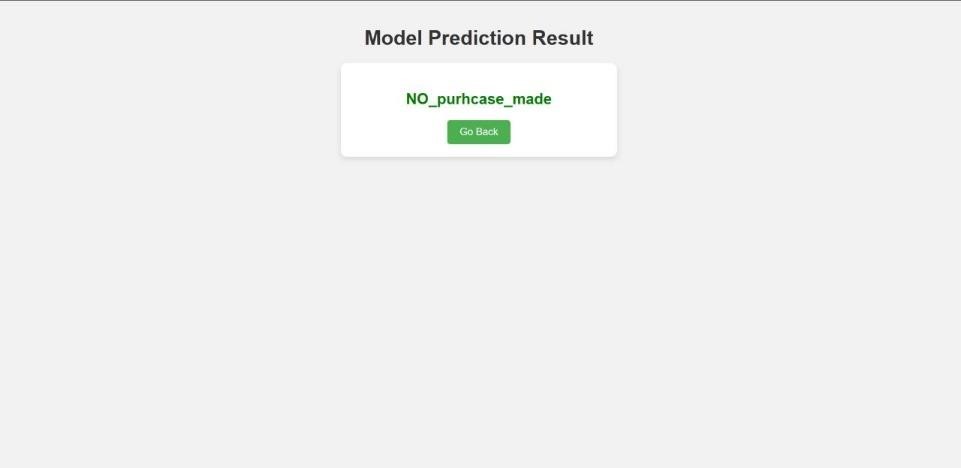
1. Versatility: Django can be used to build not only traditional web applications but also RESTful APIs, making it suitable for creating web services and integrating with other platforms.

1. Active Community :Django has a large and active community of developers, which means extensive documentation, third-party packages, and community support are available.

**USER INTERFACE :**







**CHAPTER 10**

### **ADVANTAGES & DISADVANTAGES**

**Advantages of the proposed solution:**

* Accurate prediction of customer behaviour during online shopping, enabling businesses to optimize their marketing strategies and improve conversion rates.

* Insights into customer preferences and browsing patterns, allowing businesses to tailor their product offerings and enhance the overall shopping experience.

* The ability to make data-driven decisions based on customer segmentation and targeted marketing.

**Disadvantages of the proposed solution:**

* Dependency on the quality and availability of data for accurate predictions.

* The need for continuous monitoring and updating of the predictive model as customer behaviour and preferences change over time.

* Ethical consideration regarding the collection and usage of customer data, ensuring privacy and compliance with data protection regulations

### **10.1 APPLICATIONS:**

The proposed solution can be applied in various areas, including:

E-commerce platforms: Predicting customer behaviour to optimize product recommendations, personalize marketing campaigns, and improve customer satisfaction.

Digital marketing:

Targeting specific customer segments with tailored advertisements and promotional offers.

Market research:

Analyzing online shopping behaviour to identify trends, preferences, and patterns for strategic decision-making.

Customer relationship management:

Understanding customer preferences to provide personalized customer support and enhance customer loyalty

**CHAPTER 11**

## **Conclusion**

In conclusion, the "Online Shoppers Intention" dataset provides valuable insights into user behaviour on an e-commerce platform. We analyzed various features related to online shopping activities and identified key trends that can inform business decisions.

Our analysis highlighted the significant impact of special days, such as weekends and holidays, on online shopping behaviour. It revealed that returning visitors demonstrated higher engagement and conversion rates compared to new visitors.

Bounce rate emerged as a critical factor affecting user engagement, with high bounce rates associated with lower chances of revenue generation. Additionally, different traffic sources showed varying degrees of influence on revenue generation.

While this report focused on exploratory data analysis and insights, there is potential for further work involving machine learning models to predict online shoppers' intentions. Deploying predictive models using web frameworks like Django can offer a user-friendly interface for visualizing model predictions and insights.

Despite the insights gained, the dataset's limitations, such as its size and missing information, need to be considered. Further research using larger and more up-to-date datasets could yield more robust findings.

To enhance the online shopping experience, businesses may consider implementing realtime monitoring, personalized user experiences, and A/B testing on web page layouts and features.

Overall, the "Online Shoppers Intention" dataset serves as a foundation for understanding user behaviour in the e-commerce domain and offers valuable guidance for businesses aiming to optimize revenue generation and improve their online presence..