SMART INTERNZ EXTERNSHIP PROJECT DOCUMENTATION

VELLORE INSITUTE OF TECHNOLOGY, VELLORE

PROJECT TITLE	ONLINE SHOPPERS INTENSIONS

TEAM MEMBERS DETAILS – TEAM 512

NAME	MAIL ID	COLLEGE REG NO
PAVITHRA. A	Pavithra.a2020@vitstudent.ac.in	20MID0210
NARENDIRAN. R	narendiran.2020@vitstudent.ac.in	20MID0212
SRIDHARAN. C	sridharan.c2020@vitstudent.ac.in	20MID0185
DINESH KUMAR. S	dineshkumar.s2020a@vitstudent.ac.in	20MID0152

ONLINE SHOPPERS INTENSION

1. INTRODUCTION

1.1 OVERVIEW

The goal of this research is to forecast a customer's behaviour when shopping online, including whether they will buy something or just window browse. The project attempts to train and evaluate the data using classification techniques including Logistic Regression, Random Forest, and the clustering method K-Means before choosing the most accurate model for prediction. The forecast can offer useful information about consumer behaviour and help firms better understand client preferences and the online buying experience.

1.2 PURPOSE

The purpose of this project is to develop a predictive model that can accurately determine whether a customer is likely to make a purchase or engage in window shopping while performing online shopping activities. By analyzing various factors and utilizing classification algorithms, the project aims to provide valuable insights to businesses, enabling them to make informed decisions regarding their marketing strategies, product offerings, and overall customer experience. With this predictive capability, businesses can optimize their resources and tailor their approaches to effectively target potential customers and enhance their conversion rates.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

Data analysis and machine learning experts have shown significant interest in addressing the challenge of predicting customer behavior 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 behaviors indicating a higher propensity to make a purchase.

Previous Purchasing Behavior: 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 behavior.

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 behavior. Machine Learning Algorithms: Other studies focused on utilizing machine learning algorithms for customer behavior 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.

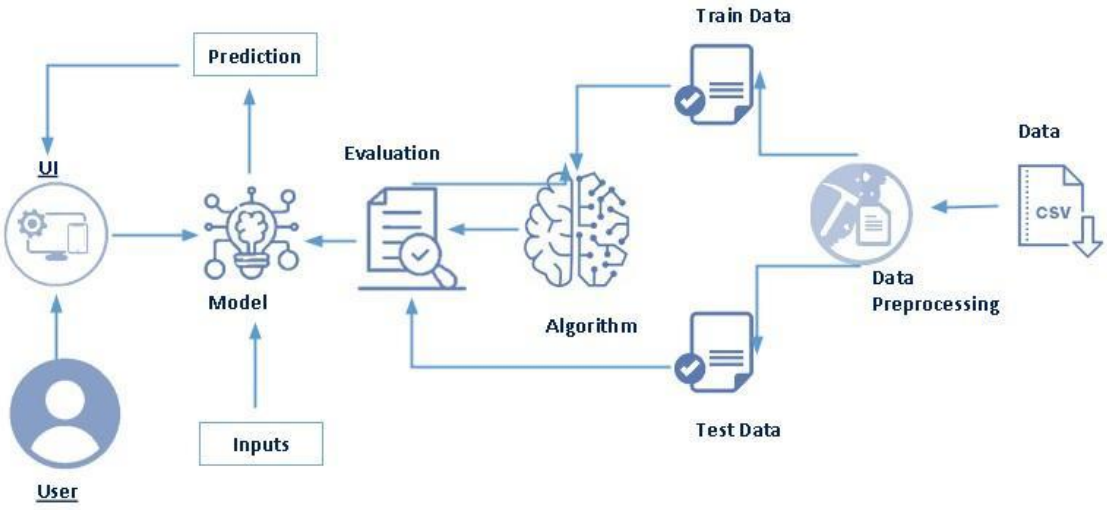
2.2 PROPOSED SOLUTION

In this project, we propose to utilize classification algorithms, specifically Logistic Regression, Random Forest, KNN, SVM, Naïve Bayes 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 pkl format, allowing for easy integration into existing systems and future use. This will enable businesses to make real-time predictions on customer behavior during online shopping, thereby optimizing their decision-making processes and maximizing their chances of success.

3. THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 Hardware / Software designing

HARDWARE	1. COMPUTER SYSTEM 2. INTERNET CONECTIVITY
SOFTWARE	1. JUPITER NOTEBOOK 2. FLASK 3. WORD 4. DATASET MANAGEMENT 5. PYTHON LANGUAGE AND LIBRARIES

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 Preprocessing: 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 behavior) and remove irrelevant or redundant features.

Data Splitting: Split the preprocessed 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.

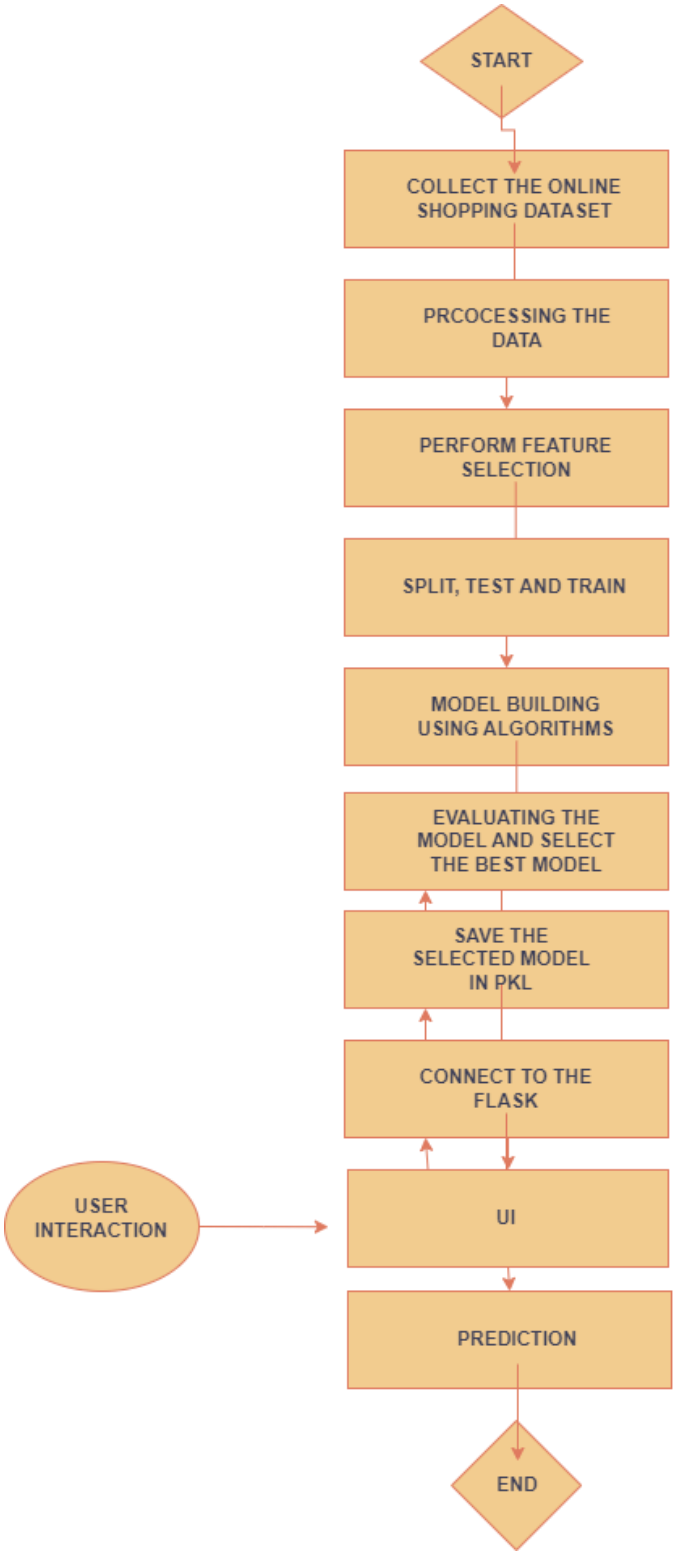
Model Training: Apply classification algorithms such as Logistic Regression, Random Forest, and K-Means clustering to train predictive models. Configure the models with appropriate parameters and hyperparameters. Train each model on the training dataset using the selected features.

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 behavior during online shopping.

Model Selection and Saving: Select the best-performing model based on the evaluation results. Saving and selecting the model in the pkl (pickle) 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.

5. FLOWCHART



6. RESULTS:

The final findings or outputs of the project include the following:

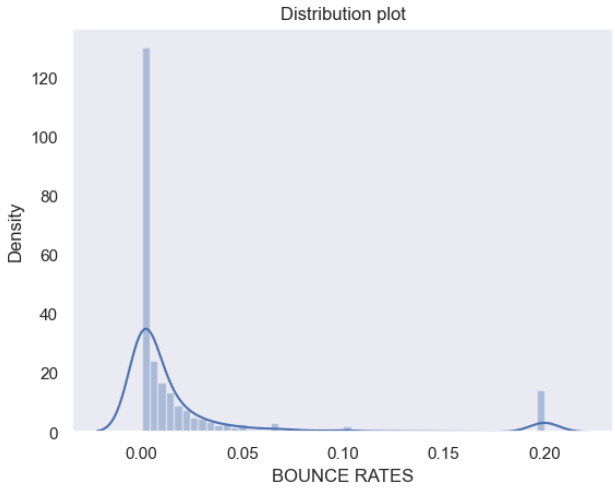
DATASET INFO:

```
In [155]: #data type
          df.info()
```

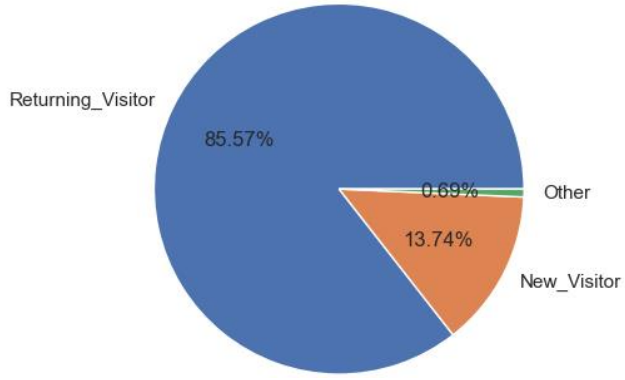
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12330 entries, 0 to 12329
Data columns (total 18 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Administrative           12330 non-null  float64
 1   Administrative_Duration  12330 non-null  float64
 2   Informational            12330 non-null  float64
 3   Informational_Duration   12330 non-null  float64
 4   ProductRelated           12330 non-null  float64
 5   ProductRelated_Duration  12330 non-null  float64
 6   BounceRates              12330 non-null  float64
 7   ExitRates                12330 non-null  float64
 8   PageValues               12330 non-null  float64
 9   SpecialDay               12330 non-null  float64
 10  OperatingSystems         12330 non-null  float64
 11  Browser                  12330 non-null  int64
 12  Region                   12330 non-null  float64
 13  TrafficType              12330 non-null  float64
 14  Weekend                  12330 non-null  bool
 15  Revenue                  12330 non-null  bool
 16  Month                    12330 non-null  int32
 17  VisitorType              12330 non-null  int32
dtypes: bool(2), float64(13), int32(2), int64(1)
memory usage: 1.4 MB
```

UNIVARIATE ANALYSIS:

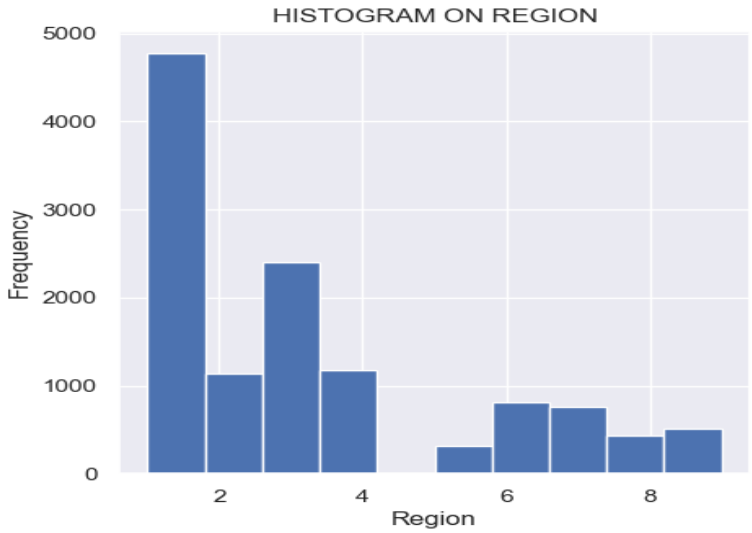
1. SNS DISTRIBUTION PLOT



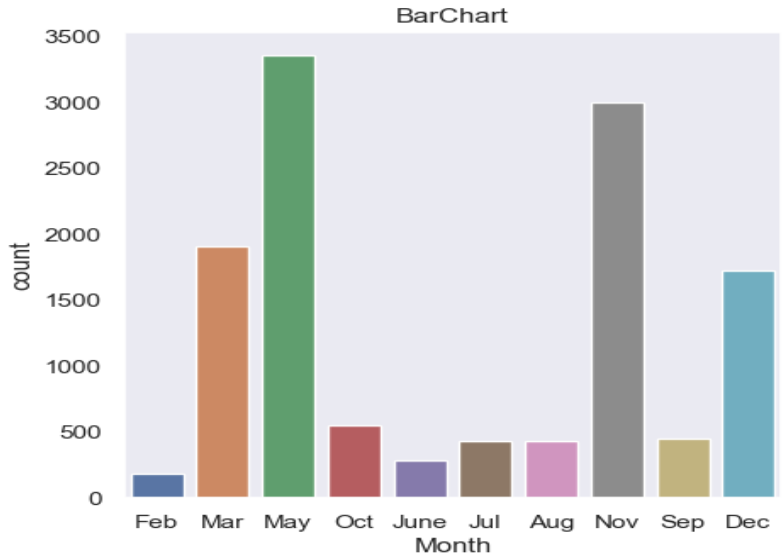
2. Pie plot



3. Histogram

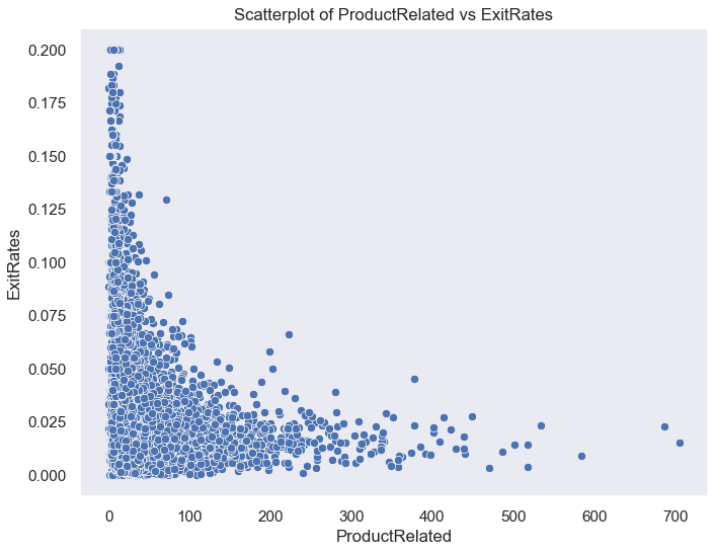


4. BarChart

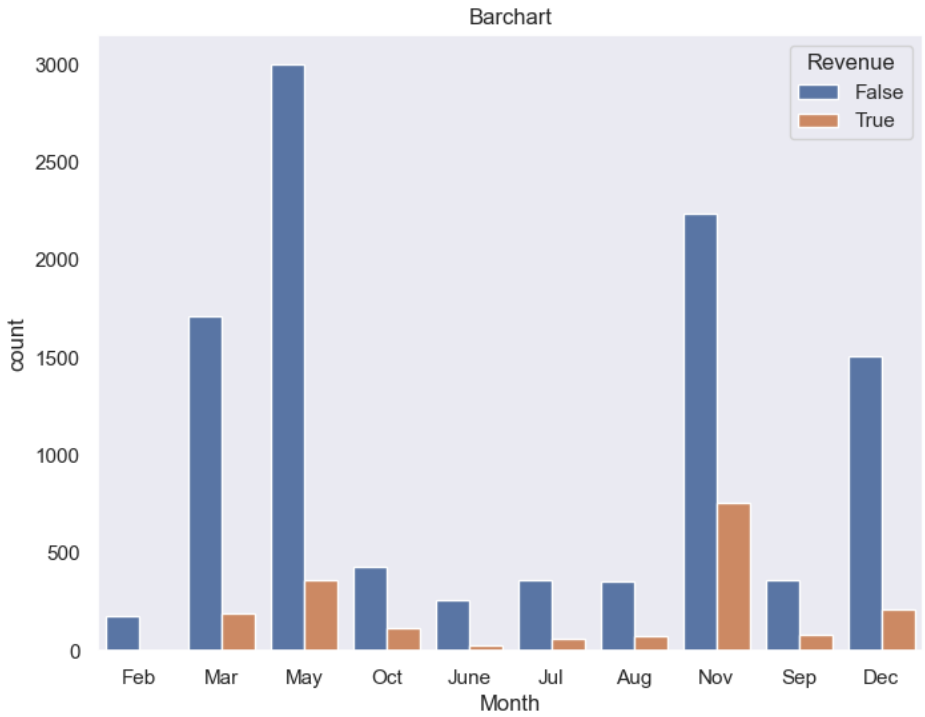


BIVARIATE ANALYSIS:

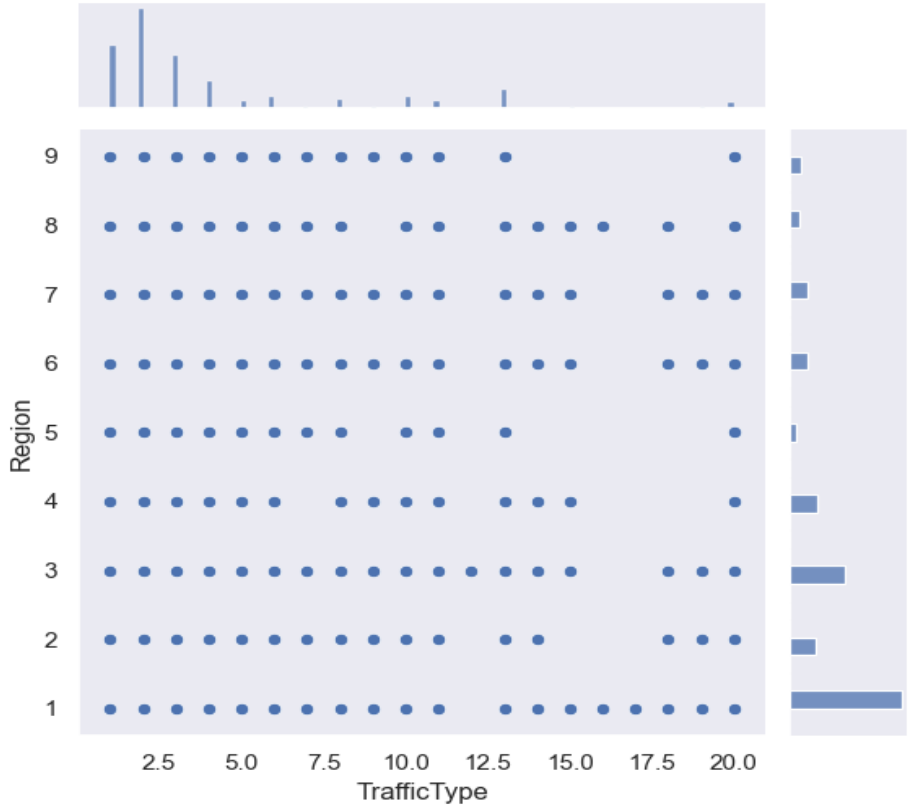
1. SCATTER PLOT



2. BARCHART

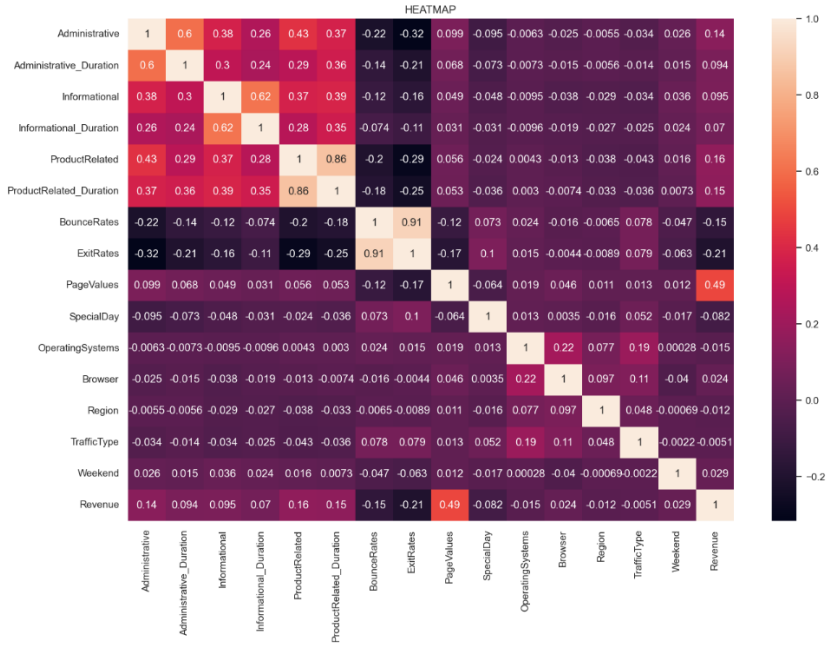


3. JOINT PLOT

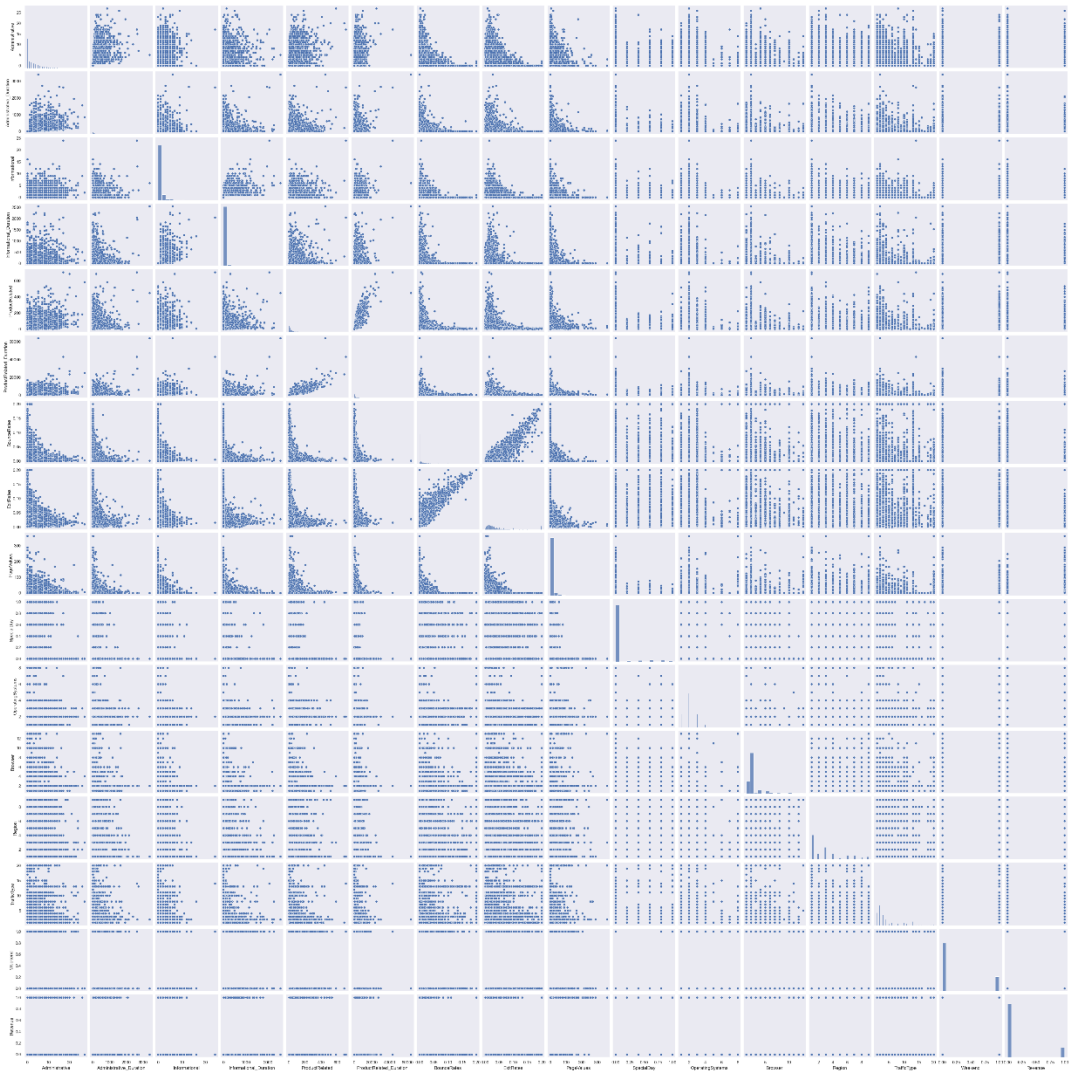


MULTIVARIATE ANALYSIS:

1. HEAT MAP



2. PAIR PLOT

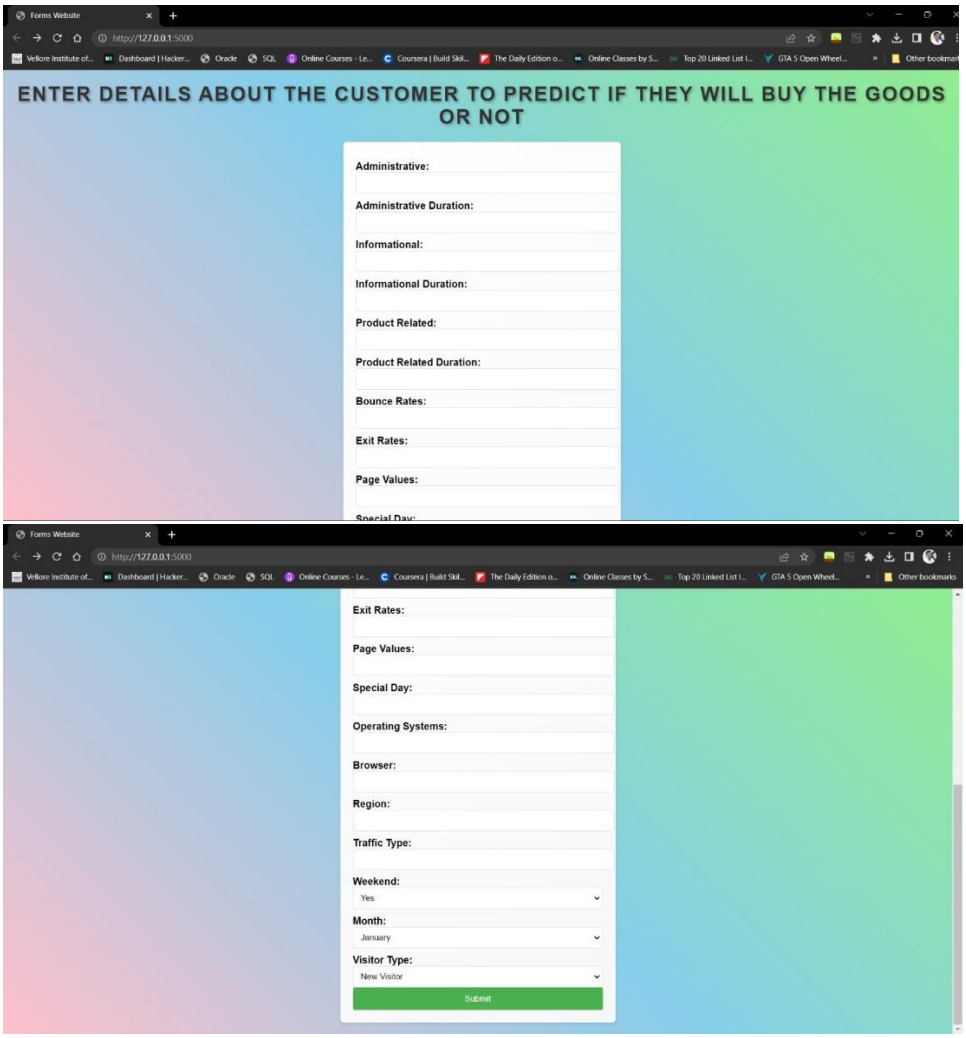


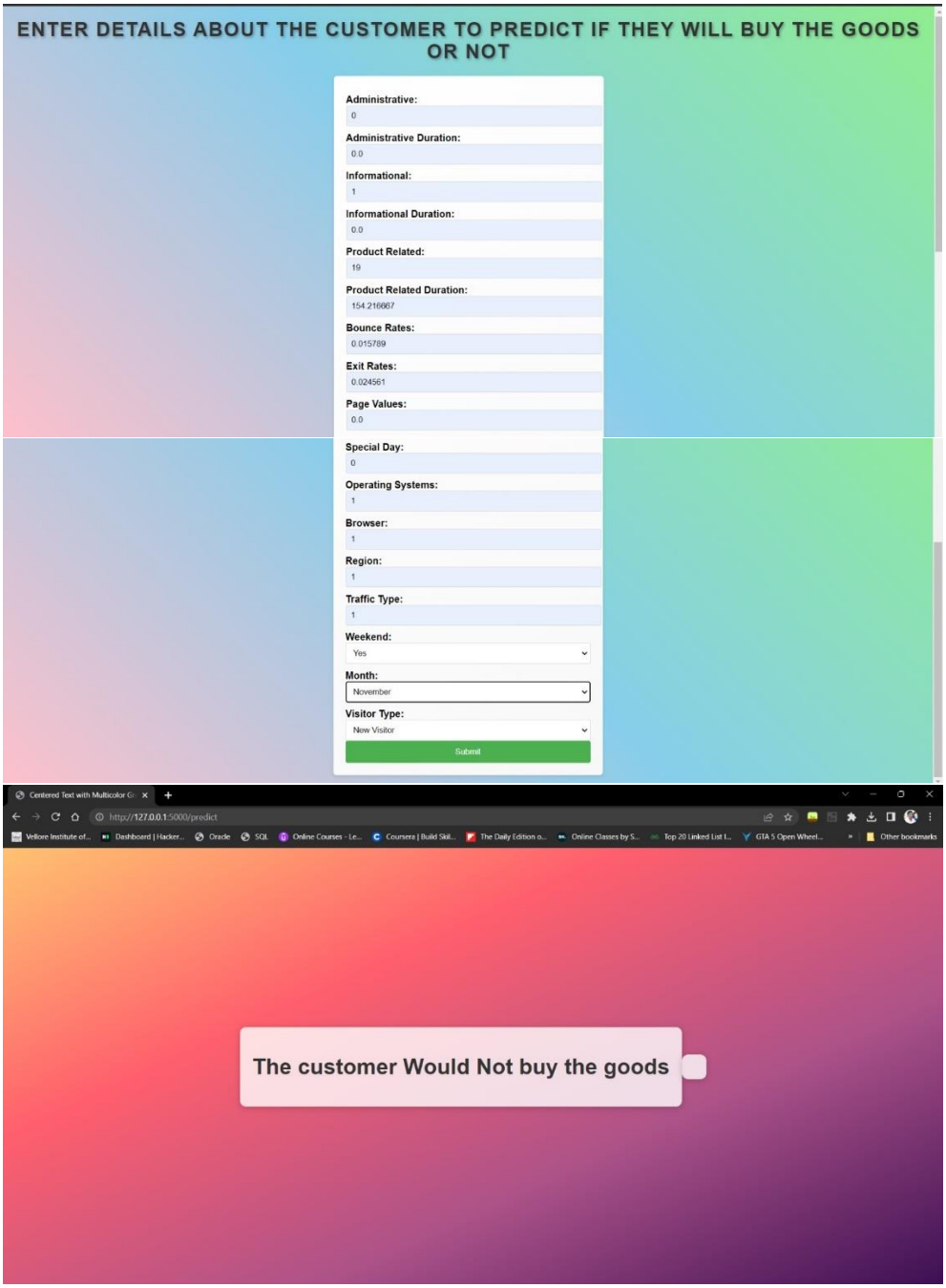
PREDICTED RESULT :

`	ACCURACY
1. Random Forest	86
2. KNN	81
3. SVM	84
4. Naïve Bayes	83
5. Logistic Regression	83

Based on the observed accuracy scores, the Random Forest classifier achieved the highest accuracy of 86% among the tested algorithms. Therefore, it is selected as the preferred model for predicting the outcome

FLASK CONNECTION AND WEB PAGE UI AND OUTCOME:





7. ADVANTAGES & DISADVANTAGES

Advantages of the proposed solution:

- Accurate prediction of customer behavior 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 behavior and preferences change over time.
- Ethical considerati ons regarding the collection and usage of customer data, ensuring privacy and compliance with data protection regulations.

8. APPLICATIONS

   The proposed solution can be applied in various areas, including:
   E-commerce platforms: Predicting customer behavior 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 behavior 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.

9. **CONCLUSION**

   In conclusion, this project aimed to develop a predictive model using classification algorithms to determine whether a customer is likely to make a purchase or engage in window shopping during online shopping activities. By analyzing relevant features and training the model, valuable insights can be obtained regarding customer behavior. The selected model can assist businesses in optimizing their marketing strategies, product offerings, and overall customer experience, leading to improved conversion rates and customer satisfaction,and to Predict the Revenue in Webpage by the users activity.

10. FUTURE SCOPE

The project's future scope includes the following enhancements:

Incorporating more advanced machine learning algorithms and techniques to improve prediction accuracy.

Integration with real-time data sources to enable up-to-date predictions and dynamic decision-making.

Utilizing natural language processing (NLP) techniques to analyze customer reviews, feedback, and sentiments for a comprehensive understanding of customer preferences.

Expanding the solution to consider multiple online shopping platforms and market segments for a broader application scope.

Conducting A/B testing and experimentation to validate the effectiveness of different marketing strategies based on the predicted customer behavior.

****