Customer Behaviour Analysis and Spending Prediction using Machine Learning

Narendra Bharambe[1], Tanuja Patil[2], Aishwarya Bavaskar[3], Minal Waykole[4], Dr. Dinesh. D. Patil[5]
[1,2,3,4]*Department of Computer Science and Engineering, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal – 425203, Dist. Jalgaon, Maharashtra, India*
[5]*Associate professor and Head Department of Computer Science and Engineering, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal – 425203, Dist. Jalgaon, Maharashtra, India*

Abstract—In today's digital business environment, organizations generate large amounts of customer shopping data through e-commerce platforms, retail stores, mobile applications, and online payment systems. Analysing customer purchasing behaviour manually has become difficult due to the continuously increasing volume of customer data. This paper presents a machine learning-based system for Customer Behaviour Analysis and Spending Prediction using data analytics techniques. The proposed system analyses customer shopping patterns and predicts customer spending amounts based on factors such as age, gender, product category, season, payment method, previous purchases, and purchase frequency. The project integrates data preprocessing, Exploratory Data Analysis (EDA), SQL-based analysis, Power BI dashboard visualization, machine learning model development, and Streamlit web application deployment into one complete workflow. The Random Forest Regressor algorithm is used for prediction because of its high accuracy and reduced overfitting capability. The machine learning model is evaluated using R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results demonstrate that the developed system generates accurate spending predictions and meaningful business insights. The Streamlit application provides real-time prediction functionality through an interactive user interface. The proposed system helps businesses improve customer understanding, targeted marketing, decision-making, and overall business profitability.

Index Terms—Customer Behaviour Analysis, Machine Learning, Random Forest Regressor, Spending Prediction, Data Analytics, Power BI, SQL, Streamlit.

I. INTRODUCTION

Customer behaviour analysis is one of the most important applications of data analytics and machine learning in modern business environments. Businesses generate a huge amount of customer shopping data every day through online shopping platforms, retail stores, banking systems, and digital payment applications. Understanding customer purchasing behaviour helps organizations improve marketing strategies, enhance customer satisfaction, increase sales, and make better business decisions. However, manual analysis of large customer datasets is time-consuming, difficult, and inefficient. Therefore, intelligent systems based on machine learning are required to automate customer analysis and generate accurate business insights.

The project titled "Customer Behaviour Analysis and Spending Prediction using Machine Learning" was developed to analyse customer shopping patterns and predict customer spending amounts using machine learning techniques. The system combines data analytics, business intelligence, visualization, and predictive modelling to create an intelligent customer analysis platform. The project helps businesses understand customer purchasing behaviour and identify spending trends effectively.

The dataset used in this project contains customer-related information such as age, gender, product category, purchase amount, season, payment method, previous purchases, and purchase frequency. Data preprocessing techniques including handling missing values, removing duplicate records, and encoding categorical variables were performed to improve

dataset quality. Exploratory Data Analysis (EDA) was carried out using visualization techniques such as histograms, bar charts, pie charts, and heatmaps to understand customer purchasing patterns and shopping trends.

The Random Forest Regressor algorithm was used for customer spending prediction because of its high accuracy, stability, and ability to handle large datasets efficiently. The machine learning model was trained and tested using customer shopping data to generate reliable spending predictions. SQL was used for customer data analysis and report generation, while Power BI dashboards were developed for interactive business visualization. A Streamlit web application was also created to provide real-time spending prediction through a user-friendly interface.

The main objective of this project is to analyse customer behaviour, predict customer spending patterns, and help businesses make intelligent decisions using machine learning and data analytics technologies.

II. LITERATURE REVIEW

Customer behaviour analysis and spending prediction have become important research areas in machine learning and data analytics. Many researchers have developed intelligent systems to analyse customer purchasing patterns and improve business decision-making processes.

Breiman [1] introduced the Random Forest algorithm, which is widely used for classification and regression tasks because of its high prediction accuracy, stability, and reduced overfitting. The study proved that Random Forest performs efficiently on large datasets and generates reliable prediction results.

Pedregosa et al. [2] presented Scikit-learn, a machine learning library in Python that provides efficient tools for predictive analytics and data mining. The library supports multiple machine learning algorithms and is widely used in customer behaviour analysis projects.

Sharma and Gupta [3] developed a customer spending prediction system using machine learning techniques. Their research demonstrated that predictive models can effectively identify customer spending patterns and improve marketing strategies. The study also highlighted the importance of data preprocessing and feature selection for improving prediction accuracy.

Kumar and Singh [4] focused on business intelligence and predictive analytics for customer behaviour analysis. Their work explained how data visualization and machine learning techniques help organizations understand customer preferences, generate business insights, and improve customer relationship management.

Hunter [5] introduced Matplotlib as a visualization tool for creating graphical representations of data. Visualization techniques such as bar charts, histograms, and heatmaps are useful for understanding customer purchasing behaviour and identifying shopping trends.

Waskom [6] discussed the importance of Seaborn for statistical data visualization. Seaborn provides advanced visualization methods that improve exploratory data analysis and help researchers identify meaningful patterns within datasets.

Based on the literature survey, it is observed that machine learning algorithms, business intelligence tools, and visualization techniques play a significant role in customer behaviour analysis and spending prediction systems. The existing research motivated the development of the proposed project using Random Forest Regressor, SQL, Power BI, and Streamlit technologies.

III. PROBLEM STATEMENT

In modern business environments, organizations generate a large amount of customer shopping and transaction data through retail stores, e-commerce platforms, and digital payment systems. Analyzing this data manually is difficult, time-consuming, and inefficient. Businesses often face challenges in understanding customer purchasing behaviour, identifying spending patterns, and predicting future customer spending accurately.

Traditional analysis methods are unable to process large datasets efficiently and fail to provide real-time business insights. Due to the lack of intelligent prediction systems, organizations struggle to improve customer targeting, marketing strategies, inventory management, and customer satisfaction.

Therefore, there is a need to develop an intelligent system capable of analyzing customer behaviour and predicting customer spending using machine learning techniques. The system should automatically process customer data, identify purchasing trends, generate

business insights, and provide accurate spending predictions to support effective business decision-making processes.

IV. PROPOSED SYSTEM

The proposed system, "Customer Behaviour Analysis and Spending Prediction using Machine Learning," is designed as an intelligent business analytics platform for analysing customer shopping behavior and predicting spending patterns.

The system consists of the following modules:

1. Customer Data Collection Module
2. Data Preprocessing Module
3. Exploratory Data Analysis (EDA) Module
4. SQL-Based Customer Analysis Module
5. Power BI Dashboard Module
6. Machine Learning Prediction Module
7. Random Forest Regressor Module
8. Streamlit-Based User Interface Module

The system collects customer shopping data, preprocesses it, and performs analysis using visualization and SQL techniques. The Random Forest Regressor algorithm is used to predict customer spending amounts accurately. Power BI dashboards provide interactive business insights, while the Streamlit web application allows users to enter customer details and receive real-time spending predictions through a user-friendly interface.

V. METHODOLOGY

The methodology of the proposed system involves multiple steps for analysing customer behavior and predicting spending patterns using machine learning techniques.

1. Data Collection:

Customer shopping behavior data was collected from the Customer Shopping Behavior Dataset containing attributes such as age, gender, product category, purchase amount, season, and payment method.

2. Data Preprocessing:

The collected dataset was cleaned by handling missing values, removing duplicate records, encoding categorical data, and selecting important features for model training.

3. Exploratory Data Analysis (EDA):

Visualization techniques such as bar charts, histograms, heatmaps, and pie charts were used to analyse customer purchasing behavior and spending trends.

4. SQL and Power BI Analysis:

SQL queries were used for customer data analysis and business insight generation, while Power BI dashboards were created for interactive visualization and reporting shown in figure 1.

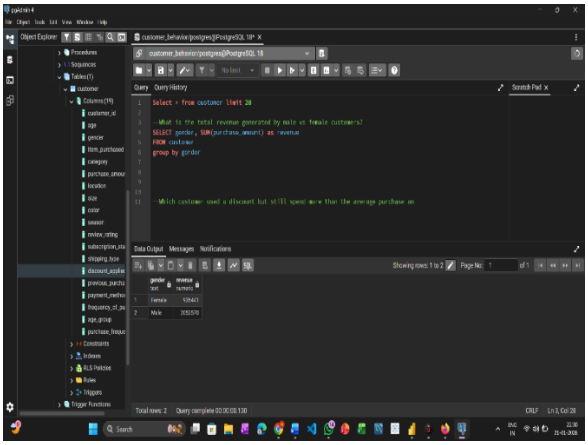

Figure 1 SQL quires

5. Machine Learning Model Development:

The Random Forest Regressor algorithm was used to train the prediction model using customer shopping behavior data. The dataset was divided into training and testing data for evaluation.

6. Model Evaluation:

The model performance was evaluated using metrics such as R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) figure 2.

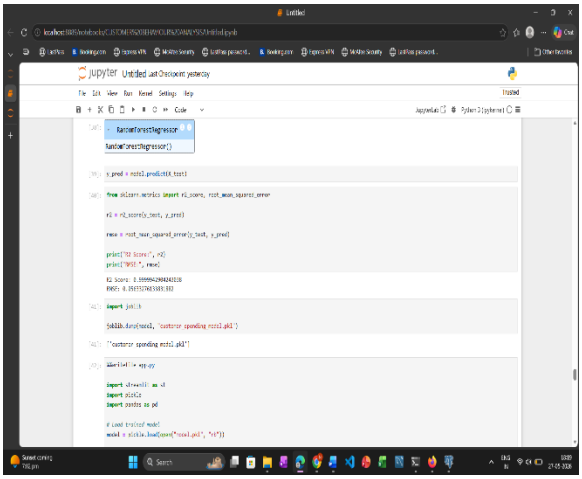

Figure 2 Train Model Accuracy

7. Streamlit Application Development:

A Streamlit web application was developed to provide real-time customer spending predictions through a user-friendly interface.

VI. RESULT AND DISCUSSION

The developed system successfully analyzed customer shopping behavior and generated accurate customer spending predictions using machine learning techniques. The Random Forest Regressor algorithm performed effectively and produced reliable prediction results with minimal prediction error.

During Exploratory Data Analysis (EDA), several important customer behavior patterns were identified. The analysis showed that middle-aged customers contributed more to overall spending, frequent customers generally spent higher amounts, and digital payment methods were widely preferred. Seasonal variations and product categories also significantly influenced customer purchasing behavior.

Figure 3 shows SQL analysis and Power BI dashboards provided meaningful business insights through interactive visualizations such as spending trends, category-wise sales analysis, payment method distribution, and gender-based customer analysis. These visualizations improved business understanding and decision-making processes.

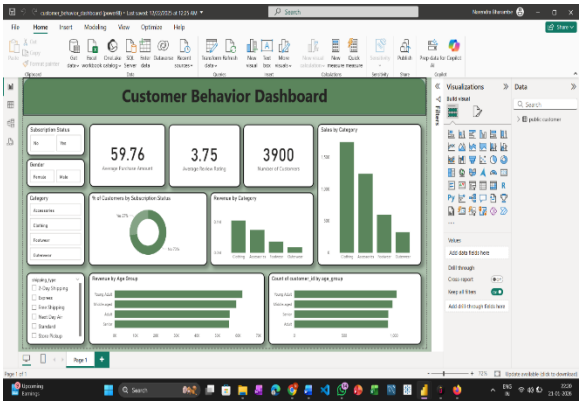

Figure 4 Customer Behaviour Analysis (Power BI dashboards)

The machine learning model was evaluated using performance metrics such as R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The evaluation results indicated that the developed model achieved good prediction accuracy and successfully identified relationships between customer features and spending amounts.

The Streamlit web application also worked successfully by allowing users to enter customer details and receive real-time spending predictions through a simple and user-friendly interface figure 4. Overall, the proposed system achieved its objectives effectively and demonstrated the practical implementation of machine learning in customer behavior analysis and business intelligence systems.

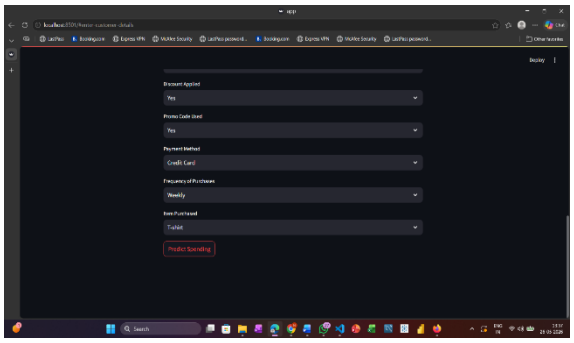

Figure 4 Real Time Prediction (Streamlit.)

VII. CONCLUSION

The "Customer Behaviour Analysis and Spending Prediction using Machine Learning" project was successfully developed and implemented to analyze customer shopping behavior and predict customer spending amounts using machine learning techniques. The system combined data analytics, SQL, Power BI, machine learning, and Streamlit technologies to generate meaningful business insights and real-time spending predictions.

The Random Forest Regressor algorithm successfully identified relationships between customer features and spending patterns, producing accurate and reliable prediction results. Exploratory Data Analysis (EDA) and Power BI dashboards helped in understanding customer purchasing trends, category preferences, seasonal effects, and payment method usage.

The Streamlit web application provided a simple and user-friendly interface for entering customer details and generating instant spending predictions. The developed system improved customer analysis, automated prediction processes, and business decision-making capabilities.

Overall, the project demonstrated how machine learning and business analytics can transform

customer shopping data into valuable business insights and support intelligent decision-making in modern business environments.

VIII. FUTURE SCOPE

The developed system can be further enhanced by integrating real-time customer data for faster and more accurate business analysis. Future versions of the project can use advanced machine learning and deep learning algorithms such as XGBoost, Artificial Neural Networks (ANN), and Deep Learning models to improve prediction accuracy and system performance.

The system can also be deployed on cloud platforms like AWS, Microsoft Azure, or Google Cloud for better scalability, security, and accessibility. In future, the application can be converted into a mobile-based system using Flutter or React Native to improve user accessibility.

Additional features such as personalized product recommendation systems, AI-based customer segmentation, and Big Data integration using Hadoop or Apache Spark can also be implemented. Power BI dashboards may be improved with live data visualization, dynamic filtering, and AI-generated insights for better business intelligence.

Overall, the project has strong future applications in e-commerce platforms, retail businesses, banking systems, marketing companies, and customer relationship management systems.

ACKNOWLEDGMENT

The authors thank Principal Dr. R. B. Barjibhe, Project Guide, Associate Professor & Head of Department Dr. Dinesh. D. Patil, of the Department of Computer Science and Engineering as well as valuable support from our department project coordinator Prof. A. P. Ingale, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, for their valuable guidance, support, and access to resources throughout this project.

REFERENCES

[1] M. P. Geetha and D. Karthika Renuka, "Deep Learning Architecture Towards Consumer Buying Behaviour Prediction Using Multitask Learning Paradigm," Journal of Intelligent & Fuzzy Systems, Vol. 46, No. 1, Pp. 1341–1357, 2024.

[2] E. Deniz and S. Çökekoğlu Bülbül, "Predicting Customer Purchase Behavior Using Machine Learning Models," Information Technology In Economics and Business, Vol. 1, No. 1, Pp. 1–10, 2024.

[3] N. Romanenko, K. Sharma, And S. Verma, "Prediction of Financial Customer Buying Behavior Based on Machine Learning," Journal of Artificial Intelligence General Science, Vol. 5, No. 1, Pp. 125–131, 2024.

[4] M. Madanchian, "Generative Ai for Consumer Behavior Prediction: Techniques and Applications," Sustainability, Vol. 16, No. 22, Pp. 9963–9975, 2024.

[5] J. Lin, "Application of Machine Learning In Predicting Consumer Behavior and Precision Marketing," Plos One, Vol. 20, No. 5, Pp. 1–15, 2025.

[6] D. T. Braithwaite, M. Cavalcanti, R. A. Mcever, Et Al., "Your Spending Needs Attention: Modeling Financial Habits with Transformers," Arxiv Preprint Arxiv:2507.23267, Pp. 1–18, 2025.