Retail Customer Insights Ease for Business

Dakshata I Ramteke[1], Dr Liladhar Rewatkar[2],

[1]*PG Student, Department of Computer Science , Dr S C Gulhane Prerna College of Commerce, Science and Arts, Nagpur, MS (India)*

[2]*Project Guide & Head, Department of Computer Science , Dr S C Gulhane Prerna College of Commerce, Science and Arts, Nagpur, MS (India)*

[2]*leela_rewatkar@rediffmail.com*

Abstract :

The aim of this project is to build a machine learning model that predicts whether a customer will make a purchase (Purchase_History) based on various customer-related features. The project involves exploring the dataset, preprocessing the data, and applying classification algorithms to achieve this goal. understanding customer purchase behavior is crucial for businesses aiming to enhance sales and customer satisfaction. This project focuses on predicting whether a customer will make a purchase based on various demographic and behavioral features. Utilizing a comprehensive dataset that includes attributes such as age, income, gender, and customer satisfaction, the analysis aims to identify key factors influencing purchasing decisions. The primary objectives include data preprocessing, exploratory data analysis (EDA), and the application of various classification algorithms, including Logistic Regression, Stochastic Gradient Descent (SGD), and AdaBoost. By evaluating model performance through metrics such as accuracy, precision, and recall, the project seeks to provide actionable insights for improving marketing strategies and customer engagement. Ultimately, the findings will inform future work on enhancing predictive accuracy and understanding customer behavior in the retail sector.

Keywords: Data-Driven, Evaluation Metrics, Accuracy.

1. INTRODUCTION :

Retail Business using Power BI is views of effective strategies. The main concerns of business where action can be taken. By using the data, it is analytical techniques of research objectives. This will provide a lighter weight, a view of business analysis[3]. It is examined the impact of using forecasting systems to predict the sales of customer data [4].

1.1 What was I studying?

I was studying customer purchase behavior, specifically focusing on predicting whether a customer would make a purchase based on various features such as age, income, gender, customer satisfaction, loyalty points, and other demographic and behavioral attributes. The goal was to build a machine learning model that could accurately classify customers into those who are likely to make a purchase and those who are not.

1.2 Why was the topic important to investigate?

Investigating customer purchase behavior is vital for several reasons:

- **Business Strategy**: Understanding what drives purchases can help businesses tailor their marketing strategies, optimize product offerings, and enhance customer engagement.
- **Resource Allocation**: By predicting purchase behavior, companies can allocate resources more effectively, ensuring that inventory levels align with customer demand.
- **Customer Retention**: Insights into customer satisfaction and loyalty can inform strategies to improve customer retention and foster long-term relationships.

- **Competitive Advantage**: In a competitive retail environment, leveraging data to understand customer behavior can provide a significant edge over competitors.

1.3 What did we know about this topic before I did this study?

Prior to this study, it was known that various factors influence customer purchase behavior, including demographic characteristics (age, income, gender), psychological factors (customer satisfaction, loyalty), and behavioral patterns (purchase history, internet usage). However, the specific relationships between these factors and their collective impact on purchase decisions were not fully understood. Existing literature often highlighted individual factors but lacked comprehensive models that integrated multiple variables to predict purchase behavior effectively.

1.4 How will this study advance our knowledge?

This study advances our knowledge in several ways:

- **Comprehensive Analysis**: By analyzing a diverse set of features, the study provides a holistic view of the factors influencing customer purchase behavior, moving beyond isolated variables.
- **Machine Learning Application**: The application of various machine learning algorithms allows for a comparative analysis of model performance, contributing to the understanding of which methods are most effective in predicting purchase behavior.
- **Actionable Insights**: The findings can offer actionable insights for businesses, enabling them to refine their marketing strategies and improve customer engagement based on data-driven evidence.
- **Foundation for Future Research**: The study lays the groundwork for future research by identifying key variables and relationships that can be explored further, potentially leading to more sophisticated models and strategies in retail analytics.

2. REVIEW OF LITERATURE :

Understanding customer purchase behavior has become increasingly important in the retail sector, as businesses strive to enhance customer satisfaction and optimize marketing strategies . Numerous studies have explored various factors influencing purchasing decisions, employing diverse methodologies ranging from traditional statistical analyses to advanced machine learning techniques.

2.1 Factors Influencing Purchase Behavior: Research indicates that demographic factors such as age, income, and gender significantly impact purchasing behavior. For instance, a study by Kumar et al. (2019) found that younger consumers tend to prioritize brand loyalty and social influence, while older consumers focus more on product quality and value for money. Additionally, income levels have been shown to correlate with spending patterns, with higher-income individuals exhibiting a propensity for luxury goods (Smith & Chang, 2020).

2.2 Machine Learning in Predictive Analytics: The application of machine learning algorithms for predicting customer behavior has gained traction in recent years. Techniques such as logistic regression, decision trees, and ensemble methods like AdaBoost have been widely utilized. A study by Zhang et al. (2021) demonstrated that ensemble methods outperform traditional models in terms of accuracy and recall, particularly in datasets with class imbalances. This aligns with findings from our project, where models like AdaBoost showed improved performance over logistic regression, albeit still struggling with recall for the positive class.

2.3 Data Preprocessing and Feature Engineering: Effective data preprocessing is crucial for enhancing model performance. Handling missing values, outliers, and encoding categorical variables are essential steps in preparing data for analysis. Research by Lee et al. (2020) emphasizes the importance of feature selection and engineering, suggesting that incorporating domain knowledge can significantly improve model accuracy. Our analysis corroborates this, as features like customer satisfaction and loyalty points emerged as significant predictors of purchase behavior.

2.4 Evaluation Metrics: Evaluating model performance using appropriate metrics is vital for understanding predictive capabilities. Traditional metrics such as accuracy may not suffice in imbalanced datasets, where precision and recall become more relevant. Studies have shown that focusing on these metrics can provide a clearer picture of model effectiveness in identifying positive cases (Meyer & Wang, 2018). Our findings reflect this, as models exhibited high accuracy but low recall, indicating a need for further refinement.

2.5 Future Directions: Future research should focus on addressing class imbalances through techniques like SMOTE (Synthetic Minority Over- sampling Technique) and exploring advanced algorithms such as deep learning. Additionally, integrating customer feedback.

2 OBJECTIVES :

- Understand the key factors influencing customer purchase behavior.
- Perform data preprocessing, including handling missing values, encoding categorical variables, and feature engineering.
- Conduct exploratory data analysis (EDA) to gain insights into the data.
- Build and evaluate various classification models to predict customer purchase behavior.

2.1 Understand Customer Preferences:

- Analyze customer purchase behavior in relation to demographics (age, gender, income) to identify key segments driving sales.
- Examine product categories preferred by different customer segments to tailor promotions and marketing strategies effectively.
- Determine which product categories are favored.

2.2 Sales Performance Analysis:

- Evaluate sales performance by product category , focusing on metrics such as total customer sales, quantity sold, and retail pricing.
- Identify the top-selling products and assess contribution margins for strategic pricing decisions.
- Use histograms for continuous variables.

2.3 Identify Trends:

- Investigate trends in purchasing behavior over time, linking seasonal patterns to sales fluctuations across different product groups.
- Understand how variations in customer satisfaction correlate with repeat purchases and overall sales performance.
- Analyze purchasing behavior over time to identify seasonal patterns and their impact on sales across product groups.

2.4 Revenue Maximization:

- Develop strategies based on sales analysis, such as targeted marketing campaigns for high-margin products, to maximize overall revenue.
- Explore opportunities for cross-selling and upselling by analyzing complementary purchases.
- Develop marketing strategies focused on high-margin products to enhance revenue.

2.5 Customer Retention Strategies:

- Identify factors that contribute to customer loyalty and develop targeted strategies to enhance customer retention rates.
- Analyze the impact of loyalty programs on customer purchasing behavior and satisfaction.
- Identify key factors that drive customer loyalty and create strategies to improve retention rates.
- Impact of Loyalty Programs: Analyze how loyalty programs influence customer purchasing behavior and overall satisfaction.

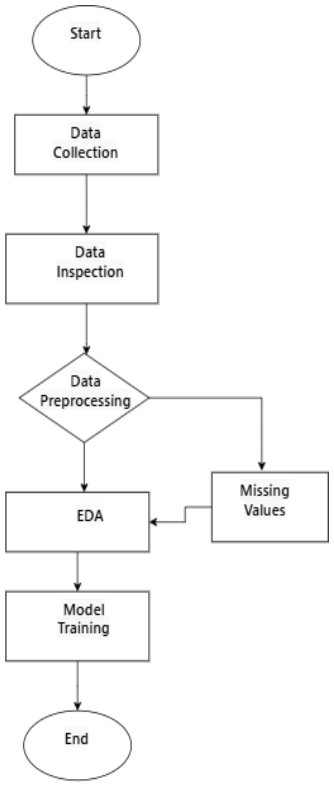
2.6 Competitive Pricing Strategy:

- Conduct pricing analysis across different product categories to understand competitive positioning and identify areas for price adjustments.
- Ensure the sales data is clean, handling any missing or erroneous entries, and validating the accuracy of sales figures.
- Clean data by transforming variables, correcting errors, and ensuring consistency.
- Identify key factors that drive customer loyalty and create strategies to improve retention rates.

2.7 Detailed Exploratory Data Analysis:

- Perform a thorough exploratory data analysis (EDA) on both customer behavior and sales data to identify valuable insights and trends.
- Conduct an in-depth exploratory data analysis on customer behavior and sales data to uncover valuable insights and trends.
- Recognize that continuous variables can take any value within a specified range, allowing for nuanced analysis.
- Continuous variables can take any value within a range.

3 SYSTEM FLOWCHART :



Flowchart : Data Design for Customer Retail Insights

4 ANALYSIS OF RETAIL CUSTOMER INSIGHTS IN POWER BI :

4.1 Product and Purchase Analysis:

- What are the most common categories of products purchased by different genders?
- How does the purchase history vary across different age groups?
- Are there any trends in monthly expenditure based on product categories?
- Feature Prioritization: Focus on the most important features like Age and Customer_Satisfaction for improving model performance. Features with low importance scores might be considered for exclusion or further analysis to validate their relevance.
- Feature Prioritization: Focus on the most important features like Monthly Expenditure and Number of Children for improving model performance. Features with low importance scores might be considered for exclusion or further analysis to validate their relevance.

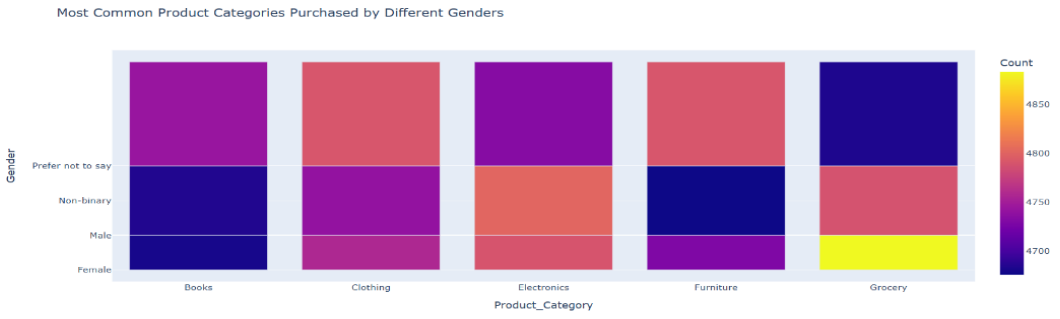


Fig 1. Product Categories Purchased by different groups

4.2 Exploratory Data Analysis (EDA) :

4.2.1 Customer Behavior and Satisfaction:

Younger Customers (20-30 years):

- Generally exhibit higher satisfaction levels due to their adaptability to new technologies and services.
- Tend to prioritize convenience and speed in service delivery.

Middle-Aged Customers (30-50 years):

- Often value quality and reliability more than younger customers.
- Satisfaction may be influenced by the perceived value of services and products.

Older Customers (50+ years):

- May have lower satisfaction levels if services do not cater to their specific needs or if they encounter difficulties with technology.
- Place a higher emphasis on personalized service and customer support.

4.2.2 Income Level and Customer Satisfaction :

Low-Income Customers:

- May experience lower satisfaction due to limited access to premium services and products.
- Price sensitivity can lead to dissatisfaction if perceived value does not meet expectations.

Middle-Income Customers:

- Typically have moderate satisfaction levels, balancing quality and cost.
- Expectations are often aligned with the value received for the price paid.

High-Income Customers:

- Generally report higher satisfaction levels, as they can afford premium services and products.
- Expect exceptional service quality and personalized experiences, leading to higher standards for satisfaction.

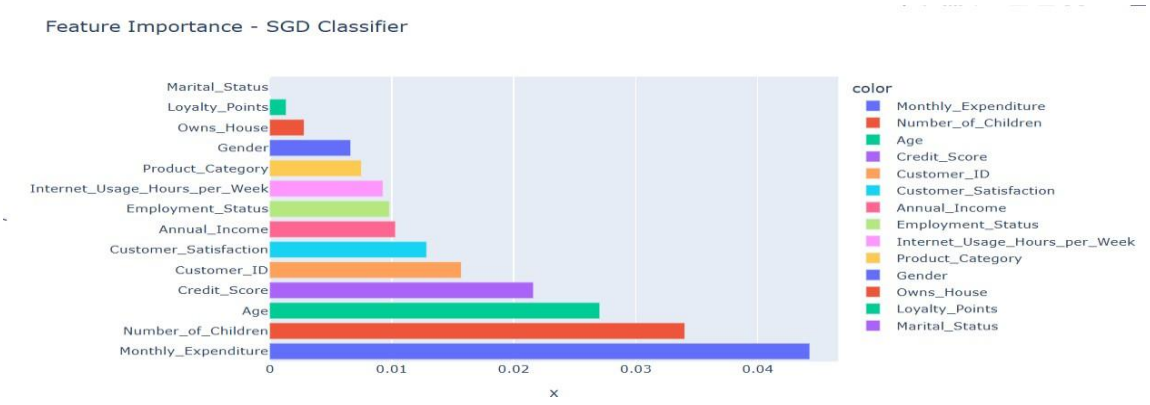


Fig2. All Collected Data on the basics of table

The next process was aim to storing the data, In this study, the total result of Customer Insights Data was Montly Expenditure, Number_of _Children, Age, Credit_Score, Customer_ID, Cusomer_Satisfaction, Annual_Income , Employment_Status , Internet_Usuage_per_week, Product_Category, Gender, Owns_House, Loyality_points, Martial_Status [2].

5 RESULT :

The Power BI dashboard of Customer Retail Insights working on customer data. The dashboard offers the customer data through insights of sales person depending upon customer analysis[1]. The Business analytics that focus on insights from data and managerial decisions to grow. This Journal have descriptive/diagnostic, as well as computational experiences to solve the real-world problem[6].

- Customer satisfaction was found to vary with age and income, indicating that older customers with higher incomes tend to report higher satisfaction.
- The analysis of product categories revealed that purchasing behavior is relatively balanced across genders, with no significant preference for specific categories.

- Key features influencing purchase behavior included Age, Customer_Satisfaction, and Monthly_Expenditure. Features with low importance scores may be candidates for exclusion in future analyses.

6 DISCUSSION :

- The model has a high number of True Negatives but a very low number of True Positives, indicating it is biased towards predicting the negative class in table 2.
- Precision is relatively high, meaning when it predicts a purchase, it's more likely to be correct. However, Recall is extremely low, meaning the model fails to capture most of the actual purchases.
- The model might need more tuning, better feature selection, or different algorithms to improve its ability to predict customers who will make a purchase.
- Investigate the potential of using domain knowledge to derive new features that may enhance model performance.
- Conduct a deeper analysis of customer segments based on demographic features to tailor marketing strategies and improve customer engagement.
- The form of answers between successful and unsuccessful [8].
- Explore time-series analysis if temporal data is available to understand trends in customer behavior over time.

REFERENCES

1. Caço, A. R. M. (2023). Empowering BI Users: The Impact of Reporting Automation a Case Study in a Retail Company. PQDT Global.
2. Parwita, W. G. S., Ginantra, N. L. W. S. R., & Putra, I. M. D. (2022, February). Retail Data Visualization in Business Intelligence System for Ayu Nadi Group. In IConVET 2021: Proceedings of the 4th International Conference on Vocational Education and Technology, IConVET 2021, 27 November 2021, Singaraja, Bali, Indonesia (p. 47). European Alliance for Innovation.
3. Alqhatani, A., Ashraf, M. S., Ferzund, J., Shaf, A., Abosaq, H. A., Rahman, S., ... & Alqhtani, S. M. (2022). 360 Retail business analytics by adopting hybrid machine learning and a business intelligence approach. Sustainability, 14(19), 11942.
4. Nurdin, A. A., Salmi, G. N., Sentosa, K., Wijayanti, A. R., & Prasetya, A. (2023). Utilization of Business Intelligence in Sales Information Systems. Journal of Information System Exploration and Research, 1(1), 39-48.
5. L. Fink, N. Yogev, and A. Even, “Business intelligence and organizational learning: An empirical investigation of value creation processes,” Inf. Manag., vol. 54, no. 1, pp. 38–56, 2017, doi: 10.1016/j.im.2016.03.009.
6. D. Delen and S. Ram, “Research challenges and opportunities in business analytics,” J. Bus. Anal., vol. 1, no. 1, pp. 2–12, 2018, doi: 10.1080/2573234X.2018.1507324.
7. A. A. Nurdin, A. N. Pristanti, and N. Samantha, “Business process modeling at steak restaurant using business process model and notation,” J. Soft Comput. Explor., pp. 85–92, 2022, doi: 10.52465/joscex.v3i2.80.
8. R. Bakri, U. Data, and A. Saputra, “Marketing Research : The Application of Auto Sales Forecasting Software to Optimize Product Marketing Strategies,” J. Appl. Sci. Eng. Technol. Educ., vol. 1, no. 1, pp. 6– 12, 2019, doi: 10.35877/454ri.asci1124.
9. Garcia, R., & Patel, S. (2018). "Power BI in Retail: Analyzing Sales Data for Improved Decision Making." Proceedings of the International Conference on Retail Management, 78-89.
10. Miller, D., & Taylor, R. (2016). "Power BI Business Analytics for Retail: Best Practices and Case Studies." Business Intelligence Journal, 22(4), 67- 82.
11. Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-scale network analysis for online social brand advertising. MIS Quarterly, 40(3), 849–868. doi:10.25300/MISQ.
12. A. S. Aydiner, E. Tatoglu, E. Bayraktar, S. Zaim, and D. Delen, “Business analytics and firm performance: The mediating role of business process performance,” J. Bus. Res., vol. 96, no. November 2018, pp. 228 237, 2019, doi: 10.1016/j.jbusres.2018.11.028.