2024 International Conference on ABC in Computing Technologies and Information Sciences (ICTIS)

# Consumer Behaviour Prediction

|  |  |  |
| --- | --- | --- |
| ABC KHAN  *Computer Engineering*  *University of Engineering and Technology*  ABC, ABC  abc@abc.edu | ABC KHAN  *Computer Engineering*  *University of Engineering and*  *Technology*  ABC, ABC  abc@abc.edu | ABC KHAN  *Computer Engineering*  *University of Engineering and*  *Technology*  ABC, ABC  abc@abc.edu |

***Abstract*— This paper introduces a predictive modeling framework to forecast consumer behavior using demographic and socioeconomic factors, with a focus on Naive Bayes classification. Leveraging a dataset of customer profiles sourced from Kaggle, the study examines correlations between age, salary, and purchasing decisions. Through preprocessing and exploratory data analysis, insights are gained into patterns within the data. Utilizing Multinomial Naive Bayes, the study achieves good accuracy in predicting purchasing decisions. However, challenges in accurately predicting minority classes are noted, suggesting avenues for future research. Visualization techniques are employed to illustrate model performance, concluding with recommendations for enhancing predictive accuracy in real-world consumer behavior analysis.**

**Keywords**: ***Predictive modeling, Consumer behavior, Naive Bayes classification, Data analysis.***

I. INTRODUCTION

Consumer behavior prediction holds significant importance in contemporary business strategies, enabling organizations to anticipate market trends, tailor marketing initiatives, and optimize resource allocation. Leveraging demographic and socioeconomic factors, businesses can gain valuable insights into consumer preferences and purchasing decisions. In this study, we introduce a predictive modeling framework that focuses on Naive Bayes classification to forecast consumer behavior. The research study explores the correlations between demographic variables, such as age and salary, and purchasing decisions. We conducted preprocessing and exploratory data analysis to uncover underlying patterns within the data. By combining data-driven methodologies with advanced classification techniques, we aim to provide actionable insights for businesses seeking to understand and anticipate consumer behavior.

*A. Research Aim*

The research aims to develop a predictive modeling framework using demographic and socioeconomic factors to forecast consumer behavior.

*B. Research Objectives*

1. To develop a predictive modeling framework using Naive Bayes classification for forecasting consumer behavior.
2. To explore the correlations between demographic variables (such as age and salary) and purchasing decisions.
3. To conduct preprocessing and exploratory data analysis to uncover underlying patterns within the dataset.
4. To apply data-driven methodologies and advanced classification techniques to provide actionable insights for businesses.

*C. Research Questions*

1. How effective is Naive Bayes classification in predicting consumer behavior based on demographic and socioeconomic factors?
2. What are the correlations between age, salary, and purchasing decisions among consumers?
3. What underlying patterns within the dataset can be uncovered through preprocessing and exploratory data analysis?
4. How can data-driven methodologies and advanced classification techniques be utilized to provide actionable insights for businesses seeking to understand and anticipate consumer behavior?

II. LITERATURE REVIEW

Consumer behavior prediction has been a topic of significant interest in both academia and industry due to its crucial role in informing business strategies and decision-making processes. Consumer behavior prediction is essential for businesses to understand market structures and purchasing patterns. Big data analytics offers a powerful means to analyze complex data and derive key insights. Liu, H. (2023) [1] compares different analysis approaches and scenarios for understanding consumer behavior. They explore different factors that impact consumer behavior, like psychological, social, cultural, personal, and economic etc. Machine learning and big data analysis are discussed as effective tools for predicting consumer behavior, offering insights into market trends, optimization, and decision-making. Traditional analysis methods are compared with big data analysis, highlighting the scalability and objectivity of big data, despite challenges such as data manipulation. The paper concludes that big data analysis holds promise for future applications in consumer behavior prediction, offering valuable insights for businesses of all sizes.

Alizadeh [2] examines the role of artificial intelligence (AI) in analyzing the behavior of consumers within the marketing domain. They highlighted the significance of AI technologies in analyzing large datasets related to consumer performance, including behaviors, surveys, and purchase history. The paper discusses how machine learning can forecast the behavior of customers, aiding in marketing decisions and advertising strategies. It suggests that AI can optimize marketing efforts by creating targeted advertisements and personalized experiences for customers, ultimately increasing customer engagement and improving marketing outcomes. Furthermore, the paper explores predictive modeling techniques facilitated by AI, emphasizing their benefits in inventory management, enhancing customer experience, and increasing sales and revenue. The research aims to evaluate the effectiveness of AI technologies in predicting consumer behavior in the online retail industry, providing insights into their impact on marketing strategies and outcomes.

Valecha, H., Varma [3] explores the use of machine learning, specifically the Random Forest algorithm, to predict consumer behavior in the modern era of technology. It emphasizes the importance of anticipating market trends to understand consumer behavior in a competitive environment where trends are volatile. The research aims to establish a model that predicts consumer behavior based on various parameters such as environmental, organizational, individual, and interpersonal factors. By leveraging unique feature engineering and employing a time-evolving Random Forest classifier, the study seeks to provide the best predictions of consumer behavior regarding purchasing decisions. The results indicate that the Random Forest classifier outperforms other machine learning algorithms for the prediction of accuracy.

The paper by Mathur, N., Kumar, S [4] explores the use of machine learning and data mining techniques to build models of consumer behavior, emphasizing the need for accurate methods and strategies. It highlights the challenges in forecasting customer behavior and the importance of tailored models to address specific questions over time. Despite the complexity, most consumer behavior models are simplified, often overlooking relevant factors, which can lead to less reliable forecasts. The study reviews existing literature on analyzing consumer behavior using machine learning and data mining approaches, with a focus on implementation in Python due to its ease of use and the importance of accuracy, error rate, and precision.

Quynh, T. D., & Dung, H. T. T [5] delve into the realm of customer behavior analysis, particularly focusing on predicting customer decisions regarding the acceptance of restaurant coupons recommended by in-vehicle systems. The study utilizes machine learning models and data preprocessing techniques to address the challenges posed by categorical attributes and missing values in the dataset. By employing various classification models and handling missing data effectively, the proposed approach outperforms previous methods. The paper also sheds light on the impact of variables on customer decisions, offering valuable insights for marketers.

## III. METHODOLOGY

*A. Data Preprocessing*

Data preprocessing began with a thorough examination of the dataset to ensure data cleanliness, revealing no missing values. The 'User ID' column, deemed irrelevant for predictive modeling, was consequently dropped. Following this, the categorical variable 'Gender' underwent encoding into numerical values ('Male' represented as 0 and 'Female' as 1) to facilitate analysis. Recognizing the substantial variance in feature ranges, particularly evident in 'Age' and 'EstimatedSalary', feature scaling was executed using MinMaxScaler to standardize the data appropriately.

*B. Exploratory Data Analysis (EDA)*

Descriptive statistics played a pivotal role in the analysis, with mean, median, minimum, and maximum values calculated for continuous variables such as 'Age' and 'EstimatedSalary'. These statistics offered valuable insights into how the data was centered and dispersed, helping us grasp its distribution and unique characteristics. Complementing the numerical summaries, visualization techniques were employed to explore the dataset further. Histograms, box plots, count plots, and kernel density estimation (KDE) plots were generated to visually represent the distribution of features and explore potential relationships among them.

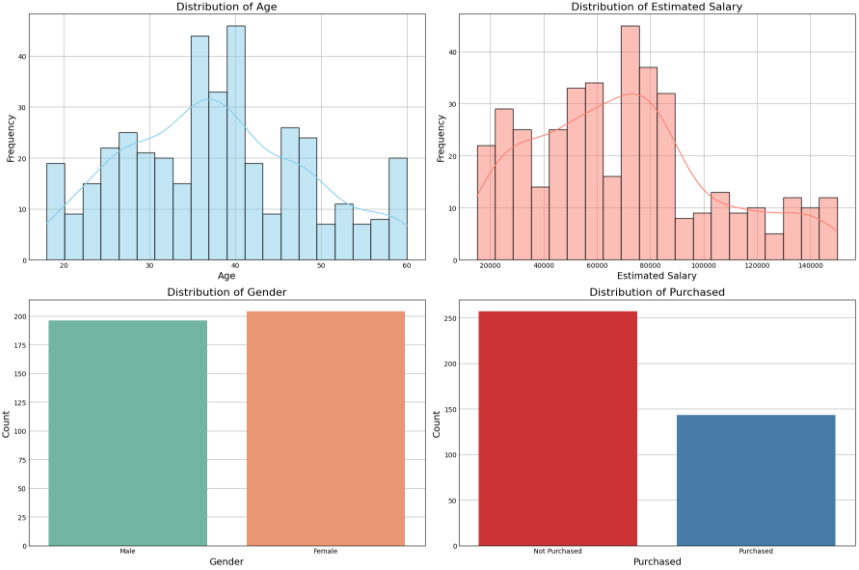


Figure 1 Distribution of features

By visualizing the feature distribution we got to know about the presence and spread of the feature. These patterns help us to understand the extent to which each feature varies.

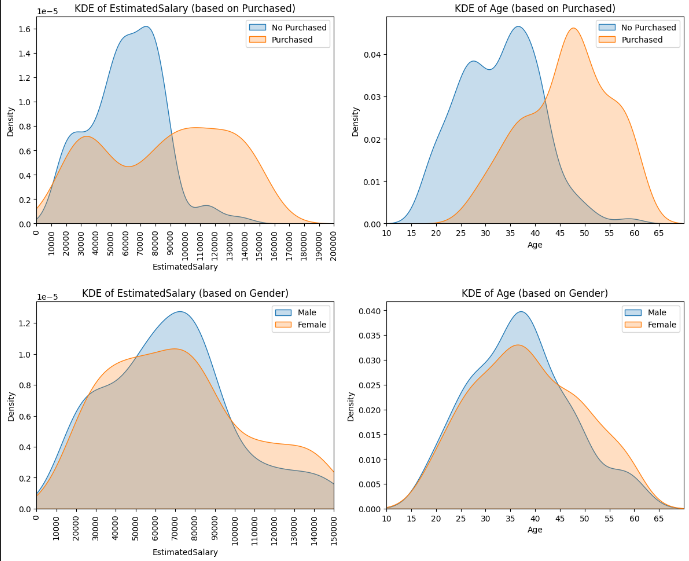


Figure 2 Visualizing Estimated Salary and Age by Purchase & Gender

Through our analysis, we got to know about the pattern: which different features show like we got to know individuals with incomes ranging from 40000, they aren’t making purchases. Moreover, those who do decide to buy something are older than those people who aren’t deciding about purchasing. Also, we discover, individuals from age 43 and above show a higher inclination towards purchasing products. And we observed that there's little difference between males and females in their purchasing decisions.

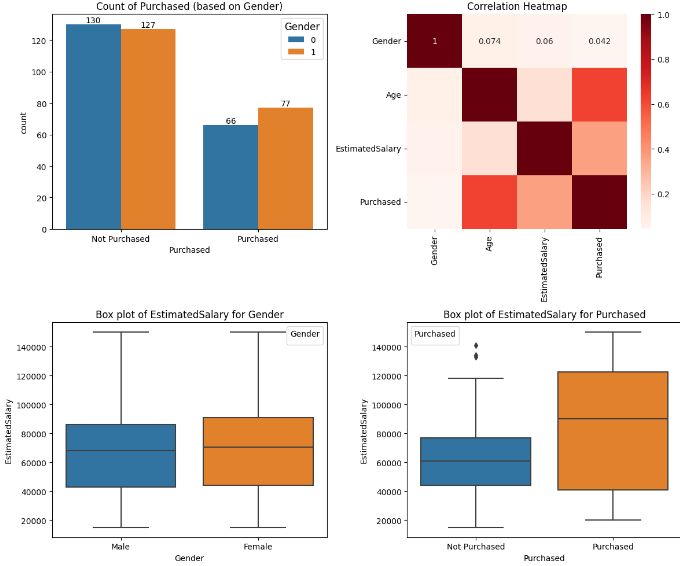


Figure 3 Exploring Purchasing and Gender: Count, Correlation, Salary

This analysis tells us all about purchasing we got to know that among the people who decide to purchase a product, there are more females than males, but among the people who do not decide to purchase a product, there are more males, we looked at the correlation and find out that the highest correlation is between Purchased and Age. we also find out the average estimated salary of people who decide to purchase a product is higher than people who do not decide to purchase a product, our analysis also depict that average estimated salary of males and females does not differ much.

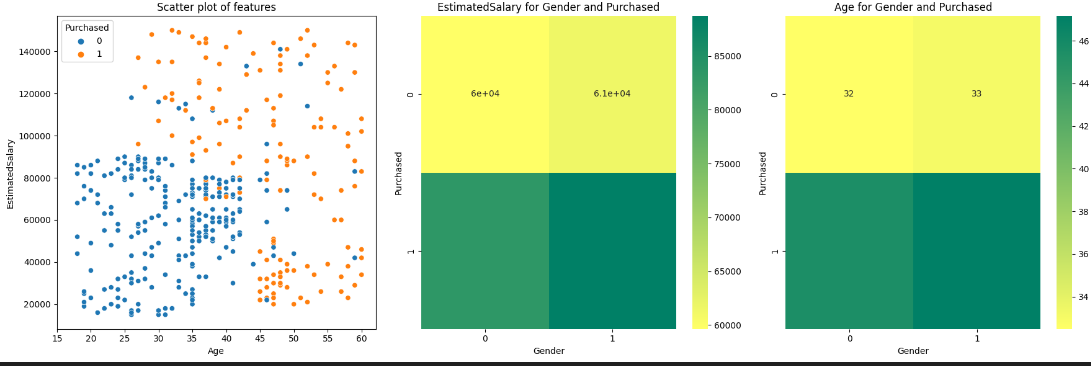


Figure 4 Relationship Exploration

We got to know about different ages people, their salaries, and their purchasing power we found out that people who are younger and have a low salary, mostly do not decide to purchase a product, and people with a salary of more than 100000, regardless of their Age, mostly decide to purchase a product, and people over the age of 45, regardless of their salary, are more likely to purchase a product.

*C. Model Development*

For feature selection, all variables except the target variable 'Purchased' were considered for model development. The Multinomial Naive Bayes classification algorithm was selected due to its aptness for handling both discrete and continuous data. Subsequently, the dataset was split into training and testing sets in an 80-20 ratio, and the Multinomial Naive Bayes model was trained on the training data. Evaluation of the model was conducted on the testing set, utilizing diverse metrics such as accuracy, recall, precision, and F1-score. Additionally, confusion matrices and classification reports were generated to comprehensively assess the performance of the model.

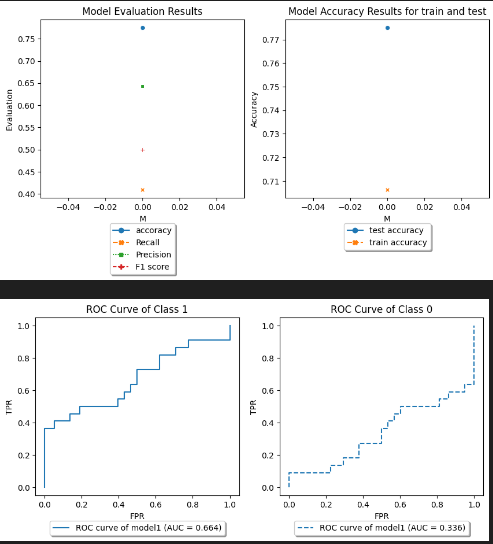


Figure 5 Model Evaluation and Performance Metrics

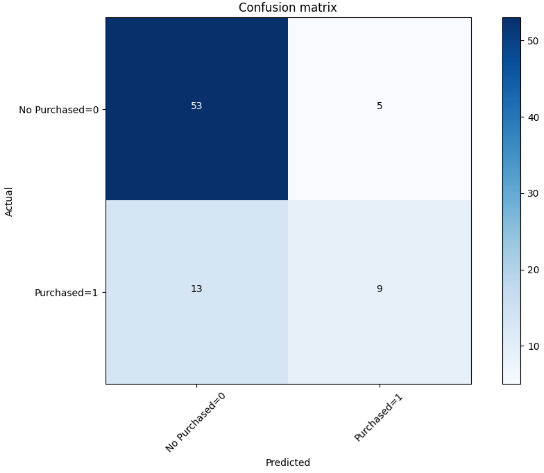


Figure 6 Confusion Matrix

*D. Model Evaluation and Interpretation:*

In terms of model performance, it demonstrated commendable accuracy in predicting purchasing decisions. Nevertheless, challenges arose in accurately predicting minority classes, suggesting potential areas for improvement. The model's interpretation highlighted age and estimated salary as pivotal factors influencing purchasing decisions. To gain insights, various visualizations like scatter plots and 3D scatter plots were utilized to interpret model predictions and determine feature importance, all this helped us to see the patterns in the data and figure out which factors are most important.

## IV. CONCLUSION

In this study, we introduced a predictive modeling framework leveraging demographic and socioeconomic factors to forecast consumer behavior, with a focus on Naive Bayes classification. Through preprocessing and exploratory data analysis, we gained insights into the correlations between age, salary, and purchasing decisions. Our findings indicate that while the Multinomial Naive Bayes model achieved good accuracy in predicting purchasing decisions, challenges were encountered in accurately predicting minority classes. This suggests avenues for future research to enhance predictive accuracy in real-world consumer behavior analysis. Visualization techniques, including scatter plots and 3D scatter plots, were instrumental in interpreting model predictions and determining feature importance. Our study underscores the significance of age and estimated salary as pivotal factors influencing purchasing decisions. Overall, this research contributes to the growing body of knowledge on consumer behavior prediction and provides valuable insights for businesses seeking to understand and anticipate market trends.

REFERENCES

[1]. Liu, H. (2023). Consumer Behavior Prediction in the Big Data Era: a Comparison Analysis. BCP Business & Management EMFRM, 38.

[2]. Alizadeh, H., Kashani, H. N., Filshoor, M. J., & Khameneh, A. P. (2023, November). Evaluation of consumer behavior prediction based on artificial intelligence in marketing

[3]. Valecha, H., Varma, A., Khare, I., Sachdeva, A., & Goyal, M. (2018). Prediction of Consumer Behaviour using Random Forest Algorithm. Presented at the 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON). doi:10.1109/UPCON.2018.8597070.

[4]. Mathur, N., Kumar, S., Joshi, T., & Dhuliya, P. (2022). Analyzing Consumer Behavior Predictions: A Review of Machine Learning Techniques. Presented at the 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM). doi:10.1109/ICACCM56405.2022.10009209.

[5]. Quynh, T. D., & Dung, H. T. T. (n.d.). Prediction of Customer Behavior using Machine Learning: A Case Study. International School-Vietnam National University, Hanoi, Vietnam.

Top of Form