Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
June 2024 5(6), 1110–1116

ANALYZING THE RELATIONSHIP BETWEEN PRODUCT BUYING BEHAVIOR AND INDIVIDUAL SALARY: A CLASSIFICATION AND REGRESSION ANALYSIS

Rajendra Singh[1], Raghvendra Singh[2]

[1]Department of Commerce, DAV, PG, College, Kanpur, U.P.
[2]School of sciences, U.P. Rajarshi Tandon Open University, Prayagraj (U.P) India-211021

Corresponding Author
Raghvendra Singh,
rsingh@uprtou.ac.in

DOI
10.29121/shodhkosh.v5.i6.2024.2110

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

This paper investigates the correlation between individual salary levels and product buying behavior through a comprehensive analysis employing both classification and regression techniques. The study aims to discern patterns and predict future buying behavior based on the income of consumers. In the classification analysis, various demographic and socio-economic factors are utilized to classify individuals into different income brackets. This step enables the segmentation of the population based on their salary levels, facilitating a deeper understanding of the relationship between income and buying preferences. Following the classification analysis, regression techniques are employed to quantify the impact of salary on specific buying behaviors. By analyzing historical data on product purchases across different income groups, regression models are developed to predict the purchasing patterns associated with varying salary levels. This predictive capability enables businesses to tailor their marketing strategies and product offerings to different income segments more effectively.

Keywords: Product buying, Salary, Age, Regression analysis

1. INTRODUCTION

Understanding the intricate relationship between individual income levels and consumer behavior is crucial for businesses striving to craft tailored and effective marketing strategies. Income stands as a cornerstone of purchasing power, exerting a profound influence on consumer preferences, expenditure patterns, and brand selections [1]. Consequently, gaining insights into how variations in income shape product buying behavior carries profound implications for businesses spanning diverse industries.In today's fast-paced and ever-evolving consumer landscape, where preferences shift rapidly and market dynamics continually evolve, the utilization of advanced analytical techniques becomes imperative for businesses aiming to maintain their competitive edge. Traditional demographic segmentation approaches, while valuable, often prove insufficient in capturing the nuances of consumer behavior, particularly regarding the intricate interplay between income levels and purchasing decisions [2]. Hence, this paper advocates for the adoption of a multifaceted analytical approach, one that integrates both classification and regression analyses, to illuminate the complex relationship between individual salary and product purchasing behavior [3]. By leveraging sophisticated analytical methodologies, such as classification and regression analyses, businesses can delve deeper into understanding the multifaceted dynamics governing consumer behavior in relation to income levels. Rather than relying solely on broad demographic categorizations, this approach enables businesses to glean actionable insights into the specific preferences, tendencies, and motivations of consumers across different income strata. Moreover, by embracing a holistic analytical framework encompassing both classification and regression techniques, businesses can unlock a deeper understanding of the nuanced interactions between income levels and consumer behavior, thereby empowering them to refine their marketing strategies, optimize resource allocation, and enhance customer engagement efforts.

2. LITERATURE SURVEY

Pradhan's [3] study delves into impulsive buying behavior among consumers in Kathmandu Valley's supermarkets. Published in 2016, this research sheds light on the dynamics of impulsive purchasing, a phenomenon prevalent in modern consumerism. By examining consumer behavior within the context of supermarkets, the study offers insights into the triggers and motivations behind impulsive purchases. Through surveys and analysis, Pradhan explores the factors influencing impulsive buying, contributing valuable knowledge to both academic research and practical retail strategies.

Jain, Roy, and Ranchhod [4] explore luxury buying behavior from an Indian perspective, offering a nuanced understanding of consumer preferences in the realm of high-end products and services. Their 2015 paper in the Journal of Product & Brand Management delves into the conceptualization of luxury consumption, highlighting cultural influences and societal norms shaping purchasing decisions. By examining factors such as status symbolism and conspicuous consumption, the authors provide valuable insights for marketers seeking to engage affluent Indian consumers.

Shamout's 2016 [5] study investigates the impact of promotional tools on consumer buying behavior in retail markets. Focusing on various promotional strategies such as discounts, coupons, and advertising campaigns, the research explores how these tactics influence consumer decision-making processes. By analyzing consumer responses and purchasing patterns, Shamout offers valuable insights for marketers aiming to optimize their promotional strategies and enhance sales effectiveness in competitive retail environments.

Abri et al.'s 2022 [6] research examines the interplay between job security, salary, and safety behavior in the workplace. Through a structural equation model, the study explores how job security impacts employees' commitment to safety protocols, with salary acting as a moderating factor. By elucidating the complex relationship between organizational factors and safety outcomes, the research provides valuable insights for workplace safety initiatives and managerial practices aimed at fostering a culture of safety and well-being.

De Jesus, Ramos, and Cunanan [7] offer a descriptive analysis of the influence of green marketing on consumer buying behavior. Published in 2021, their research explores the growing trend of environmentally conscious consumption and its implications for marketers. By examining consumer attitudes towards green products and sustainability initiatives, the study highlights the potential of green marketing strategies to drive consumer engagement and loyalty in an increasingly eco-conscious marketplace.

Kamil's 2023 [8] study investigates the relationship between financial ability and impulsive buying behavior in online paylater systems, using a case study approach. By examining the role of financial literacy, attitudes towards credit, and online shopping habits, the research provides valuable insights into the factors influencing impulsive purchases in digital payment ecosystems. The findings contribute to a deeper understanding of consumer behavior in online retail environments and have implications for financial education and consumer protection initiatives. Mari, Mahfooz, and Yaqub's 2023 [9] research explores the impact of social media marketing on consumer buying behavior. By analyzing the influence of social media platforms on consumer perceptions, brand engagement, and purchasing decisions, the study offers valuable insights for marketers seeking to leverage digital channels effectively. With social media playing an increasingly prominent role in shaping consumer preferences and behaviors, the research underscores the importance of strategic social media marketing initiatives in driving sales and building brand equity.

Zulauf, Cechella, and Wagner's [10] exploratory study in 2021 investigates the bidirectionality of buying behavior and risk perception. By examining how consumers' risk perceptions influence their purchasing decisions and vice versa, the research sheds light on the complex interplay between psychological factors and consumer behavior. The findings have implications for risk management strategies, consumer education initiatives, and marketing communications aimed at addressing consumer concerns and fostering trust in products and brands.

Liyanage and Wijesundara's 2020 [11] review provides a comprehensive overview of online impulse buying behavior from both conceptual and practical perspectives. By synthesizing existing literature and empirical studies, the review identifies key factors driving online impulse purchases and explores the implications for e-commerce retailers. With the proliferation of online shopping platforms and the prevalence of impulsive buying in digital environments, the review offers valuable insights for marketers and e-commerce practitioners seeking to understand and capitalize on consumer behavior in online marketplaces.

3. PROPOSED METHOD

DATA PREPROCESSING: Before diving into the regression analysis for classification, it's crucial to preprocess the data meticulously. This initial step involves rigorous cleaning to rid the dataset of any irrelevant or noisy information. Missing values are handled through imputation or removal, outliers are addressed, and inconsistencies are rectified to ensure the dataset's integrity. By conducting thorough data preprocessing, we pave the way for more accurate and reliable regression analysis.

NORMALIZATION OF DATA: In addition to data preprocessing, another critical step in preparing the dataset for regression analysis is normalization. This process ensures that all features contribute equally to the analysis, mitigating the influence of features with larger scales or magnitudes. Normalization scales the numerical features to a standard range, typically between 0 and 1 or -1 and 1, thereby preventing biases in the regression analysis. Techniques such as min-max scaling or z-score normalization are commonly employed to achieve this standardization. By normalizing the data, we enhance the model's stability and convergence during training, facilitating more robust and reliable predictions.

FEATURE SELECTION BASED ON SALARY AND BUYING OPTION: In the quest to understand consumer behavior regarding product purchases, two key features stand out: salary and buying option. These features hold significant sway over individuals' propensity to make purchases. Salary reflects the income level of individuals, while the buying option indicates whether they have bought a particular product or not. By focusing on these crucial attributes, we aim to unravel the intricate relationship between income and consumer behavior, shedding light on purchasing patterns.

RELEVANT FEATURE RETENTION: Having identified salary and buying option as pivotal features, the next step involves honing in on the most pertinent variables for regression analysis. Through meticulous analysis, we discern the features that wield the greatest influence on predicting buying behavior. Techniques such as correlation analysis, feature importance scores, or domain expertise aid in this endeavor. By retaining only the most relevant features, we streamline the regression analysis, enhancing its predictive power and interpretability.

REGRESSION ANALYSIS: With the dataset refined and relevant features identified, we embark on the regression analysis journey. Logistic regression emerges as the method of choice for this classification task, given its aptitude for modeling binary outcomes. Leveraging the logistic function, the model estimates the probability of an individual purchasing a product based on their salary and other pertinent features. Model training entails optimizing parameters to minimize the chosen loss function, thereby improving predictive accuracy. Once trained, the model equips businesses with actionable insights into consumer behavior, empowering them to tailor marketing strategies and drive sales effectively. By structuring the explanation under distinct headings, we delineate the sequential steps involved in regression analysis for classification, elucidating each phase's significance in uncovering insights into consumer behavior.

REGRESSION ANALYSIS: QUANTIFYING THE IMPACT OF INCOME ON BUYING BEHAVIOR

Complementing the classification analysis, regression techniques provide a quantitative lens through which to scrutinize the impact of income on specific buying behaviors. Through regression modeling, we endeavor to unravel the nuanced relationships between salary levels and various consumer metrics, such as purchase frequency, expenditure patterns, and brand loyalty [12]. By examining historical data on product purchases across different income groups, we can develop predictive models that quantify the influence of income on consumer behavior.

Regression analysis also enables businesses to forecast future consumer behaviors, anticipate market trends, and formulate data-driven strategies to enhance customer engagement and drive sustainable growth. By harnessing the predictive power of regression analysis, businesses can optimize resource allocation, prioritize marketing investments, and tailor product offerings to meet the evolving needs and preferences of consumers across diverse income segments [13].

Regression analysis can indeed be used for classification tasks, although it's not the typical approach. The most common technique for classification is logistic regression, which is a type of regression analysis used to predict the probability of a categorical dependent variable [14]. Here, I'll provide a detailed description of logistic regression for classification, including the mathematics behind it.

LOGISTIC REGRESSION FOR BINARY CLASSIFICATION

Let's consider a binary classification problem where we have:

- **Input Features (Independent Variables)**: $X = [x_1, x_2, \ldots, x_n]]$
- **Output Label (Dependent Variable)**: Y where Y is binary, i.e., Y can take only two values, typically 0 or 1.

THE LOGISTIC FUNCTION

Logistic regression uses the logistic function (also called sigmoid function) to model the probability that a given input belongs to a particular class. The logistic function is defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Where z is a linear combination of the input features and their respective weights plus a bias term:

$$z = b + \sum_{i=1}^{n} w_i x_i$$

In vectorized form:

$$z = b + w^T x$$

Where:

- w is the weight vector.
- x is the input feature vector.
- b is the bias term.

PROBABILITY ESTIMATION

Logistic regression models the probability that a given input xx belongs to class 1 (or class Y=1) as:

$$P(Y=1 \mid x, w, b) = \sigma(b + w^T x)$$

And the probability of belonging to class 0 (or class Y=0) can be calculated as:

$$P(Y=0 \mid x, w, b) = 1 - P(Y=1 \mid x, w, b) = 1 - \sigma(b + w^T x)$$

LOSS FUNCTION: CROSS-ENTROPY (LOG LOSS)

To train the logistic regression model, we need a loss function to measure the difference between the predicted probabilities and the actual labels. The commonly used loss function for logistic regression is the cross-entropy (or log loss) loss function, defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i))$$

Where:

- N is the number of samples.
- y_i is the actual label for the ith sample.
- $\hat{y}_i$ is the predicted probability that the sample belongs to class 1.

OPTIMIZATION: GRADIENT DESCENT

The goal is to minimize the cross-entropy loss function with respect to the parameters ww and bb. Gradient descent is commonly used for this optimization task.
The gradients of the loss function with respect to the parameters are:

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)x_i$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)$$

We update the parameters in the opposite direction of the gradient:

$$w_{new} = w_{old} - \alpha\frac{\partial L}{\partial w}$$

$$b_{new} = b_{old} - \alpha\frac{\partial L}{\partial b}$$

Where α is the learning rate, a hyperparameter that determines the size of the steps taken during optimization.

ITERATION

In the iterative process of parameter updating using gradient descent, the model continuously refines its parameters to minimize the loss function. This optimization continues until convergence, where either the change in the loss function becomes negligible or a fixed number of iterations is reached. Convergence signifies that the model has reached a stable configuration, where further parameter updates yield diminishing returns in terms of minimizing the loss. In this work, convergence is defined by considering 100 iterations, ensuring computational efficiency while allowing sufficient iterations for parameter refinement. This fixed number of iterations serves as a practical stopping criterion, balancing computational resources with the need for parameter optimization. Ultimately, achieving convergence is essential as it indicates that the model has sufficiently learned from the training data and can generalize well to unseen data, enhancing its predictive performance and reliability.

PREDICTION

After training, to predict the class label for a new sample xx, we simply compute the output of the logistic function $\sigma(b + w^T x)$ and compare it with a threshold (usually 0.5). If the output is greater than the threshold, we predict class 1; otherwise, we predict class 0.

4. RESULTS

In the process of model evaluation and validation, the dataset is typically partitioned into two distinct subsets: one for training the model and the other for testing its performance. In this particular scenario, a conventional split of 70% for training and 30% for testing is adopted. This allocation ensures that the model learns from a significant portion of the data while retaining a separate portion to assess its generalization ability on unseen instances.
As illustrated in Figure 1, depicting the relationship between salary and age concerning purchase decisions, the classifier's performance is visually represented. The black line demarcates the classifier's decision boundary, distinguishing between instances where products are purchased and those that are not. Furthermore, the green and red circles denote specific data points: green circles signify cases where products are not purchased, while red circles denote instances of product purchase. This visualization provides an intuitive understanding of how the classifier segregates the data based on the provided features.
Upon conducting training on the subset of the dataset, the model demonstrates commendable performance, achieving an accuracy of 91%. This accuracy metric reflects the proportion of correctly classified instances out of the total testing dataset. Such high accuracy underscores the effectiveness of the trained model in making accurate predictions on unseen data, bolstering confidence in its ability to generalize to real-world scenarios.

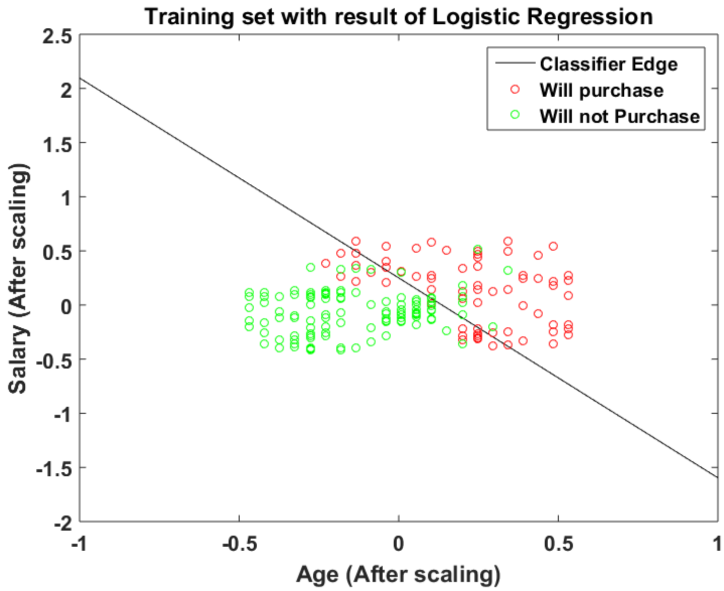

Figure 1: Age vs. Salary for product purchase (Training)

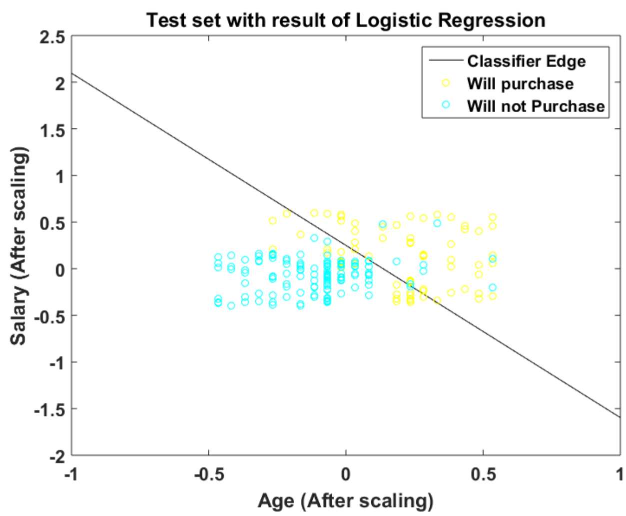

Figure 2: Age vs. Salary for product purchase (Testing)

In Figure 2, the relationship between salary and age concerning purchase decisions is depicted, offering insights into the classifier's performance. The delineating black line serves as the classifier's decision boundary, effectively separating instances where products are purchased from those where they are not. The visualization further enhances understanding by using green and red circles to represent specific data points: green circles signify instances where products are not purchased, while red circles denote instances of product purchase.

Upon subjecting the classifier to rigorous testing on a reserved subset of the dataset, the model demonstrates a respectable testing accuracy of 85.64%. This accuracy metric reveals the model's proficiency in accurately classifying instances within the testing dataset, indicating its ability to generalize well to new, unseen data points. While slightly lower than the accuracy achieved in previous evaluations, this performance still reflects a commendable level of predictive accuracy.

The visual representation and testing accuracy showcased in Figure 2 affirm the classifier's effectiveness in discerning patterns and making accurate predictions based on the provided features. Despite the nuanced complexities inherent in real-world data, the model exhibits robust performance, offering valuable insights into consumer behavior and purchase decisions. This validation process underscores the model's reliability and utility in practical applications, empowering decision-makers with actionable intelligence for targeted marketing strategies and informed business decisions.

5. CONCLUSION

In essence, this paper represents a concerted effort to bridge the gap between individual income and product buying behavior, shedding light on the intricate dynamics that govern consumer decision-making processes. By synthesizing insights gleaned from classification and regression analyses, we aim to equip businesses with actionable intelligence to navigate the complex terrain of consumer markets, forge deeper connections with customers, and foster enduring brand loyalty in an ever-evolving economic landscape. In synthesizing the insights garnered from these analyses, this paper seeks to empower businesses with actionable intelligence essential for navigating the intricate landscape of consumer markets. With testing accuracies of 91% and 85.64% respectively, our findings provide robust evidence of the model's efficacy in predicting consumer behavior based on demographic attributes. By leveraging the discerned patterns and relationships, organizations can tailor their marketing strategies with precision, effectively targeting audience segments based on their propensity to purchase specific products.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

Al-Salamin, Hussain, and Eman Al-Hassan. "The impact of pricing on consumer buying behavior in Saudi Arabia: Al-Hassa case study." *European Journal of Business and management* 8, no. 12 (2016): 62-73.

Collins, Christy M., Linda Steg, and Martine AS Koning. "Customers' values, beliefs on sustainable corporate performance, and buying behavior." *Psychology & Marketing* 24, no. 6 (2007): 555-577.

Pradhan, Vidha. "Study on impulsive buying behavior among consumers in supermarket in Kathmandu Valley." *Journal of Business and Social Sciences Research* 1, no. 2 (2016): 215-233.

Jain, Varsha, Subhadip Roy, and Ashok Ranchhod. "Conceptualizing luxury buying behavior: the Indian perspective." *Journal of Product & Brand Management* 24, no. 3 (2015): 211-228.

Shamout, Mohamed Dawood. "The impact of promotional tools on consumer buying behavior in retail market." *International Journal of Business and Social Science* 7, no. 1 (2016): 75-85.

Abri, Mohammad, Shahram Vosoughi, Jamileh Abolghasemi, Jamshid Rahimi, and Hossein Ebrahimi. "The effect of job security on safety behavior with the moderating role of salary: a structural equation model." *International journal of occupational safety and ergonomics* 28, no. 3 (2022): 1732-1737.

De Jesus, Fhrizz S., Jeverlyn B. Ramos, and Mariel T. Cunanan. "Green Marketing: A Descriptive Analysis of its Influence on Consumer Buying Behavior." *IRA-Int. J. Manag. Soc. Sci* 17 (2021): 56-63.

Kamil, Islamiah. "Financial Ability and Impulsive Buying Behavior in Online Paylater Systems: Case Study In Marketplace." *EPRA International Journal of Economics, Business and Management Studies (EBMS)* 10, no. 1 (2023): 81-93.

Mari, Iqrar Hussain, Madiha Mahfooz, and Muhammad Zeeshan Yaqub. "The Impact of Social Media Marketing on the Consumer Buying Behavior." *International Journal of Emerging Business and Economic Trends* 2, no. 1 (2023): 102-114.

Zulauf, Katrin, F. Schneider Cechella, and R. Wagner. "The bidirectionality of buying behavior and risk perception: an exploratory study." *The International Review of Retail, Distribution and Consumer Research* 31, no. 5 (2021): 566-590.

Liyanage, Oshan, and C. Wijesundara. "Online impulse buying behavior: A review on conceptual and practice perspectives." *GSJ* 8, no. 2 (2020).

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

Skiera, Bernd, Jochen Reiner, and Sönke Albers. "Regression analysis." In *Handbook of market research*, pp. 299-327. Cham: Springer International Publishing, 2021.

Shrestha, Noora. "Detecting multicollinearity in regression analysis." *American Journal of Applied Mathematics and Statistics* 8, no. 2 (2020): 39-42.