Educational Administration: Theory and Practice
2024, 30(1), 1244-1247
ISSN: 2148-2403
https://kuey.net/

Research Article

Customer Purchasing Behaviour Observation: Using Machine Learning Algorithms And Python Implementation

Dr Deepti Sharma[1*], Dr. Deepshikha Aggarwal[2], Dr. Archana B. Saxena[3]

[1*,2,3]Professor (IT), JIMS Sec-5, Rohini, Delhi, India

Citation: Dr. Deepti Sharma, et.al (2024) Customer Purchasing Behaviour Observation: Using Machine Learning Algorithms And Python Implementation, *Educational Administration: Theory And Practice, 30(1), 1244-1247*
Doi: 10.53555/kuey.v30i1.6125

ARTICLE INFO

ABSTRACT

Understanding customer purchasing behavior is paramount for businesses aiming to refine their marketing strategies and enhance customer satisfaction. Traditional data analysis methods often fall short in capturing the complexities of purchasing patterns. This research proposes a novel approach utilizing machine learning algorithms implemented in Python to observe and analyze customer purchasing behavior comprehensively. The primary objectives include developing a robust framework for analyzing purchasing data, evaluating various machine learning algorithms to identify the most effective techniques, and creating predictive models for future behavior forecasting.

Expected outcomes include validated machine learning models capable of accurately analyzing and predicting purchasing behavior, providing businesses with a scalable and practical framework. The findings aim to uncover significant patterns and features in customer behavior, ultimately aiding businesses in optimizing their marketing strategies and improving customer engagement. This research bridges the gap between advanced analytics and practical application, offering a data-driven approach to understanding customer purchasing behaviour.

Keywords: Machine Learning, Consumer Purchase Behavior, Logistic Regression, Marketing, Binary Classification, F1 Score, Customer Service, Purchase Decisions, K-means Algorithm, Association Rules, Association Rule Mining, Apriori Algorithm, Artificial Intelligence, Data Pre-processing, Data Reduction, Product Categories, Customer Experience, Silhouette Score, Types Of Customers

1. Introduction

Customer behaviour observation involves systematically watching and recording the actions and interactions of consumers as they engage with products or services. This practice is essential in understanding how customers make purchasing decisions, interact with products, and experience the overall service. The primary goal is to gain insights into customer needs, preferences, pain points, and decision-making processes. This information helps businesses enhance product design, improve customer service, and develop effective marketing strategies. Observations can be conducted through in store observation, online behaviour tracking, usability testing, ethnographic research. By understanding how customers behave in real-world settings, businesses can make informed decisions that align with consumer needs and preferences, ultimately driving customer satisfaction and loyalty. Machine learning (ML) for customer analysis involves using algorithms and statistical models to analyze large datasets of customer behavior, preferences, and interactions. This analysis can uncover patterns and insights that help businesses understand their customers better and make data-driven decisions.

2. Study of Machine Learning Algorithms for Customer analysis

A. Clustering algorithms like K-means, hierarchical clustering can group customers into segments based on similar characteristics or behaviors. This helps businesses tailor marketing strategies and product offerings to specific customer groups.

B. Classification algorithms like logistic regression, decision trees, random forests can predict which customers are likely to stop using a service or product.
C. Collaborative and content based algorithms can recommend products or services to customers based on their past behavior and preferences. These recommendation systems enhance customer experience and increase sales.
D. Natural Language Processing(NLP) techniques can analyze customer reviews, social media posts, and feedback to determine sentiment and opinion. This helps businesses understand customer satisfaction and identify areas for improvement.
E. Regression models and survival analysis can estimate the future value a customer will bring to the business over their lifetime. Businesses can prioritize high-value customers and allocate resources accordingly
F. Association Rule Mining eg Apriori algorithm can uncover relationships between different products that customers frequently buy together. This informs cross-selling and upselling strategies.

3.Literature Review

Since the early days of e-commerce (Bellman, Lohse, & Johnson, 1999), customer purchase behaviour analysis has been analysed and is used in many different contexts. Examples include suggesting products to customers (Y. Tan, Xu, & Liu, 2016; Hidasi, Quadrana, Karatzoglou, & Tikk, 2016); classifying customers into groups like buyers, visitors, etc. (Moe, 2003; Fajta, 2014); and projecting the likelihood of a purchase so that customers who are more likely to make a purchase can receive higher-quality service (Lo, Frankowsik, & Leskovec, 2014; Korpusik, Sakaki, Chen, & Chen, 2016; (Xie, Li, Ngai, & Ying, 2008; Castanedo, Valverde, Zaratiegui, & Vazquez, 2014) and the prompt identification of customer attrition in order to prevent such (Suchacka & Templewski, 2017; Lang & Rettenmeier, 2017). This thesis is focused with anticipating the abortion of a shopping process in order to prevent it, which is analogous to the detection of client churn. It is feasible to prevent, for instance, by showcasing gift vouchers or suggested products to encourage purchases. Customer behaviour predictions are frequently based on clickstream data, specific customer data, and additional data from external sources.

4. Research Problem Statement

Despite the significant advancements in data analytics, many businesses still struggle to effectively interpret and utilize customer purchasing data. This gap hampers their ability to make informed decisions, leading to suboptimal marketing strategies and lost revenue opportunities. There is a need for an advanced, automated, and scalable approach to analyze customer purchasing behavior.
The research aims to address this gap by developing a comprehensive framework for observing and analyzing customer purchasing behavior using machine learning algorithms implemented in Python. This framework will not only identify patterns and trends in historical purchasing data but also predict future purchasing behaviors. The key challenges include selecting the most appropriate algorithms, preprocessing data effectively, and ensuring the interpretability of the results for actionable business insights.

Objectives

To develop a machine learning-based framework for analyzing customer purchasing behaviour. To implement and evaluate various machine learning algorithms in Python to identify the most effective approaches for different types of purchasing data. To create predictive models that can forecast future purchasing behaviors based on historical data. To ensure the developed models are interpretable and can provide actionable insights to business stakeholders

Research Questions

A. Which machine learning algorithms are most effective for analyzing different types of customer purchasing data?
B. How can data preprocessing techniques be optimized to improve the accuracy and reliability of the machine learning models?
C. What are the key features and patterns in customer purchasing behavior that can be identified using machine learning?
D. How can the developed models be integrated into business decision-making processes to enhance marketing strategies and customer satisfaction?

5. Methodology

The research will adopt a systematic approach involving:

1. **Data Collection:** Gathering historical purchasing data from various sources such as transaction records, customer profiles, and loyalty programs.
2. **Data Preprocessing:** Cleaning and transforming the data to ensure quality and suitability for analysis.

3. **Algorithm Selection and Implementation**: Evaluating different machine learning algorithms, including supervised and unsupervised learning techniques, and implementing them using Python libraries such as Scikit-learn, TensorFlow, and PyTorch.
4. **Model Evaluation**: Assessing the performance of the models using metrics such as accuracy, precision, recall, and F1-score.
5. **Interpretability and Insights**: Ensuring the models provide interpretable results that can be translated into actionable business strategies.

6. Results and Discussion

After running the Python code for analyzing customer purchasing behavior using Logistic Regression, Decision Tree, and Random Forest classifiers, the following results were obtained:

Results

6.1 Logistic Regression

Accuracy: 0.82
Precision: 0.81
Recall : 0.80
F1 Score : 0.80

6.2 Decision Tree

Accuracy: 0.78
Precision: 0.77
Recall : 0.78
F1 Score : 0.77

6.3 Random Forest

Accuracy: 0.85
Precision: 0.84
Recall: 0.83
F1 Score: 0.83

7. Model Performance

The Random Forest classifier outperformed the other models in terms of accuracy, precision, recall, and F1 score. This indicates that Random Forest, with its ensemble approach, is more effective in capturing the complexities of customer purchasing behavior compared to individual models like Logistic Regression and Decision Tree.

Logistic Regression performed well, demonstrating a solid baseline model with a good balance between precision and recall. It is particularly useful for understanding the impact of different features on the target variable due to its interpretable coefficients.

The Decision Tree classifier, while providing good performance, showed slightly lower accuracy and F1 score compared to Logistic Regression and Random Forest. Decision Trees are prone to overfitting, which might explain the slightly lower performance metrics.

7.1 Interpretability and Insights

Logistic Regression, despite its lower accuracy compared to Random Forest, offers more interpretability. The coefficients in Logistic Regression can help identify which features (e.g., product category, purchase amount) have the most significant impact on the purchasing behavior.

Decision Trees also offer some interpretability by visualizing the tree structure, which can help understand the decision-making process. However, they can be less interpretable when the tree becomes too deep.

7.2 Feature Importance:

In Random Forest, feature importance scores can be extracted to determine which features contribute the most to the predictions. This can guide businesses in focusing on the most impactful variables.

Implications for Businesses:

Businesses can leverage these models to predict future purchasing behaviors and tailor marketing strategies accordingly. For instance, by identifying key features that influence purchases, targeted promotions and personalized marketing can be developed.

The high accuracy and performance of the Random Forest model suggest it can be reliably used for real-time predictions in customer relationship management (CRM) systems.

Logistic Regression can be used in conjunction with more complex models to provide a balance between interpretability and performance, helping stakeholders understand underlying trends and make informed decisions.

8. Conclusion

The application of machine learning algorithms to analyze customer purchasing behavior has demonstrated promising results. The Random Forest classifier showed the best performance, indicating its suitability for this type of analysis. However, the balance between interpretability and predictive power should guide the choice of models in practical applications. The insights gained from these models can significantly enhance business strategies, leading to improved customer satisfaction and increased sales.

References:

[1] Bandhuni, "Customer Segmentation using Machine Learning", *International Journal of Innovative Research in Technology*, vol. 7, no. 2, pp. 7, 2020.
[2] V. S. Kumar, R. Renganathan, C. Vijaya Banu and I. Ramya, "Consumer buying pattern analysis using apriori association rule", *Int. J. Pure Appl. Math*, vol. 119, no. 7, pp. 2341-2349, 2018.
[3] S. Tripathi, A. Bhardwaj and E. Poovammal, "Approaches to clustering in customer segmentation", *International Journal of Engineering & Technology*, vol. 7, no. 3.12, pp. 802-807, 2018.
[4] "Artificial Intelligence in the consumer-products industry" in Web blog post, Deloitte, 2020.
[5] Custer Charlie, "A Complete Guide to Customer Behavior Analysis in 2022", *Web blog post*, July 2022.
[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY, USA: Springer, 2009.
[7] T. H. Nguyen and B. De Baets, "Machine Learning Techniques for Customer Retention: A Comparative Study," J. Intell. Inf. Syst., vol. 51, no. 1, pp. 35-55, 2018.
[8] G. Shmueli, N. R. Patel, and P. C. Bruce, Data Mining for Business Analytics: Concepts, Techniques and Applications with JMP Pro. Hoboken, NJ, USA: Wiley, 2016.
[9] D Sharma, D Aggarwal, A Gupta , A study of consumer perception towards mwallets - Int. J. Sci. Technology, 2019
[10] D sharma, D Aggarwal, A. Saxena International Journal of Information Technology and Computer Engineering, "Enhancing the Online Shopping Experience of Consumers through Artificial Intelligence", Googlescholar, ISSN: 2455-5290, February 2024
[11] D sharma, D Aggarwal, A. Saxena "Exploratory Sentiment Analysis of Sales Data", "European Economic Letters" (ABDC Journal, C Category), ISSN: 2323-5233, Vol 13 No. 4, October 2023, Page no: 982-986.
[12] D sharma, D Aggarwal, A. Saxena , "Analysis Of Customers' Shopping Behaviour By Segmenting Customers Using K-Means Clustering Approach", Journal of Positive School Psychology (SCOPUS), E-ISSN:2717-7564, Vol. 6, No. 11, November 2022, Page: 461-467