WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Prediction of Consumer Behavior Using Machine Learning Techniques

Md Sirajul Huque[a]*, M Mounika[b]

[a]*Assistant Professor, Department of Computer Science & Engineering, Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.*

[b]*Assistant Professor, Department of Computer Science & Engineering, TKR College of Engineering and Technology , Hyderabad, Telangana ,India.*

A B S T R A C T

Reviews, ratings, and comments submitted by users can be examined to yield more insightful data for business purpose. Understanding the needs of the customer and foreseeing their future intentions with regard to the service are made easier by analyzing such consumer behavior. Marketing professionals always struggle with figuring out what customers want and need and, ideally, how to anticipate those demands. Researchers have used machine learning to solve various issues in consumer behavior analysis. To find the most accurate answer to a particular problem, we must research machine learning models and data pretreatment methods. Given a set of information regarding the customer's driving environment, the challenge is to forecast the choice of consumers with regard to whether they accept a restaurant voucher provided by an in-vehicle system. The fact that most characteristics are categorical and there are missing values makes the situation more challenging. We looked into ways to manage missing values, convert categorical variables to numerical attributes, and then apply different classification models. In order to comprehend an organization's demographics, this study uses data-driven marketing technologies including data visualization, machine learning, and natural language processing. Collaborative filtering, neural networks, and sentiment analysis are other techniques we use to create recommender systems.

Keywords: Data visualisation, sentiment analysis, classification models, customer behaviour, machine learning.

1. Introduction

Analysis of customer behaviour is crucial for businesses. It aids businesses in comprehending the demands and wishes of their customers. As a result, the business may provide the consumer a better service or a product that suits their needs. The corporation can boost revenues and improve its bottom line thanks to consumer behaviour research. These days, digital transformation has an impact on all business operations. Information systems may be used to gather data about items and consumers. In the current technological era, e-commerce usage has significantly expanded. Online shopping is more popular than traditional market shopping. [1] In e-commerce, customer generated data in the form of ratings and reviews of a specific product may be utilised as a method of product publicity and legitimacy. These ratings and reviews help consumers make decisions about whether or not to purchase a certain product. Such information may consist of opinions expressed either favourably or unfavourably by users of the product.

To obtain insights and comprehend the needs and intents of their customers, e-commerce companies might benefit from a thorough study of this user-generated information. We can use machine learning algorithms to plot precise visual representations of this type of customer behaviour. Such visual analyses would provide a more thorough examination of the dataset in order to comprehend and draw specific inferences about the general customer behaviour on the e-commerce platform. Machine learning's branch of natural language processing, called natural language processing (NLP), is used to analyse text and detect favourable or negative consumer evaluations[9]. Therefore, the information produced in the form of reviews can affect how buyers behave, including their intents to buy a product. Real-time suggestions may be generated using this simple analysis to provide customers a tailored purchasing experience. This will entice customers to make larger purchases, increasing the organisations' total profit. The amount of study in this area is little, despite the fact that machine learning holds promise for tackling challenges involving the prediction of consumer behaviour. Additionally, before

* *Corresponding author.* Tel.:+91 9963828634
E-mail address: mailtosirajul@gmail.com

using a machine learning model, data must be pre-processed. Furthermore, there isn't a good approach for every dataset. To enhance the findings for a particular dataset, research into pre-processing techniques and machine learning models is thus required. With the use of cutting-edge technology, the context for improving consumer behaviour prediction has been expanded in numerous ways. Though there is a vast field of machine learning available, there is currently little study on particular business solutions that might help address challenges. This study is a novel strategy that enhances the accuracy of several prediction models. As a result, the marketers will have a broad and distinct understanding of how machine learning may benefit them and additional options for using machine learning models to grow their firms.

2. Literature Study

Since corporations now gather consumer data, predicting client purchasing behavior is a significant subject that has been studied by several researchers. Many methods have been used to forecast customers' buying patterns. The majority of them made use of data mining and machine learning methods. A review of customer churn prediction was also given by authors Santharam and Krishnan (2018)[2]. 18 modelling approaches from diverse sectors were covered in this article. One sector where consumer purchase behaviour has been extensively researched is telecommunications. Studies that sought to precisely anticipate customer attrition were presented by Momin et al. in 2019. Different algorithms, including K-nearest neighbour, decision tree, random forest, logistic regression, naive Bayes, and ANN, were compared. ANN had the highest accuracy, with 83.8%. E-commerce is another area that is expanding quickly. In this field, several researchers have worked[8]. According to the attention, interest, search, action, and share (AISAS) model, Kachamas et al. (2019) suggested an analytical tool for online retailers to forecast the behaviour of customers. The naive Bayes approach was used to construct the classification model. More than 86% of accuracy was reached by the model. The majority of clients who purchase goods on an online marketplace are once-off bargain seekers. Charanasomboon and Viyanon (2019) offered a technique for recurring buyer prediction to find devoted consumers. The model was built using random forest, gradient boost, XgBoost, and leave-one-out.

In addition, Ullah et al. (2019) suggested a prediction model to pinpoint consumers who churn as well as the causes of customer churn. Different classification techniques were applied to categorize consumers that churned, and a random forest (RF) algorithm worked well, with an accuracy rate of 88%. Amin et al. (2018) compared and evaluated the effects of cutting-edge data transformation using a just-in-time customer churn prediction (JIT-CCP) model. To determine the JIT-CCP model's performance, they ran three tests [3].

In their 2018 article, Alloghani et al. covered the use of machine learning in software engineering education and student performance prediction. Seven approaches and two datasets have been employed. For the first dataset, the neural network provided the greatest accuracy, while for the second dataset, the random forest provided the best accuracy. Similar to this, Liu et al. (2018) addressed the issue of data scarcity and conducted a holistic study of temporal aspects[4]. This study combined temporal and geographical elements in a Euclidean space by using POI prediction. A DMETS model was proposed in light of this. Metric embedding was used to build a model that was more adaptable to different contextual circumstances. The authors also assessed performance using additional models, such as the BPR-MF, PRME-G, WMF, WWO, and ST-RNN, and found that these models produced better outcomes on benchmark datasets like Foursquare and Gowalla. Following that, a machine learning-based technique for maintaining the position confidentially of roaming PBS users was suggested. The authors used a decision tree and K-nearest neighbour learning methods to determine the user's position. In order to predict users' destinations using this approach, including their position track sequence, the hidden Markov model was used. Finally, 90% of the position confidentiality for this task was reviewed by PBS (Sangaiah et al. 2019)[5].

A class-specific extreme learning machine (CS-ELM) was suggested in 2018 by researchers to address the binary class imbalance problem utilising class-specific regularisation parameters. In 38 different datasets obtained from the KEEL dataset repository, this strategy produced a reduced computational cost and greater accuracy than a prior weighted extreme machine learning model (Raghuwanshi and Shukla 2018).Based on earlier research, this article thoroughly covers each algorithm as well as the geometrical and mathematical reasoning behind machine learning classification methods. Customer behaviour has been examined in this work as either retail or hotel, restaurant, and café (HoReCa).

3. Methodology

In its widest sense, consumer behavior refers to how customers choose, decide, utilise, and dispose of goods and services. It encompasses any vertical's persons, groups, or organisations.It offers useful information and insights into the thoughts, feelings, and preferences of consumers that influence their purchasing decisions. As a result, the consumers receive value, marketers are better able to understand their demands, and the business makes money. Consumer behaviour can be predicted by: Customers are segmented into smaller groups depending on their purchasing habits. This aids in the division of duties, which in turn aids in the localization of the market. Through the application of statistical methods, we may forecast future consumer behaviour by analysing historical data from the past.

The UCI-Irvin machine learning data repository is used in the proposed study to process data from wholesale clients. Eight characteristics make up the dataset, with the target class being the channel (HoReCa and retail clients). A total of 440 instances (298 HoReCa customers and 142 retail customers) are present (Cardoso 2014). To determine the class of the consumers, many classification techniques are applied[6].

The following steps make up the entire analytical process:

- collect the data
- Data preparation
- Feature selection

- Making the model fit
- Evaluating the model's precision

3.1 Machine Learning Method:

We employ machine learning classification techniques in this study. Predictive modelling is the process of roughly estimating the mapping function from discrete input variables to output variables. The main objective is to identify the category or class to which the new data belongs. We used a variety of classification techniques for the studies, which were chosen because of their wide usage, well-understood behaviour, and positive performance on a number of categorization tasks. The classifiers that may learn different functional forms of classification borders as well as classifiers that are based on different hypotheses on the relationships between various characteristics were the subjects of our investigation. A variety of classification models, such as decision trees, random forests, and support vector.

4. Conclusion and future scope

Data mining has a significant role to play in e-commerce. In comparison to earlier research, this study outlines the most significant supervised machine learning algorithms and achieves a higher accuracy on this dataset. Machine learning techniques work well for small datasets, whereas deep learning with more parameters is recommended for large datasets. The static datasets are the only ones supported by supervised learning methods. If there is a dynamic dataset, the time series or deep learning models should be the main emphasis. Customer research is crucial for all businesses. Future product sales must be made to the targeted client in order to satisfy the demands of opposed customers. Therefore, a consumer analysis is crucial.

The highest degree of accuracy was 92.42%. To improve the accuracy of personality prediction, the diversity of the data may be modified. Although most of the significant machine learning algorithms have been examined, a model's accuracy may be increased by combining new hybrid algorithms with a fresh dataset that contains more examples. Deep learning developments have made it possible to assess client purchase patterns using video data. In smart malls, a more creative solution may be offered to anticipate or propose the goods that shoppers would buy based on their need.

Explainable AI will allow models to behave as white boxes in the future by increasing their transparency and helping users comprehend the reasoning behind consumer segmentation. Defining the key characteristics for each classification is a brand-new problem that has to be solved.

REFERENCES

1. N. Asghar. 2016. Challenge using the Yelp dataset: predicting review ratings. Preprint for arXiv is arXiv:1605.05362.
2. Verstraeten, G., Buckinx, W., and Van den Poel, D. Utilizing the internal transactional database to forecast client loyalty. 32(1):125–134 in Expert Systems with Applications.
3. Y. Ding, W. S. DeSarbo, D. M. Hanssens, K. Jedidi, J. G. Lynch Jr., & D. R. Lehmann (2020). Measurements and techniques used in marketing analysis in the past, present, and future.promotional letters, https://doi.org/10.1007/s11002-020-09527-7.
4. Dou, X. (2020). Using ensemble learning, forecast and analyse online purchasing behaviour.The 5th IEEE International Conference on Cloud Computing and Big Data Analytics (ICCCBDA 2020) will take place in 2020. (pp. 532–536). https://doi.org/10.1109/icccbda49378.2020.9095554.
5. Kaefer, Frank, C. M. Heilman, and S. D. Ramenofsky 2005 Application of a neural network to customer categorization to enhance the timing of direct marketing*gn*,*236*, 54.
6. Orogun Adebola BO (2019) A machine learning technique to predicting customer behaviour in the digital economy. Int. J. Innov. Res. Sci. Eng technological 8(8):8391-8402
7. Wei, Z. Adeniyi, and Yongquan (2016) K-nearest neighbour is used in an automated online usage data mining and recommendation system (KNN) classification strategy Application Computing Inform 12(1):90-108
8. S. Agatono-Kustrin and R. Beresford (2000) basics of artificial intelligenceThe use of neural network (ANN) modelling in pharmaceutical research. 22(5):717-727 J Pharm Biomed Anal
9. Ahmad N, Maqsood I, Khan R, and Ali J (2012) Decision trees and random forests. Int J Comp Science 9 (5). http://ijcsi.org/papers/IJCSI-9-5-3-272-278.pdf
10. Hussain A, Mustafna J, Baker T, Alloghani M, Al-Jumeily D, and AljaafAJ (2018) Machine learning applications for teaching software engineering and early performance prediction of students. computer and communications in