ISSN: 2633-4828 Vol. 6 No.1, January, 2024

International Journal of Applied Engineering & Technology

MACHINE LEARNING PARADIGM FOR CUSTOMERS BEHAVIORAL MONITORING IN E-COMMERCE PLATFORMS

Riyadh Mejbel Yaseen Al-Dulaimi[1*], Amel Meddeb Makhlouf[2] and Ahmed Fakhfakh[3]

[1,2,3]NTS'COM National School of Electronic and telecommunication of Sfax, University of Sfax , Tunisia

[3]Laboratory of signals, System, Artificial Intelligence and Networks, Digital Research Center of Sfax, University of Sfax, Tunisia

[1]riadmajbal2@gmial.com , [2]amel.makhlouf@enetcom.usf.tn and [3]ahmed.fakhfakh@enetcom.usf.tn

ABSTRACT

With the advent of internet technology and the growth of mobile communications, the global economy has witnessed a shift towards digital and software-based markets. In response, significant efforts have been made to establish a virtual marketplace that offers an exceptional shopping experience to every customer. In the realm of e-commerce platforms, customer behavior plays a pivotal role. Companies are keenly interested in understanding how different customer segments respond to their products. To extract valuable insights from marketing data, data mining technologies such as clustering techniques, classification, and prediction are employed. In this study, we analyze the impact of data mining on e-commerce, collecting the necessary data through a questionnaire-based review. The findings reveal that data mining can significantly enhance the performance of e-commerce by enabling intelligent and efficient prediction of customer behaviors. Moreover, this technology also brings several advantages to end-customers, improving their overall experience. Three algorithms namely: Naïve Bays (NV), Fandom Forest (RF), K-nearest Neighbour (KNN) are used to classify the customers behaviours data. Results of RF algorithm are found optimum in terms of behaviours prediction accuracy (96.342%).

Key words: - E-marketing, Shopping, Clustering, Neural Network, Mean, Standard Deviation, HTML, Event Prediction, Random Forest, Naïve Bays, K-nearest Neighbour.

1 INTRODUCTION

In today's rapidly changing world, various levels of the economy have experienced significant impacts, leading experts to term it globalization or inclusivity [1]. At the heart of this ongoing phenomenon, which gained momentum in the latter half of the twentieth century, lies information and communications technology (ICT) [2][3]. Despite being broad and deep concepts on their own, information and communication are intricately intertwined. Without information and its technologies, communication would be limited, and the development of communication means would be hindered. Similarly, information would lack the importance it holds today without communication and its technologies [4].

The cohesion between information and communication has had a profound influence on today's world, characterized by the breakdown of spatial and temporal barriers, largely thanks to the internet—an outcome of information technology. The internet has enabled global communication and information exchange anytime and anywhere, profoundly reshaping the way institutions operate. It has brought forth new rules, particularly in the realm of business [5][6]. One of the crucial aspects is recognizing the customer as the key to excellence. Institutions strive to create value for customers by continuously meeting their evolving needs and understanding their preferences even before they are expressed. Achieving this necessitates leveraging information technology optimally, with E-commerce serving as a central component aligned with information technology within organizations. Information technology serves as a vital link between institutions and their environment, and its successful utilization, especially via the internet, requires the availability of physical and knowledge infrastructures [7][8].

In today's economy, which is increasingly oriented towards E-commerce and the internet transactions, The importance of e-commerce has grown. Understanding customer thinking is essential for developing effective marketing strategies. Each customer category has distinct requirements, necessitating marketing planners to

address these factors. Through customer analysis behaviors, One can get an insight into their needs and thought processes. Various methods such as web logging, online surveys, and personal interviews are employed to collect relevant data. However, the volume of data for analysis can be overwhelming, surpassing the capabilities of conventional methods and standard computer programs. Extracting knowledge from big data poses a significant challenge for marketing professionals. Data mining technology can help address this challenge, although factors like cost, privacy, and reliability also come into play.

2 CUSTOMER TENDENCY PREDICTION

Customer tendency prediction technology plays a crucial part in enhancing the effectiveness of e-commerce sites. It includes utilizing various data analysis techniques, machine learning algorithms, and predictive modeling to understand and anticipate customer behavior and preferences. By analyzing past customer interactions, browsing patterns, purchase history, Demographics of the country and other respective data, E-commerce sites can gain valuable understand into individual customer preference and tendencies [9].

a. **Personalized Recommendations:** through analyzing customer conduct and history buy, e-commerce sites can provide individualized product recommendations to individual customers. These recommendations are based on the customer's preferences, browsing history, and similar profiles of other customers [10].

b. **Predictive Analytics:** Predictive analytics techniques are used to forecast customer tendencies and predict future actions. By applying machine learning algorithms to historical data, e-commerce sites can anticipate customer behavior, such as the likelihood of making a purchase, preferred product categories, and even the potential value of a customer over time [11].

c. **Customer Segmentation:** Customer tendency prediction technology helps in segmenting customers into different set based on their favorite, behaviors, and characteristics. This segmentation allows e-commerce sites to customize market strategies, advertising promot and resultant offerings to specific customer segment [12].

d. **Abandoned Cart Recovery:** By analyzing customer behavior, including cart abandonment patterns, e-commerce sites can identify customers who have abandoned their shopping carts and implement strategies to recover those sales. This may involve sending personalized emails with incentives or reminders to complete the purchase [13].

e. **Dynamic Pricing:** Customer tendency prediction technology can be used to determine optimal pricing strategies. By analyzing customer data, market trends, and competitor pricing, e-commerce sites can adjust prices dynamically based on customer segments, demand, and other relevant factors [14].

f. **Churn Prediction:** E-commerce sites can use customer inclination prediction technology to identify customers who are at risk of changing or ceasing their participation in the platform. By identifying such clients in advance, targeted retention strategies can be implemented to prevent churn [15].

So-to-say, customer tendency prediction technology empowers e-commerce sites to provide personalized experiences, enhance customer satisfaction, and improve sales and conversion rates. By taking advantage of advanced analytics and machine learning techniques, e-commerce businesses can gain a competitive edge by understanding and catering to customer preferences and tendencies effectively.

3 MACHINE LEARNING

Technologies of Machine learning are widely used to predict the shopping tendencies of customers in e-commerce. Through analyzing the huge amounts of data collected from customer interplay, browsing behavior and buying history, and other relevant factors, models of Machine learning can make accurate predictions about customers ' future shopping behaviors. Here's how machine learning is used for customer shopping tendency prediction in e-commerce:

a. **Data Collection:** E-commerce platforms collect data on customer interactions, including website visits, product views, cart additions, purchases, and demographic information. This data forms the foundation for training predictive models.

b. **Feature Extraction:** Relevant features are extracted from the collected data to represent customer behavior and preferences. These features can include factors such as browsing duration, numeric of visits, product categories viewed, frequency of purchase, and customer demographics.

c. **Model Training:** Machine learning models, such as decision trees, random forests, support vector machines, or neural networks, are trained using historical data. The models learn the patterns and relation between the extracted features and the corresponding shopping tendencies of customers.

d. **Feature Engineering:** To improve model performance, feature engineering techniques are applied to transform or combine existing features. This step helps uncover more meaningful patterns and capture the underlying complexities of customer behavior.

e. **Model Evaluation and Validation:** The trained models are evaluated using validation data to assess their performance in predicting customer shopping tendencies accurately. We can used evaluation metrics Various , such as accuracy, precision, recall, or the area under the curve (AUC), can be used to measure the effectiveness of the model.Prediction and Recommendation: Once the model is deemed reliable, it can be deployed to predict customer shopping tendencies for new or existing customers. Based on these predictions, e-commerce platforms can provide personalized recommendations, targeted promotions, and tailored shopping experiences to customers.

f. **Continuous Improvement:** The predictive models are continuously updated and refined as new data becomes available. This iterative process allows the models to adapt to changing customer behaviors and preferences over time, improving their accuracy and effectiveness.

By taking advantage of machine learning algorithms, E- commerce platforms can gain valuable look into customer shopping trends. This enables them to enhance customer satisfaction, optimize marketing strategies, and increase conversion rates by delivering personalized experiences and targeted recommendations.

4 RELATED WORK

By deploying machine learning and predicting customer behaviors in e-commerce, implementing a behavior tracking subsystem is critical for the successful implementation of an e-commerce portal. Based on the results of a previous survey, "machine learning" was initially aimed at replicating the learning capabilities of the human brain. Over time, machine learning algorithms have advanced to cope with complex computational tasks, including forecasting, grouping and classifying huge amounts of data. There are three types of machine learning algorithms: supervised learning, unsupervised learning, and reinforced learning.

We will use three algorithms, the first of which the Naive Bayes method is a common and simple probability classification algorithm based on Bayes ' theorem, assuming that the features are conditionally independent . Despite its" naive "assumption, Bayes' naivety has been surprisingly effective in many real-world applications, especially in text classification and spam filtering. Secondly, the nearest neighbor algorithm is a simple and versatile supervised learning algorithm used for classification and regression tasks. It works based on the principle that data points are more likely to belong to the same category if they are closer in feature space.Thirdly, the random forest algorithm is a group learning method that combines multiple decision trees to improve accuracy and durability. Each tree is trained on a random subset of data and features, and the final prediction is made by compiling the predictions of individual trees.

Through it, a marketing model can be created by collecting and extracting data and predicting the customer's inclinations, the idea of improving the product or reprogramming the marketing plan based on the customer's

desire until reaching new results that enable us to have an integrated marketing model that meets the beneficiary's ambition

5 MODEL IMPLEMENTATION

Implementing a behavior tracking subsystem is crucial for the successful implementation of an e-commerce portal. Based on the findings of a previous survey study, understanding the behaviors of customers on the e-commerce platform is vital for improving the portal's profitability. To achieve this, machine learning techniques are utilized, which are explained in detail below.

Learning is an inherent part of various activities in our lives, such as cooking, repairing equipment, playing sports, and many others. Prior knowledge and skills are necessary for the brain to effectively perform these actions. The human brain stands out as an intelligent machine with its rapid signalling and information processing capabilities, enabling advanced learning and data analysis through millions of specialized cells.

Initially, "machine learning" aimed to replicate the learning abilities of the human brain. Over time, machine learning algorithms have advanced to handle complex computational tasks, including prediction, clustering, and categorization of vast amounts of data. There are three types of machine learning algorithms: supervised learning, unsupervised learning, and reinforcement learning.

Prediction poses a significant challenge for machine learning since it relies on past data to anticipate future events. The equation below provides a more comprehensive explanation:

$$y[n] = a_1 + x[n]a_2 \quad (1)$$

y[n] represents the predicted future event at a specific time point n, while x[n] refers to the input data for the machine learning model. The learning coefficients a1 and a2 are established using the same machine learning paradigm. By adjusting these coefficients, the learning process determines the accuracy of y[n]. The primary objective of the machine learning model is to minimize the error by optimizing the learning parameters (a1, a2). A more detailed overview of the learning process is illustrated in Figure 1.

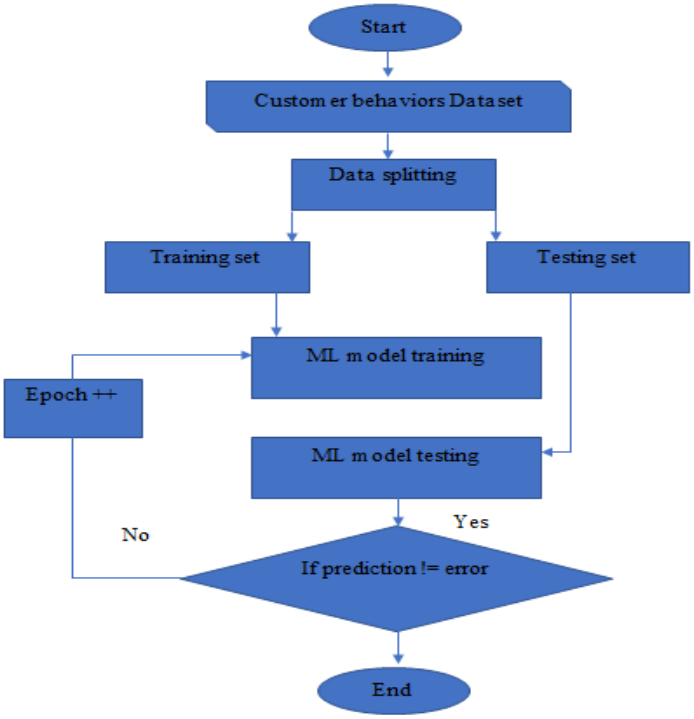

Figure 1: Training of algorithm in order to perform the customers behaviors tracking.

In summary, the implementation of a behavior tracking subsystem in an e-commerce portal is crucial, and machine learning techniques play a vital role in analyzing and predicting customer behaviors. By leveraging these approaches, the portal can enhance its profitability and provide a more personalized and efficient user experience.

5.1Experimental Results

5.1.1 Naive Bayes Method

The Naive Bayes algorithm is a popular and simple probabilistic classification algorithm based on Bayes' theorem, with the assumption that the features are conditionally independent given the class label. Despite its "naive" assumption, Naive Bayes has been surprisingly effective in many real-world applications, especially in text classification and spam filtering.

Usage in E-commerce Dataset Processing: In e-commerce, Naive Bayes can be utilized for various tasks, including sentiment analysis, product categorization, and spam detection. For example, it can be employed to analyze customer reviews and classify them as positive, negative, or neutral, helping businesses understand customer satisfaction levels. Additionally, Naive Bayes can categorize products based on their descriptions, titles, or attributes, simplifying product organization and search on e-commerce platforms. Furthermore, the algorithm can identify and filter out spam emails or reviews, ensuring that customers receive relevant and trustworthy information.

5.1.2 K Nearest Neighbor (KNN) Algorithm

K Nearest Neighbor is a simple and versatile supervised learning algorithm used for classification and regression tasks. It works based on the principle that data points are more likely to belong to the same class if they are closer in the feature space.

Usage in E-commerce Dataset Processing: KNN finds applications in e-commerce datasets for tasks such as recommendation systems, customer segmentation, and anomaly detection. For instance, it can be used to build collaborative filtering recommendation systems, suggesting products based on the preferences and behaviors of similar users. Moreover, KNN can cluster customers based on their purchase history or browsing behavior, helping businesses tailor marketing strategies to different customer groups. Additionally, the algorithm can identify unusual patterns or behaviors in e-commerce data, such as detecting fraudulent transactions or abnormal customer activities.

5.1.3 Random Forest Algorithm

Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy and robustness. Each tree is trained on a random subset of the data and features, and the final prediction is made by aggregating the predictions of individual trees.

Usage in E-commerce Dataset Processing: Random Forest is particularly effective in e-commerce datasets for tasks like product recommendation, customer churn prediction, and sales forecasting. It can be used to build product recommendation systems by predicting customer preferences based on various features and historical data. Moreover, the algorithm can help predict whether a customer is likely to churn (stop using the service or platform), allowing businesses to take proactive measures to retain valuable customers. Additionally, Random Forest can be employed to forecast future sales based on historical sales data, market trends, and other relevant factors, assisting in inventory management and resource planning.

5.2 Dataset

The dataset contains behavior data spanning 7 months, from October 2019 to April 2020, obtained from a large multi-category online store. Each row in the dataset represents an event, with all events being associated with products and users. The relationship between products and users follows a many-to-many model. The data was collected through the Open CDP project, utilizing an open-source customer data platform. The dataset is compromised of the following features:

Event_time	Time when event happened at (in UTC).
Event_type	Only one kind of event: purchase.
Product_id	ID of a product
Category_id	Product's category ID
Category_code	Product's category taxonomy (code name) if it was possible to make it. Usually present for meaningful categories and skipped for different kinds of accessories.
Brand	Downcased string of brand name. Can be missed.
Price	Float price of a product. Present.
ser_id	Permanent user ID.

Those were employed to carry out the identical goal of predicting user or client behavior. The prediction accuracy for each algorithm is evaluated in order to determine the ratio of properly expected outcomes to all results. The accuracy of the sickness prediction is given in Equation 2.

$$AC = \frac{C_P}{T_D} * 100\% \quad (2)$$

Last but not least, AC is the ACcuracy measure's percent determined for the output outcomes, where Cp denotes the quantity of accurate judgements and Tp denotes the total quantity of decisions, including both accurate and faulty ones.

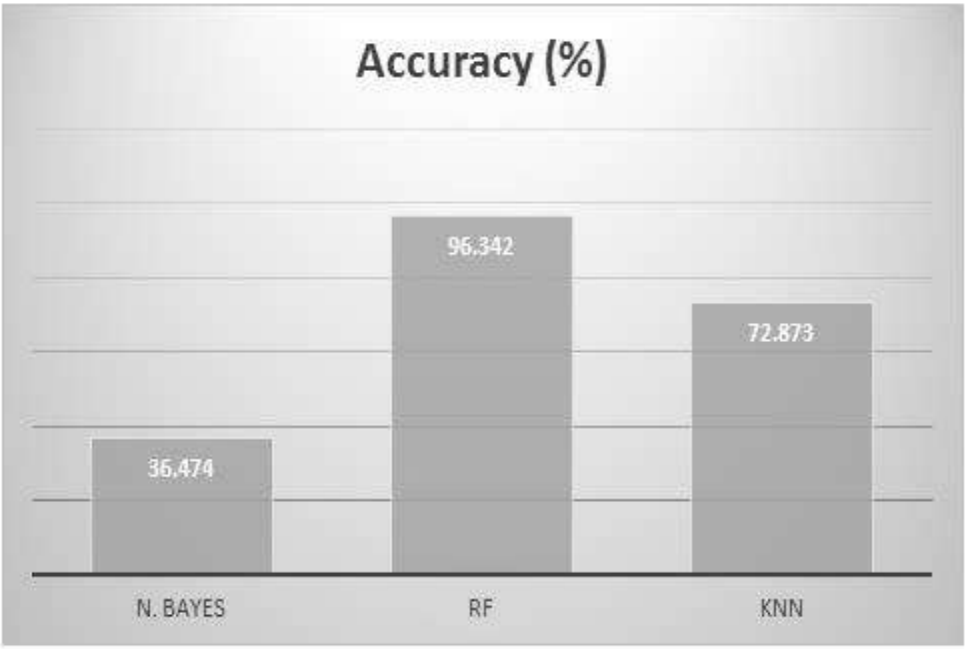

Figure 2: Measures of accuracy were utilized in the project's tools.

5.2.1 Prediction Time

A distinct performance indicator is being created for each strategy. The effectiveness of any processing tool mostly hinges on how rapidly it can handle data. It is essential to the success of many applications, including live streaming and real-time software. The first three algorithms used to predict the behavior of the dataset (among the Machine Learning algorithms). We are keeping an eye on the results as well as the calculated interval between the algorithm's beginning and end. The K Nearest Neighbor method is demonstrated to offer the smallest amount of time, while the Random Forest technique is proved to require the longest time. The time and the differences in time between the algorithms are:

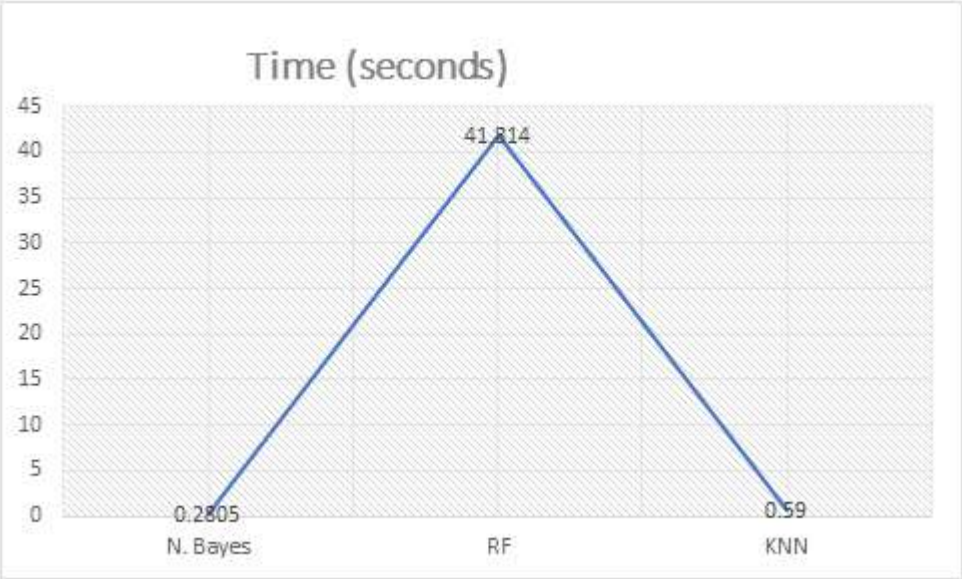

Figure 3: Measures of prediction time.

5.2.2 MSE

The effectiveness of deep learning and machine learning paradigms is significantly impacted by the MSE (Mean Square Error) performance measure because it reveals the level of inaccuracy in the results[7]. Or, to put it another way, the root mean square error and the mean square error abstract the degree of precision in the findings. Figure 4 provide statistical and visual representations of this measure.

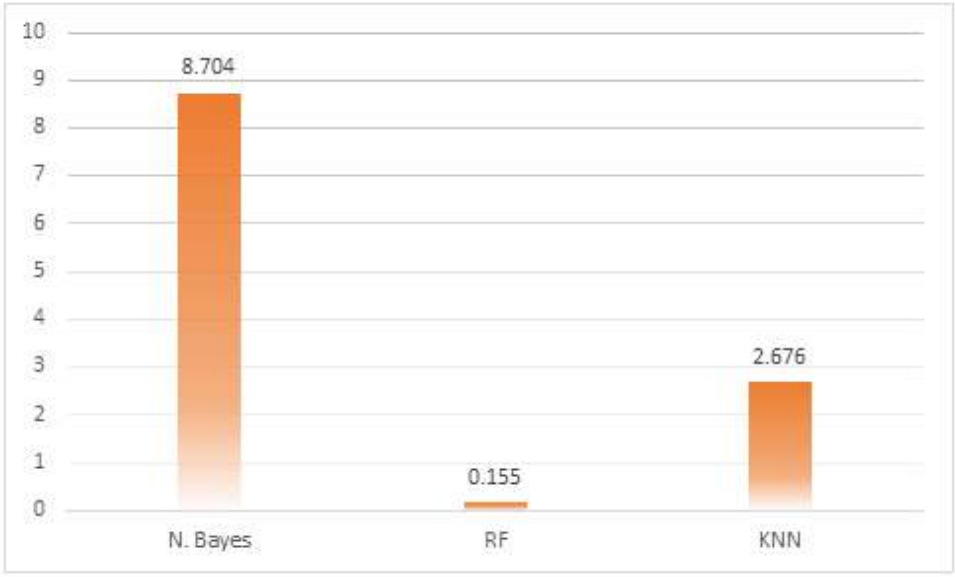

Figure 4: Measures of MSE.

5.2.3 MAE

Deep learning and machine learning models are influenced a lot by this performance metric. The mean absolute error (MAE) shows the level of inaccuracy in the results that. Or in other words, the root means the absolute error and the mean square of the error abstracts the degree of accuracy of the results. Figure 5 provides statistical and visual representations of the average absolute error.

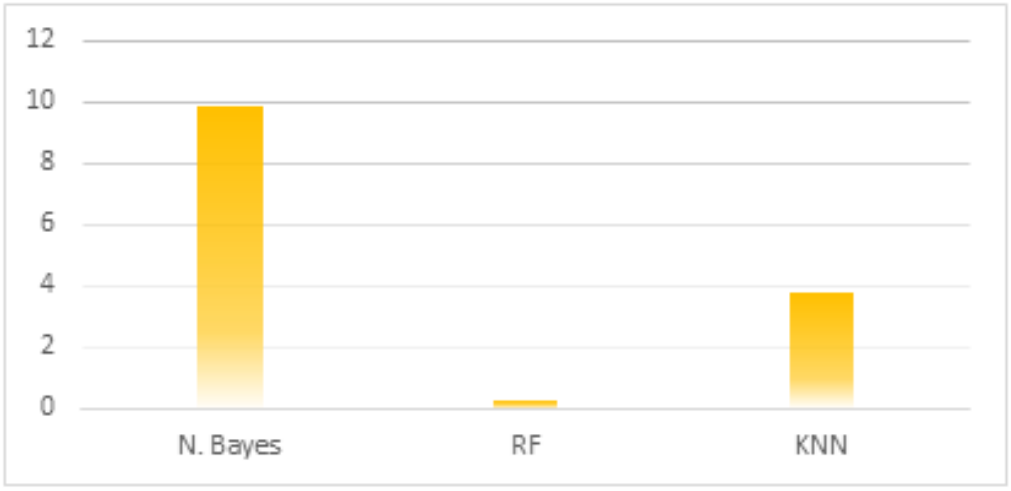

Figure 5: Measures of MAE.

5.2.4 RMSE

A key performance indicator for comparing the effectiveness of the machine learning and deep learning paradigms is root mean square error. The degree of inaccuracy in the results is shown by the root mean square error. To make the root mean square error easier to comprehend, smaller figures are given. The root mean square error is presented analytically and graphically in Figure 6.

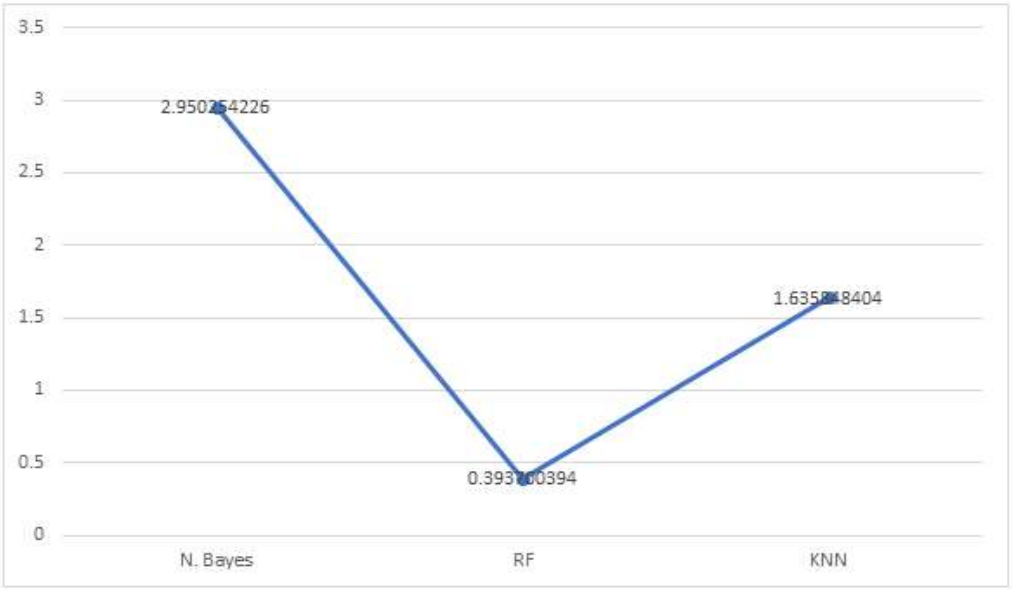

Figure 6: Measures of RMSE.

6 CONCLUSION

Data mining has become a crucial aspect of modern businesses due to advancements in communication and technology. However, mining and extracting knowledge from the large volumes of data collected daily from customer activities on shopping portals present significant challenges. Therefore, the utilization of data mining technologies, such as clustering, classification, prediction, and tools like neural networks, requires a comprehensive understanding of both technical and economic considerations.

Additionally, it is suggested that customer behaviour data be collected through the use of cookies on website pages. Cookies are cost-efficient and highly trusted for predicting customer behaviours, Random Forest has proved a reliable performance in customer behaviours prediction with 96.34% in 41.814 seconds. Overall, this study highlights the significance of data mining in E-Commerce and provides insights into the preferences and considerations of industry professionals regarding its implementation.

7 REFERENCES

[1] Bhowmick, S., Datta, S., & Dey, S. (2020). Customer behavior analysis in E-commerce using machine learning algorithms. In International Conference on Computational Intelligence in Data Science (pp. 99-109). Springer.

[2] Zhang, W., Zheng, X., Qi, Y., Wang, X., & Li, Q. (2019). Predicting customer behavior in E-commerce using machine learning algorithms. Journal of Retailing and Consumer Services, 51, 272-283.

[3] Zhu, X., & Liu, B. (2019). Customer purchase prediction in E-commerce using machine learning techniques. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 503-515). Springer.

[4] Luo, Y., Li, X., & Wang, S. (2020). Customer behavior prediction in E-commerce using ensemble machine learning algorithms. Electronic Commerce Research and Applications, 39, 100912.

[5] Li, X., Wang, Y., Wang, Y., & Lin, J. (2018). A novel customer behavior analysis framework for E-commerce using machine learning techniques. Expert Systems with Applications, 112, 149-160.

[6] Wang, X., Gao, F., & Li, H. (2017). Customer behavior prediction in E-commerce using deep learning techniques. Expert Systems with Applications, 82, 114-123.

[7] Huang, L., Guo, X., Guo, Y., & Dong, M. (2018). Customer behavior analysis in E-commerce using LSTM recurrent neural networks. International Journal of Information Management, 43, 100-110.

[8] Chang, J., Chen, H., & Zhu, Y. (2017). Customer behavior prediction in E-commerce using machine learning techniques based on browsing and transaction records. Electronic Commerce Research and Applications, 24, 62-74.

[9] Liu, S., Li, L., Liu, Q., & Guan, J. (2019). Customer behavior prediction in E-commerce using machine learning algorithms with feature selection. Neural Computing and Applications, 31(9), 4921-4931.

[10] Wang, Z., He, M., Wang, M., & Sun, S. (2019). A customer behavior prediction model for personalized recommendation in E-commerce. Complexity, 2019, 1-12.

[11] Yan, X., Li, Y., Zhang, L., & Zhang, L. (2018). Customer purchase behavior prediction in E-commerce using deep learning. In International Conference on Internet of Things and Machine Learning (pp. 229-238). Springer.

[12] Liu, Y., & Shen, D. (2018). Customer behavior analysis and prediction in E-commerce using machine learning algorithms. In 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 42-47). IEEE.

[13] Niazi, M. A., Rashid, N. A., & Abas, H. (2019). Predictive modeling for customer behavior in E-commerce using machine learning techniques. In International Conference on Information Retrieval and Knowledge Management (pp. 92-103). Springer.

[14] Hwang, H. G., Han, I., & Lee, S. (2019). Customer churn prediction in E-commerce using machine learning techniques. Journal of Business Research, 104, 236-246.

[15] Yang, Y., & Wei, X. (2018). Customer segmentation in E-commerce using machine learning techniques. In International Conference on Advanced Data Mining and Applications (pp. 366-378). Springer.