e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 9, September 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 9, September 2024| DOI: 10.15680/IJMRSET.2024.0709018

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Prediction of Customer Loyalty in E-Commerce using Machine Learning Models

Mahesh K N, Rosemary R, Punith Rao Gayakwad S, Prof. Bhavana B R

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

PG Student, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

Assistant Professor, Department of M.C.A, Surana College (Autonomous), Kengeri, Bangalore, India

ABSTRACT: In competitive e-commerce, customer loyalty is a critical factor for long-term success. Loyal customers not only generate steady income but also act as brand promoters, reducing marketing costs and increasing organizational growth. However, predicting customer loyalty remains a difficult challenge due to the versatile and dynamic nature of consumer behavior. This study aims to predict customer loyalty using machine learning techniques on an 'e-commerce consumer behavior and purchasing' dataset. In this paper, we implement and compare Random Forest, SVM, and ANN models to accurately identify loyal customers and provide valuable insights to enhance customer retention strategies. Our findings reveal that all three models demonstrate strong performance with minimal differences in accuracy. Consequently, the models are deemed optimal, with SVM, ANN, and RF achieving accuracy rates of 94.10%, 94.05%, and 93.10%, respectively.

KEYWORDS: Machine Learning , Random Forest, Support Vector Machine, Artificial Neural Netwok, Customer Loyalty.

I. INTRODUCTION

Client reliability is crucial for e-commerce as it drives repeat purchases, lowers acquisition costs, and encourages positive word-of-mouth that attracts new customers. Loyal customers generate higher conversion rates and provide a stable revenue stream, which facilitates more accurate business forecasting and planning. They also contribute valuable feedback for product and service enhancements, boosting overall brand confidence and customer relationships. Ultimately, loyalty supports continuous growth and profitability, making it an essential element for success in the competitive e-commerce sector.

For businesses to thrive, analyzing and investigating customer retention is vital, particularly in the context of increasing rivalry in well-established industries. Different approaches are used to explore this crucial component of economic sustainability[1].Here, identifying churn is essential for forecasting which customers are likely to depart from the company or cancel a service subscription by reviewing their interaction and behavior with the product. Many existing studies concentrate on predicting churn utilizing machine learning models, including SVM, Logistic Regression, Decision Tree, KNN, Random Forest and XGBoost. In contrast, this paper focuses on developing models to project customer loyalty, which is crucial for corporations to enhance their services and provide greater benefits to their customers**.**

II. LITERATURE REVIEW

Churn prediction models were designed and assessed through the adoption of Random Forest (RF) and Logistic Regression (LR) approaches for businesses on e-commerce marketplaces. Tested three different methods: under-sampling, over-sampling, and no preprocessing at all. F-Scores of 0.76, 0.71, and 0.92 were obtained by RF models by stratified cross-validation, whereas F-Scores of 0.84, 0.68, and 0.69 were obtained by LR models. The potency of different machine learning algorithms is impacted by preprocessing strategies, as demonstrated by these studies[2].Research in e-commerce analytics has effectively utilized machine learning algorithms to predict various business metrics, such as customer churn, annual spending, and on-time product delivery. Studies have demonstrated that the Support Vector Machine model, in particular, excels in these predictive tasks, achieving accuracies of 83.45\% for customer churn and 68.42\% for on-time delivery predictions. Key factors influencing these predictions include purchase frequency, product categories, and average session length. These findings provide valuable insights for optimizing customer retention strategies and enhancing revenue generation in the e-commerce sector[3].A classification model has been proposed to effectively differentiate between churners and non-churners within the e-commerce sector. The study employed Naive Bayes, K-Nearest Neighbors along with Decision Tree algorithms to evaluate various models using real customer data. Among the algorithms assessed, the Decision Tree emerged as the most impactful, achieving a notable accuracy of 93.6\%. The model's optimal performance was noted with a tree depth of 14, underscoring its robustness in predicting customer churn. These results highlight the practical benefits of utilizing Decision Trees to enhance customer retention strategies in e-commerce environments[4].

This study conducted a thorough assessment of Random Forest, k-Nearest Neighbors (kNN) and Logistic Regression algorithms for estimating customer churn in the e-commerceindustry. Employing a dataset sourced from Kaggle, which details customer churn within a prominent online E-commerce company,each algorithm's performance was reviewed in the context of its suitability to the dataset's characteristics. Logistic Regression showed moderate predictive capability but faced challenges given its linear assumptions. kNN emerged as the top performer with high sensitivity (98.22\%) and specificity (96.35\%), surpassing other models. Random Forest excelled in specificity (95.83\%) but demonstrated slightly lower sensitivity (91.75\%). These findings underscore kNN's effectiveness in churn prediction and highlight the varied strengths within machine learning models in e-commerce analytics[5].Anticipating customer turnover in B2C e-commerce was explored using LR and SVM models with customer behavior data. The research demonstrated that segmentation via k-means clustering notably enhanced prediction accuracy across various metrics, with SVM outperforming LR in predictive performance. However, the study faced limitations due to reliance on a single dataset and a limited set of predictive variables. Future research could benefit from comparative analyses across multiple datasets and the inclusion of a broader range of customer behavior metrics to improve predictive accuracy and real-world applicability[6].The capabilities of random forest and logistic regression models in predicting churn among e-business clients was investigated using historical data and various attributes. The findings revealed that both models are capable of accurately forecasting e-business turnover, demonstrating their potential for proactively managing customer retention strategies in the digital marketplace[7]

A comprehensive comparison of five machine learning techniques Artificial Neural Network, Naive Bayes,Adam Deep Learning, Random Forest, and Support Vector Machine (SVM) reveals their effectiveness in predicting customer churn within Brazilian e-commerce. The evaluation, conducted using a benchmarked dataset, employs assessment metrics such as accuracy, sensitivity, specificity, true positive, and true negative rates.Within the series of models Random Forest classifier, when optimized through feature selection via Neighborhood Component Analysis, achieves the highest prediction accuracy at 99.35\%. This highlights Random Forest's superior performance in accurately forecasting customer churn, offering actionable insights for improving customer retention strategies in the digital commerce sector[8].Using a dataset from Tehran's leading online food ordering service, this article explores the elements leading to customer churn and develops strategies for retention. The analysis centers on predicting customer churn by leveraging online properties and user behavior data. The findings reveal that Gradient Boosted Trees surpass other techniques, achieving a degree of accuracy of 86.90\%.This underscores the model's effectiveness in accurately determining customer retention risk, providing essential insights for improving retention efforts within Tehran's online food delivery industry[9].This analysis utilizes the Decision Trees algorithm on data from a Korean influencer marketing agency to anticipate customer actions . By leveraging the significant impact influencers have on their followers, the study a Forecast accuracy of 90\%, as measured by the F-measure. These findings demonstrate the worth of using influencers in social media marketing to effectively forecast and address customer churn within brand promotion strategies[10].Customer churn, a critical issue for businesses, is examined through a spectrum of data mining tools in conjunction with machine learning techniques. This study specifically evaluated the the efficacy of the gradient boosting model. The model achieved atraining accuracy of 93\% and atesting accuracy of 91\%, indicating its robustness in projecting customer churn. These results highlight gradient boosting technique's potential for accurately forecasting customer patterns and aiding in retention strategies[11].

© 2024 IJMRSET | Volume 7, Issue 9, September 2024| DOI: 10.15680/IJMRSET.2024.0709018

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III.METHODOLOGY

In this research, several machine learning models are used to predict client loyalty in an e-commerce dataset. Knowing which customers will be loyal to a firm is crucial for creating effective strategies that increase client happiness and retention, which boosts revenue and expansion. With the aim of this, a methodical process that included data collection, preprocessing, features selection, model training, parameter adjustment, and evaluation was used. The rigorous procedure used to develop and evaluate the predictive models is described in this methodology section.

A. Data Collection

This research utilized data from the "e-commerce consumer behavior and purchasing" dataset that was obtained from Kaggle. https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis In-depth records of consumer interactions with an online store are encompassed in this dataset. These records include purchase history, product categories, payment options, customer demographics, and loyalty metrics. The dataset used is ideal for studying and forecasting client loyalty since it offers a deep insight into consumer behavior. The study aims to identify significant trends and insights that can guide strategies to improve customer satisfaction and retention in the e-commerce industry by utilizing this large dataset.

B. Data Preprocessing

By getting the dataset ready for analysis and model training, data preparation is essential. The initial phase of this research work comprised the following key actions:

1. **Handling Missing Values:** Missing data may skew results and diminish the efficiency pertaining to the model. The dataset was verified to be complete and consistent by assigning a value of 0 to any absent values in the 'Returns' column.
2. **Parameter Tuning:** We performed parameter tuning to define customer loyalty by creating a 'target' column. To determine whether a customer is labeled as 'loyal' or 'not loyal', we used two criteria: If the count of unique Customer IDs associated with a customer exceeded 5 and the total number of returns was fewer than 3, the customer was labeled as 'loyal'. Conversely, if these conditions were not met, the customer was categorized as 'not loyal.' This approach allowed us to effectively segment clients categorized by their shopping behavior and return activity, which was crucial for accurately assessing customer commitment in our predictive models.
3. **Feature Selection:** The behavior of a model might be adversely affected by duplicated or irrelevant features. For the purpose of addressing the most useful parameters for forecasting customer loyalty, several columns from the feature set, like "Customer ID," "Purchase Date," "Customer Name," and "Gender," were eliminated in this investigation.

C. Data Splitting

To test the models on unobserved data, the sampling dataset was split into a testing set (20%) and a training set (80%). The testing data was leveraged to evaluate the models' performance on fresh, untested data, while the training set allowed the models to discover patterns among the data. This 80/20 split is a standard procedure that guarantees the models can prevent overfitting and generalize well, offering a trustworthy assessment of their capacity to estimate customer allegiance in real-life circumstances.

D. Model Training

Three predictive models in machine learning were trained on the dataset used for training:

1. **Random Forest Model:** An effective ensemble learning technique called the Random Forest Classifier builds several decision trees during training and combines their results to reduce overfitting and increase prediction accuracy. A random sample from the dataset is employed to train each decision tree within the ensemble, and the mean of each tree's predictions or the most frequent vote is utilized to make the final decision. This model is incredibly precise, resilient, and resistant to overfitting because to this method, which takes benefit from diversity among the trees to lower variance and bias especially when working with intricate datasets which possess a large quantity of attributes.

© 2024 IJMRSET | Volume 7, Issue 9, September 2024| DOI: 10.15680/IJMRSET.2024.0709018

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. **Support Vector Machine (SVM):** Strong classification algorithms like SVM find the appropriate hyper plane to divide data into distinct classes. The nearest data points for every class or support vectors, are selected so as to maximize the margin, or interval, between them and the hyper plane. Especially in situations where the categories are well-separated, this margin maximization aids SVM in achieving high accuracy and generalization. Because of its adaptability and ability to handle both linear and non-linear classification using kernel functions, SVM is useful for a variety of classification applications**.**
3. **Artificial Neural Network (ANN):** ANN is a model whose architecture and operation are modeled after biological neural networks. It is composed of layers upon layers of networked nodes, or neurons, that process incoming data and, during training, become able to uncover trends. After applying a weight and an activation function to the input, each neuron transfers the outcome to the subsequent layer. Because ANNs are so good at identifying intricate, non-linear correlations in data, they are a good fit for jobs like NLP, image recognition, and predicting consumer loyalty. They can represent complex relationships and patterns that more basic models could overlook because of their capacity to learn from and generalize from enormous datasets.

E. Model Evaluation:

Accuracy: Accuracy measures the proportion of correctly predicted instances out of the total instances in the dataset. It is calculated by dividing the number of correct predictions by the total number of predictions, providing an overall indicator of model performance.

Classification Report: Provides detailed metrics including Precision, Recall, F1-score, and Support for each class.

- **Precision:** The ratio of correctly predicted positive instances to the total predicted positives.

$$\mathbf{Precision} = \frac{\mathbf{TruePositives}}{\mathbf{TruePositives+FalsePositives}}$$

- **Recall (Sensitivity) :** The ratio of correctly predicted positive instances to all actual positives.

$$\mathbf{Recall} = \frac{\mathbf{TruePositives}}{\mathbf{TruePositives+FalseNegatives}}$$

- **F1- Score :** The harmonic mean of precision and recall, balancing both metrics.

$$\mathbf{F1-score} = \mathbf{2} \times \frac{\mathbf{Precision \times Recall}}{\mathbf{Precision+Recall}}$$

Confusion Matrix: A confusion matrix is a matrix that provides an overview of how well a machine learning model performs on a given set of test data. Derived from the model's predictions, it is a way of illustrating the ratio of accurate and inaccurate cases.

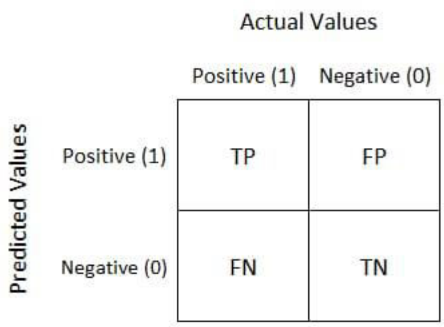

Fig 1. Confusion Matrix

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. IMPLEMENTATION

This research examines "e-commerce consumer behavior and purchasing" dataset, which consists of 250,001 customer transaction records. By leveraging this dataset, we gain valuable insights into customer purchasing patterns. To predict customer loyalty, we apply several predictive modeling techniques, to assess models such as Artificial Neural Networks (ANN), Random Forest, and Support Vector Machines (SVM), and analyze their performance.

While performing the model on the dataset, we encountered an issue where the same Customer ID appeared multiple times for different product categories. While attributes such as Purchase Date, Product Price, Quantity, Total Purchase Amount, Payment Method, Customer Age, and Returns had unique values for each entry, the repetition of Customer IDs created challenges during model evaluation. This led to confusion in predicting customer loyalty, as the model struggled to account for the same customer across multiple transactions. Consequently, the accuracy of predictions and overall functionality of the models were negatively impacted.

A data consolidation strategy was employed to tackle the concern regarding duplicate Customer IDs inside the dataset. First, we combined the dataset using the Customer ID to aggregate the transaction details. We accomplished this by adding up the components of key attributes like Quantity, Total Purchase Amount, and Returns. Every customer's behavior was seen in detail due to this aggregation. We then integrated the old dataset on the Customer ID with this consolidated data to formulate a new dataset that showed aggregated values for distinct customer records. As each customer was represented by a single, detailed record, this approach effectively handled the duplication problem and improved the reliability of our model evaluation.

49,674 records represent the recently assembled dataset. We used significant parameters including Quantity, Total Purchase Amount, Returns, and Age to train the models, with target serving as the prediction parameter. We split the data into training and test sets, dividing 80% from the dataset for training and 20% for testing, using the train_test_split function. To be more precise, the models were trained on 39,739 records, while 9,935 records were set aside for testing. By splitting the data in this manner, it was feasible to ensure where the model could be trained on a significant percentage concerning the data and maintain a representative sample for testing and evaluation.

A. Random Forest Model

We have implemented Random Forest model to forecast customer loyalty. Using this method, ten decision trees are trained, each made from a random subset of the data. Combining the results from each of these different trees yields the final prediction. The model's 93.09% accuracy rate indicates this is a good predictor of client retention. This exceptional level of accuracy shows the model’s effectiveness and how well it can generalize from the data, making efficient utilization of the ensemble approach to manage the elaborate structure of the dataset.

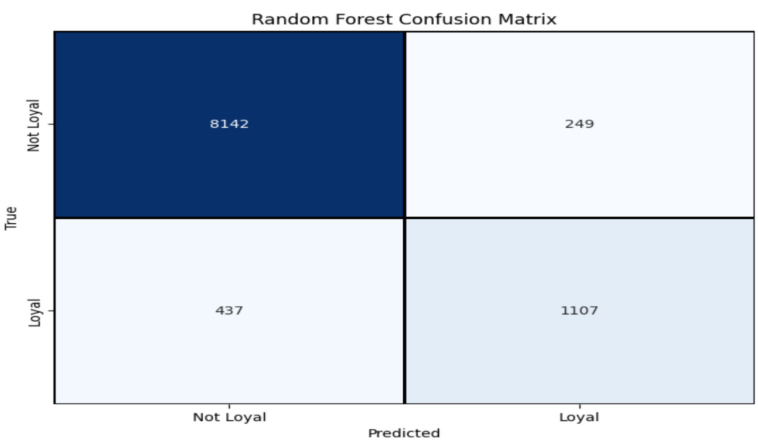

Fig. 2. Random Forest Confusion Matrix

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

As indicated by the confusion matrix of Figure 2, the model correctly classified 8,142 instances of "Not Loyal" clients and 1,107 instances of "Loyal" customers. However, it misclassified 249 "Not Loyal" customers as "Loyal," and 437 "Loyal" customers were incorrectly predicted as "Not Loyal." These results highlight the model's effectiveness in identifying "Not Loyal" customers, while exhibiting some challenges in accurately predicting "Loyal" customers.

B. Support Vector Machine

The SVM model was trained on the dataset to distinguish customer loyalty by identifying the best hyper plane that divides the data points into distinct categories. It iteratively adjusts its parameters to maximize the margin between classes, thus enhancing its predictive capability. After evaluating the model, it achieved a measurement of accuracy at 94.10%, indicating that it effectively captured the patterns within the dataset and successfully classified loyal and non-loyal customers extracted from the selected features. This high accuracy reflects the model's ability to generalize well to the data.

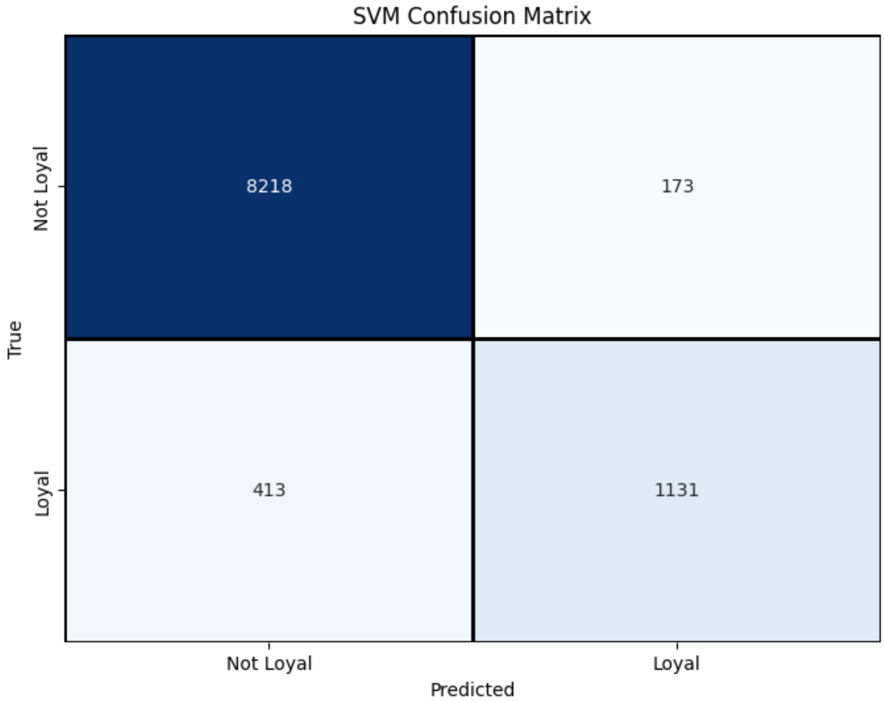

Fig. 3. SVM Confusion Matrix

As per the confusion matrix in Figure 3, SVM model correctly classified 8,218 instances of "Not Loyal" customers and 1,131 instances of "Loyal" customers. However, it misclassified 173 "Not Loyal" customers as "Loyal," and 413 "Loyal" customers were incorrectly predicted as "Not Loyal." These results indicate that SVM model demonstrates strong performance in determining "Not Loyal" customers while maintaining relatively remarkable accuracy in predicting "Loyal" customers.

C. Artificial Neural Network

Leveraging several layers of neurons, the ANN model was served to process the attributes within the dataset. Using back propagation, the model iteratively optimized the predictions during training by adjusting its internal weights. The accuracy rate of the model was 94.05% after training. This high accuracy shows that the ANN was successful in understanding the relationships between the input variables and customer loyalty, which allowed for precise predictions to be made on the test data.

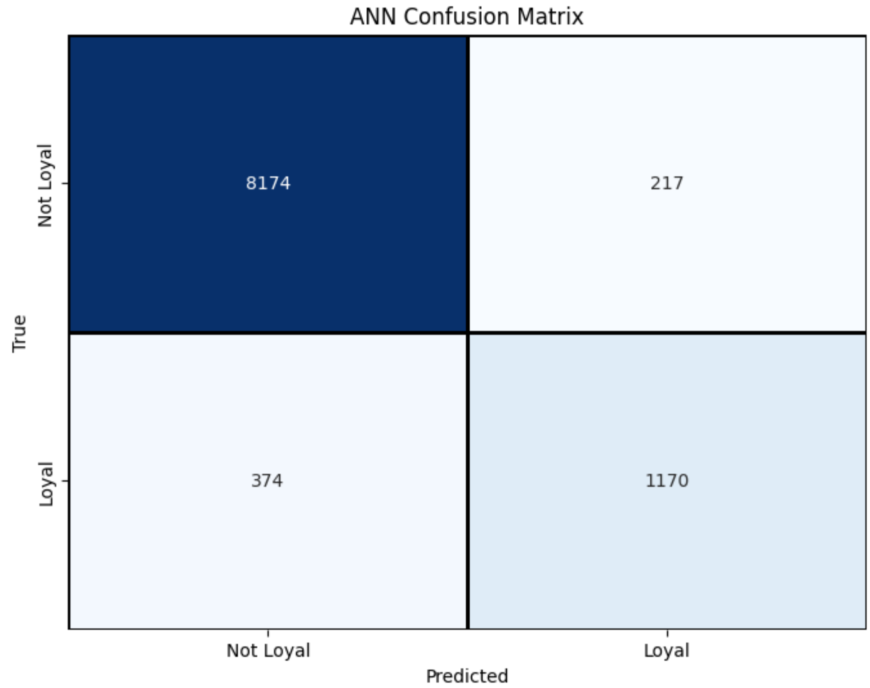

Fig. 4. ANN Confusion Matrix

Referring to the confusion matrix in Figure 4, ANN model correctly classified 8,142 instances of "Not Loyal" customers and 1,107 instances of "Loyal" customers. However, it misclassified 249 "Not Loyal" customers as "Loyal," and 437 "Loyal" customers were incorrectly predicted as "Not Loyal." These results highlight the model's effectiveness in identifying "Not Loyal" customers, while exhibiting some challenges in accurately predicting "Loyal" customers.

D. Discussion

The table below summarizes the results of the SVM ,Random Forest, and ANN models in predicting customer loyalty, comparing metrics such as the level of Accuracy, Recall, ,Precision and F1-score. These values offer insights into each model's effectiveness, helping to recognize the utmost reliable approach for distinguishing between loyal and non-loyal customers.

Metrics	Random Forest	SVM	ANN
Accuracy	0.9309	0.9410	0.9405
Precision (Class 0)	0.95	0.95	0.96
Precision (Class 1)	0.82	0.87	0.84
Recall (Class 0)	0.97	0.98	0.97
Recall (Class 1)	0.72	0.73	0.76
F1-Score (Class 0)	0.96	0.97	0.97
F1-Score (Class 1)	0.76	0.79	0.80

Table : Comparison of Random Forest, SVM, and ANN Models

1. **Accuracy :** In evaluating model accuracy, the Support Vector Machine achieved the highest performance with 94.10%, followed closely by the Artificial Neural Network (ANN) at 94.05%, and Random Forest model at 93.10%. These results highlight that the Support Vector Machine model provides the most accurate predictions for

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

customer loyalty. While the ANN and Random Forest models also performed well, their slightly lower accuracies suggest that the SVM is marginally more effective in classifying customer loyalty within the dataset.

2. **Precision :** In comparing precision for classifying customer loyalty, the Support Vector Machine (SVM) model showed the highest performance for both classes, with a precision of 0.95 for class 0 (non-loyal customers) and 0.87 for class 1 (loyal customers). The Artificial Neural Network (ANN) also performed well, achieving a precision of 0.96 for class 0 and 0.84 for class 1. The Random Forest model had slightly lower precision, with 0.95 for class 0 and 0.82 for class 1. These outcomes suggest that the SVM model is marginally more precise in distinguishing between loyal and non-loyal customers in comparison with the ANN and Random Forest models**.**
3. **Recall :** In terms of recall, the Support Vector Machine (SVM) outperformed the other models with a recall of 0.98 for class 0 and 0.73 for class 1. The Artificial Neural Network (ANN) followed, achieving a recall of 0.97 for class 0 and 0.76 for class 1. The Random Forest model had the lowest recall, with 0.97 for class 0 and 0.72 for class 1. These results highlight that the SVM excels in identifying class 0, while the ANN is slightly better at identifying class 1.
4. **F1-Score:** In terms of F1-Score, the Artificial Neural Network (ANN) performed the best for both classes. For Class 0, ANN achieved an F1-Score of 0.97, matching the Support Vector Machine (SVM) and slightly outperforming Random Forest, which also had an F1-Score of 0.96. For Class 1, ANN led with an F1-Score of 0.80, followed by SVM with 0.79, and Random Forest with 0.76.These results indicate that while both SVM and ANN excel in identifying Class 0, ANN provides the highest overall performance in identifying Class 1, making it the most balanced model in terms of F1-Score.

V .CONCLUSON AND FUTURE WORK

In conclusion, the e-commerce industry emphasizes significant weight on the capacity to anticipate customer loyalty because it has a direct impact on strategic decision-making, customer retention initiatives, and overall business success. Based on an extensive dataset, this study analyzes the effectiveness regarding three predictive models in predicting customer loyalty: Random Forest, Artificial Neural Networks, and Support Vector Machine. The analysis illustrates that the SVM and ANN models perform well in this field, with respective accuracies of 94.10% and 94.05%. These models performed well in differentiating between loyal and disloyal customers, which gave rise to crucial insights to develop focused marketing campaigns. By contrast, the Random Forest model had marginally lower performance, with an accuracy of 93.10%. Overall, all the models performed effectively, with SVM and ANN showing slightly better accuracy. Future work will involve conducting a longitudinal analysis to monitor shifts in customer retention over time and assess how effectively the models predict loyalty in evolving, real-world conditions.

REFERENCES

[1] Nagaraj, P., et al. "E-Commerce Customer Churn Prediction Scheme Based on Customer Behaviour Using Machine Learning." 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2023.
[2] Öztürk, Mehmet Emin, Akasya Akyüz Tunç, and Mehmet Fatih Akay. "Machine learning based churn analysis for sellers on the e-commerce marketplace." International Journal of Mathematics and Computer in Engineering (2023).
[3] Al Rahib, Md Abdullah, et al. "Customer data prediction and analysis in e-commerce using machine learning." Bulletin of Electrical Engineering and Informatics 13.4 (2024): 2624-2633.
[4] Shi, Aoxia , Chia Yean Lim, and Sau Loong Ang. "Customer Churn Analysis for Live Stream E-Commerce Platforms by Using Decision Tree Method." International Conference on Advanced Communication and Intelligent Systems. Cham: Springer Nature Switzerland, 2023.
[5] Aljifri, Ahmed. "Predicting Customer Churn in a Subscription-Based E-Commerce Platform Using Machine Learning Techniques." (2024)
[6] Xiahou, Xiancheng, and Yoshio Harada. "B2C E-commerce customer churn prediction based on K-means and SVM." Journal of Theoretical and Applied Electronic Commerce Research 17.2 (2022): 458-475.
[7] Rupesh, Ayyapureddi Siva Sai, and Advin Manhar. "E-Business Churn Prediction Model Using Machine Learning." (2023).

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[8] Baghla, Seema, and Gaurav Gupta. "Performance evaluation of various classification techniques for customer churn prediction in e-commerce." Microprocessors and Microsystems 94 (2022): 104680.
[9] Raeisi, Shamim, and Hedieh Sajedi. "E-commerce customer churn prediction by gradient boosted trees." 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2020.
[10] Kim, Sulim, and Heeseok Lee. "Customer churn prediction in influencer commerce: An application of decision trees." Procedia Computer Science 199 (2022): 1332-1339.
[11] Lubis, Arif Ridho, et al. "Comparison of Model in Predicting Customer Churn Based on Users' habits on E-Commerce." 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE, 2022.