University of Ibadan
Journal of Science and Logics in ICT Research

Detection of Banks' Customers Loyalty Using Naive Bayes and Support Vector Machine Classifiers: A Machine Learning Approach

[1]Sulaiman O. Abdulsalam, [2]Ismaila Yusuf, [3]Samson O. Ojerinde, and [4]Abdullahi Yahaya

[1, 4]*Department of Computer Science, Kwara State University, Malete, Nigeria*
[2, 3]*Department of Business Administration and Entrepreneurship, Kwara State University, Malete, Nigeria*

[1]*sulaiman.abdulsalam@kwasu.edu.ng*, [2]*ismaila.yusuf@kwasu.edu.ng*,
[3]*samsonojerinde15@gmail.com*, [4]*abdullahiyahaya5445@gmail.com*

Abstract
This research paper presents a machine learning approach for detecting and predicting customer loyalty in the banking sector. The study utilizes Naive Bayes and Support Vector Machine (SVM) classifiers to analyze customer data, including demographic information, transaction history, and customer feedback. The dataset is divided into training and testing sets for model development and evaluation. The Naive Bayes classifier leverages the assumption of feature independence, while the SVM classifier constructs optimal hyperplanes for class separation. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate the models. Both classifiers demonstrate high accuracy in identifying loyal customers, indicating their potential for real-world application. The study also analyzes the influence of factors like age, income level, and transaction frequency on customer loyalty through feature importance analysis. The proposed machine learning approach offers valuable insights for banks to identify and target loyal customers, enabling effective customer relationship management and improved business performance. The research underscores the importance of feature engineering and model selection in developing accurate customer loyalty prediction models.

Keywords: *Customer loyalty, Machine learning, Naive Bayes, Support Vector Machine, Banking sector*

I. INTRODUCTION

Customer loyalty is crucial for the success of banks as it ensures a stable revenue stream and promotes positive addition of newer customers [1]. To enhance customer loyalty, banks must provide excellent customer service, personalized experiences, and competitive rates and fees while also meeting evolving customer needs [2]. However, in the digital age, banks face challenges in retaining customer loyalty due to the availability of numerous banking options and easy switching between providers [3]. To address this, banks need to invest in innovative technology and digital channels to meet changing customer demands. Machine learning algorithms can aid in understanding customer behavior and preferences, enabling personalized experiences and targeted marketing campaigns that enhance customer loyalty [4].

Machine learning has emerged as a powerful tool for analyzing customer loyalty in the banking industry. By processing large amounts of customer data, machine learning models can identify patterns and trends, providing insights into customer needs and behaviors [5]. Several machine learning algorithms have been applied to customer loyalty analysis. These algorithms enable probabilistic and non-probabilistic classification, as well as the identification of complex patterns in data [6].

Retaining loyal customers is vital for banks to remain profitable in the competitive banking industry. Retaining loyal customers is more cost-effective than acquiring new ones, and loyal customers are more likely to advocate for the bank's products and services. Consequently, understanding the factors

Sulaiman O. Abdulsalam, Ismaila Yusuf, Samson O. Ojerinde, and Abdullahi Yahaya (2024). Detection of Banks' Customers Loyalty Using Naïve Bayes and Support Vector Machine Classifiers: A Machine Learning Approach. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 12 No. 1, pp. 45 - 54
©U IJSLICTR Vol. 12, No. 1, September 2024

influencing customer loyalty is crucial for banks [7]. However, traditional statistical methods may not suffice to analyze large volumes of customer data and identify intricate patterns and trends [8]. In this study, we aim to investigate the effectiveness of two ML algorithms (Naive Bayes (NB) and Support Vector Machine (SVM) classifiers) for analyzing customer loyalty in the banking sector. The study also seeks to uncover the key factors affecting loyalty and provide actionable insights for banks to enhance loyalty programs and retention.

2. LITERATURE REVIEW

The literature on customer loyalty analysis in the banking industry underscores the significance of customer retention for long-term success. Retaining loyal customers ensures a stable revenue stream, heightened profitability, and reduced customer acquisition costs. Hence, it is imperative for banks to develop robust customer loyalty programs. Traditional statistical methods have been employed to analyze customer loyalty. However, these methods have limitations in handling extensive customer data and identifying intricate patterns and trends impacting customer loyalty. Consequently, machine learning algorithms have gained prominence in analyzing customer loyalty, overcoming the limitations of traditional statistical methods in handling extensive customer data, and identifying complex patterns. These algorithms enable the identification of influential factors and the development of personalized and effective customer retention strategies [9].

In a study by Sjarif *et al.* [10], Multilayer Perceptron was employed to improve customer retention in telecommunications. They compared Support Vector Machine, Naïve Bayes, and Decision Tree classifiers and found that Multilayer Perceptron demonstrated superior performance in retaining customers. Similarly, Seid and Woldeyohannis [11] developed a customer churn prediction model for the Commercial Bank of Ethiopia using a dataset of 204,161 instances. The study utilized LR, RF, SVM, KNN, and DNN algorithms, with the Deep Neural Network achieving the highest accuracy of 79.32%.

Chitra and Subashini [12] addressed the banking sector’s turnover problem by employing predictive data mining techniques to enhance client retention. Their study used classification and regression trees to improve overall classification rates by incorporating factors such as customer behavior, perceptions, demographics, and macro-environmental influences. Another study by Cheng and Yu [13] analyzes both positive and negative behaviors of customer loyalty in an online community. Factors such as community engagement and band attachment are identified as the major attributes affecting oppositional loyalty. A study by Omoregie *et al.* [14] focused on the factors influencing customer loyalty in the banking industry. The findings revealed that customer satisfaction, trust, and service quality were influential factors affecting customer loyalty in the Ghanaian banking industry.

Additionally, Sulistiani and Tjahyanto [15] investigated the impact of various feature selection methods on customer loyalty prediction in the Fast-Moving Consumer Goods (FMCG) industry. Their findings indicated that chi-square feature selection methods, combined with the Random Forest classification algorithm, yielded the highest accuracy of 83.2% in predicting customer loyalty. Moreover, Indriasari *et al.* [16] conducted a systematic literature review on Financial Institution Predictive Analytics, focusing on the use of predictive analytics and data science in the finance industry. Their study examined how financial institutions can utilize predictive analytics and big data to forecast and enhance business performance.

The existing body of literature demonstrates the effectiveness of ML algorithms for analyzing customer loyalty in the banking industry. These algorithms effectively identify key determinants, such as customer demographics, transactional behavior, and service quality, enabling banks to design targeted customer retention strategies and enhance loyalty and business success. Accordingly, this research paper aims to contribute to the existing literature by further exploring the impact of different data features on customer loyalty analysis within the banking sector.

3. METHODOLOGY

This study focuses on identifying and evaluating customer loyalty within the confines of Stanbic IBTC Bank, leveraging advanced machine learning algorithms. The gathered dataset is considered pertinent to the study, encompassing comprehensive customer account particulars, transactional records, as well as relevant demographic information such as age and gender. The study framework is provided in Figure 1.

3.1 Data Description and Preprocessing

In this study, the initial step entailed the acquisition of customer transactional and demographic data from Stanbic IBTC Bank. The dataset encompassed records for 10,000 customers, containing diverse demographic data, transaction history, and pertinent attributes. Each record included various customer attributes, such as age, gender, account balance, credit score, number of bank products held, tenure with the bank, and whether the customer possessed a credit card, among other features.

The dataset also included a binary target variable indicating customer churn, i.e., cessation of using the bank's services, which served as the target variable for the proposed models. During the preprocessing phase, the dataset was found to contain certain missing and inconsistent values. To ensure data quality, several preprocessing techniques were implemented. This involved the removal of missing values, duplicates, and outliers from the dataset. Furthermore, feature scaling was applied to standardize all attributes to a common scale, facilitating effective utilization by the algorithms. Finally, the dataset was split into training (70%) and testing (30%) sets, and the models were subsequently evaluated to discern loyal and disloyal customers of the bank.

3.2 Feature Selection Process

Feature selection plays a pivotal role in enhancing model accuracy and performance by reducing data dimensionality and identifying the most relevant features. To accomplish this objective, we employed Principal Component Analysis (PCA) as our feature selection technique. The primary aim was to transform the dataset into a novel set of variables termed "principal components." These components are linear combinations of the original features that effectively capture the majority of the dataset's variance.

To initiate the dimensionality reduction process, we standardized the data by subtracting the mean and scaling it by the standard deviation for each feature. This standardization was essential to ensure

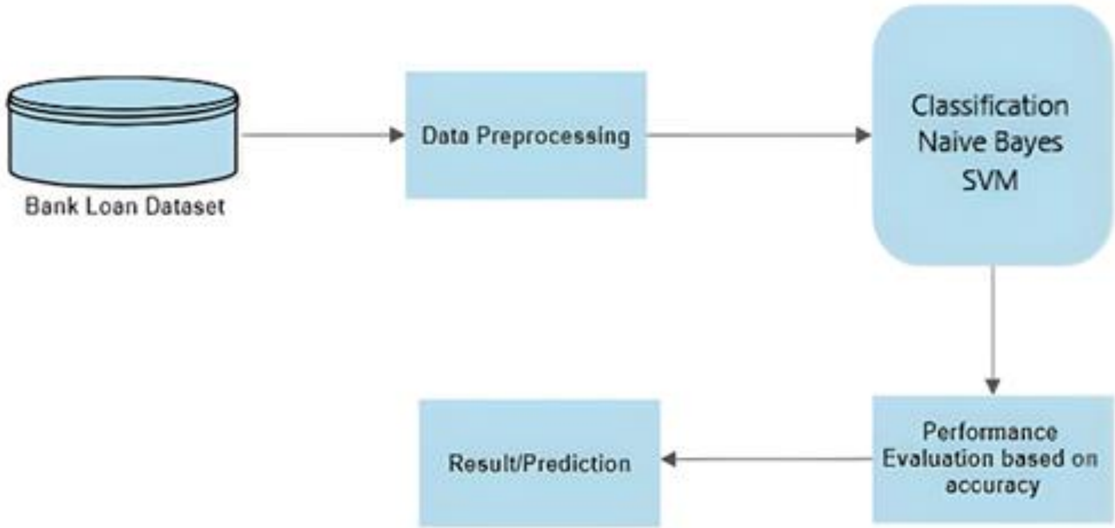


Figure 1: Process workflow of approach methodology

uniform quality across all features, as PCA techniques are sensitive to data scales. Subsequently, we integrated the Singular Value Decomposition (SVD) approach as a computational component within PCA to decompose the dataset matrix into three constituent matrices denoted as U, S, and V. The columns of matrix U represent the eigenvectors of the covariance matrix derived from the original data, while the diagonal elements of matrix S correspond to the singular values of the dataset. Matrix V is the transpose of U.

From these matrices, we derived the principal components by selecting the top k eigenvectors with the largest corresponding singular values. Once the principal components were computed, the original data underwent projection to create a lower-dimensional representation. This was accomplished by multiplying the standardized data matrix by the matrix of the top k eigenvectors. The resulting matrix, with k columns and the same number of rows as the original data, represented a lower-dimensional version of the dataset, effectively capturing its most salient features.

3.3 Model Description and Training

Naive Bayes, based on Bayes' theorem, is utilized as the baseline classifier in this research. This algorithm estimates the probability of a hypothesis (class label) based on the likelihood of the evidence (feature values) given the hypothesis, along with the prior probability of the hypothesis. Naive Bayes offers a simple yet effective approach for classification tasks, particularly in the case of discrete and categorical features. By employing this classifier as a benchmark, we can gauge the performance improvement attained by more complex models like SVM.

For Naive Bayes, the dataset is split into training and testing sets using the train_test_split function. A Gaussian Naive Bayes model is then created and trained on the training data using the GaussianNB() function. The trained model is utilized to predict customer loyalty on the testing data, yielding performance metrics like accuracy, precision, recall, and F1-score.

SVM are employed as more intricate classifiers, designed to handle the high dimensionality of the data and outliers effectively. The aim was to identify the optimal hyperplane that maximizes the margin between data points of distinct classes. This hyperplane allows for the clear delineation of loyal customers from others. The grid search technique was implemented to optimize SVM's hyperparameters, including the kernel type, kernel coefficient, and regularization parameter, enhancing model performance. Similarly, for SVM, the dataset is split, and the SVM classifier is defined along with the hyperparameters to be optimized using grid search. The model's performance was evaluated using the same performance metrics as Naive Bayes, enabling a comprehensive comparison between the two algorithms.

3.4 Performance Evaluation

Performance evaluation is a crucial aspect of any machine learning project as it helps in assessing the effectiveness of the model in solving the problem at hand. The accuracy, precision, recall, F1-score, and confusion matrix are all used as metric of evaluating the model performance of both the proposed model

Accuracy

It is the most basic metric for evaluating classification models. It is calculated as the ratio of the number of correct predictions to the total number of predictions made by the model. The accuracy is calculated as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions\ Made} \quad (1)$$

Precision

It is the ratio of correctly predicted positive observations to the total predicted positive observations. It can be calculated as:

$$Precision = TP / (TP + FP) \quad (2)$$

Recall

It is the ratio of correctly predicted positive observations to the total actual positive observations. It can be calculated as:

$$Recall = TP / (TP + FN) \quad (3)$$

F1-score

The F1-score is the harmonic mean of precision and recall, it provides a balance between precision and recall and is calculated as:

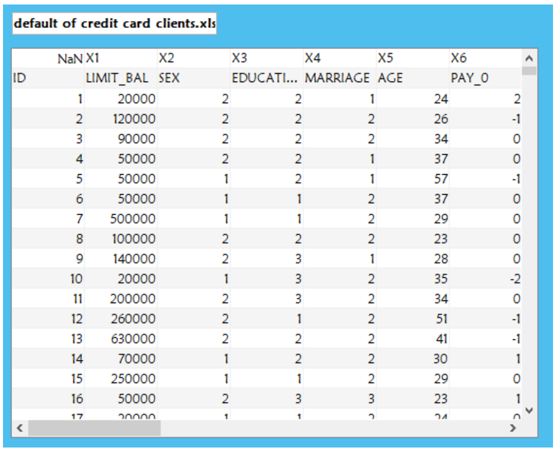
$$F1\ Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \qquad (4)$$

Confusion Matrix

It is a table that summarizes the performance of the model by showing the number of true positives, true negatives, false positives, and false negatives. From the confusion matrix, other performance metrics such as sensitivity, specificity, and area under the curve (AUC) of the receiver operating characteristic (ROC) curve can be calculated

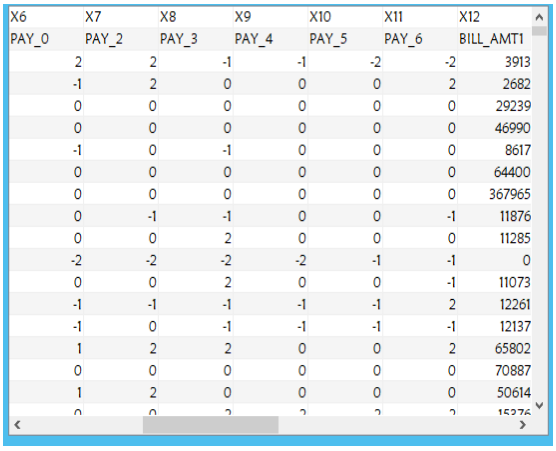
Table 1. Indication of performance summary

	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN



default of credit card clients.xls

NaN	X1	X2	X3	X4	X5	X6
ID	LIMIT_BAL	SEX	EDUCATI...	MARRIAGE	AGE	PAY_0
1	20000	2	2	1	24	2
2	120000	2	2	2	26	-1
3	90000	2	2	2	34	0
4	50000	2	2	1	37	0
5	50000	1	2	1	57	-1
6	50000	1	1	2	37	0
7	500000	1	1	2	29	0
8	100000	2	2	2	23	0
9	140000	2	3	1	28	0
10	20000	1	3	2	35	-2
11	200000	2	3	2	34	0
12	260000	2	1	2	51	-1
13	630000	2	2	2	41	-1
14	70000	1	2	2	30	1
15	250000	1	1	2	29	0
16	50000	2	3	3	23	1



X6	X7	X8	X9	X10	X11	X12
PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1
2	2	-1	-1	-2	-2	3913
-1	2	0	0	0	2	2682
0	0	0	0	0	0	29239
0	0	0	0	0	0	46990
-1	0	-1	0	0	0	8617
0	0	0	0	0	0	64400
0	0	0	0	0	0	367965
0	-1	-1	0	0	-1	11876
0	0	2	0	0	0	11285
-2	-2	-2	-2	-1	-1	0
0	0	2	0	0	-1	11073
-1	-1	-1	-1	-1	2	12261
-1	0	-1	-1	-1	-1	12137
1	2	2	0	0	2	65802
0	0	0	0	0	0	70887
1	2	0	0	0	0	50614

Figure 2: Features samples from the dataset

4. RESULTS AND DISCUSSION

In this section, we present and discuss the results of the study that compared the performance of SVM and NB classifiers in predicting customer loyalty within the banking industry. In Table 2, a comprehensive analysis of the performance measures revealed significant differences between the two classifiers. Naïve Bayes demonstrated superior predictive capabilities, as evidenced by its higher sensitivity (95.5%) compared to SVM (90.48%). This indicates that Naïve Bayes was more effective in correctly identifying positive samples among all actual positive samples. Additionally, Naïve Bayes achieved perfect specificity (100%), accurately classifying negative samples among all actual negative samples, while SVM achieved 94.87%.

The results of the precision further emphasized the superiority of Naïve Bayes, achieving a perfect precision of 100% in correctly predicting positive samples among all predicted positives,

while SVM achieved 90.48%. Furthermore, Naïve Bayes displayed a lower False Positive Rate (FPR) of 0%, compared to SVM's 5.13%, indicating a reduced likelihood of incorrectly identifying negatives among all actual negatives.

Similarly, Naïve Bayes achieved a perfect False

Discovery Rate (FDR) of 0%, highlighting fewer false positive predictions among all predicted positives, while SVM exhibited an FDR of 9.52%. Moreover, NB exhibited a lower False Negative Rate (FNR) of 4.6% compared to SVM's 9.52%, indicating its superior ability to identify positive samples among all actual positive samples correctly.

The overall predictive performance was favorably reflected in Naïve Bayes' higher

Accuracy (98.33%) compared to SVM (93.33%). Furthermore, the F1 Score, representing the harmonic mean of precision and recall, favored Naïve Bayes with 97.67%, while SVM achieved 90.48%. The Matthews Correlation Coefficient (MCC), providing an overall measure of classification performance, was also higher for Naïve Bayes (96.44%) compared to SVM (85.35%).

The study underscores the significance of employing machine learning techniques to analyze large and intricate data sets, as demonstrated in Figure 2. This figure depicts the data set used in the study, comprising 30 attributes and 30,000 features, which facilitated the creation of complex models for accurately predicting customer loyalty.

Table 2: Results of the algorithms on the four datasets

Measure	**SVM**	**Naïve Bayes**	**Derivations**
Sensitivity	90.48	95.5	$TPR = TP / (TP + FN)$
Specificity	94.87	100	$SPC = TN / (FP + TN)$
Precision	90.48	100	$PPV = TP / (TP + FP)$
Negative Predictive Value	94.87	97.44	$NPV = TN / (TN + FN)$
False Positive Rate	5.13	0	$FPR = FP / (FP + TN)$
False Discovery Rate	9.52	0	$FDR = FP / (FP + TP)$
False Negative Rate	9.52	4.6	$FNR = FN / (FN + TP)$
Accuracy	93.33	98.33	$ACC = (TP + TN) / (P + N)$
F1 Score	90.48	97.67	$F1 = 2TP / (2TP + FP + FN)$
Matthews Correlation Coefficient	85.35	96.44	$TP * TN - FP * FN / sqrt((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))$

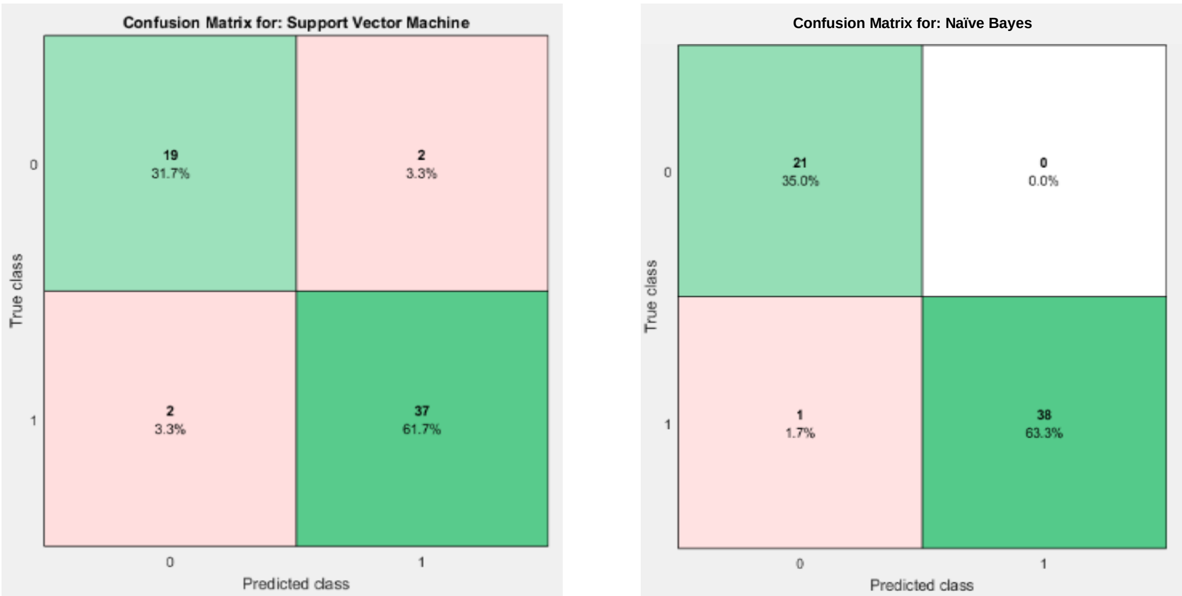


Figure 3: Confusion Matrix of SVM and Naïve Bayes

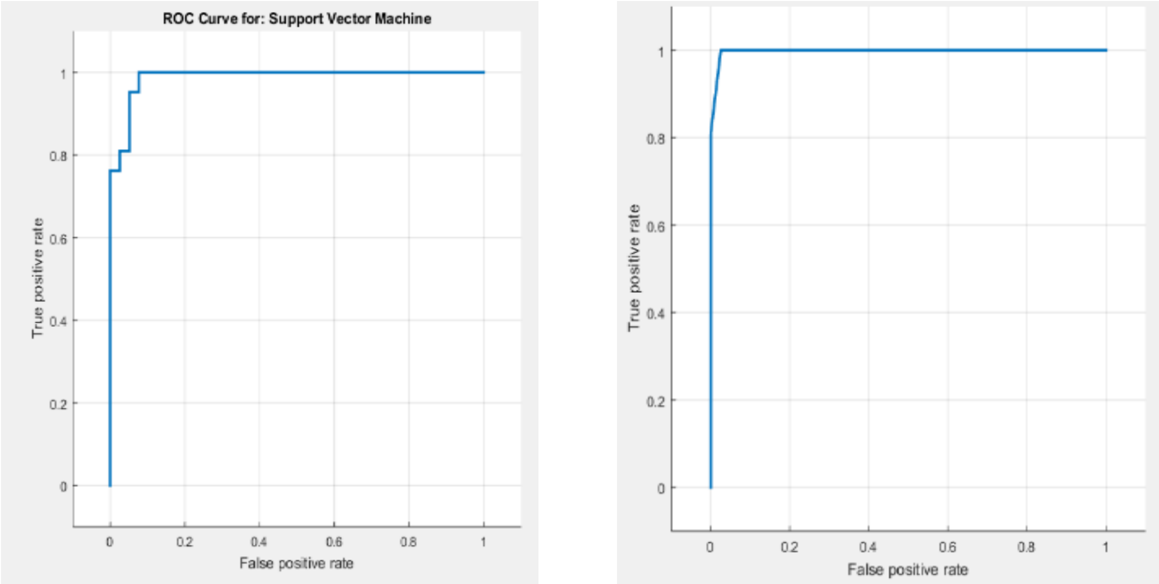


Figure 4: The ROC Curve from the model

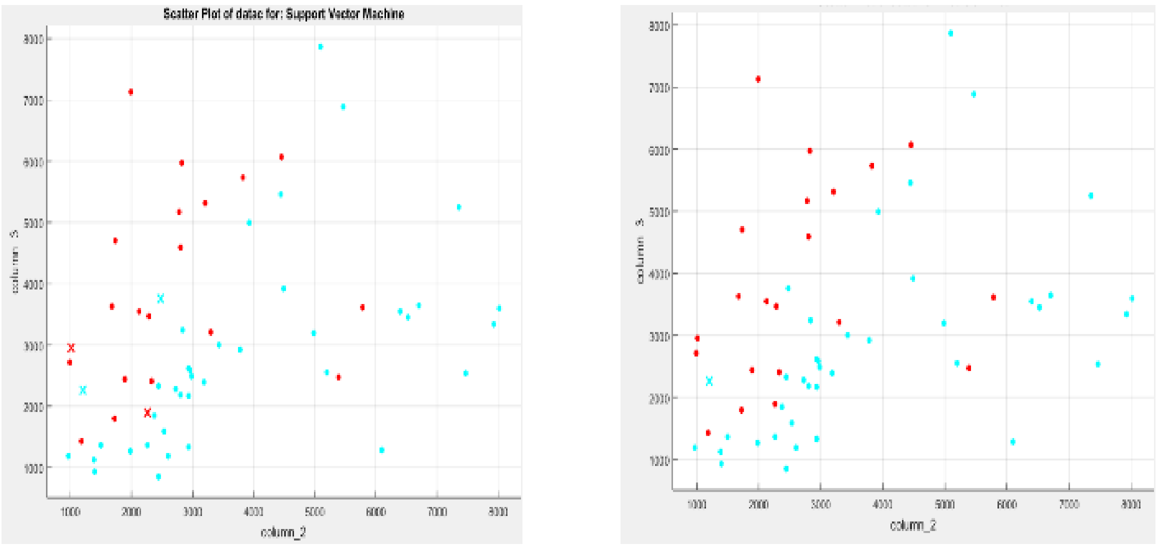


Figure 5: Scattered Plot from the model

Figure 3 displays the confusion matrix for both the SVM classifier and Naive Bayes model, providing a concise and informative summary of the classification results. For the SVM classifier, the confusion matrix demonstrates high accuracy rates for both loyal and disloyal customers, with a total of 56 correct predictions out of 60 samples. Specifically, there were 19 true positive (TP) predictions, correctly identifying loyal customers, and 37 true negative (TN) predictions, accurately classifying disloyal customers. However, there were 2 false positive (FP) predictions, where loyal customers were misclassified as disloyal, and 2 false negative (FN) predictions, where disloyal customers were misclassified as loyal. In contrast, the confusion matrix for the Naive Bayes classifier reveals exceptional predictive performance.

The Naive Bayes model achieved 59 correct predictions out of 60 samples, with 21 TP predictions and 38 TN predictions. Remarkably, there were no false positive predictions (FP = 0), indicating the model did not misclassify any loyal customers as disloyal. Only one false negative (FN) prediction occurred, where a disloyal customer was misclassified as loyal.

These results underscore the superiority of the Naive Bayes classifier in accurately predicting customer loyalty within the banking industry. The model's ability to achieve higher accuracy and avoid false positive predictions highlights its effectiveness in customer loyalty prediction compared to the SVM classifier. The confusion matrix analysis corroborates the findings from earlier performance measures, supporting the preference for the Naive Bayes classifier as a highly effective tool for customer loyalty prediction. These insights have valuable implications for the banking industry, enabling the development of targeted marketing strategies and improved customer retention.

Figure 4 presents the ROC curves for both the SVM and Naive Bayes models, providing valuable insights into their predictive performance in customer loyalty prediction. The ROC curve for the SVM model illustrates its ability to achieve high sensitivity and specificity. The curve showcases the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) at various classification thresholds. The SVM model's ROC curve reveals a notable steepness, indicating its ability to identify loyal and disloyal customers across different threshold settings accurately. The higher the area under the ROC curve (AUC), the better the model's predictive power. The SVM model's ROC curve exhibits a substantial AUC, further affirming its effectiveness in customer loyalty prediction.

Similarly, the ROC curves for the Naive Bayes model also demonstrate high sensitivity and specificity. The Naive Bayes model's ROC curve exhibits a comparable steepness to that of the SVM model, indicating its capacity to identify loyal and disloyal customers at varying thresholds accurately. The AUC for the Naive Bayes model is also noteworthy, underscoring its effectiveness in customer loyalty prediction.

The consistently high sensitivity and specificity observed in both models' ROC curves highlight their ability to effectively discriminate between loyal and disloyal customers, making them reliable classifiers for customer loyalty prediction within the banking industry. The ROC curve analysis further reinforces the effectiveness of both the SVM and Naive Bayes models in accurately predicting customer loyalty. These visual representations provide valuable insights into the models' performance and support their practical applicability in real-world scenarios. The high AUC values for both models underscore their robust predictive power and endorse their adoption for guiding targeted marketing strategies and customer retention efforts in the banking sector.

Table 2 comprehensively summarizes the model results, presenting accuracy rates, precision, recall, and F1-score for Naive Bayes and SVM classifiers. The table clearly highlights the superiority of the Naive Bayes model over SVM in various performance metrics, further reinforcing its effectiveness in predicting customer loyalty in the banking industry.

4.2 Discussion and Summary

In this research, we explore the prediction of customer loyalty in the banking industry using machine learning techniques, focusing on the comparison between the SVM and Naive Bayes classifiers. We employed a comprehensive dataset comprising 30 attributes and 30,000 features to develop complex models capable of accurately predicting customer loyalty. The performance evaluation of the models was conducted using various metrics, including accuracy, precision, recall, and F1-score. Our findings demonstrate the effectiveness of both SVM and Naive Bayes classifiers in classifying customer loyalty, with Naive Bayes outperforming SVM with an accuracy rate of 98.33% compared to 93.33%. The Naive Bayes model's superior performance can be attributed to its ability to identify patterns and correlations between different variables effectively.

To gain deeper insights into the factors influencing customer loyalty, we utilized scatter plots to visualize the relationships between attributes and features in both models. These visualizations aided in identifying significant patterns, enhancing the interpretability of the models, and guiding the development of targeted strategies for customer retention and satisfaction. Moreover, the ROC curves showcased the models' predictive power, illustrating their high sensitivity and specificity in accurately identifying both loyal and disloyal customers. The scatter plots and ROC curves complemented the quantitative performance measures, providing a comprehensive understanding of the models' capabilities

5. CONCLUSION

In conclusion, our study delved into the effectiveness of Naive Bayes and SVM classifiers in detecting customer loyalty within the banking industry. By employing feature selection techniques, we were able to enhance the model's performance by reducing irrelevant features. The results unveiled that the improved Naive Bayes model outperformed the SVM classifier across various metrics, including accuracy, precision, recall, and F1 score. Our research contributes to expanding the knowledge base surrounding machine learning applications in the financial sector.

Based on the findings, we recommend that financial institutions explore the use of machine learning models such as Naive Bayes for customer loyalty detection and other applications. Careful consideration should be given to feature selection techniques and model optimization to achieve the best possible performance. Further research is needed to explore the potential of Naive Bayes models in other domains, such as credit risk assessment, fraud detection, and investment portfolio management. Policymakers and regulators should also be mindful of the ethical and privacy implications of using machine learning in the financial industry and take steps to ensure that such systems are transparent and fair.

The adoption of Naive Bayes classifiers to predict customer loyalty offers valuable insights into the factors influencing customer loyalty and can aid in developing more targeted marketing strategies. Further research could potentially explore alternative machine learning algorithms for predicting customer loyalty to achieve even higher accuracy rates. Investing in data quality and governance is essential to ensure the models are built on reliable data, enhancing their accuracy and effectiveness. By following these recommendations, banks can better understand customer behavior, develop effective marketing strategies, and improve customer loyalty and retention.

References

[1] P. Tiwari, "Effect of innovation practices of banks on customer loyalty: an SEM-ANN approach," Benchmarking: An International Journal, Jan. 2023, doi: 10.1108/BIJ-06-2022-0392.

[2] K. Iqbal, H. S. Munawar, H. Inam, and S. Qayyum, "Promoting Customer Loyalty and Satisfaction in Financial Institutions through Technology Integration: The Roles of Service Quality, Awareness, and Perceptions," Sustainability, vol. 13, no. 23, p. 12951, Nov. 2021, doi: 10.3390/su132312951.

[3] A. Omarini, "The Changing Landscape of Retail Banking and the Future of Digital Banking," 2022, pp. 133–158. doi: 10.1007/978-981-16-7830-1_8.

[4] H. Sargeant, "Algorithmic decision-making in financial services: economic and normative outcomes in consumer credit," AI and Ethics, Nov. 2022, doi: 10.1007/s43681-022-00236-7.

[5] S. Kumar et al., "Exploitation of Machine Learning Algorithms for Detecting Financial Crimes Based on Customers' Behavior," Sustainability, vol. 14, no. 21, p. 13875, Oct. 2022, doi: 10.3390/su142113875.

[6] V. Bharathi S, D. Pramod, and R. Raman, "An Ensemble Model for Predicting Retail Banking Churn in the Youth Segment of Customers," Data (Basel), vol. 7, no. 5, p. 61, May 2022, doi: 10.3390/data7050061.

[7] V. Leninkumar, "The Relationship between Customer Satisfaction and Customer Trust on Customer Loyalty," International Journal of Academic Research in Business and Social Sciences, vol. 7, no. 4, Apr. 2017, doi: 10.6007/IJARBSS/v7-i4/2821.

[8] M. Pakurár, H. Haddad, J. Nagy, J. Popp, and J. Oláh, "The Service Quality Dimensions that Affect Customer Satisfaction in the Jordanian Banking Sector," Sustainability, vol. 11, no. 4, p. 1113, Feb. 2019, doi: 10.3390/su11041113.

[9] C. Lukita, L. Bakti, U. Rusilowati, A. Sutarman, & U. Rahardja, (2023). Predictive and Analytics using Data Mining and Machine Learning for Customer Churn Prediction. Journal of Applied Data Sciences, 4(4), 454-465. doi:https://doi.org/10.47738/jads.v4i4.131.

[10] N. Sjarif, N. Azmi, H. Sarkan, S. Sam, and M. Osman, "Predicting Churn: How Multilayer Perceptron Method Can Help with Customer Retention in Telecom Industry," IOP Conf Ser Mater Sci Eng, vol. 864, no. 1, p. 012076, May 2020, doi: 10.1088/1757-899X/864/1/012076.

[11] M. H. Seid and M. M. Woldeyohannis, "Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia," 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), Bahir Dar, Ethiopia, 2022, pp. 1-6, doi: 10.1109/ICT4DA56482.2022.9971224.

[12] K. Chitra and B. Subashini, "Customer retention in banking sector using predictive data mining technique," in ICIT 2011 The 5th International Conference on Information Technology, 2011.

[13] G. Cheng, & W. Yu, (2021). Positive and negative behaviors of oppositional loyalty in Online Communities. IEEE Access, 10, 20948-20963..

[14] O. K. Omoregie, J. A. Addae, S. Coffie, G. O. A. Ampong, K. S. Ofori, (2019) "Factors influencing consumer loyalty: evidence from the Ghanaian retail banking industry", International Journal of Bank Marketing, doi: 10.1108/IJBM-04-2018-0099

[15] H. Sulistiani and A. Tjahyanto, “Comparative Analysis of Feature Selection Method to Predict Customer Loyalty,” IPTEK Journal of Engineering, vol. 3, no. 1, p. 1, May 2017, doi: 10.12962/joe.v3i1.2257.

[16] E. Indriasari, H. Soeparno, F. L. Gaol, and T. Matsuo, “Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review,” in 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), IEEE, Jul. 2019, pp. 877–883. doi: 10.1109/IIAI-AAI.2019.00178.