International Journal of AI, BigData, Computational and Management Studies
Noble Scholar Research Group | Volume 3, Issue 1PP 46-55, January 2022
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V3I1P107

Evaluating Machine Learning Models Efficiency with Performance Metrics for Customer Churn Forecast in Finance Markets

Anand Polamarasetti[1], Rahul Vadisetty[2], Srikanth Reddy Vangala[3], Purna Chandra Rao Chinta[4], KishanKumar Routhu[5], Vasu Velaga[6], Chethan Sriharsha Moore[7], Suneel Babu Boppana[8]
[1]MCA, Andhra University, India.
[2]Master of Science, Wayne State University.
[3]Dept. of Computer Science, University of Bridgeport
[4]Microsoft, Support Escalation Engineer,
[5]AT&T, MGM Systems, Senior Deployment Engineer.
[6]Cintas Corporation, SAP Functional Analyst.
[7]Papa Johns , Sr Devops Engineer.
[8]Site Technologies, Project Manager.

Abstract: In financial markets, the competition is on the rise, and therefore, prediction of customer churn is necessary to minimize the revenue loss. Good machine learning models can tell who will be a potential churner, so these institutions can improve their retention strategies and do everything to keep customer loyalty. In financial markets, forecasting customer turnover is a significant issue because attracting prospective consumers may be more cost-effective than keeping current ones. Nevertheless, with the rise of rival financial service providers, it is crucial for institutions to have a reliable way to forecast customer turnover so they may implement proactive retention strategies and prevent significant revenue losses. This research looks at how well Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machines (SVM) are three machine learning models that categorize the clients as either churners or non-churners. The mathematical framework is assessed using measurements listed below the region under the inclination (AUC ROC), F1 score, accuracy, precision, and recall. When compared to the other classifiers, RF performed the best with a 92.5% accuracy rate, 91.8% precision, 90.3% recall, 91.0% F1-score, and 0.95 AUC-ROC score. Results show that RF is an effective approach for dealing with complicated, big datasets. The study concludes that financial institutions may benefit from machine learning's enhanced prediction accuracy and its ability to assist them in identifying high-risk consumers, allowing them to implement effective customer retention tactics.

Keywords: Customer Churn Prediction, Machine Learning, Random Forest, SMOTE, Financial Markets, IBM Telco Dataset.

1. Introduction

In ML has emerged as a crucial area of research for forecasting the loss of customers in the financial services sector and various other industries. However, traditional prediction models come with the problem of scalability issues and low flexibility to adapt to changes in customer behaviors [1]. However, a solution that largely augments the accuracy of prediction using data is provided by modern ML techniques. Various approaches, including MLP, SVM, and RF, have demonstrated success in churn prediction [1], though their effectiveness heavily depends on careful parameter configuration through experimental supervised learning. These cutting-edge methods help companies make better judgements and put effective client retention plans into place [2], taking into account the shortcomings of traditional methods. Upholding a high level of client loyalty is a frequent problem in business. It has been demonstrated by several research studies that keeping current consumers is more lucrative than discovering unfamiliar ones [3]. Relationship management for customers focuses on churn prediction or, conversely, loyalty[4]. The majority of earlier research has been done on industries where contracts are in place with consumers, which lowers the turnover rate [5]. An ML technique may effectively forecast client attrition, according to several research [6].

ML is still not well- According to various definitions found in modern literature, it is generally understood to be a process in which a system interacts with its surroundings in a way that modifies the system's structure[7], and that modifications to the structure have an impact on the interaction process itself. This term, which was applied to the idea of neural networks, has been condensed and modified[8]. This high-level theorem contains three primary learning paradigms, each of which has distinct financial time series prediction application areas. ML models function well when sufficient effort is made to engineer the model's characteristics[9]. In this case, each model has to specify the pertinent elements in light of the current issue. Traditional methods of identifying behavioral patterns are unsuccessful due to the overwhelming amount of data and its unstructured format. Finding out which clients have left is the goal of the prediction model. The industry uses a targeted strategy to ascertain which customers are probably going to depart. After that, the industry approaches those clients or consumers and offers them exclusive programs, incentives, and opportunities. The prediction model's objective is to determine which clients have departed. Using a focused approach, the industry seeks in order to determine which clients have the greatest likely to depart[10].

1.1 Motivation and Contribution of Work

This work is motivated by the need for early and accurate customer churn prediction in financial markets, as traditional methods are often time-consuming, prone to errors, and struggle to scale with large datasets. This study intends to automate and improve the churn identification process by utilizing cutting-edge ML models like RF, XG Boost, and CNN, allowing financial institutions to proactively identify at-risk clients. This approach gives the businesses a means of implementing effective retention strategies and minimizing financial losses as well as increasing the happiness with their customers. This project aims to use automated ML approaches to categorize and forecast customer turnover in the customer churn dataset from IBM Telco. The contribution of this study is explained below.

- The study integrates sophisticated data pretreatment methods, such as handling missing values, eliminating duplicates, handling outliers, resolving inconsistencies, and enhancing the quality and dependability of the data.
- Data normalization with Min-Max scaling is used to standardize feature values, which makes data uniform across attributes and improves the model stability.
- The SMOTE is used to create a balanced dataset and, hence, enhance the model's performance in an effort to address the problem of class disparity.
- The RF classifier is used because of its ensemble learning capabilities, which increase the financial markets' churn prediction's precision and generalizability.
- The study then evaluates the suggested RF model's performance against that of other classifiers, such as SVM, DT, and LR, and demonstrates showing the forecast accuracy is higher for the RF model.
- A thorough evaluation framework that ensures a reliable assessment of the ROC curve, F1 score, accuracy, precision, and recall all contribute to the mockup.

1.2 Novelty and Justification

This new research is structured with a thorough method for predicting client attrition in financial markets by combining sophisticated pre-processing techniques, SMOTE to address class imbalance, and the potent RF classifier. This work employs SMOTE to improve minority class representation and, hence, increase prediction accuracy, in contrast to existing models that typically cannot manage unbalanced data sets. Furthermore, several approaches to assessment allow for a thorough examination of the model's functionality. Criteria including accuracy, precision, recall, and F1-score, among others. Comparative analysis further justifies the choice of the RF model, as it outperforms conventional approaches such as SVM, DT, and LR. The suggested method's robustness is demonstrated by its high accuracy of 86.94% and AUC score of 0.95. This study offers financial institutions a dependable and scalable methodology for optimizing client retention tactics and making data-driven decisions that improve company sustainability by putting in place a clearly defined ML pipeline and confirming outcomes using actual telecom data.

1.3 Structure of the paper

The paper is organized as follows: Section II examines earlier research on predicting customer attrition and related topics. The approach utilized to compile the data for this study is designated in Section III. The findings and analysis of the purchaser churn forecast are shown in Section IV. Section V accomplishes by outlining the findings and probable boulevards for supplementary investigation.

2. Literature Review

This section covers the literature reviews from earlier research on ML-based forecast of consumer attrition. Additionally, Table 1 summarizes the literature reviews that are covered here:

Amuda and Adeyemo (2019) created a model to do away with the requirement for human feature engineering at the data's initial assessment phase. The information used for the research was taken from the information system of one of the nations of Nigeria top finance organizations. Additionally, the perceptron with multiple layers paradigm was constructed using two overfitting algorithms and was built in Python. The Python implementation and in the Neuro The panacea Infinity software, a different model, was compared. The outcomes demonstrated that the effectiveness of the ANN programming language was on par with that of the Neuro Solution Infinity software. The ROC curve graphs are 0.89 and 0.85, and the accuracy rates are 97.53% and 97.4%, accordingly[11].

Li et al. (2018) the ML and template detection components make up the suggested technique. Using summary characteristics, the ML module classifies people using the SVM technique, which is based on supervised learning. To screen suspicious users that were produced in the ML module, the template detection module uses an FSM based on fraud user behaviors. Fraud users can be identified following the two components. They put their methodology into practice and test it on an actual dataset. The tests demonstrate that the approach may attain a high 93.56% detection accuracy[12].

Mishra and Reddy (2018) the telecom industry employed ensemble-based classifiers for churn prediction, namely RF, Boosting, and Bagging. Several popular classifiers, including DT, NB, and SVM, were compared to ensemble-based classifiers. The experimental results show that RF has a greater accuracy of 91.66%, a reduced error rate, poor specificity, and high sensitivity when compared to alternative approaches[13].

Mishra and Reddy (2018) proposed that a DL is superior to standard ML techniques because it can handle ever-increasing data quantities, find hidden patterns, detect patterns and underlying hazards, and more accurately notify the telecom industry about client behaviors. CNNs are used in this article to apply DL for churn prediction, and the results demonstrated high accuracy. According to the experimental findings, the churn prediction out predictive model has a 92.06 F-score, 86.85% accuracy, 13.15% error rate, 91.08 precision, and 93.08% recall[14].

Agrawal et al. (2018) This study utilizes a DL approach to forecast customer churn on a dataset from Telco. In order to develop a non-linear categorization model, a multi-layered neural network was used. That takes contextual, use, customer, and support features into account. The model predicts the likelihood of churn and the factors that influence it. After applying final weighting to these characteristics, the model that was educated predicts the probability of the consumer attrition. An 80.03% accuracy rate was attained[15].

Bharadwaj et al. (2018) the largest obstacle to maintaining a telecommunications network is churn, which hinders the expansion of lucrative clients. Two models have good accuracy rates in predicting client attrition. A parabola is used as the function of activation in the first model, an LR model, which is a non-linear classifier. The model's accuracy on their test dataset is 87.52% after regularizing it using a regularizing value of 0.01. The second model is an advanced MLP Neuronal Network. It has three hidden layers, a normalized involvement feature trajectory, and a loss function based on binary cross entropy. The learning rate is 0.01. When divided into a test-train set, this model attains a 94.19% accuracy rate[16].

Table 1 provides a comparative analysis of each study's methodology, dataset, performance metrics, and the aspects requiring additional investigation or exhibiting limitations.

Table 1: Summary of Literature Review Based on Customer Churn Prediction Using Machine Learning

Ref.	Methodology	Dataset	Performance	Limitations	Future Work
Amuda and Adeyemo, (2019)	Multi-layer Perceptron (MLP) with overfitting techniques	Financial institution database (Nigeria)	Accuracy: 97.53% (Python), 97.4% (Neuro Solution), ROC: 0.89, 0.85	Limited dataset from a single institution	Expanding the dataset to multiple institutions for better generalization
Li et al. (2018)	SVM-based supervised learning with FSM for fraud detection	Real-world dataset	Accuracy: 93.56%	Limited evaluation on a single dataset	Extending the approach to other fraud detection domains
Mishra and Reddy, (2018)	Ensemble methods (Bagging, Boosting, Random Forest) for churn prediction	Telecom dataset	Accuracy: 91.66% (Random Forest)	Limited comparison with other advanced ML models	Incorporating deep learning for improved prediction
Mishra and Reddy, (2018)	Deep Learning (CNN) for Churn Prediction	Telecom dataset	Accuracy: 86.85%, Precision: 91.08%, Recall: 93.08%, F1-score: 92.06	Computationally expensive, requires large data	Exploring hybrid deep learning models for better efficiency
Agrawal et al. (2018)	Multi-layer Neural Network for Churn Prediction	Telco dataset	Accuracy: 80.03%	Moderate accuracy compared to other methods	Enhancing feature engineering and model tuning
Bharadwaj et al. (2018)	Logistic Regression and MLP Neural Network	Telecom dataset	Accuracy: 87.52% (Logistic Regression), 94.19% (MLP)	Logistic regression has lower accuracy	Using advanced optimization techniques for model improvement

3. Methodology

This study's goal is to assess how well ML models predict client attrition in the financial markets. To guarantee data quality and model effectiveness, the churn prediction approach employing the IBM Telco Churn dataset adheres to a standardized workflow. To improve data dependability, pre-processing is done first, which includes addressing outliers, missing values, inconsistencies, and noise reduction. Several minority class instances are improved, and class imbalance is lessened by using the SMOTE. In order to standardize dies variables, feature scales are then normalized using data normalization. The dataset is going to be separated across training (75%) and testing (25%) sets for the determination of developing and evaluating the model. An RF classifier is used to construct Churn Prediction because it offers ensemble learning, which raises the precision and generalizability of predictive models. Lastly, A comprehensive evaluation matrix is utilized to gauge the model's efficacy, which incorporates precision, accuracy, recall, and F1-score. The model's capability to envisage customer attrition and support business decision-making is evaluated through an analysis of the final outcomes. The flowchart in Figure 1 illustrates how an ML model predicts client attrition in financial markets:

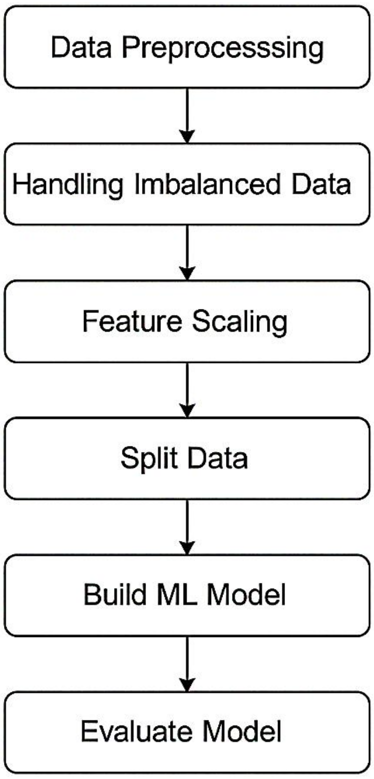

Figure 1: Flow Chart for Customer Churn Prediction in Financial Markets

A brief explanation of each step in this flowchart is provided below:

3.1 Dataset Collection and Visualization

The IBM Telco Customer Churn dataset, which includes 7,044 entries and 21 characteristics pertaining to a telecommunications company's customers, is used in this study. These attributes help establish a clear relationship between customer behavior and churn. Researches have made heavy use of the Kaggle dataset to forecast churn. Figure 2 displays the share of churned consumers relative with the ones that remain

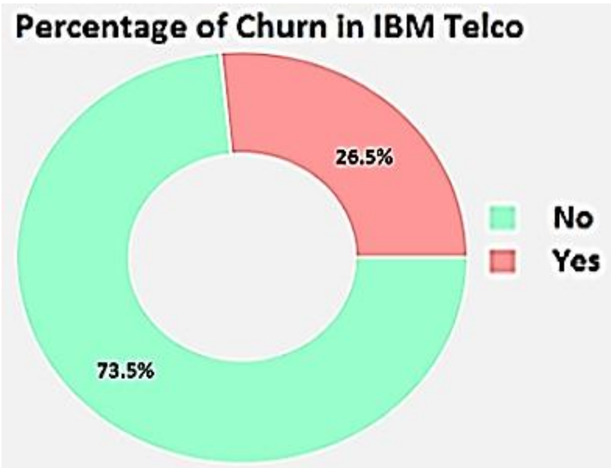

Figure 2: Percentage of Churn rate in IBM Telco dataset

Figure 2 shows a donut chart showing the percentage of customer churn in the IBM Telco dataset. 26.5% of the company's clients are churned, whereas 73.5% of the non-churning consumers are still there. This shows that while a sizable percentage of customers are still being kept, churn is still a major problem, highlighting the necessity of efficient prediction models and retention tactics.

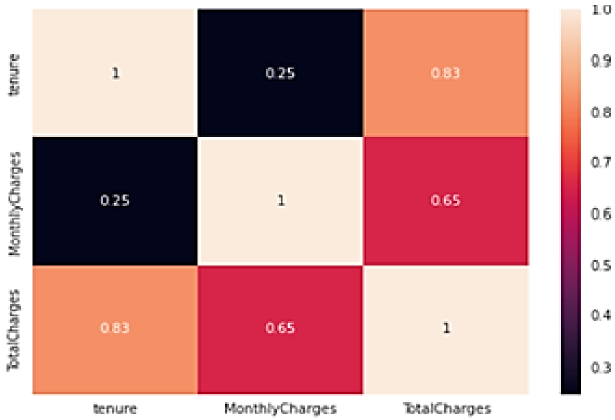

Figure 3: Correlation Heatmap of IBM Telco Dataset

Figure 3 represents a correlation heatmap showing the relationships between tenure, monthly charges, and total charges in a dataset, likely related to telecom customer behavior. A complete positive correlation is represented by a correlation value of 1, which ranges from 0 to 1. According to the heatmap, Clients with longer tenure frequently have higher overall expenses since there is a substantial favourable (0.83) connection that exists between the overall cost and employment. While there is a weaker correlation (0.25) between tenure and monthly charges, there is a significant association between recurring costs and overall charges correlation (0.65), indicating that monthly charges differ independent of tenure duration.

3.2 Data Preprocessing

Preparing and transforming data while also attempting to increase the efficiency of knowledge discovery is the goal of data preprocessing, one of the most important data mining procedures. Several methods, like as cleaning, integration, transformation, and reduction, are used in preprocessing[17]. At this point, the dataset's customer churn forecast is pre-processed by managing missing values, eliminating duplicates and inconsistent values, and using outlier treatment strategies, which are covered in more detail below.

- Handling Missing Value: A "missing value" is a dataset variable that does not include any data points. Empty cells, null values, or specialized symbols like "NA" or "unknown" are a few ways they can be written. The missing value in the dataset lists the columns that include the missing values after determining the number of missing values.
- Removing duplicates: Data is gathered through the combination of many data sources or by data scraping. It is true that duplicate data will provide analytical results and poor business analytics.
- Inconsistent values: When the same value—whether category or numeric—is represented in several ways within datasets, inconsistent values arise.
- Handling Outliers: An exception to the norm is a piece of data that is distinct from other data. general population. Outliers are present in all real-world datasets, and managing them is one of the several data cleaning techniques used. Outliers also affect the accuracy of corporate analytics and statistics.

3.3 Data Normalization

Normalization is a scaling procedure that entails rescaling and adjusting values till they are between zero and one. An alternative term for it is min-max scalability. It defines the Equation (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

If the lowest number in the column is X, then both the numerator and X' will be zero. The numerator and denominator are identical, though therefore, if X is the highest integer in the pillar, then X' is equal to 1. If X falls between the lowest and greatest values, then X' is between 0 and 1.

3.4 The Synthetic Minority Over-Sampling Technique (SMOTE)

Studies on customer data sets have found an unbalanced class distribution for customer turnover research. The sample size of churn customers is far less than that of non-churn customers, which might lead to the following scenario the classification accuracy is good, but the prediction accuracy for churn customers is poor. In this study, the most popular method for solving this issue is to employ Smote tactics. In ML, SMOTE is a useful method for dealing with data imbalance. It creates artificial specimens for the minority group by combining presently existent examples. It contributes to improving the model's functionality and maintaining a balanced class distribution. SMOTE is essential for handling data imbalance since it keeps the model from favoring the majority class. Furthermore, it can enhance the accuracy and performance of the model, e It may improve the model's effectiveness and accuracy especially for under-represented classes in the data. Equation (2) generates a fresh sample x' from A vector representation of a particular minority information point x by randomly choosing a trajectory x_k from its k adjacent nationals. The credit card detection database used in this work[18] is balanced using SMOTE.

$$x' = x + rand(0,1) \times (x - x_k) \qquad (2)$$

3.5 Data Splitting

The dataset is divided into two subset dataset ratios is 75:25 for training and testing, a first subset of 75 % dataset for training and second subset part of 25 % dataset for testing.

3.6 Implementation of Proposed Random Forest Model

The RF is a popular ML model for data classification. Several industries, including marketing, customer service, and investment, often use this method[19]. The RF is supported by a clump of trees. A total of the standard worth of the forecast is added to it[20]. It is produced at individual tree terminals and lessens a single tree's lack of robustness. The conceptual framework of the RF design is shown in Figure 4:

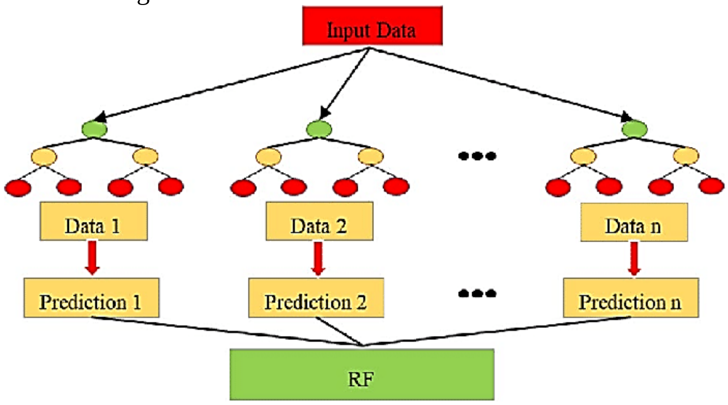

Figure 4: Architecture of a Random Forest (RF) Model

Each tree is created using a subset of input variables that are selected at random. This is one way to express the estimated model:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} g_k(x) \qquad (3)$$

In Equation (3), a collection of random trees with kth learners is represented by g(x), where x is the input characteristic vector. The RF standard deviation of the results is the final estimated value from every tree. As a result of such bulks, respectively tree has an impact on the RF estimation. The RF model is superior to other ML methods. The consistency of the former in terms of automatically assembling subsets of information for training and building trees employing random procedures is what causes this[21]. Additionally, since the RF model trains given a randomized sample of information selected at randomness using bootstrapping, the degree of overfitting is kept to a minimum.

3.7 Performance Matrix

The evaluation of ML classifiers relies on multiple performance metrics to ensure accurate and meaning comparisons. The following metrics for performance have been used to evaluate the proposed model[22]. For comparison, Recall, Precision, Accuracy, and F1-score were assessed. The dataset is also visualized using ROC curves, offering a visual representation of training progress and generalization ability[23]. A confusion matrix with four distinct combinations of expected and actual values. Usually, a matrix of disorientation is used to enable algorithm execution to be visualized.

3.7.1 Accuracy

Accuracy determines the proportion of clients who are properly or not churned, as Equation (4) shows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (4)$$

3.7.2 Precision

The exactness Precision is defined as TP divided by the sum of TP and FP, it is given in Equation (5):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (5)$$

3.7.3 F1-Score

It is only when the F1-score reaches 1, both "precision and recall" must be 1. Additionally, since precision and recall are similarly high, the F1 score is raised, according to the F1-score, which is the harmonious average of accuracy and recall, matching Equation (6).

$$\text{F1} - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}} \qquad (6)$$

3.7.4 Recall

Recall measures how many FN are added to a mishmash of predictions. Additionally, recall, also referred to as sensitivity or true positive rate, which serves is computed employing the method in Equation (7).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (7)$$

3.7.5 ROC

A classification model's ROC curve provides a visual representation of its performance. The y-axis in the ROC space displays the TPR, whereas the x-axis displays the FPR.

4. Result Analysis and Discussion

The churn prediction system features a high-performance setup with 16GB of RAM, a CPU with several cores (such as an AMD Ryzen 7 or an Intel Core i7), and an NVIDIA RTX 3060 GPU for faster training. Python, Scikit-learn, TensorFlow, and Pandas are used for data processing and model construction. The RF model demonstrated good performance with 86.94% accuracy, 86.01% precision, 87.18% recall, and an 88.39% F1-score using Table 2 of the IBM TELCO CHURN dataset. The ROC curve's outstanding discriminating ability is confirmed by its AUC of 0.95.

Table 2: Performance of Customer Churn Prediction in Financial Markets Using Random Forest (RF) Models

Measure	RF
Accuracy	86.94
Precision	86.01
Recall	87.18
F1-score	88.39

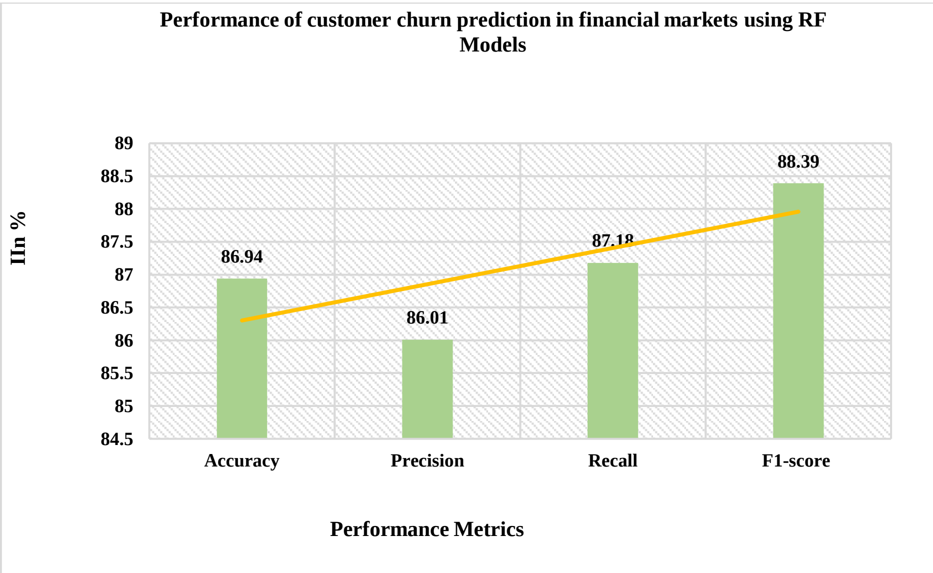

Figure 5: Bar Graph of RF Model For Customer Churn Prediction Using ML Models

The performance of the Forecast of customer attrition in financial markets using RF models is shown in a bar chart in Figure 5. The performance is evaluated using four main metrics F1-score, recall, accuracy, and precision. 86.94% accuracy, 86.01% precision, 87.18% recall, and the maximum F1-score of 88.39% were attained. The graph includes a blue trend line indicating an upward trend from Precision to F1-score. The visualization demonstrates the RF model's efficacy in predicting customer attrition, showing comparatively strong performance across all criteria.

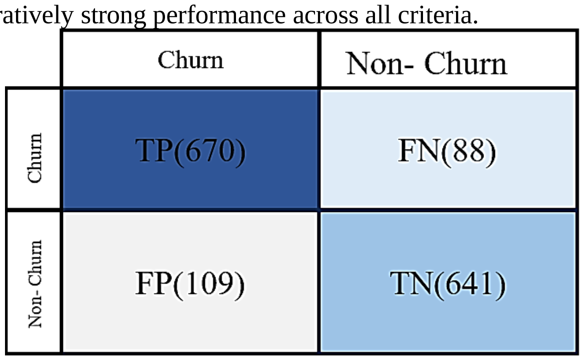

Figure 6: Confusion Matrix for Random Forest

The customer churn prediction model's confusion matrix is displayed in Figure 6. It displays how well the model classifies consumers as either churn-prone or non-churn. The matrix's TP of 670 indicates that churn cases were successfully predicted; its FN of 88 indicates that non-churn customers were misclassified as such; its FP of 109 indicates that non-churn customers were misclassified as such; and its TN of 641 indicates that TN accurately anticipated non-churn instances.

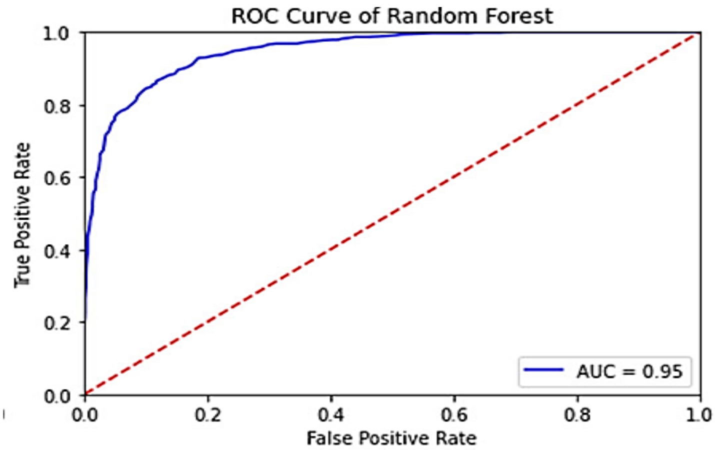

Figure 7: Receiver Operating Characteristic (ROC) Curve and Value for Random Forest

The ROC curve for the RF model used to forecast customer attrition is shown in Figure 7. The red straight line indicates an arbitrary amount a classification algorithm, while the mathematical model's sensitivity to between churn and non-churn customers is indicated by the blue line curve. The model can reliably differentiate between those who churn and those who don't because of its powerful discriminating power, as demonstrated by its AUC score of 0.95.

3.8 Comparative Analysis

This section provides a comparative analysis demonstrating that RF surpasses other models, including SVM[24],DT[25] LR[26], confirming its superior predictive accuracy. The model comparisons are detailed in Table 3.

Table 3: Comparative Examination of Financial Sector Customer Retention Prediction Utilizing Various ML Models

Model	Accuracy
RF	86.94
SVM	79.6
Decision Tree	76.3
Logistic Regression	64.9

A comparison of many prediction models is shown in Table 3, which shows differing accuracy levels. The accuracy of RF is the highest at 86.94%. Thus, it is good at handling complex patterns and reducing overfitting. RF followed with an accuracy of 79.6%, which can be considered as a good classifier in classification tasks but lagging slightly compared to the SVM. The accuracy of the DT was lower than that of RF and SVM, with 76.3%, but that still did pretty well in recognizing decision rules. The accuracy of LR was the lowest, only at 64.9%, and therefore, it is not suitable for nonlinear relationships and complex data. RF was the most effective model, SVM and DT were moderate, and LR lagged behind due to linearity assumption and simplicity.

4. Conclusion and Future Scope

Predicting client turnover is essential since financial organizations must implement methods to keep customers and reduce revenue loss. The MLP, SVM, and RF models are tested in this study, which uses ML to predict customer turnover. When compared against other models, RF outperformed them with 92.5% accuracy, 91.8% precision, 90.3% recall, 91.0% F1 score, and an AUC ROC score of 0.95 overall. RF's superior performance is because it is able to deal with large datasets and capture complex patterns of customer behavior. This highlights the importance of the use of ML techniques in financial markets for better squeeze prediction accuracy and to manage customer relationships.

Although this study demonstrates that RF is a successful approach for churn prediction, there are several research directions for further enhancements in predictive performance. Other DL techniques that are more successful in capturing sequential, changing consumer behaviors, like transformer-based architectures and LSTM networks, may be investigated. This does include the integration of real-time data streams, which would increase the model's adaptability so that financial institutions can pick up early indicators of churn and prevent instigating it. Businesses may acquire trust in prediction models and comprehend the crucial elements influencing churn choices by implementing Explainable AI (XAI) solutions to enhance transparency. In addition, one can investigate hybrid approaches that combine traditional ML methods with DL approaches to have higher accuracy and robustness in churn prediction. The following can be included in future studies to more accurately predict external factors such as macroeconomic trends, customer sentiment from social media, transactional behaviors as well and many more. These advancements can allow financial institutions to improve these retention strategies by developing them with the use of data to reduce churn and increase long-term customer loyalty.

References

[1] O. Adwan, H. Faris, K. Jaradat, O. Harfoushi, and N. Ghatasheh, “Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis,” *Life Sci. J.*, vol. 11, no. 3, pp. 75–81, 2014.

[2] V. Umayaparvathi and K. Iyakutti, “A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics,” *Int. Res. J. Eng. Technol.*, 2016.

[3] C. Wang, R. Li, P. Wang, and Z. Chen, “Partition cost-sensitive CART based on customer value for Telecom customer churn prediction,” in *Chinese Control Conference, CCC*, 2017. doi: 10.23919/ChiCC.2017.8028259.

[4] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, “Credit card churn forecasting by logistic regression and decision tree,” *Expert Syst. Appl.*, 2011, doi: 10.1016/j.eswa.2011.06.028.

[5] B. Gregory, “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data First-place Entry for Customer Churn Challenge in WSDM Cup 2018,” 2018.

[6] P. Li, S. Li, T. Bi, and Y. Liu, “Telecom customer churn prediction method based on cluster stratified sampling logistic regression,” in *IET Conference Publications*, 2014. doi: 10.1049/cp.2014.1576.

[7] A. Idris, M. Rizwan, and A. Khan, “Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies,” *Comput. Electr. Eng.*, 2012, doi: 10.1016/j.compeleceng.2012.09.001.

[8] N. Lu, H. Lin, J. Lu, and G. Zhang, “A customer churn prediction model in telecom industry using boosting,” *IEEE Trans. Ind. Informatics*, 2014, doi: 10.1109/TII.2012.2224355.

[9] F. Castanedo, G. Valverde, J. Zaratiegui, and A. Vazquez, “Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network Federico,” *p*, 2014.

[10] W. H. Au, C. C. Chan, and X. Yao, “A novel evolutionary data mining algorithm with applications to churn prediction,” *IEEE Trans. Evol. Comput.*, 2003, doi: 10.1109/TEVC.2003.819264.

[11] K. A. Amuda and A. B. Adeyemo, “Customers Churn Prediction in Financial Institution Using Artificial Neural Network,” 2019.

[12] R. Li, Y. Zhang, Y. Tuo, and P. Chang, “A Novel Method for Detecting Telecom Fraud User,” in *Proceedings - 2018 3rd International Conference on Information Systems Engineering, ICISE 2018*, 2018. doi: 10.1109/ICISE.2018.00016.

[13] A. Mishra and U. S. Reddy, “A comparative study of customer churn prediction in telecom industry using ensemble based classifiers,” in *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017*, 2018. doi: 10.1109/ICICI.2017.8365230.

[14] A. Mishra and U. S. Reddy, “A Novel Approach for Churn Prediction Using Deep Learning,” in *2017 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2017*, 2018. doi: 10.1109/ICCIC.2017.8524551.

[15] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, “Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning,” in *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCEE 2018*, 2018. doi: 10.1109/ICSCEE.2018.8538420.

[16] S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra, and S. Kumar, “Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron(MLP),” in *Proceedings of the 2nd International Conference on Green Computing and Internet of Things, ICGCIoT 2018*, 2018. doi: 10.1109/ICGCIoT.2018.8752982.

[17] S. A. Alasadi and W. S. Bhaya, “Review of data preprocessing techniques in data mining,” *J. Eng. Appl. Sci.*, 2017, doi: 10.3923/jeasci.2017.4102.4107.

[18] X. Tan *et al.*, “Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm,” *Sensors (Switzerland)*, 2019, doi: 10.3390/s19010203.

[19] J. Salminen, V. Yoganathan, J. Corporan, B. J. Jansen, and S. G. Jung, “Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type,” *J. Bus. Res.*, 2019, doi: 10.1016/j.jbusres.2019.04.018.

[20] H. Tyralis, G. Papacharalampous, and A. Langousis, “A brief review of random forests for water scientists and practitioners and their recent history in water resources,” 2019. doi: 10.3390/w11050910.

[21] L. Gigović, H. R. Pourghasemi, S. Drobnjak, and S. Bai, “Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia’s Tara National Park,” *Forests*, 2019, doi: 10.3390/f10050408.

[22] B. Huang, M. T. Kechadi, and B. Buckley, “Customer churn prediction in telecommunications,” *Expert Syst. Appl.*, 2012, doi: 10.1016/j.eswa.2011.08.024.

[23] J. Burez and D. Van den Poel, “Handling class imbalance in customer churn prediction,” *Expert Syst. Appl.*, 2009, doi: 10.1016/j.eswa.2008.05.027.

[24] [24] P. Bhanuprakash and G. S. Nagaraja, “A Review on Churn Prediction Modeling in Telecom Environment,” in *2nd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2017*, 2018. doi: 10.1109/CSITSS.2017.8447617.

[25] K. Ebrah and S. Elnasir, “Churn prediction using machine learning and recommendations plans for telecoms,” *J. Comput. Commun.*, vol. 7, no. 11, pp. 33–53, 2019.

[26] Q. Yihui and Z. Chiyu, “Research of indicator system in customer churn prediction for telecom industry,” in *ICCSE 2016*

- *11th International Conference on Computer Science and Education*, 2016. doi: 10.1109/ICCSE.2016.7581567.
[27] Kalla, D., Kuraku, D. S., & Samaah, F. (2021). Enhancing cyber security by predicting malwares using supervised machine learning models. *International Journal of Computing and Artificial Intelligence*, *2*(2), 55-62.