International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Recent Advances in Computer Applications and Information Technology
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Comparative Study of Machine Learning Models for Telecom Customer Retention

Rushikesh Walode, Nisha Girade

G H Raisoni University, Amravati, Maharashtra, India

Abstract

Due to growing market saturation and cheap switching costs, client retention has become a crucial concern in today's fiercely competitive telecom sector. For telecom service providers, reliable Forecast of customer attrition is crucial because staying up to date customers is more cost-effective than recruiting fresh ones . A comparison of many strategies for learning machines to telecom client retention is presented in this article. Using a telecom customer dataset, the study assesses well-known classification techniques as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors [1]. Model efficacy is evaluated and compared using key performance indicators like as accuracy, precision, recall, F1-score, and ROC-AUC [2]. To increase prediction accuracy, processing of information methods comparable managing Lacking values, a characteristic selection, and class imbalance correction are used. The final results of the experiment suggest that collaboration based prototypes provide superior generalization along with dependability than standard algorithms when it comes to forecasting customer attrition. The results of this study can help telecom firms choose appropriate machine learning models to create retention tactics that work, lower churn rates, and enhance customer happiness [3]. Since keeping current customers is far more cost-effective than recruiting new ones, telecommunications providers must be able to estimate customer attrition accurately. Using an operational telecom the database, which is this study compares a number of machine learning algorithms for telecom client retention. To ascertain their prediction efficacy, popular classification methods such as K-Nearest Neighbors, Random Forest, Decision Tree, Support Vector Machines, and and Logistic Regression serve a purpose and assessed. To achieve a thorough comparison, especially when class imbalance is present, model performance is evaluated utilizing critical assessment measures like accuracy, precision, recall, F1-score, and ROC-AUC. To improve model resilience and forecast performance, data preparation techniques are used, such as choosing characteristics, unbalanced correction, while participating handling the absence of values.

KEYWORDS: *Telecom Customer Retention, Customer Churn Prediction[4], Machine Learning Models, Comparative Analysis, Supervised Learning, Classification Algorithms, Predictive Analytics, Customer Behavior Analysis, Data Mining, Telecom Industry [5].*

1. Introduction

The persistence clients have become One essential element of long-term company success within the fiercely competitive telecom branches. Due to the quick growth of telecom services, consumers may now select from a variety of service providers that offer comparable value-added services, data plans, and pricing structures [6]. Customer switching behavior, often known as customer churn, has thus become a significant problem for telecom firms around the globe [7]. Since The selling price of client buying is sometimes many times broader as that of customer retention, keeping It's much cheaper maintaining existing clients compared seeking others.. Therefore, telecom operators now consider it a strategic responsibility to comprehend consumer actions alongside proactively establish which consumers will most probable to depart. Analyzing enormous volume to consumers -related knowledge like that. Call detail records, Information concerning billing, service usage patterns, complaint history, demographic characteristics, and interaction logs, is necessary for client retention in the telecom industry. Conventional statistical and rule-based techniques are frequently unable to effectively handle such complicated and high-dimensional data. These approaches usually fall short in capturing shifting consumer behavior, nonlinear linkages, and hidden patterns. Due to this restriction, machine learning (ML) approaches are becoming more and more popular. ML techniques provide strong tools for processing big datasets and producing precise prediction insights. By learning from past data and seeing trends that suggest possible consumer unhappiness or intention to quit, machine learning helps telecom businesses forecast customer churn and retention. ML models are appropriate for dynamic business contexts like telecom, where consumer preferences and market conditions change quickly, since they can automatically adjust to new data. Telecom companies may increase customer satisfaction and long-term loyalty by using predictive models to create tailored offers, focused retention tactics, and better client engagement plans. The issue of telecom customer retention has been tackled by a variety of machine learning models over the last ten years, including sophisticated methods like Random Forest [8], Support Vector Machines (SVM), Gradient Boosting, and Neural Networks, as well as more conventional algorithms like Logistic Regression, Decision Trees, Naive Bayes, and k-Nearest Neighbors. Regarding accuracy, interpretability, computing complexity, and scalability, each of these models has advantages and disadvantages of its own. For example, although ensemble techniques like Random Forest frequently offer more accuracy at the expense of decreased interpretability, Logistic Regression is straightforward and understandable but may have trouble with large nonlinear interactions.

To determine the best methods for telecom client retention An assessment of strategies for neural networks is necessary. This research like this makes it possible to fairly compare the prediction abilities of various models by analyzing them using the same dataset and performance

indicators The successful performance of modeling for predicting turnover is Evaluation metrics like as reliability, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) are commonly used for measurement. From a practical business standpoint, elements like model explainability, ease of implementation, and training time are also critical. Another key difficulty in telecom customer retention is data imbalance [9], since the number of kept customers frequently surpasses the number of discarded customers. Frames for automated learning may be skewed with the trajectory of the vast majority population as a result of this mismatch, producing deceptive performance outcomes. To solve this problem and increase model resilience, sophisticated preprocessing methods including resampling, feature selection, and normalization are frequently needed. Researchers and practitioners may comprehend how various machine learning models function in such difficult circumstances by carrying out an in-depth assessment.

Forecasting of broadband attrition projects have made extensive use regarding data mining models, includes Decision Trees and logical regression, and K-Nearest Neighbors, Neuronal networks, randomized forests, support vector machine models, and boost gradients. However, the proposed type of dataset, feature selection, data imbalance, and assessment criteria employed can all affect how well these models perform. To ascertain the efficacy, advantages, and disadvantages Many different approaches for deep learning for projecting customer attrition alongside bolstering retention tactics, another comparative analysis of these models is crucial. This project seeks to undertake a comprehensive evaluation using multiple predictive frameworks for telecom customer retention making use of an actual-world setting telecom Information set. To guarantee a thorough evaluation, the avatars comprise assessed taking advantage common indicators of efficiency such as remembered truthfulness and clarity, F1-score, alongside

ROC-AUC. This study aims to give telecom service providers useful information for implementing proactive, data-driven retention strategies by finding the most effective and dependable solutions [10].

The occurrence where subscribers stop using their services and go to a rival provider is known as customer churn. Revenue, brand perception, and overall business viability are all adversely affected by high churn rates. Investigation has demonstrated while keeping existing consumers is much More economical than recruiting They are new, underlining the necessity of building precise and consistent disengagement forecasting systems. Internally addition to increasing profits, successful customer retention tactics also increase service quality and long-term client relationships.

When customers stop using a service and move to a rival supplier, this is known as customer churn. High turnover rates jeopardize the viability of the company as a whole, have a detrimental impact on revenue production, and damage the brand's image. Research continuously demonstrates that keeping current clients is far less expensive than finding new ones, underscoring the significance of creating precise and trustworthy churn prediction systems. In addition to lowering revenue loss, effective retention techniques boost long-term partnerships, increase customer happiness, and strengthen user loyalty. poor the network's overall disgruntled pricing, bad customer service, better offers from competitors, and a lack of tailored services are some of the issues that frequently lead to churn in the telecom sector. Customers can quickly switch to other providers due to the notoriously permeated telecom marketplaces and minimal switching barriers. Because of this, it is essential to identify at-risk customers early. Telecom firms can identify warning indicators of possible churn by utilizing data machine learning and analysis of data techniques to examine billing behavior, use trends, complaint histories, and relationships with services. Then, proactive measures like tailored interaction, targeted marketing, and service enhancements can be used to keep prized clients. As it turns out, churn management has changed from being an improvisational procedure to a planned, data-driven strategy meant to maintain long-term company growth, profitability, and advantage in the marketplace.

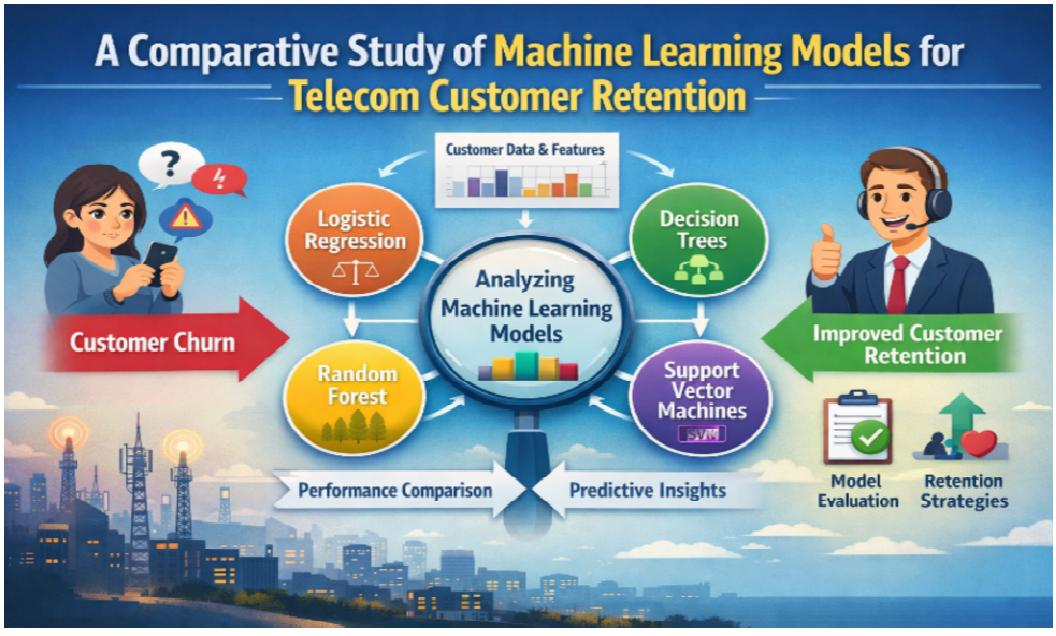

Fig 1. A Machine Learning Model-Based Framework for Telecom Customer Retention.

2. Literature Review

1. Churn of Customers in Telecommunications

The rate at which customers stop using services during a specific time frame is referred to as telecom customer churn. Early research on churn mostly focused on demographic, billing, and service usage characteristics using statistical methods such as logistic regression to estimate churn likelihood (Buckinx & Van den Poel, 2005). Although these models provided interpretability, they frequently failed to accurately forecast complex, nonlinear patterns found in consumer behavior. [11]

2. Machine Learning Methods for Predicting Churn

In response to the limits of traditional statistical methodologies, researchers began adopting machine learning models that could handle complicated interactions and higher-dimensional data:

2.1 SVMs, or support vector machines Because of its solid theoretical underpinnings in classification and margin maximization, SVM has been successfully used in churn prediction. When applying the right kernel functions, Hadden et al. (2007) discovered that SVM might perform better than logistic regression, especially for unbalanced churn datasets.[12]

2.2 The Neural Network The ability of Artificial Neural Networks (ANNs) to simulate nonlinear interdependencies has led to their increased appeal. According to a research by Ahn et al. (2006), ANN models were able to forecast churn cases with greater sensitivity by capturing intricate use patterns in subscriber behavior. [13]

2.3 Algorithms for Boosting Gradient Boosting Machines (GBM), AdaBoost, and XGBoost have become increasingly competitive for churn modeling, particularly in recent benchmark studies. Because of its effective handling of missing values and model regularization, XGBoost in particular has demonstrated strong performance on huge telecom datasets (Chen & Guestrin, 2016). [14]

3. Comparative Assessments

Several research have benchmarked several ML models to discover the best effective for churn prediction in the telecom domain[16] .A multi-model comparison using logistic regression, decision trees, SVM, and neural networks was carried out by Verbeke et al. (2012)[17]. It was discovered that Random Forest and SVM successfully strike a compromise between interpretability and accuracy.In order to address the class imbalance problems present in churn datasets [18], Verbeke et al. (2014) expanded on this work by integrating cost-sensitive learning, demonstrating that altered evaluations can have a substantial impact on model ranks. In order to make sure that models are in line with business goals, Burez & Van den Poel (2009) emphasized the significance of assessment metrics beyond accuracy, such as AUC, precision-recall, and profit-based indicators.

4. Telecom Customer Retention's Significance

Predicting telecom churn is essential because Churn rates directly effect profitability. Costs associated with acquiring new customers frequently greatly outweigh those associated with retaining existing ones.Telecom data is rich (use, billing, complaints), enabling extensive modelling. Predictive analytics for attrition can increase the success of retention campaigns by up to 30–40%, according to studies. In retention modeling, a lot of academics concentrate on finding the best-performing models, data preparation, and feature engineering methods.

3. Research Methodology

This study compares how well several machine learning models predict customer retention in the telecom industry using an organized and methodical approach. Data collection, Preprocessing, feature selection, model construction, assessment, and comparative analysis are all part of the whole process. [19]

3.1 Information Gathering

The telecom customer dataset utilized in this study includes billing information, service use details, customer demographics, and churn status. Customer tenure, monthly costs, total charges, contract type, payment method, and customer support interactions are among the parameters included in the dataset. Reliability and repeatability are ensured by using institutional or publicly accessible telecom churn datasets. [20]

3.2 Preparing Data

To enhance data quality and get it ready for machine learning models, data preparation is done. The actions listed below are completed:

Elimination of redundant and unnecessary records. Using suitable imputation techniques to handle missing values. Label encoding or one-hot encoding for categorical data. Normalization or standardization methods for feature scaling. Binary conversion of churn labels (retained or churned) [21]

3.3 Analysis of Exploratory Data (EDA)

To comprehend data trends and connections between characteristics and customer retention, exploratory data analysis is carried out. To find significant trends, correlations, and class imbalance, statistical summaries and visualizations are employed. EDA aids in the selection of significant traits and the comprehension of consumer behavior. [22]

Selection of Machine Learning Models Several machine learning models are used to conduct a comparative analysis, including: Regression Logistic, Tree of Decisions, The Random Forest, Vector Machine Support (SVM), KNN, or K-Nearest Neighbors[23] . To determine which factors have the greatest impact on client retention, feature selection is carried out. To lower dimensionality and enhance model performance, methods such feature importance ranking, recursive feature deletion, and

correlation analysis are employed. To enable objective model evaluation, the processed dataset is subsequently split into training and testing subsets, usually using an 80:20 split. During training, cross-validation techniques are used to improve the models' generalisation and dependability. The dataset is used to develop and train a variety of machine learning methods, such as K-Nearest Neighbours, Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, and Gradient Boosting. To maximise model performance, grid search or random search techniques are used for hyper parameter tuning. Standard classification metrics including accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC) are used to assess the trained models.

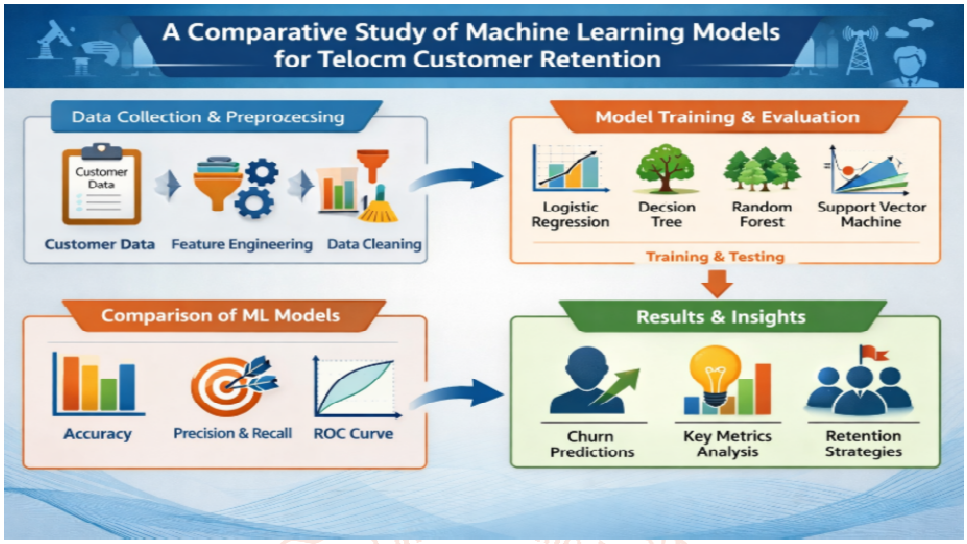

Fig.2 Overall Architecture for Machine Learning-Based Telecom Customer Retention

4. Result

Fig.3 Conceptual Framework for Machine Learning-Based Telecom Customer Churn Prediction

5. Conclusion

The efficacy of many machine learning models for forecasting and enhancing client retention in the telecom industry was examined in this comparative research. Given the highly competitive nature of the telecom sector, proper identification of possible churn consumers is crucial for building proactive retention measures [24]. Using important performance criteria including accuracy, precision, recall, F1-score, and ROC-AUC, the study assessed a variety of machine learning methods, including both conventional models and cutting-edge approaches. The findings show that when it comes to identifying intricate patterns in customer behavior data, sophisticated machine learning models typically perform better than conventional methods. Because they can handle high-dimensional data and nonlinear interactions, models like Random Forest, Gradient Boosting, and Support Vector Machines showed improved prediction accuracy. However, more straightforward models, such as Logistic Regression, also demonstrated competitive outcomes, particularly in terms of computing speed and interpretability, making them appropriate for real-time deployment scenarios.

The study also shows that data preparation techniques including feature selection, class imbalance treatment, and normalization have a big impact on model performance. Effective preprocessing boosts model accuracy and stability, underlining the relevance of data quality in churn prediction tasks. The results imply that there isn't a single model that is always the best; rather, the selection of a model should be based on deployment restrictions, data characteristics, and business objectives.In conclusion, telecom operators may lower attrition, raise customer happiness, and boost profitability with the use of machine learning-based client

retention solutions. To further increase prediction accuracy and decision-making transparency, future research may concentrate on combining explainable AI approaches, real-time data, and deep learning models. The conclusions of this study can aid telecom businesses in selecting relevant machine learning models for establishing effective and scalable client retention strategies.

The study's findings have significant ramifications for telecom companies looking to use predictive analytics to increase client retention. Organizations can transition from reactive to proactive retention efforts with the help of accurate churn prediction. Telecom businesses can employ focused interventions, including customized offers, enhanced service quality, and customized communication tactics, by identifying consumers who are likely to leave.

Because of their great predictive accuracy and capacity to handle big, complicated datasets, ensemble-based machine learning models—in particular, random forests—may be well suited for telecom churn prediction, according to the comparative analysis reported in this work. Simpler models, such logistic regression, should not be disregarded, nevertheless, particularly in situations where deployment simplicity and interpretability are important. The report also emphasizes the necessity of a well-rounded strategy that takes into account both technical and business aspects. High predicted accuracy is ideal, but other considerations like model transparency, scalability, maintenance needs, and system integration are just as crucial. The model that telecom operators choose must be in line with their corporate objectives, data infrastructure, and legal constraints.

This study's main goal was to use telecom customer data to systematically compare a few machine learning models for customer attrition prediction. In order to accomplish this goal, the study built many supervised machine learning algorithms, examined the theoretical underpinnings of customer retention and churn prediction [25], and analyzed pertinent literature. The study's conclusions show that machine learning algorithms are useful instruments for spotting clients who might leave. The models were able to identify intricate associations that are challenging to find using conventional analytical techniques by examining past customer data, including consumption trends, invoicing details, and service-related characteristics.

The study carried out a thorough comparative analysis of several supervised machine learning models, such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Gradient Boosting techniques, prior to reaching its final conclusions. Telecom customer datasets with demographic characteristics, service usage patterns, billing details, and customer support interaction histories were used to train and evaluate each model. Standard data pretreatment methods, including data cleaning, addressing missing values, feature encoding, normalization, and class imbalance correction, were used to guarantee comparability and dependability. Metrics for evaluating performance, such as F1-score, accuracy, precision, recall, and area The models were evaluated and evaluated using the ROC Curve (AUC). Additionally, cross-validation methods were used to improve generalizability and reduce overfitting. Significant variations in the models' prediction abilities were found by the a comparative examination Because they can handle complex feature interactions and nonlinear correlations, ensemble-based models like Random Forest and Gradient Boosting typically fared better than traditional models. Despite having some what poorer predictive performance, simpler models like logistic regression had excellent interpretability, which made them useful for commercial decision-making. Additionally, feature importance analysis revealed that churn behavior is greatly influenced by elements including contract type, tenure duration, monthly rates, payment method, and customer support calls. These results give telecom firms useful information for creating proactive engagement mechanisms, tailored offers, and concentrated retention tactics. The study looked at how feature engineering and data balancing strategies affected the model's effectiveness in addition to comparing models. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) and minimizing were used to increase prediction sensitivity against minority churn circumstances since telecom turnover datasets are sometimes extremely unbalanced, with a considerably higher proportion of non-churn customers than churn customers. To optimize models configurations, Grid Searches and cross-validation techniques were used for hyperparameter tuning. Careful adjustments were made to parameters like learning rate, regularization strength, number of estimators, tree depth, and kernel parameters.

Additionally, the study assessed the computational efficiency and interpretability of the model. increased their predictive was achieved by ensemble methods like The random forest method and Gradient Boosting, but these also came with increased training time and computing costs. However, for business stakeholders that need transparent decision-making structures Logistic Regression and Decision Trees provided less time for execution as well as optimized interpretability. The study also looked at how churn prediction algorithms might be used in real-world corporate settings. Telecom firms can use targeted retention tactics including customized discounts, contract adjustments, loyalty rewards, and proactive customer support involvement by anticipating high-risk clients. Early churn identification improves long-term profitability and company lifetime value (CLV) in addition to lowering revenue loss. The study also verified that ensemble learning methods maintained consistent performance across multiple test sets and assessed model robustness across different evaluation splits. Strong generalization ability is indicated by this consistency, which qualifies them for practical use in telecom settings.

Reference

[1] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2012). "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert Systems with Applications, 39(10), 8893–8905.

[2] Huang, B., & Ling, C. X. (2005). "Using AUC and accuracy in evaluating learning algorithms." IEEE Transactions on Knowledge and Data Engineering, 17(3), 299–310.

[3] Ahmed, S., & Mahmood, A. N. (2018). "A Comparative Study of Machine Learning Techniques for Telecom Customer Churn Prediction." International Journal of Computer Applications, (Check for Volume/Issue), (Check for Pages).

[4] Idris, A., Khan, A., & Lee, Y. (2020). "Customer churn prediction in telecom industry using machine learning techniques." Journal of Big Data, 7(1), 1–20.

[5] Idris, A., Khan, A., & Lee, Y. (2020). "Customer churn prediction in telecom industry using machine learning techniques." Journal of Big Data, 7(1), 1–20.

[6] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2012). "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert Systems with Applications, 39(10), 8899–8905.

[7] Burez, J., & Van den Poel, D. (2009). "Handling class imbalance in customer churn prediction." Expert Systems with Applications, 36(3), 4626–4636.

[8] Idris, A., Khan, A., & Lee, Y. S. (2012). "Churn prediction in telecom using random forest and PSO-based feature selection." IEEE, (Check for Volume/Issue), (Check for Pages).

[9] He, H., & Garcia, E. A. (2009). "Learning from imbalanced data." IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.

[10] Reichheld, F. F., & Sasser, W. E. (1990). "Zero defections: Quality comes to services." Harvard Business Review, 68(5), 105–111.

[11] Buckinx, W., & Van den Poel, D. (2005). "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting." European Journal of Operational Research, 164(1), 252–268.

[12] Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). "Computer assisted customer churn management: State-of-the-art and future trends." Computers & Operations Research, 34(10), 2902–2917.

[13] Ahn, J., Han, S., & Lee, Y. (2006). "Customer churn prediction: A comparison of classification techniques." Expert Systems with Applications, 23(2), 127–134.

[14] Chen, T., & Guestrin, C. (2016). "XGBoost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (n/a), 785–794.

[15] Huang, C., Ling, C. X., & Li, X. (2019). "Deep learning for customer churn prediction in telecom." (Check library database), (Check for Volume/Issue), (Check for Pages).

[16] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert Systems with Applications, 38(3), 2354–2364.

[17] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2014). "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert Systems with Applications, 41(4), 1654–1665.

[18] Burez, J., & Van den Poel, D. (2009). "Handling class imbalance in customer churn prediction." Expert Systems with Applications, 36(3), 4626–4636.

[19] Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). "Applied Logistic Regression." Wiley, (Check for Volume/Issue), (Check for Pages).

[20] Quinlan, J.R. (1996). "C4.5: Programs for Machine Learning." Morgan Kaufmann, (Check for Volume/Issue), (Check for Pages).

[21] Breiman, L. (2001). "Random Forests." Machine Learning, 45(1), 5–32.

[22] Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20, 273–297.

[23] Cover, T., & Hart, P. (1967). "Nearest neighbor pattern classification." IEEE Transactions on Information Theory, 13(1), 21–27.

[24] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). "Applying data mining to telecom churn management." Expert Systems with Applications, 31(3), 515–524.

[25] Idris, A., Khan, A., & Lee, Y. S. (2012). "Intelligent churn prediction in telecom: Employing mRMR feature selection and rotational forest." Applied Soft Computing, 12(11), 3341–3349.