2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC) | 979-8-3503-7651-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICEC62498.2024.10808863

Telecom Customer Churn Prediction using enhanced Machine Learning Classification Techniques

Goldy Verma

Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

goldy.verma@chitkara.edu.in

***Abstract*— For telecommunications corporations, customer attrition is a chief trouble that outcomes in large revenue losses and better costs of customer acquisition. by using putting in place cantered retention programmes, corporations who can as it should be forecast attrition can reduce these losses. in this work, improved device learning classification methods are used to predict customer attrition inside the telecom sector. Three classifiers were used to build predictive models: Decision Tree, Random Forest, and K-Nearest Neighbours (KNN). The models produced, in turn, accuracy of 79%, 82%, and 79%. The Random Forest classifier proved to be the only approach for forecasting viable client churn due to its superior performance in extra correctly predicting such turnover. This work gives a reliable approach for predicting consumer turnover by the usage of machine gaining knowledge of classifiers like selection Tree, Random Forest, and KNN. because those models are so accurate—the Random wooded area specifically—telecom corporations may additionally better manage client happiness, reduce churn risks, and maximise retention efforts.**

***Keywords*— Artificial Intelligence, Machine Learning, Decision Tree, Random Forest, and K-Nearest Neighbours (KNN), Telecom Customer Churn Classification, Classifier Training**.

I. INTRODUCTION

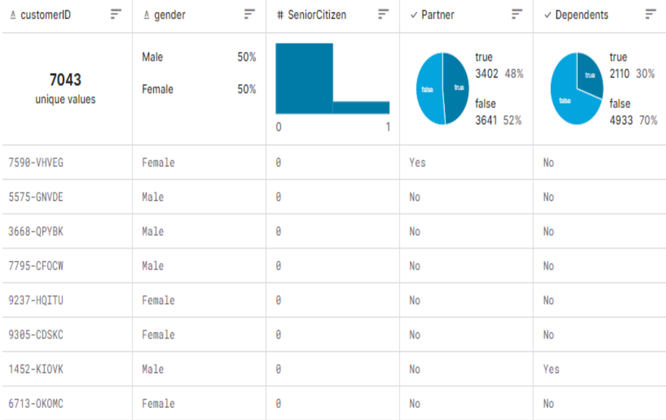
Maintaining profitability and corporate growth in the increasingly competitive telecom industry mostly depends on keeping customers. When a customer cancels their service, a trend known as customer churn, there are major issues and significant financial losses. Telecom companies are increasingly turning to advanced data analytics and machine learning techniques to forecast and lower churn as well as to develop focused retention plans that can increase customer happiness and loyalty. The need of precise churn prediction has motivated the investigation of several machine learning techniques. In order to improve the churn, forecast accuracy, Of and colleagues [1] offer a hybrid architecture that integrates many techniques. Their approach underlines that the complex patterns associated with consumer behaviour must be described by merging several models. A deep learning-based model called ChurnNet employs neural networks to improve telecom sector forecast, according to Saha et al. [2]. They found that deep learning handles huge telecom datasets with complicated properties well.KNN, Decision Trees, and Random Forest are used to anticipate churn. Wagh et al. [3] discuss these methods highlight each of their advantages and disadvantages in relation to telecom data. Comparative research, including that of Das and Mahendher [4], clarifies the performance of different machine learning methods and provide important information about their suitability for churn prediction. Apart from the conventional machine learning techniques, ensemble learning techniques have become more popular. Wahul et al. [5] look into how merging several models and using the benefits of various methods might increase the forecast accuracy. This ensemble approach is especially well adapted to reduce the variance and bias associated to single models. In their adaptive learning approach, Amin and colleagues [6] use evolutionary computation and Naïve Bayes to dynamically adapt to changing client behaviours, hence improving prediction reliability.KNN is useful for categorization beyond churn prediction. For adaptability and robustness in multiple domains, Agarwal and colleagues [7] study its use in network security. KNN's cross-domain use shows its promise as a telecom churn classifier. Finally, practical methods for forecasting telecom customer attrition are shown via the integration of state-of-the-art machine learning techniques. Telecom companies that use models like Decision Trees, Random Forest, and KNN to gain a better grasp of customer behaviour can develop effective retention strategy. The dynamic client turnover in the telecom industry will require these models to be updated and modified on a regular basis.

II. LITERATURE

Studies in the telecom enterprise is sizeable considering purchaser turnover without delay impacts sales and efforts to retain customers. numerous systems learning techniques had been studied to elevate the accuracy and dependability of fashions for churn prediction. the use of analysis of variance (ANOVA), Babatunde and associates [8] investigated a way to classify fashions of customer churn prediction. Their research validated how ANOVA would possibly successfully distinguish the giant features from the less enormous ones, consequently enhancing the model's overall performance. With this approach, it's far made clean how critical characteristic choice is to constructing trustworthy churn forecast fashions. Kavitha and friends [9] examined resource vector machines, choice bushes, and logistic regression as device mastering techniques for forecasting consumer attrition inside the telecom region. Their outcomes showed that, at the same time as each approach had benefits, ensemble strategies—which combine several algorithms—frequently produced higher outcomes. This realisation is in keeping with the greater trendy fashion in device getting to know to apply ensemble techniques to optimise the benefits of numerous models. the usage of BERT for function extraction and logistic regression, Mittal and co-workers [10] identified phishing domain names. They used methodologies that demonstrate the flexibility and application of logistic regression and sophisticated feature extraction techniques in a variety of classification problems, including customer churn prediction, even though their work was on a different domain. Customer attrition was predicted by Saheed and Hambali [11] using machine learning and information gain filter feature selection methods. They found that by removing superfluous and unnecessary information, feature selection techniques might greatly improve the predicted accuracy of machine learning models. For customer churn prediction, Beeharry and Fokone [12] put up a hybrid method utilising machine learning technique. Their method more advantageous prediction accuracy via combining many algorithms to seize the difficult styles in consumer statistics. The tendency of blending several gadget gaining knowledge of methods to improve version overall performance is first-class proven by using this hybrid technique. In an ensemble-based totally technique to beautify purchaser churn prediction, Bilal et al. [13] incorporated clustering and class algorithms. The fact that this combination made it possible to become aware of churn patterns more exactly underscores the advantages of ensemble mastering for managing a spread of facts functions. Labhsetwar [14] concentrated on methods of supervised mastering for forecasting client attrition. The paintings highlighted the performance of supervised getting to know fashions in detecting churn trends and offered an intensive assessment of several strategies, therefore advancing understanding of their blessings and downsides in churn prediction. The studies indicate, in summary, that function choice, hybrid fashions, and ensemble strategies are fantastically valued in the telecom quarter to enhance the precision and dependability of client churn prediction. this research taken together spotlight how device mastering techniques are continually changing and how important they're to growing models that forecast churn.

III. INPUT DATASET

The Kaggle open-source platform offered this study's dataset, which comprises a wide range of telecom business customer data needed to anticipate customer attrition. The CSV dataset provides numerous classifications for churn prediction and client profiling.These classes are customerID, gender, SeniorCitizen status (indicating if the customer is a senior citizen), Partner (indicating if the customer has a partner), Dependents (indicating if the customer has dependents), tenure (the length of the customer's association with the company), PhoneService (indicating if the customer has a phone service), MultipleLines (indicating if the customer has multiple lines), InternetService (denoting the type of internet service), OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, and StreamingMovies (reflecting the availability of matching services).Additionally covered within the dataset are, one-year, and two-year agreement types, PaperlessBilling reputation, PaymentMethod, MonthlyCharges (the quantity paid monthly the patronmonthly each month), TotalCharges (month-to-monththe monthmonthly quantity charged), and Churn repute (which suggests whether or not or now not the month-to-month has churned). collectively, those month-to-month give incisive facts on service subscriptions, billing alternatives, and cusmonthmonthly demographics, which allows the development of prediction models that allow telecom businesses month-to-month expand effective consumer retention techniques.



customerID	gender	SeniorCitizen	Partner	Dependents
7043 unique values	Male 50% Female 50%	0 1	true 3402 48% false 3641 52%	true 2110 30% false 4933 70%
7590-VHVEG	Female	0	Yes	No
5575-GNVDE	Male	0	No	No
3668-QPYBK	Male	0	No	No
7795-CFOCW	Male	0	No	No
9237-HQITU	Female	0	No	No
9305-CDSKC	Female	0	No	No
1452-KIOVK	Male	0	No	Yes
6713-OKOMC	Female	0	No	No

Fig. 1. CSV type dataset utilized for Telecom Customer Churn prediction

IV. PROPOSED METHODOLOGY

In this study, the suggested approach for forecasting customer churn in the telecom business entails utilising upgraded machine learning classification algorithms. This methodology was developed in order to increase accuracy. In particular, the development of prediction models makes use of three different classifiers: the Decision Tree, the Random Forest, and the K-Nearest Neighbours (KNN) algorithm. Preprocessing is achieved at the dataset together with client traits like gender, Senior Citizen reputation, tenure, service subscriptions, billing choices, and churn fame on the way to deal with lacking values, encode categorical variables, and normalise numerical traits. To higher assess the version's overall performance, the dataset is next break up into schooling and trying out sets. A good way to maximise the classifiers' ability for prediction, they're skilled on the training set and then first-rate-tuned via the application of methods together with go-validation and hyperparameter tuning. After education, the fashions are tested at the testing set using performance measures along with accuracy, precision, don't forget, and F1-score. This assessment comes following the education level. The effects display that, of the classifiers, the Random woodland achieves the very best accuracy of 82%, outperforming the decision Tree and KNN classifiers, which both obtain accuracies of 79%. This suggests that due to its superior overall performance, which emphasises its capability to extra accurately pick out destiny consumer churn, the Random Forest classifier is appropriate to be used in the telecom sector for the motive of churn prediction.

A. Decision Tree Classification Technique

The Decision Tree classifier is applied to the look at paper to predict telecom client turnover. Want trees are widely used supervised learning techniques for categories. With centre nodes standing in for characteristic choices and leaf nodes for elegance labels, they iteratively divided the dataset into subgroups depending on the most significant tendencies to create a tree-like structure. The interpretability and simplicity of Decision Trees make them perfect for churn prediction. By discovering the most important dataset properties, they illuminate churn reasons. Decision Trees can analyse telecom consumer data using numerical and categorical data. The telecom dataset is used to educate the decision Tree classifier to are expecting consumer attrition inside the research article. The selection Tree version predicts patron attrition with the aid of analysing tenure, carrier subscriptions, and billing choices. The version's accuracy, precision, keep in mind, and F1-score monitor its churner detection skills. selection trees are essential to developing telecom patron churn prediction fashions.

B. *Random Forest Classification Technique*

Random Forest classifier is applied in the telecom field purchaser churn prediction article. Several decision trees are used by the Random Forest ensemble getting to know approach to generate a credible prediction model. Random forest decision trees make different forecasts when they are proficient with a part of the dataset. The very final forecast is obtained via voting or tree forecast averaging. Ideally suited for churn prediction, random forest classifiers can manage overfitting, missing values, and high-dimensional date. This work employs the Random Forest classifier to anticipate customer attrition based on tenure, service subscriptions, and billing choices. Train the model using the telecom dataset, and evaluate its F1-score, accuracy, precision, and recall. Telecom customer churn prediction can be used to the 82% accuracy of the Random Forest classifier in predicting probable churners. The approach of the research work is based on the Random Forest classifier, which contributes to the provision of precise and trustworthy prediction models for telecom customer attrition.

C. *KNN Classification Technique*

In this paper, telecom customer churn is predicted using the K-Nearest Neighbours (KNN) classifier. Simple yet effective, KNN is a categorization supervised learning technique. A class label depending on the majority class is applied to the K nearest data points in feature space to a sample. Churn prediction unearths KNN classifiers to be ideal because of their non-parametric nature and capability to comprehend nearby statistics traits. Their dislike of noisy statistics and difficult decision-making procedures qualify them for assessing many facets of telecom buyer information. In this paper, customer attrition is predicted using the KNN classifier based on tenure, service subscriptions, and billing choices. Train the model on the telecom dataset, and evaluate its recall, accuracy, precision, and F1-score. Telecom customer churn prediction is improved and future churners are detected with 79% accuracy by the KNN classifier. The KNN classifier is essential to the approach of the research work since it can predict telecom customer turnover simply and efficiently.

V. RESULTS

Three classifiers—Decision Tree, Random Forest, and K-Nearest Neighbours (KNN)—are assessed in the findings part of the study article on telecom customer churn prediction. The models reach, in turn, accuracy of 79%, 82%, and 79%. Interestingly, with an accuracy of 82%, the Random Forest classifier seems to be the best approach in predicting possible customer attrition . These consequences underline the want present day correctly predicting churn in the telecom enterprise by making use of cutting edge machine learning methods present day strategies. furthermore improving our information of the way well each classifier approaches the problem trendy patron turnover are the performance metrics accuracy, don't forget, and F1-score.

A. *Distribution Plot Analysis*

To give more understanding of the distribution of crucial traits between churned and non-churned clients, a Distribution Plot evaluation is finished within the findings a part of the studies have a look at on telecom customer churn prediction. substantial versions are shown via this study within the distribution styles of parameters like tenure, month-to-month prices, and total expenses between the two corporations. More precisely, compared to non-churned clients, churned customers often have shorter tenures and higher monthly fees. these effects emphasise how vital those characteristics are to differentiating among churners and non-churners and emphasise their potential as critical signs of client churn. Such Distribution Plot analysis enhances knowledge of the capabilities of the dataset and courses the introduction of extra precise churn forecast models in the telecom area as shown in Fig. 2.

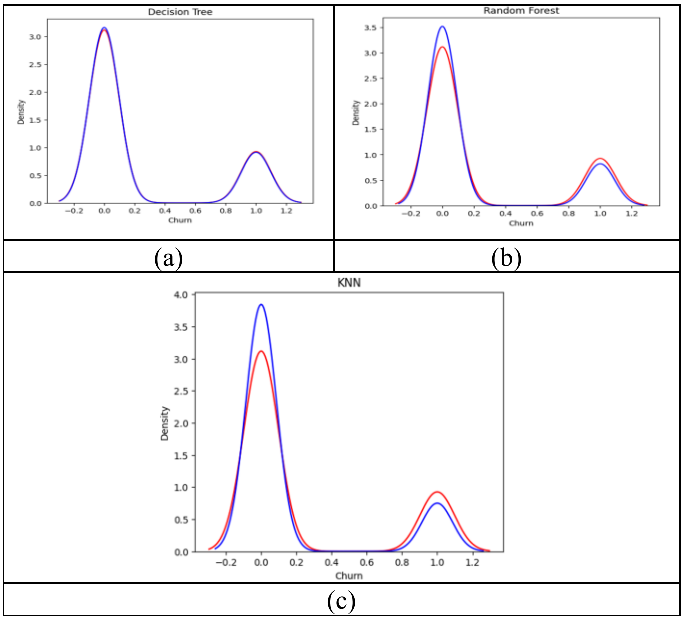


Fig. 2 Distribution Plot Analysis for (a) Decision Tree (b) Random Forest (c) KNN classifier respectively

B. Classification Report Analysis

(a) Decision Tree

The Desicion Tree class record analysis presents accuracy, consider, and f1-score information for churned (1) and non-churned (0) clients. class zero (non-churned) has a precision of 0.86, that means 86% of cases are well categorized. class 1 (churned) has 0.54 accuracy, which means 54% of instances are well recognized.Class 0 recall is 0.87, suggesting that 87% of non-churned cases are properly identified, whereas class 1 recall is 0.53, indicating 53%. Class 0 has a f1-score of 0.86 and class 1 0.54, balancing accuracy and recall. The Decision Tree classifier detects 79% of examples with an accuracy of 0.79. The classifier's macro average f1-score is 0.70, indicating its performance across both classes. The weighted average f1-score, which accounts for class imbalance, is 0.79. These metrics assess the Decision Tree classifier's customer churn prediction performance As shown in Fig. 3

Decision Tree Classification Report:

	precision	recall	f1-score	support
0	0.86	0.87	0.86	1081
1	0.54	0.53	0.54	321
accuracy			0.79	1402
macro avg	0.70	0.70	0.70	1402
weighted avg	0.79	0.79	0.79	1402

Fig. 3 Decision Tree Classification Report Analysis

(b) Random Forest

The Random Forest classification report analysis shows accuracy, recall, and f1-score metrics for churned (1) and non-churned (zero) customers. elegance 0 (non-churned) has a precision of 0.86, meaning 86% of cases are well labeled. elegance 1 (churned) has a precision of zero.63, that means sixty three% of instances are properly identified.. Class 0 recall is 0.91, reflecting 91% of actual non-churned instances properly categorised, whereas class 1 recall is 0.52, reflecting 52% of actual churned instances correctly classified. Class 0 has a f1-score of 0.89 and class 1 0.57, balancing accuracy and recall. The Random Forest classifier detects 82% of occurrences with an accuracy of 0.82. The classifier's macro average f1-score is 0.73, indicating its performance across both classes. The weighted average f1-score, which accounts for class imbalance, is 0.81. These metrics assess the Random Forest classifier's customer churn prediction performance as shown in Fig. 4.

Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.86	0.91	0.89	1081
1	0.63	0.52	0.57	321
accuracy			0.82	1402
macro avg	0.75	0.71	0.73	1402
weighted avg	0.81	0.82	0.81	1402

Fig. 4 Random Forest Classification Report Analysis

(c) KNN

The KNN classification report analysis provides accuracy, take into account, and f1-score statistics for churned (1) and non-churned (0) customers. class zero (non-churned) has a precision of zero.eighty four, which means eighty four% of instances are nicely recognized. class 1 (churned) has a precision of zero.56, which means fifty six% of instances are well categorised. Recall for class 0 is 0.91, meaning that 91% of non-churned instances are properly identified, while for class 1, 0.40, 40% are. Class 0 has a f1-score of 0.87 and class 1 0.47, balancing accuracy and recall. KNN classifier accuracy is 0.79, meaning it classifies 79% of examples correctly. The classifier's macro average f1-score is 0.67, indicating its performance across both classes. The weighted average f1-score, which accounts for class imbalance, is 0.78. These metrics assess the KNN classifier's customer churn prediction performance as shown in Fig. 5.

KNN Classification Report:

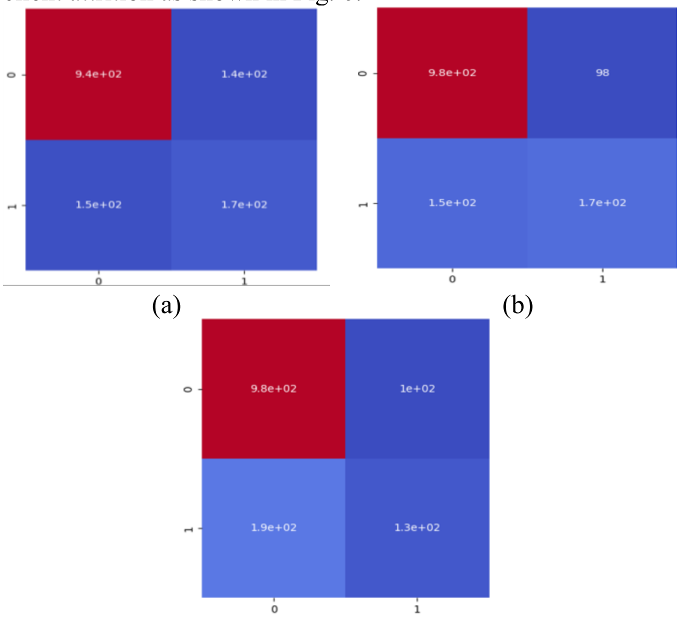
	precision	recall	f1-score	support
0	0.84	0.91	0.87	1081
1	0.56	0.40	0.47	321
accuracy			0.79	1402
macro avg	0.70	0.65	0.67	1402
weighted avg	0.77	0.79	0.78	1402

Fig. 5 KNN Classification Report Analysis

C. Confusion Matrix Analysis

The confusion matrix research shows that Decision Tree, Random Forest, and KNN classifiers predict customer turnover differently. To assess model accuracy and error rates, each matrix shows the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. Decision Tree classifiers perform well in distinguishing clients who have left and those who have not using balanced TP-TN distributions. However, misclassifications increase FP and FN rates. Higher TP and TN counts imply better Random Forest classifier performance in identifying clients who departed and those who did not.

The KNN classifier may misclassify non-churned customers due to its higher FN rates. These insights help choose classifiers by highlighting their pros and cons in predicting client attrition as shown in Fig. 6.



(c)

Fig. 6 Confusion Matrix Analysis for (a) Decision Tree (b) Random Forest (c) KNN classifier respectively

VI. CONCLUSION

In this work, we anticipated client attrition inside the telecom quarter using three state-of-the-art device studying categorization methods: Decision Tree, Random Forest, and K-Nearest Neighbours (KNN). We found that, with an accuracy of eighty-two%, the Random woodland classifier executed better than the other models, with the choice Tree and KNN classifiers each achieving an accuracy of seventy-nine%. The results show that machine learning models may be used to forecast customer turnover rather well; the Random Forest classifier is the most dependable tool for this purpose. Telecom firms may obtain important insights into consumer behaviour and more precisely identify at-risk clients as well as create focused plans to improve customer retention by using the predictive capabilities of these models. A good way to beautify model overall performance even further, destiny study may additionally have a look at integrating different traits and information assets. extra proactive and dynamic consumer interaction processes is probably made possible by way of telecom groups receiving timely alerts from these fashions implemented in real-time structures. The findings of this observe emphasise how important machine studying is to solving the urgent problem of patron attrition, which in flip allows the telecom area preserve steady organization growth and consumer happiness.

REFERENCES

[1] Ouf, S., Mahmoud, K.T. and Abdel-Fattah, M.A., 2024. A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry. Journal of Big Data, 11(1), pp.1-27.

[2] Saha, S., Saha, C., Haque, M.M., Alam, M.G.R. and Talukder, A., 2024. ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry. IEEE Access.

[3] Wagh, S.K., Andhale, A.A., Wagh, K.S., Pansare, J.R., Ambadekar, S.P. and Gawande, S.H., 2024. Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization, 14, p.100342.

[4] Das, D. and Mahendher, S., 2024. Comparative Analysis Of Machine Learning Approaches In Predicting Telecom Customer Churn. Educational Administration: Theory and Practice, 30(5), pp.8185-8199.

[5] Wahul, R.M., Kale, A.P. and Kota, P.N., 2023. An ensemble learning approach to enhance customer churn prediction in telecom industry. International Journal of Intelligent Systems and Applications in Engineering, 11(9s), pp.258-266.

[6] Amin, A., Adnan, A. and Anwar, S., 2023. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. Applied Soft Computing, 137, p.110103.

[7] Agarwal, M., Gill, K.S., Chauhan, R., Kapruwan, A. and Banerjee, D., 2024, March. Classification of Network Security Attack using KNN (K-Nearest Neighbour) and Comparison of different Attacks through different Machine Learning Techniques. In 2024 3rd International Conference for Innovation in Technology (INOCON) (pp. 1-7). IEEE.

[8] Babatunde, R., Abdulsalam, S.O., Abdulsalam, O.A. and Arowolo, M.O., 2023. Classification of customer churn prediction model for telecommunication industry using analysis of variance. IAES Int J Artif Intell, 12(3), pp.1323-9.

[9] Kavitha, V., Kumar, G.H., Kumar, S.M. and Harish, M., 2020. Churn prediction of customer in telecom industry using machine learning algorithms. International Journal of Engineering Research & Technology (2278-0181), 9(05), pp.181-184.

[10] Mittal, K., Gill, K.S., Chauhan, R., Singh, M. and Banerjee, D., 2023, December. Detection of Phishing Domain Using Logistic Regression Technique and Feature Extraction Using BERT Classification Model. In 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) (pp. 1-5). IEEE.

[11] Saheed, Y.K. and Hambali, M.A., 2021, October. Customer churn prediction in telecom sector with machine learning and information gain filter feature selection algorithms. In 2021 International Conference on Data Analytics for Business and Industry (ICDABI) (pp. 208-213). IEEE.

[12] Beeharry, Y. and Tsokizep Fokone, R., 2022. Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. Concurrency and Computation: Practice and Experience, 34(4), p.e6627.

[13] Bilal, S.F., Almazroi, A.A., Bashir, S., Khan, F.H. and Almazroi, A.A., 2022. An ensemble-based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry. PeerJ Computer Science, 8, p.e854.

[14] Labhsetwar, S.R., 2020. Predictive analysis of customer churn in telecom industry using supervised learning. ICTACT Journal on Soft Computing, 10(2), pp.2054-2060.