Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods

Muteb Zarraq Alotaibi
Department of Computer Science, College of Computer and Information Science, Majmaah University, 11952, Saudi Arabia
441104767@s.mu.edu.sa

Mohd Anul Haq
Department of Computer Science, College of Computer and Information Science, Majmaah University, 11952, Saudi Arabia
m.anul@mu.edu.sa (corresponding author)

Received: 14 April 2024 | Revised: 23 April 2024 | Accepted: 25 April 2024

 DOI: https://doi.org/10.48084/etasr.7480

ABSTRACT

This study investigates customer churn, which is a challenge in the telecommunications sector. Using a dataset of telecom customer churn, multiple classifiers were employed, including Random Forest, LGBM, XGBoost, Logistic Regression, Decision Trees, and a custom ANN model. A rigorous evaluation was conducted deploying cross-validation techniques to capture nuanced customer behavior. The models were optimized by hyperparameter tuning, improving their customer churn prediction results. An ensemble averaging method was also adopted, achieving an accuracy of 0.79 and a recall of 0.72 in the test data, which was slightly lower than that of the LGBM, XGBoost, and Logistic Regression. These findings contribute to the development of more reliable churn prediction models to ameliorate the customer retention rates and the operational performance of the service providers.

Keywords-customer churn; XG-Boost; ensemble method; logistic regression; deep learning

I. INTRODUCTION

During the past 20 years, the telecommunications sector has seen substantial changes, moving from simple voice communication to a wide range of digital services, such as mobile data and the Internet [1]. With hopes of lower latency, higher speeds, and the advent of new technologies, like the Internet of Things (IoT), the 5G technology represents a significant turning point [2]. In the highly competitive telecommunications industry, customer retention is becoming increasingly important as a business strategy. Since client acquisition expenses are claimed to be five times higher than retention costs, it is sometimes more cost-effective to retain current customers than to attract new ones [3]. In addition, devoted consumers are more inclined to make larger purchases and act as brand ambassadors enhancing a company's image. Customer churn is a major issue in the ever-changing telecommunications sector since it has a direct effect on long-term viability and profitability. This study follows sophisticated ensemble techniques, including Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoot), to address the important problem of predicting customer churn. These techniques, well-known for their precision in managing intricate and substantial datasets, were implemented in a telecommunications dataset to decipher the subtleties of customer churn. Through rigorous feature selection, hyperparameter tuning, and model comparison, these models can go beyond simple prediction and provide practical guidance to create strategies that effectively retain customers. Customer churn is defined as the termination of a customer's connection to a telecommunication company [4-5]. There may be several causes for this, including shifting client requirements, rivals offering better deals, or unhappiness with services [6]. Since customer behavior varies so much, predicting churn is difficult, and conventional statistical techniques often fail to capture the subtleties required for precise predictions [7]. Churn prediction is made more difficult by the dynamic nature of the telecom sector, which is characterized by the continuously changing offerings and consumer expectations [8]. Given these difficulties, there is an increasing demand for sophisticated prediction techniques to examine large datasets and spot trends that point to churn risk. Multiple ensemble model approaches have shown the potential to increase prediction accuracy [9]. Ensemble approaches, such as Random Forest, Gradient Boosting, and XGBoost can build on the advantages of individual prediction models and minimize their disadvantages. These methods provide a more detailed and sophisticated examination of consumer information, excelling in managing the intricate and nonlinear connections in client behavior patterns [10]. Ensemble approaches also improve prediction resilience, reduce overfitting, and enhance generalizability to fresh, untested data. This attribute is essential in the rapidly changing telecom sector, where consumer preferences and market dynamics can change suddenly. By focusing on aggregating several predictive models to ameliorate accuracy and resilience, ensemble ML techniques constitute a paradigm shift away from single-model approaches. These techniques, including bagging, boosting, and stacking among others, combine the advantages of several different individual models to upgrade the overall prediction performance [11]. They operate on the tenet that a collection of weak learners may become a strong one, producing stronger and more reliable predictions. Ensemble approaches can be applied for churn prediction in the telecom industry, as they are suited to handle this complexity, given the industry's dependence on large and diverse datasets [12].

II. RELATED WORKS

In [13], DT, XGBoost, RF, and Gradient Boosting Machine (GBM) were put into service to predict customer attrition or turnover in the telecommunication industry, with XGBoost yielding the most effective results. The algorithms were trained and tested on a Syriatel dataset. This study also added Social Network Analysis (SNA) elements to the prediction model, increasing the Area Under Curve (AUC) score from 84 to 93.3%. In [14], a hybrid GA-XGBoost model was engaged along with the SHAP framework for interpretation to address the problem of client turnover in the telecom industry. A genetic algorithm was applied to optimize the model and then compare it against traditional ML methods, such as GBDT, KNN, DT, and XGBoost. In [15], churn prediction was examined in one of Indonesia's leading fixed broadband companies deploying ensemble learning approaches, like RF and XGBoost. In addition, an application was designed to assist decision-makers in designing targeted retention strategies, increasing the effectiveness of Customer Relationship Management (CRM) systems, and improving product and service quality. The optimized XGBoost algorithm achieved the best results, with 98.82% accuracy and 87.48% recall rate. This study also included a risk-based approach to measure client attrition, which allows an organization to prioritize its retention efforts. In [16], a hybrid firefly algorithm was proposed for churn prediction on huge telecom data, outperforming the normal Firefly method with an accuracy of 86.38%. In [17], the GA-XGBoost model significantly outperformed other approaches, and SHAP values were utilizefd to highlight the relevance of features and how they contribute to the model. In [18], interesting future research directions were proposed, such as incorporating game theory to increase accuracy. In [19], a neural network-based method was suggested to predict customer attrition, achieving an accuracy of more than 92%. In [20], a two-step model was recommended to forecast attrition in the prepaid telecom sector, dividing the client pool into four distinct categories according to Recency, Frequency, and Monetary (RFM) features to obtain a consistent definition of churn for this specific market, and then using four algorithms to create predictive models for each cluster. This study shed light on how to deal with the challenges of churn in a noncontract-based market such as the prepaid mobile telecommunications.

TABLE I. SUMMARY OF RELATED WORKS

Ref.	Method	Application	Dataset	Evaluation Metric	Metric Value	Limitation
[13]	DT, RF, GBM, XGBoost	Syriatel	Syriatel customer data	AUC	93.3%	Specific to Syriatel, lacks comparative analysis
[14]	GA-XGBoost, SHAP	General telecom companies	Provincial telecom operator	Not explicitly mentioned	Original dataset: 96.12%, New dataset: 98.09%	Not explicitly mentioned
[15]	XGBoost, LR, DT, NB	General telecom companies	Original and newly engineered features	Not explicitly stated	Original dataset: 96.12%, New dataset: 98.09%	Not explicitly mentioned
[16]	Hybrid Firefly algorithm	Telecom churn prediction	Not explicitly mentioned	ROC, PR, F-score, accuracy	Accuracy: 86.38%	Future research directions mentioned, no explicit limitations
[17]	Neural Network	Customer churn prediction	2,427 customers, 20 variables	Not explicitly mentioned	Accuracy > 92%	Not explicitly mentioned
[19]	Clustering & Classification	Prepaid telecom churn	4 clusters (RFM features)	Gain measure	DT > NN (top 10, 20%)	Varying Decision Tree performance
[20]	LR, NB, DT, Multilayer Perceptron, SVMs, etc.	Land-line telecom churn	Various feature subsets	ROC curves, AUC	SVM and C4.5 were best. AUC for new features higher	Computational cost for MLP and DMEL
[21]	SGD, RF, GB, AdaBoost, Stack	Customer churn prediction	Orange telecom churn dataset	AUC, accuracy, F1-score, precision, recall	Stack: AUC 92.45%, accuracy 97.65%, F1-score 96.98%	Limited to the telecom churn dataset, lacks industry-specific factors
[22]	33 classifiers (single, homogeneous, and heterogeneous ensembles)	Customer churn prediction	11 different datasets	AUC, accuracy, F1, top-decile lift	Heterogeneous ensemble with simulated annealing highest in AUC	Benchmark study, not application-specific
[23]	GAMensPlus	Customer churn prediction	Six real-life projects	AUC, top-, lift index, decile lift, accuracy	Competitive but not specified	Not explicitly stated
[24]	XGBoost, LightGBM, CatBoost, Hybrid Resampling	Customer churn prediction	Telecom churn dataset	ROC-AUC, F1-score	Highest F1-score with XGBoost and SMOTE Tomek links	Single dataset used, generalizability to be confirmed
[25]	RF, XGBoost	Fixed broadband company in Indonesia	Company customer data	Accuracy, recall	Accuracy: 98.82%, Recall: 87.48%	Lacks competitor categorization, suggests advanced methods for future research

In [21], four classification algorithms were compared to predict customer loss in the telecommunications industry, displaying that GB achieved the best results. In [22], a detailed benchmark was performed on customer churn prediction, comparing 33 classifiers across 11 different datasets. Heterogeneous ensembles consistently outperformed homogeneous ensembles and single classifiers. This study performed virtual annealing algorithm selection among the best to determine a heterogeneous ensemble in terms of AUC and predicted maximum revenues. In [23], a novel ensemble classifier, called GAMensPlus, was proposed to forcast client churn. This model was based on GAMs and included generalized feature importance ratings and bootstrapping credibility zones to smooth splines. The model was evaluated on six real-world churn prediction projects and outperformed various benchmarks, including Bagging, RSM, RF, and LR. This study concluded that the GAMensPlus model not only enables effective identification of churners, but also aids in understanding the sources of churn, improving its organizational acceptance. In [24], recent ML techniques were followed to predict client retention. This study deployed hybrid resampling techniques, entailing SMOTE Tomek Links and SMOTE-ENN, as well as collaborative learning approaches, such as XGBoost, LightGBM, and CatBoost. The results disclosed that the ensemble learning algorithms combined with hybrid resampling approaches outperformed standard methods. In [25], SMOTE was used with XGBoost to classify customers into four churn risk clusters.

III. METHODOLOGY

This study explores the potential applications of ensemble approaches for customer attrition prediction in the telecom industry by employing and evaluating XGBoost, GB, RF, and other ensemble techniques. The goal is to comprehend how these techniques may be utilized to precisely forecast churn, which would help telecom companies create more successful customer retention plans. This study also explores the subtleties of various approaches, weighing their advantages and disadvantages in light of the telecom industry's dynamic and data-intensive environment. Figure 1 portrays the method followed.

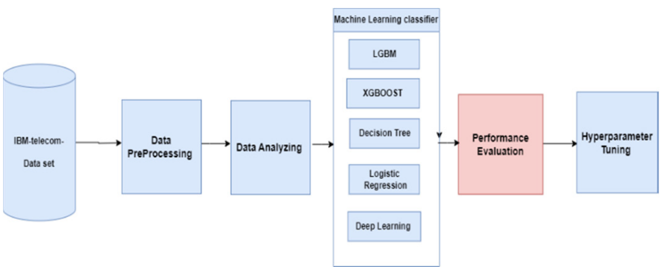

Fig. 1. Diagram of the method followed.

A. *Dataset Description*

Despite the efforts to obtain authentic data from local sources within Saudi Arabia in partnership with telecom companies, this could not be achieved due to privacy policy. A highly relevant dataset was acquired from IBM, called "The Telco Customer Churn Data," which pertains to a fictional telecommunications company that provided home phone and internet services in California. While the dataset's origin differs from the target region, its relevance and availability make it a suitable choice to develop and evaluate predictive models for customer churn in the telecom sector. This dataset manifests which customers have left, stayed, or signed up for their service. It consists of 7,043 rows (customers) and 21 columns (features), with 3 columns containing numeric values and 18 columns containing categorical values.

1) *Dependent Variable*

Churn: Indicates customers who have recently discontinued the service.

2) *Independent Variables*

- Demographic data include the gender, age range, partner, and dependent status of customers.
- Services subscribed to by each customer, such as phone service, multiple lines, internet service, online security, and online backup.
- Customer account details: represent attributes related to customer behavior, namely payment method, preferences for paperless billing, monthly charges, and total charges.

Table II outlines the specifics of these features.

TABLE II. DATASET DESCRIPTION

#	Feature	Description
1	Gender	Whether the customer is male or female
2	Senior Citizen	Whether the customer is a senior citizen (Yes/No)
3	Partner	Whether the customer has a partner (Yes/No)
4	Dependents	Whether the customer has dependents (Yes/No)
5	Tenure	Number of months the customer has stayed with the company
6	Phone Service	Whether the customer has phone service (Yes/No)
7	Multiple Lines	Whether the customer has multiple lines (Yes/No or No phone service)
8	Internet Service	Customer's internet service provider (DSL, Fiber optic, No)
9	Online Security	Whether the customer has online security (Yes/No or No internet service)
10	Online Backup	Whether the customer has online backup (Yes/No or No internet service)
11	Device Protection	Whether the customer has device protection (Yes/No or No internet service)
12	Tech Support	Whether the customer has tech support (Yes/No or No internet service)
13	Streaming TV	Whether the customer can access streaming TV services (Yes/No or No internet service)
14	Streaming Movies	Whether the customer can access streaming movie services (Yes/No or No internet service)
15	Contract	Indicates the type of contract (month-to-month, one year, two years)
16	Paperless Billing	Indicates whether customers opted for paperless billing (Yes/No)
17	Payment-Method	Indicates the payment method (electronic check, bank transfer (automatic), credit card (automatic)
18	Monthly Charges	Indicates the total monthly subscription cost for customers
19	Total-Charges	Indicates total charges paid by customers so far
20	Churn	Indicates whether customers have churned or not

B. Data Preprocessing

Data preprocessing is essential to clean, organize, and prepare data for ML models. Without it, the system's performance can be affected negatively due to inaccurate results, model instability, overfitting or underfitting, reduced model performance, and increased computational inefficiency. Data preprocessing involves the following four key steps.

1) Data Cleaning

The dataset underwent thorough cleaning to ensure its suitability for modeling. Null values were identified and addressed using Pandas' isna().sum() method. Specifically, missing values in the 'TotalCharges' column were filled with the median value using fillna(). The 'TotalCharges' column was also converted to a numeric data type after addressing non-numeric entries and handling the null values discovered in the 'total charges' column.

2) Irrelevancy

Irrelevant features were identified and removed. Features such as 'Row number' and 'customer ID' were considered irrelevant and removed. This step streamlined the dataset and optimized classifier performance.

3) Transformation

Data transformation is vital for refining and structuring data to improve their quality and compatibility within applications. Specific conversions were performed for important categorical features, such as the 'Gender' feature, mapping 'Female' to 0 and 'Male' to 1. Such transformations ensure data uniformity and coherence and protect against potential issues, such as null values, duplicates, incorrect indexing, and incompatible formats. These steps were applied for 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProte', 'StreamingTV', 'StreamingMovies', 'Contract', and 'PaperlessBilling'.

C. Exploratory Data Analysis

Regarding the social aspect of the customers, the gender distribution and the senior citizen field were analyzed. Half of the customers were male and the other half were female. Interestingly, 16% of the customer base comprises senior citizens. Figure 2(a,b) shows the distribution of customers according to gender and age. Considering the age of the customers, it is worth exploring whether it affects their decision to discontinue their relationship with the company. Figure 2 (c) illustrates the distribution of contract durations among customers, revealing that a predominant proportion opts for month-to-month contracts. Interestingly, the data indicates a parity between customers selecting 1- and 2-year contracts, with both options attracting an equal number of subscribers. Figure 2 (d) exhibits the distribution of individuals within the sample based on their partnership with yes and no representing 48.3 and 51.7%, respectively.

D. Model Building

Five ML algorithms were employed to analyze the preprocessed data in the model-building phase. An ensemble method combines multiple ML models to improve overall performance and predictive accuracy. By leveraging the strengths of different algorithms, ensemble methods can mitigate individual model weaknesses and capture complex patterns in the data more effectively. This study used the DT, LR, DL, LBGM, and XGBoost models. After applying these algorithms, a comparative analysis of the results was performed. Additionally, the performance and accuracy of these algorithms were optimized by hyperparameter tuning.

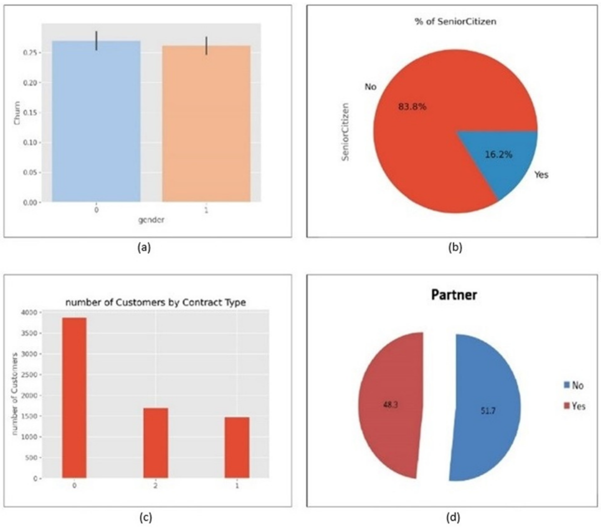

Fig. 2. Customer details: (a) gender, (b) senior citizens, (c) contract durations, and (d) partner rates.

1) The Decision Tree (DT) Algorithm

This statistical model is implemented in supervised learning for classification and regression tasks. It classifies data and represents the results in a flowchart in the form of a tree structure. The DT model flows through a query structure from the root until it reaches the leaf, representing one class.

2) Logistic Regression (LR)

This is a popular supervised learning algorithm that predicts the categorical dependent variable deploying a given set of independent variables.

3) Extreme Gradient Boosting (XGBboost) Model

This is a scalable, distributed Gradient-Boosted Decision Tree (GBDT) ML library that provides parallel tree boosting. It is the leading ML library for regression, classification, and ranking problem trees by combining predictions from multiple trees.

4) Light-GBM (LGBM)

This is a gradient-boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient, with faster training speed, higher efficiency, lower memory usage, better accuracy, support for parallel and GPU learning, and the ability to handle large-scale data.

5) ANN-based Model

The ANN-based model was constructed using the Sequential API provided by TensorFlow and Keras. It had three densely connected layers: two hidden layers with 64 and 32 units respectively, both activated by the Rectified Linear Unit (ReLU) function, and an output layer with a single unit activated by the sigmoid function.

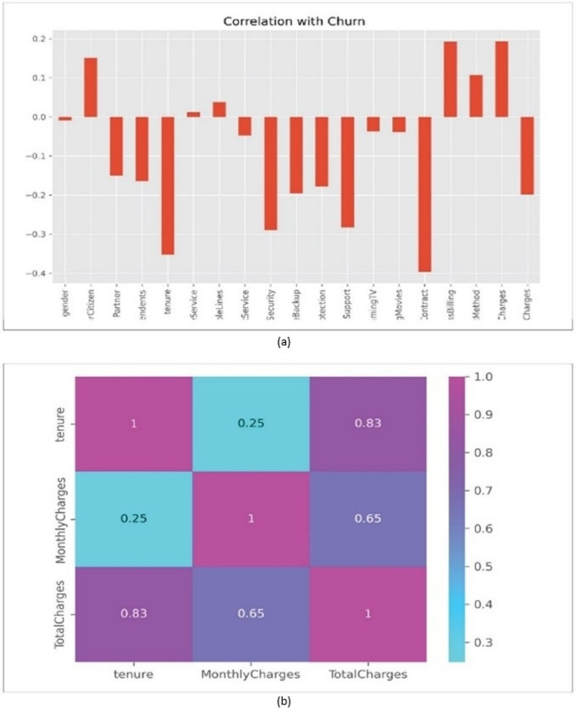

Fig. 3. Correlation coefficient values between (a) categorical and (b) numerical variables.

IV. RESULTS AND DISCUSSIONS

After preprocessing, four ML models were trained on the dataset utilizing 75% of the refined dataset. The remaining 25% was used for testing. The accuracy of each model was evaluated employing a 10-fold cross-validation process. The models were also optimized applying hyperparameter tuning. An ANN model was built and compared with an ensemble averaging method of the ML models. The DT model had an accuracy of 0.78 and a recall of 0.64. The LGBM classifier had an accuracy of 0.80 and a recall of 0.65. The XGBoost algorithm had an accuracy of 0.78 and a recall of 0.69. The LR model had an accuracy of 0.80 and a recall of 0.71. The ANN model was trained for 30 epochs with a batch size of 32 and a validation split of 20%. It achieved an accuracy of 0.79 and a recall of 0.72 on the test data, with a threshold of 0.5 used to convert the predicted probabilities to binary predictions.

Hyperparameter tuning was performed deploying GridSearchCV to optimize the result and gain the best accuracy and performance. GridSearchCV exhaustively searches through a grid of hyperparameters to find the best combination, ensuring optimal model performance. Unlike RandomizedSearchCV, which samples a fixed number of combinations, GridSearchCV guarantees finding the best hyperparameters within the specified grid, making it a reliable choice despite being more computationally intensive. Table III depicts the tuned hyperparameters of the ML models along with their best accuracy.

TABLE III. OPTIMIZATION OF ML MODELS USING HYPERPARAMETER TUNING

#	Model	Best parameters	Best accuracy
1	LGBM	Learning rate: 0.01, N estimators: 200	0.80
2	XGBoost	Learning rate: 0.1 Max depth: 3 N estimator: 50	0.80
3	DT	Criterion: entropy Max depth:5	0.78
4	LR	C: 01	0.80

Figure 5 presents a comparison of the accuracy of the ML models before and after tuning. Ensemble learning approaches can be applied to various other applications beyond telecommunications. In the present investigation, the performance of each model was individually assessed. Employing averaging was chosen as the ensemble model, resulting in an accuracy of 0.795, indicative of its lower predictive performance, although capturing a broader range of patterns and generalizations. Rigorous evaluation and hyperparameter tuning optimized the models, demonstrating their effectiveness in improving customer retention rates in the telecommunications sector. However, limitations include the study's reliance on a specific dataset, potentially limiting generalizability, and the computational complexity of the ensemble method, which could pose challenges in real-time applications. Future research could address these limitations by testing the method on diverse datasets and streamlining its implementation for practical use. The future scope of this work is to apply advanced and robust models [26, 27] with a similar objective of comparing and optimizing models for various aligned areas [28-35].

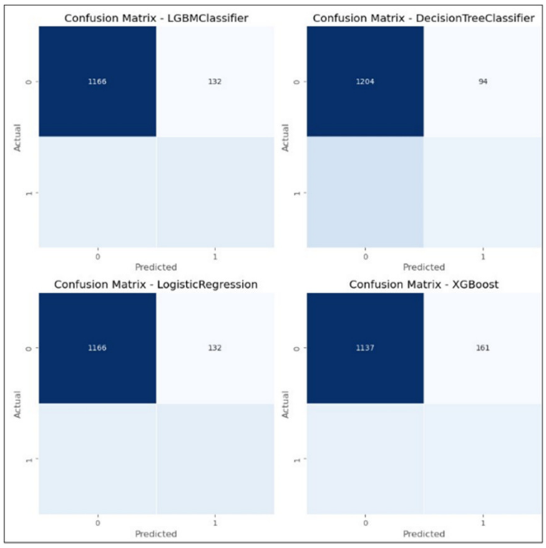

Fig. 4. Confusion matrix of ML models.

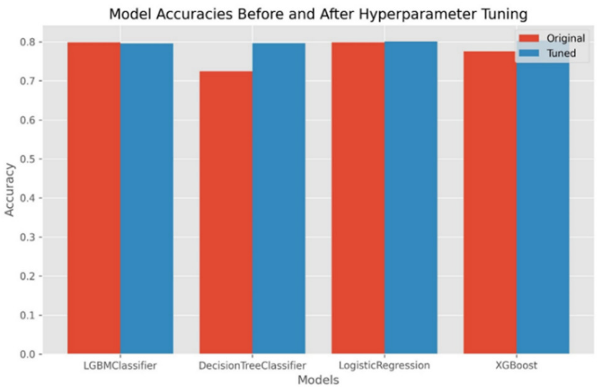

Fig. 5. Models compared before and after tuning.

V. CONCLUSION

Customer retention is crucial in the telecommunications sector, highlighting the need for early-stage churn prediction. This study developed an optimized model using various classifiers and oversampling techniques to address this challenge. Despite the dataset's modest size and imbalance, the XGBoost classifier, coupled with oversampling, achieved the highest accuracy of 80.366%. When comparing the models, the ensemble models XGBoost and LGBM outperformed the ANN model developed in this context. The findings highlight the importance of targeted retention strategies, especially for customers with month-to-month contracts and those without partners, offering valuable insights for telecom companies to improve customer loyalty. Future research could explore the potential of blockchain-based solutions to securely manage and share telecom customer data across networks [36-39].

ACKNOWLEDGEMENT

Mohd Anul Haq would like to thank the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for supporting this work under Project No. R-2024-1076.

REFERENCES

[1] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2354–2364, Mar. 2011, https://doi.org/10.1016/j.eswa.2010.08.023.

[2] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2354–2364, Mar. 2011, https://doi.org/10.1016/j.eswa.2010.08.023.

[3] A. Amin *et al.*, "Customer churn prediction in the telecommunication sector using a rough set approach," *Neurocomputing*, vol. 237, pp. 242–254, May 2017, https://doi.org/10.1016/j.neucom.2016.12.009.

[4] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining," *International Journal of Computer Science Issues*, vol. 10, no. 2, pp. 165–172, Mar. 2013.

[5] M. H. Dunham, *Data mining: Introductory and advanced topics*. Upper Saddle River, NJ, USA: Prentice Hall, 2006.

[6] F. Thabtah, "A review of associative classification mining," *The Knowledge Engineering Review*, vol. 22, no. 1, pp. 37–65, Mar. 2007, https://doi.org/10.1017/S0269888907001026.

[7] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. Cambridge, MA, USA: Morgan Kaufmann, 2022.

[8] K. Coussement and K. W. De Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning," *Journal of Business Research*, vol. 66, no. 9, pp. 1629–1636, Sep. 2013, https://doi.org/10.1016/j.jbusres.2012.12.008.

[9] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, Nov. 2015, https://doi.org/10.1016/j.ejor.2015.05.030.

[10] T. Verbraken, W. Verbeke, and B. Baesens, "A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 5, pp. 961–973, Feb. 2013, https://doi.org/10.1109/TKDE.2012.50.

[11] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, Nov. 2007, https://doi.org/10.1016/j.eswa.2006.07.007.

[12] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, Apr. 2009, https://doi.org/10.1016/j.eswa.2008.06.121.

[13] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *Journal of Big Data*, vol. 6, no. 1, Mar. 2019, Art. no. 28, https://doi.org/10.1186/s40537-019-0191-6.

[14] K. Peng and Y. Peng, "Research on Telecom Customer Churn Prediction Based on GA-XGBoost and SHAP," *Journal of Computer and Communications*, vol. 10, no. 11, pp. 107–120, Nov. 2022, https://doi.org/10.4236/jcc.2022.1011008.

[15] T. Xu, Y. Ma, and K. Kim, "Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping," *Applied Sciences*, vol. 11, no. 11, Jan. 2021, Art. no. 4742, https://doi.org/10.3390/app11114742.

[16] A. A. Q. Ahmed and D. Maheswari, "Churn prediction on huge telecom data using hybrid firefly based classification," *Egyptian Informatics Journal*, vol. 18, no. 3, pp. 215–220, Nov. 2017, https://doi.org/10.1016/j.eij.2017.02.002.

[17] A. Sharma and D. P. K. Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," *International Journal of Computer Applications*, vol. 27, no. 11, pp. 26–31, Aug. 2011, https://doi.org/10.5120/3344-4605.

[18] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, Jan. 2012, https://doi.org/10.1016/j.eswa.2011.08.024.

[19] A. T. Jahromi, M. Moeini, I. Akbari, and A. Akbarzadeh, "A Dual-Step Multi-Algorithm Approach for Churn Prediction in Pre-Paid Telecommunications Service Providers," *Journal on Innovation and Sustainability RISUS*, vol. 1, no. 2, 2010, https://doi.org/10.24212/2179-3565.2010v1i2a7.

[20] B. Q. Huang, M. T. Kechadi, and B. Buckley, "Customer Churn Prediction for Broadband Internet Services," in *Data Warehousing and Knowledge Discovery*, Penang, Malaysia, 2009, pp. 229–243, https://doi.org/10.1007/978-3-642-03730-6_19.

[21] R. M. Wahul, A. P. Kale, and P. N. Kota, "An Ensemble Learning Approach to Enhance Customer Churn Prediction in Telecom Industry," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 9s, pp. 258–266, Jul. 2023.

[22] M. Bogaert and L. Delaere, "Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art," *Mathematics*, vol. 11, no. 5, Jan. 2023, Art. no. 1137, https://doi.org/10.3390/math11051137.

[23] K. W. De Bock and D. Van den Poel, "Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models," *Expert Systems with Applications*, vol. 39, no. 8, pp. 6816–6826, Jun. 2012, https://doi.org/10.1016/j.eswa.2012.01.014.

[24] T. Kimura, "Customer Churn Predition with Hybrid Resampling and Ensemble Learning," *Journal of Management Information and Decision Sciences*, vol. 25, no. 1, 2022.

[25] A. Dhini and M. Fauzan, "Predicting Customer Churn using ensemble learning: Case Study of a Fixed Broadband Company," *International Journal of Technology*, vol. 12, no. 5, pp. 1030–1037, Dec. 2021, https://doi.org/10.14716/ijtech.v12i5.5223.

[26] A. Rahman and M. N. A. Khan, "A Classification Based Model to Assess Customer Behavior in Banking Sector," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2949–2953, Jun. 2018, https://doi.org/10.48084/etasr.1917.

[27] E. Jamalian and R. Foukerdi, "A Hybrid Data Mining Method for Customer Churn Prediction," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2991–2997, Jun. 2018, https://doi.org/10.48084/etasr.2108.

[28] M. A. Haq, "DBoTPM: A Deep Neural Network-Based Botnet Prediction Model," *Electronics*, vol. 12, no. 5, Jan. 2023, Art. no. 1159, https://doi.org/10.3390/electronics12051159.

[29] J. Gyani, M. A. Haq, and A. Ahmed, "Analyzing the Impact of Lockdown on COVID-19 Pandemic in Saudi Arabia," *International Journal of Computer Science & Network Security*, vol. 22, no. 4, pp. 39–46, 2022, https://doi.org/10.22937/IJCSNS.2022.22.4.6.

[30] M. A. Haq and A. Ahmed, "On Interesting Correlation between Meteorological Parameters and COVID-19 Pandemic in Saudi Arabia," *International Journal of Computer Science & Network Security*, vol. 22, no. 4, pp. 159–168, 2022, https://doi.org/10.22937/IJCSNS.2022.22.4.20.

[31] M. A. Haq, M. A. R. Khan, and M. Alshehri, "Insider Threat Detection Based on NLP Word Embedding and Machine Learning," *Intelligent Automation and Soft Computing*, vol. 33, no. 1, pp. 619–635, 2022, https://doi.org/10.32604/iasc.2022.021430.

[32] S. Kumar *et al.*, "Multilayer Neural Network Based Speech Emotion Recognition for Smart Assistance," *Computers, Materials & Continua*, vol. 74, no. 1, pp. 1523–1540, 2022, https://doi.org/10.32604/cmc.2023.028631.

[33] A. Alabdulwahab, M. A. Haq, and M. Alshehri, "Cyberbullying Detection using Machine Learning and Deep Learning," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, pp. 424–432, 2023, https://doi.org/10.14569/IJACSA.2023.0141045.

[34] J. Gyani, A. Ahmed, and M. A. Haq, "MCDM and Various Prioritization Methods in AHP for CSS: A Comprehensive Review," *IEEE Access*, vol. 10, pp. 33492–33511, 2022, https://doi.org/10.1109/ACCESS.2022.3161742.

[35] G. Revathy, S. A. Alghamdi, S. M. Alahmari, S. R. Yonbawi, A. Kumar, and M. Anul Haq, "Sentiment analysis using machine learning: Progress in the machine intelligence for data science," *Sustainable Energy Technologies and Assessments*, vol. 53, Oct. 2022, Art. no. 102557, https://doi.org/10.1016/j.seta.2022.102557.

[36] M. Suresh, A. S. Shaik, B. Premalatha, V. A. Narayana, and G. Ghinea, "Intelligent & Smart Navigation System for Visually Impaired Friends," in *Advanced Computing*, Kolhapur, India, 2023, pp. 374–383, https://doi.org/10.1007/978-3-031-35641-4_30.

[37] S. Merugu, K. Jain, A. Mittal, and B. Raman, "Sub-scene Target Detection and Recognition Using Deep Learning Convolution Neural Networks," in *ICDSMLA 2019*, pp. 1082–1101, https://doi.org/10.1007/978-981-15-1420-3_119.

[38] A. Bathula, S. K. Gupta, S. Merugu, and S. S. Skandha, "Academic Projects on Certification Management Using Blockchain- A Review," in *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, Hyderabad, India, Dec. 2022, pp. 1–6, https://doi.org/10.1109/ICMACC54824.2022.10093679.

[39] A. Bathula, S. Muhuri, S. K. Gupta, and S. Merugu, "Secure certificate sharing based on Blockchain framework for online education," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16479–16500, May 2023, https://doi.org/10.1007/s11042-022-14126-x.