THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 2 (2024)
https://doi.org/ 10.62019/abbdm.v4i02.175

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/
ISSN (online): 2959-0809

Data-Driven Decision-Making: Accurate Customer Churn Prediction with Cat-Boost

Bismah Hasan*, Zubair, Shahrukh Ali Shaikh, Abdul Khaliq, Ghalib Nadeem*

Chronicle

Article history

Received: May 31, 2024
Received in the revised format: June 5, 2024
Accepted: June 6, 2024
Available online: June 7, 2024

Bismah Hasan* & Abdul Khaliq are currently affiliated with the Department of Computer Science, Iqra University, Karachi, Pakistan.
Email: bismah@iqra.edu.pk
Email: abdulkhaliq@iqra.edu.pk

Zubair is currently affiliated with the Department of Control Science and Engineering, North China Electric Power University, Beijing, China.
Email: 120214200040@ncepu.edu.cn

Shahrukh Ali Shaikh is currently affiliated with the Department of Computer Science and Information Technology, NEDUET, Karachi, Pakistan.
Email: shahrukhalishaikh100@gmail.com

Ghalib Nadeem* is currently affiliated with the Department of Electrical Engineering and High-Performance Research (HPR) Group, Iqra University, Karachi, Pakistan.
Email: ghalibnadeem@iqra.edu.pk

Corresponding Author*

Abstract

In the cutthroat telecommunications industry, customer churn poses a significant threat to profitability and brand reputation. To combat this, we propose a data-driven approach to predict churn, leveraging advanced Machine Learning (ML) algorithms on a comprehensive dataset of 7043 customer records with 21 features sourced from the IEEE website. By applying one-hot encoding to transform categorical data, we optimize ML algorithm performance. Our study pits three ML algorithms against each other: Cat Boost, Stochastic Gradient Boost, and Extreme Gradient Boost. The results are striking, with Cat Boost incipient as the top performer, boasting an impressive 94% accuracy, 90% precision, 97% recall, 94% F1-score, and an AUC of 0.94. These discoveries emphasize Cat Boost's extraordinary ability to predict churn, empowering telecom companies to proactively prevent attrition, target high-risk customers, and make data-driven decisions to make the most of profits. In today's speedily developing technical landscape, correctly predicting churn is critical to meeting the ever-changing potential of customers. Our research marks a noteworthy innovation in addressing churn trials, flooring the way for long-term accomplishment in the telecom industry.

Keywords: Cat Boost, Customer Retention, Stochastic Gradient Boost, Predictive Analysis.

INTRODUCTION

In today's hyper-competitive telecommunications landscape, customer churn positions as a tough problem for continued success, where technological developments reveal at a quick speed, customer churn ruins an insistent task. The everlasting churn of customers not only imposes substantial economic losses but also smudges the brand status of telecommunications companies, as consumers exert more power than ever before prepared with adoptions, favourites, and the capacity to switch providers with a few clicks telecommunications companies find themselves in high-stakes combat for customer devotion. The costs of losing this battle are terrible: revenue destruction, brand weakening, and a reduction in market share. Identifying the serious need for advanced solutions to contest churn, this paper presents a pioneering method for churn prediction leveraging the influence of data-driven methodologies and cutting-edge machine learning (ML) algorithms (Ahmad, T et al., 2022). The telecommunications industry is an active ecosystem characterized by swift technological developments and developing customer likings. In such a fast-paced environment, companies must stay ahead of the curve to keep market significance and profitability (Rahman, M. M. Et al., 2023). However, amidst fierce struggle and ever-changing consumer behaviors, diminishing customer erosion leftovers an intimidating contest for telecom providers. Conventionally, telecom companies have trusted reactive measures to address churn, often resorting to generic preservation strategies or relying on past data analysis (Tirkkonen, T. Et al., 2021). While such methods may suggest short-term relief, they often fall short of bringing bearable solutions to moderate churn effectively. Moreover, the complete volume and complication of data generated in telecommunications operations make it progressively stimulating to derive actionable understandings using conventional analytical methods alone (Keshavarz, H. Et al., 2021). Against this backdrop, the addition of advanced ML techniques presents a capable avenue for revolutionizing churn prediction and management in the telecommunications sector. By binding the vast troves of customer data available to them, telecom companies can gain invaluable insights into customer behavior, preferences, and predictors of churn (Lottu, O. A et al., 2024). This paper endeavours to discover the transformative potential of ML-driven churn prediction models in permitting telecom providers to proactively recognize and address churn risks before they intensify.

The Churn Conundrum

Imagine a busy call center, where agents climb to hold dissatisfied customers. The churn rate—the percentage of customers who abandon ship—haunts executives' dreams. It's a faint dance: balancing achievement efforts to fascinate new subscribers while loyally holding onto existing ones (Zhao, M. Et al., 2021). But how do you forecast when a customer will raise the white flag and defect to a contestant? Enter data science—a beacon of hope in this sea of uncertainty.

A Data-Driven Approach

Our research embarks on a quest to undo the secrecies of churn. Armed with a dataset of 7043 customer histories, we investigate the details of customer behavior (Mengjing Hao. Et al., 2024). These records disclose more than mere transactions; they hold the pulse of consumer sentiment, preferences, and practice patterns. Our mission: is to build a predictive model that not only recognizes possible churners but also allows providers to take proactive actions (Mengjing Hao. Et al., 2024).

The Orchestra of Features

Our dataset orchestrates a symphony of features—21 in total. These attributes span the gamut, from basic demographics (age, gender, tenure) to service-specific metrics (monthly charges, contract type, and streaming TV usage). But it's not just about the individual notes; it's about the harmonious interplay. We apply one-hot encoding to transform categorical variables into a format that our machine-learning algorithms can outperform (Dahouda, M. K. Et al., 2021). The stage is set for our models to perform directing the players:

- Demographics: Age, gender, and tenure—the basic notes that set the stage.
- Usage Patterns: Monthly charges, total charges, and streaming TV usage—the crescendos that reveal behavior.
- Contract Details: Month-to-month contracts, one-year commitments, and two-year harmonies—the rhythm of commitment.

Key to the success of any ML model is the pre-processing of data to ensure its compatibility with algorithmic requirements. In this regard, the application of one-hot encoding emerges as a critical pre-processing step, facilitating the transformation of categorical data into a format conducive to ML analysis (Khaldi, B. Et al., 2021). By encoding categorical variables as binary vectors, one-hot encoding enables ML algorithms to effectively discern patterns and relationships within the data, thereby enhancing the predictive accuracy of the model (Johnson, J. M. Et al., 2022). But this isn't a solo performance; it's a duet with machine learning algorithms. We apply one-hot encoding, transforming categorical variables into numerical harmonies. The stage is set, the lights dimmed—the models step forward.

The Cat Boost Crescendo

Among the ensemble of algorithms, one stands out—the Cat Boost. Like a virtuoso pianist, it strikes the right chord. With an accuracy of 94%, precision of 90%, recall of 97%, and an F1-score of 94%, it dazzles the audience. The area under the curve (AUC) reaches 0.94—an applause-worthy feat. But this isn't just about numbers; it's about the symphony of insights it conducts.

Among the algorithms, one stands out—a virtuoso named Cat Boost. It's not a feline with musical aspirations; rather, it's a gradient-boosting algorithm (Hancock, J. T. Et al., 2021). Like a pianist caressing ivory keys, Cat Boost learns from mistakes, adjusts its tempo, and produces melodies of prediction. Its metrics read like a symphony program:

Accuracy: 94%—a standing ovation from statisticians.

Precision: 90%—a delicate balance between false positives and false negatives.

Recall: 97%—a safety net for catching potential churners.

F1-score: 94%—the harmonic means of precision and recall.

AUC: 0.94—an encore-worthy performance.

But numbers alone can't evoke emotion. Imagine the Cat Boost as a conductor, guiding providers toward strategic decisions:

Proactive Churn Prevention: Armed with predictions, companies can intervene before the final note. A personalized offer, a retention call—the symphony of prevention.

Targeted Retention Strategies: High-risk customers receive VIP treatment. Loyalty programs, discounts, and heartfelt apologies—the cadence of care.

Data-Driven Baton: Decision-makers no longer rely on gut feelings. They wield data-driven batons, conducting the orchestra of business.

Advantages Beyond the Scorecard

Our model isn't a mere showpiece; it's a strategic asset. Imagine the conductor's baton guiding providers toward proactive churn prevention. Armed with accurate predictions, they can intervene before the curtain falls on a customer relationship. Targeted retention strategies emerge—personalized offers, loyalty programs, and tailored communication. Decision-makers no longer rely on gut feelings; they wield data-driven swords.

The Overture to Success

This research isn't just about algorithms and metrics; it's about the long-term overture to success. By predicting churn accurately, telecommunications providers can:

- Reduce Financial Haemorrhage: Every retained customer is a victory against revenue leakage.
- Foster Loyalty: Loyal customers become brand advocates, amplifying the chorus of positive word-of-mouth.
- Navigate the Shifting Tides: As customer expectations evolve, our model equips providers to adapt swiftly.

Furthermore, this study endeavors to compare and contrast the performance of three prominent ML algorithms - Cat Boost, Stochastic Gradient Boost, and Extreme Gradient Boost - in churn prediction. Through rigorous experimentation and evaluation, we aim to ascertain the algorithm that offers the highest predictive accuracy and reliability in identifying churn risks. Preliminary findings suggest that the Cat Boost algorithm demonstrates superior performance metrics, including accuracy, precision, recall, F1-score, and area under the curve (AUC), thus underscoring its efficacy as a potent tool for churn prediction in telecommunications. Beyond the technical aspects of model development and algorithm selection, this paper also delves into the practical implications of churn prediction for telecommunications providers (Usman-Hamza, F. E. Et al., 2024).

By leveraging ML-driven churn prediction models, telecom companies stand to gain a multitude of strategic advantages, including proactive churn prevention, targeted retention strategies, and data-driven decision-making (Dhanawade, A. Et al., 2023). Armed with the foresight to anticipate and pre-empt customer churn, telecom providers can institute tailored interventions aimed at preserving customer loyalty and maximizing lifetime value. In essence, this paper represents a significant advancement in the field of churn prediction within the telecommunications industry. By pioneering a data-driven approach to churn management and showcasing the efficacy of ML algorithms in this domain (Lalwani, P. Et al., 2022), we aim to equip telecom providers with the tools and insights needed to navigate the complex landscape of customer attrition successfully. Through proactive intervention and strategic decision-making informed by predictive analytics, telecom companies can forge stronger bonds with customers, mitigate churn risks, and secure long-term viability in an increasingly competitive marketplace (Fei, D. Et al., 2021).

LITERATURE REVIEW

The authors (Agha, A. A. Et al., 2021) examined the correlation between customer satisfaction, switching barriers, consumer sentiment, and customer retention. They found a positive association between customer satisfaction and retention. Low satisfaction levels made it clear that customers were more likely to move to a rival after the reasons for customer turnover were examined. Additionally, the study discovered that lowering switching costs can contribute to higher client retention rates. Overall, the study showed that improving customer happiness and lowering barriers to change are critical for long-term client retention. The argument made by (Routh, P. Et al., 2021) was that determining the cause of customer churn is a crucial factor in determining whether or not a win-back strategy can be used to win back customers. Businesses can customize their win-back tactics to target particular problems that initially caused customers to depart by knowing the causes of customer churn. This focused strategy can improve the likelihood of gaining back lost clients and restoring confidence. Businesses may enhance their overall customer retention efforts and develop more successful win-back campaigns by studying the underlying causes of customer churn. Research on churn algorithm and model: the existing research mainly focuses on regression, neural networks, decision trees, and other algorithms. However, there is potential for further exploration in the application of machine learning techniques such as random forests, support vector machines, and deep learning models to improve the accuracy and efficiency of churn prediction. Additionally, incorporating customer behaviour analysis and sentiment analysis could provide valuable insights for predicting and preventing churn in various industries. Further research is needed to compare the effectiveness of different algorithms and models in different business contexts to develop more personalized and effective churn prediction strategies (Geiler, L. Et al., 2022).

A comprehensive analysis of the existing studies shows that, in academic circles, customer churn research is an important issue in customer relationship management, while in management practice, customer churn will bring huge losses to the profits and future development of enterprises. Therefore, businesses must understand the reasons behind customer churn and take proactive measures to retain their customers. By implementing effective strategies such as improving customer service, offering personalized experiences, and analyzing customer feedback, companies can reduce churn rates and ultimately increase their overall success in the market. Ultimately, addressing customer churn not only preserves current revenue streams but also fosters long-term customer loyalty and sustainable growth for the future (Prabadevi, B. Et al., 2023). The industry recognizes the importance of retaining existing customers over acquiring new ones due to the higher costs associated with customer acquisition. Therefore, accurate and efficient churn prediction models are crucial for telecom companies' survival and growth.

METHODOLOGY

Research Design

This study adopts a quantitative research design aimed at empirically investigating the effectiveness of machine learning (ML) algorithms in predicting customer churn within the telecommunications industry. Quantitative research methodologies are well-suited for analysing large datasets and identifying statistical patterns and relationships, aligning with the objectives of this research to develop a robust churn prediction model. Similarly, figure 1 shows the entire flow diagram of the methodology to be followed for the entire study.

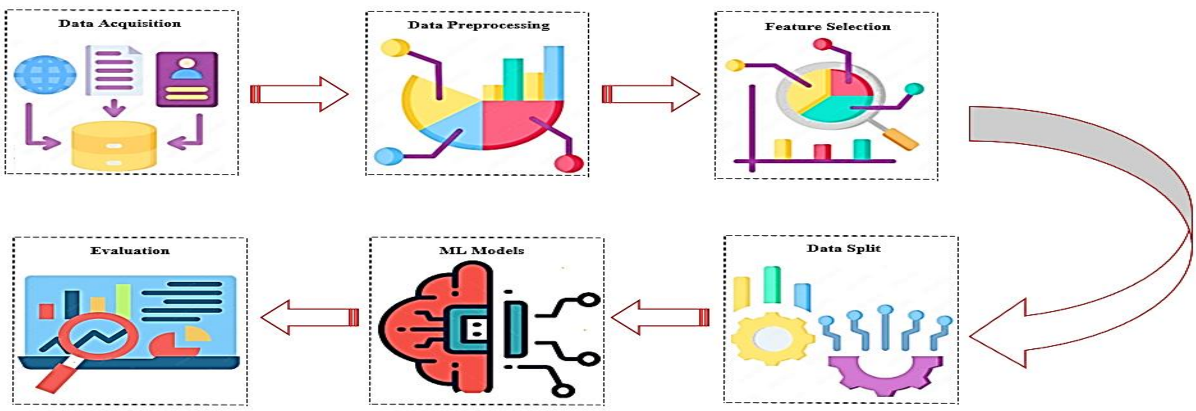

Figure 1.
Block Diagram of the entire process.

Data Acquisition

The primary source of data for this study is a rich dataset sourced from the IEEE website, comprising 7043 customer records obtained from telecommunications companies. Each customer record is meticulously annotated with 21 distinct features as shown in figure 2 with individual datatypes, encompassing a wide range of customer attributes and usage patterns. The dataset includes both churn and non-churn instances, providing a comprehensive foundation for training and evaluating the ML models.

Data Pre-processing

Before model development, the dataset undergoes rigorous pre-processing to ensure its compatibility with ML algorithms. Important pre-processing procedures include one-hot encoding to convert categorical variables into a binary format appropriate for machine learning analysis, feature scaling to normalize the range of feature values, and data cleaning to manage missing values and outliers (Nadeem, G. Et al., 2023). These pre-processing steps are essential for enhancing the predictive accuracy and generalizability of the models.

Feature Selection

After hyperparameter tuning and data pre-processing, feature selection is one of the most crucial steps in assisting the model to perform better based on input attributes. This is because the target output entirely depends on the input attributes that are more pertinent to or related to the targeted output; to handle feature selection, the correlation technique is used to identify the attributes that are most frequently correlated with the targeted output. The association is displayed in Figure 3.

```
Data columns (total 21 columns):
 #    Column            Non-Null Count  Dtype
---   ------            --------------  -----
 0    gender            7043 non-null   object
 1    SeniorCitizen     7043 non-null   int64
 2    Partner           7043 non-null   object
 3    Dependents        7043 non-null   object
 4    tenure            7043 non-null   int64
 5    PhoneService      7043 non-null   object
 6    MultipleLines     7043 non-null   object
 7    InternetService   7043 non-null   object
 8    OnlineSecurity    7043 non-null   object
 9    OnlineBackup      7043 non-null   object
 10   DeviceProtection  7043 non-null   object
 11   TechSupport       7043 non-null   object
 12   StreamingTV       7043 non-null   object
 13   StreamingMovies   7043 non-null   object
 14   Contract          7043 non-null   object
 15   PaperlessBilling  7043 non-null   object
 16   PaymentMethod     7043 non-null   object
 17   MonthlyCharges    7043 non-null   float64
 18   TotalCharges      7043 non-null   float64
 19   Churn             7043 non-null   object
 20   Tenure Bins       7032 non-null   category
dtypes: category(1), float64(2), int64(2), object(16)
memory usage: 1.1+ MB
```

Figure 2.
Dataset detail with datatypes.

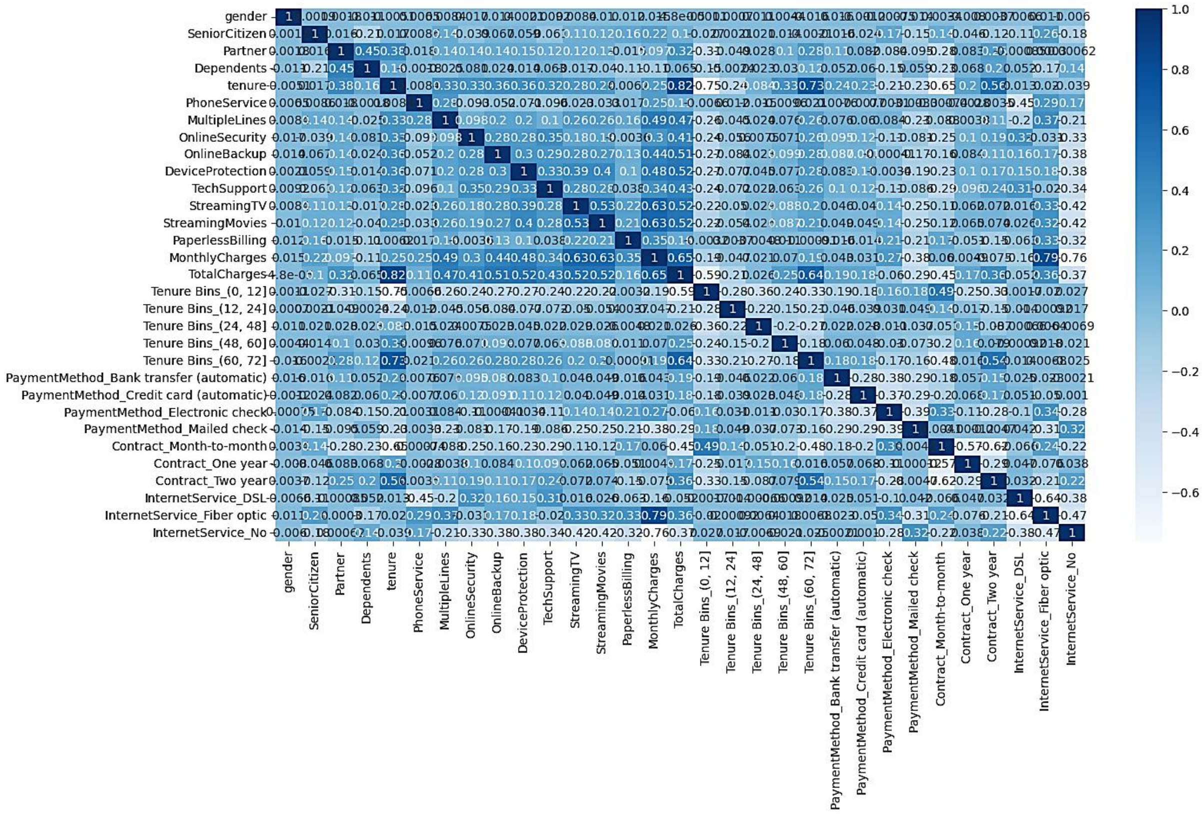

Figure 3.
Attributes correlation to the target.

ML Model Development

The study explores the performance of three prominent ML algorithms in predicting customer churn: Cat Boost, Stochastic Gradient Boost, and Extreme Gradient Boost. These algorithms are selected based on their widespread use in predictive analytics and their proven effectiveness in handling classification tasks. Model development involves training each algorithm on the pre-processed dataset and fine-tuning model parameters through techniques such as cross-validation to optimize performance metrics (Rahman, K. Et al., 2023).

Stochastic Gradient Boosting

Stochastic Gradient Boosting, also known as Gradient Boosting Machines (GBM), is a powerful ensemble learning technique that builds a series of decision trees sequentially, each tree correcting errors made by the previous one (Louk, M. H. L. At al., 2023). In this approach, decision trees are trained on random subsets of the training data, introducing an element of randomness that helps prevent overfitting.

Model Expression

The Stochastic Gradient Boosting algorithm minimizes a loss function by iteratively adding weak learners (decision trees) to the ensemble. At each iteration, the algorithm fits a new tree to the negative gradient of the loss function, effectively reducing the residual error (Wang, K. Et al., 2023). The final prediction is obtained by aggregating the predictions of all trees in the ensemble.

Relevance

Stochastic Gradient Boosting has been widely used in churn prediction tasks due to its robustness and ability to handle complex nonlinear relationships in the data. By incorporating randomness into the training process, this algorithm can effectively generalize to unseen data and produce accurate predictions of customer churn (Li, J. Et al., 2023).

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting, or XGBoost, is an optimized implementation of the gradient boosting framework known for its computational efficiency and scalability. XGBoost extends traditional gradient boosting by incorporating additional regularization techniques and a novel tree-learning algorithm, making it highly effective for large-scale datasets (Louk, M. H. L. At al., 2023).

Model Expression

XGBoost constructs an ensemble of decision trees in a sequential manner, where each tree is trained to minimize a regularized objective function. The algorithm employs a technique called gradient boosting with decision trees (Louk, M. H. L. At al., 2023), where the prediction at each iteration is the sum of predictions from all previous trees plus a learning rate multiplied by the new tree's prediction.

Relevance

XGBoost has gained popularity in churn prediction tasks due to its superior performance and ability to handle high-dimensional datasets with sparse features. By optimizing both computational speed and predictive accuracy, XGBoost offers telecom providers a powerful tool for accurately identifying customers at risk of churn (Saha, L. Et al., 2023).

CatBoost

CatBoost is a gradient-boosting algorithm specifically designed to handle categorical features efficiently. Unlike traditional gradient boosting methods, CatBoost automatically handles categorical variables without the need for explicit pre-processing, making it particularly well-suited for datasets with a mix of categorical and numerical features (Louk, M. H. L. At al., 2023).

Model Expression

CatBoost uses ordered boosting, a variation of gradient boosting that combines weak learners in a certain sequence to optimize the prediction task. To encode categorical variables in a tree-friendly manner, the algorithm integrates a novel approach for handling categorical features. This approach combines one-hot encoding and target encoding (Wang, K. Et al., 2023).

Relevance

Because of its robust performance over a wide range of datasets and its ability to handle categorical data with ease, CatBoost has become the primary choice for churn prediction in the telecom industry. Cat-Boost provides telecom companies with a workable way to estimate customer attrition with little effort by streamlining the pre-processing pipeline and increasing model accuracy (Manzoor, A. Et al., 2024).

Performance Evaluation

To evaluate the efficacy of the churn prediction models, several performance metrics are employed, including accuracy, precision, recall, F1-score, and the AUC of the receiver operating characteristic (ROC) curve (Nadeem, G. Et al., 2024). These metrics provide comprehensive insights into the models' ability to correctly classify churn and non-churn instances, as well as their overall predictive performance.

ETHICAL CONSIDERATIONS

Throughout the entire study process, ethical issues are crucial, especially when it comes to data privacy and confidentiality. All the data utilized in this research are aggregated and anonymized to guarantee that each customer's right to privacy is respected. Additionally, by helping telecommunications companies improve their client retention tactics and boost overall business success, the research's findings should help the sector as a whole.

LIMITATIONS

This study has limitations even though it aims to offer insightful information about churn prediction in the telecom sector. The dataset's representativeness, the quality and accessibility of the data, and the innate biases of machine learning algorithms are some potential drawbacks. To offer a fair assessment of the study results, these limitations are recognized and addressed.

RESULTS AND EVALUATION

The computation of the F-measure value will be the assessment approach employed in this study to assess the algorithm's performance. In terms of data pre-processing (Nadeem, G. Et al., 2024), feature selection, parameter tuning, accuracy, precision, recall, F-measures, and AUC value, all of the suggested models' values are computed before optimization. The associated algorithms are listed in Table 1 for each of these processes.

Table 1.
Before optimization accuracy metrics of ML models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Stochastic Gradient Boosting	80	77	85	81	0.80
Extreme Gradient Boosting	76	72	85	78	0.76
CatBoost	84	80	91	85	0.84

In Table 2 the performance of each model can be observed clearly. Based on the F-measure score, the Cat Boost model possesses a 97.56% F-measure score which is also the highest among the proposed models, and then followed by the Stochastic Gradient Boosting model which holds nearly the same F-measure score of 86.08% with the Cat boost. In addition, the XGB model holds a 78.08% F-measure score among the proposed models. Lastly, by evaluating the results of the models we noticed that the overall performance of the model increased after implementing the data preprocessing, feature selection (Nadeem, G. Et al., 2024), and one-hot encoding approach. Figure 4 presents the results of a classification test using three machine learning algorithms: CatBoost, Extreme Gradient Boosting, and Stochastic

Gradient Boosting. Numerous metrics, like as accuracy, precision, recall, F1-score, and AUC, are used to assess the algorithms. Both before and after optimization, the algorithms' performance is evaluated.

Table 2.
After optimization accuracy metrics of ML models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Stochastic Gradient Boosting	85	82	90	86	0.85
Extreme Gradient Boosting	76	72	86	78	0.76
CatBoost	94	90	97	94	0.94

All three algorithms exhibit solid results on the challenge; however, CatBoost outperforms the other two techniques overall. Based on the findings, performance can be significantly increased by optimizing the algorithms. The data's lower standard deviation, represented by the error bars, suggests that the accuracy metrics are genuine.

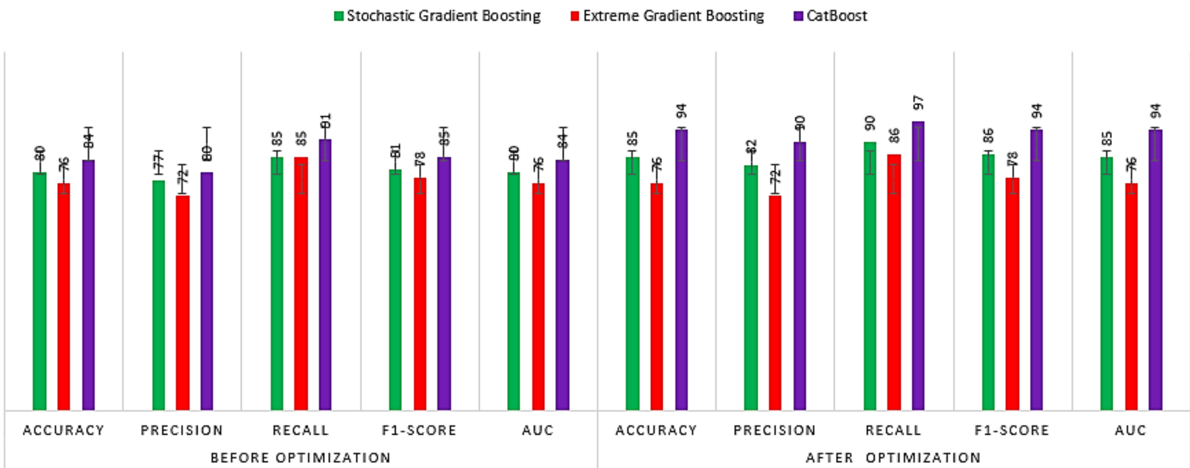

Figure 4.
Comparison of the performance of ML models before and after optimization.

CONCLUSION AND FUTURE WORK

This study examines the problems associated with customer attrition in the telecom sector, emphasizing the necessity of a proactive and anticipatory strategy. Through the integration of sophisticated machine learning algorithms, such as Cat Boost, with a customer record dataset, the research showcases the revolutionary capabilities of churn prediction models. The study achieves impressive predicted accuracy and dependability by analysing client information and utilizing algorithms such as Cat Boost. This allows providers to make data-driven decisions and implement tailored retention tactics. The comparison of prominent ML algorithms underscores Cat Boost's superiority in churn prediction, highlighting its efficacy in mitigating churn risks and fostering long-term customer loyalty. The practical implications of churn prediction include proactive churn prevention, tailored interventions, and strategic decision-making. By harnessing predictive analytics, telecom companies can navigate the complexities of customer attrition and secure sustained success in a dynamic marketplace. This research represents a significant advancement in churn prediction within the telecommunications industry, equipping telecom providers with the tools and insights needed to thrive amidst evolving consumer behaviors and fierce competition. In conclusion, this work provides the foundation for future directions in the field of customer churn prediction research. Prospective research directions encompass investigating ensemble learning methodologies, incorporating auxiliary data sources like social media interactions, and creating customized churn prediction models that cater to certain consumer categories.

DECLARATIONS

Acknowledgment: We appreciate the generous support from all the supervisors and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally in the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent

REFERENCES

Agha, A. A., Rashid, A., Rasheed, R., Khan, S., & Khan, U. (2021). Antecedents of customer loyalty in the telecom sector. Turkish Online Journal of Qualitative Inquiry, 12(9).

Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical feature encoding. IEEE Access, 9, 114381-114391.

Dhanawade, A., Mahapatra, B., & Bhatt, A. (2023, November). A smote-based churn prediction system using machine learning techniques. In 2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI) (Vol. 1, pp. 1-7). IEEE.

Fei, D. (2021). Chinese telecommunications companies in Ethiopia: The influences of host government intervention and inter-firm competition. The China Quarterly, 245, 186-207.

Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. International Journal of Data Science and Analytics, 14(3), 217-242.

Hancock, J. T., & Khoshgoftaar, T. M. (2021). Gradient boosted decision tree algorithms for medicare fraud detection. SN Computer Science, 2(4), 268.

Johnson, J. M., & Khoshgoftaar, T. M. (2022). Encoding high-dimensional procedure codes for healthcare fraud detection. SN Computer Science, 3(5), 362.

Keshavarz, H., Mahdzir, A. M., Talebian, H., Jalaliyoon, N., & Ohshima, N. (2021). The value of big data analytics pillars in the telecommunication industry. Sustainability, 13(13), 7160.

Khaldi, B., Babelhadj, Z., & Khemgani, S. Classification of Brain Tumor using some approaches of Machine Learning (Doctoral dissertation, UNIVERSITY OF KASDI MERBAH OUARGLA).

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 104(2), 271-294.

Li, J., Bai, X., Xu, Q., & Yang, D. (2023). Identification of Customer Churn Considering Difficult Case Mining. Systems, 11(7), 325.

Lottu, O. A., Ezeigweneme, C. A., Olorunsogo, T., & Adegbola, A. (2024). Telecom data analytics: Informed decision-making: A review across Africa and the USA. World Journal of Advanced Research and Reviews, 21(1), 1272-1287.

Louk, M. H. L., & Tama, B. A. (2023). Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system. Expert Systems with Applications, 213, 119030.

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A Review of Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners. IEEE Access.

Mengjing Hao. (2024). WA_Fn-UseC_-Telco-Customer-Churn. IEEE Dataport. https://dx.doi.org/10.21227/0q5y-3529

Nadeem, G., Ali, S., Hasan, B., Yasin, S., & Hasan, B. (2024). Decoding Bovine Behavior: A Machine Learning Analysis of Disease and Event Detection. The Asian Bulletin of Big Data Management, 4(1).

Nadeem, G., Rehman, Y., Khaliq, A., Khalid, H., & Anis, M. I. (2023, March). Artificial Intelligence-based prediction system for General Medicine. In 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE.

Nadeem, G., & Anis, M. I. (2024). Investigation of bovine disease and events through machine learning models. Pakistan Journal of Agricultural Research, 37(2), 102-114.

Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer churning analysis using machine learning algorithms. International Journal of Intelligent Networks, 4, 145-154.

Rahman, M. M. (2023). The effect of business intelligence on bank operational efficiency and perceptions of profitability. FinTech, 2(1), 99-119.

Rahman, K., Ghani, A., Alzahrani, A., Tariq, M. U., & Rahman, A. U. (2023). Pre-Trained Model-Based NFR Classification: Overcoming Limited Data Challenges. IEEE Access.

Routh, P., Roy, A., & Meyer, J. (2021). Estimating customer churn under competing risks. Journal of the Operational Research Society, 72(5), 1138-1155.

Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., & El-kenawy, E. S. M. (2023). Deep churn prediction method for the telecommunication industry. Sustainability, 15(5), 4543.

Tirkkonen, T. (2021). Improving Customer Experience with Artificial Intelligence, Data Analysis, and Automation in Azets.

Usman-Hamza, F. E., Balogun, A. O., Nasiru, S. K., Capretz, L. F., Mojeed, H. A., Salihu, S. A., ... & Awotunde, J. B. (2024). Empirical analysis of tree-based classification models for customer churn prediction. Scientific African, 23, e02054.

Wang, K., Lu, J., Liu, A., Song, Y., Xiong, L., & Zhang, G. (2022). Elastic gradient boosting decision tree with adaptive iterations for concept drift adaptation. Neurocomputing, 491, 288-304.

Zhao, M., Zeng, Q., Chang, M., Tong, Q., & Su, J. (2021). A prediction model of customer churn considering customer value: an empirical research of telecom industry in China. Discrete Dynamics in Nature and Society, 2021, 1-12.