Machine Learning with Applications 17 (2024) 100567

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Explaining customer churn prediction in telecom industry using tabular machine learning models

Sumana Sharma Poudel [a], Suresh Pokharel [b], Mohan Timilsina [c,*]

[a] Pokhara University, Nepal college of Information Technology, Nepal
[b] Westcliff University, Presidential Graduate School, Nepal
[c] Data Science Institute, University of Galway, Ireland

ARTICLE INFO

Keywords:
Customer churn
Explainable model
Global explainable
Local explainable
Telecommunication

ABSTRACT

The study addresses customer churn, a major issue in service-oriented sectors like telecommunications, where it refers to the discontinuation of subscriptions. The research emphasizes the importance of recognizing customer satisfaction for retaining clients, focusing specifically on early churn prediction as a key strategy. Previous approaches mainly used generalized classification techniques for churn prediction but often neglected the aspect of interpretability, vital for decision-making. This study introduces explainer models to address this gap, providing both local and global explanations of churn predictions. Various classification models, including the standout Gradient Boosting Machine (GBM), were used alongside visualization techniques like Shapley Additive Explanations plots and scatter plots for enhanced interpretability. The GBM model demonstrated superior performance with an 81% accuracy rate. A Wilcoxon signed rank test confirmed GBM's effectiveness over other models, with the *p*-value indicating significant performance differences. The study concludes that GBM is notably better for churn prediction, and the employed visualization techniques effectively elucidate key churn factors in the telecommunications sector.

1. Introduction

The service-oriented industries, such as telecommunications, face considerable challenges due to customer churn, where valuable customers are lost to competitors. As the world rapidly embraces digitization, the telecommunications sector serves as a crucial backbone. Notably, it represents a significant contributor to national income, particularly in developing countries, where it plays a substantial role in generating revenue (Liao & Lien, 2012). With its substantial business volume, telecommunications is recognized as a key industry, evident in ongoing technical advancements and a growing number of operators. Consequently, fierce competition among service providers persists (Gerpott, Rams, & Schindler, 2001), leading to the introduction of new technologies, services and strategies aimed at attracting new customer and retaining existing customers. The churn rate in this sector is approximately 2.6% monthly (Hawley, 2003). Comparing the return on investment between acquiring a new customer and retaining an existing one reveals that the latter is less expensive (Reinartz & Kumar, 2003; Yang & Peterson, 2004) and generally easier than upselling (Ascarza, Iyengar, & Schleicher, 2016). Therefore, customer retention is recognized as the most profitable strategy (Qureshi, Rehman, Qamar, Kamal, & Rehman, 2013; Wei & Chiu, 2002) and can positively influence the company's reputation, reducing marketing costs for new customer acquisition (Bolton & Bronkhorst, 1995; Reichheld & Sasser, 1990). So, it is desirable to have thorough research on customer churn and taking proactive measures in response by decision maker can provide a competitive edge to stay ahead in this competition.

The primary goal of the churn prediction is to support the creation of client retention plans in a market that is highly competitive. Churn models are made to predict which customers are likely to quit on their own will and to spot early signs of churn (Wei & Chiu, 2002). For this, companies must leverage their databases as valuable assets to comprehend customer churn behavior (Coussement & Van den Poel, 2008). Fundamentally, these databases contain information on customer service usage, billing details, and satisfaction levels. In addition to predicting customers likely to switch, companies seek to understand churn causes, which aids in profiling prone customers and devising effective retention campaigns (Leung, Pazdor, & Souza, 2021). Effective churn modeling has two important components: (i) predicting whether a specific customer will churn, (ii) discovering the reasons behind their churn, either at a local or global level. While

* Corresponding author.
E-mail addresses: sumana.192978@ncit.edu.np (S.S. Poudel), suresh.pokharel@presidential.edu.np (S. Pokharel), mohan.timilsina@insight-centre.org (M. Timilsina).

https://doi.org/10.1016/j.mlwa.2024.100567
Received 9 February 2024; Received in revised form 28 March 2024; Accepted 19 June 2024
Available online 24 June 2024

much of the existing research predominantly focuses on the first aspect. They are treating churn prediction as a binary classification task and employing various machine learning techniques around it such as feature extraction (Zhao, Gao, Dong, Dong, & Dong, 2017), feature selection (Umayaparvathi & Iyakutti, 2017), treatment of imbalanced datasets (Fujo et al., 2022), and utilizing classifiers like SVC (Cortes & Vapnik, 1995), Logistic Regression (Hosmer, Lemeshow, & Sturdivant, 2013), Random Forest (Breiman, 2001), XGBoost (Friedman, 2001), and Neural networks (Goodfellow, Bengio, & Courville, 2016). However, this alone may not suffice to fully grasp customer behavior and these ignore the second important component. These model cannot explain the reason behind the churning.

This study aims to close the research gap in the field of churning prediction by focusing not only on forecasting whether a certain customer would churn or not, but also on the reason why. For the reasoning we adapt SHapley Additive exPlanations (SHAP) to explain machine learning predictions by identifying influential customers from the training set (Lundberg & Lee, 2017a). The specific research questions (RQs) investigated are:

- **What are the best available off-the-shelf machine learning algorithms for predicting customer churn?**
  1. Which classification algorithm performs best for churn prediction in terms of different evaluation metrics?
  2. Is there a significant difference in the predictions made by these classifiers?
- **How can we explain the factors responsible for customer churn?**
  1. What are the most important predictors, and how do they influence prediction performance?
  2. Is there any interaction between the churn predictors?

Contributions: Our contributions are summarized as follows:
With this, this research's contributions are summarized as follows:

- We rigorously compared state-of-the-art supervised machine learning
algorithms for churn prediction.
- We performed statistical tests to find the most significant model for churn prediction.
- We provide explanations for each predictor corresponding to customer churn, highlighting both positive and negative contributions to churn prediction.

To the best of our knowledge, our approach is the first to generate global and/or local explanations for churn prediction. We conducted rigorous experiments to evaluate tabular machine learning algorithms using different evaluation metrics and to choose the most significant model.

The remainder of the paper is organized as follows: related work, problem definition, method description, experiments, and conclusion.

2. Related work

Recently, data mining techniques have emerged to tackle the challenging problems of customer churn in telecommunication service field (Au, Chan, & Yao, 2003; Hadden, Tiwari, Roy, & Ruta, 2007). As one of the important measures to retain customers, churn prediction has been a concern in the telecommunication industry and research (Bin, Peiji, & Juan, 2007). Majority of the research focused on churn prediction were dedicated in voice services available over mobile and fixed-line networks. In most of the cases, the features used for churn prediction in mobile telecommunication industry includes customer demographics, contractual data, customer service logs, call details, complaint data, bill and payment information (Bin et al., 2007; Hadden et al., 2007). However, the information for land-line services providers are different than mobile services (Bin et al., 2007). Some of this data is missing, less reliable or incomplete in land-line communication service providers. For instances, customer ages and complaint data, fault reports are unavailable and only the call details of a few months are available. Due to business confidentiality and privacy, there are no public datasets for churn prediction (Huang, Kechadi, & Buckley, 2012).

Customer churn prediction models have demonstrated significant value beyond telecommunications, notably within industries like digital marketing, e-commerce, and banking, where understanding and mitigating churn is equally critical. In digital marketing, the application of churn models facilitates the optimization of customer engagement and retention strategies. For instance, Ascarza (2018) in their work delve into how digital marketing efforts can be tailored to retain customers showing signs of churn, offering insights into the effectiveness of targeted interventions. In the banking sector, Miguéis, Van den Poel, Camanho, and e Cunha (2012) apply churn prediction to understand and predict customer churn concerning specific banking products and services. These references collectively highlight the broad applicability of churn prediction models across various industries, emphasizing their potential to inform and refine customer retention strategies in diverse business contexts.

Churn analysis and prediction task is also tackled from statistical modeling perspective. A very popular approach to model churn is time to event prediction (Bhattacharya, 1998; Van den Poel & Lariviere, 2004). In the context of customer attrition, the time to failure links to the churn behavior. The potential churner behavior has also been considered using structural equation modeling (Nguyen & LeBlanc, 1998; Varki & Colgate, 2001). Such technique can be of great interest for managerial decisions, as it evaluates the effect of suspected influential features on a specific customer decision, such as churn (Geiler, Affeldt, & Nadif, 2022). The variance analysis was also widely used in marketing and business areas to uncover customer behavior (Maxham, 2001; Mittal & Kamakura, 2001; Zeithaml, Berry, & Parasuraman, 1996). Financial and retail services also rely on classical T-test and Chi square statistics to forecast customer behavior and perceptions (Hitt & Frei, 2002; Mittal & Lassar, 1998). The churn prediction problem has one important issue of class imbalance (Kong, Kowalczyk, Menzel, & Bäck, 2020) that might cause biased towards the negative samples which might hinder training the machine learning models (Zhu, Baesens, & vanden Broucke, 2017). Typically, this problem occurs when the classes in a given dataset are unequally distributed between the minority and majority classes that is low number of "churners" than "non churners". Without considering this problem, effective learning process by classification algorithms will be a challenge, since the main goal is the detection of minority classes (Dwiyanti et al., 2016; Sun, Wong, & Kamel, 2009). The popular algorithms like k-nearest neighbors (k-NN) is also applied in the churn-like data however studies (Dubey & Pudi, 2013; Tan, 2005) have shown several significant drawbacks. In the context of class imbalance issues in churn prediction problem, Naive Bayes classifier also appeared to be sensitive due to the strong bias in the prior estimation (Bermejo, Gámez, & Puerta, 2011). However, Huang et al. (2012) demonstrated reasonable results using Naive Bayes method.

Earlier studies have provided for various customer churn models they have analyzed the model based on customer behavior data and used different data mining techniques (Moayer & Gardner, 2012; Naz, Shoaib, & Shahzad Sarfraz, 2018; Pushpa, 2012). In these studies, all churn prediction models were analyzed and models with the best results were presented. There are various approaches for that for example Lazarov and Capota (Lazarov & Capota, 2007) showed that a model based on the customer's lifetime value analysis is the best way to predict customer churn. Similarly Naz et al. (2018) and Bandara, Perera, and Alahakoon (2013) analyzed model based on a dataset they used and showed that a big dataset with more features causes model training and evaluation difficult. Hence, this research suggested focusing on feature selection to reduce the number of features. In terms of machine learning models the study showed that for true churn rate and false churn rate, SVM should be used and in case of churn probability, logistic regression should have been used. Similarly Ahmed and Linen (2017) proposed that using hybrid models are useful and accurate for churn prediction.

The user churn prediction is also studied from the network science perspective. Recently the studies (Ahmad, Jafar, & Aljoumaa, 2019; Huang et al., 2015; Mitrović & De Weerdt, 2020; Xu et al., 2021; Zhang, Zeng, Zhao, Jin, & Li, 2022) showed the effect of social influence on user churn. The techniques to approach this problem is categorized from two perspective. The first one is to model the network structure as a surrogate of social influence. For instance, Ahmad et al. (2019) used social network analysis to extract network-based features for machine learning model. Similarly, Yang, Shi, Jie, and Han (2018) extracted network features to cluster users in different communities and predict customer churn with a deep learning model. The second one is to model the sequential order of churn as a diffusion process and use propagation models such as inflection and stopping rule (Ji et al., 2021) and spreading propagation activation (Dasgupta et al., 2008) to simulate the diffusion process and give predictions. However, the main caveat of this method is that these approaches failed to capture the causal nature of social influence. There is also a graph-based semi-supervised effort to predict the customer-churn in telecommunication (Benczúr, Csalogány, Lukács, & Siklósi, 2007). Liu et al. (2018) propose a novel graph-based inductive semi-supervised embedding model that jointly learns the prediction function and the embedding function for user-game interaction to predict the user churn from the games.

Recent studies begin to investigate how to use causal information to build better deep learning models (Bonner & Vasile, 2018; Yoon, Jordon, & Van Der Schaar, 2018). It includes the applications to eliminate the bias between the observed data and the application scenarios and learning the causal effects to give more accurate churn predictions (Johansson, Shalit, & Sontag, 2016). The studies by Umayaparvathi and Iyakutti (2017) demonstrated that deep learning models have similar performance to conventional classifiers such as support vector machine and random forest. The transfer learning which is very popular in image classification has also been employed in the customer churn prediction (Ahmed et al., 2019). Similarly, Seymen, Dogan, and Hiziroglu (2020) proposed a novel deep learning model which is compared to logistic regression and artificial neural network models. In a similar note, Momin, Bohra, and Raut (2020) demonstrated that deep Learning enables multi-stage models to represent the data at multiple abstraction levels which reduces the time and effort of feature selection considerably as it automatically creates useful features for accurate customer churn prediction. In spite of the popularity, the deep learning models can still be considered as a black box because of the complicated architecture and there is a little visibility into its decision rationale (Colbrook, Antun, & Hansen, 2022). Furthermore, it is also ambitious to recognize problems in a machine learning model or otherwise find improvements for it if the model's behavior cannot be understood (Adadi & Berrada, 2018). EXplainable Artificial Intelligence (XAI) (Emmert-Streib, Yli-Harja, & Dehmer, 2020) is a research area that studies how to make models transparent and explainable. In terms of black box models such as random forest and artificial neural network they require the application of XAI techniques to explain the model recommendation (Leung et al., 2021).

From the above listed studies, we observed that customer churn has investigated a wide range of algorithms from white box to black box models. They have good abilities to differentiate between "churn" and "no churn" customers. However, previous studies have not primarily focus on explaining churn prediction model. Therefore, successfully discriminating between these two categories is not only the aspect that is utmost importance. For customer churn prediction, understanding of the model and its outputs is important as well to target incentives to customers who have a high risk of churning and inducing them to stay. Thus in this work, we exploit the power of XAI to uncover **local** and **global** explanation of churn prediction. In particular, these explanations will enable the understanding of machine learning reasoning for the domain expert for customer churn prediction. From global explanation, one can learn about the most important pattern learned by the machine learning model for churn prediction about training population. It helps to understand the interaction between the confounding predictors. From local explanation, it enables the reasoning that the model applied to a particular case to answer the very specific questions such as "Why customer Alex churned?" and "Why has Jane continued to subscribe the plan".

3. Solution approach

The overall solution of our approach is illustrated in Fig. 1. The main aim of this study is to assess machine learning classifiers to predict the customer churn and provide local and global explainability for those predictions. In the next section, we explained our methodology of our approach.

4. Methods

The figure above depicts the methodology of the proposed model approach for the churn prediction. The step wise working of methodology is described as below:

- **Dataset:** The input to the model is the Telecommunication dataset in any tabular format. The dataset used in the paper is from Kaggle. The dataset consists of missing data which requires cleaning. For this, the dataset is passed to data preparation and preprocessing steps.
- **Selection Criteria:** The Telecommunication dataset consist of data of both churners and non-churners. Some of field might consist of missing values as well. Such data should be handled before the data are fed into the model. Thus, in this steps missing values are drop.
- **Feature Engineering and feature selection:** The raw datasets need to be handled before fetching to the classifiers. The input datasets consists of duplicate columns and unique value columns as well. Such data does not provide any significance in the churn prediction and thus these columns are drop.
- **Encoding:** The Telecommunication dataset consist of both numeric as well as categorical data. However, all of the machine level models do not work with categorical data. Thus, numeric conversion of data need to be done before application of ML models. For handling of such categorical data one-hot encoding technique is implemented in the model. This led to the increment in the column of the dataset.
- **Hyper parameter selection:** the optimization of hyperparameters across diverse machine learning models deployed for predicting customer churn within the telecommunications sector. These models are characterized by a multitude of hyperparameters, each necessitating precise calibration to enhance model efficacy.
- **Training Models:** In our methodology, we have incorporated a suite of state-of-the-art classification algorithms to ensure robust and accurate modeling. This includes the utilization of the SVM (Cortes & Vapnik, 1995), known for its effectiveness in high-dimensional spaces, and LR (Hosmer et al., 2013), a staple for binary classification problems. Additionally, we have leveraged the Random Forest Classifier (Breiman, 2001), which excels in handling large datasets with numerous features. The GBM (Friedman, 2001) has been selected for its prowess in predictive accuracy by combining multiple weak prediction models into a strong one. Lastly, Neural Networks (Goodfellow et al., 2016) have been implemented for their unparalleled capacity to learn from complex data patterns through layers of interconnected nodes, making our approach comprehensive and powerful.

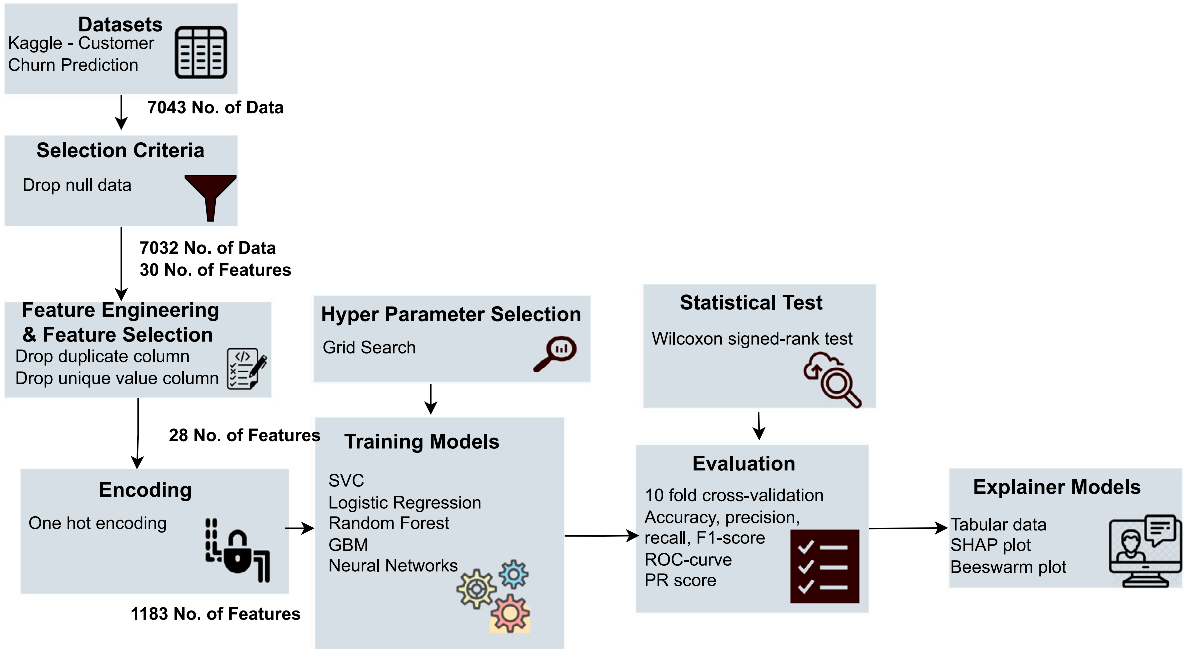

Fig. 1. An illustration of the data processing, model training, evaluation and explainer models on the customer churn data.

- **Statistical Test:** For a fair assessment of whether any model's predictive ability is statistically better or not in comparison to others Wilcoxon signed-rank test is implemented in the model. This provide a rigorous validation for model selection.
- **Evaluation Strategy:** To rigorously evaluate the performance of the classification models, we used 10-fold cross-validation strategy. This technique involves partitioning the original dataset into 10 equal-sized subsets. In each fold of the validation process, nine subsets are used to train the model, and the remaining one subset is used to test the model. This cycle is repeated 10 times, with each of the 10 subsets serving as the test set exactly once. By employing 10-fold cross-validation, we aim to achieve a more accurate and generalized understanding of the model's predictive power and ensure that classifier's performance is not dependent on a particular random split of the data. This is considered a robust method for assessing the generalizability of the model to an independent dataset.
- **Explainer models:** To elucidate the complexities of the telecommunication dataset within our study, we have integrated explainer models that substantially improve data visualization. Our approach incorporates SHAP (Lundberg & Lee, 2017b) plots for a macro-level analysis, providing global explanations of feature influences on the predictive model, alongside scatter plots for micro-level insights into individual customer behaviors. This bifurcated visualization strategy enables a comprehensive understanding of the dataset, facilitating the identification of systemic and case-specific factors influencing customer churn. Consequently, these methods enhance the interpretability of our model and strengthen the predictive accuracy regarding the key determinants of churn in the telecommunications domain.

Table 1
Summary of Attributes used in the Dataset.

S.N	Attributes	Data type
1	CustomerID	object
2	Count	int64
3	Country	object
4	State	object
5	City	object
6	Zip_Code	int64
7	Lat_Long	object
8	Latitude	float64
9	Longitude	float64
10	Gender	object
11	Senior_Citizen	object
12	Partner	object
13	Dependents	object
14	Tenure_Months	int64
15	Phone_Service	object
16	Multiple_Lines	object
17	Internet_Service	object
18	Online_Security	object
19	Online_Backup	object
20	Device_Protection	object
21	Tech_Support	object
22	Streaming TV	object
23	Streaming Movies	object
24	Contract	object
25	Paperless_Billing	object
26	Payment Method	object
27	Monthly Charges	float64
28	Total_Charges	object
29	CLTV	int64
30	Churn_Label	object

5. Feature engineering and data processing

Table 1, demonstrates the feature we used in our study. We prepared the dataset for analysis, ensuring a robust foundation for our predictive models. The dataset comprises an array of features spanning customer demographic information, account details, and service subscriptions, each playing a crucial role in understanding customer behavior and predicting churn.

In our churn prediction model, we placed significant emphasis on feature engineering to enhance the model's ability to predict customer churn accurately. Among the additional features we created, the engagement score. This composite score is derived from Tenure_Months, Monthly Charges, and Total Charges, offering customer engagement and loyalty over time. By integrating these elements, the engagement score provides a multifaceted understanding of how deeply and satisfactorily customers are connected to the services offered. Another critical feature we introduced is service utilization, which quantifies the

Table 2
Model hyperparameters.

Model	Hyperparameter tuning range	Hyperparameter
SVC	0.001641949 – 464.0812108	C
Logistic Regression	5.15E−05 - 4534347.358	C
Random Forest	9 – 20 14–20	max-depth n-estimators
GBM	5–29 5–10 auto 3–7	max-depth min-samples-leaf max-features max-leaf-nodes
Neural networks	5–9 relu adam	hidden-layer-sizes activation solver
AdaBoost	50 – 500 0.01 – 1.0	n-estimators learning-rate
XGBoost	100 - 1000 0.01 – 0.3 3 – 10	n-estimators learning-rate max-depth

total number of services a customer utilizes, including Phone_Service, Multiple_Lines, Internet_Service, among others. This feature reflects the depth of product penetration and serves as an indicator of potential customer satisfaction. A higher service utilization often suggests that customers find value in a wider range of services, potentially increasing their loyalty and decreasing their likelihood of churn.

6. Results

6.1. Model hyperparameter tuning

Table 2 illustrates the optimization of hyperparameters across diverse machine learning models deployed for predicting customer churn within the telecommunications sector. These models are characterized by a multitude of hyperparameters, each necessitating precise calibration to enhance model efficacy. Detailed in the table are the ranges of hyperparameter tuning, alongside the specific hyperparameters selected for each model, underscoring their pivotal role in refining model performance.

6.2. Experiments

Table 3 presents the specifications of the dataset employed in our studies. The use of detailed telecommunication data poses substantial challenges, primarily due to rigorous privacy regulations and proprietary limitations, which significantly hinder external analytical endeavors and innovative developments. Kaggle[1] enhances these constraints by providing anonymized datasets, thereby ensuring adherence to privacy standards while simultaneously facilitating the extraction of valuable analytical insights. The platform's dynamic community further promotes a culture of collaboration and knowledge exchange, catalyzing the development of novel solutions for intricate sector-specific issues such as churn prediction. Consequently, our study leverages this publicly accessible data to train our models and derive predictive insights from this dataset.

To assess the performance of the state-of-the-art classifier model, we utilized a comprehensive set of evaluation metrics, including Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curve, and Precision–Recall (PR) score. Table 4 summarizes the performance metrics of various machine learning models used for churn prediction tasks. Each evaluation metric is accompanied by a mean value and a standard deviation $\pm$, indicating the variability of the

Table 3
Summary of the Dataset.

Description	Dataset
Number of samples	7043
Number of features	30
% of positive samples (Churn)	26.54%
%of negative samples (Non-Churn)	73.46%
Data source	Kaggle

model's performance. The models are ranked by their Accuracy, with GBM showing the highest Accuracy of 0.81 $\pm$ 0.02 and Neural Networks the lowest at 0.74 $\pm$ 0.06. The ROC-score follows a similar trend, with GBM having the highest score. The PR-score is also highest for GBM, suggesting its superior performance across various aspects of churn prediction tasks in this evaluation. The data presented in the table reveals that the GBM model exhibits superior performance compared to other models.

We have used the Wilcoxon signed-rank test is used to determine if there is a significant difference in the predictive power of GBM compared to each of the other models when applied to the same churn prediction task. This allows for a fair assessment of whether GBM's predictive ability is statistically better or not, providing a rigorous validation for model selection.

The test results showcased in the Table 5 demonstrate that GBM significantly outperform several other supervised machine learning models in the context of churn prediction for this specific dataset. AdaBoost, with a p-value of 0.05, indicates that its difference in performance compared to GBM is on the threshold of statistical significance, suggesting a competitive but slightly less effective model than GBM in this context. XGBoost's p-value of 0.07, slightly above the conventional threshold for statistical significance, suggests that while it may offer strong predictive capabilities, it does not statistically outperform GBM to a significant degree in this dataset. Both Neural Networks and Logistic Regression, with p-values well below the 0.05 threshold, demonstrate a statistically significant difference in performance compared to GBM, indicating GBM's superior capabilities in churn prediction. The SVC's performance, with a p-value marginally above the threshold, and Random Forest, with a higher p-value, suggest a less significant difference compared to GBM, underscoring GBM's robustness and effectiveness as a churn prediction tool. This comprehensive comparison underscores the importance of selecting the right model based on the dataset's specific characteristics and the predictive task at hand. While GBM shows strong performance, the nuanced differences between models highlight the potential benefits of model ensemble approaches or further hyperparameter tuning to optimize predictive accuracy.

To further understand the effectiveness of the GBM, we utilized a confusion matrix to examine its predictive accuracy and identify the areas where the model may be making errors.

Table 6 presents the confusion matrix for the GBM model, a key tool in our churn prediction analysis. The matrix indicates that the model is highly effective at identifying customers who will remain with the service, as evidenced by the 466 true negatives. However, it also points to a notable challenge in the form of 84 false negatives, which represent customers who were predicted to stay but actually churned. While the model successfully identified 103 actual churners (true positives), it incorrectly flagged 51 loyal customers as likely to churn (false positives), suggesting a need for refinement. The GBM model's strong suit is its ability to recognize stable customers, a vital aspect of preserving a customer base and avoiding the costs associated with unwarranted retention incentives. Yet, its tendency to overlook some churners could lead to substantial customer loss if not addressed. Improving the model's sensitivity, to capture more true churn cases, and its precision, to reduce the mistaken identification of loyal customers as churners, emerges as a critical focus for advancing its utility in practical

[1] https://www.kaggle.com/

Table 4
Results of the 10 Fold cross validation of supervised machine learning classification model for churn prediction. The figure behind ± is the standard deviation.

Models	Accuracy	Precision	Recall	F1-score	ROC-score	PR-score
Neural networks	0.74 ± 0.06	0.58 ± 0.26	0.43 ± 0.31	0.41 ± 0.21	0.83 ± 0.02	0.64 ± 0.03
SVC	0.78 ± 0.01	0.68 ± 0.03	0.34 ± 0.02	0.45 ± 0.02	0.77 ± 0.02	0.57 ± 0.04
Logistic Regression	0.79 ± 0.02	0.64 ± 0.04	0.47 ± 0.06	0.54 ± 0.05	0.81 ± 0.03	0.61 ± 0.04
AdaBoost	0.79 ± 0.01	0.65 ± 0.02	0.50 ± 0.06	0.57 ± 0.03	0.82 ± 0.01	0.63 ± 0.02
XGBoost	0.80 ± 0.03	0.68 ± 0.01	0.55 ± 0.02	0.61 ± 0.03	0.85 ± 0.02	0.67 ± 0.02
Random Forest	0.80 ± 0.02	**0.71 ± 0.04**	0.43 ± 0.08	0.53 ± 0.07	0.84 ± 0.01	0.64 ± 0.03
GBM	**0.81 ± 0.02**	0.67 ± 0.04	**0.55 ± 0.03**	**0.60 ± 0.02**	**0.86 ± 0.01**	**0.68 ± 0.03**

Table 5
Wilcoxon signed rank test.

Models	Statistics	pvalue
Neural Networks	0.0	0.03125
SVC	1.0	0.0625
Logistic Regression	0.0	0.03125
AdaBoost	2.0	0.05
XGBoost	1.5	0.07
Random Forest	3.0	0.15625

Table 6
Confusion Matrix analysis for GBM.

		Prediction outcome	
		Non-churners	Churners
Actual value	Non-churners	466	51
	Churners	84	103

business scenarios. These enhancements are imperative for tailoring customer retention strategies more effectively and securing a healthier churn rate, thereby improving the business's financial performance and customer satisfaction.

Qualitative Benchmark with Other State-Of-The-Art Models: In Table 7, we introduce an innovative approach to customer churn prediction, leveraging Gradient Boosting Machines (GBM) to analyze the Kaggle customer churn prediction dataset. Our methodology achieved a ROC-Score of 0.86, positioning it competitively among state-of-the-art methods in churn prediction for the telecommunications industry. Notably, Ebrah et al.'s use of SVM on both the IBM Watson dataset and the cell2cell dataset resulted in ROC-Scores of 0.83 and 0.99, respectively, indicating a high benchmark for model performance in varied contexts (Ebrah & Elnasir, 2019). Similarly, Shrestha et al. demonstrated the efficacy of XGBoost in achieving a ROC-Score of 0.98 with data from a Telecom service provider in Nepal (Shrestha & Shakya, 2022), while Saha et al. utilized CNN and ANN models to reach a ROC-Score of 0.99 across datasets from both Southeast Asian and American telecom markets (Saha et al., 2023). These findings underscore the significant advancements in churn prediction methodologies, with SVM, XGBoost, CNN, and ANN models setting high standards for accuracy and reliability. Our GBM-based approach contributes to this evolving landscape by not only achieving a commendable ROC-Score but also by emphasizing the adaptability and effectiveness of GBM models in handling the complexities of customer churn prediction. This comparative analysis highlights our model's potential in bridging the gap between traditional machine learning techniques and the demands of modern-day churn prediction challenges.

6.2.1. Selection of most important predictors

Fig. 2 presents a beeswarm plot generated using SHAP values, which delineates the influence of various features on the GBM model's churn predictions. The plot reveals that the 'Contract_Month-to-month', 'Tenure_Months', and 'Monthly Charges' features exert the most substantial impact on the model's output, with the 'Contract_Month-to-month' feature, in particular, strongly pushing predictions towards churn. A gradation from blue to red denotes the range of feature values, with red signifying higher values. The horizontal spread of the dots reflects the magnitude of each feature's SHAP value; points to the right of the central vertical line indicate a feature's propensity to increase the likelihood of churn, while points to the left suggest a decrease. Notably, features such as 'Internet_Service_Fiber optic' and 'Payment_Method_Electronic check' predominantly contribute positively to churn predictions, whereas features like 'Online_Security_No', 'Dependents_Yes', and 'Tech_Support_No' display a mixture of positive and negative effects on the model's predictions. In the next section, we have demonstrated the top two ranked features 'Contract_Month-to-month', 'Tenure_Months' by the GBM and its interaction with the other features in the data.

6.2.2. Interaction between the churn predictors

Fig. 3 visualizes the relationship between month-to-month contracts and the provision of fiber optic internet service in the context of customer churn. The red dots represent customers who have churned (discontinued their service), and the blue dots represent those who have not churned (continued their service). The x-axis differentiates customers based on their contract type, with a particular focus on month-to-month contracts. The y-axis measures some standardized metric related to churn, possibly a probability or a churn score. From the plot, we can observe a higher density of red dots at the higher end of the month-to-month contract axis, indicating that customers with month-to-month contracts and fiber optic internet service are more likely to churn. Conversely, there are more blue dots concentrated towards the lower end of the axis, suggesting that customers without fiber optic service or with longer contract terms are less likely to churn. This implies an interaction where the likelihood of churn is amplified for customers who have fiber optic service on a month-to-month basis compared to those without such service or with more extended contracts.

Fig. 4 illustrates the relationship between customer tenure, measured in months on the x-axis, and the amount they are charged monthly, represented by the color intensity of the dots, with magenta indicating higher charges and blue indicating lower charges. The y-axis shows a standardized value metric, which might represent customer satisfaction or likelihood of churn. The pattern suggests that customers with shorter tenure and higher monthly charges (magenta dots) experience a more substantial negative impact on the standardized value metric, which could indicate lower satisfaction or higher churn risk. As tenure increases, the density of magenta dots diminishes, particularly beyond the 20-month mark, suggesting that customers with higher monthly charges either improve in their standardized value metric or possibly churn out of the service, leaving behind those more satisfied or less sensitive to the charge amount. The convergence of magenta and blue dots as tenure increases indicates that the impact of monthly charges on the standardized metric decreases over time. Customers with longer tenure, irrespective of their monthly charges, show similar values of the standardized metric, which could imply that the initial sensitivity to pricing diminishes, or that the remaining customer base has adapted to or accepted the monthly charges.

Table 7
Performance comparison of various models on telecom customer churn prediction, highlighting our GBM approach.

Reference	Dataset	Evaluation metric	Model
Yabas, Cankaya, and Ince (2012)	Orange Telecom	ROC-Score (0.653)	Random Forest
Ebrah and Elnasir (2019)	IBM Watson dataset	ROC-Score (0.83)	SVM
Ebrah and Elnasir (2019)	cell2cell	ROC-Score (0.99)	SVM
Shrestha and Shakya (2022)	Telecom service providerof Nepal	ROC-Score (0.98)	XGBoost
Saha et al. (2023)	Southeast Asian telecom industry, and American telecom market.	ROC-Score (0.99) in bothdataset	CNN and ANN
Our approach	Kaggle customer churn prediction	ROC-Score (0.86)	GBM

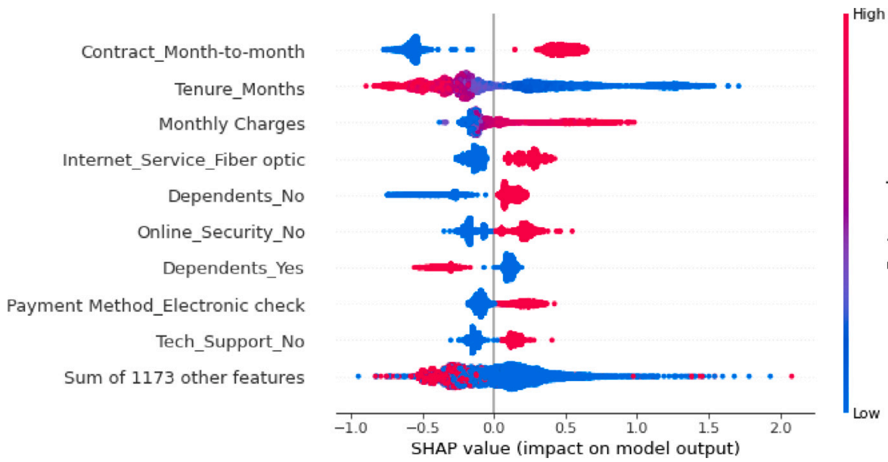

Fig. 2. Beeswarm plot for GBM model.

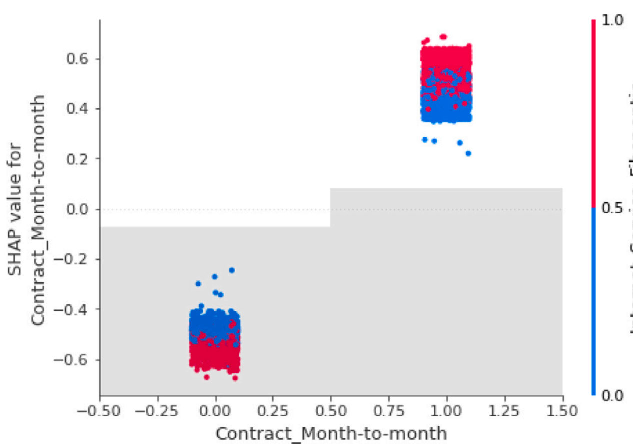

Fig. 3. Feature Interaction of Contract-Month-to-month.

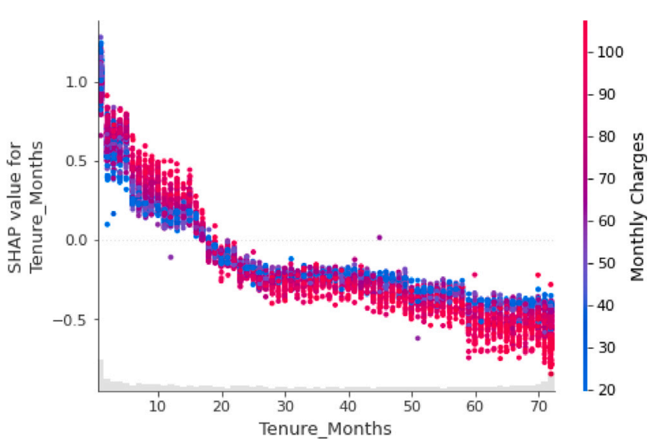

Fig. 4. Feature Interaction of Tenure-Months.

7. Discussion

The telecommunications industry is at the forefront of customer-centric strategies, where understanding and mitigating churn is not just beneficial but essential for sustaining growth and profitability. In our study, we sought to identify a machine learning model that not only excels in churn prediction but also offers clear insights into the reasons behind customer turnover. GBM model emerged as the front-runner in our analyses, substantiated by a rigorous statistical comparison using the Wilcoxon signed-rank test. The test revealed that GBM significantly outperforms Neural Networks and Logistic Regression in predicting churn, with p-values indicating the improbability of such results being due to chance.

What sets our approach apart is the incorporation of SHAP, which provided both global and local interpretability of the GBM model's predictions. Globally, SHAP values allowed us to rank features by their importance and to understand the overall direction and strength of each feature's impact on churn prediction. For instance, features like month-to-month contracts, tenure, and monthly charges were identified as key drivers of churn. Customers with short-term contracts or higher monthly charges were predisposed to churn, implying that long-term contracts and competitive pricing could be effective retention strategies.

Locally, SHAP offered insights into individual predictions, explaining why specific customers were likely to churn according to the model. This level of detail is crucial for customer relationship management, as it allows for personalized intervention strategies. For example, a customer predicted to churn due to high monthly charges could be offered a discount or a bundle package as an incentive to stay. The GBM model's ability to reveal complex interactions between features was another advantage. Through SHAP interaction values, we observed how the impact of one feature on churn could change in the presence of another feature. For example, the negative effect of a month-to-month contract on customer retention was exacerbated when combined with fiber optic internet service, suggesting that customers with this combination of services were particularly churn-prone.

The combination of GBM and SHAP explanations thus provided a powerful tool for telecom operators. Not only could they accurately predict which customers were at risk of churning, but they could also understand the underlying factors contributing to these predictions. This understanding facilitates the development of targeted strategies to retain specific customer segments, enhancing the efficiency of marketing efforts and potentially improving customer satisfaction. Incorporating these insights into business operations could lead to more nuanced customer segmentation and more effective churn prevention initiatives. For instance, identifying at-risk customers based on their usage patterns and service preferences enables the deployment of tailored communication strategies and personalized offers, thereby fostering customer engagement and loyalty.

Our work's core contribution lies in enhancing the interpretability of machine learning (ML) models for customer churn prediction, particularly through the use of SHapley Additive exPlanations (SHAP) values. The creation of unique features before data classification indeed presents a valuable avenue for research; however, it poses substantial challenges, including the need for deep domain expertise, limitations posed by data availability and quality, the balance between model complexity and interpretability, and the risk of overfitting. Our study focuses on leveraging existing, well-understood features and enriching the analysis with detailed interpretability to provide actionable insights. This approach not only aids telecom providers in identifying and addressing churn risks but also maintains the model's generalizability and robustness, carefully navigating the complexities inherent in feature engineering.

8. Conclusion

In the telecom sector, accurately predicting which customers are likely to leave the service is crucial. The ability to identify at-risk customers early on allows companies to intervene with targeted retention strategies. Machine learning models, particularly those that handle tabular data, are key to making these predictions. These models analyze customer data and can effectively forecast who might churn. This predictive power is essential for reducing churn rates, which is a persistent problem for telecom providers. Our research found that the GBM model was especially effective in this data. To confirm GBM's performance, we compared it with other advanced models using the Wilcoxon signed-rank test. The test results showed that GBM was significantly better at predicting churn. The p-value from the test helped us understand the strength of this evidence. A lower p-value indicates a more definitive difference between the models, and in our case, GBM's lower p-value confirmed its superior predictive ability. similarly, we leveraged the SHAP (SHapley Additive exPlanations) values to gain insights into the importance of different features in our predictive model. This information is invaluable for telecom companies looking to pinpoint the factors that most influence customer churn. By utilizing SHAP values, we were able to identify which specific customer attributes, such as call duration, plan type, or contract length, had the most significant impact on the churn prediction. These insights helped telecom providers tailor their retention efforts towards addressing the key factors driving customer attrition. SHAP values provided a transparent and interpretable way to analyze the model's decision-making process, making it a valuable tool for optimizing customer retention strategies in the telecommunications sector.

Funding

This work receive no funding.

Ethical approval

All data used in this work is freely available online. No other aspect of this work cause ethical issues.

CRediT authorship contribution statement

Sumana Sharma Poudel: Conducted experiments, Analysed the results, Prepared the original draft. **Suresh Pokharel:** Revised the original draft. **Mohan Timilsina:** Provided the guidance, Revised the manuscript.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank Data Science Institute, Insight Center for Data Analytics, at University of Galway Ireland for providing us constructive feedback and improvement of the manuscript.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.
Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1–24.
Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer learning and meta classification based deep churn prediction system for telecom industry. arXiv preprint arXiv:1901.06091.
Ahmed, A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. In *2017 4th international conference on advanced computing and communication systems* (pp. 1–7). IEEE.
Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, *55*(1), 80–98.
Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, *53*(1), 46–60.
Au, W.-H., Chan, K. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*(6), 532–545.
Bandara, W., Perera, A., & Alahakoon, D. (2013). Churn prediction methodologies in the telecommunications sector: A survey. In *2013 international conference on advances in ICT for emerging regions* (pp. 172–176). IEEE.
Benczúr, A. A., Csalogány, K., Lukács, L., & Siklósi, D. (2007). Semi-supervised learning: A comparative study for web spam and telephone user churn. In *In graph labeling workshop in conjunction with ECML/pKDD*. Citeseer.
Bermejo, P., Gámez, J. A., & Puerta, J. M. (2011). Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. *Expert Systems with Applications*, *38*(3), 2072–2080.
Bhattacharya, C. (1998). When customers are members: Customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, *26*(1), 31–44.
Bin, L., Peiji, S., & Juan, L. (2007). Customer churn prediction based on the decision tree in personal handyphone system service. In *2007 international conference on service systems and service management* (pp. 1–5). IEEE.
Bolton, R. N., & Bronkhorst, T. M. (1995). The relationship between customer complaints to the firm and subsequent exit behavior. *ACR North American Advances*.
Bonner, S., & Vasile, F. (2018). Causal embeddings for recommendation. In *Proceedings of the 12th ACM conference on recommender systems* (pp. 104–112).
Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.
Colbrook, M. J., Antun, V., & Hansen, A. C. (2022). The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proceedings of the National Academy of Sciences*, *119*(12), Article e2107151119.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.
Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, *34*(1), 313–327.
Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., et al. (2008). Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on extending database technology: advances in database technology* (pp. 668–677).

Dubey, H., & Pudi, V. (2013). Class based weighted k-nearest neighbor over imbalance dataset. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 305–316). Springer.

Dwiyanti, E., Ardiyanti, A., et al. (2016). Handling imbalanced data in churn prediction using rusboost and feature selection (case study: Pt. telekomunikasi Indonesia regional 7). In *International conference on soft computing and data mining* (pp. 376–385). Springer.

Ebrah, K., & Elnasir, S. (2019). Churn prediction using machine learning and recommendations plans for telecoms. *Journal of Computer and Communications*, *7*(11), 3. http://dx.doi.org/10.4236/jcc.2019.711003.

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Explainable artificial intelligence and machine learning: A reality rooted perspective. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(6), Article e1368.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Fujo, S. W., Subramanian, S., Khder, M. A., et al. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, *11*(1), 24.

Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, 1–26.

Gerpott, T. J., Rams, W., & Schindler, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, *25*(4), 249–269.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

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

Hawley, D. (2003). International wireless churn management: research and recommendations. Yankee Group report, (June), URL http://www.ams.com/cme/pdfs/yankeechurnstudy.pdf. (Accessed January 2006).

Hitt, L. M., & Frei, F. X. (2002). Do better customers utilize electronic distribution channels? The case of PC banking. *Management Science*, *48*(6), 732–748.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression: vol. 398*, John Wiley & Sons.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425.

Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., et al. (2015). Telco churn prediction with big data. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 607–618).

Ji, H., Zhu, J., Wang, X., Shi, C., Wang, B., Tan, X., et al. (2021). Who you would like to share with? a study of share recommendation in social e-commerce. In *Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 1* (pp. 232–239).

Johansson, F., Shalit, U., & Sontag, D. (2016). Learning representations for counterfactual inference. In *International conference on machine learning* (pp. 3020–3029). PMLR.

Kong, J., Kowalczyk, W., Menzel, S., & Bäck, T. (2020). Improving imbalanced classification by anomaly detection. In *International conference on parallel problem solving from nature* (pp. 512–523). Springer.

Lazarov, V., & Capota, M. (2007). Churn prediction. *Business Analysis Course. TUM Computer Science*, *33*, 34.

Leung, C. K., Pazdor, A. G., & Souza, J. (2021). Explainable artificial intelligence for data science on customer churn. In *2021 IEEE 8th international conference on data science and advanced analytics* (pp. 1–10). IEEE.

Liao, C.-H., & Lien, C.-Y. (2012). Measuring the technology gap of APEC integrated telecommunications operators. *Telecommunications Policy*, *36*(10–11), 989–996.

Liu, X., Xie, M., Wen, X., Chen, R., Ge, Y., Duffield, N., et al. (2018). A semi-supervised and inductive embedding model for churn prediction of large-scale mobile games. In *2018 ieee international conference on data mining* (pp. 277–286). IEEE.

Lundberg, S. M., & Lee, S.-I. (2017a). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc..

Maxham, J. G., III (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of Business Research*, *54*(1), 11–24.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & e Cunha, J. F. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, *39*(12), 11250–11256.

Mitrović, S., & De Weerdt, J. (2020). Churn modeling with probabilistic meta paths-based representation learning. *Information Processing & Management*, *57*(2), Article 102052.

Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, *38*(1), 131–142.

Mittal, B., & Lassar, W. M. (1998). Why do customers switch? The dynamics of satisfaction versus loyalty. *Journal of Services Marketing*, *12*(3), 177–194.

Moayer, S., & Gardner, S. (2012). Integration of data mining within a strategic knowledge management framework. *International Journal of Advanced Computer Science and Applications*, *3*(8).

Momin, S., Bohra, T., & Raut, P. (2020). Prediction of customer churn using machine learning. In *EAI international conference on big data innovation for sustainable cognitive computing* (pp. 203–212). Springer.

Naz, N. A., Shoaib, U., & Shahzad Sarfraz, M. (2018). A review on customer churn prediction data mining modeling techniques. *Indian Journal of Science and Technology*, *11*(27), 1–27.

Nguyen, N., & LeBlanc, G. (1998). The mediating role of corporate image on customers' retention decisions: an investigation in financial services. *International Journal of Bank Marketing*.

Pushpa, S. (2012). An efficient method of building the telecom social network for churn prediction. *International Journal of Data Mining & Knowled Management Process*, *2*(3), 31–39.

Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013). Telecommunication subscribers' churn prediction model using machine learning. In *Eighth international conference on digital information management* (pp. 131–136). IEEE.

Reichheld, F. F., & Sasser, W. E. (1990). Zero defeofions: Quoliiy comes to services. *Harvard Business Review*, *68*(5), 105–111.

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, *67*(1), 77–99.

Saha, L., et al. (2023). Deep churn prediction method for telecommunication industry. *Sustainability*, *15*(5), 4543.

Seymen, O. F., Dogan, O., & Hiziroglu, A. (2020). Customer churn prediction using deep learning. In *International conference on soft computing and pattern recognition* (pp. 520–529). Springer.

Shrestha, S. M., & Shakya, A. (2022). A customer churn prediction model using XGBoost for the telecommunication industry in Nepal. *Procedia Computer Science*, *215*, 652–661.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(04), 687–719.

Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, *28*(4), 667–671.

Umayaparvathi, V., & Iyakutti, K. (2017). Automated feature selection and churn prediction using deep learning models. *International Research Journal of Engineering and Technology (IRJET)*, *4*(3), 1846–1854.

Van den Poel, D., & Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*(1), 196–217.

Varki, S., & Colgate, M. (2001). The role of price perceptions in an integrated model of behavioral intentions. *Journal of Service Research*, *3*(3), 232–240.

Wei, C.-P., & Chiu, I.-T. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, *23*(2), 103–112.

Xu, F., Zhang, G., Yuan, Y., Huang, H., Yang, D., Jin, D., et al. (2021). Understanding the invitation acceptance in agent-initiated social e-commerce. In *Proceedings of the international AAAI conference on web and social media, vol. 15* (pp. 820–829).

Yabas, U., Cankaya, H. C., & Ince, T. (2012). Customer churn prediction for telecom services. In *2012 IEEE 36th annual computer software and applications conference* (pp. 358–359). http://dx.doi.org/10.1109/COMPSAC.2012.54.

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & Marketing*, *21*(10), 799–822.

Yang, C., Shi, X., Jie, L., & Han, J. (2018). I know you'll be back: Interpretable new user clustering and churn prediction on a mobile social application. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 914–922).

Yoon, J., Jordon, J., & Van Der Schaar, M. (2018). GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International conference on learning representations*.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46.

Zhang, G., Zeng, J., Zhao, Z., Jin, D., & Li, Y. (2022). A counterfactual modeling framework for churn prediction. In *Proceedings of the fifteenth ACM international conference on web search and data mining* (pp. 1424–1432).

Zhao, L., Gao, Q., Dong, X., Dong, A., & Dong, X. (2017). K-local maximum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing*, *20*, 1401–1409.

Zhu, B., Baesens, B., & vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, *408*, 84–99.