Journal of Information Systems and Telecommunication
Vol.11, No.4, October-December 2023, 331-346

http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Application of Machine Learning in the Telecommunications Industry: Partial Churn Prediction by using a Hybrid Feature Selection Approach

Fatemeh Mozaffari[1], Iman Raeesi Vanani[2], Payam Mahmoudian[1], Babak Sohrabi[1*]

1.Department of Information Technology Management, College of Management, University of Tehran, Tehran, Iran
2.Department of Industrial Management, Faculty of Management and Accounting, Allameh Tabataba'i University, Tehran, Iran

Received: 09 Jul 2022/ Revised: 04 Feb 2023/ Accepted: 07 Mar 2023

Abstract

The telecommunications industry is one of the most competitive industries in the world. Because of the high cost of customer acquisition and the adverse effects of customer churn on the company's performance, customer retention becomes an inseparable part of strategic decision-making and one of the main objectives of customer relationship management. Although customer churn prediction models are widely studied in various domains, several challenges remain in designing and implementing an effective model. This paper addresses the customer churn prediction problem with a practical approach. The experimental analysis was conducted on the customers' data gathered from available sources at a telecom company in Iran. First, partial churn was defined in a new way that exploits the status of customers based on criteria that can be measured easily in the telecommunications industry. This definition is also based on data mining techniques that can find the degree of similarity between assorted customers with active ones or churners. Moreover, a hybrid feature selection approach was proposed in which various feature selection methods, along with the crowd's wisdom, were applied. It was found that the wisdom of the crowd can be used as a useful feature selection method. Finally, a predictive model was developed using advanced machine learning algorithms such as bagging, boosting, stacking, and deep learning. The partial customer churn was predicted with more than 88% accuracy by the Gradient Boosting Machine algorithm by using 5-fold cross-validation. Comparative results indicate that the proposed model performs efficiently compared to the ones applied in the previous studies.

Keywords: Partial Churn; Churn Prediction; Machine Learning; Feature Selection; Telecommunications Industry; The Wisdom of the Crowd.

1- Introduction

Churn prediction that can be defined as identifying which customers would put an end to using a company's services may consider as the most frequent predictive task within the telecommunications industry due to several benefits it could bring in for the company such as the less cost spent on retaining the existing customer in comparison to acquiring a new one [1]. In the customer churn prediction, a churn probability would be assigned to each customer based on their historical data, which leads to identifying the targeted customers for receiving marketing retention campaigns [2].

There are several reasons which make tackling customer churn problem necessary. In today's saturated and competitive market, any company should consider the fact that customers have the option to switch to other service providers [3]. In addition to more costs involved in acquiring new customers than retaining the existing ones [4]–[6], churning might also impact the reputation of a company and could lead to its brand loss [3].

The dynamic relationship between customer satisfaction, service quality, and customer loyalty or switching behavior is the topic of many studies today. For example, satisfied customers will be more accepting of price rises which will, in turn, bring greater profits [7]. In recent years customer management research has followed major categories such as Customer Lifetime Value (CLV) which

✉ **Babak Sohrabi**
bsohrabi@ut.ac.ir

leads to the conclusion that retaining customers is the most crucial work of customer management. Moreover, due to rapid progress in computing science and data mining algorithms, new directions followed by researchers consist of applying various machine learning algorithms, such as Neural Networks, to identify valuable customers and predict customers' churn rate [8].

There are two basic approaches to reduce customer churn, namely untargeted and targeted approaches. In the targeted approach, companies will identify the customers who are likely to churn and offer direct incentives to retain them, while in the untargeted approach, superior products and mass advertisement are used to increase brand loyalty and retain customers. The targeted approach can be divided into two categories, namely, reactive and proactive approaches. In the reactive approach, the company will wait till the customer decides to terminate their subscription, and at that time, the company offers some incentives. However, in the proactive approach, the company tries to identify customers who have high risk of churn in the near future so that they can be targeted with special packages or incentives to keep them from churning. The proactive approach has the advantage of lowering incentive costs. The most important point in taking the proactive approach is to classify the customers accurately since the ineffective classification of customers will only lead to the waste of financial resources for wrongly targeted customers [9].

The customer churn prediction modeling has been studied in various industries such as retailing [10]–[14], banking [11], [15]–[17], e-commerce [18]–[20], media and social networking services [21], [22], financial services [23], [24], and telecommunications [2], [3], [25]–[31]. However, the customer churn predication still is a sophisticated process that contains many decision points for analysts [32].

In recent years, due to high costs related to acquiring new customers, saturated and dynamic market, and continuous new competitive offerings, the concentration of telecom companies has shifted to customer retention strategies [2], [33]. With an increasing interest in customer churn prediction, a wide range of machine learning classifiers have been proposed in the literature to deal with the churn prediction problem [28].

A non-contractual relationship between the customers and the companies suffers from the problem once the customers change their service provider without informing about it [10]. In such a scenario, predicting potential churners by analyzing the pattern in their behavioral features becomes a significant challenge that must be handled. Moreover, in developing a prediction model, there are various feature types with different levels of importance. Another issue that needs to be tackled is gaining an insight into which features have more predictive power and impact on customer churn. Yet another challenge is how to deal with the real-world data. Since the behavioral patterns of different churners are diverse, the company should not treat all customers in the same manner, i.e., the risk or probability of churn is not alike for all churners [3]. Therefore, embedding this difference in churn definition in the form of partial and semi-partial churners, in addition to complete churners, is a crucial point that were not considered in many previous studies in this field.

While numerous research has been conducted to develop an effective classification model and identify the customers with a high risk of churn, defining churn and, more specifically, partial churn in various industries, especially the telecom industry, still needs to be completed. In other words, as mentioned before, inaccurate classification of the customers can lead to wasting financial resources, and achieving an effective model is based on how the churn is defined and the predictability of the features. In this regard, these two points, defining partial churn and feature selection, are focused on in this study.

Encouraged by the aforementioned challenges, the key objectives of this paper are (i) defining partial churn in a new and practical way so that partial and semi-partial churners would be identified and targeted by retention strategies before total defection, (ii) identifying the most effective features in terms of greater power for discriminating between churners and non-churners by using various feature selection methods and the wisdom of the crowd as a novel approach to feature selection, and (iii) developing and evaluating an efficient churn prediction model based on the actual customer data in a telecommunications company.

Hence, this research has been designed to answer the following research questions regarding customer churn prediction:

RQ1: How can customer churn, partial churn, and semi-partial churn, as various categories with different probability of churn, be defined in the telecom company?

RQ2: What are the most predictive features that affect customer churn in the telecom company based on data mining techniques?

RQ3: How can a classification model based on advanced machine learning algorithms be developed to predict the customers with various risks of churn in the telecom company?

RQ4: What are the challenges in dealing with real data in the telecom industry, and how can they be handled to develop a prediction model?

To answer these questions, a novel definition of partial churn is presented in this study. The definition is based on customers' status, determined by usage, plan, and volume features. Four classes of customers are defined by using a data mining approach and classifying customers based on their usage behavior similarity to active or total churners.

Hence, customers can be categorized with more resolution, and retention strategies would be more purposeful. Proposing a hybrid feature selection approach is another novelty of this paper. Comparing various feature selection methods and using the wisdom of the crowd can lead to a significant increase in the accuracy of the predictive model. Finally, a predictive model based on advanced machine learning algorithms is developed, which can efficiently predict partial and semi-partial churners. The model is obtained through handling different challenges such as imbalanced datasets, data cleansing, and feature selection which are tackled in the preprocessing phase of this study. Finally, while defining the four classes of customers, the proposed churn prediction model predicts the class of new customers, i.e., test dataset, with at least 88% accuracy.

The main contribution of this study lies in the partial churn definition based on data mining techniques which can find the degree of similarity between assorted customers with active ones or churners. Also, the hybrid approach in feature selection utilized in this study sheds light on the most predictive features among many features in a real dataset.

The rest of the paper is organized as follows. In Section 2, the background and related works of the customer churn prediction concerning its different aspects are given. Section 3 specifies the methodology used in this work in terms of various steps to develop a churn prediction model. The results are given in Section 4 followed by a discussion in Section 5. Conclusion and directions for further research are in Section 6.

2- Background and Related Work

The literature review indicates about a number of researches already done in the customer churn prediction area. It also highlights techniques from basic as well as advanced algorithms that were used in different industries. This section provides an overview of related works on partial churn, feature selection for the churn prediction modeling, and classification algorithms.

2-1- Partial Churn

Although researchers often define customer churn based on the nature of the organization they examine [34]; but in general, customer churn can be defined as the customers' tendency to stop doing business with one company or organization and switch to products of another company within an assumed period [35]. One of the most fundamental points that should be considered in defining customer churn is to differentiate between contractual and non-contractual businesses since each has its own model. There are relatively few studies that define churn in non-contractual settings [36] probably due to the difficulties in determining the time when customer becomes effectively inactive, when there is no contract. From a business perspective, we can say that partial churn is even more important than the complete churn [37]. The importance of identifying partial churners is that, firstly, these customers are considered to be loyal and the loss of sales related to them, even partially, can be significant. Secondly, the partial defection can lead to complete defection in the long-term. Hence, the sooner the partial churn is detected, the more valuable it is for marketing managers to prevent them from complete churn by adopting appropriate actions [10].

Based on the literature review, most previous studies were conducted in some industries such as telecommunications and retailing, due to their non-contractual settings. Most customers show a partial defection before their complete churn, which may finally lead to a complete switch [10]. Table 1 summarizes some of the most important partial churn definitions in the literature.

Table 1: Partial churn definitions

Study	***Industry***	***Churn Definition***
[10]	Retailing	Change in transaction patterns and stopping the use of a product or service
[25]	Telecommunications	Change in customers' status
[13]	Retailing	Customers who have not made a purchase within a certain period of time or in all subsequent periods have spent less than 40% of the reference period. Also, partial churn is defined as stopping the purchase of certain goods or services
[36]	Retailing	Evaluation of a set of definitions based on economic parameters
[19]	E-commerce	Change in Length, Recency, Frequency, Monetary (LRFM) pattern and purchasing behavior
[38]	Retailing	Considering the Poisson distribution for the customer purchasing pattern, the significant difference between the rate parameter in the two time periods of reference and evaluation indicates churn
Current study	Telecommunications	The amount of similarity to either active or expired

Study	*Industry*	*Churn Definition*
		customers based on the usage, plan, and volume statuses

2-2- Feature Selection

In order to get consistent, unbiased, and a set of explanatory features, the number of features should be reduced by applying feature selection methods [39]. Filtering methods are advantageous due to their high-speed calculations, but there is no guarantee that the subset of the selected features will be optimal. Chi-square test, information gain, Fisher's Score, Anova, Minimum Redundancy Maximum Relevance (MRMR), and Linear Discriminant Analysis (LDA) are among filtering feature selection techniques [5], [33], [40]–[47]. On the other hand, wrapper methods are based on scoring possible subsets of features relying on their predictive potential. Sequential Forward Selection (SFS), Sequential Backward Elimination (SBE), and Boruta algorithm are among the most common wrapper techniques though they are intensive computationally [44], [48]–[50].
Finally, embedded algorithms have their own built-in feature selection methods. These are similar to wrapper methods in terms of dependencies on the classifier. However, embedded methods are computationally faster than wrapper ones. LASSO regression, Random Forest, and Recursive Feature Elimination-Support Vector Machine (RFE-SVM) are some examples of this type of techniques [29], [51]–[54]. Moreover, using the two-phase feature selection method by researchers is reviewed by Jain et al. (2020). The two-phase feature selection method can consist of a combination of experts-advised subset and Markov blanket discovery.
The wisdom of the crowd is an emerging concept that is used for feature selection in this research. The wisdom of the crowd concept refers to the fact that aggregated judgments by individuals are often more accurate than that of the smartest person in the crowd [55]. In this regard, crowd refers to a group of individuals with various perspectives, and wisdom contains purposeful acts, reasonable thought, and effectively dealing with an environment [55]. There are some factors that influence crowd performance, such as diversity, independence, and decentralization which were introduced by Surowiecki (2005) [56].
Possession of varying degrees of knowledge and insight explains the diversity. Independence means no influence on each member's decisions by other members, and decentralization relates to the concept that power does not fully reside in one central location [56]. Moreover, Hong et al. (2020) proposed that crowd size is a moderator between crowd characteristics and crowd performance [55]. In recent years, the wisdom of the crowd has been found to have many applications in some fields, such as the stock prediction domain and financial trade [57]–[59].
The predictive features can be used by the marketing department to implement retention strategies. While data mining techniques are used in this regard, it should be noted that using predictive features to predict and then change the churn intention can find its roots in the Theory of Planned Behaviors (TPB) introduced by Ajzen (1985). This theory, which expands the theory of reasoned action, states that many factors influence the stability of behavioral intentions. Investigating these factors sheds light on how it may be possible to prevent changes in intentions. Hence, a measure of intention is expected to allow precise prediction of intentional behavior unless the intention changes after it is evaluated but before the behavior is observed. This intention is, in turn, a function of two factors which are the attitude toward trying and the subjective norm with regard to trying [60]. The findings of this study are discussed according to this theory in the Discussion section.

2-3- Classification Algorithms

The customer churn prediction can be considered as a management science problem for which we can usually adopt a data mining approach. "Data mining is the process of discovering meaningful new correlations, patterns, and trends sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques" [61]. Since the main aim of churn prediction is to classify the type of customers (non-churner, partial churner, churner), a variety of supervised machine learning classifiers have been proposed in the literature. In this subsection, we provide an overview of some of the previous studies on churn prediction. Table 2 provides a review of the classification algorithms used in some of customer churn prediction studies in the telecommunications sector.

Table 2: Literature review on classification algorithm to predict customer churn in the telecom sector

Study	*Classification Algorithm*
[30]	Ensemble
[25]	Logistic Regression (LR)
[62]	Neural Network (NN)
[63]	NN
[64]	Support Vector Machine (SVM), NN, Decision Tree (DT)
[65]	SVM
[26]	Genetic Programming (GP) & Self-Organizing Maps
[31]	SVM, DT, Back Propagation Network (BPN)
[66]	Ordered Fuzzy Rule Induction
[67]	Fast Fuzzy C-Means & GP
[68]	Convolutional Neural Network (CNN)
[69]	Recurrent Neural Network (RNN)
[70]	CNN

Study	***Classification Algorithm***
[71]	Fuzzy classifiers
[2]	LR, DT, NN, Random Forest (RF), SVM
[72]	Long Short-Term Memory (LSTM), Gradient Boosting Tree (GBT), RF, SVM
[73]	Agent-Based Modeling and Simulation (ABMS)
[28]	RNN, LR, RF, LSTM, Probabilistic Neural Network (PNN)
[74]	DT, Naïve Bayes (NB), Rule Induction
[29]	CNN
[3]	LR, Multi-Layer Perceptron (MLP), NB, Bagging and Random Tree, AdaBoostM1, Attribute Selected, Decision Stump, RF, J48, Random Tree, Lazy learning methods (Locally Weighted Learning, lazy k-nearest neighbor)
[27]	NB, Generalized Linear Model (GLM), LR, Deep Learning (DL), DT, RF, GBT
[75]	Ensemble Classifiers (K-Nearest Neighbors, NB, RF, LR) and (Cat Boost, Gradient Boosting, Extreme Gradient Boosting)
[76]	NN, SVM, NB, RF, Adam DL

As Table 2 illustrates, advanced data mining algorithms such as bagging, boosting, and stacking are exploited in recent years due to their better performance compared to the basic ones. Boosting is a technique that improves the accuracy of the classification model by applying the functions in a series iteratively and then combining the result of each function with weighting in order to maximize the total accuracy of the prediction. Friedman (2002), constructed additive regression models by iteratively fitting a base learner or weak classifier to current pseudo-residuals with least squares. Indeed, these pseudo-residuals are the gradient of the loss function, which should be minimized [11].

Implementing Rough Set Theory (RST) which is a technique for dealing with uncertainty and for identifying cause-effect relationships in databases, has also been used for customer churn classification modeling [9], [77]. In [8], knowledge of survival analysis was applied in customer management, while Devriendt et al. (2021) used uplift modeling as prescriptive analytics, which aims a reduction in the likelihood of churn when customers are targeted with the retention campaign based on the net difference in customer behavior [78]. Using unstructured data such as text is another approach that has been used in recent years. For example, [79] utilized a call center dataset to predict customer churn risks and generate meaningful insights using interpretable machine learning with personas and customer segments, and [80] used website data holding customer complaints in this regard.

3- Data and Methodology

This section discusses the main building blocks of the proposed customer churn prediction model, which leads to identify partial churners in a telecommunications company. Since the problem is analyzed through the data mining approach, the methodology in the present study is the data science methodology introduced by IBM [81], which includes ten steps, which is an extension to the CRoss Industry Standard Process for Data Mining (CRISP-DM) method [82]. While using the data mining approach, the findings of the study are discussed through the lens of TPB in the Discussion section.

Figure 1 shows the overall framework of the data analysis. As it can be observed, this framework consists of four main steps. In the first step, a binary classifier is used to model Active versus Expired customers. Expired customers are the ones who were not using the services for more than eight months. In the second step, a partial churn definition can be derived based on the amount of similarity between any customer with either Active or Expired customers. In other words, the “statuses” of the customers are linked to the partial churn. Then, based on the partial churn definition, a multiclass problem is set up to predict whether customers are in the “partial churn” status. Finally, using Partial_Churn and Semi_Partial_Churn labels and prediction model, the most predictive features can be identified.

3-1- Data Definition

Based on the Exploratory Data Analysis (EDA), our variables and their statistical features were obtained through data summarization and visualization. Three datasets were used in this research, which contains customer data on three specific dates, i.e., August and November of 2018 and February of 2019. Table 3 presents the number of records and features of each dataset before and after data cleansing. Three dates were selected so that the customers’ status can be tracked at least in a 6-month period.

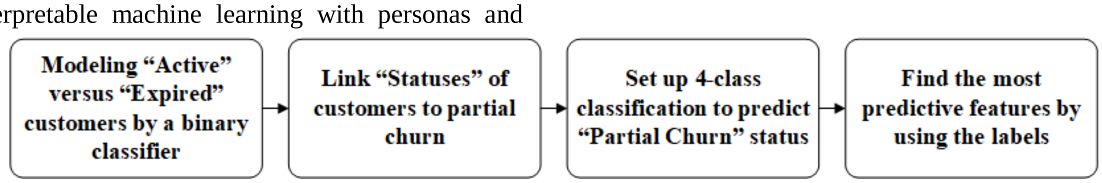

Fig. 1 Overall framework of the data analysis

Table 3: Datasets used in this research

No. of dataset	*Date*	*No. of records before data cleansing*	*No. of features before data cleansing*	*No. of records after data cleansing*	*No. of features after data cleansing*
1	August 2018	213166	110	208585	87
2	November 2018	228941	110	223970	87
3	February 2019	254994	110	249511	87

The resulted independent features consist of 31 nominal variables, nine dates, which then changed to numeric variables, and 47 numeric variables. The dependent variable, which is the STATUS of the customer, i.e., a label indicates the churned or non-churned customers, is a logical variable. In the second phase of data cleansing, some of the nominal features with too many unique values were eliminated by investigating the statistical characteristics of features. Moreover, some dates from which the Status can be derived directly were removed from the dataset, which led to 72 features. The features can be categorized into six main groups: plan-related, customer life-related, usage-related, revenue-related, sale-related, and geographic-related features- the number of each and some examples are provided in Table 4.

Table 4: Categories of features in real-world dataset used to predict customer churn

Category	*Number of features*	*Example*
Plan-related features	10	Volume, Speed, Duration
Customer Life-related features	5	Nominal Life, Real Life
Usage-related features	14	Last_Month_Usage, Last_3_Month_Usage, Burned_Traffic
Revenue-related features	19	Total_Revenue, Monthly_Revenue, Usage_Revenue
Sale-related features	9	Sale_Type, Model_Name
Geographic-related features	15	BTS_City

Training and test datasets are derived from the three datasets mentioned above. In this research, two training and two test datasets were used. The first training and test datasets were used for partial churn definition in a binary classification model, and the second ones were used in 4-class classification. The first training dataset consists of customer records whose status had not been changed during the 6-month period, i.e., from August 2018 to February 2019. In other words, these customers were labeled as Active customers during the 6-months period, which means they were using the services during this period, or they labeled as Expired, which means they were not using the services for more than eight months and maintained this status for at least six months. The second training dataset, which includes the data set of November 2018 with new labels derived from the previous step, was used for training. The second test dataset consists of the customers who were added in the three-month period and were considered new records selected for testing. In both phases, 5-fold cross-validation was applied to training datasets. Table 5 presents training and test datasets for each of the phases.

Table 5: Training and test datasets for each of the phases

phase	*No. of records of the Training dataset*	*No. of records of the Test dataset*
Partial churn definition	96934	152434
	all records from August 2018 which maintained the same status during 6 months	all records from the three dates which are not among training dataset
4-class classification	173706	43280
	all records from November 2018 which were assigned one of the four labels based on Table 6	all new records from February 2019 that were added to the dataset in the last three months

3-2- Partial Churn Definition

In this paper, different customer statuses are grouped into four: active, semi-partial churn, partial churn, and expired. In non-contractual businesses, as in our case, the churn event cannot be determined explicitly since the customer can finish the relationship with the service provider without prior notice. Therefore, we have to clearly define a churn criterion based on which partial churners can be identified before their complete defection. In other words, the churn should be predicted in advance in order to have time for customer retention strategies.

Customers with semi-partial and partial churn status are more likely to churn than customers with active status, hence; these are the customers marketing department can focus on. Optimization of the fraction of customers that should be targeted by the retention campaign can lead to maximize the profit [39]. In our case study, which is an internet service provider company, 14 statuses are defined for customers based on criteria: (i) usage behavior, (ii) having a plan (i.e., the last main plan purchased by customer), and (iii) having volume. Table 6 summarizes different groups of customers based on their statuses.

In order to define the partial churn, a binary classification model was built based on the training dataset, i.e., a dataset containing the records of active and expired customers who had not changed their status for at least a 6-month period. Therefore, Partial_Churn can be defined as a status that customers have equally shows similarity to either Active customers or Expired customers while Semi_Partial_Churn is assigned to customers with more similarity to Active customers, but they still tend to churn more than the Active ones. As explained, the definition approach is based on the similarity of customers' features with either active customers or total churners.

Table 6: Customers' status based on usage behavior, plan, and volume

Criteria	***Status***
Usage behavior	Active
	without usage 10 to 20 days (10TO20)
	without usage 20 to 30 days (20TO30)
	without usage for more than 30 days (gt30)
Plan	Without plan 0 to 1 month (0TO1)
	without plan 1 to 2 months(1TO2)
	without plan 2 to 3 months (2TO3)
	without plan 3 to 5 months (3TO5)
	without 5 to 7 months (5TO7)
	without plan 7 to 8 months (7TO8)
	without plan more than 8 months (Expired)
Volume	without volume 3 to 7 days (3TO7)
	without volume 7 to 20 days (7TO20)
	without volume more than 20 days (gt20)

Table 7 illustrates the labels which are derived after several moves back and forth between preprocessing and modeling steps in order to obtain a model with more than 99% binary classification accuracy. In Table 7, two new classes were defined: Semi_Partial_Churn and Partial_Churn. Customers labeled as Semi_Partial_Chrun are the ones whose features are more similar to the Active customers, while the customers labeled as Partial_Churn can be either Active or Expired. In other words, the probability of becoming a churner is more with the ones who labeled as Partial_Churn. Moreover, semi_partial churners tend to become churners more than the active ones. Therefore, these labels can categorize the customers with more resolution. The statuses provided in Table 7 are the same as those in Table 6 except for "gt20". This status was removed since it could not be in any of the classes mentioned above.

Table 7: Partial churn definition based on customers' status

Status	***Label***
Active 0TO1 3TO7 7TO20 10TO20	ACTIVE (classified as active with more than 95% probability)
1TO2 gt30	Semi_Partial_Churn (classified as active with 80 to 95 percent probability)
2TO3	Partial_Churn (classified as active (or expired) with nearly 50% probability
3TO5 5TO7 7TO8 Expired	EXPIRED (classified as expired with more than 95% probability)

3-3- Preprocessing

Data preparation methods often refer to the transformation of features into variables that support a particular machine learning algorithm [2]. In the process of knowledge discovery and developing a practical model to predict churn, the data preparation and the feature selection are essential steps. In other words, selecting the most valuable features can reduce over fitting as well as the complexity of the model while improving the interpretability for users [83]. The preprocessing step in this study consists of initial data cleansing, data transformation, normalization, One-Hot encoding, feature selection, and the handling of imbalanced dataset problem. Figure 2 depicts the main tasks of the data preprocessing used in this study.

Data cleansing includes finding the test records, which do not relate to customers, removing the inaccurate data caused by system bugs and data entry mistakes, replacing the inconsistent and missing values, and eliminating irrelevant and zero variance features, all done using experts' knowledge about the data. For example, most of the missing values are replaced by using other databases of the company. As a preprocessing task, the z-score normalization is applied on the dataset to make them appropriate for correlation analysis, i.e., all numerical features are standardized by removing the mean and scaling the values to unit variance. For nominal variables, One-Hot encoding is applied so that each of these variables can be represented with as many binary variables as the number of unique values of that variable. It is worth to note that most of the preparations such as reducing skewness of features distribution, normalization, and a One-Hot encoding are applied to the data for the correlation analysis purpose and not needed when the H2O package is used for modeling since H2O AutoML-function applies all the required preprocessing in accordance with each algorithm before modeling [84].

In many data mining applications such as the churn prediction, rare cases or minority classes, are of main interest [11]. It may cause classification methods to experience challenges in identifying the churners, which leads to the poor classification power. The best performance of classification techniques can be achieved when the class distribution is approximately even [39]. Considerable works have been done on handling the imbalanced dataset problem [47], [85]–[92] which categorize the methods into oversampling, synthetic data generation which is a type of oversampling, under-sampling, and cost-sensitive learning ones. The Synthetic Minority Oversampling Technique (SMOTE) is one of the synthetic oversampling techniques used in the literature [93], which is also applied in the current study due to its efficiency. Indeed, oversampling methods, on the one hand, focus on improving classifiers' performance on the minority class samples; on the other, boosting methods focus on the hard-to-learn majority class samples. Therefore, as proposed by Barua et al. (2012), boosting and oversampling together can provide an efficient option for learning the imbalanced data [85]. This approach is also useful in this study since we apply a boosting algorithm, i.e., gradient boosting machine, together with oversampling for handling the imbalanced dataset challenge. Therefore, oversampling based on SMOTE technique is adopted in this study to handle imbalanced dataset.

3-3-1- Feature Selection

In this study, feature selection has played a crucial role in achieving the results. So, as a part of the research, the result of some common feature selection methods from various categories, i.e., filtering, wrapper, and embedded methods, are compared to each other. Therefore, the correlation analysis by using Pearson correlation coefficient, Boruta, LASSO, Ridge, and Random Forest methods are used. Moreover, as mentioned before, the concept of the wisdom of the crowd is applied. The individuals who have expert knowledge of the features are from various departments, i.e., marketing, data management, and Customer Relationship Management (CRM). The important features selected by these individuals are then compared to the features selected by algorithms.

As shown in Table 8, two different approaches are used for feature selection: the algorithm-based approach and the wisdom-of-the-crowd-based approach. The features selected by the wisdom of the crowd can help us in two ways. First, in case that there is a difference between the results of various algorithms, it would be observed that using the crowd's wisdom can be helpful. Secondly, most of the features considered as important by the wisdom of the crowd are in the final list of important features with the best performance of the model in terms of accuracy. So, it can be used as a useful method that complements the feature selection algorithms' results.

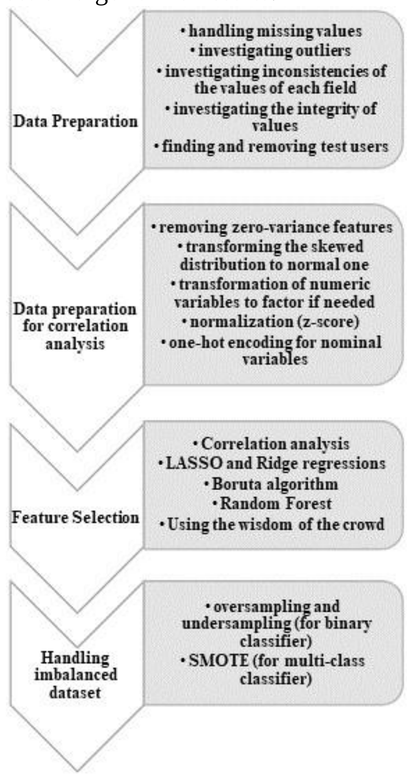

Fig. 2 Preprocessing step for developing a customer churn prediction model

Table 8 presents that some features are considered important by all methods, for example, PLAN_GROUP_NAME, some other features are considered important by some of the methods while their importance weight is low by the others, like REAL_LIFE or BURNED_TRAFFIC. Some features are not considered predictive by the wisdom of the crowd, while feature selection methods considered them important and, therefore, predictive ones such as INSTALLATION_AGENT_NAME. Based on the results by various methods, a hybrid approach was adopted. Based on this hybrid approach, different subsets of features were used to train the prediction model to find the most effective subset, leading to the highest prediction accuracy. The difference between these subsets is based on the selection criterion.

Table 8: Applying various feature selection methods

Feature	***LASSO (absolute coefficients)***	***Ridge (absolute coefficients)***	***RF (weights)***	***Boruta (confirmed/rejected and mean importance)***	***Crowd's wisdom***
CUSTOMER_TYPE	0.8272	0.2385	0.00	Rejected/ -0.2852793	✓
PLAN_GROUP_NAME	0.2057	0.0879	0.40	Confirmed/ 5.9657923	✓
LAST_MONTH_USAGE	0.0002	1.89529E-05	0.47	Confirmed/ 25.9943734	✓
LAST_MONTH_EXTRA_TRAFFIC_REVENUE	0	9.0333E-07	8.87E-05	Rejected/ 1.8492914	✓
BURNED_TRAFFIC	6.72E-06	5.78202E-07	0.00	Confirmed/ 7.5701373	✓
INSTALLATION_AGENT_NAME	0.3188	0.1434	0.07	Confirmed/ 3.9381313	-
REAL_LIFE	0	0.0045	0.72	Confirmed/ 11.5165463	✓

3-4- Gradient Boosting Machine (GBM)

Advanced data mining algorithms are those implemented in the H2O library. The H2O is an open-source advanced machine learning tool that helps us create high-performance models. The algorithms implemented in the H2O library, which have been used for the classification and the churn prediction, include Gradient Boosting Machine (GBM), Stacked Ensemble, Deep Learning (DL) as a fully connected multi-layered artificial neural network, Distributed Random Forest (DRF), Extremely Randomized Trees (XRT), Generalized Linear Model (GLM), and XGBoost [84]. After applying AutoML function on our dataset, GBM outperforms the other algorithms in terms of prediction accuracy, precision vs. recall, Area Under Curve (AUC), and logloss. In the H2O package, the GBM algorithm provided by Ellis et al. (2008) is implemented [94]. This algorithm uses distributed trees in such a way that a tree node is assigned to each row. An in-memory map-reduce task calculates statistical parameters such as the Least Mean Square Error (MSE) and uses it to make an algorithm-based decision [95].

3-5- Evaluation Measures

In this study, in order to evaluate the performance of the classification model, we have used different performance metrics such as accuracy, precision, and recall, which are derived from the confusion matrix and logloss. Area Under Curve (AUC), i.e., the area under the Receiver Operating Characteristic Curve (ROC), is another useful metric for a binary classification since it is not sensitive to imbalanced classes [27]. AUC can be defined as "the estimated probability that a randomly chosen churner has a higher posterior churn probability than a randomly selected non-churner" [2]. Therefore, we have used this metric in the partial churn definition phase.
All the preprocessing, classification, and prediction implementations were done using R language version 3.6.1 mainly with tidyverse [96], recipes [97], and H2O [84] packages. For feature selection, packages such as glmnet [98], Boruta [99], and randomForest [100] were used. The experiment was performed using the classification algorithms of the H2O package in R.

4- Results

This section explores the results obtained through the experimental analysis. Further, the results are used to evaluate the impacts of feature selection, oversampling, and classification algorithms. Figures 3 represent the visualization of comparison between various advanced machine learning algorithms in terms of ROC.
Since GBM outperforms other algorithms with respect to ROC plots, we could select GBM as the best classifier on predicting the correct customer churn for our dataset. We adopted an oversampling technique during the training to handle the imbalanced dataset problem. This led to an increase in the churn rate from 20% to nearly 50%.
Therefore, an approximately uniform distribution of the target variable affects the churn prediction performance with an improvement in the accuracy measured by the ratio of true positive/negative to the total number of samples. Table 9 presents the overall performance of the churn prediction model, which classifies the new customers into four classes with over 88% accuracy for each, and more than 97% overall accuracy rate.

Table 9: Performance of final multi-class GBM classifier

Label (target variable)	***Prediction accuracy***	***Overall accuracy***
ACTIVE	0.9908	97.6%
Semi_Partial_Churn	0.9006	
Partial_Churn	0.8843	
EXPIRED	0.9596	

The features which were selected through the mechanism explained in Section 3.3.1 are provided in Table 10 after the interpretation step in order to filter them. The results indicate that the 31-numeric features are the best subset of features in terms of accuracy, AUC, and other metrics. 27 out of these 31, which means 87% of the features are amongst the features selected by the wisdom of the crowd. In our case study, the customers classified as Semi_Partial_Churn and Partial_Churn were 10% of the total in test dataset, which equals nearly 4,000 customers who should be targeted for marketing campaigns.

4-1- Interpretation of the Results

To understand the contribution of each feature to the prediction result, we use the Locally Interpretable Model-agnostic Explanations (LIME) package [101]. This is implemented in R language to find the k-most important features which lead to the obtained result for each customer. Figure 4 depicts a sample of the LIME diagram, a local interpreter, and specifies the importance of features in different ranges. In this Section, we want to infer some insights by comparing each feature selection method's results to the results derived by the LIME interpreter. In Figure 4, the blue color indicates the feature positively affecting the category it belongs to, i.e., Active, Partial Churn, or Expired, while the red color shows the contrast. The color intensity illustrates the importance of that feature. Each feature's weight obtained by the LIME interpreter is based on that feature's role in the prediction model. Table 9 shows some of the most predictive features after the interpretation of the results.

The results indicate that among nominal features, the type of the last plan purchased by the customer has a high weight in prediction. Another predictive feature is the usage-related revenue, i.e., dividing the total revenue of the purchased plans to the usage of the customer. The total usage, the number of transactions, and the remaining volume from the last plan are also among the most predictive features. The length of the relationship between customer and company is also found to be important.

Table 10: The most predictive features for customer churn prediction

Category	***Feature***
Revenue	Last month usage-related revenue
Plan	Duration of the last plan
Usage	Remained volume of the last plan
Revenue	Last year average total revenue
Plan	Volume of the last plan
Customer Life	Life of the customer (length of the relationship with the company)
Usage	Burned traffic of the last plan
Usage	Last month usage
Plan	Cluster of the plan

5- Discussion

This section highlights the contributions made to the existing knowledge by comparing the results obtained in Section 4 to some of the previous studies, followed by the current research findings. Finally, we conclude keeping in mind the limitation to the study.

5-1- Contribution to the Existing Knowledge

As described in Section 3.2, we have defined the partial churn in a new way, which was not used by any of the previous studies, to the best of our knowledge. This definition not only exploits the status of customers based on the criteria that can be measured easily in the telecommunications industry, especially internet service provider businesses, but also by defining four classes of customers narrows the targeted customers in accordance with their risk of churning. Hence, retention campaigns would be more purposeful.

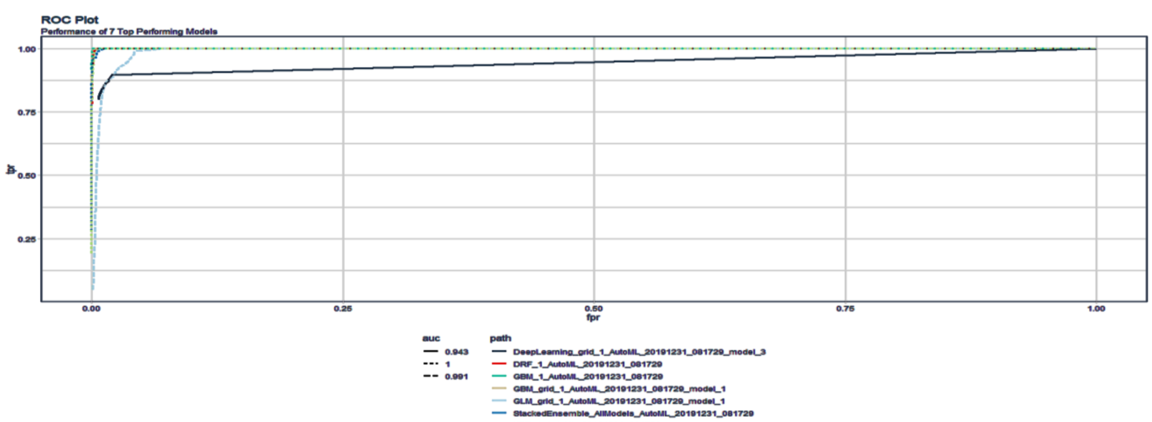

Fig. 3 Comparison between advanced classification algorithms in terms of AUC

As other studies [31] proposed, we too confirmed that boosting can improve the performance of the classifier. Moreover, gradient boosted regression trees can have a better performance compared to other advanced data mining algorithms. This is in line with the results provided by [27], and, therefore, is recommended. Moreover, we also found the features related to recency, frequency, and monetary to be among the best predictors to distinguish between different classes of customers which is in line with other studies [10], [12], [19]. Other features such as customer usage-related ones and the length of relationship are important as well. A complete list of the most important features is presented in Table 10.

The proposed feature selection approach is another important step used to achieve high accuracy and an efficient prediction model. For the feature selection approach, we used at least one method from each category of techniques in order to compare the results and select the features, which are recognized as important by most methods. Additionally, we applied the interpretation of the results and the wisdom of the crowd in the feature selection process. The results showed that the aggregation of this wise crowd's opinions can be used as a complement to the results of feature selection algorithms. Especially in the cases that the results of the algorithms contradict each other, we can rely on the wisdom of the crowd. It can also be observed that 87% of the most important features that form the best subset of features are the same as the features selected by the wisdom of the crowd. So, using the wisdom of the crowd can contribute to the feature selection procedure in two ways. First, if there is any contradiction between the results of various algorithms, it would be observed that using the crowd's wisdom can be helpful. For example, BURNED_TRAFFIC and REAL_LIFE were not selected as important features by LASSO and Ridge algorithms while confirmed as important ones by Boruta and also by the wisdom of the crowd. It can be observed that these features are among the most predictive features in Table 10, which is the final list of features. Secondly, most of the features considered as important by the wisdom of the crowd are in the final list of important features with the best performance of the model in terms of accuracy. Therefore, it can be used as a useful method that complements the feature selection algorithms' results.

Table 11 provides a performance comparison of the present study with some of the other studies in terms of overall accuracy. It can be observed from comparative results that our prediction model performs efficiently as compared to the previously used techniques. The last but not the least, since our experiment was based on the real-world dataset, we faced many challenges to tackle.

As a matter of fact, the results can be considered as applicable to the business. As mentioned, according to the Theory of Planned Behavior, behavioral attitude, subjective norms, and perceived behavioral control influence intention toward the behavior. This cognitive model has become applicable in many research areas, such as predicting loyalty intention. Based on various studies, it was found that behavior attitude is the strongest predictor of intention, while some authors have resembled switching costs as the perceived behavioral control of the TPB because of its ability to predict customer loyalty [102].

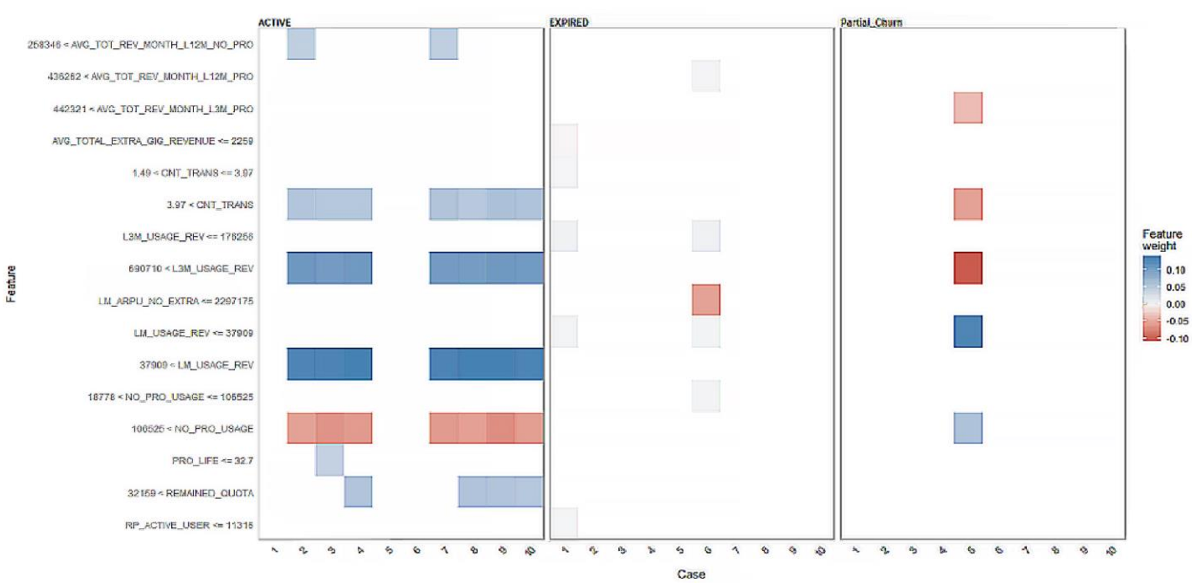

Fig. 4 Interpretation of prediction results with LIME

In recent years, due to the increasing availability of data on customer activities, more complex metrics have been developed to describe customer behavior and how behavioral attributes can be linked to customer retention and company performance [103]. These behavioral attitudes influence the churn intention and can be used as early signs of churn, which can be handled by different retention strategies that affect behavioral attitudes. Based on the findings of this study, Table 10 indicates the most predictive features extracted by the hybrid feature selection approach and interpretation of the prediction model. Among these features, 3 of them, such as last month usage, burned traffic, and remained volume of the last plan, directly relate to customer usage behavior, while the features in plan and revenue categories indirectly indicate this behavior. In other words, by investigating these signs in customer usage behavior, the service provider can offer personalized plans and incentives to customers to influence their attitudes and retain them.

Table 11: Comparison of predictive performance of proposed and previous approaches

Model	***Accuracy***					
	Present study	[31]	[19]	[27]	[3]	[76]
NB	-	-	-	0.7	0.48	0.98
GLM	0.838	-	-	0.8	-	
LR	-	-	-	0.8	0.71	
DL	0.89	-	-	0.7	-	
RF (DRF)	0.941	-	-	0.7	0.89	0.99
Stacked Ensemble/DT Ensemble	0.974	-	0.97	-	-	
GBT (GBM)	0.976	-	-	0.8	-	
Bagging + Random Tree	-	-	-	-	0.89	
BPN/ANN	-	0.95	0.935	-	-	0.974
SVM-Radial Basis Function (RBF)	-	0.96	-	-	-	0.976
SVM-POLY (polynomial kernel)	-	0.968	-	-	-	
DT/DT-C5.0	-	0.95	0.94	0.7	-	
J48	-	-	-	-	0.88	
Adam						0.98

5-2- Implications and Limitations

The results of the present study have some implications for the practice in the churn prediction area of the telecommunications company, such as to improve satisfaction by targeting customers, which are identified as partial or semi-partial churners as well as managing customer's expectations by specifying effective features on customer churn by using proposed feature selection approach through combining various feature selection algorithms and complement it by the wisdom of the crowd. Like others, this study has shortcomings and limitations as well. For that matter, all experimental evaluations are based on a specific real-world dataset and confined to the telecom industry so that not only the results can be verified by other studies but it can also be applied in practice. It must be noted that some of the very important sources of data, such as those on marketing campaigns and call centers, are not used in the process of churn prediction because of their incompleteness.

6- Conclusion and Future Work

The churn prediction in the present telecom market is a compelling issue of the CRM [3] which can be conducted by identifying likely churn customers and providing competitive offers to them. In this article, we first dealt with this problem and its different related aspects by using the real-world data extracted from various sources of a telecommunications company. Then, we proposed that partial churners, i.e., potential ones with medium to high risk of churn, could be identified prior to their decision to a complete churn by analyzing the patterns e.g., customers' usage, revenue-, plan-related features. We used state-of-the-art classification techniques for the customer churn prediction problem for a real-world dataset of an internet service provider company. It is clear from the comparative results that the gradient boosting machine performed better than other classification algorithms in predicting the customer churn. Further, this work sheds some light on the features that should be considered as more important, and it is observed that customer's usage- and plan-related features have more importance and predictive power than other types. While investigating the impact of oversampling technique SMOTE on the performance of the prediction model, the results of the current study suggested that the classifiers could achieve a significantly improved performance applying an oversampling method which also supported the findings of some previous studies [27], [28]. Consequently, this study is unique when it comes to the partial churn definition and its feature selection approach, which uses the feature selection algorithms complemented by the wisdom of the crowd, leading to the high accuracy of the prediction model.

Still some areas need further study. In any future research, we would like to use other feature types such as marketing data and other advanced machine learning algorithms such as CNN and RNN suggested by [29] and [28] respectively.

References

[1] S. Mitrović, B. Baesens, W. Lemahieu, and J. De Weerdt, "On the operational efficiency of different feature types for telco Churn prediction," Eur. J. Oper. Res., vol. 267, no. 3, pp. 1141–1155, 2018.
[2] K. Coussement, S. Lessmann, and G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," Decis. Support Syst., vol. 95, pp. 27–36, 2017.
[3] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE Access, vol. 7, pp. 60134–60149, 2019.
[4] J. Dyche, The CRM handbook: A business guide to customer relationship management. Addison-Wesley Professional, 2002.
[5] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in 2012 15th International Multitopic Conference (INMIC), 2012, pp. 23–27.
[6] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Syst. Appl., vol. 38, no. 3, pp. 2354–2364, 2011.
[7] L. Geiler, S. Affeldt, and M. Nadif, "An effective strategy for churn prediction and customer profiling," Data Knowl. Eng., vol. 142, p. 102100, 2022.
[8] Y. Chen, L. Zhang, Y. Zhao, and B. Xu, "Implementation of penalized survival models in churn prediction of vehicle insurance," J. Bus. Res., vol. 153, pp. 162–171, 2022.
[9] M. Makhtar, S. Nafis, M. A. Mohamed, M. K. Awang, M. N. A. Rahman, and M. M. Deris, "Churn classification model for local telecommunication company based on rough set theory," J. Fundam. Appl. Sci., vol. 9, no. 6S, pp. 854–868, 2017.
[10] W. Buckinx and D. Van den Poel, "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," Eur. J. Oper. Res., vol. 164, no. 1, pp. 252–268, 2005.
[11] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," Expert Syst. Appl., vol. 36, no. 3, pp. 4626–4636, 2009.
[12] A. Dingli, V. Marmara, and N. S. Fournier, "Comparison of Deep Learning Algorithms to Predict Customer Churn within a Local Retail Industry," Int. J. Mach. Learn. Comput., vol. 7, no. 5, 2017.
[13] V. L. Miguéis, D. Van den Poel, A. S. Camanho, and J. F. e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," Expert Syst. Appl., vol. 39, no. 12, pp. 11250–11256, 2012.
[14] V. L. Miguéis, A. Camanho, and J. F. e Cunha, "Customer attrition in retailing: an application of multivariate adaptive regression splines," Expert Syst. Appl., vol. 40, no. 16, pp. 6225–6232, 2013.
[15] Y. Chen, Y. R. Gel, V. Lyubchich, and T. Winship, "Deep ensemble classifiers and peer effects analysis for churn forecasting in retail banking," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2018, pp. 373–385.
[16] N. Glady, B. Baesens, and C. Croux, "Modeling churn using customer lifetime value," Eur. J. Oper. Res., vol. 197, no. 1, pp. 402–411, 2009.
[17] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Syst. Appl., vol. 36, no. 3, pp. 5445–5449, 2009.
[18] N. Gordini and V. Veglio, "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry," Ind. Mark. Manag., vol. 62, pp. 100–107, 2017.
[19] A. D. Rachid, A. Abdellah, B. Belaid, and L. Rachid, "Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context," Int. J. Electr. Comput. Eng., vol. 8, no. 4, p. 2367, 2018.
[20] A. Tamaddoni, S. Stakhovych, and M. Ewing, "The impact of personalised incentives on the profitability of customer retention campaigns," J. Mark. Manag., vol. 33, no. 5–6, pp. 327–347, 2017.
[21] I. Adaji and J. Vassileva, "Predicting churn of expert respondents in social networks using data mining techniques: a case study of stack overflow," in 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 2015, pp. 182–189.
[22] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert Syst. Appl., vol. 34, no. 1, pp. 313–327, 2008.
[23] D. F. Benoit and D. Van den Poel, "Improving customer retention in financial services using kinship network information," Expert Syst. Appl., vol. 39, no. 13, pp. 11435–11442, 2012.
[24] M. Á. de la Llave, F. A. López, and A. Angulo, "The impact of geographical factors on churn prediction: an application to an insurance company in Madrid's urban area," Scand. Actuar. J., vol. 2019, no. 3, pp. 188–203, 2019.
[25] J.-H. Ahn, S.-P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," Telecomm. Policy, vol. 30, no. 10–11, pp. 552–568, 2006.
[26] H. Faris, B. Al-Shboul, and N. Ghatasheh, "A genetic programming based framework for churn prediction in telecommunication industry," in International Conference on Computational Collective Intelligence, 2014, pp. 353–362.
[27] A. S. Halibas, A. C. Matthew, I. G. Pillai, J. H. Reazol, E. G. Delvo, and L. B. Reazol, "Determining the Intervening Effects of Exploratory Data Analysis and Feature Engineering in Telecoms Customer Churn Modelling," in 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1–7.
[28] J. Hu et al., "pRNN: A recurrent neural network based approach for customer churn prediction in telecommunication sector," in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4081–4085.
[29] M. Karanovic, M. Popovac, S. Sladojevic, M. Arsenovic, and D. Stefanovic, "Telecommunication Services Churn Prediction-Deep Learning Approach," in 2018 26th Telecommunications Forum (TELFOR), 2018, pp. 420–425.

[30] A. Lemmens and C. Croux, “Bagging and boosting classification trees to predict churn,” *J. Mark. Res.*, vol. 43, no. 2, pp. 276–286, 2006.

[31] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” Simul. Model. Pract. Theory, vol. 55, pp. 1–9, 2015.

[32] E. Lima, C. Mues, and B. Baesens, “Monitoring and backtesting churn models,” Expert Syst. Appl., vol. 38, no. 1, pp. 975–982, 2011.

[33] A. Amin et al., “Customer churn prediction in the telecommunication sector using a rough set approach,” Neurocomputing, vol. 237, pp. 242–254, 2017.

[34] A. Hiziroglu and O. F. Seymen, “Modelling Customer Churn Using Segmentation and Data Mining.,” in DB&IS, 2014, pp. 259–271.

[35] V. Bhambri, “Data mining as a tool to predict churn behavior of customers,” Int. J. Manag. Res., pp. 59–69, 2013.

[36] M. Clemente-Císcar, S. San Matías, and V. Giner-Bosch, “A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings,” Eur. J. Oper. Res., vol. 239, no. 1, pp. 276–285, 2014.

[37] T. Mutanen, V. Österlund, and R. Kinnunen, “Monitoring service adaptation and customer churn in the beginning phase of a new service,” in Fourth International Conference on Data Analytics, DATA ANALYTICS 2015, 2015, pp. 69–73.

[38] D. Ringbeck, D. Smirnov, and A. Huchzermeier, “Proactive Retention Management in Retail: Field Experiment Evidence for Lasting Effects,” Available SSRN 3378498, 2019.

[39] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New insights into churn prediction in the telecommunication sector: A profit driven data mining approach,” Eur. J. Oper. Res., vol. 218, no. 1, pp. 211–229, 2012.

[40] A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” J. Big Data, vol. 6, no. 1, p. 28, 2019.

[41] B. Bonev, F. Escolano, and M. Cazorla, “Feature selection, mutual information, and the classification of high-dimensional patterns,” Pattern Anal. Appl., vol. 11, no. 3–4, pp. 309–319, 2008.

[42] A. De Caigny, K. Coussement, and K. W. De Bock, “A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees,” Eur. J. Oper. Res., vol. 269, no. 2, pp. 760–772, 2018.

[43] T.-H. Hsu, C.-C. Chen, M.-F. Chiang, K.-W. Hsu, and W.-C. Peng, “Inferring potential users in mobile social networks,” in 2014 International Conference on Data Science and Advanced Analytics (DSAA), 2014, pp. 347–353.

[44] S. Maldonado, Á. Flores, T. Verbraken, B. Baesens, and R. Weber, “Profit-based feature selection using support vector machines–General framework and an application for customer retention,” Appl. Soft Comput., vol. 35, pp. 740–748, 2015.

[45] A. K. Meher, J. Wilson, and R. Prashanth, “Towards a large scale practical churn model for prepaid mobile markets,” in Industrial Conference on Data Mining, 2017, pp. 93–106.

[46] K. B. Subramanya and A. Somani, “Enhanced feature mining and classifier models to predict customer churn for an E-retailer,” in 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, 2017, pp. 531–536.

[47] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, “Feature selection with high-dimensional imbalanced data,” in 2009 IEEE International Conference on Data Mining Workshops, 2009, pp. 507–514.

[48] M. B. Kursa and W. R. Rudnicki, “Feature selection with the Boruta package,” J Stat Softw, vol. 36, no. 11, pp. 1–13, 2010.

[49] H. Li, C.-J. Li, X.-J. Wu, and J. Sun, “Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine,” Appl. Soft Comput., vol. 19, pp. 57–67, 2014.

[50] H. Xu, Z. Zhang, and Y. Zhang, “Churn prediction in telecom using a hybrid two-phase feature selection method,” in 2009 Third International Symposium on Intelligent Information Technology Application, 2009, vol. 3, pp. 576–579.

[51] K. Cao and P. Shao, “Customer churn prediction based on svm-rfe,” in 2008 International Seminar on Business and Information Management, 2008, vol. 1, pp. 306–309.

[52] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature extraction: foundations and applications, vol. 207. Springer, 2008.

[53] Y. Li and G. Xia, “The explanation of support vector machine in customer churn prediction,” in *2010 International Con*ference on E-Product E-Service and E-Entertainment, 2010, pp. 1–4.

[54] Y. Saeys, T. Abeel, and Y. Van de Peer, “Robust feature selection using ensemble feature selection techniques,” in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2008, pp. 313–325.

[55] H. Hong, Q. Ye, Q. Du, G. A. Wang, and W. Fan, “Crowd characteristics and crowd wisdom: Evidence from an online investment community,” J. Assoc. Inf. Sci. Technol., vol. 71, no. 4, pp. 423–435, 2020.

[56] J. Surowiecki, The wisdom of crowds. Anchor, 2005.

[57] W. Pan, Y. Altshuler, and A. Pentland, “Decoding social influence and the wisdom of the crowd in financial trading network,” in 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 2012, pp. 203–209.

[58] A. Bari, P. Peidaee, A. Khera, J. Zhu, and H. Chen, “Predicting financial markets using the wisdom of crowds,” in 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), 2019, pp. 334–340.

[59] X. Wu, Q. Ye, Y. Jin, and Y. Li, “Wisdom of Experts and Crowds: Different Impacts of Analyst Recommendation and Online Search on the Stock Market.,” in PACIS, 2019, p. 129.

[60] I. Ajzen, “From intentions to actions: A theory of planned behavior,” in Action control, Springer, 1985, pp. 11–39.

[61] D. T. Larose and C. D. Larose, Discovering knowledge in data: an introduction to data mining, vol. 4. John Wiley & Sons, 2014.

[62] C.-F. Tsai and Y.-H. Lu, “Customer churn prediction by hybrid neural networks,” Expert Syst. Appl., vol. 36, no. 10, pp. 12547–12553, 2009.

[63] P. C. Pendharkar, “Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services,” Expert Syst. Appl., vol. 36, no. 3, pp. 6714–6720, 2009.

[64] B. Q. Huang, T.-M. Kechadi, B. Buckley, G. Kiernan, E. Keogh, and T. Rashid, “A new feature set with new window techniques for customer churn prediction in land-line telecommunications,” Expert Syst. Appl., vol. 37, no. 5, pp. 3657–3665, 2010.
[65] C. Orsenigo and C. Vercellis, “Combining discrete SVM and fixed cardinality warping distances for multivariate time series classification,” Pattern Recognit., vol. 43, no. 11, pp. 3787–3794, 2010.
[66] N. Kamalraj and A. Malathi, “An Ordered Fuzzy Rule Induction Based Churn Mining For Telecom Industry,” ICIREIE 2015, p. 17, 2015.
[67] B. Al-Shboul, H. Faris, and N. Ghatasheh, “Initializing genetic programming using fuzzy clustering and its application in churn prediction in the telecom industry,” Malaysian J. Comput. Sci., vol. 28, no. 3, pp. 213–220, 2015.
[68] J. Zaratiegui, A. Montoro, and F. Castanedo, “Performing highly accurate predictions through convolutional networks for actual telecommunication challenges,” arXiv Prepr. arXiv1511.04906, 2015.
[69] A. Rodan and H. Faris, “Echo state network with SVM-readout for customer churn prediction,” in 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2015, pp. 1–5.
[70] A. Wangperawong, C. Brun, O. Laudy, and R. Pavasuthipaisit, “Churn analysis using deep convolutional neural networks and autoencoders,” arXiv Prepr. arXiv1604.05377, 2016.
[71] M. Azeem, M. Usman, and A. C. M. Fong, “A churn prediction model for prepaid customers in telecom using fuzzy classifiers,” Telecommun. Syst., vol. 66, no. 4, pp. 603–614, 2017.
[72] F. Khan and S. S. Kozat, “Sequential churn prediction and analysis of cellular network users—A multi-class, multi-label perspective,” in 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1–4.
[73] D. Bell and C. Mgbemena, “Data-driven agent-based exploration of customer behavior,” Simulation, vol. 94, no. 3, pp. 195–212, 2018.
[74] L. M. Qaisi, A. Rodan, K. Qaddoum, and R. Al-Sayyed, “Customer churn prediction using data mining approach,” in 2018 Fifth HCT Information Technology Trends (ITT), 2018, pp. 348–352.
[75] Y. Beeharry and R. Tsokizep Fokone, “Hybrid approach using machine learning algorithms for customers’ churn prediction in the telecommunications industry,” Concurr. Comput. Pract. Exp., p. e6627, 2021.
[76] S. Baghla and G. Gupta, “Performance Evaluation of Various Classification Techniques for Customer Churn Prediction in E-commerce,” Microprocess. Microsyst., vol. 94, p. 104680, 2022.
[77] M. A. Khan, M. A. I. Khan, M. Aref, and S. F. Khan, “Cluster & rough set theory based approach to find the reason for customer churn,” Int. J. Appl. Bus. Econ. Res, vol. 14, no. 1, pp. 439–455, 2016.
[78] F. Devriendt, J. Berrevoets, and W. Verbeke, “Why you should stop predicting customer churn and start using uplift models,” Inf. Sci. (Ny)., vol. 548, pp. 497–515, 2021.
[79] N. N. Y. Vo, S. Liu, X. Li, and G. Xu, “Leveraging unstructured call log data for customer churn prediction,” Knowledge-Based Syst., vol. 212, p. 106586, 2021.
[80] B. Erkayman, E. Erdem, T. Aydin, and Z. Mahmat, “New Artificial intelligence approaches for brand switching decisions,” Alexandria Eng. J., vol. 63, pp. 625–643, 2023.
[81] J. B. Rollins, “Foundational methodology for data science,” Domino Data Lab, Inc., Whitepaper, 2015.
[82] P. Chapman et al., “The CRISP-DM user guide,” in 4th CRISP-DM SIG Workshop in Brussels in March, 1999.
[83] I. Guyon and A. Elisseeff, “An introduction to feature extraction,” in Feature extraction, Springer, 2006, pp. 1–25.
[84] M. Landry and B. Angela, “Machine Learning with R and H2O,” Mt. View, CA, 2018.
[85] S. Barua, M. M. Islam, X. Yao, and K. Murase, “MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 2, pp. 405–425, 2012.
[86] P. Cao, O. Zaiane, and D. Zhao, “A measure optimized cost-sensitive learning framework for imbalanced data classification,” in Biologically-Inspired Techniques for Knowledge Discovery and Data Mining, IGI Global, 2014, pp. 48–75.
[87] V. Effendy and Z. K. A. Baizal, “Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest,” in 2014 2nd International Conference on Information and Communication Technology (ICoICT), 2014, pp. 325–330.
[88] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches,” IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev., vol. 42, no. 4, pp. 463–484, 2011.
[89] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in International conference on intelligent computing, 2005, pp. 878–887.
[90] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008, pp. 1322–1328.
[91] T. M. Khoshgoftaar, M. Golawala, and J. Van Hulse, “An empirical study of learning from imbalanced data using random forest,” in 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), 2007, vol. 2, pp. 310–317.
[92] X.-Y. Liu, J. Wu, and Z.-H. Zhou, “Exploratory undersampling for class-imbalance learning,” IEEE Trans. Syst. Man, Cybern. Part B, vol. 39, no. 2, pp. 539–550, 2008.
[93] N. V Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, “SMOTEBoost: Improving prediction of the minority class in boosting,” in European conference on principles of data mining and knowledge discovery, 2003, pp. 107–119.
[94] J. Elith, J. R. Leathwick, and T. Hastie, “A working guide to boosted regression trees,” J. Anim. Ecol., vol. 77, no. 4, pp. 802–813, 2008.
[95] M. Malohlava and A. Candel, “Gradient boosting machine with H2O.” H20 Booklet, http://docs. h2o. ai/h2o/latest-stable/h2o-docs/booklets …, 2017.
[96] H. Wickham and M. H. Wickham, “Package tidyverse,” Easily Install Load ‘Tidyverse, 2017.
[97] M. Kuhn and H. Wickham, “Recipes: preprocessing tools to

create design matrices.” 2018.
[98] J. Friedman, T. Hastie, R. Tibshirani, and B. Narasimhan, “Package ‘glmnet,’” CRAN R Repositary, 2021.
[99] M. B. Kursa, W. R. Rudnicki, and M. M. B. Kursa, “Package ‘Boruta.’” 2020.
[100] S. RColorBrewer and M. A. Liaw, “Package ‘randomForest,’” Univ. California, Berkeley Berkeley, CA, USA, 2018.
[101] T. L. Pedersen and M. Benesty, “Package ‘lime.’” 2018.
[102] N. Hasbullah, A. J. Mahajar, and M. I. Salleh, “The conceptual framework for predicting loyalty intention in the consumer cooperatives using modified theory of planned behavior,” Int. J. Bus. Soc. Sci., vol. 5, no. 11, 2014.
[103] M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock, “Behavioral modeling for churn prediction: Early indicators and accurate predictors of custom defection and loyalty,” in 2015 IEEE International Congress on Big Data, 2015, pp. 677–680.