CUSTOMER CHURN PREDICTION ANALYSIS IN A TELECOMMUNICATION COMPANY WITH MACHINE LEARNING ALGORITHMS*

Zeynep Uyar Erdem[1]†, Banu Çalış Uslu[2], Seniye Ümit Fırat[3]

Marmara University Engineering Faculty, Industrial Engineering Department, İstanbul
ORCID No: http://orcid.org/0000-0003-4626-125X
Marmara University Engineering Faculty, Industrial Engineering Department, İstanbul
ORCID No: http://orcid.org/0000-0001-8214-825X
İstanbul Gedik University Engineering Faculty, Industrial Engineering Department, İstanbul
ORCID No: http://orcid.org/000-0002-0271-5865

Keywords

Customer Churn Analysis, Telecommunication, Data-mining, Customer Relation Management

Abstract

The purpose of this study is to provide a descriptive analysis of the assessment of machine learning algorithms to an effective customer churn prediction (CCP) methodology. In the rapidly developing field of Customer Relation Management (CRM), to propose a convenient CCP methodology for retaining the customers who tend to churn, a set of data-mining analyses has been conducted to predict customer churn from a bulky dataset from customers with specific attributes in a telecommunication company by using machine learning (ML) algorithms built in an open-source data mining software, WEKA. Throughout the study, a set of experimental analyses regarding customer churn prediction are conducted by using residential, corporate, and combined datasets with the number of incidences of 195712, 32905, and 228617 respectively a private telecommunication company in Turkey. Six data mining algorithms have been evaluated to predict the customer churn status: Logistic Regression, Naive Bayes, J48, and ELM schemes such as RandomForest, Bagging and Boosting. RandomForest uses RandomTree, whereas Bagging uses J48 as a base learner. The experimental analyses are conducted with real-world datasets acquired from the company's historical database to validate some decision trees' effectiveness and ensemble ML classifiers to determine the likelihood of future churning customers based on such data mining analyses implemented for CCP. The results show that the J48 outperforms Naïve Bayes based on all datasets, and it provides very similar results as the Logistic Regression classifier scheme. Besides, since Bagging has not solved the large-sized database and J48 has given similar accurate results in the residential and complete data sets, the J48 decision tree classifier can be chosen and Bagging for customer churn prediction.

MAKİNE ÖĞRENME ALGORİTMALARI İLE BİR TELEKOMUNİKASYON ŞİRKETİNDE MÜŞTERİ KAYIP TAHMİNİ ANALİZİ

Anahtar Kelimeler

Müşteri Kayıp Analizi, Bilişim, Veri madenciliği, Müşteri İlişkileri Yönetimi

Öz

Bu araştırmanın amacı, makine öğrenimi algoritmalarının değerlendirilmesinin etkili bir müşteri kayıp tahmini (MKT) metodolojisine yönelik açıklayıcı bir analizini sağlamaktır. Hızla gelişen Müşteri İlişkileri Yönetimi (MİY) alanında, ayrılma eğiliminde olan müşterileri elde tutmak için uygun bir MKT metodolojisi önermek için, belirli müşterilerden açık kaynaklı bir veri madenciliği yazılımı olan WEKA'da oluşturulan makine öğrenimi algoritmalarını kullanarak bir telekomünikasyon şirketinden gelen anonim büyük bir veri setinden müşteri kaybını tahmin etmek için bir dizi veri madenciliği analizi yapılmıştır. Çalışma boyunca, Türkiye'deki özel bir telekomünikasyon şirketinden sırasıyla 195712, 32905 ve 228617 müşteri sayılarına sahip bireysel, kurumsal ve birleşik veri setleri kullanılarak müşteri kayıp tahminine ilişkin bir dizi deneysel analiz yapılmıştır. Müşteri kayıp durumunun tahmini için altı veri madenciliği algoritması değerlendirildi: Lojistik Regresyon, Naive Bayes, J48 ve RandomForest, Bagging ve Boosting gibi ELM şemaları. RandomForest, RandomTree'yi kullanırken, Bagging, temel öğrenme olarak J48'i kullanmaktadır. Deneysel analizler, MKT için uygulanan bu tür veri madenciliği analizlerine dayalı olarak gelecekteki müşteri kayıplarının olasılığının belirlenmesi için bazı karar ağaçlarının ve topluluk makine öğrenme sınıflandırıcılarının etkinliğini doğrulamak için şirketin tarihsel veri tabanından elde edilen reel veri kümeleri ile gerçekleştirilir. Sonuçlar, J48'in tüm veri kümelerine göre Naive Bayes'ten daha iyi performans gösterdiğini ve Lojistik Regresyon sınıflandırıcı şemasına çok benzer sonuçlar verdiğini göstermektedir. Ayrıca, Bagging büyük boyutlu veritabanını çözmediğinden ve J48, bireysel ve eksiksiz veri setlerinde benzer doğru sonuçlar verdiğinden, J48 karar ağacı sınıflandırıcısının yanı sıra müşteri kaybı tahmini için Bagging seçilebilir.

Araştırma Makalesi		Research Article	
Başvuru Tarihi	: 23.02.2021	Submission Date	: 23.02.2021
Kabul Tarihi	: 14.10.2021	Accepted Date	: 14.10.2021

* This study was produced from Zeynep Uyar Erdem's master's thesis conducted at Marmara University, Institute of Pure and Applied Sciences, under the supervision of Prof. Dr. Seniye Ümit Fırat and co-advisory of Dr. Banu Çalış Uslu.

†Sorumlu yazar; e-posta : zeynepuyar@marun.edu.tr

1. Introduction

In today's competitive business environment, considering that customers have many service provider options with many alternatives, it can be easily seen that they can change the direction of the service they received and their service provider quite quickly (Amin, Al-Obeidat, Shah, Adnan, Loo, and Anwar, 2019). Considering that customers are the most critical assets for businesses, companies' problem is how to analyze loyal or lost customers (Kelvin, Cindy, Charles, Leonardo, and Yennimar, 2020). Specifically, for telecommunication companies, a lost customer can be defined as the person who terminates all business partnerships with the company by closing all of their accounts (Karvana, Yazid, Syalim, and Mursanto, 2019).

The more number of loyal customers a company maintains by not considering short term profits, the more advocates and the better reputation they have in the marketplace, which proves that Customer Churn Prediction (CCP) to be a substantial problem, especially in the telecommunication industry (TCI) domain (Ganesh, Arnold, and Reynolds, 2000). Although the retention of an existing customer is more efficient in comparison to acquiring a new one, it is also very challenging since TCI also suffers due to too intense contention, the concentrated marketplace, adaptive conditions, and initiating new appealing offers along with recent advanced development of technology (Amin et al., 2019).

On the other hand, abundant data from TCI customers in terms of connection/disconnection, demographics, financial details, and other user attributes offer an excellent opportunity for the researchers in the fields of Machine Learning (ML) and Data Mining (DM) (Pelin Biçen, 2002) to foster CCP field with new predictive models to eventually provide aid tools to practitioners in TCI.

Thorough data analysis is required to develop an analytical method or utilize one of the currently developed methods to maintain a long-lasting relationship with the existing customers. It is significant to clarify the root cause of churned customers that may be due to one or more such reasons as dissatisfaction (Baier, Kühl, Schüritz, and Satzger, 2020), the higher cost compared to the competitor's (Yulianti and Saifudin, 2020), low service quality (Celik and Güler, 2019), lack of desirable features (Deligiannis and Argyriou, 2020), privacy concerns (Bandara, Fernando, and Akter, 2020), brand power (Al-Mashraie, Chung, and Jeon, 2020), and so on.

Therefore, the concept of customer churn analysis (Çiçek and Arslan, 2020) is essential, especially in the telecommunication and banking sectors which have a membership-based income model. These business sectors are required to sustain their stability to survive by preventing any income loss (Fırat, and Biçen, 2003). To achieve this, issues such as measuring the level of customer loyalty, determining the most important factors affecting customer churn, and predicting customer tendency should be identified through the customer churn analysis. In this context, many prediction models based on machine learning tools such as a decision tree algorithm have been developed in the literature to predict this possible future loss of income (Ahmad, Jafar, and Aljoumaa, 2019; Al-Mashraie et al., 2020; Machado, Karray, and De Sousa, 2019; Sjarif, Yusof, Wong, Ya'akob, Ibrahim, and Osman, 2019). These analyses are focused on issues such as keeping the customer loyal to the company or retaining the customers from the decision of churn.

In this paper, customer churn analysis is carried out by using authenticity. However, anonymous Big Data collected from customers of a private company in the telecommunication sector and the methodology of data mining are discussed to detect possible customers in the leaving process and prevent customer churn. Keeping the current customers not only takes less time and expenditure in comparison to gaining new customers, but this also leads the companies to avoid more significant problems with regards to their stability. An appropriate analysis of the current data can be performed through the data mining methods, and the customers can be decomposed appropriately and specified. Consequently, the results from the customer churn analysis through data mining algorithms such as decision tree can indicate when the customers are classified correctly, and appropriate diverse plans are handled for each customer type according to the customer behavior against the situation that they face with, the attributes of the incidences can be predicted and changed to good for the company.

From the Customer Relation Management perspective, to propose a helpful CCP technique for holding the customers who tend to churn, this paper covers a set of data-mining analyses that has been conducted to predict customer churn from big data in terms of services offered to the customers of a telecommunication company to find out the most suitable ML scheme built-in WEKA. For this purpose, six data mining algorithms have been evaluated for CCP analysis: Logistic Regression, Naive Bayes, J48, and ELM schemes such as RandomForest, Bagging, and Boosting. RandomForest uses RandomTree, whereas Bagging uses J48 as a base learner. The experimental datasets acquired from the historical database of the TCI Company to validate the effectiveness of some ML classifiers such as decision tree and ensemble algorithms for determination of likelihood of customers who tends to leave based on such data mining analyses. The ML schemes have been run first on testing the dataset with two options to assess their stability. Then, ZeroR scheme is used as a baseline to benchmark the accuracy performance of the other schemes. Thus, ZeroR baseline data-mining scheme is run on each dataset to have the least classification accuracy that the other data mining schemes must provide. Then Logistic Regression, Naive Bayes, and J48 decision tree schemes are run and followed by the ELM schemes such as RandomForest, Bagging, and AdaBoostM1.

Furthermore, this study offers assistance for possible future implementations and potential further research applications to support the researchers in different fields and practitioners in different sectors. It fosters the CCP field by the most efficient predictive models, which can improve the forecasting performance to provide aid tools and comparative information regarding the ML schemes to practitioners in TCI.

The following parts of the paper mainly include five sections. In section 2, the detailed literature review about decision tree, logistic regression, and ensemble learning algorithms. In Section 3, we presented the management of churning customers, introduced a variety of data mining and machine learning techniques. In Section 4, we proposed our methodology for predicting the churning customers. Analysis of applied methodology is described in Section 5, and we concluded the article with discussion in Section 6.

2. Literature Review

A growing body of studies has been recently carried out on customer churn in the telecommunication market (Li, Yang, Yang, Lu, and Lin, 2016; Preeti K. Dalvi, Siddhi, Khandge, Ashish Deomore, Aditya Bankar, 2016; Yeboah-Asiamah, Narteh, and Mahmoud, 2018). These studies are highly dependent on data mining and machine learning technologies.

Data mining (Er Kara, Oktay Fırat, and Ghadge, 2020; Özekes, 2003) is the operation of discovering critical confidential data at the big data sets (Gürsoy, 2010; Şen, Körük, Serper, and Çalış Uslu, 2019). It can both use classical statistical methods and machine learning practices. Various ML algorithms such as Multicriteria Decision Aiding (MCDA) (Es, Hamzacebi, and Firat, 2018), clustering analysis (Anuşlu and Fırat, 2019; Avni Es, Hamzacebi, and Oktay Firat, 2018), and grey cluster analyses (Karakoç, Avni Es, and Firat, 2019) as an economic indicator that classifies different groups have been continuously developed to aid decision making in many fields in the age of Industry 4.0 (Anuşlu and Fırat, 2020).

Furthermore, Many ML approaches for data processing are available in the literature, such as Naïve Bayesian (Çiçek and Arslan, 2020), Artificial Neural Network (ANN) (Dahiya and Bhatia, 2015; Gülpınar, 2013; Mitchell, 1997), Neuro-Fuzzy Classifier (Abbasimehr, 2011), Support Vector Machines (Oh and Sohn, 2009), Decision Tree (Ahmad et al., 2019; Dahiya and Bhatia, 2015; Hassouna, Tarhini, Elyas, and Abou Trab, 2015), Genetic Algorithms (Goldberg and Holland, 1988), and Ensemble models (Coussement, Lessmann, and Verstraeten, 2017; Wang, Xu, and Hussain, 2019) to forecast whether a customer is a churner in anticipated future based on his/her activities.

Additionally, Bayesian networks are currently one of Customer Analytics' primary tools (Sauro, 2015) for representing probabilistic knowledge from measuring customer attribution. General probability distributions could be concisely represented by Bayes' algorithm over a set of propositional stochastic quantities (Russell, 2013). Besides, the probabilities of hypotheses could be calculated based on Bayes' theorem. The Naive Bayes classifier algorithm that calculates explicit probabilities for hypotheses for estimating nominal values of unobserved variables is among the most practical approaches to the ML problems, such as CCP problems (Çiçek and Arslan, 2020).

On the other hand, decision trees can be considered an unstable learning scheme because where a small difference in the training data may result in a large deviation in the model. In other words, slightly changing the training data helps obtain an entirely various decision tree structure in a decision tree. However, with Naïve Bayes, small changes in the training set do not deviate from the result of Naïve Bayes or instance-based learning schemes. Therefore Naïve Bayes can be considered as a "stable" ML method. It can produce probabilities, and it can also function concerning probabilities.

The decision tree algorithm is the most effective approach that has been used as a classifier and forecasting tool in data mining studies. Decision tree methods existing in the literature are mainly ID3, C4.5, ASSISTANT, PUBLIC, CART, CN2, SLIQ, SPRINT, and so on (Anyanwu, 2011; Mitchell, 1997; Salzberg, 1994). ID3 is the most typical decision tree learning scheme which was originated in the concept learning system (CLS). It has addressed many accuracy issues improving its efficiency by avoiding issues such as over-fitting, mistraining due to missing values, and so on; hence, it was finally evoked as C4.5 (C5.0) so that it could deal with continuous attributes (Dai, Zhang, and Wu, 2016). When the tree does not classify correctly, all the given training parts of the dataset, it selects another training dataset adding to the original portion, and then it repeats this procedure until the decision set correctly avails. ID3 algorithm refines the search further to the tree below a leave node as long as the new erroneous example varies uncertainly from the other examples associated with the node, ID3 attempts until it eventually succeeds in finding a new decision attribute. Another approach in decision tree algorithms is reduced error pruning to avoid over-fitting (Pham, Prakash, Singh, Shirzadi, Shahabi, Tran, and Bui, 2019). In this approach, three nodes are iteratively pruned. The pruning approach always selects to remove the node to most increase the accuracy over the validation set, and it continues to prune such nodes until their removal results in the tree's inaccuracy over the validation set. The over-fitting issue can also be prevented by another successful method called Rule-post pruning that could find high-accuracy hypotheses. The approach first infers the decision tree from the training set, and then it grows the tree until the training subset agrees with minimal or no permutation allowing the over-fitting. Then the learned tree is adopted into an equivalent set of rules, from the root to a leaf node, it generates one rule for each path. To eventually improve the accuracy, it follows next by pruning (generalizing) each rule by simply removing any preconditions. Finally, it classifies the pruned rules by their accuracy and considers them in the same order when sorting later examples (Mitchell, 1997).

The number of open-source software has increased gradually with the development of the independent software movement since the software that has emerged could produce extremely safe results. Some of the open-source software is called SPSS Clementine (Altay, 2005), KNIME, Tanagra, RapidMiner, and WEKA, and they can contribute to the data mining solution of companies of all sizes operating in different sectors as well as in academic fields (Çınar and Silahtaroğlu, 2015). To evaluate, Dahiya and Bahatia (Dahiya and Bhatia, 2015) compared two data-mining algorithms –for churn prediction analysis – such as Decision Tree (J48 or, i.e., C4.8) and Logistic Regression, which are built-in WEKA machine learning software. They found out that the J48 Decision Tree classifier algorithm has given slightly better-précised prediction compared to the Logistic classifier. J48 pruned decision tree algorithm generally offers fairly good accuracy compared to its counterpart algorithms. For some datasets, J48 may outperform even Ensemble (committee) techniques. WEKA (Waikato Environment for Knowledge Analysis) data mining software is rich in terms of the decision tree, Bayes, and ensemble learning algorithms; nevertheless, it does not involve any ANN algorithm, whereas RapidMiner (Çelik and Başarır, 2017; Geetha and Nasira, 2014) does. Both WEKA and RapidMiner have support vector machines (SVM) and ensembles as classifiers for data-mining the big data. They both also have a user-friendly graphical user interface (GUI) for eventually conducting graphical analysis in the post-processing stage.

Decision tree algorithms have top-down recursive induction by selecting attributes for the root node, and it generates branches for each possible attribute value (Grabczewski, 2014). Then it breaks up instances of a dataset into subsets, one for each branch extending from the node. It recursively repeats this for each branch in the decision tree using only instances that reach the branch. Once all instances have the same class, the algorithm stops. The approach to how the algorithm decides which one of the attributes ends up with the purest nodes on pursuing the purity for eventually having the smallest tree is by utilizing the information theory (Yulianti and Saifudin, 2020) based on entropy, which is to be explained in more details in the following section.

Decision tree algorithms such as J48 and Decision Stump use the measure of entropy for classification (Ullah et al., 2019). A boosting ensemble algorithm in WEKA called AdaBoostM1 (adaptive boosting), by default, uses Decision Stump as its base learner. With boosting, the accuracy generally improves as the number of iterations increases and then flattens out. The idea with Bagging, several different decision structures are desired to be produced, such as using J48 to produce decision trees, and then with several different training sets of the same size, sampling the original training set can generate slightly different decision trees. The set could be sampled through replacement, which means that sometimes two of the same [instances] might be sampled in Bagging. Several different training sets could be produced, and then a model could be built for each one – such as a decision tree – employing the same or some other ML schemes. Then voting combines the predictions made by different models, or in the regression situation, the numeric result could be averaged rather than voting on it.

A combination of such algorithms as ANN, DT, Bayes, etc., has also been developed called ensemble models that combine multiple algorithms including decision tree regression and classification methods as well as any other models mentioned previously to gain better predictive performance. Ensemble learning methods (ELM) have received remarkable attention (Karakurt, Erdal, Namlı, Yumurtacı-Aydoğmuş, and Türkkan, 2013; Wang et al., 2019), and the performance of the regression and classification problems have had considerable advancement by using ELM techniques in recent years. Techniques based on regression are mostly associated with good results in CCP. The ensemble learning methods can be classified into four groups: Bagging, randomization, boosting, and stacking; each works with different approaches. Running different accompanying ML algorithms, all producing classifiers for the same problem, and letting them vote so that an unknown test instance is classified, can often improve the predictive performance. In general, ELM forecasting models such as bagged (bootstrap aggregated) regression trees (BRT) and gradient boosted regression/decision trees (GBDT) as well as bagged artificial neural network (BANN) and gradient boosted artificial neural network (GBANN) are among the most popular ELM techniques proposed to reduce the prediction error of ML algorithms (Karakurt et al., 2013; Zhang, Zhang, and Wang, 2008).

There are various algorithms, tools, and techniques for predicting and management customer churn in the literature and practice. The objective of this research is to develop a comparative analysis to customer churn prediction model for a telecommunication provider by using a few data mining classification techniques. The data set containing the information of more than ten years customers were used to measure the model performances by using these techniques such as ZeroR, Logistic Regression, Naïve-Bayes, Random Forest, Bagging, and Adaboost. The main contribution to the literature is to compare different algorithms and to reveal the effective one for data mining classification techniques. Best of our knowledge, there is no study covers these types of comparison with real-life case study from the telecommunication sector.

3. Churn Management and Related Data Mining Techniques

In this section, customer loss management is discussed and the classifier algorithms used herein in predicting customer churn are discussed in more detail, including their mathematical bases, origins, and features in the built-in schemes of the WEKA data mining open-source software package.

3.1 Churn Management

If a customer ends the contract with a company and begins a new one with another company, it's a churned customer (Richeldi and Perrucci, 2002). Customer churn is closely related to the loyalty of the customer. In nowadays economy, the only way is not to reduce the prices to gain customer loyalty. Accordingly, increase loyalty through the addition of new assets to the products has been a norm in various industries (Rud, 2001).

The aim of the companies to keep the customers in the company is to determine customers with leaving thoughts and cost analysis to keep them in the company. The most important point when making cost analysis is to define the customers with leaving thoughts.

To avoid customer churn, the data should be analyzed. This is the most effective way to ensure profitability. In this section, the methods by which data can be examined will be explained.

3.2 Decision Tree

Decision tree algorithms use this theory to choose the best attribute by checking the information gain: knowing the value of an attribute, the information amount in bits is taken into account by subtracting the distribution entropy after the split from that before the split (Witten, 2020).

The decision tree algorithm is strongly recommended for classification and prediction tasks due to having several advantages (Song and Ying, 2015):

- The relationship between the attributes and the target variable is simplified.
- Interpretation of the algorithm is easy.
- It is capable to deal with missing values and biased data.
- It cannot be affected by outliers.

J48 is a decision tree algorithm that carries out top-down induction based on information theory, creating a recursive divide-and-conquer strategy by selecting and splitting an attribute at its root node, and the best attribute for this is chosen by information gain. Then this leads to the generation of a branch corresponding to each possible attribute value. This procedure eventually divides the instances into subsets, one corresponding to each branch extending from the root node. Recursively this procedure repeats concerning each branch, selecting an attribute at each node, which uses only instances reaching the branch. This recursive event carries on until all instances have the same class then it stops.

Once J48 runs, it also provides probabilities in the result panel on WEKA – the negative and positive probabilities, respectively, where the errors can also be seen. These are all different probabilities that are internally used by J48 to perform the pruning operations, which was the same approach as discussed in ID3's pruning.

3.3 Logistic Regression

Logistic regression is a linear classification model which is used to describe the relationship between one or more independent variables and a dependent variable that exhibits a binary structure. In a logistic regression model, a logistic function is used to compute the probabilities of the possible discrete outcomes of given input variables. It computes the probability that an instance belongs to a class or not. If the predicted probability that a sample belongs to a particular class is greater than 50%, that sample is classified as that particular class.

Respectively, a linear function and a threshold are calculated by the linear regression method, which has a linear sum. For probabilities prediction, Logistic regression performing the logit transform is a popular and influential ML method that works internally with probabilities as Naïve Bayes does.

Logistic regression is beneficial when the dependent target variable is binary and the independent definitive variables are continuous (Banks, and Said, 2006). Thus, it becomes a candidate model that can be used for customer churn estimation.

3.4 Ensemble Learning

Ensemble models combine multiple algorithms in a machine learning task to compensate errors of a single model and thus the overall model performance increases (Sagi and Rokach, 2018).

Several reasons for the increase in the prediction performance are as follows:

- The combination of different models decreases the overfitting.
- Better data representation and fitting while working nonlinear data sets.
- Mitigation of the class imbalances.
- Increase in computational performance.

Mainly in the Boosting (AdaBoost) categories, Bagging, Randomizing (Random Forest, Random Tree), Stacking, many ELM algorithms exist in the open-source data-mining software.

In Bagging: The training set is sampled to have another set at the same size by choosing the Bag size of 100%, but in each time sampled "with replacement" which means that different sets of the same size, but each set may include the duplicate of the original training. This bagging event happens to build a model concerning each one – using the same ML scheme, then predictions can be combined by voting on classifiers for which one to bag. In the bagging approach, the settings of the random-number seed and the number of bagging iterations can also be set.

The randomizing approach: The training data is not randomized in the "RandomForests" algorithm.

Depending on what the algorithm is, the way how the algorithm is randomized. Random forests – using decision tree algorithms that select the best attribute for splitting each time, for example, J48. This randomizing procedure selects a few of the best options rather than selecting the very best and then randomly picking among the classifier algorithms that give different trees every time. Generally, to gain better performance, the decision trees are bagged and then randomized, and eventually, the result is bagged.

Boosting is iterative: The performance of previously built models influences new models by basically creating a model, and then the instances that are misclassified by that model are taken into account. These are the hard instances to classify; extra weight is put on those instances that get wrong to make a training set for producing the next model in the iteration. The new model is encouraged to become an "expert" for misclassified instances nominated by all the earlier models. They are eventually combined by voting along with weighting models concerning their performance.

The stacking ensemble learning method: base learners are not combined by voting herein, but through a meta-learner that is another learner scheme uniting the base learners' output. The base learners are called level-0 models, whereas the meta-learner is a level-1 model. The classifying predictions performed by the base learners become the input to the meta-learner. Typically, different ML schemes are used as the base learners to gain various experts at different things. Special care is required to be paid in the way data is generated to train the level-1 model, which may involve somewhat cross-validation.

4. Methodology

This study adhered to the research and publication ethics.

In this study, we declare that we comply with scientific and ethical principles. Customer churn analysis is carried out by using anonymous in this paper. However, anonymous data is collected from customers of a private company in the telecommunication sector, and the methodology of data mining is discussed to detect possible churning customers and retain customer churn. For this purpose, three sets of data are analyzed, dividing the primary dataset into two: the residential members and corporate members. The third data set is a combination of both groups. In all datasets, we aim at predicting the same nominal classes that are churn status, whether active or passive, along with the other seven attributes the same in all. However, in the combined dataset, one more nominal attribute is added concerning the customer type: residential and corporate, as shown in Table 1.

Table 1
Attributes and Incidences of Three Datasets

Attributes	Residential	Corporate	Combination
Total failure	+	+	+
Unrepaired	+	+	+
Membership day	+	+	+
Failure Count	+	+	+
Average Repair Time	+	+	+
Customer Lifetime	+	+	+
Customer Type			+
Churn Status	+	+	+
Number of incidences	195712	32905	228617

4.1 Data mining software, WEKA

On the test options panel of WEKA, there are four test schemes such as 1. Use training set, 2. Supplied Test Set, 3. Cross-Validation, and 4. Percentage Split. The first test option uses the training dataset for testing it, which sometimes gives the best prediction precision percentage; however, it is not much recommended by the experts. For the second option, a dataset for testing needs to be uploaded into the software since we have not prepared or been supplied with a test set. In running most of the classifiers, we have not chosen this testing option. Instead, the other three options are generally chosen for all the datasets and the best results obtained are presented in the following section. In cross-validation, the default value is ten folds, which means the dataset is divided into ten pieces and one piece is used for testing, and the rest is used for training, and the model runs, and this repeats ten times each time the test piece shifts to the next one. This causes us to take a longer time to obtain the classification results compared to others; however, this testing option is more robust and generally gives the best prediction percentage. In percentage split, the default value is 66%, which means two-third of the dataset is used for training, and one-third of it is for testing as the model runs only once. We have selected a 90% percentage split in this option for our data mining procedure. We have not experienced any overfitting issue in any data mining schemes because we have not seen much difference among the results from each test option.

4.2 Datasets

Table **2** presents an example of the data anonymously recorded by the telecommunication company based on customer membership length, the number of failures, and their response length to repair in their internet connection service. This dataset is required to be determined to comprehend the probability of customer churn based on failure in the service.

Table 2
Sample of a Dataset Used for Customer Churn Prediction Analysis

Total failure no	Unrepaired	Membership day	Unit time failure no	Ave repair hrs	Customer Lifetime	Customer type	Churn Status
0	0	1	0	(1,000)	1-10 days	Residential	PASSIVE
0	0	1	0	(1,000)	1-10 days	Corporate	PASSIVE
0	0	1	0	(1,000)	1-10 days	Corporate	PASSIVE

The numbers of active and passive residential and corporate customers against the customer lifetime are presented in Table 3 and Table 4, respectively.

Even though customer lifetime of 3-5 years and 5-10 years have a good enough percentage of the grand totals, 10+ years-long customer lifetime is in the band of 2.5-3 % of grand totals of both residential and corporate customers. The loyalty of the customers' needs to be increased by increasing their satisfaction with the service through avoiding or at least minimizing the matters of disturbance

Table 3
Numbers of the Active and Passive Residential Customer Concerning the Membership Time Interval

Customer Lifetime	Active	Passive	Grand Total
10 + years	1,913	3,458	5,371
1-10 days	748	735	1,483
11-20 days	922	608	1,530
1-3 years	48,385	21,774	70,159
1-6 months	8,689	9,968	18,657
21-30 days	666	2,177	2,843
3-5 years	45,413	9,886	55,299
5-10 years	21,561	3,631	25,192
6-12 months	9,495	5,683	15,178
Grand Total	137,792	57,920	195,712

Table 4
Numbers of the Active and Passive Corporate Customer Concerning the Membership Time Interval

Customer Lifetime	Active	Passive	Grand Total
10 + years	401	654	1,055
1-10 days	108	217	325
11-20 days	172	182	354
1-3 years	5,839	4,249	10,088
1-6 months	1,535	3,113	4,648
21-30 days	101	415	516
3-5 years	3,258	3,567	6,825
5-10 years	3,711	2,556	6,267
6-12 months	1,567	1,260	2,827
Grand Total	16,692	16,213	32,905

4.3 Metrics

We use different metrics while evaluating the performance of our models. These are Accuracy and Confusion Matrix for classification and Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE) for regression models.

Accuracy is the ratio of the correctly classified instances to the total instance count. It is the most preferred metric while measuring classifier performance. It can be used when the distribution of the target dataset is balanced.

The confusion matrix is one of the easiest tools used to measure the performance of classification algorithms. It clearly shows correct and incorrect predictions. We cannot call the confusion matrix a measurement by itself, but many measurement methods emerge from this matrix. It is very useful when the target dataset is imbalanced.

Mean Absolute Error (MAE) calculates the average absolute distance between predicted and true values of the target variable. Outliers cannot easily affect MAE unlike in MSE, therefore it is preferred if our dataset has many unresolved outliers. If the unit of the model errors differs, this metric cannot be used.

Root Mean Squared Error (RMSE) is a square root of averaged squared error between predicted and actual values of the target variable. Since MSE is prone to outliers, square rooted version, RMSE is preferred over MSE. If the unit of the model errors differs, this metric cannot be used.

Relative Absolute Error (RAE), is a division of two calculations where absolute error between predicted and actual values is the numerator and absolute error between actual values and average of actual values is the denominator. RAE is used for the comparison of models when the model error units are different and it is also robust to outliers.

Root Relative Squared Error (RRSE) is the square root of a division of two calculations where squared error between predicted and actual values is the numerator and squared error between actual values and the average of actual values is the denominator. RRSE is used for the comparison of models when the model error units are different.

5. Analysis and Findings

In this section, data mining analysis for churn prediction is conducted by using the experimentally obtained datasets from customers of a telecommunication company. ZeroR scheme is used as a baseline to benchmark the accuracy of the other schemes used in the CCP since their accuracy must be higher than the ZeroR. Thus, ZeroR baseline data mining scheme is run on each dataset to have the least classification percentage to benchmark what accuracy the other data mining schemes are supposed to provide. Then Logistic Regression, Naive Bayes, and J48 decision tree schemes are run and followed by the ELM schemes such as Random Forest, Bagging, and AdaBoostM1.

Since the Bagging scheme requires a large heap size in running Java in the background and when sometimes the memory size may not be enough, these schemes are not compatible to solve the models for large-sized datasets, e.g., in this particular case, the residential and combined dataset with around 200,000 instances. As a solution to this issue, the instances in the datasets of residentials and combined are shuffled by using the unsupervised incidence filter named 'Randomize' with a Random seed number of 42 (arbitrarily chosen by default) and then truncated from its 50% and 60% respectively through another unsupervised incidence filter called 'RemovePercentage' so that the datasets became prepared to be handled by the computationally expensive schemes such as Bagging.

As shown in the results presented in Table 5.

J48 outperforms Naïve Bayes based on all datasets. It provides very similar results as the Logistic Regression classifier scheme in terms of both the accuracy and classifying the instances into the nominal attribute of Churn Status. The accuracy is determined through the confusion matrix. The classified instances indicate how many of the customers are classified to be in inactive and active status. Indeed, Naïve Bayes seems to be worse than even ZeroR in terms of accuracy in two datasets.

One can conclude that J48 is the right choice for predicting the customer churn for these datasets; however, further investigation into its stability is required.

Table 5
Comparison of Accuracy and Number of Classified Instances From Naïve Bayes and J48 Decision Tree Data Mining Algorithms Based on Three Datasets

Datasets	Result variable	ZeroR		Logistic Regression		Naïve-Bayes		J48	
		Passive	Active	Passive	Active	Passive	Active	Passive	Active
Corporate	Accuracy %	51		61		61		67	
	Class. Instant.	0	16692	9045	11062	15933	3976	11466	10454
Residential	Accuracy %	70.4		73		65.4		75	
	Class. Instant.	0	68864	8002	63757	7156	14598	18228	128150
Combination	Accuracy %	68		71		67		74	
	Class. Instant.	0	61866	10321	54840	18813	42422	29439	138935

As presented in Table 6, AdaBoost using the Decision Stump decision tree as a classifier based on the weight threshold of 100 has provided the lowest accuracy for all the datasets. Therefore, other schemes should be further investigated to evaluate the customer churn choosing the best data mining scheme among the ELM classifiers used in this study.

Table 6
Comparison of Accuracy and Number of Classified Instances from Elm Data Mining Algorithms Random Forest, Bagging, and Adaboost Based on Three Datasets

Datasets	Result variable	Random Forest		Bagging		AdaBoost	
		Passive	Active	Passive	Active	Passive	Active
Corporate	Accuracy %	75		73		61	
	Classified Instances	12527	12178	12544	11375	15801	4289
Residential	Accuracy %	77		75		73	
	Classified Instances	11808	63423	8553	64518	4993	66405
Combination	Accuracy %	77		73.5		71	
	Classified Instances	14778	55714	10840	56340	7166	57607

Random Forest datamining scheme classifies the dataset for constructing a forest of random trees bagging with 100 iterations using Random Tree as the base-learner scheme.

Table 7 and 8 regarding the Random Forest classifier scheme are presented to show its stability. The result summaries and the confusion matrices are obtained from the 'Corporate dataset' by running the Random Forest scheme based on testing the dataset choosing '10-fold cross-validation' and 'use training set' options. Since the accuracy from the testing options has given a large discrepancy from each other, the scheme may not be considered stable for this dataset even though evaluation with 'on training set' test mode has provided relatively somewhat high accuracy compared to the other ELM classifiers.

By varying the test mode, Random Forest was attempted on the residential (relatively larger) data set. One of the results from the evaluation of the training set has given the accuracy of 77%. Another result in terms of the accuracy of 74% was obtained from the evaluation on cross-validation, which is considerably different results. In conclusion, The Random Forest data mining scheme may not be the best option among other ELM algorithms for predicting customer churn since it seems unstable. However, in the same test, Bagging provides much better stabilization by giving almost the same results of accuracy as the test option is varied.

Table 8 presents the confusion matrices that show the numbers of correctly and incorrectly classified attributes from the Random Forest scheme testing based on both options. Even though passive and active customer numbers seem different due to the different accuracy and misclassified customers, the ratio between the passive and active customer numbers are very similar: 10983/10877=1.01 and 12527/12178=1.03.

Table 7
Comparison of the Result Summaries From Random Forest Scheme Dataset Based on Selection for Testing

	Random Forest			
	Use Training Set		10 Fold Cross-Validation	
Correctly Classified Instances	24705	75.08%	21860	66.43%
Incorrectly Classified Instances	8200	24.92%	11045	33.57%
Kappa statistic	0.5019		0.3289	
Mean absolute error	0.3238		0.3774	
Root mean squared error	0.3991		0.4643	
Relative absolute error	64.77%		75.49%	
Root relative squared error	79.84%		92.86%	

Also, similarly, the ratio of the passive and active customer numbers concerning the correctly classified instances (Table 8) are very similar: 10983/21860=0.502 and 12527/24705=0.507. The Bagging scheme has been chosen as J48 as a base learner. Bagging scheme with 73% accuracy has also classified the corporate customers into active and passive classes with similar manner at close accuracy to the Random Forest scheme with 75%: 12544 / 11375 = 1.102 and 12544 / (11375 + 12544) = 0.524. The customer churn rate is around the band of 50% in predictions. In residential customers, this ratio is much lower in terms of classified instances by the Bagging scheme, which means the number of incidences in the churn status of passive is much lower than active. The ratio given in Table 7 is simply 8553/64518=0.13.

Table 8
Confusion Matrices For Random Forest Classifier Testing Based On Cross-Validation and Use Training Set

Random Forest Confusion Matrix				
Cross-Validation		Use Traning Set		
a	b	a	b	classified as
10983	5230	12527	3686	a = PASSIVE
5815	10877	4514	12178	b = ACTIVE

6. Discussions and Conclusions

In this study, CCP analysis is carried out by using a dataset collected from customers of a private company in the telecommunication sector. The data mining methodologies that can be utilized to detect and prevent possible customers who tend to churn are discussed. By using the experimental data from a private telecommunication company in Turkey, a set of data-mining analyses classifying a nominal attribute with regards to CCP has been conducted. Six data mining algorithms have been evaluated to predict the customer churn status: Logistic Regression, Naive Bayes, J48, and ELM schemes such as RandomForest, Bagging, and Boosting. RandomForest has used RandomTree, whereas the Bagging J48 is a base learner. All algorithms have been run on testing with two options to assess their stability. ZeroR scheme is used as a baseline to benchmark the accuracy performance of the other schemes. Thus, the ZeroR baseline data mining scheme is run on each dataset to have the least classification accuracy the other data mining schemes are supposed to provide. Then Logistic Regression, Naive Bayes, and J48 decision tree schemes are run and followed by the ELM schemes such as RandomForest, Bagging, and AdaBoostM1.

RandomForest seemed to be unstable as providing considerably different results in terms of classes and prediction accuracy as the decision tree algorithms can suffer from instability with some datasets; therefore, they are required to be evaluated with this concern. The other ELM scheme named Bagging using J48 as a base learner has provided somewhat good accuracy with proper stability as compared to others as well as J48 itself, especially in classifying the nominal attribute of churn status of the corporate dataset. However, the Bagging scheme is computationally much more expensive, requiring much higher performance and much longer time than J48. Since Bagging has not solved the large-sized database, it was edited by shuffling and truncating via relevant unsupervised incidence filters. Therefore, since J48 has given similar accurate results in the residential and complete (the combination of residential and corporate) data sets, depending on the capabilities in terms of computer performance and time, J48 decision tree classifier can be chosen as well as Bagging for customer churn prediction of such case studies.

The originality of this article is that it demonstrates the effectiveness and efficiency of data mining classification algorithms based on real data. Findings and results show that J48 performs better than other tested algorithms such as ZeroR, Logistic Regression, Naïve-Bayes, Random Forest, Bagging, and Adaboost in predicting churn analysis. The performance of the churn prediction model is evaluated by using validation metrics such as Accuracy, Root Relative Squared Error (RRSE), Relative Absolute Error (RAE), the confusion matrix, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). In addition to the academic contribution of research, it also provides a practical model to analyze customer churn for TCI.

As a future study, algorithms of ELM with regards to ANN such as a bagged artificial neural network (BANN) and gradient boosted artificial neural network (GBANN) that are embedded in RapidMiner can be suggested to attempt as a classifier for CCP analyses in comparison to Decision Tree and other ELM forecasting algorithms such as bagged (bootstrap aggregated) regression trees (BRT), and gradient boosted regression/decision trees (GBDT).

The contribution of researchers

In this study, all authors simultaneously realized the determination of the subject, research design, definition of research methodology, data acquisition, data cleaning, findings, interpretation of results, and article writing. Zeynep Uyar Erdem

literature review and implementation the case in weka software; Banu Çalış Uslu reporting and control; Seniye Ümit Fırat identifying algorithms and final control have contributed to the steps.

Conflict of interest

Conflict of interest was not declared by authors.

References

Abbasimehr, H. (2011). A Neuro-Fuzzy Classifier for Customer Churn Prediction. *International Journal of Computer Applications*, *19*(08), 35–41. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.206.4022&rep=rep1&type=pdf

Ahmad, A. K., Jafar, A., and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1). Doi: https://doi.org/10.1186/s40537-019-0191-6

Al-Mashraie, M., Chung, S. H., and Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers and Industrial Engineering*, *144*, 106476. Doi: https://doi.org/10.1016/j.cie.2020.106476

Altay, T. (2005). *Knowledge Discovery in Databases and Data Mining Techniques: An Applied Study*. Marmara University Institute of Science. Retrieved from: https://www.proquest.com/openview/a3d2536939ffb88eb9aa2caa391be13a/1?pq-origsite=gscholar&cbl=2026366&diss=y

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., and Anwar, S. (2019). Customer churn prediction in the telecommunication industry using data certainty. *Journal of Business Research*, *94* (March), 290–301. https://doi.org/10.1016/j.jbusres.2018.03.003

Anuşlu, M. D., and Fırat, S. Ü. (2019). Clustering analysis application on Industry 4.0-driven global indexes. *Procedia Computer Science*, *158*, 688–695. Doi: https://doi.org/10.1016/j.procs.2019.09.037

Anuşlu, M. D., and Fırat, S. Ü. (2020). Ülkelerin Endüstri 4.0 Seviyesinin Sürdürülebilir Kalkınma Düzeylerine Etkisinin Analizi. *Endüstri Mühendisliği*, *1*(0), 44–58. Retrieved from: https://dergipark.org.tr/en/pub/endustrimuhendisligi/issue/52861/651168

Anyanwu, M. N. (2011). Comparative Analysis of Serial Decision Tree Classification Algorithms. *Journal of Computer Science*, *3*(3), 230–240. Retrieved from: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1040.5524&rep=rep1&type=pdf

Avni Es, H., Hamzacebi, C., and Oktay Firat, S. U. (2018). Assessing the logistics activities aspect of economic and social development. *International Journal of Logistics Systems and Management*, *29*(1), 1–16. Retrieved from: https://www.inderscienceonline.com/doi/abs/10.1504/IJLSM.2018.088577

Baier, L., Kühl, N., Schüritz, R., and Satzger, G. (2020). Will the customers be happy? Identifying unsatisfied customers from service encounter data. *Journal of Service Management*. Doi: https://doi.org/10.1108/JOSM-06-2019-0173

Bandara, R., Fernando, M., and Akter, S. (2020). Explicating the privacy paradox: A qualitative inquiry of online shopping consumers. *Journal of Retailing and Consumer Services*, *52*(May), 101947. Doi: https://doi.org/10.1016/j.jretconser.2019.101947

Banks, D. L., & Said, Y. H. (2006). Data mining in electronic commerce. *Statistical Science*, *21*(2), 234-246. Doi: https://doi.org/10.1214/088342306000000204

Biçen, P., and Fırat, S. Ü. (2003). Knowledge Discovery in Databases KDD and Data Mining An Application of Customer Segmentation Analysis in Banking Sector. *International Statistical Institute 54 Th Session*. Retrieved from: https://www.isi-web.org/isi.cbs.nl/iamamember/cd3/abstracts/papers/3275.pdf

Biçen, Pelin. (2002). *Veri madenciliği: Sınıflandırma ve tahmin yöntemlerini kullanarak bir uygulama / Data mining: Application by using predictive and classification modelling*. Yıldız Teknik Üniversitesi / Sosyal Bilimler Enstitüsü. Retrieved from: http://dspace.yildiz.edu.tr/xmlui/handle/1/6868

Celik, H., and Güler, M. (2019). *The Importance of Customer Loyalty With Corporate Governance in the Telecommunication Sector*. 50–73. Doi: https://doi.org/10.4018/978-1-5225-9265-5.ch003

Çelik, U., and Başarır, C. (2017). The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner. *Alphanumeric Journal*, *5*(1), 45–45. Doi: https://doi.org/10.17093/alphanumeric.290381

Çiçek, A., and Arslan, Y. (2020). Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması. *İleri Mühendislik Çalışmaları ve Teknolojileri Dergisi*, *1*(1), 13–19. Retrieved from: https://dergipark.org.tr/en/pub/imctd/issue/56439/765347

Çınar, A. and Silahtaroğlu, G. (2015). Veri Madenciliği Teknikleri ile Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi. Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi , 33 (2) , 309-330 . Retrieved from https://dergipark.org.tr/tr/pub/muiibd/issue/499/4516

Coussement, K., Lessmann, S., and Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. In *Decision Support Systems* (Vol. 95). Elsevier BV Doi: https://doi.org/10.1016/j.dss.2016.11.007

Dahiya, K., and Bhatia, S. (2015). Customer churn analysis in the telecom industry. *2015 4th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2015*, 1–6. Doi: https://doi.org/10.1109/ICRITO.2015.7359318

Dai, Q., Zhang, C., and Wu, H. (2016). Research of Decision Tree Classification Algorithm in Data Mining. *International Journal of Database Theory and Application*, *9*(5), 1–8. Doi: https://doi.org/10.14257/ijdta.2016.9.5.01

Deligiannis, A., and Argyriou, C. (2020). Designing a Real-Time Data-Driven Customer Churn Risk Indicator for Subscription Commerce. *International Journal of Information Engineering and Electronic Business*, *12*(4), 1–14. Doi: https://doi.org/10.5815/ijieeb.2020.04.01

Er Kara, M., Oktay Fırat, S. Ü., and Ghadge, A. (2020). A data mining-based framework for supply chain risk management. *Computers and Industrial Engineering*, *139* (December 2018). https://doi.org/10.1016/j.cie.2018.12.017

Es, H. A., Hamzacebi, C., & Firat, S. U. O. (2018). GRA-TRI: *A Multicriteria Decision Aid Classification Method based on Grey Relational Analysis.* Journal of Grey. Retrieved from: System, 30(3). https://www.researchgate.net/profile/Seniye-Firat/publication/327233146_GRA-TRI_A_multicriteria_decision_aid_classification_method_based_on_grey_relational_analysis/links/5e7314c3a6fdcc37caf64567/GRA-TRI-A-multicriteria-decision-aid-classification-method-based-on-grey-relational-analysis.pdf

Fırat, S. Ü., and Biçen, P. (2003). Veri Madenciliği Tekniklerini Kullanarak Banka Müşterileri Bölümlendirmesi ve Kredi Skorlama Modeli. *Türkiye İstatistik Kurumu İstatistik Araştırma Dergisi*, *2*(2), 135–150. Retrieved from: https://journal.tuik.gov.tr/archive/05_2003_02_02_agustos.pdf

Ganesh, J., Arnold, M. J., and Reynolds, K. E. (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *Journal of Marketing*, *64*(3), 65–87. Doi: https://doi.org/10.1509/jmkg.64.3.65.18028

Geetha, A., and Nasira, G. (2014). Artificial neural networks' application in weather forecasting – Using RapidMiner. *International Journal of Computational Intelligence and Informatics*, *4*(3), 177–182. Erişim Adresi: http://www.periyaruniversity.ac.in/ijcii/issue/Vol4No3December2014/IJCII 4-1-152.pdf

Goldberg, D. E., and Holland, J. H. (1988). Genetic Algorithms and Machine Learning. *Machine Learning*, *3*(4), 95–99. Retrieved from: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/46947/10994_2005_Article_422926.pdf

Grabczewski, K. (2014). *Meta-learning in decision tree induction* (Vol. 1). Cham: Springer International Publishing. Retrieved from: https://link.springer.com/book/10.1007%2F978-3-319-00960-5#about

Gülpınar, V. (2013). Yapay Sinir Ağlari Ve Sosyal Ağ Analizi Yardimi İle Türk Telekomünikasyon Piyasasinda Müşteri Kaybı Analizi. *M U Iktisadi ve Idari Bilimler Dergisi*, *34*(1), 331–350. Retrieved from: https://dergipark.org.tr/en/pub/muiibd/issue/497/4465

Gürsoy, U. T. Ş. (2010). Customer churn analysis in telecommunication sector. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, *39*(1), 35–49. Retrieved from: https://dergipark.org.tr/en/pub/iuisletme/issue/9249/115711

Hassouna, M., Tarhini, A., Elyas, T., and Abou Trab, M. S. (2015). Customer Churn in Mobile Markets: A Comparison of Techniques. *International Business Research*, *8*(6), 224–237. Doi: https://doi.org/10.5539/ibr.v8n6p224

Karakoç, Ö., Avni Es, H., and Firat, S. Ü. (2019). Evaluation of the development level of provinces by grey cluster analysis. *Procedia Computer Science*, *158*, 135–144. Doi: https://doi.org/10.1016/j.procs.2019.09.036

Karakurt, O., Erdal, H. I., Namlı, E., Yumurtacı-Aydoğmuş, H., and Türkkan, Y. S. (2013). Comparing Ensembles Of Decision Trees And Neural Networks For One-day-ahead Stream Flow Predict. *Scientific Research Journal (SCIRJ), Volume I, Issue IV, November 2013*, *1*(4), 43–55. Retrieved from : http://scirj.org/rp/files/original/7d69762103d3029c9d708025 6b0676ec.pdf

Karvana, K. G. M., Yazid, S., Syalim, A., and Mursanto, P. (2019). Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry. *2019 International Workshop on Big Data and Information Security, IWBIS 2019*, 33–38. Doi: https://doi.org/10.1109/IWBIS.2019.8935884

Kelvin, K., Cindy, C., Charles, C., Leonardo, D. P., and Yennimar, Y. (2020). Customer Churn's Analysis In Telecommunications Company Using Fp-Growth Algorithm: Customer Churn's Analysis In Telecommunications Company Using Fp-Growth Algorithm. *Jurnal Mantik*, *4*(2), 1285–1290. Retrieved from: http://iocscience.org/ejournal/index.php/mantik/article/view/933

Li, H., Yang, D., Yang, L., Lu, Y., and Lin, X. (2016). Supervised massive data analysis for telecommunication customer churn prediction. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, 163–169. Doi: https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.35

Machado, M. R., Karray, S., and De Sousa, I. T. (2019). LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *14th International Conference on Computer Science and Education, ICCSE 2019*, *ICCSE*, 1111–1116. Doi: https://doi.org/10.1109/ICCSE.2019.8845529

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math. Retrieved from: https://uvaml1.github.io/2020/slides/1.2_WhatIsMachineLearning.pdf

Oh, C. K., and Sohn, H. (2009). Damage diagnosis under environmental and operational variations using unsupervised support vector machine. *Journal of Sound and Vibration*, *325*(1–2), 224–239. Doi: https://doi.org/10.1016/j.jsv.2009.03.014

Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, *2*(3), 65–82. Retrieved from: http://www.ticaret.edu.tr/uploads/yayin/dergi/d3/M00041.pdf

Pham, B. T., Prakash, I., Singh, S. K., Shirzadi, A., Shahabi, H., Tran, T. T. T., and Bui, D. T. (2019). Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena*, *175*(December 2018), 203–218. Retrieved from: https://www.sciencedirect.com/science/article/abs/pii/S0341816218305538

Preeti K. Dalvi, Siddhi K. Khandge, Ashish Deomore, Aditya Bankar, V. A. K. (2016). Analysis of Customer Churn Prediction in Telecom Industry using Decision Trees and Logistic Regression. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, *88*(5), 613–618. Doi: https://doi.org/10.1109/CDAN.2016.7570883

Richeldi, M. and Perrucci , A., (2002). Churn Analysis Case Study, Telecom Italia Lab. Torino. 3-6. Retrieved from: https://www-ai.cs.tu-dortmund.de/PublicPublicationFiles/richeldi_perrucci_2002b.pdf

Rud, O.P. , (2001). *Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management*, John Wiley and Sons, Inc. Canada. Retrieved from: https://books.google.com.tr/books?hl=tr&lr=&id=HYveCsbG3ZoC&oi=fnd&pg=PT8&dq=Rud,+O.P.+,+(2001).+Data+Mining+Cookbook:+Modeling+Data+for+Marketing,+Risk+and+Customer+Relationship+Management,+John+Wiley+and+Sons,+Inc.+Canada&ots=lGQRawr2M5&sig=XTBpWBt4ngbyt-jDXRZy4aarIv8&redir_esc=y#v=onepage&q&f=false

Russell, S. (2013). Chapter 4 : Machine Learning. In *Handbook of Perception and Cognition* (Vol. 14, Issue 510). Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.1192&rep=rep1&type=pdf

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1249. Doi: https://doi.org/10.1002/widm.1249

Salzberg, S. L. (1994). Book Review: C4 . 5: Programs for Machine Learning by J. Ross Quinlan. *Machine Learning*, *240*, 235–240. Retrieved from: http://server3.eca.ir/isi/forum/Programs%20for%20Machine%20Learning.pdf

Sauro, J. (2015). *Customer Analytics for Dummies* (2015th ed.). John Wiley and Sons, Inc. Retrieved from: https://books.google.com.tr/books?hl=tr&lr=&id=l-U1BgAAQBAJ&oi=fnd&pg=PA3&dq=Sauro,+J.+(2015).+Customer+Analytics+for+Dummies+(2015th+ed.).+John+Wiley+and+Sons,+Inc.&ots=XL1ktvDX-a&sig=n0JTUG7C8V0LB6p00kpBO4ZJJ6c&redir_esc=y#v=onepage&q&f=false

Sjarif, N. N. A., Yusof, M. R. M., Wong, D. H. Ten, Ya'akob, S., Ibrahim, R., and Osman, M. Z. (2019). A customer Churn prediction using Pearson correlation function and K nearest neighbor algorithm for the telecommunication industry. *International Journal of Advances in Soft Computing and Its Applications*, *11*(2), 46–59. Retrieved from: http://188.247.81.52/PapersUploaded/2019.2.4.pdf

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/

Şen, E., Körük, E., Serper, N., & Çalış Uslu, B. (2019). Big Data Analytics and Simulation for Better Strategic Management. *Journal of Current Research on Engineering, Science and Technology*, *5*(2), 1-12. Retrieved from: http://www.jocrest.com/Makaleler/56307035_5-2-1-sen-vd..pdf

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., and Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*, 60134–60149. Doi: https://doi.org/10.1109/ACCESS.2019.2914999

Wang, Q. F., Xu, M., and Hussain, A. (2019). Large-scale Ensemble Model for Customer Churn Prediction in Search Ads. *Cognitive Computation*, *11*(2), 262–270. https://doi.org/10.1007/s12559-018-9608-3

Witten, I. H. (2013). Data mining with weka. Department of Computer Science University of Waikato New Zealand. Retrieved from: https://cs.famaf.unc.edu.ar/~laura/Weka work shop/Slides.pdf

Yeboah-Asiamah, E., Narteh, B., and Mahmoud, M. A. (2018). Preventing Customer Churn in the Mobile Telecommunication Industry: Is Mobile Money Usage the Missing Link? *Journal of African Business*, *19*(2), 174–194. Doi: https://doi.org/10.1080/15228916.2018.1440462

Yulianti, Y., and Saifudin, A. (2020). Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes. *IOP Conference Series: Materials Science and Engineering*, *879*(1). Doi: https://doi.org/10.1088/1757-899X/879/1/012090

Zhang, C. X., Zhang, J. S., and Wang, G. W. (2008). An empirical study of using Rotation Forest to improve regressors. *Applied Mathematics and Computation*, *195*(2), 618–629. Doi: https://doi.org/10.1016/j.amc.2007.05.010