A Stacking-Based Data Mining Solution to Customer Churn Prediction

Shabankareh, M.J, Shabankareh, M.A., Nazarian, A., Ranjbaran, A. and Seyyedamiri, N.

This is an accepted manuscript of an article published by Taylor & Francis in Journal of Relationship Marketing, DOI: 10.1080/15332667.2021.1889743.

The final definitive version is available online:

https://doi.org/10.1080/15332667.2021.1889743

A Stacking-Based Data Mining Solution to Customer Churn Prediction

Abstract

In today's competitive world, organizations are in a constant struggle to retain their current customers while attracting new customers through various methods. Customer churn is a major challenge in different industries and companies. Despite their initial successful attempts at attracting customers, organizations soon face the fact that their current customers may turn away towards their rivals. By identifying churn candidates, organizations will be able to guarantee their future success by revising their customer relationship management policy. Analyzing the data of the telecommunications industries, this study provided an effective early-churn-detection solution using modern techniques by stacking data mining algorithms. Research findings indicate that integrating support vector machines (SVMs) with the chi-square automatic interaction detection (CHAID) decision tree can yield the best outcome. The results show the proper accuracy of the proposed churn prediction solution. In addition, stacking contributed to improved customer churn detection results.

Keywords: Stacking, Data Mining, Customer Churn, CHAID Decision Tree, Customer Relationship Management, Predicting Models

1. Introduction

In recent years, many organizations and industries have planned to change from being product-oriented to customer orientation which could be as a result of increase in customer knowledge and their influence on other buyers(Oh et al., 2018). This could help organizations in different ways including improve firm's cash flow as well as contribute to its stakeholders' wealth and finally lead to customer satisfaction (Peng et al., 2015). It is extremely important for organizations to retain existing, valuable customers as well as being able to attract new customers. When newly-emerged companies rely more on the acquisition of new customers, experienced companies focus more on the retention of current customers(Han et al., 2020)and The 20/80 rule mentiones that 80% of the firm's business comes from 20% of existing customers (Qiu et al., 2015). According to Chu et al (2007)costs of acquiring a new customer is five to ten times greater than that of retaining an existing one. As a matter of fact, existing customers could potentially provide much more benefits for organization both financially and non-finically (loyalty) as compare to new customers (Russo et al., 2016). According to the studies conducted by marketing experts in this filed, it has been estimated that companies could lose nearly 25% of their existing customers and 75bilion on average every year due to negligence (Kolin, 1982; Hyken, 2018). Considering this facts, enterprises are getting more interested in customer retention and preventing from customer churn. Therefore, the goal of customer churn management is to mitigate the detriment of customer churn and increase the profit of valuable customer retention.

According to Kisioglu & Topcu (2011) if customers stop doing business with the existing provider and discontinue their subscriptions and subscribe to another one, this type of customers are called a 'churner'. They further argued that churn is a major problem for companies with many customers, e.g., credit card providers or insurance companies. It has been argued that it is common among all industries that a sharp growth in competition could result customer churn which create a great concern for the companies (Shirazi & Mohammadi, 2019). According to Sharma (2017) the main reasons for customer churn can be identified as dissatisfaction, higher cost, low quality, lack of features, higher referrals, lack of engagement and privacy concern among customers. As a result the prevention of customer churn certainly will have a significant role in decreasing organizational costs, increasing market share, and improving the position of companies (Amin et al., 2019). Furthermore, at the same time, retention of a customer is much less costly than attraction of a new customer. According to Zhao et al (2009) if the customer churn rate decreases by 1%, corporate profit could potentially increases by 6%; this is important fact in profitability and growth of any companies. Therefore, in order to develop an accurate and effective model that help to reduce customer churn there are some steps that need to be taken by managers (Coussement & Van den Poel, 2008). Firstly, data collected by organization should be utilized in order to analyze customer behavior, attitudes and attributes that could potentially help to reduce customer churn. Then by analyzing the data collected those customers with high churn likelihood must be identified for further analysis and finally, appropriate policies and regulations should be formulated for this group of customers. According to Dejaeger et al (2012) customer retention benefits organizations both financially and non-financially more than attracting new customer as:

1- It is five to six times costlier to attract new customers than to retain current customers.
2- Existing customers are more profitable because they are less sensitive to marketing activities of other rivals; therefore, they are less costly to provide services for.
3- Losing customers because of declined sales is considered costly (Kisioglu & Topcu., 2011; Dejaeger et al., 2012)

There are several methods proposed by scholars in the literature including both quantitative and qualitative to measure and tackle customer churn. There are several quantitative methods that can be employed to face the customer churn problem. One of the most recent and novel one is data mining (Celik, 2019). In fact, data mining methods can be utilized to help both researchers and practitioners to develop prediction models to mitigate customer churn. In order to get better results scholars have proposed to integrate methods to achieve more appropriate and accurate model. One of the most innovative and effective method of integrations to use is the stacking technique which ensemble multiple models and helps to increase the accuracy of prediction models and improves the output.

This study adopted the stacking technique in order to enhance the accuracy of the models developed for customer churn management. Although different papers have discussed and analyzed the integration of several methods and algorithms, no study has apparently been conducted on the analysis of different stacking methods in customer churn. As a result, it could be argued that this study is first in its kind that to implement and evaluate the stacking process for various algorithms. This goal was achieved by employing six different algorithms, namely neural networks of different structures (RBF and MLP), k-nearest neighbor (KNN), support vector machine (SVM), and various types of decision trees such as C5, CHAID, and C&R.

Contributions of this study are two folds, firstly, in this study has offered an effective solution not only to predict customer churn but also to reduce it by using stacking method. This method has not been never used as a tool to predict customer churn and as a result this study is first in its kind. Secondly, this study also present interpretable results by using decision tree which help researchers to be able to analyses the causes for customer churn as well as its effect on organization itself that could potential help them to make correct decisions to reduce the negative effect of customer churn. The rest of paper is consisted of following sections: the next section includes theoretical background and related works, after that we have presented methodology and results, next section includes discussion and conclusion and finally in the last section we have presented our recommendation for future research.

2. **Theoretical Background**

2.1. Concept of Customer Churn

In the today's competitive world, many businesses are coming to a full realization of the importance of the customer-oriented business strategy for sustaining their competitive edge and maintaining a stable profit level which received mainly from customer income(Tsai & Lu, 2009). Gallo (2014) discusses in Harvard Business Review (HBR) that making a new customer is anywhere from 5 to 25 times more expensive than retaining an existing one which clearly means companies don't have to only spend time and resources to find new clients whereas there is a need to keep and satisfy the existing ones. Hence, there is logical and beneficial for organizations to make their utmost effort to satisfy their existing customers in order to improve customer retention (Agarwal, A., Harding, D.P. and Schumacher, 2004; Kotler, 2001)

Customer churn defined as the percentage of customers that stopped doing business with a company and they no longer want purchase products and services from an entity during a period of time and move to rivals (Hadden et al., 2007; Jahanzeb & Jabeen, 2007). Churn can be caused by specific problems such as technological developments, economic issues, qualitative factors, service type and coverage, and even a bad experience of encountering the telephone center employees of an organization (Kim, Park, & Jeong, 2004). Based on Olafsson et al (2019) There are 2 different types of customer churns including: voluntary and forced churn. The former refers to a situation that established customers decide not to continue purchasing or being a customer of the organization, on the other hand, the latter refers to those established customers who are no longer preferred customers and the company cuts the relationships with them. Moreover, according to Burez & Poel, (2008) voluntary churners are two groups: commercial churners and financial churners. Commercial churners are those customers that do not renew their fixed contract when is due for renewal as a result of different reasons including poor services, product problems, Cash-flow Crises. On the other hand, financial churners are those customers who stop honoring the contract due to fact they no longer can afford the services.

Burez & Poel (2007) indicate that there are two basic approaches in managing customer churn namely targeted and untargeted. Untargeted approaches rely on superior product and mass advertising to increase brand loyalty and retain customers. However, on the other hand, targeted approaches rely on identifying customers who are likely to churn, and then either provide them with a direct incentive or customize a service plan to satisfy their needs and convince them to stay. Moreover, Burez & Poel (2007) further argued that the targeted approach itself can be divided into two separate methods of reactive and proactive in managing customer churn. The former method, reactive method, is the situation when a company waits till customers themselves ask the company for cancelation of the service provided. In this situation, the company would normally offer an incentive to the customer to convince them stay. However, on the other hand, when a company adopts a proactive approach, it tries to identify customers who are likely to churn before they do so. In this situation, the company tends to provide special programs or incentives for the potential customers who are likely to churn in order to keep the customers from churning. Targeted proactive programs have some potential advantages including having lower incentive costs for company. However, at the same time this system may also be proven to be wasteful if churn predictions were inaccurate as this may put extra financial burden on companies without being able to achieve the desirable outcome which is customer retention. Therefore, it is extremely important to build an accurate customer-churn prediction model to avoid wasting resources.

2.2. Predictive modeling and Stacking method

Predictive modeling, which is perhaps the most used subfield of data mining, draws from statistics, machine learning, database techniques, pattern recognition, and optimization techniques. Predictive models in real world are useful tools for organizations and have been used in different industries and business including insurance, fraud detection in banks, text mining, telecommunication and many others. (Hong & Weiss., 2001). One of the regular and exact approaches in predictive modeling is Stacking (Czarnowski & J□drzejowicz, 2018; Mastelini et al., 2017). Stacking is a Metalearning based method which ensemble multiple classifications or regression model. This approach, also known as Stack Generalization, was first introduced by Wolpert (1992).The classification based on this method consists of two steps in which the output of one level of classifier in the first stage is used as input for the second level. In other words, it could be said that the prediction of classifier in one level are used as features for another classifier. Figure 1 shows the structure of used Stacking method in this study.

...Figure1...

In this method in order to improve the overall performance, and eventually end up with a model which is better than any sole intermediate model, we would selectively implement different models that each is best and capable of learning one specific aspect of the problem and not the whole problem but collectively are able to learn the complete problem. Therefore, we can build multiple learners and use them to build an intermediate prediction model with one prediction for each learned model. Therefore, in this method the aim is to create a new model that is based on several other models which has the advantages all those models combined and it has been designed to solve the specific problem or achieve specific aim (Ghalejoogh et al., 2020; Ribeiro & dos Santos Coelho, 2020;Wolpert, 1992).

2.3. Data Mining in Customer Relationship Management

In recent years, data generation and collection technologies have significantly changed and rapidly grown. Thus, organizations do not face the data collection challenges that used to be the major concern however, having said that, they still need to deal with the ability to extract useful information from data (Lee et al., 2004). Data mining is a process for exploring the collected data to help discover hidden patterns in the data that is not possible or extremely difficult with traditional methods. In this method, research aim to acquire proper information from a large amount of data and converting them into useful knowledge for organizations (Schuh et al., 2019).

The customer relationship management (CRM) system is an instrument that designed to help companies attract and retain their customer base by serving as the best way to analyze customer information. With this respect, data mining methods newly emerged trends and tools could serve companies to achieve high level customers' information and relevant data analysis in this competitive and global market that could potentially differentiate them from their rival and provide them with competitive advantages (Ngai et al., 2009;Rygielski et al., 2002). In addition, according to Ngai et al (2009) the application of data mining techniques in CRM is worth pursuing to develop a customer-oriented economy not only for large and resourceful companies but also for small and medium enterprises. In fact, it can be argued that the most important goals of CRM are to identify different groups of customers and formulating realistic and achievable marketing, sale, and services plans based on the needs and characteristics of each group. In modern days, sophisticated customers can be identified through the data that generated online using different medium such social media or other online platform. As a result, using this available data would potentially help organizations to analyze and understand their customers' behavior, expectation, needs and wants. This would help organizations to classified their customers based on their behavior, expectations, needs and wants and in order to do that they use various methods of classification for different purposes (Kotler, 2001). Using data mining would enable firms to gain a competitive advantage, making personalization, get economic benefits and help to build an intimate relationship between businesses and their customers (Rygielski et al., 2002).

2.4. Empirical Research Background

In this section we are planning to explore some major studies conducted by other scholars on customer churn using different data mining method by explaining the algorithms used dataset collected and results of the studies.

The first study considered was the study by Ferreira et al., (2004) who employed a multilayer perceptron for the study by using C4.5 decision trees, hierarchical neuro-fuzzy systems, and a data mining instrument named the Rule Evolver based on the genetic algorithm implemented on a telecommunications dataset in Brazil. They found that neural networks of 15 hidden layers performed the best classification and as a result of that company is able to save cost from customer churn.

In another study by Buckinx & Van Den Poel, (2005), they utilized logistic regression and random forest techniques in order to separate loyal and disloyal customers. They concluded that personal characteristics of customers such as the length of customer relationship, mode of payment, buying behaviour across categories, usage of promotions and brand purchase behavior could be used as better functions to separate loyal customer behavior than RFM variables (Recency, Frequency, and Monetary value) in order to identify churn candidates.

In the study of telecommunication company Wang et al., (2009) used a decision tree to explore 60000 transaction records from 4000 customers communications during a three-month period. According to the rules obtained from their model, the customer's final connection to the network and the number of network connections were identified as the variables helping predict customer churning or loyalty behavior. In another study in telecommunication industry in South Korea Ahn et al., (2006) used the data of transactions and payments, including 5789 records: 5137 retained, 652 churned, made by subscribers to analyze their churn factors and concluded that dissatisfaction, change costs, and service use rates affected clients' decisions of staying or leaving. Furthermore, they argued that some churn determinants such as Customer dissatisfaction, Switching costs, customer related variables, service usage and customer status influence customer churn, both directly or indirectly through a customer's status change.

In the study of online multimedia service companies Tsai & Chen (2010) explored more than 37000 records of customer interactions on a telecommunication company providing the MOD services in Taiwan through a decision tree method, they identified network connection time and discount rate of services as the most important factors predicting customer churn in the studied company. Furthermore, Meghyasi & Rad, (2020) analyzed customer churn in Iranian telecommunication giant, Irancell Co. using data mining methods. For the purpose of the study they employed genetic and neural network techniques where the genetic method optimized the neural network structure in the proposed approach to achieve higher accuracy. As a result of that the resultant accuracy of the proposed method was improved and achieved95.5% on the studied dataset. Ahmed et al., (2019) analyzed data collected from telecommunication industry in Pakistan to identify transitional learning on customer churn. In order to do that they proposed two steps method wherein the first step, pre-trained convolutional networks were employed to perform the transitional learning process. Telecommunications industries usually benefit from vector data, but in this study, convolutional network capacities were utilized to convert vector data into two-dimensional matrices. In the second step, the predicted results of this method were added to the end of the main information vector and then the final model was developed on the resultant data through the genetic and AdaBoost technique. Findings show that the acceptable accuracy of the proposed solution the prediction accuracy obtained was 75.4% and 68.2%, while the area under the curve was 0.83 and 0.74, respectively.

In one major study Hadden et al., (2007) analyzed the techniques and models used in customer churn management and found some very interesting results. They argued that , decision trees are the most popular algorithm that have been used by scholars for customer churn and other methods including Fuzzy and neural network have gained less popularity or being ignored which according to Hadden et al., (2007) should be used more frequently to predict customer churn.

By adopting a decision tree method, Liou, (2009)analyzed the data collected from of an airline company which he managed to develop a method for customer churn behavior prediction for three different classes of customers, loyal , churn candidates, and churning customers. In order to identify the customer churning behavior predictors, he used a novel decision rule approach called the VC-DRSA where some variables were introduced as major predictors for customer churning behavior including average income, employment in the public sector, prices of products, and use of products during flights.

In another study of the customer churn problem in public telecommunication industry, Sjarif et al., (2019) employed Pearson product-moment correlation coefficients and k-NN to analyze data collected using Kaggle dataset. In their proposed method, Pearson product-moment correlation coefficients were utilized in data preprocessing to extract effective features whereas, the k-NN technique ($2 \leq k \leq 20$) was employed to develop and evaluate many models to reduce effect of the noise on the classification Finally, the result showed that the K Nearest Neighbor (k-NN) algorithm performs well compared to the others with the accuracy for training is 80.45% and testing 97.78%. Moreover, in another study of Telecommunication Company, Sandhya et al., (2019) addressed the customer churn problem by recommending an Intersection-Randomized Algorithm (IRA) using Map Reduce functions to avoid data duplication in the mobile user call data. In addition, they also used the agent-based model (ABM) to predict the complex mobile user behaviour to prevent customer churn with a particular telecommunication service provider. The results show that their proposed method increased accuracy due to the dynamic nature of agents and introduced certain rules based on customer behavior.

Zhou et al., (2019) proposed a novel model named DLCNN, a deep learning approach to address the customer churn problem in an online music service and the sample consists of 625850 customers. Their proposed method benefited from a long/short-term memory (LSTM) and a convolutional neural network (CNN) to analyze time series data where it enables researcher to learn hidden patterns in sequential data. Furthermore, they also employed the XGBoost method to construct new features from the existing features and concluded cross-layer connections learn more complex features automatically which can make the model perform better. Moreover, in another study of service provider in financial industry Mena et al., (2019) analyzed deep learning solutions of RFM data and assessed the performance of neural networks for churn modeling. According to their results, using RFM data along with the LSTM network would result in better outcomes than the logistic regression method. Moreover, using the features extracted by the LSTM for the regression method increased the regression accuracy by 25percent as compared to a model with only static features.

In the study of an Airline company Zhu et al., (2019)employed a method based on the random forest and the LASSO model for customer churn prediction. In their proposed method, the LASSO method was utilized to select a subset of analysis features and then the selected subset was used for training the random forest technique. According to the results, the simultaneous use of LASSO and random forest techniques would provide higher prediction accuracy and stronger generalization ability than each of these methods separately in both computation and performance accuracy.

Idris et al., (2019) proposed a method based on the integration of genetic and AdaBoost methods to solve the customer churn problem in two standard telecom data sets. In their paper, searching capabilities of genetic programming (GP) and classification capabilities of AdaBoost are integrated in order to evolve a high-performance churn prediction system having better churn identification abilities and stated that the search power of the genetic method and the classification features of the AdaBoost method can be employed to provide an effective churn-prediction solution. Therefore, for this purpose, different subsets of analysis data were selected by the genetic method and modeled on the selected data through the AdaBoost technique. In addition, they also used a sampling technique using the PSO algorithm to solve the problem of data imbalance in telecommunications companies. Their results offer better understanding and learning of churners (ChPGPAB system yields 0.91 AUC and 0.86 AUC) and identified underlying factors responsible for churn behavior of customers.

Smith et al., (2000) used a data mining methodology in the study of insurance industry, which views the knowledge discovery process within a holistic framework utilizing hypothesis testing, statistics, clustering, decision trees, and neural networks at various stages. The purpose of that study was to demonstrate the potential of capability data mining as a means for achieving market growth and profitability. Firstly, they found that price have a significant impact on the policy holder's decision to renew or terminate their contract. Secondly, they used undirected data mining to analyze the claim patterns by clustering the data to reveal the natural data structures. Groups with unacceptably high cost ratios (contributing to low profitability) can be identified in this manner. Finally, the results of these two analyses were combined into a framework for determining the optimal premium price for an individual policy: one that balances the opportunity for profit with the need to retain the customer.

In conclusion, based on the thorough literature review on different methods used to predict customer churn, different papers have used or offered several integrating methods and algorithms to predict customer churn. However, having said that, there is no study that offer analyzing customer churn by using different methods adopting stacking techniques. As a result, it can be argued that this is the first study of its kind that to implement and evaluate various algorithms using the stacking process. By learning from previous implementations, the selected method in this study offers a new technique that to integrate support vector machines with the CHAID algorithm. This proposed integrated method was never used in any previous studies in this area and it is for the first time ever that has been adopted for marketing area to predict customer churn.

3. Research Methodology

In the data mining process, there are four Steps that must be followed carefully. The first step is data preparation which aims to provide proper input for the model learning step. In this step, unprocessed data are first extracted from all existing data sources and then they are being processed in a separate step. The main purpose of this step is to identify potential problems and errors exist in data after data collection phase and try to resolve them before progressing to the next stage (Jiawei Han, Micheline Kamber, 2000). In the next step, based on the nature of data various algorithms are employed to develop a model learning that help to identify different orders exist in the data to enable researcher to explore hidden knowledge exist in the data. In the model evaluation step, various criteria such as accuracy are employed to test the model(s) used in the modeling step. In all cases, what is important for researcher and authors that data mining models should help and enhance the decision-making process (Jiawei Han, Micheline Kamber, 2000).

3.1. Research Dataset

The data used in this study was based on the existing dataset from large telecommunications company including demographic information as well as information on the specifications of services received by clients. This dataset specifically highlighted the clients who left the company for different reasons. The dataset accessible through https://www.ibm.com where information of around 7043 clients are provided for analysis.

Table (1) shows the dataset features and their types. The type of each feature depends on the values assigned to that feature. Different types such as numerical (including only numbers), nominal (discrete values of strings and numbers), and letters can be assigned to any of those features.

…Table1…

3.2. Data Preparation

Data preparation is a prerequisite step in data extraction process where researchers aim to delete missing values before exploring and extracting hidden knowledge from data. In another word, the presence of missing values prevents the development of learning models and as a result not being to gain knowledge hidden in data. According to Kapil, (2018) there are various methods to deal with missing values such as closest value, mean value, median value. Given the fact that the dataset used in this study has eleven records with missing values, these records have been excluded from analysis quite easily.

Another step of data preparation is to convert multi-value features into numerical features. Although multi-value features can be used in some algorithms, these features are not supported by neural networks. Therefore, data normalization step is considered as another important step in data preparation process. For the purpose of this study and in order to follow data normalization step, authors adopted the Z-transform method as method of data normalization. In addition, it has proposed in order to simplifying the modeling process, researchers could also use sampling techniques as analysis of the large data could be times consuming and costly. In fact, drawing sample from the large population could help researcher to have better and stronger analysis when there is hardware equipment limitations.

3.3. Modeling

In this section, different classification algorithms have been used to develop a model for the analysis of the dataset used in this study. These algorithms have been used are as follow;
Neural Network Method: Inspired by the complicated features of the brain in which billions of neurons are responsible to process information in parallel, the artificial neural network was designed and developed (Rygielski et al., 2002). Recent researches have shown that artificial neural networks have many capacities in classification and pattern recognition which are inspired by biological systems, artificial neural networks that enable researchers to detect unknown and complicated patterns of data and samples analysis (Zhang et al., 1998). This study utilized neural networks of different structures such as RBF with different networks and MLP with different numbers of layers and neurons.

Decision Trees: Decision trees are among the most useful display methods of classifiers. Considerable amount of contents are allocated to this form of classification (Espejo et al., 2010). The decision tree algorithm is an important theory used in many applied areas and fields including Marketing, decision making (Li-qunLin, 2014); this algorithm repeatedly decomposes and divides a dataset of objects through depth-first search methods. The decision tree classification technique is implemented in two phases: tree development and tree pruning. For the purpose of this researcher, researchers adopted three types of decision trees, namely C5, CHAID, and C&R Support Vector Machine: The SVM is a non-statistical binary classifier that has been receiving a great deal of attention in recent years. In this method, all samples are used along with an optimization algorithm to acquire the samples of class boundaries. These samples are then utilized to determine an optimal linear decision-making boundary for the separation of classes; it is solved by square programming problems (Cristianini et al., 2000).
k-Nearest Neighbor: The k-NN search algorithm is one of the most common and used classification and learning techniques that was introduced by Fix and (Fix & Hodges, 1989). It is known as an effective and simple algorithm aiming to find the nearest *k* in the existing training data. When a sample appeared, it will be classified to the most closest and similar category (Adeniyi et al., 2016).
Table (2) shows an overview of advantages and disadvantages of all data mining algorithms used in this study:

…Table2…

3.4. *Implementation and Analysis*

The proposed solutions were implemented in IBM SPSS Modeler, which provides researchers with different data mining algorithms. In order to evaluate these algorithms the accuracy criterion method was employed. Moreover, the cross-validation technique was utilized to ensure about the performance of methods. In cross-validation method, training data are divided into K sections along the training process where each time, one section is used for evaluation when other sections are used for training. The process of testing each method is repeated several times, and at the end the results are averaged.

In SPSS Modeler, different processes are performed through the designed operators. According to Figure (1), GetData operator is employed to load data of different formats. Data preparation processes are performed through Auto Data Preparation and Type operators. C&R tree and SVM models were used in this study. Then Filter and Partition operators were employed to develop appropriate datasets and perform the cross-validation process. In addition, Merge operator was utilized to merge the SVM results yielded on different datasets. After that, they were averaged and shown by Table operator. Since the cross-validation method was implemented only on a part of existing algorithms in IBM SPSS Modeler, this operator was applied in a relatively complicated manner shown in Figure (2).

…Figure2…

3.5. *Results*

For the purpose of this study, few parameters of some algorithms were adjusted in the instrument; therefore, the results of different methods were reported by considering different parameters. The parameters have been used in this study were chosen among different parameters by using trial and error method in each classifier to identify most suitable ones. In this study, basic algorithms were first analyzed separately. Table 3 shows the analysis results including the parameters of each algorithm, which includes best result, worst result and mean result. Each of the basic algorithms has specific parameters such as in the k-NN algorithm, three different values of 3, 5 and 8 were used whereas in the C&R tree and CHAID algorithms, Max Tree Depth shows the maximum tree depth.

In both tree algorithms, three different tree depths of 2, 5, and 10 were used. The number of layers and the number of neurons were adjusted parameters in the neural network algorithm (for instance, MLP (5,5) means using the MLP neural network of two layers, including five neurons each). On the other hand, four types of layers having different neurons were utilized in this study. In the neural network algorithm, the number of RBFs shows the number of RBF types of neural networks used in two different quantities. Furthermore, three different types of kernels (Sigmoid, RBF, and Linear) were used in the SVM algorithm

According to Table (3), better results are obtained if the depths of C&R and CHAID trees are higher in the decision tree algorithm, if the value of k is smaller in k-NN, and if there are larger numbers of neurons, layers, and networks in the neural network algorithm. Table (3) indicates the higher accuracy of the SVM than other algorithms when Kernel=RBF. Since the sampling process was considered stochastic for the implementation of cross validation, the results were reported in the worst, best, and mean ranges. In the next step, these algorithms were integrated through the stacking process.

…Table3…

3.6. Stacking

The modeling process can benefit from the capacities of several algorithms simultaneously to improve the modeling results. Stacking is one of such methods that do that. In this method, one of the algorithms is first applied to data and then the algorithm execution results are added to the main dataset which normally, include the labels predicted by the algorithm. After that, the new data set will be transferred to the main algorithm to benefit simultaneously from the main data and the labels predicted by the first method (Wolpert, 1992).

In this study, the best results obtained from basic algorithms (Table 3) were merged to use the stacking method. Table (4) shows the results of implementing of the stacking method.

…Table4…

In the next stage, based on the ordering of the results from the best to worst, results were yielded by SVM + CHAID, SVM + C5, SVM + C&R Tree, SVM + Neural NET, and SVM + k-NN hybrid algorithms showing mean accuracies of 85.8%, 0.854, 0.849, 0.828, and 0.812, respectively. The SVM + CHAID hybrid method had an appropriate performance even in the worst case. Figure (3) shows the stacking results.

…Figure3…

4. Further Evaluation

There is no doubt that the best solution can be achieved by analyzing different dataset in order to ensure the accuracy of results; however, since there is limited access to reputable datasets, the proposed method was implemented on different dimensions to analysis the data used in this study. In order to evaluate the results further, the method proposed for this study used to further investigate different aspects of data and results were shown In table (5) where that the results indicate that the larger the data type the better the results.

…Table5…

Generally speaking, the CHAID method has stronger and more sophisticated structure as compared to other method such as SVM where potentially the predicted results are better and more trustworthy.

Given the fact that the outcome of the decision tree is sometimes too large where it is not possible to be covered in one paper, as a result some of the main results have been reviewed here:

1. One of the most important features used in developing a tree is the contract. It could be argued that customer churn rates are largely related to short-term contracts.
2. If the Internet connection is of an older technology such as DSL, the churn rate is lower in the face of a new technology such as the optical fiber.
3. The individuals who have a shorter period of working experience will be more churn candidates.

Figure (4) indicates the results of implementing of the CHAID method.

…Figure4…

5. Discussion and Conclusion

The proposed process is based on stacking data mining algorithms. From previous studies it is obvious that stacking can help to develop better prediction models in different areas like customer churn (Ghalejoogh et al., 2020; Mastelini et al., 2017; Ribeiro & dos Santos Coelho, 2020). According to the implementation results, stacking methods can help improve the customer churn identification results in different organizations. Unfortunately, many industries and companies do not record and keep appropriate or sufficient data to determine the customer churn status. The telecommunications industry is one of the rare industries that collecting this information properly (Jahanzeb & Jabeen, 2007). Hence, the data of the telecommunications industry were analyzed in this study to offer an effective churn-detection solution. In this study, authors decided to analyzed data from one major telecommunication company were data could provide better view and understanding of customer churn and a result better and more effective churn- prediction method could be drawn(Amin et al., 2019). The following section summarizes the result, discusses implications and suggests areas for further study.

First, the best results were found in this study were the mixed implementation of decision tree and support vector machine methods could provide the best results in the case of this study. The research results can be applied to other training data in customer churn. Given the nature of data mining techniques, it would be false to state that the proposed method will definitely yield the best results for every dataset because data mining algorithms usually depend on training data sets. Therefore, they may produce different results if different datasets are used for training.

Second, decision tree methods used in this study were interpretable, their implementation results can be employed to achieve a set of interpretable rules and identify the relationships between numerous characteristics in the dataset (Bach et al., 2018). The results show that analyzing the structure of a decision tree generated by this method prove to be more efficient and effective to provide better insight into the causes of customer churn. Based on the results which arise from CHAID, customers with short-term contracts are on the top of list of churn candidates. Therefore, it is recommended that managers should provide motivations, considerable discounts, and customer loyalty plans to make mid-term and long-term contracts feasible so that the churn rate decreases. Surprisingly, the data show that, the use of new technologies as a negative impact and it is considered as a cause of customer churn. Although this problem requires accurate analysis and further research by using some other means of data collection including survey questionnaires or interviews. But, having said that, it appears that the company chosen for this study does not provide clients with up-to-date and appropriate facilities and technology as compared to its rival. As a result, the clients who benefit from modern communication methods will prefer other companies.

Third, according to the results, data mining methods can be utilized to achieve an appropriate accuracy in customer churn prediction. The results were obtained from the use of all characteristics of each customer apart from customers' ID. This shows that the simultaneous use of demographic characteristics and features of services, used by clients, can lead to the development of accurate models in this case. As mentioned before, the main aim of this research was to investigate the reasons that could potentially increase 'customer churn' as well as shows how new methods of data mining could help both researchers and practitioners to both collect and analyses data more efficiently and effectively. For the purpose of this study, authors have proposed a new and innovative data mining method that could potentially be used in future studies by both academics and scholars. Furthermore, it is crucial to emphasis the importance of these kind methods to measure customer churn for both students and academics in the field of business management and computer science.

There is no doubt customers are considered as the most important assets regardless of the industry or size of the organization and previous studies have always emphasized that the retention of old customers is much more beneficial and cheaper than the attraction of new customers in a competitive markets like telecommunication. From a managerial point of view, and in order to enhance organizational performance, it is essential to develop a method of proper accuracy to predict customer churn. Therefore, organizations should always pay special attention to the identification of the customers having high likelihood of churn. By identifying churn candidates early, it is possible to prevent the loss of organizational assets.

6. Recommendation for future work

To continue this study and expand research results, it is recommended to use the data of domestic organizations on customer churn. It is also recommended to merely employ specific algorithms such as different types of decision trees with interpretable results in order to conduct an accurate analysis of causes of customer churn and identify the most important factors affecting customer churn. This problem may decrease the accuracy of final results, although it can give a proper interpretation of causes of customer churn. It is also possible to benefit from evolutionary methods in the feature selection process to develop different models with the help of different subsets of data. Therefore, it is possible to determine the priorities of features and even delete the features that have no effect in the process of identifying customer churn. If the extra features are deleted, it will be possible to optimize the process of storing the information required by users. In the stacking process, novel deep learning methods can be employed to compare their results with those of the proposed method by merging different layers of different structures. Moreover, this study further suggest that future studies could use different combining methods such as bagging or stacking by combining more than two algorithms.

7. conflict of interest

There are no relevant financial or non-financial competing interests to report.

References:

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, *12*(1), 90–108. https://doi.org/10.1016/j.aci.2014.10.001

Agarwal, A., Harding, D.P. and Schumacher, J. R. (2004). Customer relationship management (CRM) & corporate renaissance. *McKinsey Quarterly*, *3*, 80–91.

Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). *Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry*. 1–10. http://arxiv.org/abs/1901.06091

Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, *30*(10–11), 552–568. https://doi.org/10.1016/j.telpol.2006.09.006

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

Bach, M. P., Dumičić, K., Žmuk, B., Ćurlin, T., & Zoroja, J. (2018). Internal fraud in a project-based organization: CHAID decision tree analysis. *Procedia Computer Science*, *138*, 680–687. https://doi.org/10.1016/j.procs.2018.10.090

Buckinx, W., & Van Den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, *164*(1), 252–268. https://doi.org/10.1016/j.ejor.2003.12.010

Burez, J., & Poel, D. Van Den. (2007). *CRM at a pay-TV company□: Using analytical models to reduce customer attrition by targeted marketing for subscription services*. *32*, 277–288. https://doi.org/10.1016/j.eswa.2005.11.037

Burez, J., & Poel, D. Van Den. (2008). *Separating financial from commercial customer churn□: A modeling step towards resolving the conflict between the sales and credit department*. *35*, 497–514. https://doi.org/10.1016/j.eswa.2007.07.036

Celik, O. (2019). *Comparing to Techniques Used in Customer Churn Analysis*. *4*(1), 30–38.

Chris Rygielski a, Jyun-Cheng Wang b, David C. Yen a, ∗. (2002). Data mining techniques for customer relationship management. *Technology in Society*, *4*(24), 483–502. https://doi.org/10.1016/S0160-791X(02)00038-6

Chu, B. H., Tsai, M. S., & Ho, C. S. (2007). Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems*, *20*(8), 703–718. https://doi.org/10.1016/j.knosys.2006.10.003

Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems* *with Applications*, *34*(1), 313–327. https://doi.org/10.1016/j.eswa.2006.09.038

Czarnowski, I., & J□drzejowicz, P. (2018). An approach to data reduction for learning from big datasets: Integrating stacking, rotation, and agent population learning techniques. *Complexity*, *2018*. https://doi.org/10.1155/2018/7404627

Dejaeger, K., Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector□: A profit driven data mining approach New insights into churn prediction in the telecommunication sector□: A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229. https://doi.org/10.1016/j.ejor.2011.09.031

Espejo, P. G., Ventura, S., & Herrera, F. (2010). A survey on the application of genetic programming to classification. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(2), 121–144. https://doi.org/10.1109/TSMCC.2009.2033566

Ferreira, J. B., Vellasco, M., Pacheco, M. A., & Barbosa, C. H. (2004). Data mining techniques on the evaluation of wireless churn. *European Symposium on Artificial Neural Networks*, *December 2013*, 483–488.

Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review*, *57*(3), 238–247. https://doi.org/10.2307/1403797

Gallo, A. (2014). The Value of Keeping the Right Customers. *Harvard Business Review*. https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

Ghalejoogh, G. S., Kordy, H. M., & Ebrahimi, F. (2020). A hierarchical structure based on Stacking approach for skin lesion classification. *Expert Systems with Applications*, *145*, 113127. https://doi.org/10.1016/j.eswa.2019.113127

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, *34*(10), 2902–2917. https://doi.org/10.1016/j.cor.2005.11.007

Han, H., Jongsik, Y., & Hyun, S. S. (2020). Nature based solutions and customer retention strategy: Eliciting customer well-being experiences and self-rated mental health. *International Journal of Hospitality Management*, *86*(February 2019). https://doi.org/10.1016/j.ijhm.2019.102446

Hong, S. J., & Weiss, S. M. (2001). Advances in predictive models for data mining. *Pattern Recognition Letters*, *22*(1), 55–61. https://doi.org/10.1016/S0167-8655(00)00099-4

Hyken, S. (2018). *Businesses Lose $75 Billion Due To Poor Customer Service*. Forbes. https://www.forbes.com/sites/shephyken/2018/05/17/businesses-lose-75-billion-due-to-poor-customer-service/#69506dff16f9

Idris, A., Iftikhar, A., & Rehman, Z. ur. (2019). Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Cluster Computing*, *22*(s3), 7241–7255.

https://doi.org/10.1007/s10586-017-1154-3

Jahanzeb, S., & Jabeen, S. (2007). Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. *Journal of Database Marketing & Customer Strategy Management*, *14*(2), 120–129. https://doi.org/10.1057/palgrave.dbm.3250043

Jiawei Han, Micheline Kamber, J. P. (2000). *Data Mining: Concepts and Techniques* (Third). Morgan Kaufmann. https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques

Kapil, A. R. (2018). *Methods of Missing Value Treatment and their effect on the Accuracy of Classification Models*. *October*. https://doi.org/10.13140/RG.2.2.31137.86881

Kisioglu, P., & Topcu, Y. I. (2011). Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. *Expert Systems with Applications*, *38*(6), 7151–7157. https://doi.org/10.1016/j.eswa.2010.12.045

Kolin, P. (1982). *Successful Writing at Work* (11th ed.). University of Southern Mississippi.

Kotler, P. (2001). *Marketing Management, Millenium Edition* (Tenth). Prentice-Hal. https://www.perspectiva.md/ro/files/biblioteca/Kotler-Marketing Management Millenium Edition.pdf

Lee, S. C., Suh, Y. H., Kim, J. K., & Lee, K. J. (2004). A cross-national market segmentation of online game industry using SOM. *Expert Systems with Applications*, *27*(4), 559–570. https://doi.org/10.1016/j.eswa.2004.06.001

Li-qunLin, ping S. (2014). An Algorithm of Multi-Variable Decision Tree Based on Genetic Programing. *International Conference on Artificial Intelligence and Software Engineering*, 314–=

Liou, J. J. H. (2009). A novel decision rules approach for customer relationship management of the airline market. *Expert Systems with Applications*, *36*(3 PART 1), 4374–4381. https://doi.org/10.1016/j.eswa.2008.05.002

Mastelini, S. M., Santana, E. J., Cerri, R., & Barbon, S. (2017). DSTARS: A multi-target deep structure for tracking asynchronous regressor stack. *Proceedings - 2017 Brazilian Conference on Intelligent Systems, BRACIS 2017*, *2018-Janua*, 19–24. https://doi.org/10.1109/BRACIS.2017.30

Meghyasi, H., & Rad, A. (2020). *Customer Churn Prediction in Irancell Company Using Data Mining Methods Customer Churn Prediction in Irancell Company by Using Data Mining*.

Mena, C. G., De Caigny, A., Coussement, K., De Bock, K. W., & Lessmann, S. (2019). *Churn Prediction with Sequential Data and Deep Neural Networks. A Comparative Analysis*. 1–12. http://arxiv.org/abs/1909.11114

Nello Cristianini, J. S.-T. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511801389

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, *36*(2 PART 2), 2592–2602. https://doi.org/10.1016/j.eswa.2008.02.021

Oh, Y., Lee, J., & Kim, N. (2018). The contingency value of the partner firm's customer assets in a business-to-business relationship. *Industrial Marketing Management*, *73*(December 2017), 47–58. https://doi.org/10.1016/j.indmarman.2018.01.020

Olafsson, S., Li, X., & Wu, S. (2019). *Operations research and data mining*. *187*(2008), 1429–1448. https://doi.org/10.1016/j.ejor.2006.09.023

Peng, C. L., Lai, K. L., Chen, M. L., & Wei, A. P. (2015). Investor sentiment, customer satisfaction and stock returns. *European Journal of Marketing*, *49*(5–6), 827–850. https://doi.org/10.1108/EJM-01-2014-0026

Qiu, H., Ye, B. H., Bai, B., & Wang, W. H. (2015). Do the roles of switching barriers on customer loyalty vary for different types of hotels? *International Journal of Hospitality Management*, *46*, 89–98. https://doi.org/10.1016/j.ijhm.2015.01.015

Ribeiro, M. H. D. M., & dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Applied Soft Computing Journal*, *86*, 105837. https://doi.org/10.1016/j.asoc.2019.105837

Russo, I., Confente, I., Gligor, D. M., & Autry, C. W. (2016). To be or not to be (loyal): Is there a recipe for customer loyalty in the B2B context? *Journal of Business Research*, *69*(2), 888–896. https://doi.org/10.1016/j.jbusres.2015.07.002

Sandhya, N., Samuel, P., & Chacko, M. (2019). Feature intersection for agent-based customer churn prediction. *Data Technologies and Applications*, *53*(3), 318–332. https://doi.org/10.1108/DTA-03-2019-0043

Schuh, G., Reinhart, G., Prote, J. P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. *Procedia CIRP*, *81*, 874–879. https://doi.org/10.1016/j.procir.2019.03.217

Sharma, R. R. (2017). Evaluating Prediction of Customer Churn Behavior Based On Artificial Bee Colony Algorithm. *International Journal Of Engineering And Computer Science*, *6*(1), 20017–20021. https://doi.org/10.18535/ijecs/v6i1.32

Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, *48*(February), 238–253. https://doi.org/10.1016/j.ijinfomgt.2018.10.005

Sjarif, N. N. A., Yusof, M. R. M., Wong, D. H. Ten, Ya'akob, S., Ibrahim, R., & Osman, M. Z. (2019). A customer Churn prediction using Pearson correlation function and K nearest neighbor algorithm for telecommunication industry. *International Journal of Advances in Soft Computing and Its*

Applications, *11*(2), 46–59.

Smith, K. A., Willis, R. J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining□: a case study. *Journal of the Operational Research Society*, *51*(5), 532–541. https://doi.org/10.1057/palgrave.jors.2600941

Tsai, C. F., & Chen, M. Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, *37*(3), 2006–2015. https://doi.org/10.1016/j.eswa.2009.06.076

Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, *36*(10), 12547–12553. https://doi.org/10.1016/j.eswa.2009.05.032

Wang, Y. F., Chiang, D. A., Hsu, M. H., Lin, C. J., & Lin, I. L. (2009). A recommender system to avoid customer churn: A case study. *Expert Systems with Applications*, *36*(4), 8071–8075. https://doi.org/10.1016/j.eswa.2008.10.089

Wolpert, D. (1992). Stacked Generalization (Stacking). *Neural Networks*, *5*, 241–259.

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhao, X., Wang, Y., & Cha, H. (n.d.). A new prediction model of customer churn based on PCA analysis. *2009 1st International Conference on Information Science and Engineering, ICISE 2009*, 4657–4661. https://doi.org/10.1109/ICISE.2009.100

ZHOU, J., YAN, J., YANG, L., WANG, M., & XIA, P. (2019). Customer Churn Prediction Model Based on LSTM and CNN in Music Streaming. *DEStech Transactions on Engineering and Technology Research*, *aemce*, 254–261. https://doi.org/10.12783/dtetr/aemce2019/29520

Zhu, Q., Yu, X., Zhao, Y., & Li, D. (2019). Customer churn prediction based on LASSO and Random Forest models. *IOP Conference Series: Materials Science and Engineering*, *631*(5). https://doi.org/10.1088/1757-899X/631/5/052008

Table1. Features of the Studied Dataset

No.	Feature	Description
1	customerID	This field shows the customer ID, which is a combination of numbers and letters.
2	Gender	This field indicates the customer gender (male or female). This field is nominal.
3	SeniorCitizen	Is the client a first-class citizen or not? This field takes two values: “0” and “1.”
4	Partner	The client’s marital status: “Yes” or “No”
5	Dependents	Status of dependents: “Yes” or “No”
6	Tenure	The client’s period of use. This field is numerical.
7	PhoneService	The client’s phone service subscription status: “Yes” or “No”
8	MultipleLines	The client’s multiline service subscription status: “Yes” or “No” or “No Phone Line”
9	InternetService	This field shows client’s Internet status, including various values such as “DSL” and “Fiber.”
10	OnlineSecurity	Does the client use the online security services provided by the company?
11	OnlineBackup	Does the client use the backup services provided by the company?
12	DeviceProtection	Does the client use the device protection services provided by the company?
13	TechSupport	Does the client use the technical support services provided by the company?
14	StreamingTV	Does the client use the TV streaming services provided by the company?
15	StreamingMovies	Does the client use the movies streaming services provided by the company?
16	Contract	The client’s contract has been set “monthly”, “annual”, or “biennial”.
17	PaperlessBilling	Was billing on paper or paperless?
18	PaymentMethod	How has the client paid the bills?
19	MonthlyCharges	How much has the client paid monthly?
20	TotalCharges	How much has the client paid in total?
21	Churn	This field shows the client churn status in two values: “Yes” or “No”

Table2. Advantages and Disadvantages of used Data Mining Algorithms in this study

Method	**Advantages**	**Disadvantages**
Decision Tree	• This method is used for both linear and nonlinear problems. • Tree-based methods outperform linear regression techniques for multi-value variables. • This method does not need data preprocessing. • This method gives a proper intuitive perception of data.	• High probability of data overfitting; the tree pruning process can be utilized to solve this problem. • Sensitivity to outliers • In complicated datasets, the tree structure is oversized. • Encountering continuous data, a part of useful information is lost.
SVM	• The SVM is used for both linear and nonlinear problems. • Despite a large number of features and a small number of training samples, the SVM performs properly.	• If there is large number of outliers in data, this method will not perform properly.
k-NN	• The k-NN is a nonparametric method. Since nonparametric methods have fewer assumptions on a problem, they are less flexible than other methods. • Simple implementation	• The k-NN method is relatively slow. • Selecting the right value for k has a significant effect in results.
Neural Network	• If trained with the right number of data, a neural network usually provides a better accuracy. • Ability to learn complicated problems	• Neural networks normally need a large number of training data. • A neural network is a relatively slow technique.

Table3. Implementation Results of Basic Algorithms

Worst	**Best**	**Mean**	**Parameter**	**Algorithm**
0.80	0.819	0.81	---	**C5**
0.785	0.805	0.79	Max Tree Depth(3)	**C&R Tree**
0.724	0.809	0.779	Max Tree Depth(5)	**C&R Tree**
0.780	0.80	0.791	Max Tree Depth(10)	**C&R Tree**
0.822	0.842	0.83	K = 3	**k-NN**
0.815	0.833	0.825	K = 5	**k-NN**
0.800	0.821	0.808	K =8	**k-NN**
0.71	0.754	0.735	MLP(5)	**Neural Net**
0.727	0.745	0.739	MLP(10)	**Neural Net**
0.732	0.748	0.741	MLP(5,5)	**Neural Net**
0.769	0.81	0.792	MLP(10, 10)	**Neural Net**
0.716	0.739	0.729	RBF(5)	**Neural Net**
0.717	0.742	0.734	RBF(10)	**Neural Net**
0.756	0.801	0.785	Max Tree Depth(3)	**CHAID**
0.778	0.802	0793	Max Tree Depth(5)	**CHAID**
0.793	0.805	0.795	Max Tree Depth(10)	**CHAID**
0.832	0.865	0.85	Kernel = RBF	**SVM**
0.717	0.745	0.733	Kernel = Sigmoid	**SVM**
0.791	0.812	0.797	Kernel = Linear	**SVM**

Table4. Results of Implementing the Stacking Method

Worst	**Best**	**Mean**	**Proposed Models**
0.853	0.864	0.858	**SVM+CHAID**
0.847	0.858	0.854	**SVM+C5**
0.855	0.842	0.849	**SVM+ C&R Tree**
0.831	0.798	0.828	**SVM + Neural Net**
0.785	0.834	0.812	**SVM+ k-NN**

Table5. Implementing the Proposed Method on Data of Different Dimensions

Worst	**Best**	**Mean**	**Quantity of Records**	**Models**
0.843	0.859	0.849	2000	SVM+ CHAID
0.841	0.860	0.847	4000	SVM+ CHAID
0.851	0.861	0.855	6000	SVM+ CHAID
0.853	0.864	0.858	Entire dataset	SVM+ CHAID

Figure1. Structure of used Stacking method

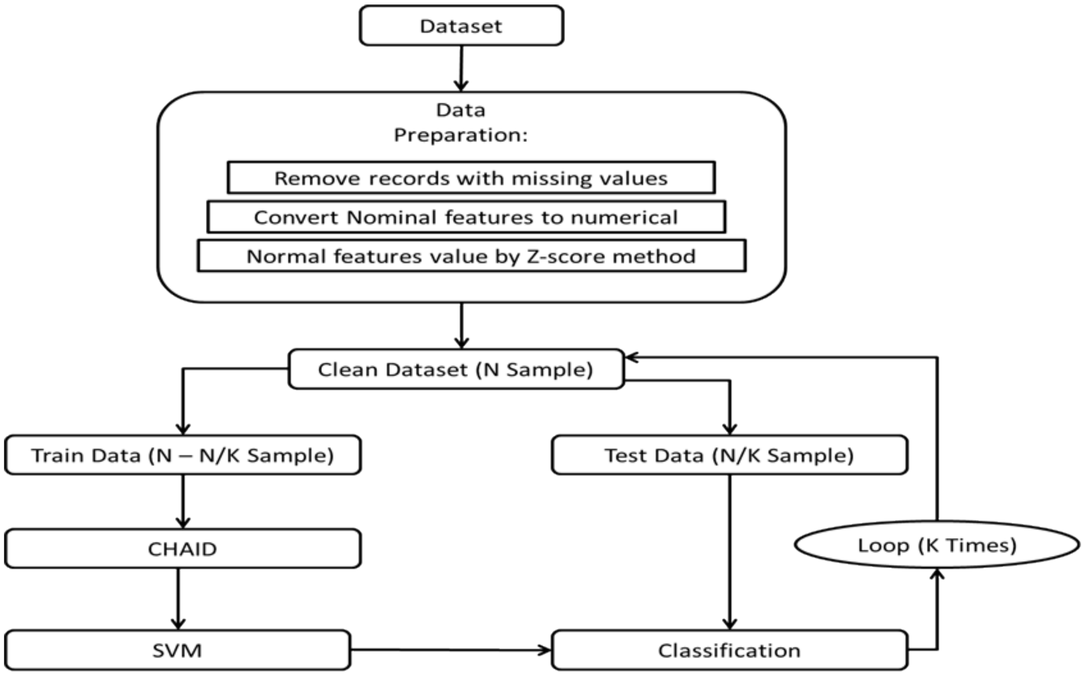


Figure2. Implementing the Cross-Validation Process in IBM SPSS Modeler (K=5. Staking process by using tree and SVM methods)

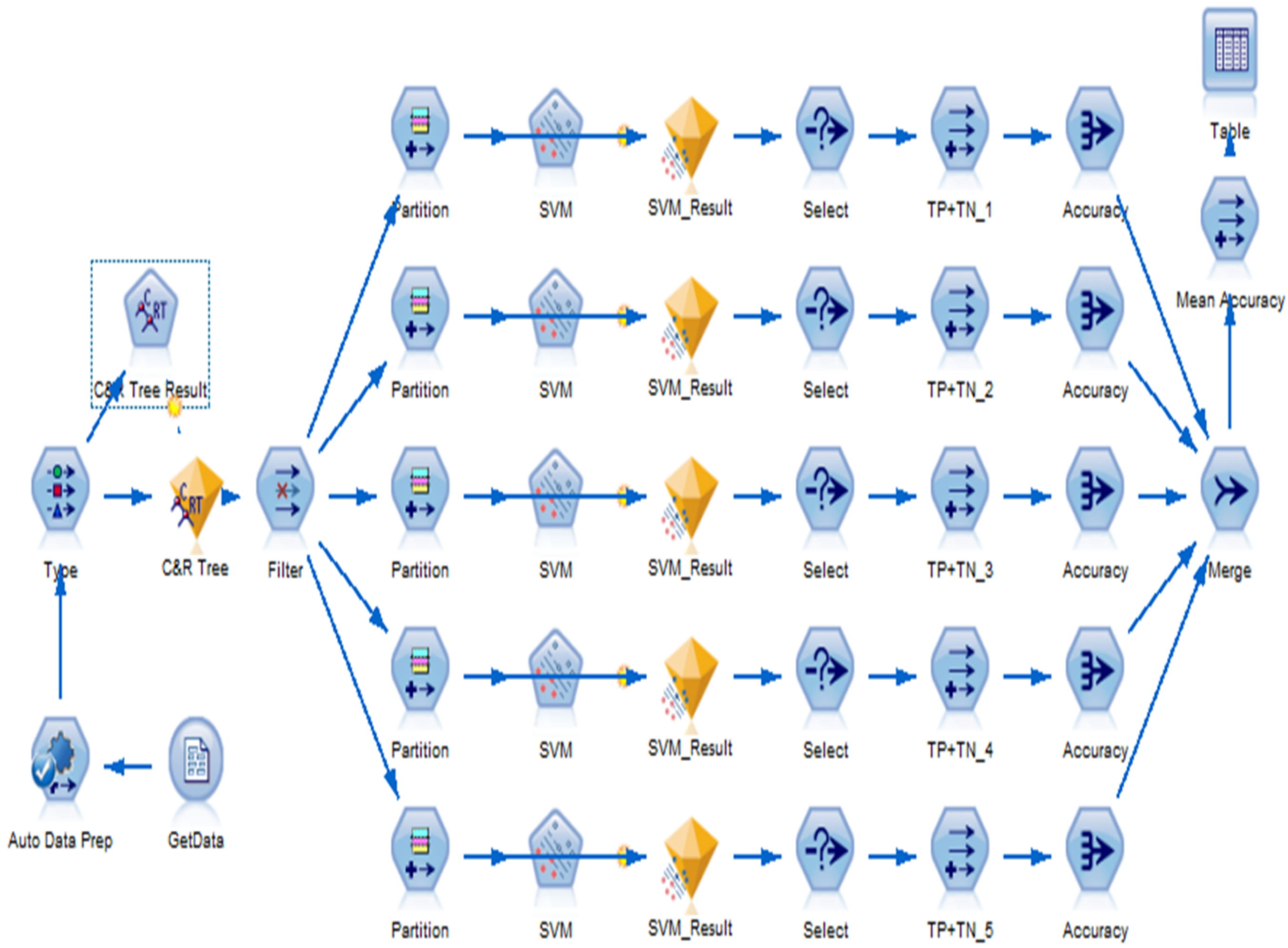


Figure3. Bar Chart of Stacking Results

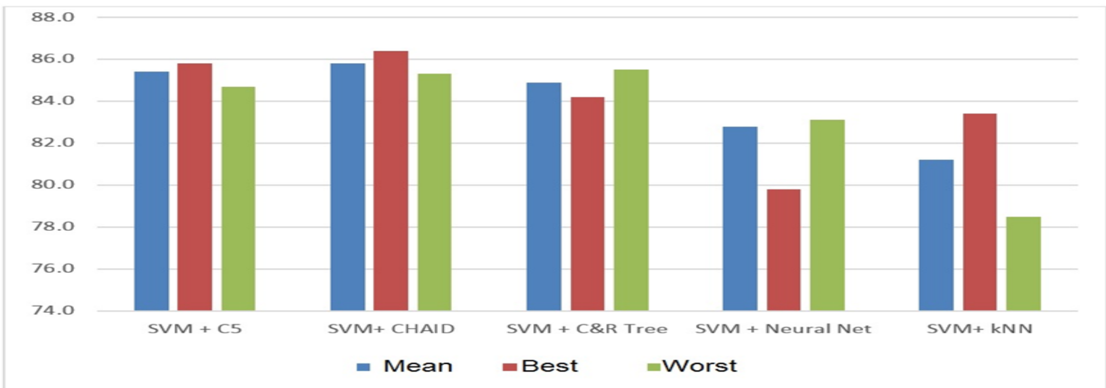
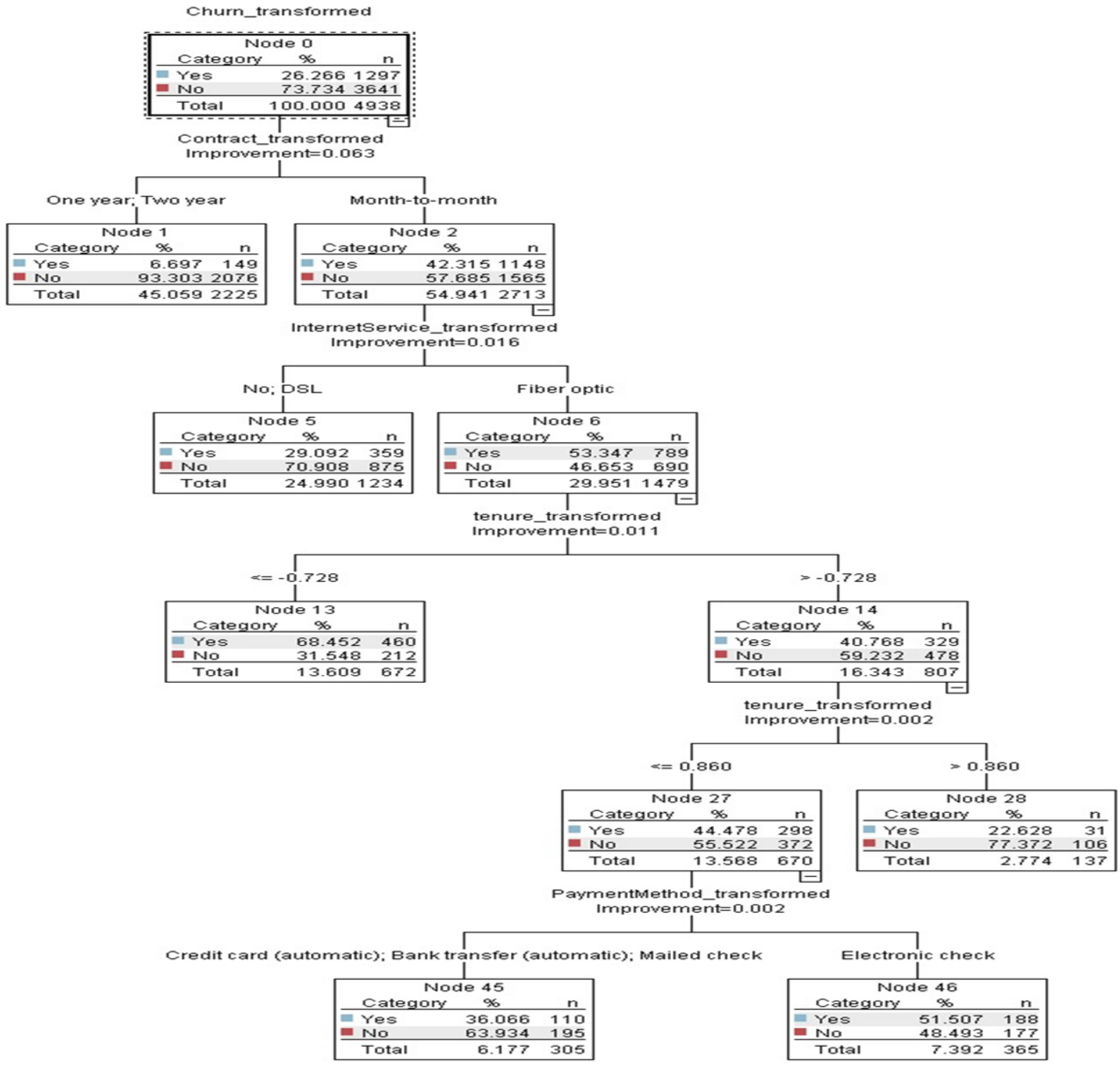


Figure4. Result of implementing of the CHAID method IN further evaluation


Churn_transformed
Node 0
Category % n
Yes 26.266 1297
No 73.734 3641
Total 100.000 4938
Contract_transformed
Improvement=0.063
One year; Two year
Month-to-month
Node 1
Category % n
Yes 6.697 149
No 93.303 2076
Total 45.059 2225
Node 2
Category % n
Yes 42.315 1148
No 57.685 1565
Total 54.941 2713
InternetService_transformed
Improvement=0.016
No; DSL
Fiber optic
Node 5
Category % n
Yes 29.092 359
No 70.908 875
Total 24.990 1234
Node 6
Category % n
Yes 53.347 789
No 46.653 690
Total 29.951 1479
tenure_transformed
Improvement=0.011
<= -0.728
> -0.728
Node 13
Category % n
Yes 68.452 460
No 31.548 212
Total 13.609 672
Node 14
Category % n
Yes 40.768 329
No 59.232 478
Total 16.343 807
tenure_transformed
Improvement=0.002
<= 0.860
> 0.860
Node 27
Category % n
Yes 44.478 298
No 55.522 372
Total 13.568 670
Node 28
Category % n
Yes 22.628 31
No 77.372 106
Total 2.774 137
PaymentMethod_transformed
Improvement=0.002
Credit card (automatic); Bank transfer (automatic); Mailed check
Electronic check
Node 45
Category % n
Yes 36.066 110
No 63.934 195
Total 6.177 305
Node 46
Category % n
Yes 51.507 188
No 48.493 177
Total 7.392 365