Customer churn prediction by hybrid neural networks

Chih-Fong Tsai a,*, Yu-Hsin Lu b

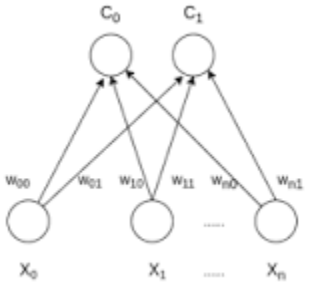
Introduction:

Churn prediction and management is very important for enterprises in the competitive market to predict possible churners and take proactive actions to retain valuable customers and profit. Therefore, to build an effective customer churn prediction model, which provides a result of accuracy, has become a research problem in recent years.

Description and related work done :

This paper spoke about two different hybrid data mining techniques by neural networks to examine their performances for telecom churn prediction. In particular, back-propagation artificial neural networks (ANN) and self-organising maps (SOM) are considered. Consequently, ANN + ANN and SOM + ANN hybrid models are developed,here the first component of the hybrid models aims at filtering out bad data or outliers. Then, the representative data as the outputs are used to create the prediction model. To make the result more accurate,this paper also includes two kinds of fuzzy testing sets based on the outliers identified by ANN and SOM, respectively, i.e. the first technique of the hybrid models. The experimental results indicate that the hybrid models outperform the single neural network baseline model in terms of prediction accuracy and the Type I and II errors. In particular, the ANN + ANN hybrid model performs the best. However, when the two fuzzy testing sets are used, the SOM + ANN hybrid model does not perform better than the baseline ANN model.

What does a self-organising map do?



SOM is used for clustering and mapping (or dimensionality reduction) techniques to map multidimensional data onto lower-dimensional data which allows people to reduce complex problems for easy interpretation. SOM has two layers, one is the Input layer and the other one is the Output layer

Data mining techniques:

- In order to establish an effective and accurate customer-churn prediction model, many data mining methods have been recently considered .
- The two primary goals of data mining in practice tend to be description and prediction. Description focuses on finding human-interpretable patterns describing the data. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest

Artificial neural network:

- Artificial neural networks (ANN) attempt to simulate biological neural systems which learn by changing the strength of the synaptic connection between neurons upon repeated stimulations by the same impulse.
- Neural networks can be distinguished into single-layer perceptrons and multilayer perceptrons (MLP). The multilayer perception consists of multiple layers of simple, two taste, sigmoid processing nodes or neurons that interact by using weighted connections.
- In addition, the neural network contains one or more several intermediary layers between the input and output layers. Such intermediary layers are called hidden layers and nodes embedded in these layers are called hidden nodes.

  Research methodology:
  Dataset
  Model development
  - The baseline
  - ANN+ANN
  - SOM +ANN

BASELINE:

- We use the original dataset to train a MLP neural network as the baseline ANN model for comparisons, which is similar to related work.
- In addition, four different learning epochs (50, 100, 200, and 300) and five different hidden layer nodes (8, 12, 16, 24, and 32) are used in order to obtain the best ANN baseline model. Table 2 shows the settings of the learning epochs and numbers of hidden layer nodes. As a result, there are twenty different ANN models developed for comparisons

ANN+ANN:

The first hybrid model is based on cascading two ANN models, in which the first one performs the data reduction task and the second one for churn prediction. That is, the original training set is used to ‘test’ the fist created ANN model, which is based on the ‘best’ baseline model identified.

SOM+ANN:

A self-organising map (SOM), which is a clustering technique, is used for the data reduction task. Then, the clustering result is used to train the second model based on ANN.

CONCLUSION:

Therefore, we can conclude that the hybrid model by combining two ANN techniques can perform better than the baseline model and the hybrid model by combining SOM and ANN. In addition, the ANN + ANN hybrid model performs more stably than the other two models.

Customer churn prediction a machine learning approach

Approach:

- With the advancement in the field of machine learning and artificial intelligence, the possibilities to predict customer churn has increased significantly.in this paper it was divided into 6 phases.
- 2 phases - data preprocessing and feature analysis
- 3rd - feature selection (gravitational search algorithm)
- 4-the data has been split into two parts: train and test set in the ratio of 80% and 20% respectively. In the prediction process, most popular predictive models have been applied, namely, logistic regression, naive bayes, support vector machine, random forest, decision trees, etc. on train sets as well as boosting and ensemble techniques are applied to see the effect on accuracy of models.

Introduction :

Inorder to maintain customers:

Among all the strategies, retention of existing customers is least expensive as compared to others. In order to adopt the third strategy, companies have to reduce the potential customer churn i.e., customer movement from one service provider to another. The main reason for churn is the dissatisfaction of the consumer service and support system.

The key to unlock solutions to this problem is by forecasting the customers which are at risk of churning

The customer churn models aim to identify early churn signals and try to predict the customers that leave voluntarily. Thus many companies have realised that their existing database is one of their most valuable assets and according to Abbasdimehr, churn prediction is a useful tool to predict customers at risk.

Description:

In this work, to tackle this problem we have used the following Machine Learning techniques: (1). Logistic Regression, (2) Naive Bayes, (3) Support Vector Machine, 123 Customer churn prediction system: a machine learning approach 273 (4) Decision Trees, (5) Random Forest Classifier, (6) Extra Tree Classifier and Boosting Algorithm such as Ada Boost, XGBoost & CatBoost. Furthermore, for better understanding of the data, the data have been pre-processed and important feature vectors have been extracted using gravitational search algorithm (GSA).

To use suitable Machine learning methods, the linearity of the data has also been checked and analysed

Advantage of proposed technique over the existing:

- We have applied a gravitational search algorithm to perform feature selection and to reduce the dimensions of the data-set,due to which prediction accuracy has increased.
- After preprocessing of data, we have applied some of the famous machine learning techniques which are used for predictions to prevent overfitting.
- Then we have used the power of ensemble learning in order to optimise algorithms and achieve better results, the obtained accuracy was high
- Then we have evaluated the algorithms on the test set using a confusion matrix and AUC curve, where obtained results are properly evaluated.

Few algorithms and analysis used in the paper such as

- Gravitational search algorithm
- Exploratory data analysis (EDA)

Machine learning models used and explained:

- Regression analysis-logistic regression analysis
- Naïve Bayes
- Support vector machine
- Decision trees
- Random forest classifier
- Extra tree classifier(ensemble learning approach)
- Boosting algorithm: adaboost
- XGBoost classifier
- CatBoost classifier

System architecture:

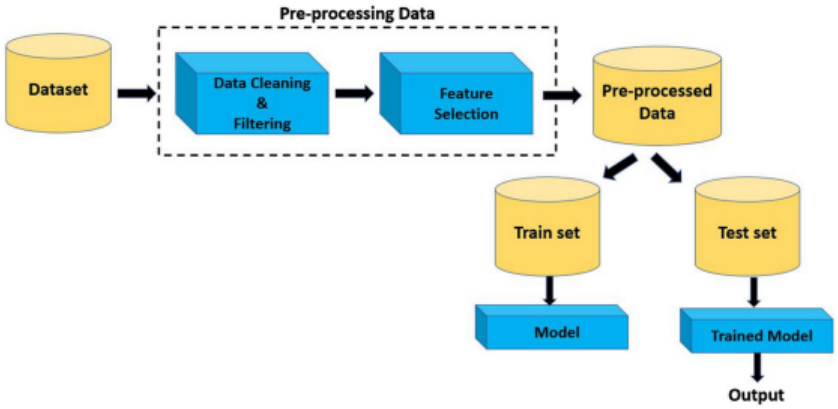


Fig. 3 System Architecture

Description of proposed model:

Phase 1: Identification of most suitable data (variance analysis, correlation matrix, outliers removal, etc).
Phase 2: Cleaning & Filtering (handling null and missing values)
Phase 3: Feature Selection (using GSA).
Phase 4: Development of predictive models (Logistic Regression, SVM, Naive Bayes, etc.).
Phase 5: Cross validation (using k-fold cross validation).
Phase 6:evaluation of predictive models(confusion matrix and auc curve)
Error rate calculated as:
Error Rate = (F P + F N)/(T P + T N + F P + F N).

Confusion matrix performance indicators:

Recall
Accuracy
F-measure
Precision

Conclusion:

- Through this research paper we provide a comparative study of Customer Churn prediction in Telecommunication Industry using famous machine learning techniques
- The experimental results show that two ensemble learning techniques that are Adaboost classifier and XGBoost classifier gives maximum accuracy with respect to others with an AUC score of 84% for the churn prediction problem with respect to other models. They outperformed other algorithms in terms of all the performance measures such as accuracy, precision, F-measure, recall and AUC score.

Results obtained:

Finally, the obtained results on the test set have been evaluated using confusion matrix and AUC curve. It was found that Adaboost and XGboost Classifier gives the highest accuracy of 81.71% and 80.8% respectively. The highest AUC score of 84%, is achieved by both Adaboost and XGBoost Classifiers which outperforms over others

Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behaviour

PURPOSE:

This paper aims to provide a predictive framework of customer churn through six stages for accurate prediction and preventing customer churn in the field of business.

DESIGN/METHODOLOGY:

6 STAGES:

- collection of customer behavioural data and preparation of the data;
- the formation of derived variables and selection of influential variables, using a method of discriminant analysis;
- selection of training and testing data and reviewing their proportion;
- the development of prediction models using simple, bagging and boosting versions of supervised machine learning
- comparison of churn prediction models based on different versions of machine-learning methods and selected variables;
- providing appropriate strategies based on the proposed model.

FINDINGS:

According to the results, five variables, the number of items, reception of returned items, the discount, the distribution time and the prize beside the recency, frequency and monetary (RFM) variables (RFMITSDP), were chosen as the best predictor variables.

LIMITATIONS:

The research data were limited to only one grocery store whereby it may not be applicable to other industries; therefore, generalising the results to other business centres should be used with caution

PRACTICAL IMPLICATIONS:

- Business owners must try to enforce a clear rule to provide a prize for a certain number of purchased items. Of course, the prize can be something other than the purchased item.
- Store owners must consider a discount for a certain amount of purchase from the store.
- They have to use an exponential rule to increase the discount when the amount of purchase is increased to encourage customers for more purchase.
- The managers of large stores must try to quickly deliver the ordered items

STUDY:

Another innovation of the current study is the comparison of machine-learning methods with their boosting and bagging versions, especially considering the fact that previous studies do not consider the bagging method. The other reason for the study is the conflicting results regarding the superiority of machine-learning methods in a more accurate prediction of customer behaviours, including churning. For example, some studies introduce ANN,SVM.

The current study identifies the best prediction method specifically in the field of store businesses for researchers and the owners. Moreover, another innovation of the current study is using discriminant analysis for selecting and filtering variables which are important and effective in predicting churners and non-churners, which is not used in previous studies. Therefore, the current study is unique considering the used variables, the method of comparing their accuracy and the method of selecting effective variables.

RESULT:

The proposed model with accuracy of 97.92 per cent, in comparison to RFM, had much better performance in churn prediction and among the supervised machine learning methods, artificial neural network (ANN) had the highest accuracy, and decision trees (DT) was the least accurate one.

The results show the substantially superiority of boosting versions in prediction compared with simple and bagging models.

A Survey on Customer Churn Prediction using Machine Learning Technique

Saran Kumar A

Introduction:

- A binary Classification task which differentiates churners from non-churners.
- Research says that acquiring new customers costs five to six times more than retaining existing ones.
- The system would also detect the Reasons behind the churn.
- Two type of approaches:
    - Reactive : provide Incentive when the customer requests to cancel their service relationship.
    - Proactive : Finding the customers who are likely to churn before they actually do, Thus providing them special incentives .
- Three types of Machines learning Techniques that can be equipped :
    - Supervised
    - Unsupervised
    - Semi-Unsupervised

Literature Survey:

- MODEL #1 :

  Drawback of general SVM – It does not reveal knowledge gained during training in human understandable form .

  Hence a Hybrid approach was taken to overcome the issue which Constituted of three phases :

    - SVM – RFE (Recursive feature elimination) used for reducing the feature set .
    - Dataset with reduced features are then used to obtain SVM models and support vectors are extracted .
    - Decision tree with Naive Bayes Classifier .
- MODEL #2 :

  Another Hybrid neural networks technique contained two phases :
    - ANN for performing data reduction and omitting the unrepresentative data .
    - Self-organising maps with ANN is done on the reduced data .
    - Result after testing the model : **Hybrid Model definitely out forms the single classifier**.

- MODEL #3 :
    - Boosting Algorithms to separate data into two clusters based on weight assigned by boosting algorithm .
    - Logistic Regression is used as a basis learner and a churn prediction model is created for each cluster .
    -
- MODEL #4 :
    - Imbalance in characteristics of the customer data is dealt with by random sampling method thus improving the SVM model.
    - The Imbalance above rises due to the low proportion of churners.
- MODEL #5 :
    - A CUSUM (Cumulative SUM) chart was used to monitor individual customer's Inter Arrival Time and used a Bayesian Model to capture the heterogeneity of customers .
- MODEL #6 :
    - Here , An Ensemble classifier is used composed of Rotation forest and Rotboost as modelling techniques for customer churn prediction .
    - Rotboost is the combination of Rotation forest and Adaboost and rotation Forests are used for feature extraction inorder to turn the input data for training base classifiers .
- MODEL #7 :
    - two Genetic Algorithms based on neural network models to predict customer churn .
    - First one used cross entropy based criteria to predict churn rate .
    - Second Algorithm directly increases the prediction accuracy of churn.

Optimal Customer Churn Prediction System Using Boosted Support Vector Machine

Saran Kumar, S.Viswanandhne,S.Balakrishnan

Abstract :

- The paper talks about boosting of SVM which improves the precision of produced rules from SVM . Since Supported adaptations have high precision and execution than non-helped forms.

Introduction :

- In this work, a nitty gritty plan is worked out to change over crude client information into helpful information which coordinates the displaying of purchasing conduct and thusly to change over this important information into learning , prescient information mining strategies are embraced.
- SVM first ventures the information into a higher dimensional element space and tries to locate the straight edge in the new component space

Research Methodology :

- Data Processing :
    - There are three key phases of this process
        - Training Set Extraction
        - Feature Attribute determination
        - Filtering strategies
    - Every diminishment framework is subsequently partitioned into two sections: the preparation dataset and the testing dataset. Each of the preparation dataset utilises the related input highlights and falls into two classes: typical (+1) and strange (−1).
    - 3 steps of training set feature selection (for finding the optimal number of features)
        - The “Training feature selection is to select n (a preset large number) sequential features from the input X”. This leads to n sequential feature sets F1,F2,.....Fn.
        - The “n sequential feature sets F1, …, Fk, …, Fn,($1 \leq k \leq n$) to find the range of k, called Ω, within which the respective (cross-validation

classification) error ek is consistently small (i.e., has both small mean and small variance).
        - Within Ω, find the smallest classification error ek = min ek. The optimal size of the candidate feature set, n*, is chosen as the smallest k that corresponds to e*” .
    - SVM with Adaboost Classifier Construction .
        - The main idea is to expel the undesirable highlights .
        - the positioning score is given by the parts of the weight vector w of the SVM .
        - where yk is a class label of sample xk and the summation is taken over all the samples. ak is the lagrange multipliers involved in maximising the margin of separation of the classes . AdaBoost is adaptive in nature ie. In each round m=1,...,M .

Experimental Results :

- The Accuracy comparison for SVM and Boosted SVM is :

Conclusion :

Combining SVM with Adaboost solves the problem of high dimensionality for classification of churn prediction.

Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models

SCOTT A. NESLIN, SUNIL GUPTA, WAGNER KAMAKURA, JUNXIANG LU, CHARLOTTE H. MASON

Overview:

This paper is an analysis on how methodologies contribute to the accuracy of customer churn predictive models. The data is based on a tournament where various academics downloaded data and built various models to predict customer churning.

The main questions being asked in this paper is:

- Does method make a difference? Are differences in predictive accuracy across various techniques managerially meaningful?
- Do models have staying power? Can a model estimated at time t predict customer churn at time t + x, where x is some later time period?
- Which methods work best? How do the various statistical techniques, variable selection approaches, and time allocation strategies contribute to predictive accuracy? What overall approaches are likely to be successful?

The results from the tournament were determined by Lift current, Lift future, Gini current, Gini future.The tournament indicated that logistic regression (used by 45%) and decision trees (23%) were the most common estimation techniques, but neural nets (11%), discriminant analysis (9%), cluster analysis (7%), and Bayes (5%) were used as well. Most participants (88%) explored more than one estimation technique.

The results from the paper suggest the following:

- Logit and tree approaches are positively associated with predictive performance. Because the factors are orthogonal, they are two independent approaches, and both tend to do well.

- The practical approach is the “middle-of-the-road.” The coefficient for this variable is often not significantly different from zero at conventional levels, but the trend is clear: Participants who score high on this factor tend not to do as well as the logit and tree modellers but do better than the discriminant and explain modellers.

- The discriminant and explain approaches do not do as well. Often, the coefficients are not significantly different from zero, but the signs are consistently negative

Table 2
FACTOR LOADINGS

		Logit	*Trees*	*Practical*	*Discriminant*	*Explain*
Estimation	Logit	.695	−.409	.207	−.378	.089
	Neural	−.642	−.072	−.103	−.181	−.196
	Tree	−.020	.698	−.164	−.045	−.061
	Discriminant	−.116	−.186	−.19	.872	.072
Variable selection	EDA	.375	−.610	.153	−.409	.028
	Theory	.086	−.113	.178	.050	.797
	Sense	−.015	−.132	.633	.255	.152
	Stepwise	.930	−.506	−.003	−.065	.560
	Factor	−.292	.057	.017	−.214	.787
	Cluster	.214	−.156	−.068	.519	.658
Relative time	Downloading	.059	−.068	.811	−.060	.145
	Data cleaning	.127	−.387	−.436	−.477	−.087
	Creating variables	−.221	−.679	−.061	.119	−.006
	Estimation	.268	.786	.233	.288	.004
	Preparing prediction files	−.748	−.183	.239	.047	.180
Total time	Total	−.162	.513	−.675	−.122	.017
Subdivde	Subdivide	−.036	−.136	−.423	.085	−.028
Vars	Number of variables	.012	.460	.085	.689	−.336
Exploration	Number of techniques explored	.198	.037	.194	.002	.836
Practitioner	Practitioner respondent	.657	.344	.355	−.184	−.049

Notes: The five factors accounted for 66.0% of the variance in the above 20 items.

Conclusion:

The results suggest several important findings.

First, methods do matter. The differences observed in predictive accuracy across submissions could change the profitability of a churn management campaign by hundreds of thousands of dollars.

Second, models have staying power. They suffer very little decrease in performance if they are used to predict churn for a database compiled three months after the calibration data.

Third, researchers use a variety of modelling “approaches,” characterised by variables such as estimation technique, variable selection procedure, number of variables included, and time allocated to steps in the model building process.

Customer churn prediction using improved balanced random forests:

YAYA XIE , XIU LI, E.W.T. NGAI , WEIYUN YING

Overview:

This paper proposes a novel learning method called improved balanced random forests (IBRF) and demonstrates its application to churn prediction.It investigates the effectiveness of the standard random forests approach in predicting customer churn, while also integrating sampling techniques and cost-sensitive learning into the approach to achieve a better performance than most existing algorithms.

The proposed method incorporates both sampling techniques and cost-sensitive learning, which are two common approaches to tackle the problem of imbalanced data. By introducing ‘‘interval variables”, these two approaches alter the class distribution and put heavier penalties on misclassification of the minority class.The interval variables determine the distribution of samples in different iterations to maintain the randomicity of the sample selection, which results in a higher noise tolerance. It allows ineffective and unstable weak classifiers to learn based on both an appropriate discriminant measure and a more balanced dataset. It therefore can achieve a more precise prediction.

Improved Balanced Random Forests (IBRF) :

The paper proposes IBRF by combining balanced random forests and weighted random forests.To combine these two methods, it introduce two ‘‘interval variables” m and d, where m is the middle point of an interval

and d is the length of the interval. A distribution variable a is randomly generated between m - d = 2 and m + d = 2, which directly determines the distribution of samples from different classes for one iteration. The main reason for introducing these variables is to maintain the random distribution of different classes for each iteration, which results in higher noise tolerance. By contrast, balanced random forests draw the same number of samples from both the majority and minority class so that the classes are represented
equally in each tree.

The algorithm takes as input a training set D = { (X1,Y1) , ……. (Xn,Yn)}, where Xi, i = 1 ... n is a vector of descriptors and Yi is the corresponding class label. The training set is then split into two subsets D+ and D-, the first of which consists of all positive training samples and the second of all negative samples.

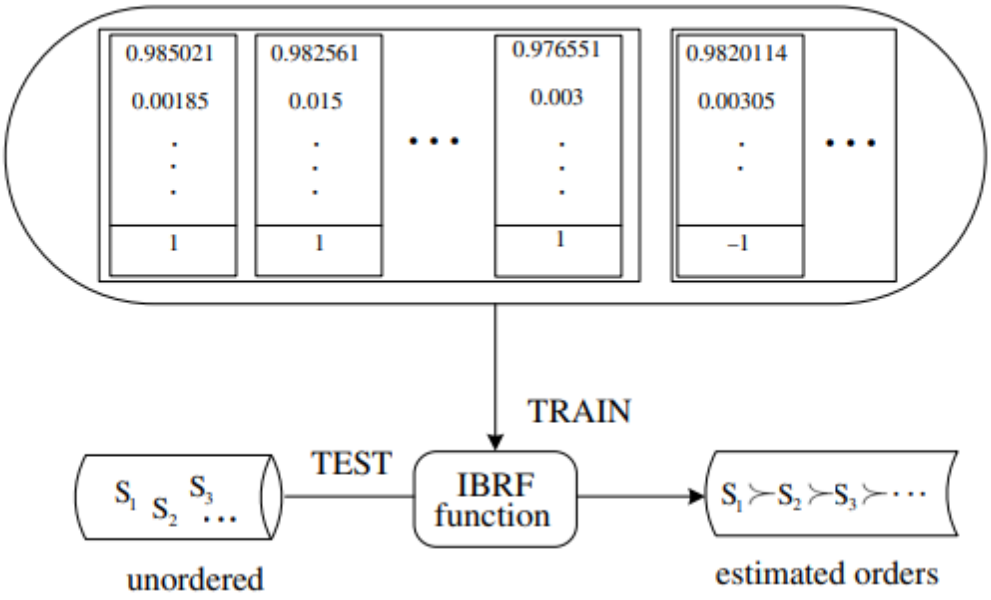


Fig. 1. Framework of IBRF.eps.

The framework of IBRF is shown in Fig. 1. Let Si denote the testing sample. The training inputs of the ranking problem are samples from D+ and D- provided with the information that the negative samples should be ranked higher than positive ones. The samples which are most prone to churn are ranked higher in output.

Findings :

To evaluate the performance of the proposed method, IBRF,the author applies it to a real-world database. A major Chinese bank provided the database for this study. The data set, as extracted from the bank's data warehouse, included records of more than 20,000 customers described by 27 variables. A total of three major descriptor categories are explored which are personal demographics, account level, and customer behavior.They are identified as follows:

- Personal demographics is the geographic and population data of a given customer or, more generally, information about a group living in a particular area.
- Account level is the billing system including contract charges, sales charges, mortality, and expense risk charges.
- Customer behaviour is any behaviour related to a customer's bank account.

A comparison of results from IBRF and other standard methods, namely artificial neural network (ANN), decision tree (DT), and CWC-SVM (Scholkopf, Platt, Shawe, Smola, & Williamson, 1999), is shown in Fig 2.We can observe a significantly better performance for IBRF in Fig. 2. Hence we conclude that IBRF achieves better performance than other algorithms

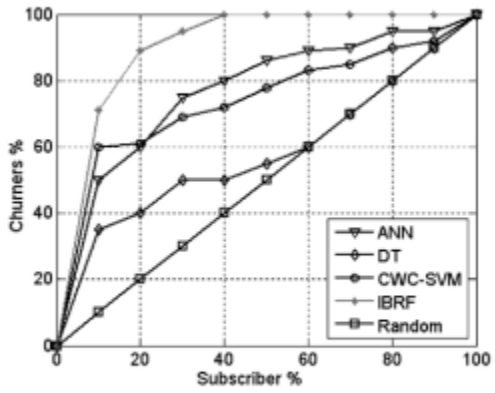


Fig. 2. Lift curve of different algorithms.eps.

Conclusion:

This paper proposes a novel method called IBRF to predict churn in the banking industry. IBRF has advantages in that it combines sampling techniques with cost-sensitive learning to both alter the class distribution and penalise more heavily misclassification of the minority class. The best features are iteratively learned by artificially making class priors equal, based on which best weak classifiers are derived**.**

Experimental results on bank databases have shown that our method produces higher accuracy than other random forests algorithms such as balanced random forests and weighted random forests. In addition, the top-decile lift of IBRF is better than that of ANN, DT, and CWC-SVM. IBRF offers great potential compared to traditional approaches due to its scalability, and faster training and running speeds.

Applying Data Mining to Customer Churn Prediction in an Internet Service Provider

AFAQ ALAM KHAN SANJAY JAMWAL M.M.SEPEHRI

Overview:

This paper talks about clustering users as per their usage features and incorporating that cluster membership information in classification models.

Literature used in the paper:

Customer attrition is an important issue for any company and is easiest to define in subscription based businesses, and partly for that reason, churn modelling is most popular in these businesses [5]. Long-distance companies, Mobile phone service providers, Insurance companies, Cable companies (Pay-TV) [6], financial service companies, Internet service providers, newspapers, magazines, and some retailers all share a subscription model where customers have a formal, contractual relationship which must be explicitly ended.

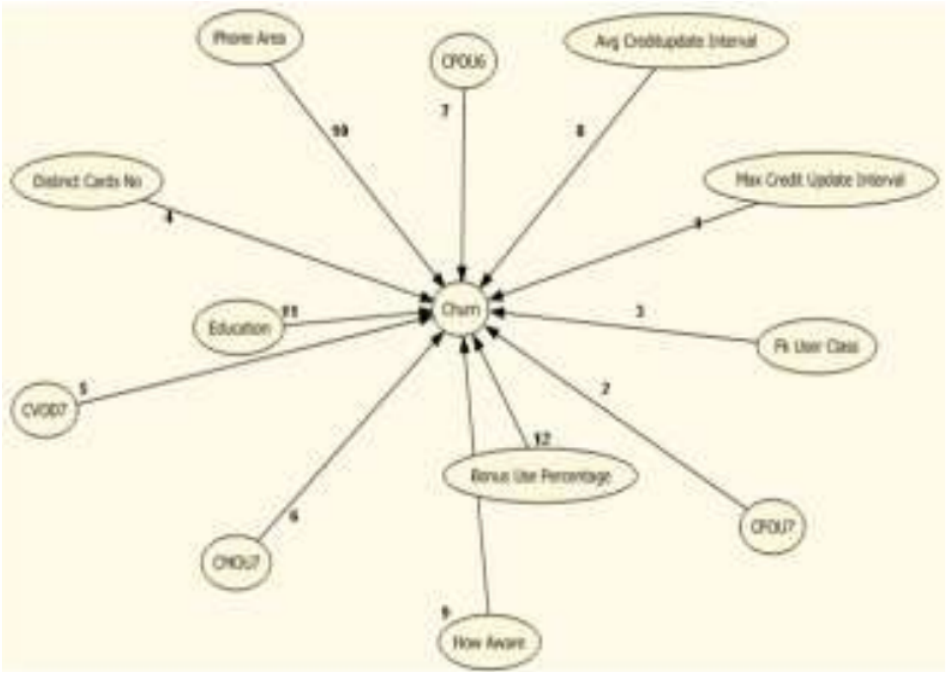
The main statement proposed in this paper:

Ways of integrating data mining into customer churning

Sectors targeted:

- Mobile service providers
- Banking and insurance
- ISP (outside iran)
- ISP (in iran)

The below diagram represents the findings in the paper. It shows all the attributes that affected churning.



Conclusion:

All the features used for the churn prediction were either demographic features or billing or usage features. Purpose of our second objective was to get an insight about the importance of these three types of features in the churn prediction.
We came to this conclusion that the demographic features have the lowest affect on the churn prediction. But it was not easy to come to the conclusion about the billing and usage features.