Research Paper

CHURN PREDICTION WITH ENSEMBLE CLASSIFIERS FOR TELECOM SECTORS

Faiza Hassan Mohamed[†], Mustafa Cem Kasapbaşı[††]
[†] Industrial Engineering Department, Istanbul Commerce University, Istanbul, Turkey
[††] Computer Engineering Department, Istanbul Commerce University, Istanbul, Turkey
faaizadiini93@gmail.com, mckasapbasi@ticaret.edu.tr
0000-0002-0632-7800, 0000-0001-6444-6659

Atıf/Citation: Mohamed F. H., Kasapbaşı M.,C. (2021). Churn Prediction with Ensemble Classifiers for Telecom Sectors. *Journal of Technology and Applied Sciences* 4(1), 57-71.

ABSTRACT

Churn Prediction has been implemented in the researches and published works using different advanced mechanisims including Machine Learning, Data Mining, and Hybrid mechanism. These mechanisms support big companies and small businesses to classify and predict churning customers to be able retaining them to stay with their company using their services. Also, helps top managers and decision makers to take reliable decisions and Customer Relation Management CRM department too. In this study, a telecom sector churn dataset named Orange is used for customer churn prediction. Ensemble classifiers are used AdaBoostM1, PCA, Gain Ratio, Info Gain, Bagging in combination with J4.8, Naïve Bayes, Logistic Regression, Random Forest, KNN, LMT (Logistic model Tree). Highest accuracy of %94 is obtained by combination of bagging and J4.8. The results are compared with other studies as well and this study performed as good as the surveyed literature and surpassed in same cases.

Keywords: Churn Prediction, KNN, LMT, CRM, PCA

Telekom Sektörleri için Topluluk Sınıflandırıcılarla Ayrılma Tahmini

ÖZET

Literatürde, Makine Öğrenimi, Veri Madenciliği ve Hibrit teknikleri gibi farklı teknikler kullanılarak Ayrılma/Çalkalanma Tahmini gerçekleştirilmiştir. Bu teknikler, şirketleri ve işletmeleri, hizmetlerini kullanarak şirketlerinde kalabilmeleri için müşterileri belirleme ve tahmin etme ayrıca ayrılan müşteri konusunda destekler. Üst düzey yöneticilerin ve karar vericilerin güvenilir kararlar almasına ve Müşteri İlişkileri Yönetimi (CRM) departmanına da yardımcı olur. Bu çalışmada, müşteri kaybını tahmin etmek için Orange adlı bir telekom sektörü ayrılan müşteri veri seti kullanılmıştır. Topluluk sınıflandırıcıları AdaBoostM1, PCA, InfoGain, Gain Ratio, Bagging ile birlikte J4.8, Naive Bayes, Lojistik Regresyon, Rastgele Orman, KNN, LMT (Lojistik model Ağacı) sınıflandırıcıları kombinasyonları ile birlikte kullanılır. Torbalama ve J4.8 kombinasyonu ile en yüksek % 94 doğruluk elde edilir. Sonuçlar diğer çalışmalarla da karşılaştırılmış ve bu çalışma araştırılan literatür kadar iyi performans göstermiş ve bazı vakalarda daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: Ayrılama Tahmini, kNN, LMT, CRM, PCA

Geliş/Received : 6.01.2021
Gözden Geçirme/Revised : 25.01.2021
Kabul/Accepted : 26.01.2021

1. INTRODUCTION

Customer Churn Prediction is a prediction to know the customer who ceases the subscription of the company. Digital companies which provide services through online and offline are mostly facing this risk case of losing a subscribed customers who subscribed their services, such as , Telecommunication sector, Banking sector, Technology and ICT sector. The problem of churn are increasingly boosting machine learning and data mining research areas, because they are the most applicable research fields and also provide most suitable solutions using developed tools and techniques to solve the churn problem. Data mining techniques allow companies to learn more about their customers, using customer information which collected from customer transactions and stored in the company database, that collected data from the customer and stored in the database, allow company to learn more about customers, their interests, and know more about their situation, whether they are satisfied and liked the services of the company they subscribed to use it or not, so as the world is moved digitally and online today, it is hotline research topic and tool, which most of the researchers are working on, and most of digital companies in the world are interesting and using it, as it is helping to get accurate and reliable customer data, which help them to take accurate and best decision through enhancing, developing, and continuing their businesses. Churn prediction which is knowing how many customers are left or unsubscribed the company, and also prediction the customers who are likely to churn, is one of hot topics in Data mining and Machine learning areas of today's researches.

Churn Prediction has been implemented in the researches and published works using different advanced mechanisims including Machine Learning, Data Mining, and Hybrid mechanism. These mechanisms support companies and businesses to classify and predict churning customers to be able retaining them to stay with their company using their services. Also, helps top managers and decision makers to take reliable decisions and Customer Relation Management CRM department too.

In general, the information which used to build models for churn prediction in mobile telecommunication companies includes customer demographics like age, gender, contractual data, tenure, call details, complaint data, billing information (Killer and Kotler,2016; Huang et al,2012).

2. LITERATURE REVIEW

Telecom sector is a sector which customer play an important role for existing their work and revenue too, so the customer care and satisfaction must be the biggest goals and plans of the company to retain them and stay their customers, otherwise customers will churn and go to use other competitive services. Churn prediction is a problem most of telecom companies facing today, when the customer churn the company faces huge financial loss as losing the customer loses the profit too, because when your customer leave you and go to another competitor of your company your company faces trouble of losing both revenue and customer too. For that reason, a lot of studies being published to solve this problem which more telecom companies nowadays facing, and researchers working on providing technologies and designing models to predict churners and then to prevent and retain churners for providing them needed services and solving their problems. Andreea Dumitrache and Monic Mihaela Mear Matei done study on prediction customers who are going to defect in a Romanian mobile telecommunication companies. The churn analysis developed for post-paid customers. Study used in logistic regression to predict churn and a solution based on smooth bootstrap technique to correct for drawbacks of imbalanced classes. In the study it analyzed churn behavior on a sample of 10701 subscribers randomly selected from a database of a large telecommunication companies operating on the Romanian market.

Businesses investing in marketing high budget and create marketing campaigns to attract customers, but, the thing that al the businesses realized and agreed is losing customer costs more than it costs attracting new customer, it is very easy to attract a new customer to your company and subscribe your service, but to experience losing an existing customer is real lose which you losing the profit you were getting from that customer and also the customer itself which was a number in your subscribers. So, that need forced decision makers and business analysts to find creative way to find or know the customers who are likely to churn, to analyse the causes, after that, the retaining process can be easily done, before the losing process happen.

Irfanullah et al (Irfanullah et al. 2019), published a study in churn prediction model which they used in the study they have done Random Forest, in their study, analysed the machine learning techniques and factor classification in telecommunication sector. The study, has used clustering technique which allowed to know easily the customers who are likely to churn, or churning customers, also, the caused reasoned for their churn, or to decide to unsubscribe the service of their subscribed telecom company. İn this study, the have implemented feature selection by using with correlation attribute ranking filter and information gain. The implemented model classified first the customer data using classification algorithms, where random forest achieved better and resulted with 88.63% correctly classified instances. Assigning retention procedures is the most significant job of Customer Relationship Management CRM to stop churners and customers who are likely to churn their company subscription. This study, has specified churn elements that are serious in defining the original causes source and

the reason of churn. Results found out after this model has implemented using random forest classifier produced better churn classification.

Data mining is a new technology procedure ,which enables discovering unexpected data samples, to assist in the prediction of the upcoming directions. Recent times, this technology has been implemented in most of the world wide businesses, as businesses moved to digital, it is needed to apply the new technologies like data mining to solve business problems in general and develop business industries which can not get away from using this technique, it is used for both, to solve businesses problems, and at the same time, to enhance and develop the services provided by the businesses to their end line customers to satisfy them. Decision Tree, is straightforward tool of data mining process, which used for predictions and to find out upcoming expectations. Decision Tree has different algorithms for generating the decision tree, these algorithms include C4.5, ID3, and other algorithms, which has been used to implement with different software tools.

Nijahwan et al (Nijahwan et al. 2019), have done a study, which they concentrated on implementing data mining in telecommunication sector, to enable predicting churn attitude of subscribed customers. They have used in their published work a data which collected from surveymonkey.com for mining purpose. After that, the collected data has been cleaned and processed, after that process, decision treehas generated to predict the customers who are likely to churn . results found out after that, even showed some reasonable facts to point out, for example, if the customer who subscribed a specific telecom company and using the same number for two years and more duration time is less likely to churn and unsubscribe that telecom company to switch another company and subscribe new telecom company, that found out fact, strengthened the fact of retaining the existing customers and keeping them harder than attractin a new customer, so that found out data help telecom companies to work on how they treate well and satisfy those existing customers and not allow them to churn for providing their needed and interested services..

Mihrimah Ozmen (Ozmen et al. 2020), in the study they have presented the importance of customer management and how telecommunication companies struggling the competition between them every company is trying to learn more about their customers and be able to manage them to keep them from churning and switching to another one.

V. Umayaparvathi and K. Iyakutti (Umayaparvathi and Iyakutti. 2012), have worked and published a study on data mining techniques in telecommunication company's churn prediction. Telecommunication companies facing many challenges in the competitive market, where every company changes rapidly as new technologies effecting to rapid changes, so that, customers are continiously looking for their need provider company to swicht for a reason they finding whatever kind of service they needed, that switching stresses tlecom companies to compete each other for the fear of losing their existing customer, each company trying to offer new and advanced services to their customers to prevent them from churning and switching to another company, constantly working on what staisfying most their service subscribed customers. So that, the work these researchers have published , searched the application of data mining procedures to predict customers who are likely going to churn and effect of attribute selection on classifying the churn.

Adnan Idris et al (Idris et al. 2013), have intended a study on an intelligent churn prediction system for telecommunication sector by utilizing functional feature extraction technique and ensemble method. İn their study, they have used ensemble classifications with minmum redundency and maximum relevence mRMR, also, they have used, fisher's ratio and f-score methods to model the telecom churn prediction issue.

Utku Yabas et al (Yabas et al. 2012), have published a study on customer churn prediction for telecom services, they have tasked on data mining methods, for a purpose of to properly predict the customers who are likely to churn, while they have the tention of to unsubscribe their current subscribed telecom service provider and to move and subscribe another similar one but different in terms of services and customer needs offered by it. They implemented their research task by using Orange Telecom dataset , which is one of the available telecom datasets in the internet, and a lot of researchers have used and studies have been done using it in churn prediction research issues.

Azeem and Usman (Azeem and Usman. 2018), have published a study churn prediction problem in telecommunication sector, which they not only focused on modeling the churn prediction and customer churners, but, they also, have implemented tools for customer retention. As the existing literatures have limitations, and the churn prediction issue getting serious problem which effecting telecom companies both of losing customer and revenue too, they have tried and implemented new narrative model, to be exact able to obtain the purpose of correct classification and to obtain the goal of retaining the churners. They worked on both narratively, to find the problem and also to draw the solution which is how company can retain customers to not churn and stay to use the service.

Alae Chouiekh and El Hassan Ibn El Haj (Chouiekh and El Haj. 2020), have published work on churn prediction problem. İn their work, they have tried new mechine learning model, as they used a narrative mthod by implementing deep convolutional neural network, which they applied on the dataset they have implemented in their research as experiment to identfy customer churn. They have found outthat deep convolutional neural network achieved better results and performed better than other previously used machine learning algorithms.

Adnan Amin et al (Amin et al), have done a tudy on customer churn prediction, which they specially focused on finding out the real solution for churn prediction problem which most of telecom service providers suffering from, they found out , there must be clear detailed and prove reasons behind the customer churn. Knowing that, finding the unseen factors must be the priority and it helps more to find the causes of the problem to be able to solve it later on. They have worked on determining relevence and dataset samples to know the unseen factor.
Yasser Khan et al (Khan et al. 2019), have don customer churn prediction study, as churn prediction recent years has been serious problem which telecom companies facing as digital worl experiencing very fast growing digital and updates, studying customer behaviour through customer data is the most important tasks businesses working and focusing on it. The competition is higly increasing, so, in this study, the authors have used Artificial Neural Network ANN approach to predict the customers who are likely to churn from the company to move to another one.
Clement Kirui et al (Kirui. 2013), have published a study on customer churn prediction in mobile telephony industry which they have used probabilistic classification in data mining. İn their study, they aimed to enhance the ability of telecom companies to know customers who are likely to churn, and to get that work done, they have used in their research customer transactions recorded to have data to have detail customer information which enables them to know better and identify best results whether there is a churn or not. They have exmained the new set of features of the customer data by using Naive Bayes and Bayesian Network probabilistic data mining algorithms, and then they have compared the found out results to the results that gained from C4.5 and Decision Tree classification algorithms.
J. Vijaya and E. Sivasankar (Vijaya and Sivansankar. 2018), have published research article on computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecom sector. They have intended a methodology using rough set theory to classify functional characteristics for telecommunication churn prediction.
Zhong and Li (Zhong and Li. 2019), have published research paper on churn prediction, by using authentic customer customer call data, they have evolved the convolutional neural network predictive model in their research study to classify telecom churn prediction problem.
Ruiyun Yu et al (Yu et al. 2018), have published research study, which they intended a particle classification optimization based BP network for telecommunication customer churn prediction algorithm.
Awodele Oludele et al (Oludele. 2020), have published study in enhanced churn prediction in telecommunication industry. They have used in their research, a Mrkov Chain Model, to sample the customer churn prediction. The Markov Chain Model gives more adaptability than most other possible models, and can easily incorporate variables which most of other models can not easily do.
Hossam Faris (Faris. 2018), have published researchy study in churn prediction. As telecommunication service provider companies been facing customer churn problem, study analyzed and found out that import of churn prediction problem solution is knowing the customers who likely churning before they took the action and switch to another company, and doing that strategy it is needed powerful prediction model, so in this study, researchers have intended an intelligent hybrid model which it based on particle sworm optimization and feedforward neural network for churn prediction.
Recent years, more research studies have been done for customer churn prediction problem using ensemble classifiers, also, dofferent models have propsed, to predict customer churn prediction problem, but, the issue still has been developing and researchers working on it, more studies been innovating new techniques and models for solving the problem, and some greate insights found out, genetic programming approach is one of proposed models to solve churn prediction problem.

3. MATERIAL AND METHODOLOGIES

Every research must have methodology to follow, and to answer how the research data collected and analyzed, we will discuss under this topic the data that we have collected to use on our research, and also the methodology that we have proceeded to get accurate results and solve the telecom churn prediction which was our research question.

3.1 Data Description

As we have been working on telecom churn prediction problem, we did not find any data provided by any company, then we have tried to find the available data sets on research platforms and on the internet. Fortunately, we have found Orange telecom data set which have been used for some previous researches done on the field.
In our research, we have used the Orange Telecom DataSet which we have found from Kaggle platform, and we tested and trained the data by using Weka which is a program for analyzing and interoperating the results of research data. We have trained and tested our dataset number of methodologies that we have applied and got

different results. We will discuss the methodologies that we have applied on our research, one by one with deep details and explanations with the visual results that we have got during our research.

3.2 Methodologies

3.2.1 Naïve Bayes

Naïve Bayes is Bayes Theorem based classification method. This method uses independent assumption, which it is getting the results weather you have some missing values or lost some data, it is not making values and some parts of that depend on others, which will make the progress complicated and not proceed, instead it would process the available assumptions independently to produce accurate results. If some features missed or unknown, so other features could progress the process they not be depend on others to get or to know, thee existing could contribute enough the probability. Naïve Bayes method is easy to use also, and very useful for large data sets, large sized data sets and complex prediction problems use this method for its ability of easy to use. Bayes theorem provides a kind of calculating posterior probability, which is well known and useful way to calculate and to do probabilities and predictions. Naïve Bayes performs well in multi class prediction, as we mentioned above because it is useful for processing large sets of data, it is because of the ability of multi class prediction it has. It is better than other models and methods in terms of performance as its independence assumption. There are four applications of Naïve algorithms which are, first one is real time prediction, the second one is multi class prediction, the third one is text classification or spam filtering also known as sentiment analysis, and the fourth and the last one is recommendation system. The detailed information about this method can be found two text books which are mentioned the refrences list, one is " Introduction to Data Mining" written by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar., and the other one is " Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann" written by H. Witte, E. Frank, M. Hill, and C. Pal.

3.2 J48

J48 classification algorith is one of machine learning algorithm. This classifier, is used for predictions and to solve classification related problems, it generates a decision tree to list down in the nodes of the tree the assumptions and drow out the possibilities. J48 classifier results better accurately comparing with other classification algorithms.

3.3 Random Forest

Random Forest is a machine learning algorithm, which used for classification. Forest means compressed, so, Random Forests works as to create more decision trees, compressed to find out the best possible solution. This classification algorithm is flexible which can do both of regression and classification.

Random forests has different applications include: recommendation engine, image classification and feature selection. It is highly accurate and robust method. It uses mean decrease impurity (MDI) for calculation of the important of each feature. It contains set of multiple trees. The detailed information about this method can be found two text books which are mentioned the refrences list, one is " Introduction to Data Mining" written by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar., and the other one is " Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann" written by H. Witte, E. Frank, M. Hill, and C. Pal.

3.4 k Nearest Neighbour (kNN)

K Nearest Neighbour is a machine learning algorithm, which can use for both classification and regression. KNN is an algorithm which classifies data point according to the similar relation of it. So many application tried with this method because of its effectiveness, non-parametric and easy to implementation properties. It's a classification which classifies instances based on their similarities. It is called in Weka IBK. The detailed information about this method can be found two text books which are mentioned the refrences list, one is " Introduction to Data Mining" written by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar., and the other one is " Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann" written by H. Witte, E. Frank, M. Hill, and C. Pal.

3.5 Logistics

Logistic Regression is a machine learning algorithm, which is used to allocate examinations to a seperate group of classes. İt works as binary classification model, which uses mostly to result two possible results, for example yes or no. This classification algorithm used by most of online transaction dealing companies, because of it is simplicity and easyness.

The detailed information about this method can be found two text books which are mentioned the refrences list, one is " Introduction to Data Mining" written by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar., and the

other one is " Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann" written by H. Witte, E. Frank, M. Hill, and C. Pal.

3.6 Decision Stump

Decision stump is a machine learning algorithm, which structurally contains one main node of decision tree. That main single node connects directly to the leaves which are the rest of the tree tsructure in the decision stump.

3.7 Logistic Model Tree LMT

Logistic Model Tree LMT is a machine learning algortihm, which is used for classification and to solve churn prediction problems. İt works together combined with logistic regression and decision tree.

3.8 AdaboostM1

AdaboostM1 classifier , is the first successful boosting sophisticated for binary classification. Adaboost used to improve and support the excution of any machine learning algorithm.

3.9 Principal Component Analysis (PCA)

Principal Component Analysis PCA classifier is a popular technique which used for today's research areas of pattern recognition and visual classifications. This statistic method allows for minimizing the distance of large sets of data and summerizes in to small visual patter, which can be easily shown visually. PCA is strong data analysis, as it has functional computational technque.

3.10 Gain Ratio

Gain Ratio is amendement of information gain which minimizes its bias. Gain Ratio improves the information gain by ensuring how much information needed to express which branch a sample belongs to.

3.11 Information Gain

Information Gain determines the amount of information about the class that an attribute can provide, so, the attribute which has highest information gain will split first. İnformation Gain is crucial solution that decision tree algorithm uses.

3.12 Bagging

Bagging is a machine learning algorithm that created to develop the settlement and precision of machine learning algorithms used in regression and classification.

4. EVALUATIONS AND DISCUSSION

In all evaluation process Weka Software is utilized for implementation. As validation, K-fold cross validation is used to obtain confusion matrix. With Confusion matrix many evaluation metric is calculated namely: True Positive Rate (TP Rate/Recall), False Positive Rate (FP Rate), Precision, F-1 Score (F Measure), MCC (Matthews correlation coefficient), ROC (Receiver Operator Characteristic) Area and PRC (Precision Recall Curve) Area. The formulations of TP Rate, FP Rate, Precision, F1 Score and MMC metrics are given from Eq. 1-5 respectively. In these Equations TP means True Positive, TN means True Negative, FP means False Negative and FN means False negative. For further description one may consult (Wikipedia, 2021)

$$TP\ Rate = \frac{TP}{TP+FN} \quad (1)$$

$$FP\ Rate = \frac{FP}{FP+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (4)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

4.1 Implementation Results Evaluation

Table 1 indicates the results of Naïve Bayes classifier in combination with AdaBoost M1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 1 for Naïve Bayes best result is obtained when Gain Ratio is used in conjunction with as indicated in the table.

Table 1. Naïve Bayes Classifier with ensemble methods

Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.85	0.534	0.84	0.844	0.342	0.75	0.851
PCA	0.859	0.849	0.879	0.795	0.095	0.651	0.818
GainRatio	0.882	0.459	0.872	0.875	0.471	0.793	0.879
InfoGain	0.882	0.459	0.872	0.875	0.471	0.793	0.879
Bagging	0.879	0.485	0.867	0.871	0.448	0.796	0.881

Table 2 indicates the results of J48 classifier in combination with AdaBoost M1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 2 for J48 best result is obtained when Bagging is used in conjunction with as indicated in the table.

Table 2. J48 Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.894	0.5	0.882	0.882	0.495	0.867	0.915
PCA	0.841	0.579	0.827	0.833	0.289	0.652	0.807
GainRatio	0.916	0.33	0.911	0.913	0.634	0.827	0.898
InfoGain	0.916	0.33	0.911	0.913	0.634	0.827	0.898
Bagging	0.934	0.309	0.931	0.93	0.708	0.9	0.943

Table 3 indicates the results of Random Forest classifier in combination with AdaBoost M1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 3 for Random Forest best result is obtained when Gain Ratio is used in conjunction with as indicated in the table.

Table 3 Random Forest Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.88	0.696	0.877	0.845	0.369	0.777	0.871
PCA	0.861	0.831	0.845	0.802	0.135	0.764	0.861
GainRatio	0.871	0.724	0.853	0.834	0.295	0.82	0.895
InfoGain	0.871	0.724	0.853	0.834	0.295	0.82	0.895
Bagging	0.87	0.759	0.859	0.825	0.27	0.816	0.894

Table 4 indicates the results of KNN classifier in combination with AdaBoost M1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 4 for KNN best result is obtained when Bagging is used in conjunction with as indicated in the table.

Table 4. kNN k=1 Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.82	0.697	0.793	0.804	0.147	0.554	0.772
PCA	0.811	0.787	0.76	0.782	0.033	0.512	0.755
GainRatio	0.815	0.736	0.778	0.793	0.101	0.54	0.763
InfoGain	0.815	0.736	0.778	0.793	0.101	0.54	0.763
Bagging	0.832	0.721	0.794	0.809	0.148	0.638	0.809

Table 5 indicates the results of Logistic Regression classifier in combination with AdaBoost M1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 5 for Logistic Regression best result is obtained when Gain Ratio is used in conjunction with as indicated in the table.

Table 5. Logistic Regression Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.847	0.666	0.817	0.827	0.238	0.62	0.801
PCA	0.847	0.675	0.816	0.826	0.231	0.701	0.831
GainRatio	0.847	0.666	0.817	0.827	0.238	0.688	0.821
InfoGain	0.847	0.666	0.817	0.827	0.238	0.688	0.821
Bagging	0.844	0.684	0.812	0.822	0.214	0.681	0.814

Table 6 indicates the results of Decision Stump classifier in combination with AdaBoostM1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 6 for Decision Stump best result is obtained when Bagging is used in conjunction with as indicated in the table.

Table 6. Decision Stump Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.858	0.638	0.831	0.838	0.291	0.799	0.874
PCA	0.858	0.849	0.808	0.795	0.056	0.614	0.795
GainRatio	0.853	0.613	0.831	0.838	0.296	0.631	0.803
InfoGain	0.853	0.613	0.831	0.838	0.296	0.631	0.803
Bagging	0.865	0.672	0.838	0.839	0.295	0.737	0.845

Table 7 indicates the results of LMT classifier in combination with AdaBoostM1, PCA, Gain Ratio, Bagging and Info Gain is used to classify the churn data set. As it could be seen from the Table 7 for LMT best result is obtained when Bagging is used in conjunction with as indicated in the table.

Table 7. LMT Classifier with ensemble methods

Ensemble Method/ Attribute Selector	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Ada boostM1	0.906	0.393	0.898	0.906	0.9	0.577	0.82	0.903
PCA	0.862	0.672	0.833	0.862	0.836	0.282	0.76	0.867
GainRatio	0.906	0.376	0.899	0.906	0.901	0.583	0.859	0.918
InfoGain	0.906	0.376	0.899	0.906	0.901	0.583	0.859	0.918
Bagging	0.919	0.373	0.914	0.919	0.913	0.634	0.886	0.93

4.2 Discussion

In order to compare our study with others which had used the same dataset, Table 8 is compiled. As it could be seen from the tablet that proposed methods are as good as the other methods and in some aspect it is better. On the other hand since in this study many comparison metrics are used only a few of them were common with others.

Table 8. Comparison with other studies

	TP Rate	F1 Geometric average Sensitity and Precision	Method Orange Data set
(Jain et al. 2020)	85.2385 %	%98	Logistic regression and Logistic Boost
(Azeem et al. 2018)	98%*	-	Fuzzy based Classifiers
	95%	85%	Random forest,
This Study	93%	90%	J48,

5 CONCLUSIONS

In this study, data mining classification is performed for Churn analysis in Telecom sector. In order to better understand and clarify the effects of the different methods namely J48, Naïve Bayes, Logistic Regression, Random Forest and Decision Dump they are applied on the data set. Moreover ensemble methods such as

AdaBoostM1, Bagging, PCA, Gain Ratio, Info Gain are used in conjunction with aforementioned methods. It has been shown that better results are obtained in classification of the churn data set when applied with the ensemble methods. Bagging together with J48 has better result then the compared ones. Success of the methods can be attributed to the implementation and algorithm of the methods. While J48 is entropy based decision tree, Bagging is simply uses bootstrap aggregating. These results also compared with some of the studies using the same data set as well. The contribution of the study is introducing new aspect of analysis and better understanding to the Churn data set. To the best of our knowledge there is no other study that utilizes comparative results of aforementioned methods.

REFERENCES

Amin, Al-Obeidat, Shah, Adnan, Loo and Anwar. (2019) Customer Churn Prediction in Telecommunication Industry under uncertain situation. Journal of Business Research, V94 pp290-231

Azeem and Usman, (2018). A Fuzzy based Churn Prediction and retention Customer in Telecom Industry. International Journal of Computational Intelligence Systems, Vol. 11 (2018) 66–78

Choniekh and Ibn El Haj, (2020). Deep Convolutional Neural Network for Customer Churn Prediction Analysis. International Journal of Cognitive Informatics and Natural Intelligence Volume 14 • Issue 1 • January-March 2020

Dumirtrache and Martei, (2019). Churn Analysis in a Romanian Telecommunications company. 2019, Volume 10, Issue 4, pages: 44-53 | doi:10.18662/po/93

Faris, (2018). A Hybrid Swarm Intelligent Neural Network model for Customer Churn Prediction and Identifying the Influencing factors. Information 2018, 9, 288; doi:10.3390/info9110288

H. Jain, A. Khunteta and S. Srivastava (2020) Churn Prediction in Telecommunication using Logistic Regression and Logit Boost, Procedia Computer Science 167 101–112

H. Witten, E. Frank, M. Hill, C. Pal, Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann; 4. basım (17 Kasım 2016)

Idris, Khan and Lee, (2013). Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification. Appl Intell (2013) 39:659–672 DOI 10.1007/s10489-013-0440-x

Idris, Khan and Lee, (2012). Genetic programming and Adaboosting based churn prediction for Telecom. 2012 IEEE International Conference on Systems, Man, and Cybernetics October 14-17, 2012, COEX, Seoul, Korea

Khan, Shafiq, Naeem, Ahmed, Safwan and Hussein, (2019). Customers Churn Prediction using Artificial Neural Network in Telecom Industry. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019

Kirui, Hong, Cheruiyot and Hillary Kirui, (2013). Predicting customer churn in mobile telephony industry using probabilistic classifiers in Data Mining. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784 www.IJCSI.org

Landwher and Frank, (2004). Logistic Model Trees LMT. This is an extended version of a paper that appeared in the Proceedings of the 14th European Conference on Machine Learning (Landwehr et al., 2003).

Nijahwan, Madan and Dave, (2019). Analytical Implementation of CART using RStudio for churn prediction. Springer Nature Singapore Pte Ltd. 2019 S. Fong et al. (eds.), Information and Communication Technology for Competitive Strategies, Lecture Notes in Networks and Systems 40, https://doi.org/10.1007/978-981-13-0586-3_11

Oludele, Ben, A.C., S.O. and Seun, (2020). Enhanced churn prediction in the Telecommunication Industry. International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-8, Issue-2, March 2020 https://doi.org/10.21276/ijircst.2020.8.2.1 www.ijircst.org

Özmen, M, Aydoğan, EK, Delice, Y, Toksarı, MD., (2020) Churn prediction in Turkey's telecommunications sector: A proposed multiobjective–cost sensitive ant colony optimization, WIREs Data Mining Knowl Discov. 10:e1338. https://doi.org/10.1002/widm.1338

Pang-Ning Tan, Michael Steinbach, Vipin Kumar "Introduction to Data Mining" Pearson, 2014.
Ullah, Raza, Malik, Imran, Islam and Kim, (2019). Churn prediction model using Random Forest: Analysis of Machine Learning Techniques for churn prediction and foctor identification in Telecom Sector. VOLUME 7, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2914999

Umayaparvathi and Iyakutti, (2012). Applications of Data Mining techniques in Telecom Churn Prediction. International Journal of Computer Applications (0975 – 8887) Volume 42– No.20, March 2012

Vijaya and Sivasankar, (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve customer churn prediction in telecommunication sector. Computing (2018) 100:839860 https://doi.org/10.1007/s00607-018-0633-6

Wikipedia Confusion Matrix https://en.wikipedia.org/wiki/Confusion_matrix : last access 15.01.2021

Yabas, Cankaya and Ince, (2012). Customer churn prediction for Telecom Services. Globecom 2013 Workshop - The 5th IEEE International Workshop on Management of Emerging Networks and Services 978-1-4799-2851 4/13/$31.00 ©2013IEEE

Yu, An, Jin, Shi, Move and Liu, (2018). Particle classification optimization-based BP network for telecommunication customer churn prediction. Neural Comput & Applic (2018) 29:707–720 https://doi.org/10.1007/s00521-016-2477-3

Zhong and Li, (2019). Predicting customer churn in the telecommunication industry by Analyzing phone call transcripts with convolutions Neural Network. ICIAI 2019, March 15–18, 2019, Suzhou, China © 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-6128-6/19/03…$15.00 https://doi.org/10.1145/3319921.3319937