


Churn Prediction in a Pay-TV Company via Data Classification

Ilayda Ulku, Istanbul Kültür University, Istanbul, Turkey

 <https://orcid.org/0000-0003-0464-7007>

Fadime Uney Yuksektepe, Istanbul Kültür University, Turkey

Oznur Yilmaz, Istanbul Kültür University, Turkey

Merve Ulku Aktas, Istanbul Kültür University, Turkey

Nergiz Akbalik, Istanbul Kültür University, Turkey

ABSTRACT

In data mining, if a data set is new to the literature, the study is comparing the existing algorithms and determining the most suitable algorithm. This study is an example of this by including many quantitative analysis. Real data was obtained from a Pay-TV Company in Turkey to predict the churn behavior of the customers. The attributes such as membership period, payment method, education status, and city information of customers were used in order to predict the customers' churn status. By applying attributes selection algorithms, the most important attributes are obtained. As a result, two datasets are proposed. While one of the datasets consists of all attributes, the other one just includes the selected attributes. Many different data classification algorithms were applied to these datasets by using WEKA software. The best method and the best dataset which has the best accuracy rate was proposed to the company. The company can predict the customers' churn status and contact the right group of people for a specific campaign with a proposed user-friendly prediction methodology.

KEYWORDS

Accuracy, Attribute Selection, Churn Management, Data Classification, Data Mining, Turkey

INTRODUCTION

Nowadays, with the increasing number of companies, product diversity, and the advent of technology, competition between companies has increased. Therefore, concepts such as customer satisfaction, customer loyalty, and target group have gained value. At this point, appropriate strategies must be presented to the relevant customers in order to meet the needs of the customers. Consequently, good churn management becomes involved in customer retention. Companies can develop techniques to keep their profitable customers and increasing customer loyalty in advance.

DOI: 10.4018/IJAIML.2021010104

This article, published as an Open Access article on November 6, 2020 in the gold Open Access journal, International Journal of Artificial Intelligence and Machine Learning (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Churn management is applied in different areas such as banking, internet service providers, cosmetics, and the health sector. In this research, the loyalty status of the churn situation is discussed for a digital broadcasting platform. The real data is obtained from one of the leading and largest Pay-TV operators in Turkey.

As long as companies have more loyal customers, they will have a higher profit. With only a five percent increase in loyal customers yields in between 25 to 95 percent increase in the net present value of customers across a wide range of industries” (KhakAbi, Gholamian, & Namvar, 2010). Data mining methods are applied to distinguish the churn status in order to apply churn management actions based on customer information. Not only to increase revenues or reduce risks but also to improve customer relationships data mining process is used. By using various kinds of data mining algorithms, meaningful patterns can be obtained to predict outcomes within large data sets. With this study, significant churn prediction factors are introduced via existing data mining algorithms. The following questions are answered in this research with the help the data obtained from the company:

- What attribute can be reviewed to determine customer churn status in advance?
- Which attribute is not necessary for customer churn status?
- According to the customer information, possible churn customers will be defined?

In this way, the relationships through the data is represented and the customers are grouped according to their similar characteristics. The obtained data is extracted and grouped to get more meaningful results for the data mining analysis. The companies are interested in their customers’ churn rate in order to determine a specific campaign for a particular group of customers to avoid customer churn.

The rest of this paper is following with the background part. The existing literature on data mining and churn management in different sectors are investigated. The existing data mining algorithms are used for the Pay-TV company to increase the company’s net present value. In the next part of this chapter, the problem statement and methodology part is briefly discussed. The information gathered data is used to determine the appropriate data mining method to obtain the results. In the following part, computational results are represented. Then a future research and conclusion parts are described.

BACKGROUND

Nowadays, there are many production and service companies. They all have the common goal of gaining more customers, increasing profits, growing the company, and taking part in the global market. As a result, competition between companies is seen. Companies must constantly improve themselves to acquire new customers because customers tend to the most appropriate company for themselves. The most suitable price, quality, shape, etc. are selected. If the company does not work customer-oriented and does not follow the customer’s changing wishes, the company loses its customer. Due to increasing globalization and competition in the market, customers may not always be loyal to companies because they have many alternatives in the market. Therefore, companies are trying to keep their existing customers, instead of getting new customers. The reason for this situation is the recent research shows that the cost of getting new customers is about 5 times and the cost of customer recovery is about 10 times more expensive than the cost of keeping existing customers (Massey, Montoya-Weiss & Holcom, 2001). In addition, customer churn adversely affects the reputation of the company. As a result, keeping customers onboard has a lower cost than others and will be more profitable for the growth of the company, their finance, and also their reputation. With respect to these cases, customer churn management is a very important topic today. Consequently, for companies to be advantageous and making good management in customer churn, it is very important that

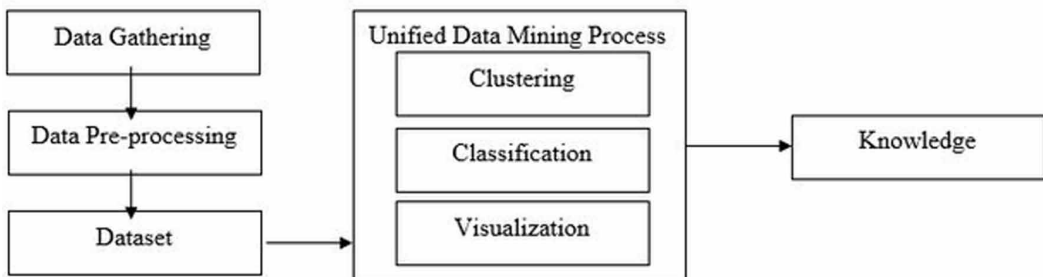
the understanding customer needs and their behaviors, in whole industries. Data mining is needed for a good churn analysis. In the next section, information is given about data mining and its applications.

Data Mining

Today, technology is developing rapidly. As a result of developing technology, the data to be stored is increasing. With so much data stored, it is very difficult to distinguish what is desired and what is important. That's why data mining is the most preferred method. Only the stored data has no meaning. This method determines the relationship, rules, and properties between stored data. Data mining helps to extract meaningful and useful information from available data. In this way, the induction process is analyzed. In other words, data mining provides valuable information by analyzing large amounts of data. Companies use this valuable information in decision-making and action plans. Data mining is an important and supportive tool that companies use to determine strategies and reach solutions on important issues. There are many areas where data mining is applied. For example; banking, medical, electronic commerce, cosmetics, engineering, sports, biology, telecommunications etc.

In the research, unified data mining process is illustrated in Figure 1 that includes clustering, classification, and visualization. Before this process the whole dataset should be pre-processed to be used in WEKA to obtain best method and best dataset.

Figure 1. Unified data mining process



Chahal examined classification algorithms by using real data in terms of the best accuracy and precision values with the help of WEKA software (Chahal, 2018).

Data Mining in Banking

Data mining is of great importance in the banking sector. There are many alternatives for customers to do their jobs. In today's increasingly competitive business environment, banks' retention of existing customers is the most important objective to be analyzed today. It is known that the biggest problem of banks is a fraud. That's why fraudsters must be identified and prevented. At this stage, banks use data mining. Thanks to data mining, it provides a great deal of assistance to companies in retaining existing customers, detecting and preventing fraudsters, and developing marketing strategies (Chitra & Subashini, 2013).

Hassani, Huang, and Silva concluded that several data mining techniques such as clustering and classification are adopted in the banking sector in order to determine fraud detection, risk management, and customer relationship management (Hassani, Huang, Silva, 2018). As more advanced technologies are developed the prevention of cybercrime attempt are increased (Lagazio, Sherif, & Cushman 2014).

Data Mining in Internet Server Provider Companies

As customer behavior plays an important role to determine campaigns, Internet service providers (ISPs) endeavor to prevent losing their customers. In 2016, Ülkü, Durak, and Üney-Yüksektepe studied about customer loyalty status and behavior in ISPs by generating a questionnaire. Then, they investigated the information about the churn possibility in Turkey in order to decide the required advertising campaigns for the customers (Ülkü, Durak, & Üney-Yüksektepe, 2016).

Data Mining in Cosmetic

Churn analysis and customer segmentation are two common application areas of data mining. The cosmetic sector is an example of these applications. In order to retain profitable customers, the data received from the sales personnel are processed and converted to identify the churn customers. By selecting the most appropriate algorithm, the churn behavior of customers is modeled. In this way, the companies retain their customers in the company with appropriate campaigns according to these studies. If we give an example to a study conducted in this sector; a brand which sold in the cosmetics store has applied data mining techniques by taking the data from the sales consultant which they accept as the expert. In this way, the algorithm that gives the most accurate customer analysis has been selected and as a result, it is envisaged to organize campaigns according to seen embranchment which include the customer arrival period, spending period, and the neighborhood (Kızılkaya Aydoğan, Gencer & Akbulut, 2008).

Data Mining in Health Sector

Using data mining is common in the health sector. With the data obtained by screening tests, the detection of diseases and disease risks can be determined. Also because of chronic diseases, the life quality of many people decreases and people die. In order to prevent these losses, many criteria such as economic, gender, age, and demographic characteristics are examined. By using data mining, the factors that may cause the disease to occur are identified and prevented (Koyuncugil & Özgülbaş, 2009). Moreover, Zhi-Gen, Jian-Ping, Hu, and Yang studied the medical image and multiple physiological analyses by using data mining technology to can provide support to the scientific management of the hospital (Zhi-Gen, Jian-Ping, Hu, & Yang, 2015).

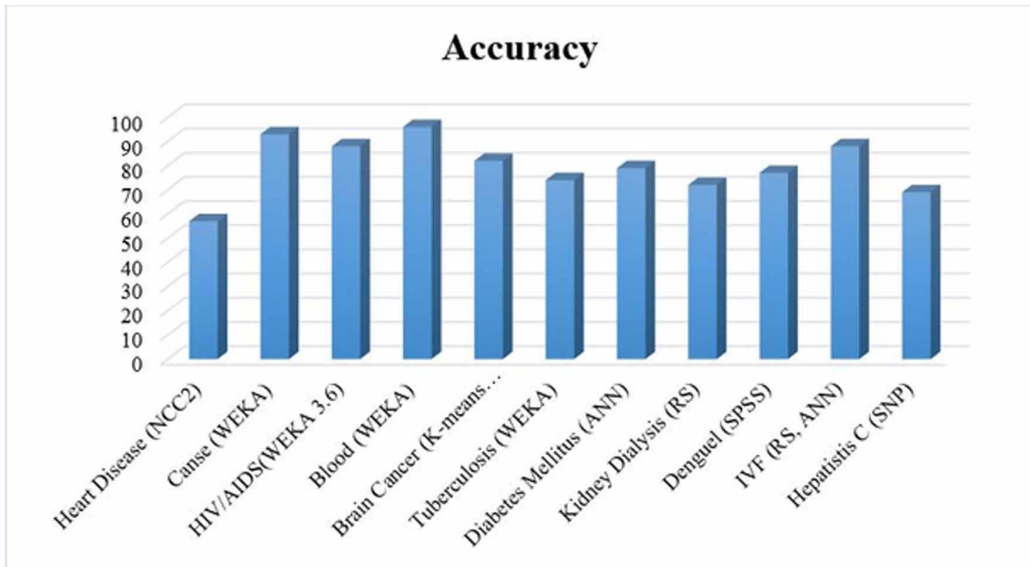
In Figure 2, accuracy levels are demonstrated according to various disease types. Mia, Hossain, Chhoton, and Chakraborty applied various types of data mining tools to the set of selected diseases in order to determine the accuracy of each particular tool (Mia, Hossain, Chhoton, & Chakraborty, 2018).

PROBLEM STATEMENT AND METHODOLOGY

Nowadays, thanks to technological advances, computers and information systems are developing and getting cheaper day by day. These developments make it possible to store and process large amounts of data and facilitate access to data. The growing batch of data is worthless in itself and needs to be processed to make sense. Data mining is defined as analyzing data from this large data stack, revealing hidden patterns between data, and summarizing useful information for users. Data Mining can be used in any field where there is a large data set based on any kind of electronic environment and a result is expected from this data set. For example, the data mining system used in the marketing area, sales forecasts, and market basket analyzes can be made by collecting and processing the customer's shopping habits and customer basic characteristics data and helps to remain competitive in the sector by developing appropriate marketing strategies.

In this research, the data mining application is performed by obtaining the information of approximately ten thousand customers from a digital broadcasting platform. This data includes customer information such as membership period, group of package, gender, city, etc. In this way,

Figure 2. Accuracy levels according to various disease types



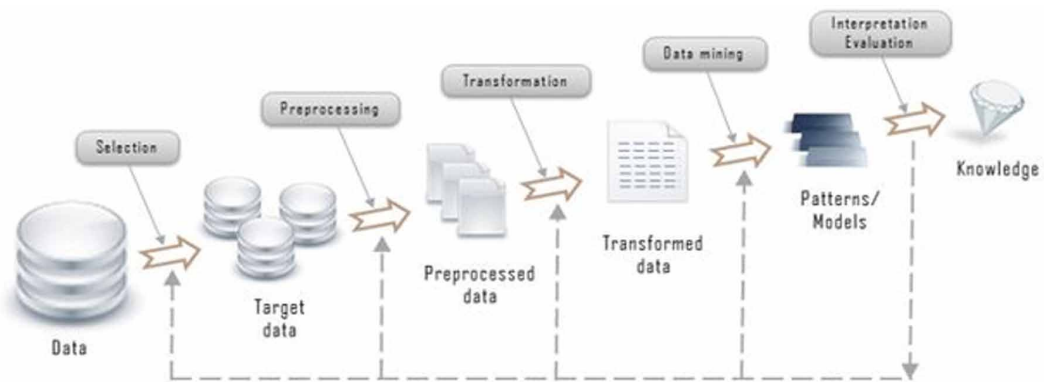
it is aimed to examine the loyalty status of customers and reach the right customer groups of the company through the right campaigns.

There are some notions in the literature close to data mining such as knowledge mining from databases, knowledge extraction, data/pattern analysis, and etc. The most well-known of these notions is The Knowledge Discovery in Databases (KDD). If the Knowledge Discovery in Databases is briefly defined; It is the process of revealing meaningful, useful, original, and specific patterns in the data. Data mining is a stage of the information discovery process (Özdemir, Aslay, & Çam, 2010). Data mining is a stage of knowledge discovery in databases. It includes the below processes as represented in Figure 3.

- **Learn the Application Domain Set:** Determining the relevant information and application objectives.
- **Create a Target Dataset:** Involves selecting a sample data set and deciding variables.
- **Data Extraction and Preprocessing:** Incorrect and missing data in the sample is eliminated and arranged according to certain rules. The data quality is improved.
- **Data Reduction and Conversion:** It includes the selection of required dimensions, finding useful properties between dimensions, size reduction, and transformations.
- **Data Mining:** Deciding which data mining methods to use for searching patterns and derive knowledge from preprocessed data.
- **Evaluation of the Discovered Knowledge:** The evaluation of the patterns found, elimination of the excess and irrelevant patterns, and arranging the result in a comprehensible manner by the users.
- **Use of the Discovered Knowledge:** It is the final step that including the comparison of the finding results with the expected results and creates a plan to monitor for implementation of discovered knowledge.

In this research, the WEKA program is used for data mining applications. Data received from the company is preprocessed. Then, unnecessary information is eliminated such as membership

Figure 3. Data mining process



numbers. Among several data mining methods, the classification method is used. In the following section, information about the classification algorithm is given.

CLASSIFICATION

The most recognized issue of data mining is to assign data to the classes by a classifier according to various attributes. At this stage, the target class in the data is estimated to explore the relevance between target values and the predictor values. Inductive learning consists of a training set of training examples and a test set of test samples. The classification involves two steps: data training and testing the model. During the training phase, the model is formed from the training set. In the testing phase, the model is verified by the test set. Validation is achieved by comparing the class found in the test sample with the class estimated through the model. Artificial Neural Networks, Trees, Bayes, K-Nearest Neighbours are some of the classification methods used in data mining studies. The following classification algorithms are used for accurate churn management.

Artificial Neural Networks

Artificial neural networks are a method inspired by how the human brain learns. Firstly, it was created for the purpose of image and voice recognition. Later, it became one of the data mining algorithms. The artificial neural networks comprise three layers: the input layer, the hidden layer, and the output layer. In the modeling stage, the input layer includes arguments, the output layer includes the variables, and the hidden layer contains the variables from the previous layer or the combined combinations of the nodes in each node as represented in Figure 4. (Lek, & Park, 2008).

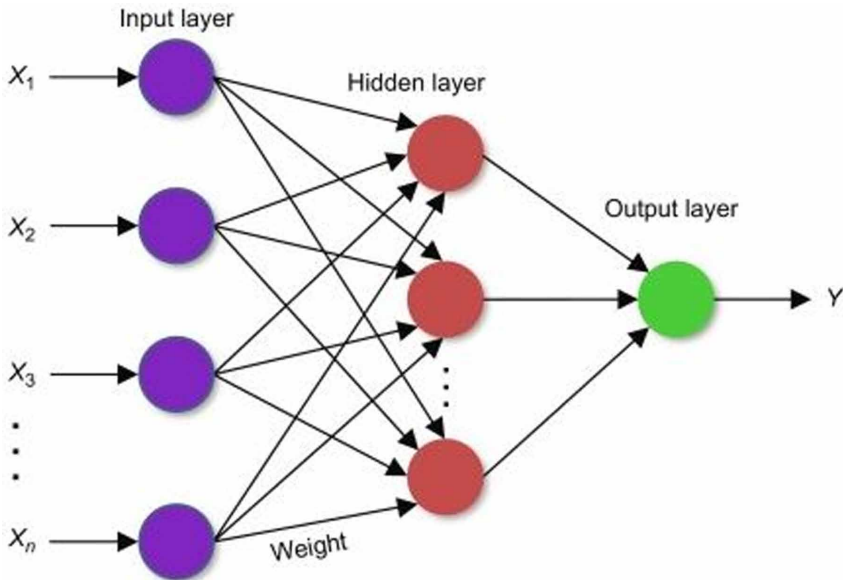
Trees

Decision trees are among the most important classification tools. Decision trees, which provide advantages in many subjects, are widely used in data mining because of their simple algorithm and easy to understand. It performs many tests while estimating in decision trees. The branches in the decision tree include the tests. To classify the target, go from root to target leaf. This path is called a rule. The following Figure 6 shows a part of the decision tree algorithm of J48 for the data of pay-TV churn status.

Bayes

Bayes' theorem is a technique that uses Bayesian classifiers in case of conditional probability. It is assumed that each training data is independent of each other and the prior probability for each

Figure 4. The modeling stage and how the nodes are input



class is calculated from the training data. Then, the class with the most likely probability is selected (Brownlee, 2016).

K-Nearest Neighbours

Another technique applied for classification is the k-nearest neighbor approach. System operation is provided by training data where k most similar training patterns are queried to make a prediction. When a prediction is requested, the computation is implemented by the querying of the training dataset with appropriate distance functions (Brownlee, 2016).

Evaluation of Classification Method

A confusion matrix is needed to evaluate the classification method. The confusion matrix is a table that is often used to measure the performance of classifiers and gives information about the predicted and actual classifications. Precision and accuracy are calculated by using the confusion matrix. It has 4 different output combinations as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These terms in Table 1 are explained:

- **True Positive:** The positive situations which are predicted correctly.
- **True Negative:** The negative situations which are predicted correctly.
- **False Positive:** The negative situations which are predicted incorrectly.
- **False Negative:** The positive situations which are predicted incorrectly.

Table 1 shows the confusion matrix for two classifiers and in this study, accuracy is calculated by using a confusion matrix. Accuracy represents the proportion of the total number of predictions that are correct. Figure 5 displays the accuracy of Pay-TV churn status among several algorithms used in WEKA.

Table 1. Confusion matrix

		Predicted classification	
		Yes	No
Actual classifications	Yes	True positives (TP)	False negatives (FN)
	No	False positives (FP)	True negatives (TN)
		Precision: $\frac{TP}{(TP + FP)}$	Negative Predictive Value: $\frac{TN}{(TN + FN)}$
		Accuracy: $\frac{TP + TN}{(TP + TN + FP + FN)}$	

IMPLEMENTATION

The University of Waikato has developed the WEKA program for data mining operations. Preprocess, classification, association, clustering and visual methods are used in the program to solve machine learning problems. In this study, the real data is obtained from a Pay-TV company. The data is classified by using the WEKA program and the accuracy values of the methods are measured and compared. The algorithm with the highest accuracy is used to find accurate results about churn situations.

The data used in this study is obtained by a Pay-TV Company that is established in 1999 in Turkey. The company serves as the first and leading platform in the field of digital television broadcasting. Pay-TV Company appeals to all age groups with its series, sports, movies, kids and family, documentaries, music, and life channels. However, as a result of increasing competition and globalization, some of the customers are churning overtime. Therefore, customer loyalty and the prevention of customer loss have gained importance for the company. In this study, in total 10,000 customers, information is used to determine the important features. 5000 customers that were churned and 5000 customers that were not churned is obtained from Pay-TV company. There are 13 attributes used that are grouped into 3 numerical attributes and 10 nominal attributes. The attribute names and types are represented in Table 2.

COMPUTATIONAL RESULTS

By applying attributes selection algorithms, the most important attributes are obtained. As a result, two datasets are proposed. While one of the datasets consists of all attributes, the other one just includes the selected attributes. Many different data classification algorithms are applied to these datasets by using WEKA software.

Data Set 1

As seen in Table 3, the highest accuracy value of data set 1 is given as 76.61% for Meta – FilteredClassifier algorithm. The methods with the highest accuracy rates from all classifier types for both dataset 1 and dataset 2 are shown with bold in the following table. Among whole classification algorithms, the highest accuracy value of data set 1 is determined as 76.93% with the Trees - J48 algorithm. Divide and conquer strategy that is recursively from top to down is called J48 (C4.5) to obtain information about the attribute at each stage (Brownlee, 2016). With the help of a visualization property of a decision tree, we are able to follow the branches that are split according to the attributes.

Table 2. Determined attributes (3 numerical, 10 nominal in total)

Attribute No	Attribute Name	Attribute Type
1	Membership Period	Nominal
2	# of Membership	Numeric
3	Group of Package	Nominal
4	Geographical Region	Nominal
5	Gender	Nominal
6	City	Nominal
7	Value Segment	Nominal
8	Sales Channel	Nominal
9	Late Payment	Numeric
10	NonPayment	Numeric
11	Payment Frequency	Nominal
12	Commitment Status	Nominal
13	Status	Nominal

Although there are several methods that are around the similar accuracy rate, the highest accuracy value is proposed to the decision makers.

Also, Patil and Sherekar reported that the efficiency and accuracy of J48 algorithm gives better results than Naive Bayes (Patil and Sherekar, 2013).

Next, attribute selection algorithms are used for using necessary attributes. City (6th attribute), Sales Channel (8th attribute), and Late Payment (9th attribute) attributes are ignored. In Table 4, attribute selection methods and selected attributes shown. By using the two data set, classification methods are used to predict churn status. Suggested data sets are represented in Table 5.

Data Set 2

As seen in Table 6, the highest accuracy value of data set 2 is 75.67% for the Functions-Logistics algorithm. Among whole classification algorithms, the highest accuracy value of data set 1 is 76.78%. The name of this algorithm is Meta – FilteredClassifier.

Best Data Set and Method

Data Set 1 has a significant factor to predict the churn situation of Pay-TV customers. The highest accuracy value is obtained from J48 Algorithm of Trees Classification. J48 is an algorithm used to generate a decision tree and J48 is an open-source Java implementation of the C4.5 algorithm in the Weka data mining tool. Decision trees are generated from a set of training data in the same way as ID3 using the concept of information entropy by C4.5 algorithm. In addition, Weka machine learning software developers described the C4.5 algorithm as a landmark decision tree program that is the machine learning workhorse probably most widely used in practice to date (Witten, Frank, Hall, 2011).

Table 7 shows that there are 2302 misclassification customers based on the selected method J48. As it is seen, 989 disloyal customers classified as loyal and 1313 loyal customers classified as disloyal.

Figure 5 shows the accuracy graph of all algorithm types used for Dataset 1. The algorithm with the highest accuracy is Trees. The lowest accuracy value of data set 1 is determined as the Misc algorithm.

The following information can be obtained from Figure 6 which represent the decision tree of the Pay-TV churn status:

Table 3. Accuracy of classification methods for data set 1

Types	Method	Accuracy (%)
BAYES	Bayes Net	72.88
	Naive Bayes	72.53
	Naive Bayes Updateable	72.53
	Naive Bayes Multinomial Text	51.6
FUNCTIONS	Logistic	75.82
	SGD	75.89
	SGD Text	51.6
LAZY	IBk	71.66
META	Iterative Classifier Optimizer	72.9
	Ada Boost M1	61.99
	Attribute Selected Classifier	70.09
	Bagging	75.55
	CV Parameter Selection	51.6
	Filtered Classifier	76.61
	Logit Boost	72.9
	Multi Scheme	51.6
	Randomizable Filtered Classifier	61.8
	Random Sub Space	75.77
	Stacking	51.6
	Vote	51.6
	Weighted Instances Handler Wrapper	51.6
	MISC	Input Mapped Classifier
RULES	Decision Table	75.82
	JRip	75.66
	OneR	61.79
	PART	74.15
	ZeroR	51.6
TREES	Decision Stump	56.63
	Hoeffding Tree	72.22
	J48	76.93
	Random Tree	71.05
	REP Tree	75.7

Table 4. Attribute selection methods and selected attributes

Method	Selected Attributes
CFSSubsetEval	4,7,11
GainRatioAttributeEval	4,11,7,12,2,1,3,5,10,9,6,8
InfoGainAttributeEval	4,11,7,1,12,6,3,10,5,2,9,8
OneRAttributeEval	4,1,12,11,3,7,5,6,10,9,8,2
ClassifierAttributeEval	12,11,4,3,2,5,6,7,10,9,8,1
CorrelationAttributeEval	11,12,3,5,1,4,2,10,7,9,8,6
SymmetricalUncertAttributeEval	4,11,7,12,1,3,6,5,10,2,9,8

Table 5. Suggested data sets and characteristics

	Attribute No	Attribute Name	Attribute Type
Data Set 1	1-13	All Attributes	Nominal, Numerical
Data Set 2	1	Membership Period	Nominal
	2	# of Membership	Numeric
	3	Group of Package	Nominal
	4	Geographical Region	Nominal
	5	Gender	Nominal
	7	Value Segment	Nominal
	10	NonPayment	Numeric
	11	Payment Frequency	Nominal
	12	Commitment Status	Nominal
	13	Status	Nominal

- Customers with the package group Premium Sport can be classified as disloyal.
- Customers with the package group Basic Sport can be classified as loyal.
- To classify customers whose package group is Basic, the number of membership attribute should be checked. If the number of membership is greater than one, it is classified as loyal. If the number of membership is less than or equal to one, the nonpayment attribute should be checked. The nonpayment attribute is classified as loyal if it is greater than zero and disloyal if it is less than or equal to zero.
- To classify customers whose package group is Cinema, the gender attribute should be checked. If the gender is female, it is classified as disloyal. If it is a male customer, the number of membership attribute should be checked. The number of membership is classified as loyal if it is greater than one and disloyal if it is less than or equal to one.

CONCLUSION

As a result of increasing population, product range, technology, and competition concepts such as customer loyalty, data mining, and big data have gained importance. Data mining tools are often used to understand and analyze customer requests and behavior. In data mining, if a data set is new to the

Table 6. Accuracy rates of classification methods for data set 2

Types	Method	Accuracy (%)
BAYES	Bayes Net	73.6
	Naive Bayes	73.29
	Naive Bayes Updateable	73.29
	Naive Bayes Multinomial Text	51.6
FUNCTIONS	Logistic	75.67
	SGD	75.16
	SGD Text	51.6
LAZY	IBk	74.4
META	Iterative Classifier Optimizer	72.9
	Ada Boost M1	61.99
	Attribute Selected Classifier	70.09
	Bagging	76.5
	CV Parameter Selection	51.6
	Filtered Classifier	76.78
	Logit Boost	72.9
	Multi Scheme	51.6
	Randomizable Filtered Classifier	71.28
	Random Sub Space	75.43
	Stacking	51.6
	Vote	51.6
	Weighted Instances Handler Wrapper	51.6
	MISC	Input Mapped Classifier
RULES	Decision Table	75.82
	JRip	75.9
	OneR	61.79
	PART	76.01
	ZeroR	51.6
TREES	Decision Stump	56.63
	Hoeffding Tree	72.97
	J48	76.66
	Random Tree	74.73
	REP Tree	76.56

Table 7. Confusion Matrix for J48

		Predicted Class	
		Disloyal	Loyal
Original Class	Disloyal	3840	989
	Loyal	1313	3836

Figure 5. Accuracy graph of all algorithm types used for Dataset 1

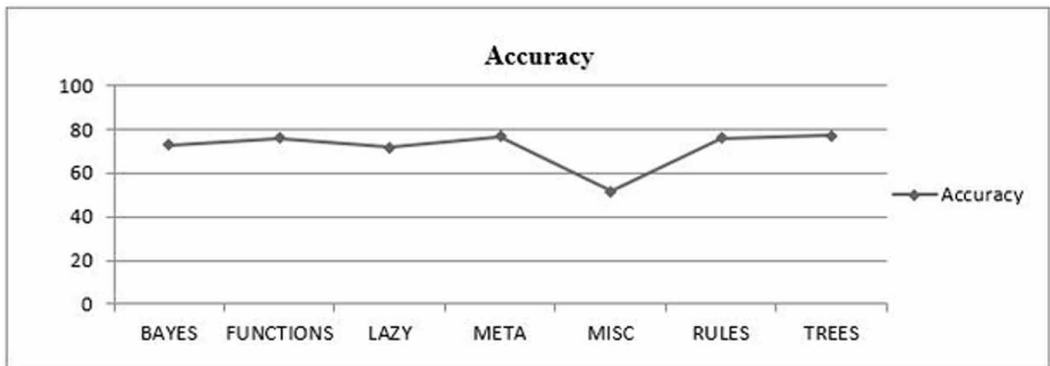
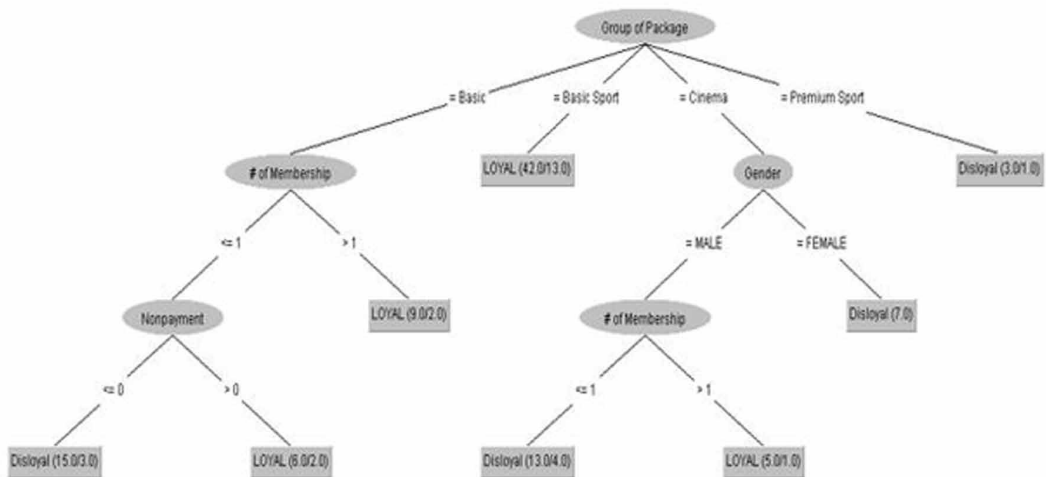


Figure 6. Decision tree of the Pay-TV churn status



literature, the study is comparing the existing algorithms and determining the most suitable algorithm. This study is an example of this. In today's competitive world, companies in various sectors in Turkey are beginning to give more importance to this issue and operate infrastructure based on data mining algorithms. Pay-TV company has various product range which includes series, sports, movies, kids and family, documentaries, music, and life channels. It caters to all age groups and genders. Therefore, it has many different customer profiles. Pay-TV should develop and renew itself in order to maintain this broad customer scale. In this study, obtained real data is used. The highest accuracy of the data

set is selected using data mining classification methods that is a decision tree-based algorithm. With the help of data mining tools, the company will be able to track customer behavior. The loyalty status of customers can be investigated and thus the company can contact the right group of people for a specific campaign by using the proposed decision tree.

FUTURE RESEARCH

As an extension of this study, more customer attributes will be obtained in order to examine the churn status in Pay-TV broadly and accurately. As long as more attributes are included, more accurate results will be obtained to provide the quality of the decision. The relationships between attributes are ignored in this study. Whether this relationship has an impact on accuracy can be examined as future research. If higher accuracy can be achieved, the results will be re-evaluated.

REFERENCES

- Brownlee, J. (2016). *How to Use Classification Machine Learning Algorithms in Weka*. Weka Machine Learning.
- Chahal, H. (2018). Comprehensive Analysis of Data Mining Classifiers using WEKA. *International Journal of Advanced Research in Computer Science*, 9, 2. doi:10.26483/ijarcs.v9i2.5900
- Chitra, K., & Subashini, B. (2013). Data Mining Techniques and its Applications in Banking Sector. *International Journal of Emerging Technology and Advanced Engineering. Certified Journal*, 3, 8.
- Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and Big Data Mining in Banking. *Big Data Cognitive Computing*, 2(3), 18. doi:10.3390/bdcc2030018
- KhakAbi, S., Gholamian, M., R., & Namvar, M. (2010). Data Mining Applications in Customer Churn Management. *2010 International Conference on Intelligent Systems, Modelling and Simulation*, 220-225. doi:10.1109/ISMS.2010.49
- Khan, D. M., Mohamudally, N., & Babajee, D. K. R. (2012). Towards the Formulation of a Unified Data Mining Theory, Implemented by means of Multiagent Systems (MASs). In *Advances in Data Mining Knowledge Discovery and Applications*. InTech.
- Kızılkaya Aydoğan, E., Gencer, C., & Akbulut, S. (2008). Churn Analysis And Customer Segmentation of A Cosmetics Brand Using Data Mining Techniques. *Journal of Engineering and Natural Sciences. Sigma*, 26, 1.
- Koyuncugil, S. A., & Özgülbaş, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri Dergisi*, 2, 2.
- Lagazio, M., Sherif, N., & Cushman, M. (2014). A multi-level approach to understanding the impact of cyber crime on the financial sector. *Computers & Security*, 45, 58–74. doi:10.1016/j.cose.2014.05.006
- Lek, S., & Park, Y. S. (2008). Artificial Neural Networks. *Encyclopedia of Ecology*.
- Massey, A. P., Montoya-Weiss, M. M., & Holcom, K. (2001). Re-engineering the customer relationship: Leveraging knowledge assets at IBM. *Decision Support Systems*, 32, 155–170. doi:10.1016/S0167-9236(01)00108-7
- Mia, M. R., Hossain, S. A., Chhoton, A. C., & Chakraborty, N. R. (2018). A Comprehensive Study of Data Mining Techniques in Health-care, Medical, and Bioinformatics. *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. doi:10.1109/IC4ME2.2018.8465626
- Özdemir, A., Çam, H., & Aslay, F. (2010). *Data Discover Process in the Database: Gümüşhane State Hospital Application*. *SU Journal of Social and Economic Research*.
- Patil, T. R., & Sherekar, S. S. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal of Computer Science and Applications*, 6, 2.
- Sumiran, K. (2018). An Overview of Data Mining Techniques and Their Application in Industrial Engineering. *Asian Journal of Applied Science and Technology*, 2(2), 947–953.
- Ülkü, İ., & Durak, M. (2016). *Churn Prediction in Internet Service Provider Companies*. Intelligent Techniques for Data Analysis in Diverse Settings. doi:10.4018/978-1-5225-0075-9.ch013
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, Elsevier.
- Zhi-Gen, H., Jian-Ping, L., Hu, L., & Yang, Y. (2015). Research and application of data warehouse and data mining technology in medical field. *Wavelet Active Media Technology and Information Processing (ICCWAMTIP) 12th International Computer Conference*, 457-460. doi:10.1109/ICCWAMTIP.2015.7494030