Düzce University
Journal of Science & Technology

Research Article

Analysis of Customer Churn in Telecommunication Industry with Machine Learning Methods

Özge Nalan Bilişik [a*], Damla Tuğba Sarp [a]

[a] *Department of Industrial Engineering, Faculty of Mechanical Engineering, Yıldız Technical University, İstanbul, TURKEY*
** Corresponding author's e-mail address: oznalan@yildiz.edu.tr*
DOI:10.29130/dubited.1061257

ABSTRACT

In today's conditions, customer loyalty has gained importance with the increase in the competitive environment between companies, the development of marketing strategies and the improvement of companies. Therefore, it is essential to acquire customers for a company to survive. Retaining an existing customer in the telecommunication sector is less costly than gaining a new customer. Customer churn analysis is the process of predicting customers with high abandonment requests by examining the offers and utilizable behaviors. Customer churn analysis provides services to develop various campaigns aiming to increase the company's loyalty by predicting the customers who are planning to move to another company. In this way, it gives the company a competitive advantage. This study aims to make predictions by developing models for customer churns through data mining and machine learning methods in the telecommunication sector. In addition, we believe that the application in this article will contribute to data analysts and academicians who will want to analyze customer churn with different data sets in telecommunication and other sectors in the future. The analysis in this study is carried out on a data set obtained from an open-access database, including 20 transaction records for the customer from 7043 customers and whether the customer left the company. Among the data mining methods, Random Forest (RF), Support Vector Machines (SVM) and Multilayer Artificial Neural Networks (ANN) are modeled in open-source Phyton environment. The results have shown that ANN has fared better at classifying customers than other machine learning methods.

Keywords: Customer Churn, Machine Learning, ANN, SVM, RF

Telekomünikasyon Sektöründe Müşteri Kaybının Makine Öğrenmesi Yöntemleriyle Analizi

ÖZ

Günümüz koşullarında şirketler arasındaki rekabet koşullarının artması, pazarlama stratejilerinin gelişmesi ve şirketlerin değişimi ve gelişimi ile müşteri ve müşteri sadakati önem kazanmıştır. Bir şirketin ayakta kalabilmesi için müşteri kazanmak önemlidir. Telekom sektöründe mevcut bir müşteriyi elde tutmak, yeni bir müşteri kazanmaktan daha az maliyetlidir. Müşteri kaybı analizi teklif ve davranışların incelenerek şirketi bırakma isteği yüksek olan müşterilerin tahmin edilmesi sürecidir. Müşteri kaybı analizi, başka bir şirkete geçmeyi planlayan müşterileri tahmin ederek, şirket bağlılığını artırmaya yönelik çeşitli kampanyalar geliştirmeye yönelik hizmetler sunmaktadır. Bu sayede firmaya rekabet avantajı sağlamaktadır. Bu çalışmanın amacı, telekomünikasyon sektöründe veri madenciliği ve makine öğrenmesi yöntemleriyle müşteri kayıplarını modelleyerek tahminlerde bulunmaktır. Ayrıca bu makaledeki uygulamanın gelecekte telekomünikasyon ve diğer sektörlerde farklı veri setleri ile müşteri kayıplarını analiz etmek isteyecek veri analistlerine ve akademisyenlere katkı sağlayacağı düşünülmektedir. Bu çalışmadaki analiz, açık erişimli bir veri tabanından elde edilen, 7043 müşteriye ait 20 işlem kaydını ve müşterilerin şirketten ayrılıp ayrılmadığını içeren bir veri seti üzerinde gerçekleştirilmiştir. Veri

Received: 21/01/2022, Revised: 24/01/2023, Accepted: 27/01/2023

madenciliği yöntemlerinden Rastgele Orman (RF), Destek Vektör Makineleri (SVM) ve Çok Katmanlı Yapay Sinir Ağları (YSA) açık kaynaklı Phyton Ortamında modellenmiştir. Sonuçlar analiz edildiğinde, YSA müşterileri sınıflandırmada diğer makine öğrenimi yöntemlerinden daha başarılı olmuştur.

Anahtar Kelimeler: *Müşteri Kaybı, Makine Öğrenmesi, YSA, DVM, RO.*

I. INTRODUCTION

An organization needs its customers to survive for a long time. Therefore, meeting customer requirements in the long term is only possible with a successful customer relationship management (CRM) process [1]. Good CRM requires being in contact with customers and both listening and understanding their requirements and needs. Strategies are developed to increase customer loyalty by offering products, services and campaigns in line with these needs,. Customer values are determined by examining purchasing behaviors with CRM (Bagheri and Tarokh, M. [2]; Farquad et al. [3]; Heinrich and Helfert [4]). In the literature, churn of customers is referred to in many different terms. The most important of these are customer wear, customer withdrawal, customer run-off and customer churn [5].

Customer churn analysis can be defined as a process that estimates customers considering canceling their subscription and takes precautions to prevent the churn [3]. In marketing, acquiring a new customer is five times more expensive than retaining an existing customer, and at the same time, the profit from a new customer is half that of an existing customer [6].

Businesses keep much valuable information belonging to their customers such as demographics, invoices, contracts, and products used in their data warehouses. From this information, it is possible to make different analyzes in line with various customer requirements, customer satisfaction, customer value or which customer may have the potential to be churned in the future [7]. In this paper, customer churn in the telecommunication sector is predicted with machine learning methods using data which consists of 20 transaction records for the customer from 7043 customers and whether the customer left the company or not

The rest of the paper is organized as follows: the literature review about customer churn, machine learning and the telecommunication sector is presented in Section 2. Then, in section 3, the proposed methodology and algorithms are explained. In section 4, the application, and analysis results are shown. Finally, the conclusion and discussion section is given.

II. LITERATURE REVIEW

In our study, we make a literature search with the keywords ‘Customer Churn’, ‘Machine Learning’ and ‘Telecom’ and present some of the studies in the following:

Ahmad et al. [8] try to help telecommunication sector operators to find out which customers tend to churn by using RF, Decision Tree, Gradient Boosted Machine (GBM) Tree and Extreme Gradient Boosting (XGBOOST) techniques. They compare these techniques results’ and obtained the best results by applying XGBOOST algorithm. Lalwani et al. [9] develop a six-phased algorithm to predict customer churn in telecommunication sector. Adaboost and XGboost algorithms accomplish the best result in their application. Adhikary and Gupta [10] analyze and compare the performances of over 100 classifiers, used in customer churn prediction for a telecommunication company. Alboukaey et al. [11] propose to predict customer churn on daily basis to consider the dynamic behavior of customers. Swetha and Dayananda [12] propose a customer churn prediction model by using XGBoost model and the model is evaluated according to some performance metrics like accuracy, graphical representation, precision and recall. Mishachandar and Kumar [13] use an approach which combines machine learning and big data analytic techniques to make customer churn prediction effectively. Ullah et al. [14] develop a model to predict customer churn for the telecommunication sector by using clustering algorithms. In addition, they analyze the churn factors that are important to determine the root causes of churn. Rani and Kant [15] compare the six most used machine learning classifiers for customer churn prediction in the telecommunication sector. Almufadi et al. [16] apply deep learning for customer churn prediction within Mobile Telephony Churn Prediction Dataset and they achieved a very high efficiency. Jyothi et al. [17] develop a customer churn model that utilizes some classification algorithms to find out churn customers and help to learn the most affecting elements behind the turnover of customers in the telecommunication sector. Pamina et al. [18] develop a model to identify the traits which influence customer churn by using three different machine learning techniques and the IBM Watson Dataset is used to predict the churn. Wassouf et al. [19] develop a methodology that helps to focus on different customer segments by most reasonable offers and services. Firstly, they segment the customers, secondly, they take the loyalty levels of each segment, thirdly they use classification algorithms based on these descriptives and finally they evaluate these models based on several criteria and develop some rules. Alzubaidi and Al-Shamery [20] conduct research by using a methodology called project pursuit Random Forest (PPForest), which uses discriminant feature analysis to predict churners/non-churners in the telecommunication sector. Rao et al. [21] develop an algorithm via machine learning to test the key performance indicators (KPIs) and weed out the troublesome sites from a network's traffic data dump spanning over a month which would help to identify the problematic websites, thus addressing the customer's problems.

Like the studies above, Malleswari et al. [22], Suguna et al. [23], Isabella Amali [24], Raja and Pandian [25] and Jaisakthi [26] study on customer churn prediction problem and use some machine learning techniques.

In addition to these studies, some authors make literature surveys about customer churn in telecommunication sector: Pamina et al. [27] conduct a literature survey to assist the researchers or data analysts to find out the most appropriate techniques to predict customer churn in telecommunication sector. Jain et al. [28] present a literature review about customer churn and its effects, the root causes of customer churn, the requirements of businesses and the method and techniques to predict customer churn. Seema and Gupta [29] conduct a literature review on customer churn prediction models for e-commerce enterprises. Finally, Sudharsan and Ganesh [30] make a literature review for the customer churn prediction in telecommunication sector for the technical review.

When the literature is reviewed, making valuable predictions for the future with machine learning has been the subject of many studies recently. In these studies, different machine learning methods are used. These methods include decision trees, ANN, RF, SVM, Naive Bayes, K-Nearest Neighbors (KNN), Regression and XG Boost. In addition, hybrid models are also included in the studies, as well as focusing on a single method. In Table 1, some of the studies, which use machine-learning techniques, are summarized with the details of the studies' scope and algorithms.

According to the market data shared by the Information Technologies and Communications Authority (Bilgi Teknolojileri ve İletişim Kurumu - BTK) in Turkey every quarter, the number of mobile number transfers in each quarter is given in Figure 1. In the third quarter of 2020, the number of mobile number transfer increased by 6.8% compared to the previous quarter and was 3.330.058. As of November 24, 2020, 146.102.648 MNT transactions were made in total [31].

Table 1. *Some studies with the detail of algorithms.*

Scope	Algorithms	Reference(s)
Customer loyalty prediction	Decision tree K-means clustering	[32]
Customer payment behaviors prediction	Association rules Clustering Decision tree	[33]
Customer dissatisfaction prediction	Boosting Decision tree Logistics regression Neural networks	[34]
Customer Retention	XG Boost	[35]
Customer Retention	Logistics regression	[36]
Customer Retention	ADA Boost Decision tree Linear discrimination analysis Logistics regression Naive Bayes Random forest Support vector machine	[37]
Customers' repurchase behaviors prediction	Random forest	[38]
Customer churn prediction	Decision tree Neural networks	[39]
Customer churn prediction	Deep learning Gradient boosted tree KNN Naive Bayes Neural networks	[40]
Customer churn prediction	Artificial neural networks	[41], [42]
Customer churn prediction	Feed-forward artificial neural networks Partial swarm optimization	[43]
Customer churn prediction	Decision tree Regression Neural networks	[44]
Customer churn prediction	Multi-layer neural networks	[45]
Customer churn prediction	Decision trees Logistics regression Artificial neural networks	[46]
Customer churn prediction	Decision trees	[47], [48]
Customer churn prediction	Support vector machine Artificial neural networks Naive Bayes	[49]
Customer churn prediction	Support vector machine	[3], [50]
Customer churn prediction	KNN Random forest XG Boost	[25]

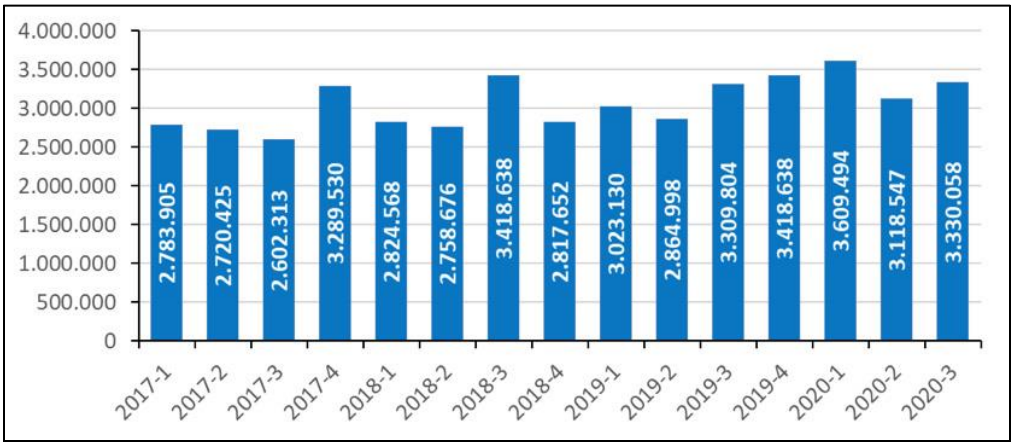

Figure 1. *The distribution of mobile number transferring.*

This study aims to-establish models that predict the customers to churn with machine learning methods on a data set with 21 attributes for 7043 customers in the telecommunication sector and compare these models. In this way, companies will be able to offer various campaigns and strategies to their customers who are likely to churn and increase their customers' loyalty to the company.

III. PROPOSED METHODOLOGY

In this section, the flowchart of the proposed methodology is given in Figure 2 and the algorithms used are explained.

A. ARTIFICIAL NEURAL NETWORK

Artificial neural networks (ANN) is a learning method that mimics the human brain. It has essential functions such as producing meaningful information from the data collected by the brain. Figure 3 indicates the working principle of ANN with layers [51].

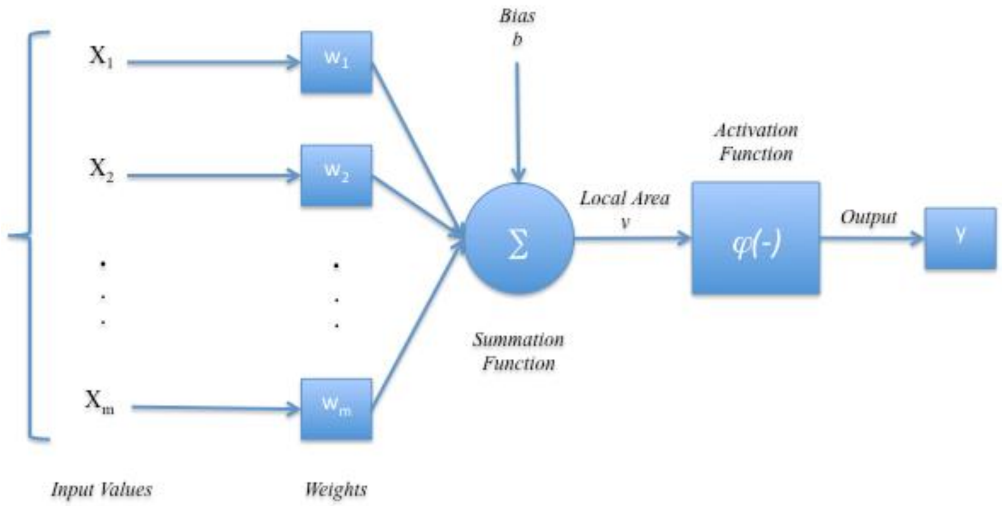

Figure 3. *Working principle of ANN with layers [51]*

Inputs ($X_1, X_2, \ldots, X_m$): It is the data transmitted to the artificial neuron from the outside or the previous layer [52].

Weights ($W_1, W_2, \ldots, W_m$): Indicates how the inputs are transferred to the output.

Addition Function: The net input of a cell is calculated with the additional functions. As expressed in Eq. (1), all entries are summed by multiplying by their own weight [53].

$$\sum_{i=1}^{m} W_i X_i = W_1.X_1 + W_2.X_2 + \cdots + W_m X_m \quad \textbf{(1)}$$

A.1. Artificial Neural Network Models

ANN models are considered in 4 groups: single-layer and multi-layer perceptrons, feed-forward and feedback artificial neural networks. It has been stated that one or two hidden layers are usually sufficient for large structures [54].

- Single-layer perceptrons have multiple inputs and a single output. The output takes values of 1 or -1 and is linear.
- Multilayer perceptrons have multiple inputs and multiple outputs. It is the structure in which non-linear neurons are connected according to different weights. It is used to solve complex and nonlinear problems.
- In feedforward networks, neurons move from input to output in regular layers. There is a bond between successive layers. The information at the input of the neural network is transferred to the hidden layer without any changes. From here, it goes to the output layer and is processed and

transmitted to the outside. In feedback networks, a neuron can have input from a different neuron in the previous layer or in a layer at the same level.

B. SUPPORT VECTOR MACHINE

In cases where the solution method is not completely clear, SVM predicts what the output will be after the inputs in the data, with the available data and information [55]. It is a successful machine learning algorithm created with statistical methods. It is a method based on statistical theories developed to solve classification and regression problems [56]. Each data attribute is represented as a point in n-dimensional space, with the value of each attribute being the value of a particular coordinate. Then, the classification is made with the hyperplane, which makes a good distinction between two classes [57]. It is possible to make nonlinear transformations in SVM with the Kernel function. In this way, linear separation can be made in high dimensions [58].

C. RANDOM FOREST

The RF algorithm is a committee algorithm. It benefits from the joint decisions of the created models. It is a fast method without a fixed pattern. In this method, it is possible to work with as many trees as desired [59].

The RF model works as follow: The model aims to generate a large number of independent decision trees as its primary purpose. When constructing each tree, the model randomly selects a particular subset of all existing observations. This situation is a type of randomness introduced by the model. In this way, each single tree created by the model will learn from independent subsets of the sample at hand [60]. To create a tree with the RF classifier, initially it is sufficient to define two parameters: the number of features to be used for each node (m) and the number of trees to be created (N) [61]. While determining the parameters, the first values are usually chosen randomly, the following values are increased or decreased according to the generalized errors. Generally, when determining the variable value, it is better to take m as the square root of the total number of variables.

The RF algorithm determines the significance criteria of the features with generalized error. That is, it measures the predicted achievements by changing places of the features and ranks them according to their importance [62]. The most important advantage of the RF method is that it can work with many variables, and its performance is fast [63].

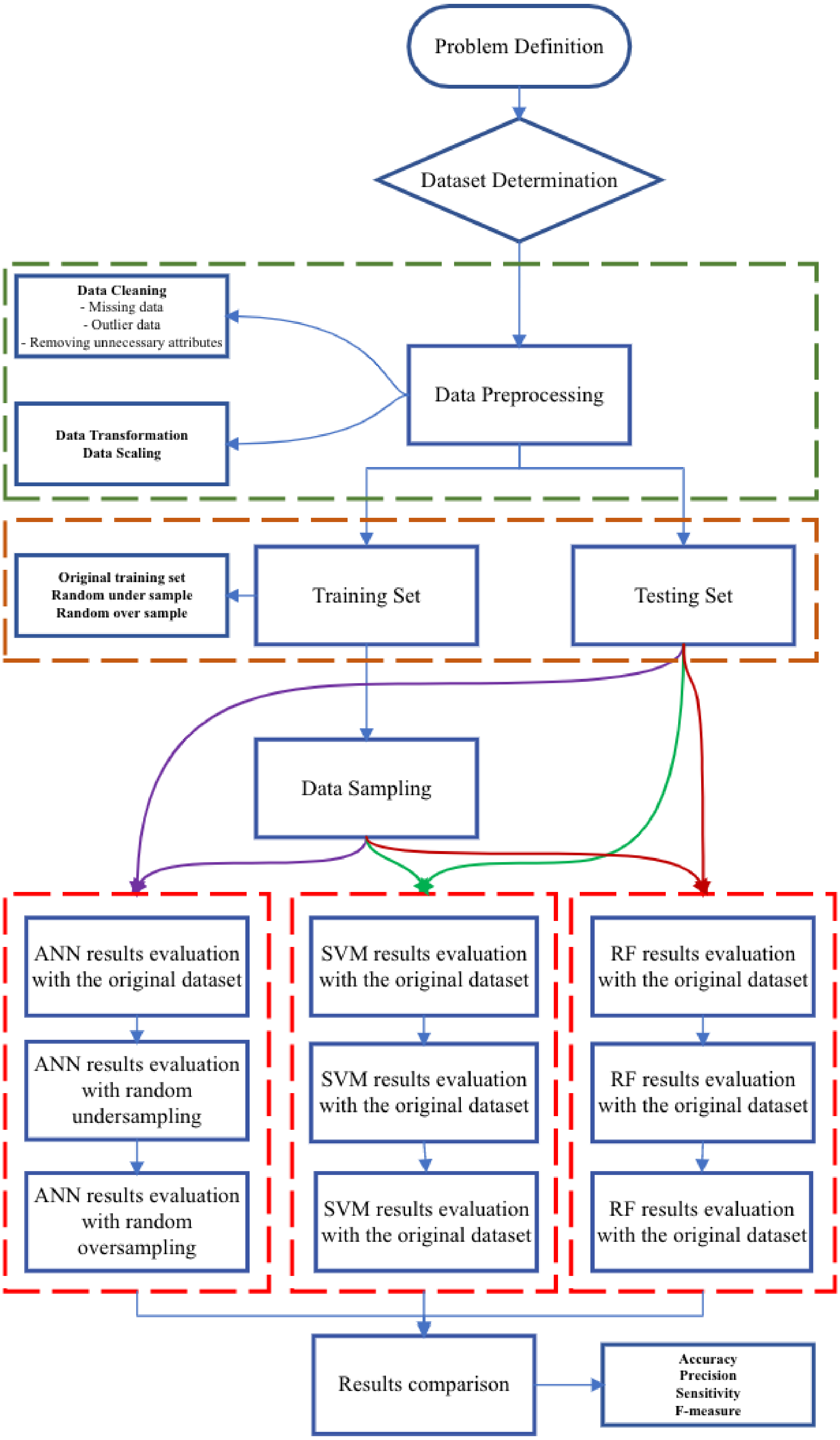

Figure 2. *The flowchart of the proposed methodology.*

IV. APPLICATION

In this study, an exemplary application of customer churn in telecommunication sector has been carried out with machine learning techniques that are widely used today, and we aim to present a data mining process to those who will work in this field. Phyton machine learning language, which has been widely used recently, has been preferred for the application. Spyder and Jupyter Notebook environments are used as interfaces. Different training sets are obtained by preprocessing the telecommunication dataset using Python. Then, various algorithms of the classification method are applied to each of the training sets. After the established models, the results of the predicted classifications are evaluated according to the confusion matrix measures, compared and the most applicable method is presented. To evaluate the performances of the applied models, the sensitivity, precision and F-measure have been used. The confusion matrix representation is shown in Table 2.

Table 2. *The representation of confusion matrix [9].*

		Predicted	
		Non-churners (S = 0)	**Churners (S = 1)**
Initial Set	**Non-churners (S = 0)**	TN	FP
	Churners (S = 1)	FN	TP

The explanation of the items in confusion matrix are as follows [9]:

- **True Positive (TP):** The number of customers those are churners and the model has predicted them correctly.
- **True Negative (TN):** The number of customers those are non-churners and the model has predicted them correctly.
- **False Positive (FP):** The number of customers who are non-churners but the model has labelled or identified them as churners.
- **False Negative (FN):** The number of customers who are churners but the model has labelled or identified them as non-churners.

4.1 DATASET

IBM's open-access dataset was used in this study [64]. Each row in the dataset represents a customer and each column contains data for the customer's attributes. The contents of the dataset are as follows:

- The class of customers churning last month is called *Churn*.
- *Services to which each customer subscribes;* phone, multiple lines, internet, online security, online backup, device protection, technical support, and streaming TV and movies.
- *Customer contract and billing information;* how long they've been a customer, contract, payment method, paperless invoicing, monthly fees, and total fees.
- *Demographic information about customers;* gender, age range, spouses, and dependents etc.

There are 7043 subscribers and 21 features for each subscriber in the data set. The binary class variable is defined as churners/non-churners. 1849 customers are labeled as churners. Figure 4 shows the customer loss numbers of the classes belonging to the data set. When we look at the difference between the classes, we see that the data set is not balanced.

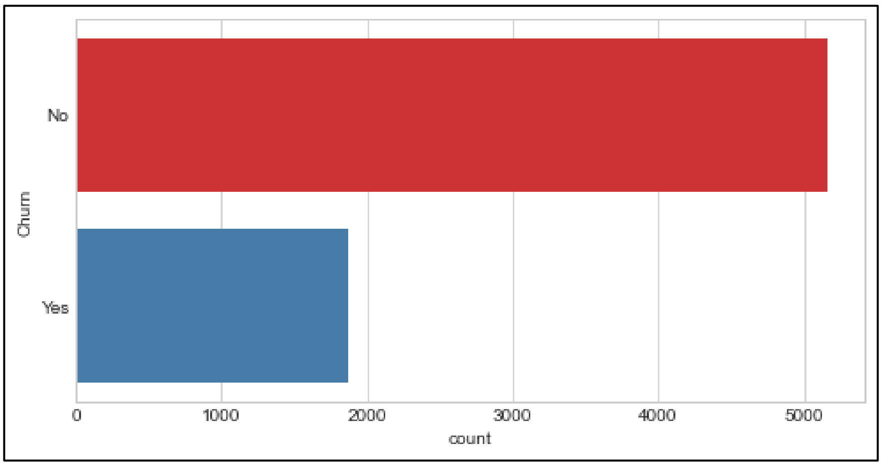

Figure 4. *Customer churn distribution in dataset.*

Dataset preprocessing

The data in the database has been preprocessed. The following operations are carried out sequentially:

- The 'Customer ID' attribute, which has a unique value for each customer in the dataset, is extracted. The reason for this is that the singular values in machine learning do not contribute to the model's training.
- It has been checked whether there are missing values in the data set. As a result of this, it is seen that the 'Total Charges' value is missing in 11 customers. Therefore, these customers are deleted from the data set as their numbers are very few. At the same time, the 'Total Charges' attribute is given categorically, although it is numeric. For this, data type conversion is applied. After this process, there are 7032 rows left in the data set.
- In Python, scikit-learn is one of the most widely used libraries in the field of machine learning. Under the preprocessing sub-library of the scikit-learn library, categorical variables need to be converted into a form suitable for machine learning algorithms at the data pre-preparation stage. For this, label encoder and dummy variable (dummy variable) are used. In this way, attributes with binary values such as yes/no were labeled with values of 0 and 1. The values of the attributes with more than two values were moved to the columns, and their binarization was carried out. For example, the contract column has three different values: monthly, annual and biennial. With the dummy function, 3 more columns have been added to the data set as contract_monthly, contract_yearly and contract_iki_yearly. Whichever contract type a customer has, 1 is printed and 0 is printed on the other contract columns in the relevant row. After this process, the number of columns in the data set reach 41.
- In the next step, Box plots are used as in Figure 5 to visualize the outliers in the data set. The fact that some values on the same data set have values less than 0 and some of them larger indicates that these distances between the data will be more effective on the results, especially the extreme data.

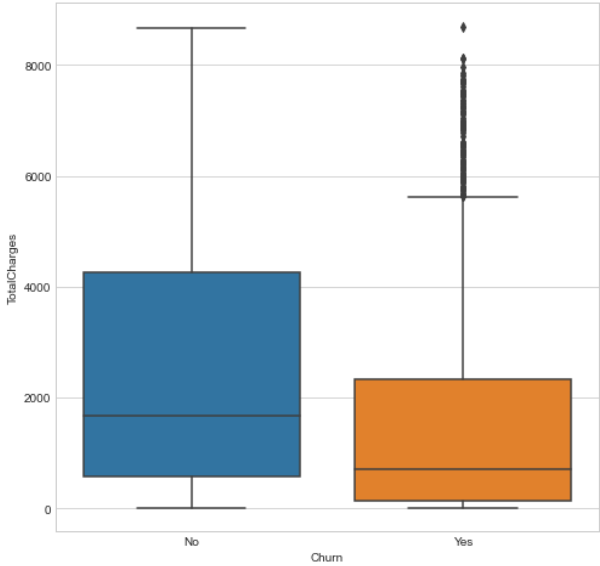

Figure 5. *Visualization of outliers for the 'TotalCharges' attribute.*

The 'Z-score' method is used to determine the extreme values of the features. Outliers have effects that weaken the learning of the model. Eq. (2) is used for the Z-Score equation [65]. After detecting outliers, these values are deleted from the data set to increase the reliability of the analysis. With this process, the number of customers in the data set is decreased to 6353.

$$x' = {(x_i - \mu_i)}/{\sigma_i} \tag{2}$$

x' = Normalized data
x_i = Input
μ_i = Average of input values,
σ_i= Standard deviation of input values
With min-max scaling, all data in the data set were converted to a range of 0 to 1. For this method, the Eq. (3) is used as follows [65].

$$x' = {(x_i - x_{\min})}/{(x_{max} - x_{\min})} \tag{3}$$

x'= Scaled data
x_i = Input value
x_{min} = Minimum number in the input set
x_{max}= Maximum number in the input set

Training and Testing Set Creation

In application, 80% of the data set is determined as the training set and the remainder as the test set. Since the data set is unbalanced, there are class differences in the training set, both unbalanced and balanced data sets are studied in order to see the effect of these differences on the result. One of the following methods can be used to analyze unbalanced data: Less data sample can be increased; excess data sample can be reduced; achievement can be measured according to different performance metrics. First, there are 5081 customers when no action is taken in the training set. Since the dataset is unbalanced, 3711 of them have not left the company, and 1370 of them have left the company. Algorithms are first run with this training set. To understand whether the results obtained here have the best success, balancing operations are performed on the training set.

With the random undersampling feature~~,~~ frequently used in the literature, the class difference in the training set is balanced. In this way, the training set is reduced to 2740 customers. 1370 of these customers do not leave the company, and the remaining 1370 are those who have left the company. Secondly, the algorithms are run with this training set. The disadvantage of the downsampling method is that valuable data can be lost.

Another method used in this study is random oversampling feature. With this approach, we sample random data from the minority class and copy it to generate more samples. In this way, 7422 customers are reached in the training set. These customers consist of 3711 who do not leave the company and the remaining 3711 who left the company. The disadvantage of the oversampling method is that it can increase the probability of overfitting. Table 3 gives information about the customer loss distribution of the training sets to be used in the models.

Table 3. *Number of customers belonging to the training sets studied*.

		Number of non-churners	Number of churners	Total Customer Number
Original dataset	Training set -1	3711	1370	5081
Random undersampling	Training set -2	1370	1370	2740
Random oversampling	Training set -3	3711	3711	7422

4.2 APPLIED MACHINE LEARNING TECHNIQUES

Considering the characteristics of the data we have and the customer's churn behavior, there are some aspects of modeling that should be considered. Customer churn behavior depends on many variables, and the interaction of these variables affects customer behavior. It has been revealed in the studies in the literature that there is a non-linear relationship between the resulting customer churn and the variables used in the prediction. For this reason, Naive Bayes, ID3 (Iterative Dichotomiser 3-Decision Tree), RF, ANN and SVM, which are nonlinear methods used in classification models, are applied. Since the accuracy values in NB and ID3 algorithms are below what they should be, they are not preferred because they do not give reliable results. For this reason, the results of RO, ANN and SVM are emphasized in the study.

4.2.1 Application and Results of Artificial Neural Networks

The Keras package, a deep learning library, has been installed in Phyton. With this package, the sequential functions are called if the model is to be used. By creating a model with the sequential function, layers are added to our model with the model.add() command and the node numbers and activation functions of these layers are determined. The model consists of three layers: The first is the input layer, the second is the hidden layer, and the other is the output layer. When entering the first layer, the number of entries is determined. A neural network architecture consisting of a single hidden layer can successfully model a nonlinear functional form. Therefore, a hidden layer is used in this study. For the number of cells in the hidden layer, the average of the cell numbers in the input and output layers, a frequently used method in the literature has been selected which corresponds to a value of 20 for the study. In the output layer, the class information will be displayed. To compile this model we created, the objective function is determined as binary_crossentropy. This objective function is frequently used in binary classification problems. The optimization algorithm used in this model is 'Adam' and the results of this model are compared according to accuracy value.

Epochs and batch size values are needed to train the model. For these values, the cluster size is given as 200 and the number of revolutions as 100. The model operated with the original dataset has an accuracy rate of 82%. The confusion matrix and other success criteria of this model are as in Table 10 and 11.

__Table 10.__ ANN confusion matrix of the model studied with the original dataset.

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	851	112
	S = 1	123	185

The confusion matrix shows TN=851, TP=185, FN=123 and FP=112. The number of correctly classified non-churners is 851 and the number of churners is 182. In total, 1033 customers are correctly classified. In addition, 112 customers are labeled as actually churners as churners as a result of ANN algorithm but they are actually non-churners and 123 customers are labeled as churners but they are actually non-churners.

__Table 11.__ ANN results of the model studied with the original data set.

	Precision	***Sensitivity***	***F-measure***
S = 0	0.87	0,88	0,88
S = 1	0,62	0,59	0,60
Weighted average	0.81	0.81	0.81

The model run with random undersampling has an accuracy rate of 80%. The confusion matrix of this model and the performance criteria of the classes are as in Table 12 and 13.

__Table 12.__ ANN confusion matrix of the model studied with random undersampling.

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	801	152
	S = 1	83	235

__Table 13.__ ANN results of the model studied with random undersampling.

	Precision	***Sensitivity***	***F-measure***
S = 0	0.90	0.84	0.87
S = 1	0,60	0.73	0,67
Weighted average	0.83	0.81	0.81

The model run with random oversampling has an accuracy rate of 80%. The confusion matrix of this model and the performance criteria of the classes are as in Table 14 and Table 15.

__Table 14.__ ANN confusion matrix of the model studied with random oversampling.

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	738	169
	S = 1	85	279

__Table 15.__ ANN results of the model studied with random oversampling..

	Precision	***Sensitivity***	***F-measure***
S = 0	0.89	0.81	0.85
S = 1	0.63	0.76	0.69
Weighted average	0.83	0.82	0.82

4.2.2 Application and Results of Random Forest Algorithm

The RF, which is an ensemble learning algorithm, of the Sklearn Library in Python has been applied to the datasets. In RF, it is necessary to determine the number of trees (*N*) that the model will create and how many random features (*m*) each tree will have. If the attribute value is significant, it will increase the correlation of the trees and cause underfitting. If the *m* value is less than necessary, it will reduce the correlation of the trees and cause overfitting. In this study, different values for *m* are tried. The most optimal result is taken at the value of 8. The number of trees (*N* value) is a parameter that affects the variance of the estimator. Higher *N* values lead to lower predictor variance, leading to better generalization performance. *N* number is given as 100 in the study. In the trials over 100, the result do not change, but the working time of the model is prolonged.

The model operated with the original data set has an accuracy rate of 80%. The confusion matrix and other performance criteria of RF model are as in Table 4 and 5. Set 0 shows non-churners, and Set 1 shows churners. In this model, the number of correctly classified non-churner customers is 830 and the number of churner customers is 185. In total, 1015 customers are correctly classified. In addition, 112 customers are labelled as churners as a result of RF algorithm but they are actually non-churners and 144 customers are labeled as churners but they are actually non-churners.

Table 4. *RF confusion matrix of the model studied with the original dataset.*

		Predicted	
		S = 0	**S = 1**
Initial Set	**S = 0**	830	112
	S = 1	144	185

Table 5. *RF results of the model studied with the original data set.*

	Precision	***Sensitivity***	***F-measure***
S = 0	0.85	0,88	0,87
S = 1	0,62	0,56	0,59
Weighted average	0.79	0.80	0.80

The model run with random undersampling has an accuracy rate of 75%. This model's confusion matrix and the sets' performance criteria are as in Table 6 and 7.

Table 6. *RF confusion matrix of the model studied with random undersampling.*

		Predicted	
		S = 0	**S = 1**
Initial Set	**S = 0**	673	269
	S = 1	67	262

Table 7. *RO results of the model studied with random undersampling*

	Precision	***Sensitivity***	***F-measure***
S = 0	0.91	0,71	0,80
S = 1	0,49	0,80	0,61
Weighted average	0.80	0.74	0.75

The model run with random oversampling has an accuracy rate of 75%. The confusion matrix and other success criteria of this model are as in Table 8 and 9.

Table 8. *RF confusion matrix of the model studied with random oversampling*

		Predicted	
		S = 0	**S = 1**
Initial Set	**S = 0**	703	239
	S = 1	78	251

Table 9. *RF results of the model studied with random oversampling*

	Precision	***Sensitivity***	***F-measure***
S = 0	0.90	0,75	0,82
S = 1	0,51	0,76	0,61
Weighted average	0.80	0.75	0.76

4.2.3 Application and Results of Support Vector Machines

The SVM algorithm is called under the sklearn library in Python. Radial Basis Function (RBF) is used in the model. This function finds SVM of infinite size and calculates how similar each point is to a certain point with a normal distribution, and classifies accordingly.
The model run with the original dataset has an accuracy rate of 79%. The confusion matrix of this model and the performance criteria of the classes are as in Table 16 and Table 17.

Table 16. *SVM confusion matrix of the model studied with the original dataset.*

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	839	103
	S = 1	168	161

The confusion matrix shows TN=839, TP=161, FN=168 and FP=103. The number of correctly classified non-churners is 839, and the number of churners is 161. In total, 1000 customers are correctly classified. In addition, 103 customers are labeled as actually churners as churners as a result of SVM algorithm but they are actually non-churners and 168 customers are labeled as churners but they are actually non-churners.

Table 17. *SVM results of the model studied with the original dataset.*

	Precision	***Sensitivity***	***F-measure***
S = 0	0.83	0,89	0,86
S = 1	0,61	0,49	0,54
Weighted average	0.78	0.79	0.78

The model run with random undersampling has an accuracy rate of 74%. The confusion matrix of this model and the performance criteria of the classes are as in Table 18 and 19.

Table 18. *SVM confusion matrix of the model studied with random undersampling.*

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	685	257
	S = 1	77	252

Table 19. SVM results of the model studied with random undersampling.

	Precision	*Sensitivity*	*F-measure*
S = 0	0.90	0.73	0.80
S = 1	0,50	0.77	0,60
Weighted average	0.79	0.74	0.75

The model run with random oversampling has an accuracy rate of 73%. The confusion matrix of this model and the success criteria of the classes are as in Table 20 and Table 21.

Table 20. SVM confusion matrix of the model studied with random oversampling.

		Predicted	
		S = 0	**S = 1**
Initial set	**S = 0**	688	252
	S = 1	83	248

Table 21. SVM results of the model studied with random oversampling.

	Precision	*Sensitivity*	*F-measure*
S = 0	0.89	0.73	0.80
S = 1	0.50	0.75	0,59
Weighted average	0.79	0.73	0.75

V. MODEL EVALUATIONS AND FINDINGS

For telecommunication dataset used in the study, firstly the data is preprocessed: incomplete and contradictory data are cleaned, data conversion processes are performed and the data set is scaled in the range of [0-1]. Two balanced training sets are created with the random undersampling and random oversampling methods available in the Phyton imblearn library because of the unbalanced dataset. The samples in the majority class are reduced to equal the number of samples in the minority class with random undersampling and the samples in the minority class are multiplied to equal the number of samples in the majority class with random oversampling. In this way, we run algorithms separately with the unbalanced training set, random undersampling and random oversampling sets. Binary classification problems are compared by applying Naive Bayes, Decision Trees (ID3), RF, ANN and SVM, which are classification algorithms in Python. Naive Bayes and ID3 algorithms are not included in the application part because they give results below expectations. The models are evaluated according to the criteria obtained from the confusion matrix. In the tables below, the results of the models are compared according to the performance metrics. When the results for each model are compared according to the accuracy rates, Table 22 shows that ANN gives better results. RF and SVM algorithms are respectively followed by ANN algorithm. The accuracy rate shows us how many of the customers that the model predicts are classified correctly.

Table 22. The success of the models according to the accuracy criterion.

	Accuracy rate		
	Original data set	*Random undersampling*	*Random oversampling*
ANN	**0.82**	**0.80**	**0.80**
RF	0.80	0.75	0.75
SVM	0.79	0.73	0.74

The sensitivity criterion indicates how many of the model's predictions as non-churners are actually true for the class of non-churners and how many of the model's predictions as churners are actually correct for the churners. The results according to the sensitivity performance metric for classes 0 and 1 are compared in Table 23. In the unbalanced data set, we see that the non-churner customers are predicted at a higher rate. This rate decreased for customers who do not churn after their datasets are balanced. It is decreased from 88% to 81% in ANN, from 88% to 71% in RF, and from 79% to 73% in SVM. The correct estimation of churner customers, while the dataset is unbalanced, is seen as improved when the dataset is balanced. This ratio is increased from 59% to 80% in ANN, from 56% to 80% in RO, and from 49% to 77% in SVM respectively.

Table 23. *The success of the models according to the sensitivity criterion for sets 0 and 1.*

		Sensitivity		
		Original data set	***Random undersampling***	***Random oversampling***
S = 0	*ANN*	**0.88**	**0.84**	**0.81**
	RF	0.88	0.71	0.75
	SVM	0.79	0.73	0.73
S = 1	*ANN*	**0.59**	**0.80**	**0.76**
	RF	0.56	**0.80**	**0.76**
	SVM	0.49	0.77	0.75

Table 24. *The success of the models according to the precision criterion for classes 0 and 1.*

		Precision		
		Original data set	***Random undersampling***	***Random oversampling***
S = 0	*ANN*	0.87	0.90	0.89
	RF	0.85	0.91	0.90
	SVM	0.83	0.90	0.89
S = 1	*ANN*	0.62	0.60	0.63
	RF	0.62	0.49	0.51
	SVM	0.61	0.50	0.50

Table 25. *Success of models according to F-Measure.*

	F-measure		
	Original data set	***Random undersampling***	***Random oversampling***
ANN	**0.81**	**0.81**	**0.82**
RF	0.80	0.75	0.76
SVM	0.78	0.75	0.75

With the above analysis, it is possible to make the following inferences. In models run with all data sets, the ANN algorithm is more successful according to the accuracy rates. However, for customer churn analysis, it may be misleading for a customer to decide to churn based only on the accuracy rate on an unbalanceed data set. Therefore, other performance measures must be considered. Therefore the accuracy rate is proportional to the number of correctly placed customers in both classes. For this reason, two more data sets are created by balancing the unbalanced data set. When these datasets are tested by training with the same algorithms, we see that the accuracy rates are slightly lower than the original dataset. Still, when we look at the sensitivity and precision criteria in predicting the customer who will leave, we see that they are successful.

In our study, we also compare the result of our models with the studies derived from literature. As seen from Table 26, our RF model has the highest accuracy, precision, sensitivity and F-measure values. Our ANN model's accuracy is %82 and Mohammad et al. (2019)'s accuracy is %86, and the sensitivity value of our model is 0,81 and the Mohammad et al. (2019)'s sensitivity value is 0,85. We can tell that their model label more accurately and can predict true positives more accurately. When we evaluate SVM results, the accuracy value of our model is lower than only Chabumba et al. (2021)'s accuracy value. In our SVM model, the other performance metrics are better than the compared studies. When comparing machine learning studies, it should be considered that the preprocessing methods used may differ and will affect the results.

Table 26. Comparison of studies in literature

		Accuracy	Precision	Sensitivity	F-measure
RF	**Our Model**	**80%**	0,79	0,80	0,80
	Halibas et al. (2019) [66]	**75%**	0,71	0,71	0,71
	Tamuka and Sibanda (2020) [67]	**79%**	0,78	0,79	0,78
	Beeharry and Fokone (2021) [68]	**80%**	0,67	0,48	0,56
ANN	**Our Model**	**82%**	0,81	0,81	0,81
	Mohammad et al. (2019) [69]	**86%**	0,78	0,85	NA
	Raut (2020) [70]	**74%**	0,70	0,84	0,76
	Hota and Dash (2021) [71]	**80%**	0,89	0,84	0,85
	Makruf et al. (2021) [72]	**79%**	0,67	0,55	0,60
SVM	**Our Model**	**79%**	0,78	0,79	0,78
	Chabumba et al. (2021) [73]	**80%**	0,68	0,45	0,54
	Makruf et al. (2021) [72]	**78%**	0,66	0,50	0,57
	Lalwani et al. (2022) [74]	**78%**	0,79	0,78	0,78

VI. CONCLUSION

In the study, the performance metrics of the results of different models are compared and it was discussed which method is more successful and applicable. Due to the unbalanced dataset, it is shown which steps are followed in the data preprocessing stages. The differences of the applied models are evaluated for three different training sets developed with different sampling methods. Suggestions are made on which criteria should be considered in the selection of a successful model besides the accuracy rate.

Since the class distribution is not equal in the data set, models are run with three different training sets by oversampling and undersampling. The results of the analysis are shown in Figure 6. When we analyze the accuracy rate of the unsampled artificial neural networks model is higher than the models with over and under sampling. This ratio tells how many of the customers, both will and will not churn, are classified correctly in total, but this model, when considered together with sensitivity, precision and F-measure, has a lower success than other models in predicting only customers who will churn.

According to the accuracy rates, we see that the ANN are followed by the RF in the second place and the SVM in the third place. Models run with original datasets have higher accuracy than balanced datasets. We can say that with the balancing process, the accuracy rates are also reduced for these models, but it has an improving effect for the sensitivity and precision measures. With these results, we see that ANN are more successful than other machine learning algorithms. When we compare the training sets of different models, it is obvious that balanced data sets give better results in classifying the customers who will leave.

The data set used in the application consists of 7043 customers. The working set can be expanded by increasing the number of customers and adding different attributes. In this study, the concept of data mining, its application areas, and machine learning methods are explained and the importance of customer churn prediction is mentioned.

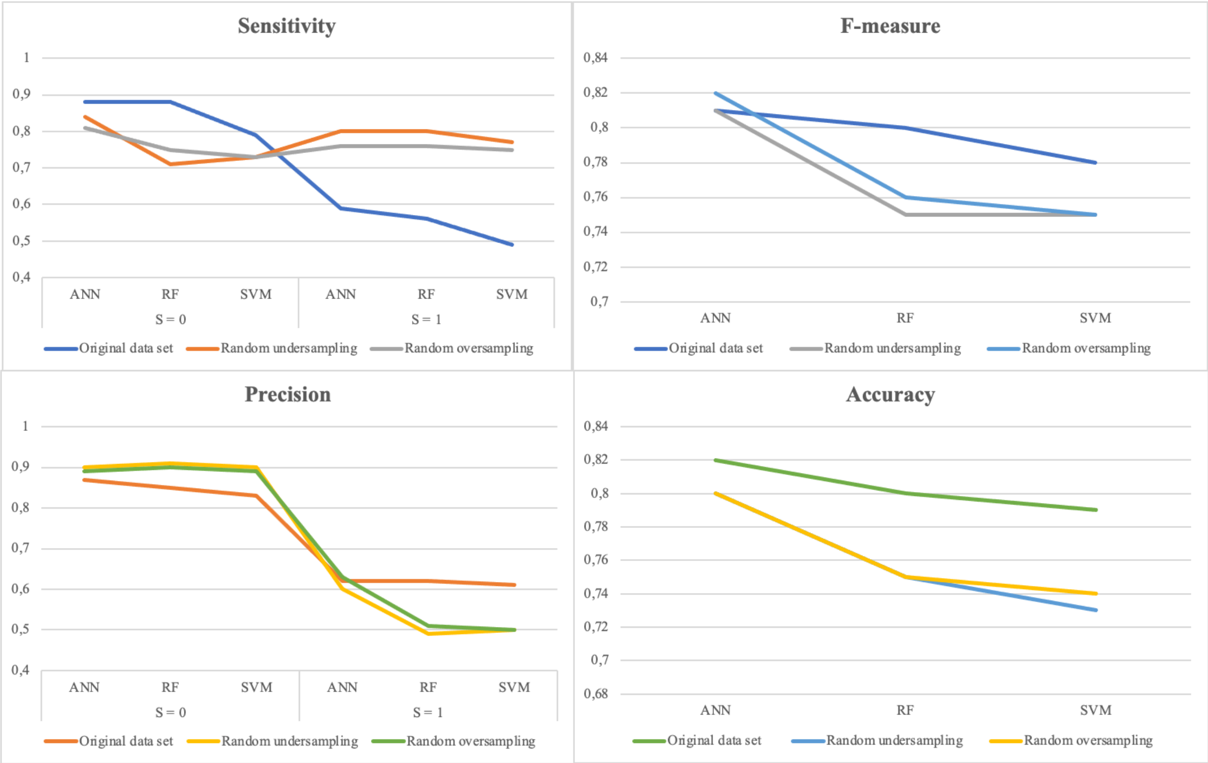

Figure 6. *The comparison of results according to datasets*

The limitations of our study can be explained as follows: in our dataset we don't have indicators which consider customers' technical, quality and price perceptions. Doing so, we can determine customers' requirements and develop some marketing strategies. In addition, the changes in market and economical conditions, customer preferences and telecommunication infrastructure will affect the results of the models. So, by using updated data and shortening the analysis period we can see the effects of these changes and provide quick responses.

In today's technology and telecommunication-based business world, customers are the most important capital of businesses. For this reason, achieving customer loyalty in the telecommunication sector, as in every sector, will enable these enterprises to have a longer life and earn high profits. The analysis made in this study will lead today's managers, who are aware of the cost of losing a customer, to determine their customers who are likely to leave the company and to determine a roadmap for these customers. In this roadmap, each customer profile can be carefully examined and customer-specific strategies can be determined.

VII. REFERENCES

[1] K. K. Tsiptsis and A. Chorianopoulos, Data mining techniques in CRM: inside customer segmentation. *John Wiley & Sons*, 2011.

[2] F. Bagheri and M. J. Tarokh, "Customer Behavior mining based on RFM model to improve the customer relationship management," *J. Ind. Eng. Manag.*, vol.1, no. 1, 2014.

[3] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Churn prediction using comprehensible support vector machine: An analytical CRM application," *Appl. Soft Comput.*, vol. 19, pp. 31–40, 2014.

[4] B. Heinrich and M. Helfert, "Analyzing Data Quality Investments in CRM-A model-based approach," *8th International Conference on Informations Quality (ICIQ), Massachusetts Institute of Technology*, pp. 80-95, 2003.

[5] S. E. Seker, "Müşteri Kayıp Analizi (Customer Churn Analysis)," *YBS Ansiklopedi*, vol. 3, no. 1, pp. 26–29, 2016.

[6] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," *Expert Syst. Appl.*, vol. 31, no. 3, pp. 515–524, 2006.

[7] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," *Comput. Oper. Res.*, vol. 34, no. 10, pp. 2902–2917, 2007.

[8] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *J. Big Data*, vol. 6, no. 1, pp. 1–24, 2019.

[9] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, 2022.

[10] D. Das Adhikary and D. Gupta, "Applying over 100 classifiers for churn prediction in telecom companies," *Multimed. Tools Appl.*, vol. 80, no. 28, pp. 35123–35144, 2021.

[11] N. Alboukaey, A. Joukhadar, and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom," *Expert Syst. Appl.*, vol. 162, p. 113779, 2020.

[12] P. Swetha and B. Dayananda, "Improvised_XgBoost machine learning algorithm for customer churn prediction," *EAI Endorsed Trans. Energy Web*, vol. 7, no. 30, p. e14, 2020.

[13] B. Mishachandar and K. A. Kumar, "Predicting customer churn using targeted proactive retention," *Int. J. Eng. Technol.*, vol. 7, no. 2.27, p. 69, 2018.

[14] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecommunication sector," *IEEE access*, vol. 7, pp. 60134–60149, 2019.

[15] B. Rani and S. Kant, "Semi-supervised learning approach to improve machine learning algorithms for churn analysis in telecommunication.", *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* ISSN 2150-7988, vol. 12, 2020, pp. 265-275

[16] N. Almufadi, A. M. Qamar, R. U. Khan, and M. T. Ben Othman, "Deep learning-based churn prediction of telecom subscribers." *Int. J. Eng. Res.*, ISSN 0974-3154, vol. 12, no. 12, pp. 2743-2748, 2019.

[17] C. R. Jyothi, N. B. Boyina, S.V. Lakshmi, B. Akhila, & A. R. Sai, "A churn detection model in telecommunication using machine learning techniques," *J. Adv. Res. Dyn. Control Syst.*, vol 12, no. 2, pp. 936-943, 2020.

[18] J. Pamina, B. Raja, S. SathyaBama, M. S. Sruthi, and A. VJ, "An effective classifier for predicting churn in telecommunication," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, 2019(a).

[19] W. N. Wassouf, R. Alkhatib, K. Salloum, and S. Balloul, "Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study," *J. Big Data*, vol. 7, no. 1, pp. 1–24, 2020.

[20] A. M. N. Alzubaidi and E. S. Al-Shamery, "Projection pursuit Random Forest using discriminant feature analysis model for churners prediction in telecom industry.," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, 2020.

[21] K. N. Rao, G. V. Nath, K. Tanmaya, B. Roja, B. Pravallika, and S. N. Aslam, "Business KPIs analysis using visual basics," *Int. J. Eng. Innov. Technol.*, vol. 8, no. 6S, 2020.

[22] M. Malleswari, R. Maniraj, and P. Kumar, "Murugan: Comparative analysis of machine learning techniques to identify churn for telecom data," *Int. J. Eng. Technol*, vol. 7, no. 3.34, pp. 291–295, 2018.

[23] R. Suguna, S. Devi, and R. Mathew, "Customer churn predictive analysis by component minimization using machine learning," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 8, pp. 3229–3233, 2019.

[24] A. Isabella Amali, R. Arunkumar, and S. Mohan, "Comparing and evaluating machine learning algorithms for predicting customer churn in telecommunication industry," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 6 Special Issue, 2019, pp. 170-178.

[25] J. B. Raja and S. C. Pandian, "An optimal ensemble classification for predicting churn in telecommunication.," *J. Eng. Sci. Technol. Rev.*, vol. 13, no. 2, 2020.

[26] S. M. Jaisakthi, N. Gayathri, K. Uma, and V. Vijayarajan, "Customer Churn prediction using stochastic gradient boosting technique," *J. Comput. Theor. Nanosci.*, vol. 15, no. 6–7, pp. 2410–2414, 2018.

[27] J. Pamina, T. Dhiliphan Rajkumar, S. Kiruthika, T. Suganya, and F. Femila, "Exploring hybrid and ensemble models for customer churn prediction in telecommunication sector," *Int. J. Recent Technol. Eng.*, vol. 8, pp. 299–309, 2019(b).

[28] H. Jain, A. Khunteta, and S. Srivastava, "Telecom churn prediction and used techniques, datasets and performance measures: a review," *Telecommun. Syst.*, vol. 76, no. 4, pp. 613–630, 2021.

[29] Seema & G.Gupta, "A critical examination of different models for customer churn prediction using data mining," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6S3, pp. 850-854, 2019.

[30] R. Sudharsan, & E. N. Ganesh, "Churn rate prediction in telecommunication systems," Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 4720-4725, 2019.

[31] Turkish Electronic Communications Industry, Quarterly Market Data Report, 2020, 3rd Quarter (Türkiye Elektronik Haberleşme Sektörü, Üç Aylık Pazar Verileri Raporu, 2020, 3. Çeyrek). Retrieved from https://www.btk.gov.tr/uploads/pages/pazar-verileri/uc-aylik-pazar-verileri-2020-3-kurumdisi.pdf

[32] S. Moedjiono, Y. R. Isak, and A. Kusdaryono, "Customer loyalty prediction in multimedia service provider company with k-means segmentation and C4. 5 algorithm," *in 2016 International Conference on Informatics and Computing (ICIC)*, 2016, pp. 210–215.

[33] C.-H. Chen, R.-D. Chiang, T.-F. Wu, and H.-C. Chu, "A combined mining-based framework for predicting telecommunications customer payment behaviors," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6561–6569, 2013.

[34] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Trans. Neural Networks*, vol. 11, no. 3, pp. 690–696, 2000.

[35] A. M. AL-Shatnwai and M. Faris, “Predicting customer retention using XGBoost and balancing methods,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 704–712, 2020.

[36] K. K.-K. Wong, “Getting what you paid for: Fighting wireless customer churn with rate plan optimization,” *J. Database Mark. Cust. Strateg. Manag.*, vol. 18, no. 2, pp. 73–82, 2011.

[37] S. F. Sabbeh, “Machine-learning techniques for customer retention: A comparative study,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, 2018.

[38] S. Mau, I. Pletikosa, and J. Wagner, “Forecasting the next likely purchase events of insurance customers: A case study on the value of data-rich multichannel environments,” *Int. J. Bank Mark.*, 2018.

[39] S. M. Kostić, M. I. Simić, and M. V Kostić, “Social network analysis and churn prediction in telecommunications using graph theory,” *Entropy*, vol. 22, no. 7, p. 753, 2020.

[40] A. Amin et al., “Cross-company customer churn prediction in telecommunication:s A comparison of data transformation methods,” *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, 2019.

[41] Y. Khan, S. Shafiq, A. Naeem, S. Hussain, S. Ahmed, and N. Safwan, “Customers churn prediction using artificial neural networks (ANN) in telecom industry,” *Editor. Pref. From Desk Manag. Ed.*, vol. 10, no. 9, 2019.

[42] P. Spanoudes and T. Nguyen, “Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors,” arXiv Prepr. arXiv1703.03869, 2017.

[43] H. Faris, “A hybrid swarm intelligent neural network model for customer churn prediction and identifying the influencing factors,” *Information*, vol. 9, no. 11, p. 288, 2018.

[44] Y. Zhang, J. Qi, H. Shu, and Y. Li, “Predicting churn probability of fixed-line subscriber with limited information: a data mining paradigm for enterprise computing,” *in Research and Practical Issues of Enterprise Information Systems*, Springer, pp. 589–590, 2006.

[45] A. Mohd Khalid, I. Mohammad Ridwan, M. Mokhairi, A. R. M Nordin, and M. Abd Rasid, “Performance comparison of neural network training algorithms for modeling customer churn prediction,” *Int. J. Eng. Technol*, vol. 7, no. 2.15, pp. 35-37, 2017.

[46] P. Wanchai, “Customer churn analysis: A case study on the telecommunication industry of Thailand,” *in 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, 2017, pp. 325–331.

[47] Y. Koçtürk, “Veri madenciliğinde bağlılık.” İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, 2010.

[48] E. Kızılkaya Aydoğan, C. Gencer, and S. Akbulut, “Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesi,” vol. 26, no. 1, *Sigma Mühendislik ve Fen Bilimleri Dergisi*, 2008.

[49] I. Brânduşoiu, G. Toderean, and H. Beleiu, “Methods for churn prediction in the pre-paid mobile telecommunications industry,” *in 2016 International conference on communications (COMM)*, 2016, pp. 97–100.

[50] T. Kamalakannan and P. Mayilvaghnan, “Optimal customer relationship management in telecalling industry by using data mining and business intelligence,” *Int. J. Eng. Technol.*, vol. 7, no. 1.1, pp. 12–17, 2018.

[51] U. Özveren, “Pem yakıt hücrelerinin yapay sinir ağları ile modellenmesi,” Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2006.

[52] L. H. Tsoukalas and R. E. Uhrig, Fuzzy and neural approaches in engineering. *John Wiley & Sons, Inc.*, 1996.

[53] M. F. Keskenler and E. F. Keskenler, “Geçmişten günümüze yapay sinir ağları ve tarihçesi,” *Tak. Vekayi*, vol. 5, no. 2, pp. 8–18, 2017.

[54] V. Gülpınar, “Yapay sinir ağları ve sosyal ağ analizi yardımı ile Türk Telekomünikasyon piyasasında müşteri kaybı analizi,” *Marmara Üniversitesi İktisadi ve İdari Bilim. Derg.,* vol. 34, no. 1, pp. 331–350, 2013.

[55] E. Alpaydin, Introduction to Machine Learning. *MIT Press*, 2020.

[56] V. Vapnik, The Nature of Statistical Learning Theory. *Springer Science & Business Media,* 1999.

[57] E. U. Küçüksille and N. Ateş, “Destek vektör makineleri ile yaramaz elektronik postaların filtrelenmesi,” *Türkiye Bilişim Vakfi Bilgi. Bilim. ve Mühendisliği Derg.*, vol. 6, no. 1, 2013.

[58] O. Kaynar, M. F. Tuna, Y. Görmez, and M. A. Deveci, “Makine öğrenmesi yöntemleriyle müşteri kaybı analizi,” *Cumhur. Üniversitesi İktisadi ve İdari Bilim. Derg.*, vol. 18, no. 1, pp. 1–14, 2017.

[59] L. Breiman & A. Cutler, Random Forest, Retrieved from http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

[60] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[61] M. Pal, “Random forest classifier for remote sensing classification,” *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.

[62] A. Liaw and M. Wiener, “Classification and regression by randomForest,” *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[63] M. Skurichina and R. P. W. Duin, “Bagging, boosting and the random subspace method for linear classifiers,” *Pattern Anal. Appl.*, vol. 5, no. 2, pp. 121–135, 2002.

[64] Telco Customer Churn Focused customer retention programs data set, [Online]. Available: https://www.kaggle.com/blastchar/telco-customer-churn

[65] Y. Beeharry and R. Tsokizep Fokone, “Hybrid approach using machine learning algorithms for customers’ churn prediction in the telecommunications industry,” Concurr. Comput. Pract. Exp., vol. 34, no. 4, 2022

[66] S. Halibas, A. Cherian Matthew, I. G. Pillai, J. Harold Reazol, E. G. Delvo, and L. Bonachita Reazol, “Determining the intervening effects of exploratory data analysis and feature engineering in telecoms customer churn modelling,” 2019 4th MEC Int. Conf. Big Data Smart City, ICBDSC 2019, 2019

[67] N. Tamuka and K. Sibanda, "Real Time Customer Churn Scoring Model for the Telecommunications Industry," 2020 2nd Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2020, 2020.

[68] Y. Beeharry and R. Tsokizep Fokone, "Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry," Concurr. Comput. Pract. Exp., vol. 34, no. 4, 2022.

[69] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop, and A. Azmi, "Customer Churn Prediction in Telecommunication Industry Using Machine Learning Classifiers," ACM Int. Conf. Proceeding Ser., 2019

[70] N.V. Raut, "A Study of Ensemble Machine Learning to Improve Telecommunication Customer Churn Prediction," Dissertation submitted in part fulfilment of the requirements for the degree of MSc Data Analytics At Dublin Business School, pp. 1–52.

[71] L. Hota and K. Dash, "55_Prediction of Customer Churn in Telecom Industry: A Machine Learning Perspective," Comput. Intell. Mach. Learn., vol. 2, no. 2, pp. 1–9, 2021.

[72] M. Makruf, A. Bramantoro, H. J. Alyamani, S. Alesawi, and R. Alturki, "Classification methods comparison for customer churn prediction in the telecommunication industry," Int. J. Adv. Appl. Sci., vol. 8, no. 12, 2021.

[73] D. R . Chabumba, A. Jadhav, and Ajoodha, R. (2021), "Predicting telecommunication customer churn using machine learning techniques," In Interdisciplinary Research in Technology and Management, pp. 625-636). CRC Press.

[74] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," Computing, vol. 104, no. 2, pp. 271–294, 2022.