A Comparison of Machine Learning Classifiers for Evaluation of Remarketing Audiences in E-Commerce

Haydar Ekelik[1]

Şenol Emir[2]

E-Ticarette Yeniden Pazarlama Kitlelerinin Değerlendirilmesi için Makine Öğrenmesi Sınıflandırıcılarının Karşılaştırması	A Comparison of Machine Learning Classifiers for Evaluation of Remarketing Audiences in E-Commerce
Öz	**Abstract**
Bu çalışmada, Türkiye'de faaliyet gösteren bir e-ticaret sitesinin kullanıcı verileri incelenmiştir. Bu kullanıcılar siteyi daha önce ziyaret eden, yani yeniden pazarlama (remarketing) kitle havuzu içerisinde bulunan kullanıcılardır. Temel amaç, yeniden pazarlama için doğru tahminler yapmak ve böylece yeni ziyaretçiler için özelleştirilmiş reklam içerikleri sunmaktır. Ziyaretçiler, e-ticaret sitesindeki önceki ziyaretlerine göre "alışveriş yapan" ve "alışveriş yapmayan" olarak etiketlendirilmiştir. Veri seti, eğitim ve test kümeleri olarak birbiriyle kesişmeyen iki bölüme ayrılmıştır. Tahmin yapmak için Yapay sinir ağlarına, sınıflandırma ve regresyon ağaçlarına (CART) ve rassal ormana (random forest) dayalı üç sınıflandırma modeli oluşturulmuş ve sınıflandırma performansları karşılaştırılmıştır.	In this study, user data of an e-commerce site operating in Turkey is examined. Users are those who have visited the site before, that is, they are in the remarketing audience pool. The main goal is to make accurate predictions for remarketing and thus offer customized ad packages for new visitors. Visitors are labeled as "Shoppers" and "Non-shoppers" based on their previous visits. The data set is divided into two portions that do not intersect with each other as training and test sets. Three classification models based on artificial neural networks, classification and regression trees (CART), and random forest are built to make predictions and then classification performances of these models are compared.
Anahtar Kelimeler: Yeniden pazarlama, Google Analytics, Yapay Sinir Ağları, Sınıflandırma ve Regresyon Ağaçları, Rassal Orman	**Keywords:** Remarketing, Google Analytics, Artificial Neural Networks, CART, Random Forest
JEL Kodları: C45, M37	**JEL Codes:** C45, M37

---

[1] Arş. Gör., İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, haydar.ekelik@istanbul.edu.tr

[2] Dr. Öğr. Üyesi, İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, senol.emir@istanbul.edu.tr

1. Introduction

Today, as the Internet becomes an inevitable tool, digital marketing plays an important role in the decision-making processes for consumers in purchasing services and products (Charlesworth, 2018: 27). Digital marketing is a type of marketing that is widely used nowadays, reaching consumers using digital channels to promote products or services. There are different types of digital marketing, including social media marketing, display advertising, search engine marketing, and so on (Afrina et. al., 2015).

Companies that enable the purchase of their products and services using digital marketing want to reach their consumers at the right time, in the right place, and at the right price. In this way, they facilitate the purchase of their products and services (Charlesworth, 2018: 27). The marketing strategy of the companies is to inform the consumer, to attract with products and services, to help them in purchasing decisions, and making loyal to the company's brand.

In digital marketing, companies focus on identifying an accurate consumer audience (target audience) based on the needs of consumers considering their marketing strategies (Patrutiu-Baltes, 2016). Marketing strategies are developed according to the target audience and thus ads that are more suitable for the target audience can be presented. Companies advertise on digital platforms (search engines, social media, websites) according to the profiles of their target audience.

In this study, a target audience analysis of users who visited the e-commerce site before via different online tools (search engines, social media, etc.) is performed. These are the users who are in the remarketing audience pool. In the audience analysis, users are divided into two categories as "Shoppers" and "Non-shoppers", according to their previous transactions on the site and after discovering the characteristics of the categories it is tried to predict the category of the new users. For the users that are predicted as "Shoppers", personalized contents are presented to motivate them to purchase. The analysis in the study is based on data obtained from the Google Analytics Platform. The results obtained by implementing Artificial Neural Networks, Decision Trees, and random forest (an ensemble method) methods on this data set are compared. The data set is divided into two distinct parts as training (for creating models) and testing (for testing the results of models). All results presented in the study are the classification performances of the methods on the test set.

2. Literature Review

To our knowledge, there is no study on the investigation of remarketing audiences for Turkish e-commerce sites. In literature, most of the studies are aimed at analyzing customer data with various methods, identifying customer segments, and thus creating effective marketing strategies. Some of these studies are listed as follows:

(Dogan et al., 2020) applied an intuitionistic fuzzy clustering algorithm to customer shopping data for customer segmentation. (Devi et al., 2020) investigated the role of customer segmentation in e-marketing. (Derevitskii et al., 2019) created a weighted graph of co-occurrences of interests of customers based on digital traces and employed Markov based and modularity-based clustering techniques on this graph data for developing marketing products based on segmentation results. They also aimed to identify consistent interest segments for cross-sell campaigns and personalized product suggestions. (Ballestar et al., 2016) analyzed consumer behavior on cashback websites to understand which factors are relevant to the decision of the customers by using structural equation modeling. (Gupta et al.,

2014) proposed a framework to predict the purchase decisions based on adaptive or dynamic pricing of products focusing on customer segments. They used a rich data set containing visit attributes, visitor attributes, purchase history, and web data. (Zhu et al., 2019) analyzed retail precision marketing strategies from the digital marketing perspectives and customer segmentation, market basket positioning, and targeted customer marketing aspects. (Levin et al., 2001) discussed the usage of segmentation to help decisions in marketing. For comparison, they utilized decision tree classifiers and judgment-based classical methods. (Arrigo et al., 2021) aimed to identify target groups by analyzing social media data for reaching an efficient marketing strategy and develop marketing communication. (Safa et. al., 2014) created behavioral profiles of users according to the shopping status by artificial neural networks. (Niu et al., 2017) classified e-commerce customers based on their search depth.

3. Conceptual Framework

3.1 Google Ad Network

Google Ads provides many ways to reach an audience. Using different ad types, ad impressions are made to users who are in the target audience (https://support.google.com/google-ads/).

Google Search Network is a group of search-related websites, images, and apps where ads can be displayed. When ads are advertised to the Google Search Network, ads can appear in search results when a search query related to one of the targeted keywords is submitted.

By using the *Google Display Network*, ads can be displayed in a way that targets a specific content (such as "comic books" or "news"), a specific audience (such as "college students" or "users who want to buy a new bicycle"), or a specific location ("Uludağ"). Google Display Network is designed to help to find the right audience. By employing targeting options, marketing-driven messages ensure potential customers are reached strategically, in the right place, and at the right time.

3.2 Remarketing Process

Remarketing is a way to reach users who have previously interacted with websites or mobile applications. It allows ads to be strategically positioned in front of these audiences on Google or partner websites such as Facebook. In this way, remarketing helps increase brand awareness and help audiences make purchases.

Figure 1: Remarketing Process

Source: https://www.dijiseo.com

In the remarketing process, the user (visitor) first comes to a website and after a while leaves this website. During the visit, a cookie is defined for each user. Then, when the user is on different websites, customized advertisements of the website that the user visited previously are presented. Customized ads are shown according to the cookie information and the categories, products the user visited previously on the e-commerce site. In this way, the user is tried to be brought back to the website.

4. Methodology and Data

4.1 Artificial Neural Networks

The idea of artificial neural networks commonly referred to as "neural networks", began to be developed after noticing that the calculations of the human brain are completely different from the computer calculations. The human brain has the ability to organize its structural components, known as neurons, much faster than the fastest digital computer (Haykin, 2008: 2).

The simplest neuron model, which includes the basic properties of a biological neural network, was proposed by McCulloch and Pitts (1943) and is still the most used model in different artificial neural network architectures (Silva et. al., 2017: 12). Figure 2 shows an artificial neuron model.

Figure 2: Artificial neuron model
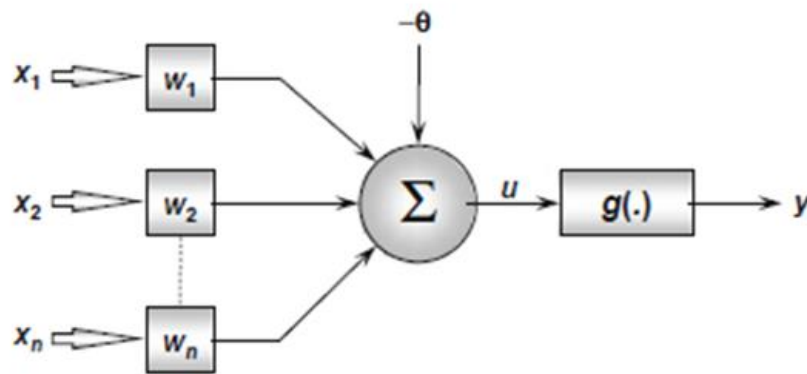
Source: Silva et. al., 2017: 12.

Details of Figure 2 are given as follow:

− **Input signals ($x_1, x_2, \ldots x_n$)** are attribute values of observations coming from the external environment.

− **Synaptic weights ($w_1, w_2, \ldots, w_n$)** are values used to weight each of the input variables. Thus, they quantify the relationships of a neuron with other neurons according to its functionality.

− **The summation function (Σ)** sums up all input signals weighted by the synaptic weights to generate an activation signal.

− **The activation threshold (θ)** is a value used to specify the appropriate threshold required for the result produced by the summation function to generate a firing value towards the neuron output.

−**The activation potential (u)** is the difference between the value of the summation function and the activation threshold. If this difference is positive i.e. u> ϴ then the neuron produces an output, otherwise, the neuron will not produce an output.

−**The activation function (g)** limits the neuron output within a suitable range of values. In other words, the activation function applied to the value of the summation function will produce output in a specified range. The activation function may vary in applications. Generally, sigmoid, linear, step, sinus, and hyperbolic tangent functions are preferred.

−**The output signal (y)** consists of the final value produced by a particular neuron for a given input signal and can also be used as input for other sequential interconnected neurons.

In feed-forward artificial neural networks, neurons are in the form of layers. The outputs of neurons in one layer are used as inputs in the next layer. The general structure of an artificial neural network is presented in Figure 3.

Figure 3: The architectural structure of the artificial neural network
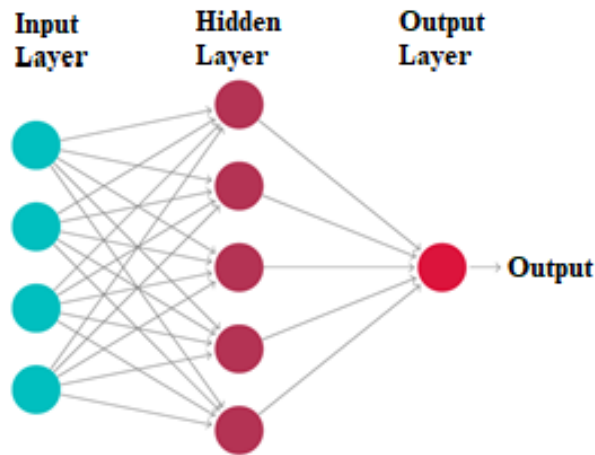
Source: Jäger, 2019.

An artificial neural network model usually consists of 3 layers as input, hidden, and output layers. In some more complicated models, the number of hidden layers may be more than one.

−Input Layer contains the input signals and sends them to the hidden layer.

−Hidden Layer combines weighted inputs. Each neural node in the hidden layer receives data from neurons from the input layer. The values are modified with the applied weights and, the output values are adjusted by an activation function.

−Output layer is the layer that the final output is calculated (Safa et. al., 2014).

Technically, the most basic task of an artificial neural network is to determine outputs that can correspond to a set of inputs (finding a functional relationship between inputs and outputs). To do this task, the network is trained with observations that belong to the relevant

concept. Generalization is the ultimate aim of the learning process. Generalization ability helps to determine outputs corresponding to new observations. For artificial neural networks to run correctly, they must first be trained on the training set and their performance must be tested on a separate test set (Öztemel, 2006: 30).

4.2 Classification and Regression Trees

CART (Classification and Regression Trees) method was developed by Leo Breiman as binary tree structures (Breiman et. al., 1984). In binary tree structures, each node has two branches. Partitioning is carried out by using towing criteria, or the Gini index. If necessary post-pruning is implemented to adjust model complexity. In the CART method branching stops when it is considered that no more information gain can be provided or when stopping criteria are met (Rokach et al.,2015: 79).

The dependent variable in classification trees is categorical, not quantitative. The different properties of the dependent variable make the difference between classification and regression trees. In the classification tree, the class prediction for a node is determined by the majority of classes of observations in that node. If the number of observations of a class is more than the other classes, the prediction of the node is determined as that class. Interpretation of the results of a classification tree deals not only with the class prediction corresponding to a specific termina1l node, but also the class ratios of training observations on those nodes (James et. al., 2014: 311).

In the process of building a classification tree, binary branching (partitioning) is used by the CART algorithm. The classification error rate is used to create binary partitions in the classification tree. In addition to the classification error, Gini index, or entropy methods are preferred in applications (James et. al., 2014: 312).

4.3 Random Forests

The bagging (bootstrap aggregating) method was developed by Leo Breiman in 1996. Bagging creates multiple subsets of the original data set and uses them to obtain an ensemble prediction. In the bagging algorithm, when prediction output is numerical then the average of the predictions is taken, when the output is a class then the majority voting of these classes is taking into account as output (Breiman, 1996). A decision tree is built for each data set obtained by using bootstrap sampling on the data set. These decision trees are parallel and independent of each other. Bagging combines the output of each decision tree into a single prediction to create an ensemble classifier, thereby attempting to improve accuracy. For classifying a new observation, each decision tree creates the class prediction for this observation, and the observation is assigned to the class which has a majority vote. As a result, bagging produces an ensemble model that outperforms a single decision tree (Rokach et al., 2015: 122).

Random forest is an ensemble learning method developed by Leo Breiman (Breiman,2001). It can be considered as an advanced version of the bagging method. In random forests, bagging is used for random variable selection purposes. New training sets are obtained from the original training set by the bootstrap method. Then, a decision tree is created using random variable selection in each new training set. The main difference between random forest and bagging is due to their random variable selection approaches. The Bagging method uses all variables for splitting, while the random forest method employs the random subspace method proposed by Ho (Ho, 1998). For example, in classification

problems, let m shows the number of variables selected for splitting, and p is the total number of variables. If $m = p$ then, it means that bagging is used. In the case of $m = \sqrt{p}$, it means that the prediction method is random forest (James et. al., 2014: 320).

Cross-validation, hold-out, or out-of-bag approaches can be used to estimate the test error of the ensemble model. In the k-fold cross-validation approach, the training set is divided randomly into k equal partition. Model training is performed on k-1 partitions, model performance is tested on the remaining partition. This procedure is repeated k-times and model performance is calculated as the mean of these k test results. The hold-out approach involves randomly partitioning training set into two non-intersecting subsets called a training set and a hold-out set. Training is performed on training set and performance evaluation is estimated on the hold-out set. In the OOB (Out-of-bag) approach, two-thirds of the observations are reserved for training, while the remaining third is reserved for testing (out bag observations), and are not used in training. Thus, obtained OOB error is interpreted as an estimate of the ensemble model's test error (James et. al., 2014:318).

4.4. Model Evaluation - Classification Tables and ROC Curves

In cases such as the data set contains much of the observations in one class (majority class), models may predict the observations as majority class and achieve a good accuracy performance. Therefore, in addition to the classification rate, alternative metrics such as Sensitivity (True Positive Rate) and Specificity (True Negative Rate) are used (Rokach et al., 2015: 36). Sensitivity measures how well the classifier recognizes positive samples, similarly Specificity measures how well the classifier recognizes negative samples. In the test set, the actual class of each observation is compared to the predicted class. A positive observation ("Shopper") correctly classified by the model is called TP (True Positive). A falsely classified positive observation (the user who shopped but was assigned to the Non-shoppers) is called FN (False-Negative). Alike, the user who did not shop but predicted as a Shopper is called FP (False Positive).

Table 1: An example classification table

Classification Table		
	Actual (Negative)	**Actual (Positive)**
Predicted (Negative)	a	b
Predicted (Positive)	c	d

Table 1 shows a standard classification table. By using the values in Table 6 classification metrics mentioned above can be calculated by the following formulas.

$- Accuracy = \frac{(a+d)}{(a+b+c+d)}$

$- False\ Classification\ Rate = \frac{(b+c)}{(a+b+c+d)}$

$- Precision = \frac{d}{(b+d)}$

$$-\ True\ Positive\ Rate(Sensivity) = \frac{d}{(c+d)}$$

$$-\ True\ Negative\ Rate(Specificity) = \frac{a}{(a+b)}$$

$$-\ F_1\ Score = 2. \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

ROC (Receiver Operating Characteristic) curves can be used to show the relationship between True Positive (sensitivity) and True Negative rates (specificity). The area under the ROC curve (AUC) provides important information for classifier performance as it is independent of the selected prevalence value and prior probabilities. The AUC comparison can be used to make a ranking among classifiers. If ROC curves intersect, the total AUC allows an average comparison between models (Lee, 2000). Thus, comparing the AUC values of the two classifiers is fairer and more informative than comparing the misclassification rates (Rokach and Maimon 2015: 39). The closer the AUC value to 1, the better the classifier model is.

4.5 Data

Google Analytics (https://analytics.google.com) provided by Google is a service that offers website owners a large number of metrics for detailed analysis of their visitors. These metrics show the most basic features, such as through which channels visitors reach the web page, which pages they visit, and how long they stay on the pages on average. Among these metrics which show the most basic features, such as which channels visitors reach the web page through, which pages they visit, and how long they stay on the pages on average. E-commerce sites also use this very important data from Google Analytics to make some very essential analyses, such as which categories of products their visitors are most interested in, which pages did the visitors who tend to buy most come to the site from, and at what point in their purchasing process visitors stop buying. Among these metrics, advertisers most consider the "Session", "Average Session Duration", and "Bounces", which have the most determinants of a user's shopping tendency. Explanations of the variables used in the study are given below. The data set was obtained from the Google Analytics platform.

- *Transaction*: This attribute shows the purchase (shopping) made on the site that is valuable to the brand. "Shoppers" are represented as 1, and "Non-shoppers" are represented as 0.

- *Session*: It is the metric that gives the number of user entries to the advertiser's site as a result of the ad click.

- *Average Session Duration*: The metric showing the average time users spend on the site. For example, if the user has visited the site 5 times and spent 250 seconds in total, the average session duration for this visitor is calculated as 50 seconds.

- *The Number of Exits (Bounces)*: The metric showing the number of times users came to the site, visited a single page, and left the site.

In the study, the data set containing the specified attribute values of 7091 visitors who are within the remarketing audience, that is visitors who have visited the analyzed e-commerce site before, were used. The data set was divided into 2 groups as 80% training and 20% test set. A predictive model was built with the training set and a test set was employed to evaluate this model. The number of observations of the "Shopper" and "Non-shopper" classes

for the dependent variable (Transaction) in the training and test sets is presented in Table 1 and Table 2, respectively.

Table 2: Number of "Shoppers" and "Non-Shoppers" in the training set.

0:(Non-shoppers)	1: (Shoppers)	Total
3.858	1.812	**5.670**

Table 3: Number of "Shoppers" and "Non-Shoppers" in the test set.

| 0:(Non-shoppers)	1: (Shoppers)	Total
977	444	**1.412**

5. Results and Analysis

Various R packages were utilized to create the models. These packages are *nnet* (Venables et. al., 2002) for artificial neural networks, *rpart* (Therneau et. al., 2019) for CART, and *randomForest* (Liaw et. al., 2002) for random forest.

5.1 Artificial Neural Networks Results

In the artificial neural network used in the analysis, there is a single hidden layer with 2 nodes. When creating the model, the weight reduction parameter *(decay)* is set to 0.005, activation function is sigmoid and the maximum number of iterations *(maxiter)* is 500 in the nnet R package. The resulting architecture is shown in Figure 4.

Figure 4: ANN Network Topology
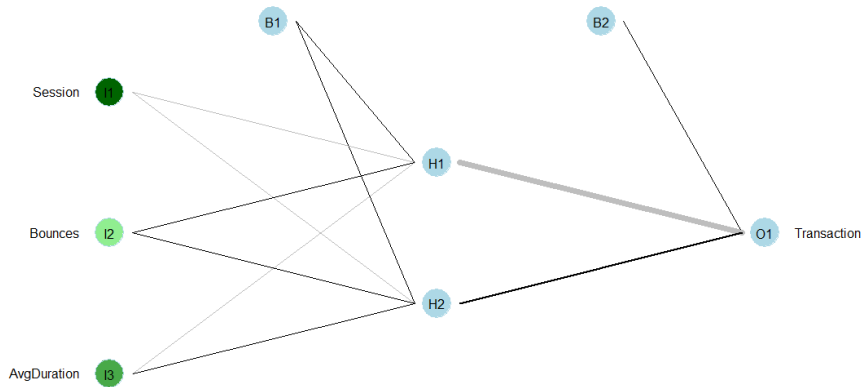
In the data set, the ratio of users who shop to total users is 0.31 so the prevalence (threshold value) is determined as 0.31 for classification. Classification tables for the results of artificial neural networks, on training and test data are presented in Table 4 and Table 5, respectively.

Table 4: Classification results for Artificial Neural Networks in Training Data

Predicted	Actual	
	0	1
0	3130	217
1	728	1595

Table 5: Classification results for Artificial Neural Networks in Test Data

Predicted	Actual	
	0	1
0	791	47
1	196	397

ROC curves and AUC values for the results of artificial neural networks, on training and test data are presented in Figure 5 and Figure 6, respectively.

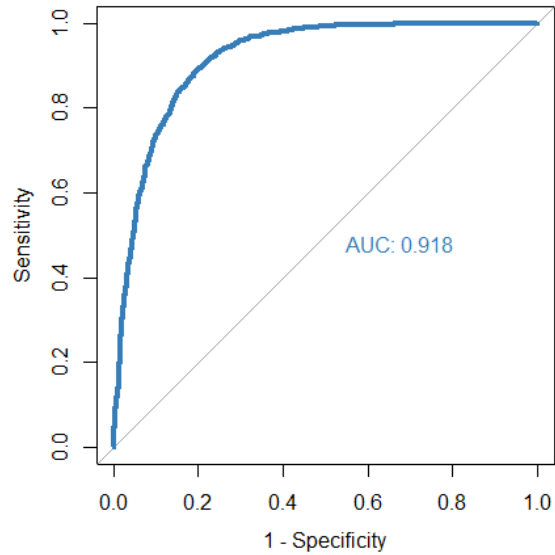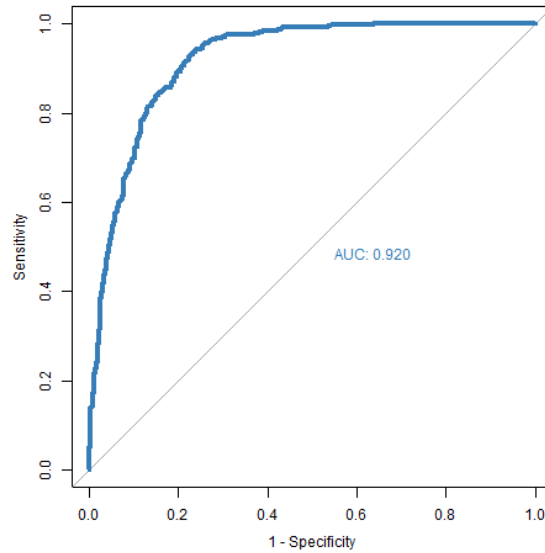Figure 5: ROC Curve for Artificial Neural Networks in Train Data

Figure 6: ROC Curve for Artificial Neural Networks in Test Data

5.2 CART Results

The following parameters are used for creating classification and regression trees in the rpart package. The *minsplit* value indicating the minimum number of observations on the leaves is set to 20 and the *cp value* (a parameter for deciding to split or not based on the decrease in Gini index) is selected as 0.01. Thus, if there is no reduction in the Gini index of 0.01 the split will not take place. Taking these parameters into consideration, the model is trained and the following results are obtained. Figure 7 presents decision tree generated by the CART algorithm.

Figure 7: CART Decision Tree
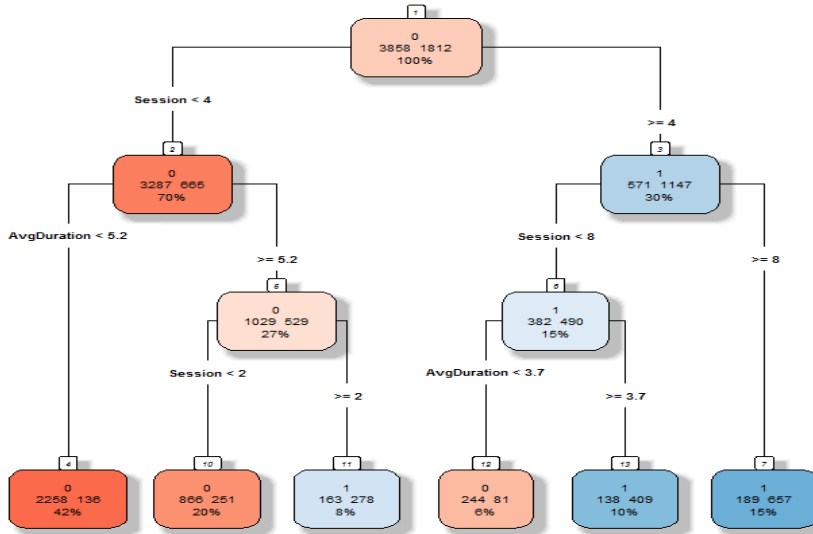
The AvgDuration variable in Figure 7 represents the logarithm of the Average Session Duration variable. It can be seen that the most important variable (root node) is "Session". Splitting criteria (Sesion<4) indicates whether the number of visiting times of a visitor is less than 4 or not. If the Session variable has a value less than 4, the left branch will be selected and splitting is continued on this branch. Here, branching is performed according to the AvgDuration variable. If the AvgDuration variable is less than 5.2 (181 seconds), the terminal node is reached, and it is seen that 136 users are shopping and 2258 users are not shopping. Similar interpretations can be made for other nodes and splitting.

In the data set, ratio of users who shop to total users is 0.31 so the prevalence (threshold value) is determined as 0.31 for classification. Classification tables for the results of CART, on training and test data are presented in Table 6 and Table 7, respectively.

Table 6: Classification results for CART method in Training Data

	Actual	
Predicted	**0**	**1**
0	3368	468
1	490	1344

Table 7: Classification results for CART method in Test Data

	Actual	
Predicted	**0**	**1**
0	854	124
1	123	320

ROC curves and AUC values for the results of CART, on training and test data are presented in Figure 8 and Figure 9, respectively. ROC Curve is similar in both training and test data.

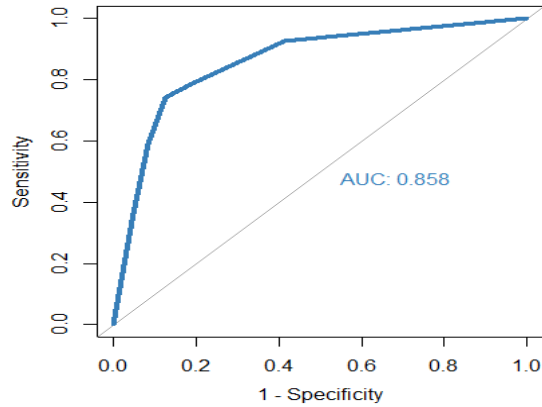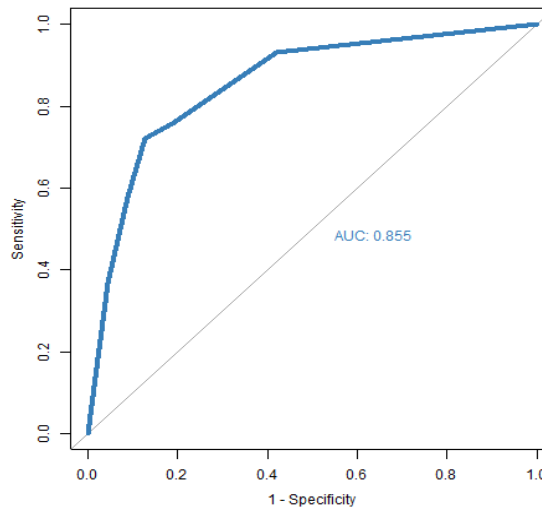Figure 8: ROC Curve for CART in Train Data

Figure 9: ROC Curve for CART in Test Data

5.3 Random Forest Results

In the randomForest R package was used. Bagging method is employed for determining the number of random variables to consider for splitting. Thus, *mtry* parameter to identify the number of random variables to be used in splitting is selected as 3 (total number of predictor variables). *Majority voting* is used for combining the predictions from multiple trees.

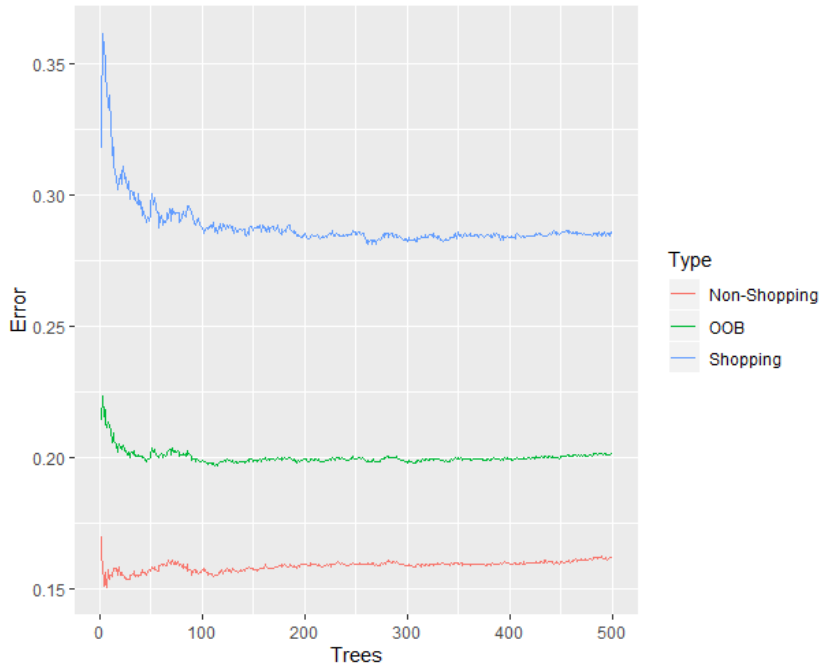Figure 10: Out-of-Bag (OOB) Error for Different Number of Trees

Figure 10 shows the Out-of-Bag (OOB) errors for ensemble models with different numbers of trees (maximum 500 trees). By inspecting the figure, it can be concluded that the minimum OOB error (0.1968) is achieved for the 115 trees (ntree=115).

Taking these parameters into account, the model is trained. Classification tables for the results of Random Forest, on training and test data are presented in Table 8 and Table 9, respectively.

Table 8: Classification results for Random Forest in Training Data

Predicted	Actual	
	0	**1**
0	3477	21
1	381	1791

Table 9: Classification results for Random Forest in Test Data

Predicted	Actual	
	0	**1**
0	779	74
1	198	370

ROC curves and AUC values for the results of Random Forest, on training and test data are presented in Figure 11 and Figure 12, respectively. ROC Curves is similar in both training and test data.

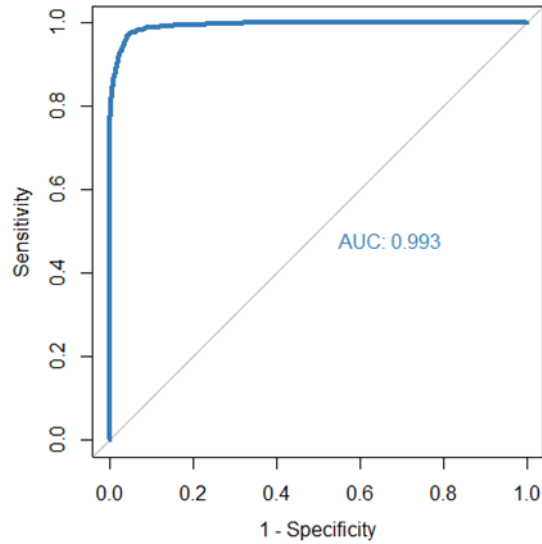Figure 11: ROC Curve for Random Forest in Training Data

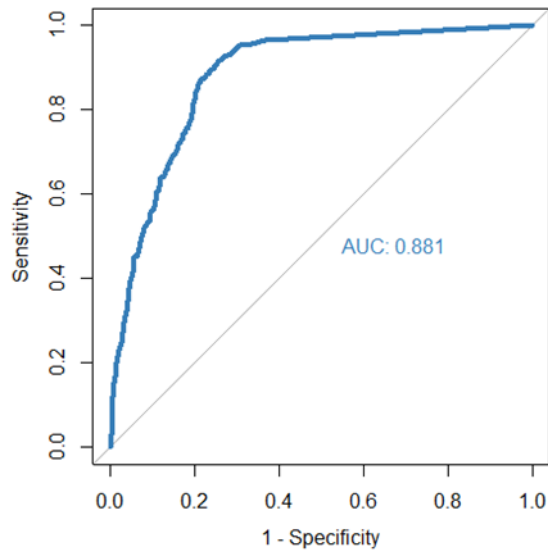Figure 12: ROC Curve for Random Forest in Test Data

Variable importance scores are included in Table 10. The most important variable is the Session variable, as in the CART algorithm. It can be seen that in the absence of Session variable, a large decrease in accuracy value occurs. The Mean Decrease Gini value measures the rate at which the tree's leaves have impurity without the corresponding variable for each

variable. The absence of AvgDuration variable causes the highest reduction in the average Gini value. In other words, having an AvgDuration variable increases impurity in terminal nodes. AvgDuration is the variable that maximally increases impurity in terminal nodes. From Table 10, it can be concluded that the Bounce variable is less important than the other two variables.

Table 10: Variable Importance Scores

Variables	Mean Decrease Accuracy	Mean Decrease Gini
Session	176.13	951.44
Bounces	9.36	163.84
AvgDuration	53.14	1190.84

The results of the three methods for different classification evaluation metrics on the training dataset are presented collectively in Table 11.

Table 11: Summary of results for different metrics in Training Data

Method	Threshold (Prevalence)	Sensitivity	Specificity	Precision	F1-Score	Error	Accuracy	AUC
Artificial Neural Networks	31.25%	88.02%	81.13%	68.66%	77.15%	16.67%	83.33%	91.80%
CART	31.25%	74.17%	87.30%	73.28%	73.72%	16.90%	83.10%	85.80%
Random Forest	31.25%	98.84%	90.12%	82.46%	89.91%	7.09%	92.91%	99.30%

By inspecting Table 11, methods can be compared with each other for different performance evaluation metrics on the training dataset. Random Forest provides the best training performance according to accuracy and AUC criteria. The results indicate that the Random Forest algorithm is more successful than other methods in detecting the relationship between predictor variables and output variable in the training set.

Similarly, the results of the methods for all metrics on the test dataset are presented collectively in Table 12.

Table 12: Summary of results for different metrics in Test Data

Method	Threshold (Prevalence)	Sensitivity	Specificity	Precision	F1-Score	Error	Accuracy	AUC
Artificial Neural Networks	31.25%	89.41%	80.14%	66.95%	76.57%	16.98%	83.02%	92.0%
CART	31.25%	72.07%	87.41%	72.23%	72.15%	17.38%	82.62%	85.5%
Random Forest	31.25%	83.33%	79.73%	65.14%	73.12%	19.14%	80.86%	88.1%

By inspecting Table 12, methods can be compared with each other for different performance evaluation metrics. For example, artificial neural networks give the highest AUC score followed by the random forest method. The algorithm with the highest specificity and precision is the CART algorithm. If the only aim is classification, artificial neural networks can be used. However, decision trees can be preferred if the aim is to explain the model and interpret the relationship between independent variables more easily. Due to the internal properties of decision tree methods, they produce more interpretable results than other methods.

5. Conclusions

In this study, we try to classify the users in the remarketing group. Results of the three methods (artificial neural networks, CART, and random forest) based on different metrics (classification rate, sensitivity, specificity, precision, and AUC value) are compared. The algorithm with the highest AUC value is artificial neural networks, followed by the random forest method. The algorithm with the highest specificity and precision is the CART algorithm. As can be seen from the results, there is not an algorithm with the highest values in terms of all metrics. The performance of algorithms may vary depending on the data set used. It is known that an algorithm cannot be the best in all domains and all data sets (Mitchell, 1980). According to the no-free lunch theorem, if one algorithm is better than the other algorithm in some data sets, there are certainly other data sets in which the opposite result is obtained. Besides, this theorem mentions that a method can give more information than other methods for a specific purpose (Wolpert et. al., 1995).

We see that the AUC value obtained from the artificial neural network algorithm is higher. If the primary aim is classification, artificial neural networks can be used. On the other hand, in addition to classification, if the aim is also to interpret the results more easily, decision trees can be preferred. Because interpretations of decision tree results are much easier than the other methods.

In conclusion, it is possible to classify a new user of an e-commerce platform as "Shopper" or "Non-shopper" by employing a prediction model and by this way to present personalized contents based on the predicted class. If the user is predicted as "Shopper", content and ad displays will be made more effective to make this user shopping.

Personalized ad impressions or different content can be shown using Google Ads advertising tools to the users who are classified as "Shopper" by classification models. For example, according to the analysis results, Session is the most important variable for the CART method and its critical value is found to be 4. Taking into consideration this result, it can be beneficial to create a remarketing pool or preparing different ad options for users who come to the site 4 or more times.

References

Afrina, Y.; Tasneem, S.; Fatema, K. (2015),"Effectiveness of Digital Marketing in the Challenging Age: An Empirical Study", International Journal of Management Science and Business Administration, Vol. 1, No. 5: 69-80.

Arrigo, E.; Liberati , C.; Mariani, P. (2021), "Social Media Data and Users' Preferences: A Statistical Analysis to Support Marketing Communication", Big Data Research, Vol. 24: 100189

Ballestar, M.; Grau, P.; Sainz, J. (2017), "Customer segmentation in e-commerce: Applications to the cashback business model", Journal of Business Research, Vol. 88: 407-414.

Breiman, L. (1996), "Bagging Predictors", Machine Learning, Vol. 24: 123-140.

Breiman, L. (2001), " Random Forests'", Machine Learning, Vol. 45: 5-32.

Breiman, L.; Jerome H.F; Richard A.O; Charles J.S (1984), Classification And Regression Trees, Chapman & Hall/CRC Texts in Statistical Science Series,

Charlesworth, A. (2018), Digital Marketing A Practical Approach, 3rd Ed., New York: Routledge.

Ciaburro, G.; Venkateswaran, B. (2017), Neural Networks with R, Birmingham-Mumbai.: Packt Publishing.

Derevitskii, I.; Severiukhina, O.; Bochenina, K. (2019), "Clustering Interest Graphs for Customer Segmentation Problems", 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 321-327.

Devi, G.; Das, D. (2020), "Role of Customer segmentation in eMarketing. Solid State Technology", Vol. 63, No. 5:, 6251-6256.

Dogan, O.; Hiziroglu, A.; Seymen, O. (2020), "Segmentation of Retail Consumers with Soft Clustering Approach", (Ed. Kahraman, C.; Cevik Onar, S.; Oztaysi, B.; Sari, I.U.; Cebi, S.; Tolga, A.C.), Intelligent and Fuzzy Techniques: Smart and Innovative Solutions, Springer: 39-46

Gupta, R.; Pathak, C. (2014), "A Machine Learning Framework for Predicting Purchase by Online Customers based on Dynamic Pricing", Procedia Computer Science, Vol. 36: 599-605.

Haykin, S. (2008), Neural Networks and Learning Machines, New Jersey: Pearson Prentice Hall.

Ho, T.K.(1998), "The random subspace method for constructing decision forests", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8: 832-944

Jäger, G. (2019), "Replacing Rules by Neural Networks A Framework for Agent-Based Modelling", Big Data and Cognitive Computing, Vol. 3, No. 4: 51.

James, G.; Witen, D.; Hastie, T.; Tibshirani, R. (2014), An Introduction to Statistical Learning with Applications in R. New York: Springer.

Lee, S. S. (2000), "Noisy Replication in Skewed Binary Classification" Computational Statistics and Data Analysis, Vol. 34: 161-195, Retrieved from https://doi.org/10.1016/S0167-9473(99)00095-X

Levin, N.; Zahavi, J. (2001), "Predictive modeling using segmentation", Journal of Interactive Marketing, Vol. 15: 2-22.

Liaw, A.; Wiener, M. (2002), "Classification and Regression by randomForest", R News, Vol. 23: 18-22

Mitchell, T. M. (1980), "The Need for Biases in Learning Generalizations". Rutgers University, Department of Computer Science.

Niu , X.; Li, C.; Yu, X. (2017), "Predictive Analytics of E-Commerce Search Behavior for Conversion", In AMCIS.

Öztemel, E. (2006), Yapay Sinir Ağları, İstanbul: Papatya Yayıncılık.

Patrutiu-Baltes, L. (2016), "Inbound Marketing - the most important digital marketing strategy", Bulletin of the Transilvania University of Braşov Series V: Economic Sciences, Vol. 9 (58),No. 2: 61-68.

Rokach, L. (2010), Pattern Classıfıcatıon Using Ensemble Methods (Serıes In Machine Perceptıon And Artificial Intelligence-Vol. 75), Singapore: World Scientific Publishing Company.

Rokach, L.; Maimon, O. (2015), Data Mining With Decision Tree Theory and Applications, 2nd Edition, World Scientific Publishing.

Safa, N.; Ghani, N.; M. A., I. (2014), "An artificial neural network classification approach for improving accuracy of customer identification in e-commerce", Malaysian Journal of Computer Science, Vol. 27, No. 3: 171-185.

Silva, I.; Spatti, D.; Flauzino, R.; Liboni, L.; Alves, S. (2017), Artificial Neural Networks: A Practical Course, Switzerland: Springer International Publishing.

Therneau , T.; Atkinson, B. (2019), "rpart: Recursive Partitioning and Regression Trees", R package version 4.1-15.

Venables, W.; Ripley, B. (2002), Modern Applied Statistics ,4th Ed., New York: Springer,

Wolpert, D.; Macready, W. (1995), "No free lunch theorems for search", Technical Report SFI-TR-95-02-010.

Zhu, G.; Gao, X. (2019), "Precision Retail Marketing Strategy Based on Digital Marketing Model", International Journal of Business and Management, Vol. 7: 33-37.

Online Resources

https://ads.google.com/intl/tr_tr/home/how-it-works/, (Access: 20.01.2020).

https://support.google.com/google-ads/answer/1722047, (Access: 20.01.2020).

https://support.google.com/googleads/answer/117120?hl=tr, (Access: 20.01.2020).

https://www.dijiseo.com/google-adwords-remarketing-yeniden-pazarlama-nasil-yapilir/,(Access: 20.01.2020).