
**POLİTEKNİK DERGİSİ**

**JOURNAL of POLYTECHNIC**

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: http://dergipark.org.tr/politeknik

# Evaluation of customer loss analysis for marketing campaigns in the banking sector

# Bankacılık sektöründe pazarlama kampanyalarına yönelik müşteri kayıp analizinin değerlendirilmesi

*Yazar(lar) (Author(s)): Recep DUR[1], Sabri KOÇER[2], Özgür DÜNDAR[3]*

*ORCID[1]: 0000-0001-8192-0172*

*ORCID[2]: 0000-0002-4849-747X*

*ORCID[3]: 0000-0002-4142-4446*

To cite to this article: Dur R., Koçer S. and Dündar Ö., "Evaluation of customer loss analysis for marketing campaigns in the banking sector",***Journal of Polytehnic***, 26(2): 759-764, (2023).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: DUR R., KOÇER S. and DÜNDAR Ö., "Evaluation of customer loss analysis for marketing campaigns in the banking sector", ***Politeknik Dergisi***, 26(2): 759-764, (2023).

**Erişim linki (To link to this article):** http://dergipark.org.tr/politeknik/archive

**DOI:** 10.2339/politeknik.1036034

# Evaluation of Customer Loss Analysis for Marketing Campaigns in the Banking Sector

***Highlights***

- *Data mining*
- *Data attribute extraction*
- *Classification*

## *Graphical Abstract*

*In a study on mobile marketing, consumer attitudes towards mobile marketing were examined. After the raw data was brought into the appropriate format, it was classified using data mining algorithms.*

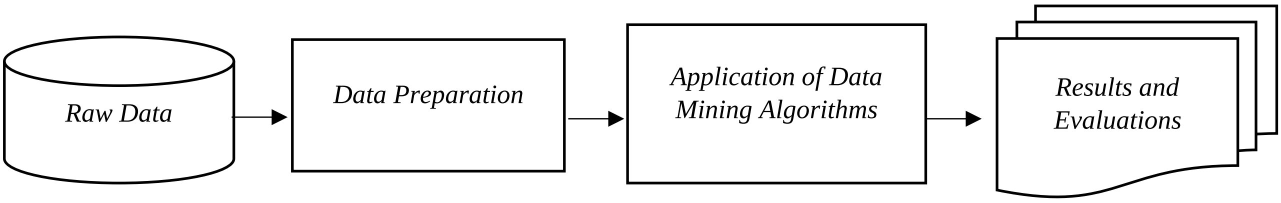


**Figure.** *Data processing steps*

## *Aim*

*To perform campaign analysis in the banking sector using data mining methods.*

## *Design & Methodology*

*Logistic Regression, Artificial Neural Networks and Support Vector Machines classification methods are used.*

## *Originality*

*In the study, 22 marketing data belonging to 29,635 different customers were used.*

## *Findings*

*Accuracy, precision, sensitivity and F-Score values were found to be close to each other.*

*Logistic regression gave much better results in terms of performance compared to other classification methods.*

## *Conclusion*

*While the accuracy, precision, sensitivity and F-Score values of the methods used were close to each other, Logistic regression gave slightly better results than other classification methods.*

## *Declaration of Ethical Standards*

*The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.*

# Evaluation of Customer Loss Analysis for Marketing Campaigns in the Banking Sector

***Araştırma Makalesi / Research Article***

**Recep DUR[1], Sabri KOÇER[2] , Özgür DÜNDAR[3]**
[1]Kuveyt Türk Participation Bank R&D Center, Turkey
[2]Department of Computer Engineering, Necmettin Erbakan University, Turkey
[3]Department of Space and Satellite Engineering, Necmettin Erbakan University, Turkey

(Geliş/Received : 13.12.2021 ; Kabul/Accepted : 28.01.2022 ; Erken Görünüm/Early View : 21.02.2022)

**ABSTRACT**

In this study, customer loss analysis conducted for mobile marketing campaigns in the banking sector. Mobile marketing is a frequently used marketing method, and SMS (Short Message Service) is the most commonly used mobile marketing tool due to its wide range of customers. However, mobile marketing activities may cause customer loss in case of irrelevant and frequent submission if customers don't want to receive advertising notifications. The data set used in the analysis includes 22 attributes belonging to 29,635 customers and class information about whether customers block marketing SMS. The obtained data set was classified by logistic regression, artificial neural networks and support vector machines, and the classification performances of the algorithms were compared. Results show that all three methods have similar accuracy, precision, recall and F-score values while Logistic regression method is slightly better than others.

**Keywords: Classification, customer churn analysis, data mining, machine learning.**

# Bankacılık Sektöründe Pazarlama Kampanyalarına Yönelik Müşteri Kayıp Analizinin Değerlendirilmesi

**ÖZ**

Bu çalışmada, bankacılık sektöründe mobil pazarlama kampanyalarına yönelik müşteri kayıp analizi gerçekleştirilmiştir. Mobil pazarlama sıklıkla kullanılan bir pazarlama yöntemi olup SMS (Short Message Service/Kısa Mesaj Hizmeti) ulaştığı geniş kullanım kitlesi sebebiyle en sık kullanılan mobil pazarlama aracıdır. Bununla beraber alakasız ve sık gönderim gibi etkenlerden dolayı müşteriler reklam bildirimi almak istemeyebilirler ve mobil pazarlama faaliyetleri sebebiyle müşteri kaybı yaşanabilir. Analizde kullanılan veri setinde 29.635 müşteriye ait 22 öznitelikler ve müşterilerin pazarlama SMS'lerini engelleyip engellemediğine dair sınıf bilgisini içermektedir. Elde edilen veri seti lojistik regresyon, yapay sinir ağları ve destek vektör makineleri ile sınıflandırılarak algoritmalara ait sınıflandırma performansları karşılaştırılmıştır. Her üç yöntemin doğruluk, kesinlik, duyarlılık ve F-Skor değerleri birbirine yakın çıkarken Lojistik regresyon diğer sınıflandırma yöntemlerinden az da olsa daha iyi sonuç vermiştir.

**Anahtar Kelimeler: Sınıflandırma, müşteri kayıp analizi, veri madenciliği, makine öğren**

## 1. INTRODUCTION

Today, with the rapid development in technology, enterprises are in search of new marketing activities. Especially the developments in mobile and internet technologies have made new efforts to increase the sales and sales of the products. In this context, mobile marketing is often used in advertising submissions used by businesses.

In the last decade, the use of Data Science, which facilitates decision-making and extraction of information contained in the digital marketing environment, has increased significantly. However, improvements should be made to improve the management of data science in digital marketing [1].

Credit products are an important part of the business of banks and other financial institutions. Messages are sent to customers at certain times to predict the desire to receive a personal loan. Irrelevant dependencies can be eliminated by conducting a time dimension and personal analysis of this approach [2-4].

Mobile marketing is a new marketing method which is evaluated in direct marketing approach aimed at providing communication to the targeted customers via mobile phones and emerged after the internet marketing methods [5,6]. Advertising with SMS is one of the most frequently used mobile marketing methods. Since SMS service is supported by almost all mobile phones, it is an important area in mobile marketing. Sending SMS with irrelevant content to customers for mobile marketing decrease customer satisfaction. In this case, the customer may exercise the right not to receive the marketing SMS by using the exit channels provided by the business. The regulation that the companies are obliged to comply with in the commercial messages to be made for marketing purposes came into force on 01/05/2015 with the Law No: 6563 on Regulation of Electronic Commerce.

**Sorumlu yazar(Corresponding Author)*
*e-posta : e-posta skocer@erbakan.edu.tr*

Within the framework of this legislation, the companies have to provide easy and free cancellation of subscription methods for the marketing SMS [7-10].

Customers who don't want to receive campaign SMS are considered as loss of customers for the businesses. In terms of mobile marketing, customer loss analysis means estimating the customers who are very likely to abandon using customer and campaign features. Customer loss has critical importance in terms of customer continuity for the sectors such as banking, telecommunications, and insurance considering the fact that keeping the existing customers often requires lower operation cost than gaining new customers [11].

In this study, estimation of customer loss is made by using logistic regression, artificial neural networks (ANN) and Support Vector Machines (SVM).

## 2. LITERATURE REVIEW

Customer loss analysis were examined by using Decision Trees and Logistic Regression methods with data pre-processing steps, such as data selection, data merging, cleaning and conversion with the customer data of a private bank [12]. According to the tests, the accuracy rate obtained with this newly developed method is 89%.

In another study, 3 models were obtained by using Support Vector Machines, Naive Bayes and Multilayer Artificial Neural Networks [13]. 75% of the dataset utilized for training purpose and 25% is used to test the model. It was observed that the highest predictive success rate was Artificial Neural Networks with 92.35%. The model with the lowest estimation success is Support Vector Machines with a ratio of 77.89%. This model has been predicted to be unsuccessful due to the small number of samples in the data set and the missing attributes.

Another study examines the effect of data preprocessing steps on customer loss analysis results in the telecommunication sector [14]. In the study, 50% of the data set was used for training, 30% for test, and 20% for selection. Estimation success was increased by 34% when data preprocessing techniques were used. The effects of Logistic Regression, Artificial Neural Networks and Support Vector Machine models with data preprocessing steps also investigated. Results show that Logistic Regression algorithm is faster than Artificial Neural Networks and Support Vector Machine.

Another study conducted a survey in Istanbul in order to examine the factors affecting the consumer attitudes towards SMS advertisements, and opinions of the participants about SMS advertising applications were taken [15]. The results show that being SMS ads are useful, functional and personalized affect the positive attitude towards mobile advertising.

In a study on mobile marketing, consumer attitudes towards mobile marketing were examined [16]. Although the attitudes of the customers towards mobile marketing activities were negative beforehand, it was determined that customer attitudes changed positively due to the fact that smart phones became more widespread and made life easier.

## 3. MATERIAL AND METHOD

### 3.1. Logistic Regression

Logistic regression analysis is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The dependent variable is categorical and the independent variable can be continuous or categorical [17]. Models with the dependent variable has only two categories are known as binary logistic regression analysis [18]. In the studies conducted, the dependent variables usually binary.

Logistics regression differentiates from linear regression having dependent variable has only two possible values. The common feature of both regression methods is they are used to estimate the dependent variable based on independent variables [19].

### 3.2. Artificial Neural Network

Artificial neural networks (ANN), which are inspired by the human brain and are characterized as information processing structures with features similar to biological neural networks [20]

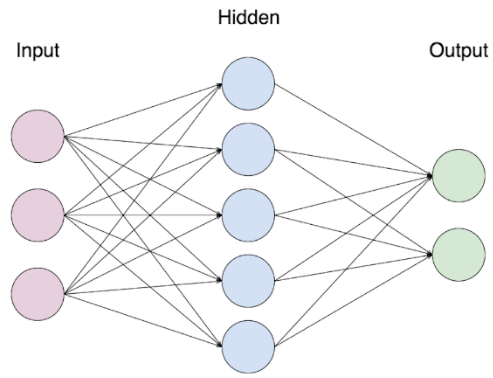


**Figure 1.** Example of artificial neural network

The convergence of neurons creates layers in a general artificial neural network system. As shown in Figure 1, there are three basic layers in an artificial neural network: input layer, hidden layer and output layer.

The input layer is the first layer that allows inputs from outside to be weighted and transmitted to the hidden layer. These inputs are referred as independent variables. The hidden layer is the part between the input and output layers. This layer has no direct connection to the external environment. They simply send the signals coming from the input layer to the output layer. The output layer is the last layer in the neural network. It enables information to transfer to the external environment. These output variables are referred as dependent variables.

In general, artificial neural networks aims such as the human brain data, to train, to learn, to generalize and to work with a large number of variables [21].

### 3.3. Support Vector Machine

Support Vector Machines (SVM) is a simple, effective and supervised machine learning algorithm that is often used for classification problems as well as regression problems. SVM was developed by Vapnik to solve the classification and regression type problems. Its main purpose is to achieve an optimal hyperplane that separates the classes from each other [22]. SVM is evaluated in two cases as linear and nonlinear support vector machines.

One of the important advantages of SVM is that it converts the classification problem into a least square optimization. In this way, the number of transactions in the learning phase is reduced and the performance is faster. This is advantageous in large data sets [19].

## 4. RESULT AND DISCUSSIONS

This study utilizes SMS data sent to the customers of a private bank for the purpose of mobile marketing between 2016-2017. There are 29,635 records in the data set.

Following preprocessing steps taken in order to make the data set meaningful.

i. Empty and damaged records removed
ii. Text converted to lowercase
iii. Turkish characters transformed to English
iv. Verbal attributes converted to numeric data

The specifications and value ranges in the data set are given in Table 1.

The data set consists of two classes representing if the customer is or is not willing to take the marketing SMS sent by the bank. 61% of the data set consists of customers who want to receive marketing SMS, while 39% do not want. The data set distribution is shown in Figure 2.

The distribution of the data set according to Gender and Education Level is shown in Figure 3 and Figure 4. As it can be seen in Figure 3, it is observed that the data set consists mainly of male customers. The level of education is mainly seen as a high school and bachelor's degree.

**Table 1.** Attributes and value ranges in the data set

| Feature | Class / Range | Record Count |
|---|---|---|
| **Age** | < - 25 | 2.686 |
| | 25-35 | 9.922 |
| | 35-45 | 9.152 |
| | > - 45 | 7.875 |
| **City** | 191 City | 29.635 |
| **Education** | Master Degree | 1.174 |
| | Bachelor Degree | 11.374 |
| | High School | 11.632 |
| | Primary School | 5.455 |
| **Has House?** | 0: No | 16.010 |
| | 1: Yes | 13.625 |
| **Has Car?** | 0: No | 22.396 |
| | 1: Yes | 7.239 |
| **Has Child?** | 0: No | 11.144 |
| | 1: Yes | 18.491 |
| **Gender** | 0: Male | 17.564 |
| | 1: Female | 12.071 |
| **Marital Status** | 1: Single | 7.916 |
| | 0: Married | 21.719 |
| **Profession** | 60 Different Jobs | |
| **Total Sms Count** | Between 0 - 24 hours | |
| **At Morning Sms Count** | Between 8 - 11 hours | |
| **At Afternoon Sms Count** | Between 12 - 15 hours | |
| **At Evening Sms Count** | Between 16 - 19 hours | |
| **At Night Sms Count** | Between 20 - 7 hours | |
| **At Monday Sms Count** | | |
| **At Tuesday Sms Count** | | |
| **At Wednesday Sms Count** | | |
| **At Thursday Sms Count** | | |
| **At Friday Sms Count** | | |
| **At Saturday Sms Count** | | |
| **At Sunday Sms Count** | | |
| **Want to Receive Marketing SMS?** | 0: Yes | 18.182 |
| | 1: No | 11.453 |

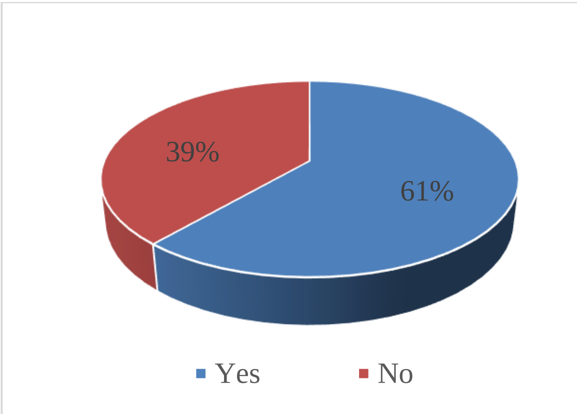


**Figure 2.** The data set classification

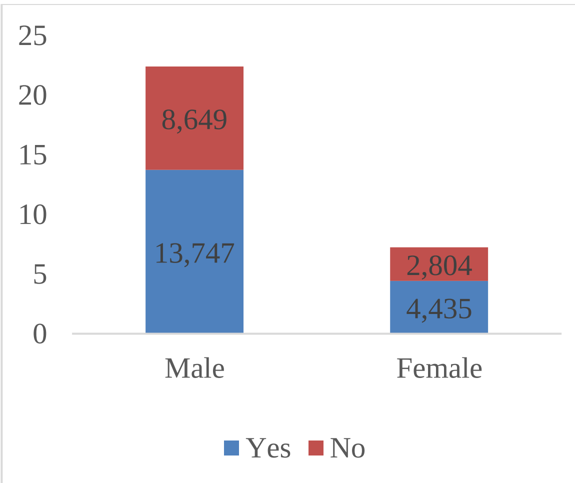


**Figure 3.** Distribution by gender

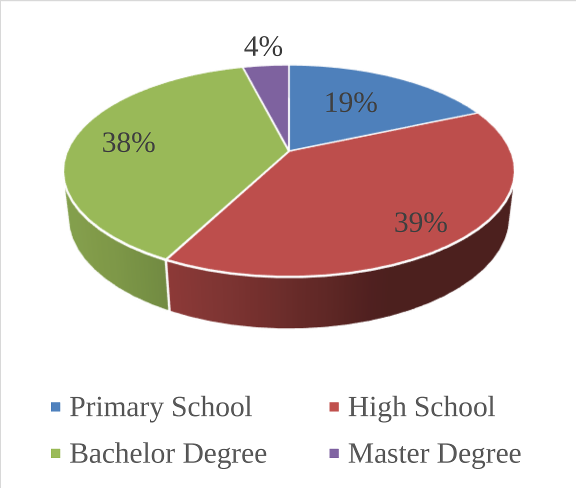


**Figure 4.** Distribution by educational level

3 different models were formed by applying 10-fold cross validation method to logistic regression, artificial neural networks, and support vector machines classification algorithms. Table 2 shows the complexity matrix values for each binary classification model. The descriptions of the values shown in the table are as follows [23-28].

i. True Positive (TP): The actual value is 1 and the estimated value is 1. For our data set; it refers to the number of customers who actually request SMS and so is model result.
ii. True Negative (TN): The actual value is 0 and the estimated value is 0. For our data set; it refers to the number of customers who actually do not want SMS and so is the model result.
iii. False Positive (FP): The actual value is 0 but the estimated value is 1. For our data set; it refers to the number of customers who do not actually request SMS but model result is vice versa.

False Negative (FN): The actual value is 1 but the estimated value is 0. For our data set; it refers to the number of customers who actually request an SMS, but model result is vice versa

**Table 2.** Confusion matrices obtained as a result of classification

| **n= 29.635** | **Estimated Value: False** | **Estimated Value: True** |
|---|---|---|
| | TN: 15.746 | FP: 2.436 |
| **Logistic Regression** | FN: 1.424 | TP: 10.029 |
| **Artificial Neural Networks** | TN: 15.511 | FP: 2.671 |
| | FN: 1.467 | TP: 9.986 |
| **Support Vector Machines** | TN: 15.671 | FP: 2.511 |
| | FN: 1.511 | TP: 9.942 |

Accuracy, Precision, Recall, F-Score values obtained from the confusion matrix are calculated as in Table 3. The following formulas were used for calculation.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

**Table 3.** Classification algorithms comparison results

| Binary Classification Algorithm | Accuracy | Precision | Recall | F-Score | Performance (sec) |
|---|---|---|---|---|---|
| Logistic Regression | 0.870 | 0.805 | 0.876 | 0.839 | 1.38 |
| Artificial Neural Networks | 0.861 | 0.789 | 0.872 | 0.828 | 56.79 |
| Support Vector Machines | 0.864 | 0.798 | 0.868 | 0.832 | 34.38 |

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F-\text{Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

The ROC (Receiver Operating Characteristics) curve and the area underneath are used to evaluate the balance between precision and recall. As ROC approaches 1, positive values are better separated from negative values and their reliability increases [24-26]. ROC curves obtained in the classification process are given in Figure 5, Figure 6 and Figure 7

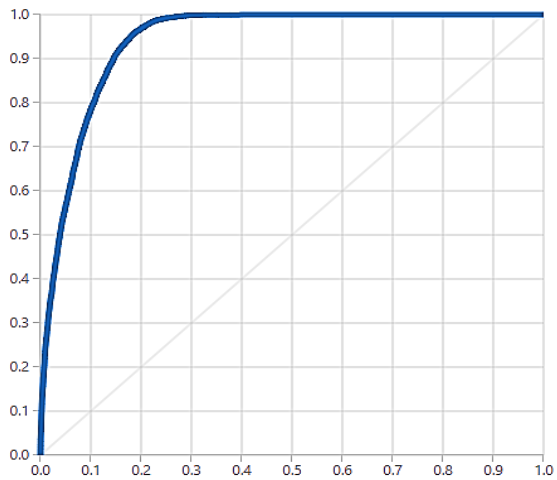


**Figure 5.** ROC curves belong to Logistic Regression

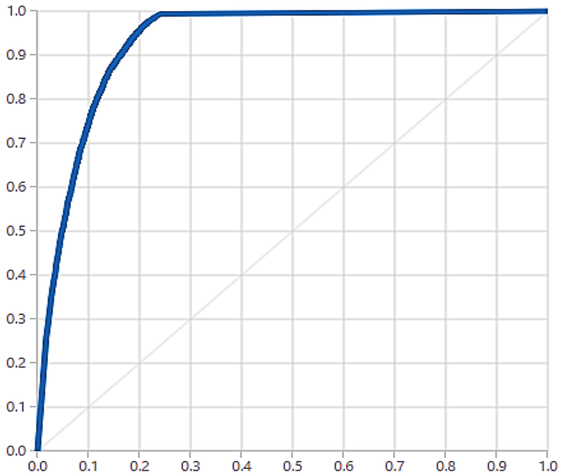


**Figure 6.** ROC curves belong to ANN

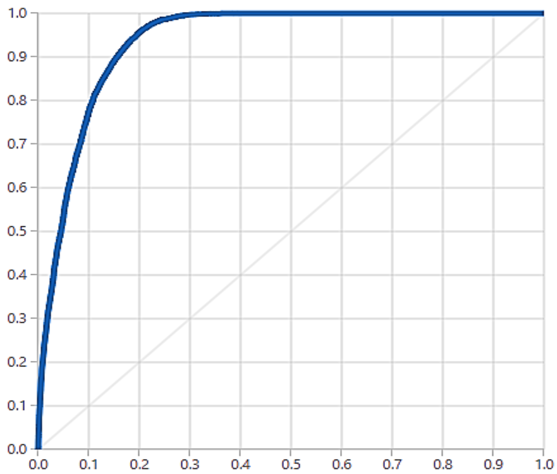


**Figure 7.** ROC curves belong to SVM

## 5. CONCLUSION

Bank direct marketing and business decisions are more important than ever to maintain the relationship with the best customers. Customer service and marketing strategies are needed for success and survival in the enterprise.

Economic conditions affect commercial organizations and banking sectors. Accordingly, marketing managers need to increase their marketing campaigns. The most difficult problems arise due to the large amount of data recording. Data mining has a key role in analyzing this large amount of data. Managers can reshape their business and campaign strategies using data mining tools.

Data mining, which has become quite widespread and important recently, is a tool that allows you to discover valuable information hidden in big data. One of the common uses of data mining is Customer Relationship Management. This method is an approach used to understand the customer's behavior and increase customer satisfaction. The aim of this study is to examine the effects of mobile marketing in the banking sector by using data mining and artificial intelligence methods that have become very common and important in recent years

Customer Loss Analysis using data mining methods is an important area of study that is gaining importance every day. Customer loss analysis allows effective and cost-effective method via preventing customer churn compare to finding new customers and it is getting popular in various sectors, such as telecommunications and banking. In this study, customer loss analysis was conducted for mobile marketing campaigns in banking sector by using Logistic Regression, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The models were compared with five criteria such as accuracy, precision, sensitivity, F-score and performance. The results of the study present that logistic regression classification algorithm is slightly more successful.

The data set used in the analysis includes 22 features of 29,635 customers and class information about whether customers are blocking marketing SMSs or not. The data sets are classified by logistic regression, artificial neural networks and support vector machines and the performance of the algorithms compared.

The results of this study are thought to guide the process mobile marketing in terms of the target population selection and the customer loss analysis. In the next study, it is planned to increase the number of features and sample size of the data set, to use data mining feature selection and classification methods.

## DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

## AUTHORS' CONTRIBUTIONS

**Recep DUR:** Performed the experiments and the analysis of the results. Also, wrote the manuscript.

**Sabri KOÇER:** Conducted the analysis and evaluation of the results.

**Özgür DÜNDAR:** Conducted the analysis and evaluation of the results.

## CONFLICT OF INTEREST

There is no conflict of interest in this study.

## REFERENCES

[1] Saura, J. R. "Using data sciences in digital marketing: Framework, methods, and performance metrics." ***Journal of Innovation & Knowledge***, 6(2), 92-102, (2021).

[2] De Caigny, A., Coussement, K., & De Bock, K. W. "Leveraging fine-grained transaction data for customer life event predictions." ***Decision Support Systems***, 130, 113232, (2020).

[3] Reddy, N. S. "Optimal feature selection and hybrid deep learning for direct marketing campaigns in banking applications", ***Evolutionary Intelligence***, 1-22, (2021).

[4] Ładyżyński, P., Żbikowski, K., & Gawrysiak, P. "Direct marketing campaigns in retail banking with the use of deep learning and random forests." ***Expert Systems with Applications***, 134, 28-35, (2019).

[5] Barutçu S." Mobil Pazarlama, Güncel Pazarlama Yaklaşımlarından Seçmeler, içinde : İ. Varinli, K. Çatı" , ***Detay Yayıncılık***, Ankara; 259-285, (2008).

[6] Arslan, R. S. , Doğru, İ. A. & Barışçı, N. Android Mobil Uygulamalar için İzin Karşılaştırma Tabanlı Kötücül Yazılım Tespiti. ***Politeknik Dergisi*** , 20 (1) , 175-189, (2017).

[7] Elektronik Ticaretin Düzenlenmesi Hakkında Kanun. Resmi Gazete (Sayı: 29166), http://www.resmigazete.gov.tr/eskiler/2014/11/20141105-1.htm [Ziyaret Tarihi: 03.02.2019].

[8] Aydın, M. A. "Müşteri Kaybı Tahmininde Sınıf Dengesizliği Problemi" . ***Politeknik Dergisi***, 1-1, (2021).

[9] Saçan, B. & Eren, T. "Dijital Pazarlama Strateji Seçimi: SWOT Analizi Ve Çok Ölçütlü Karar Verme Yöntemleri." ***Politeknik Dergisi***, 1-1, (2021).

[10] Tekerek, A. "Support Vector Machine Based Spam SMS Detection" ***Politeknik Dergisi*** , 22 (3) , 779-784, (2019).

[11] Reichheld F.F., Sasser E. "Zero Defections: Quality Comes to Services." ***Harvard Business Review***, 68; 105-111, (1990).

[12] Karaağaç Ş. S. "Churn Analysis And Churn Prediction In A Private Bank", ***Yüksek Lisans Tezi,*** Marmara Üniversitesi Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim Dalı, İstanbul, (2015).

[13] Kaynar O., Tuna M. F., Görmez Y., Deveci, M. A. "Makine öğrenmesi yöntemleriyle müşteri kaybı analizi." ***Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi.*** 18; 1 14, (2017).

[14] Coussement K., Lessmann S., Verstraeten G. "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry." ***Decision Support Systems.*** p. 27-36, (2016).

[15] Kıraç S. SMS "Reklamlarına Yönelik Tüketici Tutumları Oluşturan Faktörler," ***Yüksek Lisans Tezi,*** Bahçeşehir Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, (2012).

[16] Watson C., McCarthy J., Rowley J. "Consumer Attitudes Towards Mobile Marketing in The Smart Phone Era." ***International Journal of Information Management.*** 33; 840– 849, (2013).

[17] Hosmer D. W., Lemeshow S. "***Applied Logistic Regression."*** John Wiley & Sons, Inc., New York; (2000).

[18] Bayram N. "***Sosyal Bilimlerde SPSS İle Veri Analizi."*** Ezgi Kitabevi, Bursa, (2017).

[19] Elhan A.H. "Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması," ***Yüksek Lisans Tezi,*** Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, 4-29, (1997).

[20] Fausett L. ***Fundamentals of Neural Networks***. Prentice Hall, USA; (1994).

[21] İslamoğlu E. "Aralık Değerli Zaman Serilerinde Kullanılan Modelleme Teknikleri." ***EÜFBED Fen Bilimleri Enstitüsü Dergisi,*** 8; 178–193, (2015).

[22] Vapnik V.N. "***The Nature of Statistical Learning Theory".*** New York: Springer-Verlag; (1995).

[23] Nitze I., Schulthess U., Asche H. "Comparison of machine learning algorithms random forest, artficial neural network and support vector machine to maximum likelihood for supervised crop type classification", ***Proceedings of the 4th Geobia,*** Brazil. (2012).

[24] Akçetin E., Çelik U. "İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması." ***İnternet Uygulamaları ve Yönetimi Dergisi.*** 5; 43-56, (2014).

[25] Guliyev, H., & Tatoğlu, F. Y. "Customer churn analysis in banking sector: Evidence from explainable machine learning models***". Journal Of Applied Microeconometrics,*** 1(2), 85-99, (2021).

[26] Pamuk, Z., Yurtay, Y., & Yavuzyilmaz, O. "Establishing the potential clients using artificial neural networks". ***Balkan Journal of Electrical and Computer Engineering***, 3, 219-224, (2015).

[27] Polat, H. & Oyucu, S. Heterojen "Medikal IoT Verilerinin Depolanmasında İlişkisel Olmayan Veritabanına Dayalı Bir Yaklaşım" . ***Politeknik Dergisi*** , 22 (4) , 989-998, (2019).

[28] Ibrahım, M. H. "WBBA-KM: A Hybrid Weight-Based Bat Algorithm with K-Means Algorithm For Cluster Analysis" . ***Politeknik Dergisi*** , 1-1 (2021).