International Journal of Management Economics and Business, Vol. 18, No. 3, 2022
Uluslararası Yönetim İktisat ve İşletme Dergisi, Cilt 18, Sayı 3, 2022

Research Article / *Araştırma Makalesi*

FRAUD DETECTION BY MACHINE LEARNING ALGORITHMS: A CASE FROM A MOBILE PAYMENT SYSTEM*

Systems Analyst Özlem GÜVEN
City Card Ege Electronics, İzmir, Turkey, *(ozlemozdemirguven@gmail.com)*

Asst. Prof. Serkan ARAS
Dokuz Eylül University, FEAS, İzmir, Turkey, *(serkan.aras@deu.edu.tr)*

ABSTRACT

With the developing technology, mobile payment systems have become increasingly popular. In the public transport industry, this system has been convenient to the sector in terms of purchasing, using, carrying and storing tickets. One of the greatest challenges encountered in the mobile payment system in this sector is fraud. Fraud reduces customer satisfaction, reduces snow margins and causes severe costs for the company. Therefore, it is very important to detect and prevent fraudsters. This study is based on users using a real mobile ticketing application in USA/Kansas, a customer of Kentkart, which has a smart public transportation system. An automatic and intelligent detection system was developed using a machine learning algorithm to detect whether the users in question are fraudulent or not. For this system, the historical profiles of the variables that represent a user that the risky behavior are created. These profiles are classified using Random Forest, Support Vector Machines, Logistic Regression, K-Nearest Neighbor and Naive Bayes machine learning techniques and results are combined with simple ensemble learning methods. Users classified as frauds are automatically blacklisted in accordance with the company's management policy. Thus, the fraud costs that these users caused the company have been reduced.

Keywords: *Classification, Ensemble Learning, Fraud Detection, Machine Learning, Mobile Payment.*

MAKİNE ÖĞRENME ALGORİTMALARIYLA SAHTEKÂRLIK ALGILAMA: BİR MOBİL ÖDEME SİSTEMİ ÇALIŞMASI

ÖZET

Gelişen teknoloji ile birlikte mobil ödeme sistemleri giderek daha popüler hale gelmiştir. Toplu taşıma sektöründe bu sistem biletlerin satın alınması, kullanılması, taşınması ve saklanması açısından sektöre uygundur. Bu sektörde mobil ödeme sisteminde karşılaşılan en büyük zorluklardan biri sahtekarlıktır. Sahtekarlık, müşteri memnuniyetini azaltır, kar marjlarını düşürür ve şirket için ciddi maliyetlere neden olur. Bu nedenle sahtekarların tespiti ve önlenmesi oldukça önem taşır. Bu çalışma, akıllı toplu taşıma sistemine sahip Kentkart şirketinin müşterisi olan ABD/Kansas'ta gerçek bir mobil bilet uygulaması kullanan kullanıcılara dayanmaktadır. Söz konusu kullanıcıların sahtekar olup olmadığını tespit etmek için bir makine öğrenme algoritması kullanılarak otomatik ve akıllı bir tespit sistemi geliştirildi. Bu sistem için, bir kullanıcının riskli davranışı temsil eden değişkenlerden geçmiş

* *This study was based on the master thesis named Detecting of Mobile Payment Fraud within Machine Learning Tecnique" published in 2021.*

www.ijmeb.org ISSN:2147-9208 E-ISSN:2147-9194
http://dx.doi.org/10.17130/ijmeb.979302
Received: 05.08.2021, Accepted: 28.03.2022

profilleri oluşturulur. Bu profiller Rassal Orman, Destek Vektör Makineleri, Lojistik Regresyon, K-En Yakın Komşu ve Naif Bayes makine öğrenme teknikleri kullanılarak sınıflandırılmış ve sonuçlar basit topluluk öğrenme yöntemleri ile birleştirilmiştir. Sahtekar olarak sınıflandırılan kullanıcılar, şirketin yönetim politikasına göre otomatik olarak kara listeye alınmıştır. Böylece bu kullanıcıların şirkete yol açtığı dolandırıcılık maliyetleri düşürülmüştür.

Anahtar Kelimeler: *Sınıflandırma, Topluluk Öğrenme, Dolandırıcılık Tespiti, Makine Öğrenmesi, Mobil Ödeme.*

1. Introduction

With the developing technology, the concept of mobile payment, which enables shopping with online payment methods by connecting to a mobile network via electronic devices, has quickly entered our lives. Mobile payment is a type of payment for goods, services and invoices using wireless network and other communication technologies through a mobile device such as a mobile phone, smartphone or personal digital assistant (PDA). Payment can be made via mobile devices for various needs such as digital content, tickets, parking, transportation fees or online invoices (Dahlberg, 2008). Using the mobile payment system, it is possible to pay for goods and services quickly and easily. As in almost every sector, this payment method has been adopted in the public transportation sector. The sale and use of tickets through mobile devices have contributed to the ease of use of public transportation. Necessary processes, such as ticket printing, distribution, protection, storage and transportation, could be prevented by exploiting the mobile ticketing system. Thus, users could buy tickets anytime and anywhere with mobile devices, which are a part of their daily lives, carry the purchased tickets on their devices and use them easily through their mobile devices.

The mobile ticketing system provides many advantages to public transport passengers in terms of time, cost and ease of use. However, due to its ease of use, public transport companies faced difficulties in terms of confidentiality and reliability, such as protecting and verifying tickets, preventing violations of the user's personal rights, and keeping data anonymous. As in every system, smart transportation systems also have users who exhibit suspicious behavior, abuse system initiatives and pose risks. The fact that a device that performs transactions such as purchasing, validating and using tickets also contains information about users' daily lives makes the application vulnerable to attacks by fraudulent users. This situation reinforces users' concerns about the disclosure of sensitive data contained in the device containing the ticket and requires action. Otherwise, the intended use of the application may be abused, the application may not reach reliable users, the actual number of users cannot be estimated, the reputation of the application and the company may be damaged, and even greater financial losses may be encountered. Due to such security-related problems in the design of electronic payment systems, electronic commerce requires an effective electronic payment system. For this reason, frauds that occur or may occur in electronic commerce should be prevented and the privacy of the participants should be protected (Camenisch, 1996). Fraud reduces customer satisfaction, lowers profit margins and causes serious costs and prestige losses for the company. Therefore, detection and prevention of fraudsters is very important. Especially leaking information and accessing sensitive data by wrong people is a serious fraud issue. In an ideal payment system, sellers should not learn the real identity of their customers, users should remain anonymous,

and banks should not receive any information about the products their customers buy, other than the price. Order confidentiality and sensitive payment details must be protected from surveillance (Hassinen et al., 2006).

In this study, a method was developed that detects and prevents users who may harm the business in various ways, such as stealing, copying or leaking sensitive data produced and/or stored in the mobile application, accessing personal data, exploiting system vulnerabilities, and using the system with a stolen card or account. The developed method was used in Kentkart, a smart public transportation developer and provider company. Kentkart, a company based in Turkey, provides mass transit systems in many countries. The usage data of the smart transportation mobile application of the company used in USA/Kansas were employed. The purpose of the method was to identify fraudulent users (such as users who abuse application features and promotions, generate fake tickets, distribute or sell them, violate the personal rights of non-fraudulent users, make purchases with stolen credit/debit cards, etc.). In order to prevent fraud, users identified as fraudulent were automatically blacklisted per the company's management policy, thus preventing them from entering the system. Considering that risky users change their behavior over time by adapting to the measures required to be taken, the blacklist structure was created with a learning-based algorithm, not a rule-based one.

There are three basic stages in the developed smart detection system. The first stage is the data evaluation step. At this stage, the necessary variables were selected using the LASSO model from among the variables that make up the profiles of the users, and dimension reduction was performed using the PCA. In the second stage, to classify the behavior as fraudulent or non-fraudulent, Random Forest, Support Vector Machine, Naive Bayes, K-Nearest Neighbor and Logistic Regression algorithms were utilized. To make the final decision, voting techniques were preferred among the ensemble of methods, which are the last stage of the system, and simple majority and soft voting methods were used.

In the first part of the study, literature review about fraud analysis in ticketing and mobile ticketing is included. Then, the data set discussed in the study was introduced in the material and method section, and the machine learning methods discussed were explained. Finally, the results of the case study and the interpretation of the results are given in the results and discussion section. The study was concluded by explaining the general conclusions from the study and the future goals of the study in the conclusions section.

2. Literature Review

The ticket is used to securely verify the identity of the owner and the payment made between the authentication server and the edge server. The digital ticket includes basic necessary information such as the name of the server, the client's name, the client's web address, a random session key that validates the ticket, a timestamp from when the ticket was issued, and the lifetime of the ticket between the time the ticket was first active and the expiration date (Steiner et al., 1988). The features of the digital ticket are explained as follows (Fujimura & Nakajima, 1998):

- Reliable
- Some anonymous, some not

- Portable
- Some transferable, some not
- Can work offline
- Divisible
- Permanent
- Acceptable
- User-friendly
- Can be understood by the machine
- Transition statuses can be managed
- Can be combined

In order for the digital ticket to be fully provided, the ticket and the electronic device on which the ticket is available must be protected in terms of security and privacy. Qin et al. (2017) conducted a security and privacy analysis study on a mobile wallet. As stated, first of all, the acquisition, alteration or misuse of the payment information between the customer and the seller by any attacker should be prevented. Otherwise, the reputation of the mobile wallet and the seller can be seriously damaged due to fraudulent and abused payment information. The system administrator must determine the true identity of the malicious client and the attacker who caused this damage and take the necessary corrective and preventive action. However, the privacy of honest and reliable customers should be protected as much as possible as long as it does not involve any risk (Qin et al., 2017). Many studies argue that the privacy of users and the security of the system must be protected for a mobile payment application to operate effectively (Hassinen et al., 2006; Karnouskos et al., 2004; Linck et al., 2006; Dewan & Chen, 2005). One of the published works on this subject was done by (Pirker & Slamanig, 2012). A prepaid strategy was proposed to achieve both privacy and reliability in online applications with both micro and macro mobile payment systems (Pirker & Slamanig, 2012).

Regarding protection of the security of the system, some studies have presented complex keys randomly generated by cryptological transactions and privacy mechanisms for mobile payments as a solution (Hwang et al., 2006; Hashemi & Soroush, 2006). For the protection of the user, the necessity of fraud detection and prevention was generally advocated (Wang et al., 2016; Vlasselaer et al., 2015; Chan et al., 1999). Polla et al. (2012) argued that for fraud detection and prevention, the roles of actors in the application should be examined and the threat mechanism of each actor should be determined. Usually using fraudulent, false or misleading representation, they tend to hide their activities to prevent detection for as long as possible to maximize the effects of their fraudulent behavior (Behdad et al., 2012). For this reason, it becomes difficult to detect and prevent people attacking the system by eye, and some prevention and detection mechanisms are needed. A successful attack with today's web-based economic resources negatively affects consumer confidence and decreases consumers' desire to shop electronically (Erbacher et al., 2002).

Many security mechanisms have been developed to protect mobile payment security and privacy. Despite all the efforts made to protect them, mobile payment systems still face security challenges, such as malware detection, database attack prevention, multi-factor authentication,

identity and data breach prevention, fraud detection and prevention (Wang et al., 2016). Fraud prevention has been described by (Bolton & Hand, 2002) as stopping fraud at the first place it occurs. On the other hand, fraud detection is the process of detecting the fraud as quickly as possible after the fraud is committed. Fraud detection is activated after fraud prevention fails.

In a research study, approaches related to the prevention and detection of fraud in small businesses were discussed and fraud risks in the business were discussed in terms of employees, managers, executives, and owners (Johnson & Rudesill, 2001). Particularly, financial fraud analysis applications involving payment have been handled with very different methods in the literature. For example, in one study, binary probit and logit models and multivariate statistical techniques were used to detect false financial statements (Küçükkocaoğlu & Küçüksözen, 1997). In the same study, Artificial Neural Network Model to predict manipulative financial reporting practices of companies traded in Istanbul Stock Exchange (ISE). The results are modeled in the proposed Neural Network. In another financial fraud focused study, simulation study for fraud risk assessment process in financial statement audits was handled and a simulated fraud brainstorming session was prepared (Hess & Andiola, 2018).

Although there are many methods used for fraud detection in the literature (Wang, 2010; Jyothsna et al., 2011), the detection mechanism was frequently performed by machine learning methods, because detection is based on an intelligent and automatic system (Adewumi & Akinyelu, 2017; Ghosh & Reilly, 1994). The fraud detection problem from a machine learning point of view is a supervised classification task. Classification is done to determine if a new transaction is true or fraudulent. The training phase of supervised machine learning algorithms is designed for the corrected model that generalizes the class distribution (Acosta et al., 2017). Supervised learning can work with many algorithms. Classification algorithms such as K-Nearest Neighbor, Logistic Regression, Naive Bayes, Linear Regression and Nonlinear Regression algorithms are examples of algorithms working with supervised learning. (Abdallah et al., 2016). These methods are also frequently used in the literature for fraud detection. For example, in a study for mobile payment fraud detection, EM, KMeans, Farthest First, XMeans and MakeDensity clustering algorithms and Naive Bayes, Support Vector machines, Logistic Regression, OneR, C4.5 and Random Forest classification methods were compared. In the related study, a semi-supervised model is proposed by combining supervised and unsupervised models. When the methods used were compared, it was seen that the most effective results were in the regression models, but it was understood that the existing structure was not sufficient for the increasing fraud rates with the increasing number of transactions (Choi & Lee, 2017). On financial area, Support Vector Machine and Regression Tree methods have been used in the literature to detect fraudulent financial statements and protect the global financial market (Pai et al., 2011). On another financial fraud study, seven different classifiers, namely Support Vector Machine, Naive Bayes, Artificial Neural Network, K-Nearest Neighbor, Random Forest, Logistic Regression and Bagging, are used to detect financial accounting fraud in small and medium-sized enterprises (SMEs) (Hamal & Senvar, 2021). In another financially focused study using machine learning, credit card fraud was considered as a data mining problem in financial fraud and Naive Bayes, K-Nearest Neighbor and Logistic Regression performance were examined (Awoyemi et al., 2017). In this study, classification methods in machine learning, which are frequently used in the literature, have been applied to detect fraudulent transactions. Considering the literature, Naive Bayes, K-Nearest Neighbor, Logistic Regression, Random

Forest and Support Vector Machines methods, which have successful results in the literature, were preferred especially in financial fraud detection, since this study also focused on frauds in the mobile payment system. Contrary to most studies in the literature, instead of choosing one of the machine learning algorithms used in order not to cause information loss and to maximize the benefit provided by the models, ensemble methods were used by combining the best results.

3. Materials and Methods

3.1. Data Sets

In the study, the data obtained from Kentkart's US/Kansas customer were employed. The profiles of all users who created an account in the mobile application were considered for use in automatic blacklist management. There were 38 independent variables in total, consisting of queries that showed whether the users had a risk within the system. These variables contained categorical, binary and quantitative values. The dependent variable is a binary variable indicating whether the user is on the blacklist, consisting of 0-1 which is defined below as:

- 0: user not being on the blacklist (non-fraudulent users)
- 1: user being on the blacklist (fraudulent users)

When users created an account on the mobile application, they started as whitelist users in the application. Then, if the risky actions they exhibited were caught by rule-based controls, they were blacklisted. 1.475% of all users were blacklisted through manual controls. There were 108427 users registered in the mobile application. 1600 out of these users were on the blacklist. The remaining 106827 users were in the category of non-fraudulent users. Due to the class imbalance problem in the data set, the under sampling method was utilized and this new sample set was used rather than the main population. Therefore, 1600 people known to be on the blacklist and 6500 people who were known not to be on the blacklist were created.

The prepared sample was divided into two as training and test data sets to be used in machine learning. In order to ensure that the training data was balanced and the study provided the maximum benefit from the blacklist users, 1500 users were taken from the blacklist and 1500 users were taken from the whitelist as innocent users. The remaining 5100 users were included in the test set. In the population, users on the blacklist account for 1.475% of all users. Around 1.475% of the users in the test data should be on the blacklist to match the distribution of the population. Considering this, approximately 2% of the test data were chosen to be on the blacklist. Therefore, there were 5000 non-fraudulent users and 100 fraudulent users in the test set. Table 1 shows the number of fraudulent and non-fraudulent users in the test and training data sets.

Table 1: The Size of Datasets

	Training	Test	Total
Non-fraudulent	1500	5000	6500
Fraudulent	1500	100	1600
Total	3000	5100	8100

3.2. Random Forest (RF)

RF is a classifier of ensemble learning, which consists of the outputs of the classes obtained by combining the outputs of many decision trees. It is a decision tree-based learning algorithm that uses the classification and regression tree (CART) methodology proposed by Breiman et al. (1984). Random forests randomly select a subset of explanatory variables at each node by combining several binary decision trees. They use a collection of bootstrap samples taken from the learning sample during the combining phase (Genuer et al., 2010). Random forests differ from standard decision trees in their node division. In standard trees, node splitting occurs using the best separation between all variables. However, in a random forest, node splitting occurs using the best among subsets of randomly selected estimators at that node (Liaw & Wiener, 2002).

3.3. Support Vector Machine (SVM)

SVM is a supervised learning algorithm that can deal with complex data sets (Cortes & Vapnik, 1995). In this technique, the data is positioned on a larger-sized input area and creates an optimal separating hyperplane on this area. This linear classifier system is called the optimal separating hyperplane because it separates the classes equally and at the greatest distance. The plane that produces the greatest difference among classes is called the maximum margin hyperplane. Support vectors represent the data points closest to the maximum margin hyperplane. Each class always has at least one, and often more than one, support vector (Zareapoor et al., 2012). If the plane that provides the maximum margin hyperplane is chosen optimally, this hyperplane maximizes the generalization ability of the classification. This is true under two assumptions. First, there should be no outliers in the training data, and second, the unknown test data should fit the same distribution as the training data (Abe, 2005).

3.4. Naive Bayes (NB)

NB classifier commonly used for classification relies on Bayes' theorem, a simple probability theorem. NB classifier is obtained using the set of discriminating functions and calculates the conditional probabilities in the training set. Given the assumption that the features are independent, this classifier can easily be shown as optimal in terms of minimizing the false classification rate or zero-one loss (Domingos & Pazzani, 1997). The Bayes' rule relates to conditional probabilities and unconditional probabilities to identify the class of the cases examined under the independence condition (Lewis, 1998).

3.5. Logistic Regression (LR)

LR estimates a separating hyperplane, which is a linear function of input properties between two conditions or two classes. The purpose of a given training data set can be defined as (Ryali et al., 2010):

- Estimating the hyperplane that accurately predicts the class label of a new instance,
- Determining the subset of the most informative characteristics about class distinction.

LR is used to describe the relationship between the independent variables and the dependent variable known to be binary. According to this approach, the logistic function is

always between 0 and 1. Therefore, for the logistic model, a risk estimate of over 1 or less than 0 can never be found.

3.6. K-Nearest Neighbor (KNN)

To classify a new transaction into the normal or fraudulent class, the KNN classifier calculates the similarity between the new transaction and each of the training instances. It uses the most similar neighbors' class tags to predict the class of the new operation. The basic assumption here is that operations belonging to the same class will come together in the vector space. That is, in the KNN classifier, there is an assumption that a sample is classified into the class that is most similar to the other samples in the vector space compared to the classes in which they are located (Liao & Vemuri, 2002). Since the distance is measured in the class comparison to select the correct class, the distance measure used to define the closest neighbor of a sample directly affects the performance of the KNN. As a measure of distance, simple Euclidean distance is often used to measure differences between samples (Weinberger & Saul, 2009).

3.7. Ensemble Learning

The ensemble is a group of classifiers that combine individual decisions in some form (typically by weighted or unweighted voting) to classify new samples (Aras & Gülay, 2017). The voting procedure of an ensemble learning method treats individual classifiers as smart experts. This method aims to combine the strengths of each classifier in identifying the patterns in the problem. This procedure ensures that the classifiers are in communication with each other to construct the final decision collectively. Many papers show that simple ensemble methods generally lead to considerably better predictions than other sophisticated ensemble methods (Menezes et al., 2000). Hence, we utilized the two simplest ensemble methods in this study. The definition of these is given below.

3.8. Simple Majority (Hard Voting)

Majority voting is based on collecting votes of at least one more than half of the number of classifiers regarding the final decision. For example, if there are three classifiers and two classes, if two of the classifiers choose the 1st class, the first class with the majority is selected. This voting is an example of a weighted majority vote and assigns an equal weight 1/j to each classifier, where "j" is the number of classifiers in an ensemble. The simple voting-based ensemble learning formulation used in the study is presented in equation (1) below, with the output of the t-th classifier represented by $f_t^{(j)}(x_n)$ and the input of the j-th class denoted by x_n (Kwong et al., 2015).

$$f^j com(x_n) = \sum_{t=1}^{(T)} f_t^{(j)}(x_n), \quad f_t^{(j)}(x_n) \in \{0,1\} \tag{1}$$

3.9. Soft Voting

In this voting method, the class with more than 50% average of class probabilities wins. The approach computes the prediction of the class labels by taking the sum of the predicted probabilities into account. Therefore, more sensitive results are likely to be obtained, and it can

be obtained if the classifier outputs the probabilities of each label. The soft voting ensemble learning formulation used in the study is presented in equation (2) below, with the output of the t-th classifier represented by $f_t^{(j)}(x_n)$ and the input of the j-th class denoted by x_n. Here w_t refers to the weight of each class for classifiers (Kwong et al., 2015).

$$f^j com(x_n) = \sum_{t=1}^{(T)} w_t f_t^{(j)}(x_n), \quad f_t^{(j)}(x_n) \in \{0,1\} \tag{2}$$

3.10. Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO, developed by (Tibshirani, 1996), minimizes experimental error penalized by the regularization term and balances data tuning and regularization. LASSO aims for the irrelevant features to be zero by shrinking the model parameters and even works effectively if the number of variables is greater than the number of observations. The regularization term governs the amount of shrinkage and has a significant effect on the performance. Hence, we employed 10-fold cross-validation to find the optimal value of this parameter.

3.11. Principle Component Analysis (PCA)

PCA extracts the fundamental components or important properties of the correlation matrix in terms of eigenvectors. These vectors are linear combinations that explain the orthogonality of the observed data and the independence of the variance in the observations. Most of the observed variance in the samples can be explained with just a few major components. The transformed attributes are called principal components. It is helpful to reduce the number of original variables to smaller transformed variables if there are many highly correlated features in a data set. PCA provides a perspective on how to reduce a multidimensional data set linearly to a smaller size (Tharwat, 2016).

4. Results and Discussion

Some performance evaluation measures were utilized to assess the predictions of the classifiers examined. These performance scores were calculated from the confusion matrix obtained from the results on the test set. The confusion matrix consists of the true negative (TN), false positive (FP), false negative (FN) and true positive (TP) values of the classification. After obtaining these values, accuracy, F1-score (also known as F-score or F-measure) and Matthews correlation coefficient (MCC) metrics were computed to measure classification performance in this study.

In general, accuracy is accepted as the most reasonable performance measurement criterion in classification problems. However, when the data set is unbalanced, the accuracy criterion is no longer a reliable criterion because it provides an optimistic result on the majority class. At this point, the MCC score or F1-score can be used as a measure that can deal with imbalanced data sets. Besides, for binary classifications, it may be preferable to evaluate the ranking performance achieved by the MCC since this score is high only when the classifier can accurately predict most positive data samples and most negative data samples and includes all elements of the confusion matrix (Chicco & Jurman, 2020). If both false negatives and false positives are equally important in the application area, the F1-score can be used, which establishes a balance between the precision criterion calculated with false negatives and the

precision criterion calculated with false positives by taking the harmonic mean. For these reasons, these three performance measurements were compared to make a final decision.

Accuracy (ACC): This represents the ratio between the number of users correctly classified among fraudulent users and the number of all users. This criterion is usually known as the percentage of correctly classified users among all users. It takes a value between 0 and 1. The best measurement result is 1. Equation (3) gives the formula for accuracy.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

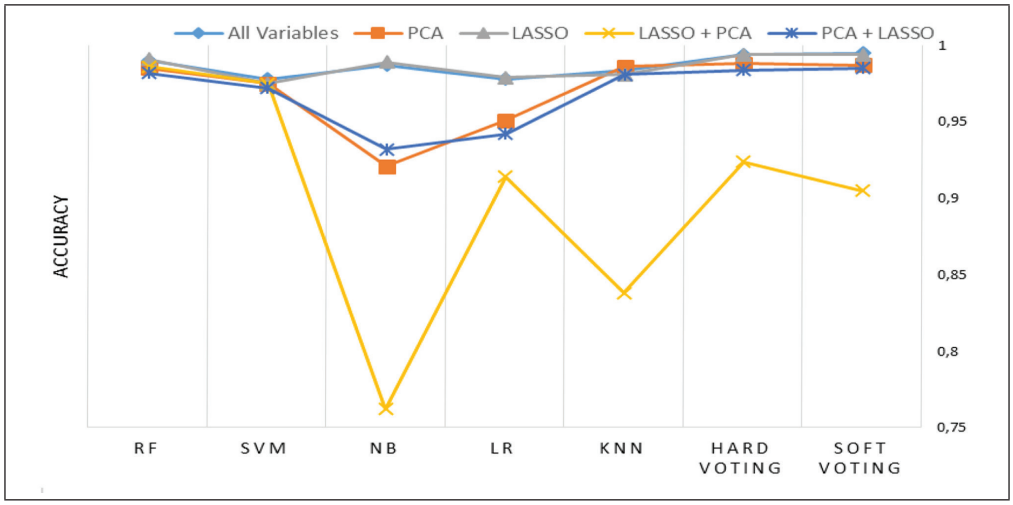
Table 2 presents the accuracy scores calculated for all methods under investigation. In this table, each column represents different cases constructed to see the effect of the techniques of dimension reduction on the results. The first column uses all variables, 38 in total, without any dimension reduction. The second column indicates the effect of feature extraction, PCA. The number of basis vectors was chosen to explain 90% of the total variance. Therefore, 23 transformed variables were used for the analysis. The third one used LASSO as a feature selection technique. Using 10-fold cross-validation, the number of the selected variables was determined as 29. The fourth column shows the sequence of LASSO plus PCA to reduce the dimension of the problem before modelling. After applying this strategy, 22 transformed variables were used for the analysis. The last column represents the sequence of PCA plus LASSO as a pre-processing strategy to handle the dimensionality problem. This strategy reduced the number of variables to 15. The rows in the table correspond to the classifiers.

As can be seen from Table 2, the highest accuracy score, which is shaded, was achieved by the ensemble of soft voting without any dimensionality reduction technique. To investigate the effect of dimension reduction techniques on each classification method, Figure 1 was generated from Table 2. Figure 1 shows that the greatest loss of information occurred in the case of LASSO+PCA. The accuracy scores decreased substantially after this data pre-processing method for NB, LR, KNN, and hard and soft voting. It is very hard to detect any performance differences between the LASSO method and the method using all variables. RF and SVM were the classifiers least affected by the number of variables. Among the single classifiers, RF generally outperformed the others.

Table 2: The Accuracy Score

	All Var.	PCA	LASSO	LASSO+ PCA	PCA + LASSO
RF	0.990	0.985	0.991	0.986	0.982
SVM	0.978	0.975	0.975	0.975	0.972
NB	0.987	0.921	0.989	0.762	0.932
LR	0.978	0.951	0.979	0.914	0.942
KNN	0.984	0.986	0.981	0.838	0.981
Hard	0.994	0.988	0.994	0.924	0.984
Soft	0.995	0.987	0.994	0.905	0.985

Figure 1: The Graph of Classification Methods in Terms of Accuracy



F1-Score: This is a combination of recall and precision criteria. These two criteria are inversely related to each other. The F1-score provides a balance between these two criteria. Therefore, this criterion takes both false positives and false negatives into consideration. It takes a value between 0 and 1. The best measurement result is 1. The F1 formula is given in equation (4).

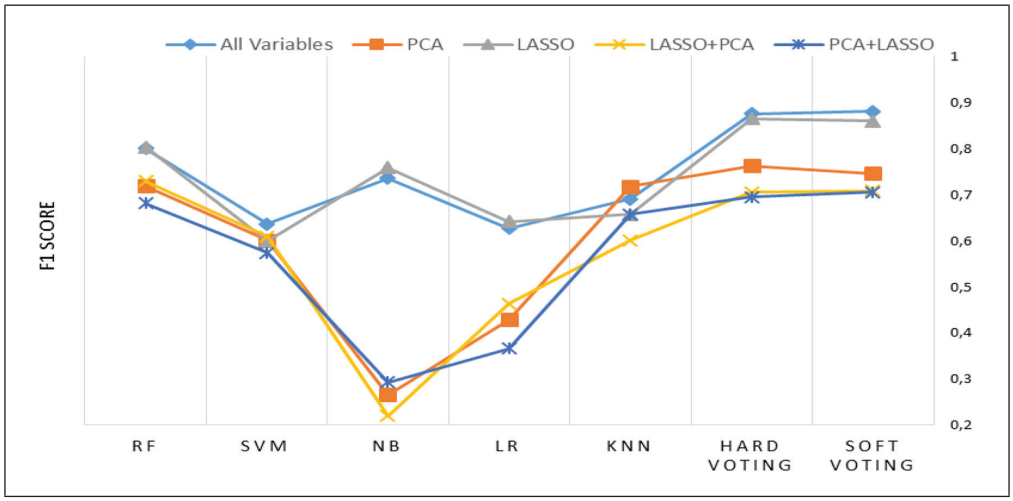
$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{4}$$

Table 3 includes the F1-score calculated for all combinations of the classifiers and the strategies for dimension reduction. According to the table, the case with the highest F1-score is the classification made by soft voting using all variables. When compared with Table 2, it can be seen that F1-score has a high power of distinguishing between different classification algorithms because the values obtained vary to a greater extent than Table 2. Considering the values calculated in Table 3, the performance change of each data pre-processing regarding dimension reduction is illustrated in Figure 2. It is evident from this figure that the LASSO and all variables exhibit very similar performance. Among the classifiers, RF was better than the other single models. However, the ensemble methods outperformed all single models, even RF. NB and LR were the algorithms most affected by the methods of dimension reduction.

Table 3: The F1-Score

	All Var.	PCA	LASSO	LASSO+ PCA	PCA + LASSO
RF	0.800	0.719	0.803	0.729	0.681
SVM	0.636	0.601	0.599	0.609	0.574
NB	0.736	0.266	0.759	0.220	0.293
LR	0.627	0.429	0.642	0.463	0.366
KNN	0.691	0.718	0.657	0.601	0.657
Hard	0.875	0.763	0.865	0.705	0.695
Soft	0.881	0.746	0.860	0.709	0.705

Figure 2: The Graph of Classification Methods in Terms of F1-Score



MCC: This measures the quality of detection rate of the classification. It ranges from -1 to 1. The best measurement result is 1. A result of -1 indicates that the prediction was completely wrong. The MCC formula is given in equation (5).

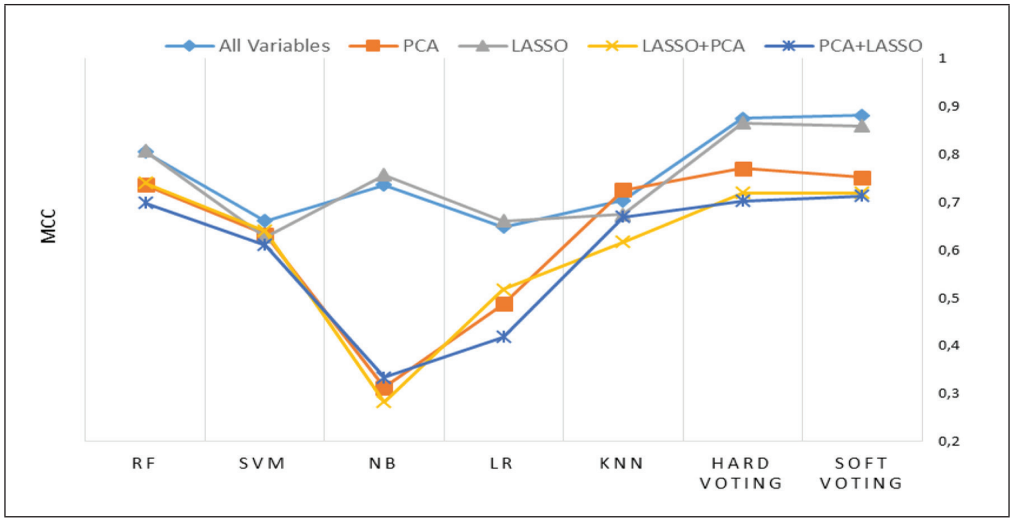
$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (FN + TN) * (FP + TN) * (TP + FN)}} \tag{5}$$

Table 4 contains the MCC scores calculated for all comparison points. The highest score was achieved again by soft voting with all variables. RF was the best single classifier in all cases examined. There was clear superiority of ensemble methods over the single classifiers except for the RF with LASSO+PCA method. Figure 3 shows a graph of classification algorithms in terms of MCC. Since the change in performance for MCC and F1-score looks very similar, the same interpretations made for Figure 2 are also valid for Figure 3.

Table 4: The MCC Score

	All Var.	PCA	LASSO	LASSO+ PCA	PCA + LASSO
RF	0.804	0.736	0.807	0.740	0.698
SVM	0.661	0.632	0.627	0.640	0.610
NB	0.736	0.313	0.757	0.283	0.334
LR	0.648	0.487	0.660	0.518	0.418
KNN	0.703	0.725	0.674	0.617	0.669
Hard	0.874	0.770	0.865	0.719	0.702
Soft	0.881	0.751	0.859	0.719	0.713

Figure 3: The Graph of Classification Methods in Terms of MCC



5. Conclusion

This study created a structure that shows the risk scores of users by automatically testing all users with a mobile or web account within the company, which is carried out with machine learning techniques. Thus, reliable and unreliable users were identified and this real need was met satisfactorily in terms of company reputation and profitability. With this study, the need to protect a real system with a high number of users from infringing users was met, the fraud existing in the system was detected and these users were automatically blacklisted. The learning-based system developed stands out due to its ability to determine new risks and people who perform attacks in a short time, effectively and with minimum deviation. The method used eliminates the burden of identifying thousands of different possible risk rules combinations in fraud detection and provides both time and cost optimization to the business where the data are used. This study increases the security of a mobile ticket application used in a real public transportation system.

In this study, an algorithm based on machine learning techniques that detected fraudulent users exhibiting risky behaviors in a mobile application was developed and these users were automatically blacklisted. In the method developed, machine learning classification techniques, variable extraction and variable selection were used. Three performance measurement scores were employed to understand which situation leads to the best prediction performance and to choose the right method to decide whether the users are fraudulent or not. As a result, a reliable and consistent result was obtained for accuracy, F1-score and MCC. The best result among all the scores was obtained using all variables without any variable selection or dimension reduction method, followed by the LASSO method. Among the single classifiers, RF stood out in terms of performance. However, it was observed that the least prediction error was obtained in soft voting, one of the ensemble learning methods. Thus, it can be said that including ensemble learning methods in the application made a meaningful contribution to the study.

In future studies on fraud analysis in mobile ticketing application, the system designed will be monitored live and performance measurements will be made over the system in real time by comparing the new fraud rates with the manual structure applied before the smart blacklist system. The number of frauds missed in the online structure and the number of users labeled as fraudulent even though they are not fraudulent will be calculated and the outputs will be compared by applying deep learning methods according to the performance measurement results and new data size by evaluating the presence of new variables according to new frauds.

Conflict of Interest

The authors have no conflicts of interest to declare.

Contribution Statement

The authors declare that they have contributed equally to this work.

Acknowledgement

This study was financially supported by Kentkart Ege Elektronik, Izmir, Turkey.

References

Abdallah, A., Maarof, M. A. & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90–113.

Abe, S. (2005). Support vector machines for pattern classification. London: Springer.

Adewumi, A. O. & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), 937–953.

Aras, S. & Gulay, E. (2017). A new consensus between the mean and median combination methods to improve forecasting accuracy. Serbian Journal of Management, 12(2), 217–236.

Awoyemi, J. O., Adetunmbi, A. O. & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI), 1–9, 29-31 October, 2017, Covenant University, Canaanland, Ota, Ogun State, Nigeria.

Behdad, M., Barone, L., Bennamoun, M. & French, T. (2012). Nature-inspired techniques in the context of fraud detection. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), 1273–1290.

Bolton, R. J. & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). Classification and regression trees–crc press. New York: Routledge.

Camenisch, J., Piveteau, J.-M. & Stadler, M. (1996). An efficient fair payment system. Proceedings of the 3rd ACM Conference on Computer and Communications Security, 88–94, New Delhi, India.

Cao, J., Kwong, S., Wang, R., Li, X., Li, K. & Kong, X. (2015). Class-specific soft voting based multiple extreme learning machines ensemble. Neurocomputing, 149, 275–284.

Chan, P. K., Fan, W., Prodromidis, A. L. & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and Their Applications, 14(6), 67–74.

Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1), 1–13.

Choi, D. & Lee, K. (2017). Machine learning based approach to financial fraud detection process in mobile payment system. IT Convergence Practice (INPRA), 5(4), 12-24.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Dahlberg, T., Mallat, N., Ondrus, J. & Zmijewska, A. (2008). Past, present and future of mobile payments research: A literature review. Electronic Commerce Research and Applications, 7(2), 165–181.

De Menezes, L. M., Bunn, D. W. & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120(1), 190–204.

Dewan, S. G. & Chen, L. (2005). Mobile payment adoption in the US: A cross-industry, crossplatform solution. Journal of Information Privacy and Security, 1(2), 4–28.

Domingos, P. & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29(2), 103–130.

Erbacher, R. F., Walker, K. L. & Frincke, D. A. (2002). Intrusion and misuse detection in large-scale systems. IEEE Computer Graphics and Applications, 22(1), 38–47.

Fujimura, K. & Nakajima, Y. (1998). General-purpose digital ticket framework. USENIX Workshop on Electronic Commerce, 177–186.

Genuer, R., Poggi, J. M. & Tuleau-Malot, C. (2010). Variable selection using random forests. Pattern Recognition Letters, 31(14), 2225–2236.

Ghosh, S. & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. System Sciences, 3, 621–630.

Hamal, S. & Senvar, O. (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for Turkish SMEs. International Journal of Computational Intelligence Systems, 14(1), 769-782.

Hashemi, M. R. & Soroush, E. (2006). A secure m-payment protocol for mobile devices. 2006 Canadian Conference on Electrical and Computer Engineering, 294–297, Ottawa, ON, Canada.

Hassinen, M., Hyppönen, K. & Haataja, K. (2006). An open, PKI-based mobile payment system. International Conference on Emerging Trends in Information and Communication Security, 86–100, Freiburg, Germany.

Hess, M. F. & Andiola, L. M. (2018). Fraud risk brainstorming at tesla motors. Issues in Accounting Education, 33(2), 19-34.

Hwang, Y.-S., Han, S. W. & Nam, T.-Y. (2006). Secure rejoining scheme for dynamic sensor networks. International Conference on Emerging Trends in Information and Communication Security, 101–114, Freiburg, Germany.

Johnson, G. G. & Rudesill, C. L. (2001). An investigation into fraud prevention and detection of small businesses in the United States: Responsibilities of auditors, managers, and business owners. In Accounting Forum, 25(1), 56-78.

Jyothsna, V., Prasad, R. & Prasad, K. M. (2011). A review of anomaly based intrusion detection systems. International Journal of Computer Applications, 28(7), 26–35.

Karnouskos, S., Hondroudaki, A., Vilmos, A. & Csik, B. (2004). Security, trust and privacy in the secure mobile payment service. 3rd International Conference on Mobile Business, 35, 12-13 July 2004, New York City, U.S.A.

Küçükkocaoğlu, G., Benli, Y. K. & Küçüksözen, C. (1997). Detecting the manipulation of financial information by using artificial neural network models. Istanbul Stock Exchange Review, 9(36), 1-26.

La Polla, M., Martinelli, F. & Sgandurra, D. (2012). A survey on security for mobile devices. IEEE Communications Surveys & Tutorials, 15(1), 446–471.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. European Conference on Machine Learning, 4–15, Chemnitz, Germany.

Liao, Y. & Vemuri, V. R. (2002). Use of k-nearest neighbor classifier for intrusion detection. Computers & Security, 21(5), 439–448.

Liaw, A. & Wiener, M. (2002). Classification and regression. R News, 2(3), 18-22.

Linck, K., Pousttchi, K. & Wiedemann, D. G. (2006). Security issues in mobile payment from the customer viewpoint. Proceedings of the Fourteenth European Conference on Information Systems, 1, 1085-1095, Göteborg, Sweden.

Melo-Acosta, G. E., Duitama-Muñoz, F. & Arias-Londoño, J. D. (2017). Fraud detection in big data using supervised and semi-supervised learning techniques. 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 1–6, Cartagena, Colombia.

Pai, P. F., Hsu, M. F. & Wang, M. C. (2011). A support vector machine-based model for detecting top management fraud. Knowledge-Based Systems, 24(2), 314-321.

Pirker, M. & Slamanig, D. (2012). A framework for privacy-preserving mobile payment on security enhanced arm trustzone platforms. 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, 1155–1160, Liverpool, United Kingdom.

Qin, Z., Sun, J., Wahaballa, A., Zheng, W., Xiong, H. & Qin, Z. (2017). A secure and privacy-preserving mobile wallet with outsourced verification in cloud computing. Computer Standards & Interfaces, 54, 55–60.

Ryali, S., Supekar, K., Abrams, D. A. & Menon, V. (2010). Sparse logistic regression for whole-brain classification of fMRI data. NeuroImage, 51(2), 752–764.

Steiner, J. G., Neuman, B. C. & Schiller, J. I. (1988). Kerberos: An authentication service for open network systems. Usenix Winter, 191–202.

Tharwat, A. (2016). Principal component analysis-a tutorial. International Journal of Applied Pattern Recognition, 3(3), 197–240.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M. & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decision Support Systems, 75, 38–48.

Wang, S. (2010). A comprehensive survey of data mining-based accounting-fraud detection research. 2010 International Conference on Intelligent Computation Technology and Automation, 1, 50–53, Changsha, China.

Wang, Y., Hahn, C. & Sutrave, K. (2016). Mobile payment security, threats, and challenges. 2016 Second International Conference on Mobile and Secure Services (MobiSecServ), 1–5, Gainesville, Florida, USA.

Weinberger, K. Q. & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 10(2), 207-244.

Zareapoor, M., Seeja, K. R. & Alam, M. A. (2012). Analysis on credit card fraud detection techniques: Based on certain design criteria. International Journal of Computer Applications, 52(3), 35-42.