Avrupa Bilim ve Teknoloji Dergisi
Sayı 20, S. 522-529, Aralık 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No 20, pp. 522-529, December 2020

Research Article

Detection of Credit Card Fraud in E-Commerce Using Data Mining

Yasin Kırelli[1*], Seher Arslankaya[2], Muhammed Taha Zeren[3]

[1] Sakarya University, Industrial Engineering Department, Sakarya, Turkey (ORCID: 0000-0002-3605-8621)
[2] Sakarya University, Industrial Engineering Department, Sakarya, Turkey (ORCID: 0000-0001-6023-2901)
[3] Turkish Aerospace Industries (TAI - TUSAŞ), Ankara, Turkey (ORCID: 0000-0001-5615-0751)

(First received 3 June 2020 and in final form 2 November 2020)

(**DOI:** 10.31590/ejosat.747399)

ATIF/REFERENCE: Kırelli, Y., Arslankaya, S. & Zeren, M. T. (2020). Detection of Credit Card Fraud in E-Commerce Using Data Mining. *European Journal of Science and Technology*, (20), 522-529.

Abstract

Credit card payment is one of the most preferred methods of e-commerce sites. Fraud orders are the biggest concerns for online shopping sites. Fraud operations affect not only customers but also companies and banks. Hence, companies should be able to classify orders and take measures against suspicious transactions. Classification is easier on the banking side because of more information about customers, but it is more difficult to determine this process on e-commerce sites. In this study, the actual order data of a private e-commerce enterprise has been examined and suspicious transactions are determined. First of all, all order data is analyzed and filtered. The best variables for classification are determined by variable selection algorithms. Afterwards, classification algorithms are applied and suspicious orders are determined with 92% success rate. Naïve Bayesian, Decision Trees and Artificial Neural Network have been used as comparative data mining methods.

Keywords: Credit Card Fraud detection, classification, data mining.

Veri Madenciliği ile E-Ticarette Kredi Kartı Dolandırıcılığının Tespiti

Öz

Kredi kartı ile ödeme, e-ticaret sitelerinin en çok tercih edilen yöntemlerinden biridir. Dolandırıcılık şüphesi olan siparişler, alışveriş siteleri için en büyük endişe kaynağıdır. Sahtekarlık işlemleri sadece müşterileri değil, aynı zamanda şirketleri ve bankaları da etkiler. Bu nedenle, şirketler siparişleri sınıflandırabilmeli ve şüpheli işlemlere karşı önlemler alabilmelidir. Bankacılık tarafında, müşteriler hakkında daha fazla bilgi olması nedeniyle sınıflandırma daha kolaydır, ancak bu süreci e-ticaret sitelerinde belirlemek daha zordur. Bu çalışmada, özel bir e-ticaret girişiminin gerçek sipariş verileri incelenmiş ve şüpheli işlemler belirlenmiştir. Öncelikle, tüm sipariş verileri analiz edilir ve filtrelenir. Sınıflandırma için en iyi değişkenler değişken seçim algoritmaları ile belirlenmiştir. Daha sonra sınıflandırma algoritmaları uygulanır ve %92 başarı oranı ile şüpheli siparişler belirlenir. Karşılaştırmalı veri madenciliği yöntemleri olarak Naive Bayesian, Karar Ağaçları ve Yapay Sinir Ağı kullanılmıştır.

Anahtar Kelimeler: Kredi Kartı Dolandırıcılığı Tespiti, Sınıflandırma, Veri Madenciliği.

[*] Sorumlu Yazar: Sakarya Üniversitesi, Endüstri Mühendisliği Bölümü, Sakarya, Türkiye (ORCID: 0000-0002-3605-8621), yasin.kirelli@ogr.sakarya.edu.tr

1. Introduction

According to 14th annual report of CyberSource, which is a Visa Company, companies reported losing an average of 0.9% of total online revenue to fraud [1], based on this information it is estimated $3.5 billons loss just for North America. According to the study in [2] using nontraditional payment channels (mobile, internet, etc.) increased the fraud transaction by 14% compared to the year 2008 which makes identification process more challenging. Furthermore, it is also an indication that fraudsters are changing their strategies and new fraud patterns are emerging as now. Therefore, all identification and classifying strategies needs to be reengineered for each new data set.

In case of this study it is aimed to develop a tool to discover the fraud transaction for an e-commerce site out of actual transaction. This classification and identification were carried out in rule base and mainly manual. In section 2 related works is going to be given, in section 3 online shopping data is going to be presented, In the section 4 classification and identification methods and results is going to be presented, in the conclusion section the results are going to be discussed. All the calculation and analysis in this study are carried out using WEKA data analysis software (Version 3.8.4) [3].

2. Related Works

Since it is widely used there has been many attempts to detect and classify fraudulent usage of credit cards by banking systems. It is a relatively easy task for banking systems than an e-commerce venture since they are in reach of much more information. Lack of relatively abundant information makes analyzing difficult for online shopping ventures and challenging. In the work of Raj and Portia, the diversity of the fraud detection methods has been analyzed and categorized in according the name of the algorithm which is generalized as Bayesian Learning, Hidden Markov models, Artificial Neural networks and their hybrid derivatives (Edwin Raj, et al., 2011).

In the study of Chan [5] a brief survey has been given about the credit card fraud detection and a cost model is proposed. Models consist of a combination of multiple learned fraud detectors. In this article, empirical studies are found to be promising.

In another study support vector machines SVM and random forest are combined and used together with well-known logistic regression to detect credit card fraud transactions [6].

Adepoju et al. (2019) inspect the execution of, supervised machine learning methods on too corrupt data on credit card fraud [19]. Vidanelage et al. (2019) discussed varios machine learning techniques using “Scikit-learn Package in Python” to find the fraudulent transactions in dataset related payment [20]. Raghavan et al. (2019) use the European (EU) Australian and German dataset and aim to benchmark deep learning and supervised machine learning methods [21]. Seemakurthi et al. (2015) describe a new approach using texts classifers to detect fraudulent texts on text-based financial documents [22].

2. Data and Implementation of Classification Methodologies

In this section first the data set unto which study is carried on is going to be presented later then the used methodologies are going to be described briefly. The relational database schema is as specified on Figure 1 and the attribute list has been created from the order table. The attribute names and what they describe in each table has been explained in Table 1.

In the field of data mining, feature subset selection is of great importance. High-dimensional data makes it difficult to test and train data mining models. Feature selection is a question of selecting a small subset of features that are necessary and sufficient. One goal of feature selection is to avoid selecting too much or too little features than necessary. If too few features are selected, the meaningful information content to be extracted from this feature group will below. On the other hand, if too many irrelevant features are selected, the effects due to noise will make it challenging to access available information. Feature selection is the process of removing unnecessary or unrelated features from the original data set. Therefore, the execution time of the classifier that processes the data is reduced, as well as the accuracy. Because irrelevant features may contain noisy data that negatively affect classification accuracy. With feature selection, meaningful information acquisition can be improved and data processing costs are reduced. Consequently, the number of features can be simplified with feature selection methods [23].

In this study, gain ratio, ChiSquared and InfoGain statistical feature selection methods are used as filter feature to simplify and sort the features. The WEKA data mining tool has been used to compare the performance of classification algorithms with feature selection methods. Default parameters have been used for each classification algorithm.

Information gain Ratio

It is one of the standard methods for feature selection. The purpose of these techniques is to eliminate irrelevant ones and the entropy value is calculated for all data. The data obtained can be obtained by using the information in a text for information acquisition, class prediction. Creating a subgroup on the class attribute selects the required attributes according to the information value obtained [24].

Chi-square MapReduce

This method is used to analyze whether the class tag or target is independent of an attribute and select the predictor variable [24].

Gain Ratio Feature Selection

The Gain Ratio has been chosen using the subset entropy value and the knowledge gain value. It is a controlled, univariate, asymmetric and entropy-based measure to eliminate the bias of knowledge gain [24].

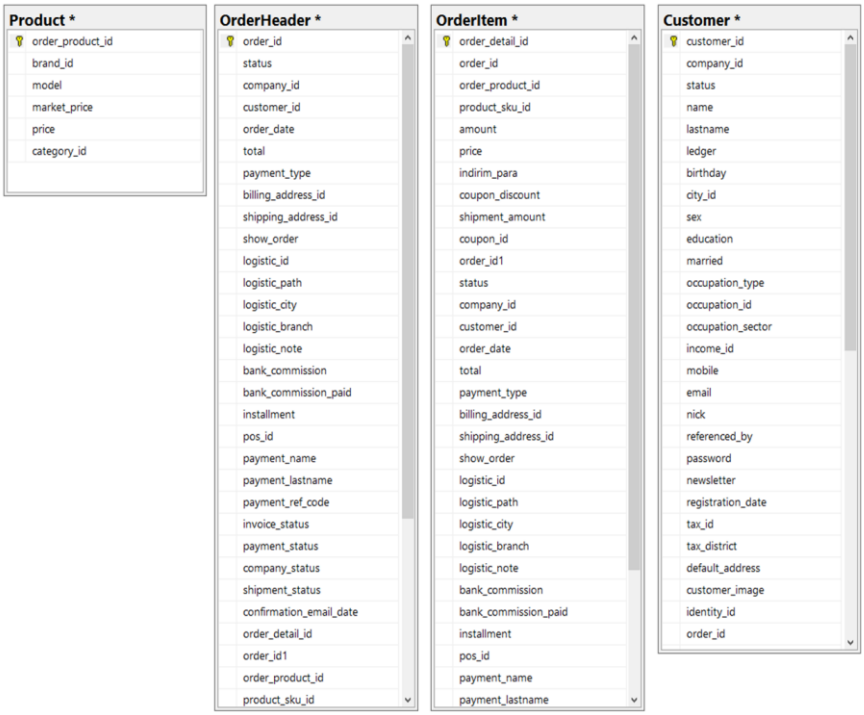

Figure 1. Relational Tables and Fields of E-Commerce Site

Table 1. Name of the attributes after pruning and order of

Name of the attributes after pruning	Meaning	Order of Attributes after Statistical Selection		
		GainRatio	ChiSquared	InfoGain
Total	Total shopping amount	2	1	1
Payment_ref_code	Special code given to order by company	3	5	4
Amount	Amount of the product	9	12	13
OrderHour	Hour of order: Normalized between 1	13	9	9
OrderDayOfWeek	Day of the Order: Normalized between 1	14	10	11
NameSurnameLen	Name and Surname length of the customer	1	2	2
Discount_money	Discount amount	15	15	15
Coupon_Discount	Coupon Discount amount	16	16	16
Shipped_Amount	Amount of Shipment	8	11	10
CouponID	ID of the coupon	17	17	17
EmailConfirmTime	Hour of conformation: Normalized between 1	11	14	14
CustomerCityID	ID of the city	5	4	6
CustomerEmailFormat	Format of the email related customer	7	7	7
OrderBrandID	Brand ID of the ordered product	6	6	5
CategoryID	Category ID of the product	4	3	3
CustomerAge	Age of the customer	10	8	8
Sex	Gender of the customer	12	12	12
isFraud	Class Attribute: Normalized to 1 or 0	Class Attribute	Class Attribute	Class Attribute

3.1. Data Sets

The raw data given by the e-commerce company consist of 38 columns for 1615 order records. The last column is entitled as is fraud which is supplied by the IT department of the e-commerce company, according to banking system declaration. After attribute selection algorithm is applied to the dataset namely GainRatio, some of the columns are discarded due to the high rate of missing values. After initial pruning the field 17 attributes out of 38 are selected for analysis, which are depicted in Table 1.

Selecting most appropriate attributes has an immense effect in performance especially for neural network classifiers [8], since the number of attributes are reduced required amount of calculation will also decrease which will result in a performance increase. In order to obtain such performance, increase the selection attribute utility of WEKA software is utilized using Gain Ratio, Info Gain and Chi-Squared algorithms. Attributes are reordered according to importance for classification as shown in Figure 2.

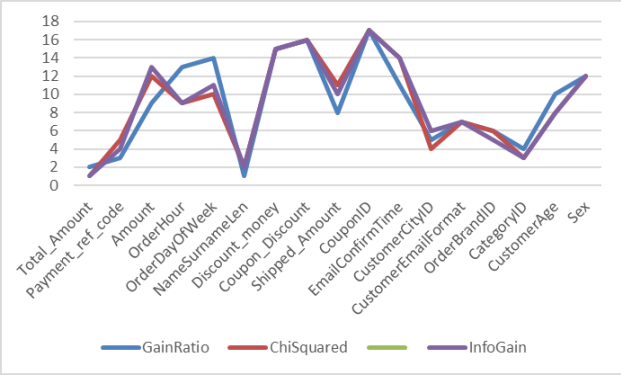

Figure 2. Importance for classification

Methodologies: Generally, classifiers can be categorized in many ways namely with being supervised or unsupervised. With unsupervised methods it is generally aimed to cluster the certain data set in unforeseen categories or groups. With supervised methods main goal is generally to determine if an instance is belonging to a certain class is given. In order to test different methodologies, different classifiers belonging relatively different realm of classification are chosen namely Naïve Bayesian, k-NN (nearest neighbour), J 48 Decision tree, ANN (artificial Neural network).

Classification algorithms have been selected by considering the cases specified respectively in the selection. Selection criteria are accuracy in general, speed of classification, tolerance to missing values, tolerance to irrelevant values, tolerance to interdependent values. Accordingly, the algorithms in the most suitable criteria for the model have been selected as: "Naïve Bayes, RBF Network, KNN Ratio, J48 Ratio".

3.1.1. Naive Bayesian

The Bayes's theorem is conditional probability calculation formula which was found by Thomas Bayes in 1812. This formula is one of the most important formula of probability theory. Naïve Bayesian theory defines the probability of an event, according to prior knowledge of conditions that could be related to the event [7].

Naïve Bayes classifier based on The Bayes's theorem. Naïve Bayes has been used and studied since the beginning of 1960. This theorem is kind of learning algorithm that can be able to work on unstable data sets. The algorithm calculates probability of each elements and classifies according to most probability rate. Naive Bayesian classifiers has assumption that the value of a spesific attribute is independent of the value of any other attribute. It is not possible to make correlations between attributes [7].

As mentioned in previous chapters Weka Software is used in ai analyses. Naïve Bayes classifier in Weka uses probabilistic Naïve Bayes classifier [8][9] which is used as descriptive and complementary classifier algorithm. Mainly make use of Bayes Rule as shown in Eq. 1:

$$\arg \max_{Y} = P(Y|\, X_1, X_2 \ldots X_n) \tag{1}$$

$$P(Y|\, X_1, X_2 \ldots X_n) = \frac{P(\,X_1, X_2 \ldots X_n|\, Y).P(Y)}{P(\,X_1, X_2 \ldots X_n)} \tag{2}$$

Naïve Bayes is based on learning from data, it means in order to learn model occurrence of every output calculated, it is named as prior (second term of nominator in Eq. 2). Likelihood probability (first term of nominator in Eq. 2) is then calculated and multiplied and divided by normalization constant (denominator term in Eq. 2).

3.1.2. Decision Tree J48

Decision tree is one of the most known nonparametric machine learning methods and this method is widely used in data mining, machine learning, expert systems and multivariate analysis. Its function is dividing and conquer approach to split the input space to sub-regions as shown Figure 3. Then it creates a model depending on these regions.

A decision tree is kind of a hierarchical structure. It implicates a root node, internal decision nodes, leaf nodes, and branches. Internal decision nodes la-bel the branches according to their test function. Leaf nodes correspond to labeled instances.

Decision tree is an easy method to apprehend. Its representation can be ex-ported to if/then rules. Decision tree algorithms run faster regarding other learning algorithms because their hierarchical structure lets elimination of some decision nodes. Instead of learning error-free model from decision tree, it is important to find the model with the simplest tree so that performance on test data can be improved [11].

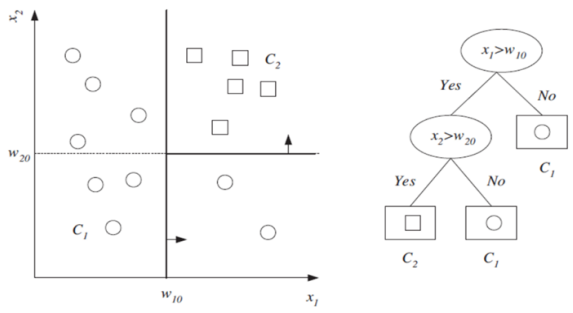

Figure 3. Decision Tree Classification [11]

In this study another classifier named J48 is also used. It is java reimplementation of well-known decision tree C4.5. A full definition of C4.5 appears as an excellent and readable book [12], along with the full source code. C5.0 is available commercially with negligible improvements.

3.1.3 *Naive* Bayes Tree

NBTree is a mixed implementation of decision trees and Naïve Bayes. It creates the tree so that leaves are Naïve Bayes classifiers for the instances that reach the leaf. Cross validation is utilized when constructing the tree to decide if a node should be split further or a Naïve Bayes model should be used instead [9].

3.1.4. *k-NN (Nearest Neighbor)*

K Nearest Neighbors (kNN) algorithms' basic aim is to utilize a dataset or database in which the data points are isolated into a few separate classes to predict the classification of characterization of another new sample point as shown Figure 4. kNN is one of the popular data mining algorithms, applied for regression and classification in the two instances [14].

kNN algorithm is a nonparametric method. And this way, mostly used for regression [15]. It is based on the idea that instance must be in a close distance when compared to its closet neighbors. k is total number of the neighbors which are going to be considering.

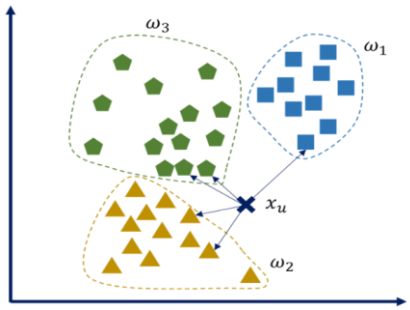

Figure 4. kNN Algorithm Classification

3.1.5. *ANN Artificial Neural Network- Multi Layer Perceptron*

Appearance of Artificial Neural Networks based on modelling of Neural Cell of Human Brain. ANN basically mathematical mimic of Biological Neural Cell.

A neuron perceptron receives multiple inputs, outputs are calculated by weighted summation. Each weight of perceptron is determined during training and calculated in relation to training data. Each output of the perceptron can be passed through an activation function or transfer function, which will be explained in the next section.

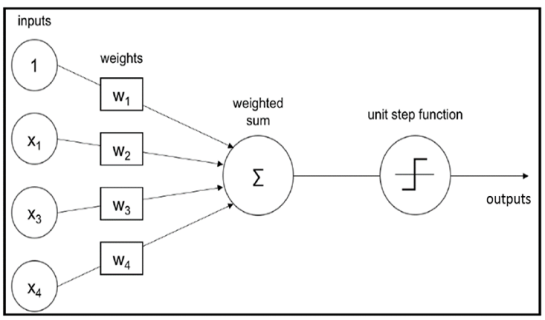

Figure 5. Diagram of Perceptron [17]

Artificial Neural Network is a combination of perceptrons and activation functions. The perceptrons are connected directly to each layer, as shown in Figure 5. Hidden layer units create a nonlinear structure. Also, this layer maps input layers to output layers in a smaller size area. This map is created with the weights of the inputs and it forms a model with match states. This map is created with the weights of the inputs, and this result pattern is called the model as shown in Eq. 3 [17].

$$Z = b + \sum_{i=1}^{n} x_i w_i \qquad (3)$$

ANNs have been widely used to model systems which are not easy modeled mathematically. They are therefore good choice of classification. Multi-Layer Perceptron is one of them that uses back propagation to classify in-stances [9].

3.1.6. *RBF Network*

RBF Network uses Gaussian radial basis function network [9]. This method basically derives the properties of hidden layers using the K-Means algorithm. This methodology has two layers and contains that not counting the input layer (i = 1 to N). It also differs from a multilayer perceptron called the hidden units as L nodes (k = 1 to L) to perform computations as shown Figure 6.

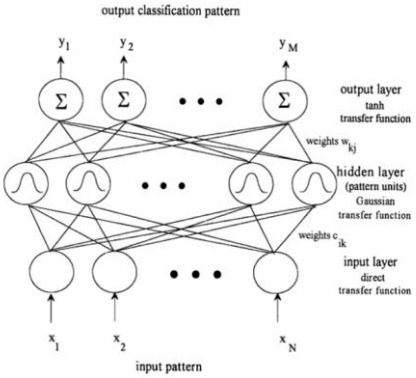

Figure 6. The architecture of a radial-basis-function network

4. Model for Classification

The basic chosen strategy is depicted in Figure 7. According to the figure first of all, raw data and attributes are converted to a suitable nominal form. Then soma attributes are trimmed manually since they contains high rate of missing value and inconsistent data. Classification Experiment A is carried out. After the further reduction in attribute is obtained by applying the attribute selection attribute of the WEKA software. Only top 2 or 3 attributes are selected, and Experiment B, Experiment C and Experiment D is carried out.

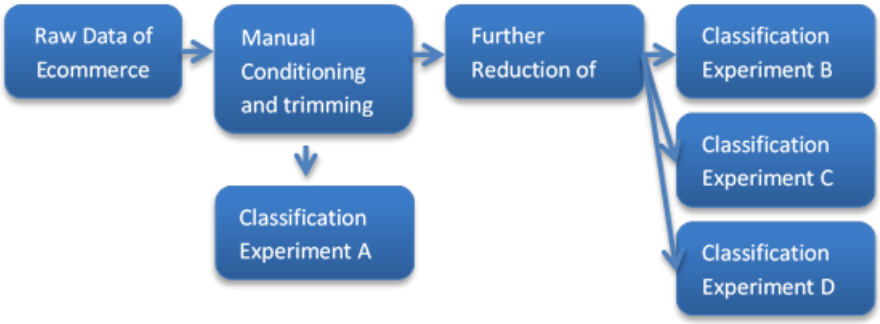

Figure 7. Classification Experiment roadmap

For Experiment set A all 17 attributes are used for six different classifiers and result of the TP accuracy is obtained. For Experiment set B only attributes total, payment_ref_code and NameSurmaneLen is used, For Experiment set C, only attributes total NameSurnameLen and coupon_discount is used, For Experiment set D only attributes total, and NameSurmaneLen is used. Last experiment set is established just for MultiLayer Perceptron classifier since the number of the unique variable are relatively less than other experiment sets. Results of F-Measure metric all 24 experiment are shown in Table 2.

Split tests are quick and great when you have many data or when it is expensive (resources or time) to train a model. A split test large dataset can produce an accurate estimate of the algorithm's actual performance. In this study test option has been selected %70 percent split of these data are used as training set and rest are used.

Table 2. Comparative results (F-Measure) of the classification experiments.

	Naïve Bayes Ratio	RBF Network	KNN Ratio	NBTree Ratio	J48 Ratio	Multilayer Perception Ratio
Experiment A	92.5 %	92.5 %	93.9 %	93.9 %	93.1 %	NA
Experiment B	91.1 %	93.2 %	95.5 %	93.1 %	87.8 %	NA
Experiment C	94.0 %	94.7 %	94.5 %	94.0 %	87.8 %	NA
Experiment D	94.4 %	94.4 %	95.0 %	94.4 %	91.7 %	94.4 %

In Table 3 TP (true positive) rate, FP (false positive) rate, precision, recall and F-measure values for the Experiment set A, B and C are represented. KNN classifier has the highest value of F-measure for experiment.

Table 3. Results for various metrics

Classifier	Experiment Set	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Naïve Bayesian	Experiment A	0.921	0.326	0.931	0.921	0.925	0.956
	Experiment C	0.948	0.528	0.946	0.948	0.94	0.963
	Experiment B	0.919	0.621	0.906	0.919	0.911	0.926
RBF Network	Experiment A	0.928	0.461	0.924	0.928	0.925	0.926
	Experiment C	0.948	0.323	0.947	0.948	0.947	0.954
	Experiment B	0.934	0.415	0.931	0.934	0.932	0.932
KNN	Experiment A	0.94	0.369	0.938	0.94	0.939	0.841
	Experiment C	0.95	0.459	0.946	0.95	0.945	0.912
	Experiment B	0.959	0.39	0.956	0.959	0.955	0.943
NBTree	Experiment A	0.948	0.551	0.948	0.959	0.939	0.828
	Experiment C	0.948	0.528	0.946	0.948	0.94	0.963
	Experiment B	0.944	0.619	0.944	0.948	0.931	0.794
J48	Experiment A	0.944	0.619	0.947	0.944	0.931	0.755
	Experiment C	0.917	0.917	0.842	0.917	0.878	0.5
	Experiment B	0.917	0.917	0.842	0.917	0.878	0.5

5. Results and Discussions

In this article, the order data of an e-commerce site consist of 1615 orders have been analyzed. %70 percent of these data are used as the training set and rest are used as test data for our models. For four different attribute sets the classification of fraud transaction is performed by the help of 6 different classifiers shown as Figure 8.

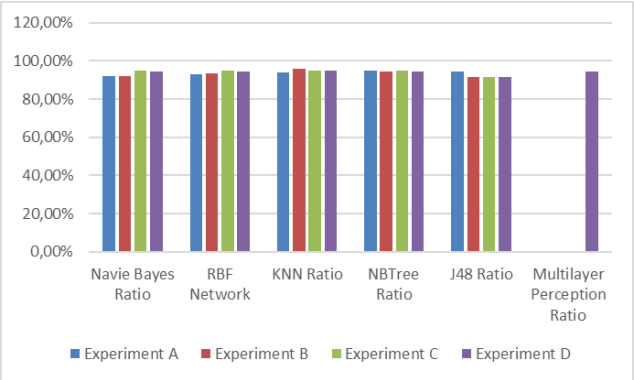

Figure 8. Comparative classifiers results

The highest percent of accuracy 95.8678% for prediction is obtained in Experiment Set B and of the KNN according to five closest neighbours as shown Figure 9.

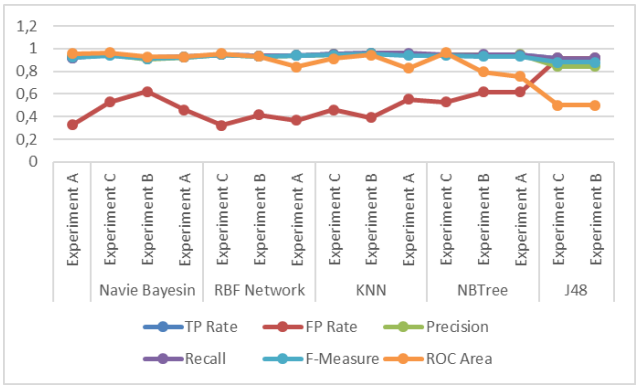

Figure 9 Values of Experiments with various metrics.

Machine learning classification requires fine tuning of parameters and a reasonably large number of samples for the dataset. Precision and correct classification takes time as well as building a model for the algorithm. Therefore, the best learning algorithm for a data set does not guarantee the precision and accuracy of another data set whose properties are logically different from the other and may not produce the same result. Therefore, it is not whether one classification algorithm is superior to others, but under what conditions a method can perform significantly better than others in each application problem. After a better understanding of the strengths and limitations of each method, performance values are compared as a solution to the problem. The goal is to identify the strengths and weaknesses of a method. The kNN, SVM, NB and RF machine learning algorithms used in the study can provide high precision and accuracy regardless of their attributes and the number of data samples. They are quick and easy procedures to implement and their results allow us to compare the performance of the methods for the predictive modelling problem [25].

This high accuracy is attributed to the nature of the data set and attribute selection procedure. Process of classification of fraud transactions has been carried out manually for the e-commerce site. Therefore, our work is going to improve their business process drastically. For future study it is intended to develop a classification software for the company.

References

[1] CyberSource a Visa Company. (2013, January) 2013 Online Fraud Report. Document.

[2] Djamila Aouada, Aleksandar Stojanovic, Björn Ottersten Alejandro Correa Bahnsen, "Feature engineering strategies for credit card fraud detection," Expert Systems with Applications, vol. 51, no. 1, pp. 134-142, June 2016.

[3] Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten Mark Hall, "The WEKA Data Mining Software: An Update," SIGKDD Explorations, vol. 11, no. 1, 2009.

[4] S. Benson Edwin Raj and A. Annie Portia, "Analysis on credit card fraud detection methods," in International Conference on Computer, Communication and Electrical Technology (ICCCET), 2011, pp. 152-156.

[5] P. K. Chan W. Fan A. L. Prodromidis and S. J. Stolfo, "Distributed data mining in credit card fraud detection," IEEE Intelligent Systems and their Applications, vol. 14, no. 6, pp. 67-74.

[6] Siddhartha Bhattacharyya, Sanjeev Jha, Tharakunnel Kurian , and J. Christopher Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602-613, February 2011.

[7] Joyce, James (2003), "Bayes' Theorem", in Zalta, Edward N. (ed.), The Stanford Encyclopedia of Philosophy (Spring 2019 ed.), Metaphysics Research Lab, Stanford University, retrieved 2020-01-17

[8] Ian H. Witten and Eibe Frank, Data Mining Practical Machine Learning Tools and Techniques, Jim Gray, Ed.: Elsevier Morgan Kaufman Publishers, 2005.

[9] N. Kwak and Chong-Ho Choi, "Input feature selection for classification problems," IEEE Transactions on Neural Networks, vol. 13, no. 1, pp. 143-159, 2002.

[10] Akbulut S., Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu.: Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, 2006.

[11] Alpaydın, E., Introduction to Machine Learning , The MIT Press, Massachusetts, 2nd edition, 2010

[12] J. R Quinlan, C4.5: Programs for machine learning.: San Francisco: Morgan Kaufmann, 1993.

[13] Kohavi, "Scaling up the accuracy of Naïve Bayes classifiers: A decision tree hybrid.," in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, 1996, pp. 202-207.

[14] O. Sutton, "Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction The k Nearest Neighbours Algorithm," pp. 1–10, 2012

[15] Wikipedia. [Online]. https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[16]Zhang, A., Lipton, Z. C., Li, M. and Smola, A. J., «Dive into Deep Learning. » http://en.diveintodeeplearning.org, 2018, [Reach Date: 21.12.2019].

[17]Shanmugamani, R. «Deep Learning for Computer Vision» Pactc. 2018. ss 6-7

[18]Baughman, D.R., Liu, Y.A., in Neural Networks in Bioprocessing and Chemical Engineering, 1995

[19]O. Adepoju, J. Wosowei, S. lawte and H. Jaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-6, doi: 10.1109/GCAT47503.2019.8978372.

[20]H. M. M. H. Vidanelage, T. Tasnavijitvong, P. Suwimonsatein and P. Meesad, "Study on Machine Learning Techniques with Conventional Tools for Payment Fraud Detection," 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), Pattaya, Thailand, 2019, pp. 1-5, doi: 10.1109/ICITEED.2019.8929952.

[21]Raghavan, Pradheepan, and Neamat El Gayar. "Fraud Detection using Machine Learning and Deep Learning." 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). IEEE, 2019.

[22]P. Seemakurthi, S. Zhang and Y. Qi, "Detection of fraudulent financial reports with machine learning techniques," 2015 Systems and Information Engineering Design Symposium, Charlottesville, VA, 2015, pp. 358-361, doi: 10.1109/SIEDS.2015.7117005.

[23]S., Rajeswari & Kannan, Suthendran. (2019). Feature Selection Method based on Fisher's Exact Test for Agricultural Data. 10.35940/ijrte.D1104.1284S219.

[24]R., Praveena & ml, Valarmathi & S., Sivakumari. (2011). Gain Ratio Based Feature Selection Method For Privacy Preservation. ICTACT Journal on Soft Computing. 01. 201-205. 10.21917/ijsc.2011.0031.

[25]Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O., Akinjobi J. "Supervised Machine Learning Algorithms: Classification and Comparison". International Journal of Computer Trends and Technology (IJCTT) V48(3):128-138, June 2017. ISSN:2231-2803. www.ijcttjournal.org. Published by Seventh Sense Research Group.