International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 29th May 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=Issue05

10.48047/IJIEMR/V10/ISSUE 05/55

TITLE: CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Volume 10, ISSUE 05, Pages: 255-263

Paper Authors **G. Chakrapani1, Perala Shravya2, Ekkey Sravanthi2, Thimmani Sukumar2, Karantok Pradeep2,Aerukonda Sai Charan2**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

G. Chakrapani[1], Perala Shravya[2], Ekkey Sravanthi[2], Thimmani Sukumar[2], Karantok Pradeep[2], Aerukonda Sai Charan[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science Engineering

[1,2]Malla Reddy Engineering College and Management Sciences, Medchal, Hyderabad

ABSTRACT: Without a doubt, credit card fraud is a form of criminal deceit. Until machine learning methods are introduced, fraud identification appears to remain a challenging task requiring a great deal of talent. Nonetheless, it is an application for the advancement of artificial intelligence and machine learning, guaranteeing that the customer's assets appear safe and unmanipulated. The entire study paper focused on a machine learning-based, efficacious fraud identification system that included a feedback mechanism. Its feedback mechanism is related to improving the efficacy and detection rate of the classifier. The most advanced methods were unable to identify the credit card transaction scams. In order to address these shortcomings, the suggested method uses Support Vector Machine (SVM) classification to identify fraudulent activity. According to the simulation findings, the suggested strategy outperforms state-of-the-art techniques in terms of classification accuracy..

Keywords: SVM, classification technique, transactions, Credit card fraud detection.

I. INTRODUCTION

It is certain that with the advent of deregulation liberalization, globalization and privatization new ways are opened for banks to enhance their revenues by diversifying their product portfolio and offerings. Technology is going to be a major player in enhancing customer delight and a place to provide uniform and integrated banking services to its clients. In fact, all major World Corporations are in the process of procuring and implementing advanced technologies with changing times and banking is no exception to this. Banks are spending considerable amount of time to spread awareness and allocating funds for the same along with offering schemes on the purchase of products/ payment through banking app linked with Debit/ Credit Cards like cash backs and discounts. With the focus of the Indian Government [1] to curb the menace of black money and use of plastic money from one’s account, the usage of mobile banking shall keep rising and more customers would choose the same over traditional banking.

Now these days digital, statistics are very easily available throughout the world because of digital online availability. All the information that also has a large volume, wide range, frequency, as well as importance is stored from small to large organizations over the cloud [2]. The whole information is available from massive amounts of sources such as followers on social media, customer order behaviors, likes, and shares. White-collar crime is the ever-increasing problem with-reaching consequences for the finance sector, business institutions as well as governments. Fraud can indeed be described as illegal deceit to gain financial benefit [3]. Enhanced card transactions had already appreciated a heavy emphasis on communication technology. When credit card transactions are by far the most prevalent form of transaction for offline and online payments, raising the rate of card fraud accelerates as well. Machine learning is the innovation of this century that eliminates conventional strategies and also can function on huge datasets [4] where humans can't immediately access. Strategies of machine learning break within two important categories; supervised learning versus unsupervised learning; Tracking of fraud can also be achieved any form and may only be determined

how to use as per the datasets. Supervised training includes anomalies to always be identified as before.

Many supervised methods [5] are being used over the last few decades to identify credit card fraud. The major obstacle in implementing ML for detecting fraud seems to be the presence of extremely imbalanced databases. Most payments are legitimate in several available evidence sets, with such an extremely small number of fraudulent ones. The significant challenges to investigators are designing the accurate as well as efficient fraud prevention framework that will be low on false positives but efficiently identifies fraud activity [6].

Throughout this study, we introduce an effective credit card fraud identification system with a feedback system, centered on machine learning techniques. That feedback approach contributes to boosting the classifier's detection rate and performance. Also analysis the performance of different classification methods [7] including random forest, tree classifiers, artificial neural networks, supporting vector machine, Naïve Baiyes, logistic regression including gradient boosting classifier approaches, on even a highly skewed credit card fraud database. This complete research paper is divided into different sections including; introduction portion, related activities, credit card fraud obfuscation techniques for machine learning, and the obstacles. Subsequently, the implementation for machine learning techniques as well as the estimation and evaluation of different performance measurement parameters are covered and then the findings of the entire research are covered and also suggested further enhancements.

The major contributions of the paper as follows:

- Preprocessing of the data has been performed, so errors in the data and malwares are effectively removed.
- The SVM method was implemented classification on public available dataset, the results shows that the proposed SVM classification gives the better performance compared to other approaches.

Rest of the paper is organized as follows; section 2 deals with the various literatures with their drawbacks respectively. Section 3 deals with the detailed analysis of the proposed method with its operation. Section 4 deals with the analysis of the results with the comparison analysis. Section 5 concludes the paper with possible future enhancements.

2. LITERATURE SURVEY

Machine learning approaches [8] play a crucial role throughout numerous efficient areas for data processing; one of them is the identification of card fraud. Through previous research, several methods were suggested to include strategies for detecting fraud through supervised methods, unsupervised methods including a hybrid strategy; that makes it necessary and know some technology involved in identifying credit card fraud and have a better understanding of the types of card fraud. Many strategies were suggested and checked. Most of them will be reviewed in the brief following.

Detection of card fraud is focused on an interpretation [9] of the card actions in purchases. Most strategies were implemented throughout the identification of card fraud like artificial neural network (ANN), genetic algorithm (GA), support vector machine (SVM), frequent item set mining (FISM), decision tree (DT), optimization algorithm for migratory birds (MBO) and process for naïve Baiyes (NB). The quantitative logistic regression and naïve bays analysis are conducted in. Bayesian and neural system output is assessed on data concerning credit card fraud [10].

Decision trees, machine learning, and logistical regression are evaluated in fraud detections of the scope.

The article [11] evaluates several innovative methods of machine learning; supporting vector machines including random forests together with logistical regression as part of an attempt towards better detect fraud when applying neural network and logistic regression to identity fraud detection issues.

Credit card identification faces many problems because fraud behavioral models [12] are complex. Which are suspicious transactions appear to look like genuine ones; card transaction sets of data are seldom accessible but extremely imbalanced (and skewed); optimum feature choice (parameters) for models; sufficient measures for test the efficiency of distorted credit card fraud database strategies. The efficiency of credit card fraud detecting becomes greatly affected by both the form of sampling approach utilized, parameter choice as well as identification techniques used [13].

Designed to detect fraudulent activity utilizing conventional manual methods seems to be time-consuming as well as incorrect, rendering such manual techniques more unrealistic to have the emergence of big data. Financial companies have also transformed into intelligent methods. Such intelligent scam methods [14] comprise methods predicated on computing intelligence (CI). Its techniques for detecting statistical scams are split into two categories: supervised and unsupervised methods. Designs are calculated of supervised techniques in fraud detection predicated on both the specimens in fraud as well as valid exchanges to classify duplicate entries when fraud and valid when statistical anomalies 'exchanges will be identified when prospective cases of fraudulent charges in unsupervised fraud detection[15].

3. PROPOSED SYSTEM

We propose a model which detects fraudulent transactions in credit card using Machine Learning techniques. The proposed model treats the fraud detection as binary classification problem. To build this system the major challenge is Class Imbalance Problem. Import the dataset from publically available Kaggle. The format of the dataset is .CSV (Comma Separated Values) file. Prepare the data by removing duplicates and verify that the dataset contains no missing values. Label encoding and one-hot encoding will handle each categorical feature in the dataset. The data consists of attributes of different scales, and several machine models may gain from rescaling the attributes to the same size for all attributes in the data. Attributes are frequently rescaled into the range between 0 and 1. MinMaxScaler is used to rescale the data. A pre-processed dataset will be available and the SVM based machine learning algorithms will be used to assess it. Separating a validation dataset to be used for subsequent confirmation of the developed model's skill. The simple approach we can use to assess the performance of a machine learning algorithm is to use different data sets for training and testing. Due to overfitting we cannot train the machine learning algorithms on the dataset and make predictions from that same dataset to evaluate machine learning algorithms. Fig. 1 represents the proposed system of fraud detection.

Data Preprocessing:

Pre-processing corresponds to the modifications made to the dataset before feeding the algorithm. Several algorithms of machine learning make assumptions about your data. It is often a very good idea to plan the data in such a way that the problem structure is better presented to the machine learning algorithms. Data Pre-processing is a way of transforming original data into a clean dataset. In simple words, when the data is gathered from varied sources, it is acquired in unprocessed form,

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

which is not feasible for evaluation. Raw data (real-world data) is always imperfect and cannot be transmitted through a model. That would trigger some errors. There are various steps in the pre-processing of data. The steps mainly such as:

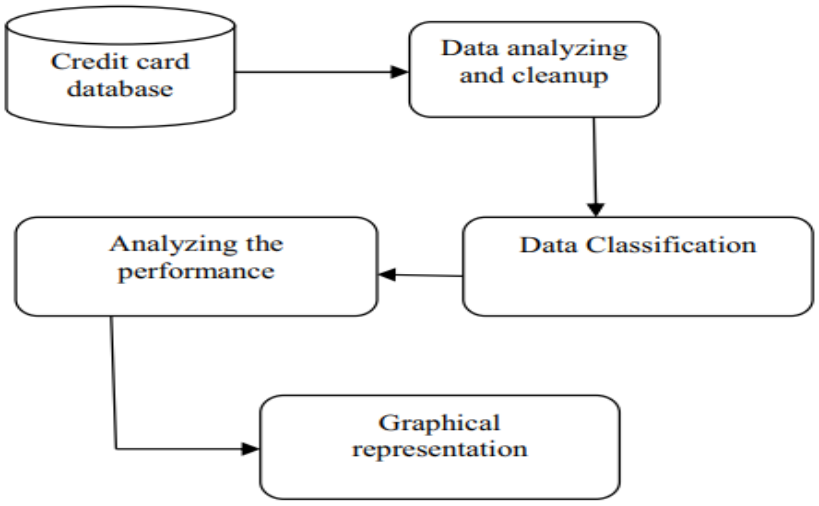


Figure 1: Proposed System architecture.

Data classification

SVM algorithm used for classification and pattern analysis. It is a classification technique to classify or predict patterns into two classes; fraud or legitimate. This Technique is used for binary classifications. SVM is used in classification as well as in pattern recognition such as face recognition, bioinformatics, and text categorization. Risk Minimization theory is developed and supported by SVM.

SVM seems to be a popular algorithm of machine learning with many successful implementations such as XGBoost. While many software optimizations have been incorporated in these implementations, when the function dimension is high and the size of data is huge, the performance and expandability are still unsatisfactory. A major reason is that they have to check all the instances of data for each feature to assess the collection of information from all potential splitting positions, which takes a long time. SVM was proposed to address this issue. SVM is a gradient boosting application that uses tree-based learning algorithms. SVM works primarily on the Histogram-based, and at the same time retains relatively accurate results. SVM is usually faster than other gradient boosting algorithms.

While it's simple to utilize decision trees for numerical features, most datasets have categorical features in reality, which are important for prediction as well. The categorical feature is a feature with a separate set of values that are not identical to each other. Before training, transforming them into numbers is the most common practice when dealing with categorical features of gradient boosting. SVM is a modern gradient boosting algorithm that handles categorical features effectively and takes advantage of handling them during training as compared with the pre-processing time. The other benefit of the algorithm is that it uses a new schema to measure the leaf values when choosing the tree structure, which further enables to minimize the overfitting. The main idea behind this algorithm is to

prevent overfitting, fights against Gradient bias, categorical features support, and provide good results with default parameters.

Table 1: Proposed algorithm
Step 1: Import libraries – Import the required libraries to the working environment. The libraries will be imported such as Numpy, Pandas, sklearn, lightgbm, catboost, etc.
Step 2: Load the data – The credit card fraud dataset of .CSV format will be loaded.
Step 3: Summarize the data – Descriptive statistics, data visualizations will be displayed.
Step 4: Pre-processing data – The data will be pre-processed by checking missing and null values, rescaling the data.
Step 5: Checking for categorical data – The categorical data will be transformed into numerical data by one-hot encoding.
Step 6: Splitting the data – The dataset can be divided into train data of 70% and test data of 30%.
Step 7: Classification: Use the SVM model to detect fraud for incoming transactions.
Step 8: Evaluate algorithms – The models of the proposed method will be applied to data to evaluate the algorithms.
Step 9: Prediction of results – The results are summarized as the accuracy and confusion matrix.

4. EXPERMENTAL RESULTS

Dataset:

The prior information on the dataset, such as its attributes, dimensions, and data types of each feature, etc., is an essential factor that helps one to perform proper operations. An offline dataset which is a publically accessible web platform named “Kaggle” is considered for the implementation of the program. The dataset is a Credit card fraud dataset that consists of several transactions. The dataset contains a combination of cases of fraud and non-fraud. CSV files are the most commonly used format for machine learning data. The dataset contains rows and columns of the following features like Merchant_id, Transaction amount, Is declined, Total Number of declines per day, is Foreign Transaction, is HighRisk Country, and is Fraudulent.

Such dataset descriptions are summarized below in Table 2.

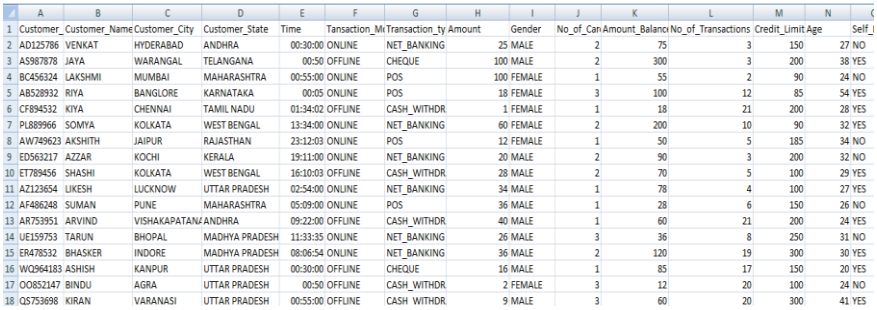
Table 2: Data-set properties

Description	Value
No. of Transactions	3075

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

No. of Attributes	11
Types of Classes	0,1

Performance evaluation:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Customer_	Customer_Name	Customer_City	Customer_State	Time	Tansaction_M	Transaction_ty	Amount	Gender	No_of_Car	Amount_Balanc	No_of_Transactions	Credit_Limit	Age	Self_I
2	AD125786	VENKAT	HYDERABAD	ANDHRA	00:30:00	ONLINE	NET_BANKING	25	MALE	2	75	3	150	27	NO
3	AS987878	JAYA	WARANGAL	TELANGANA	00:50	OFFLINE	CHEQUE	100	MALE	2	300	3	200	38	YES
4	BC456324	LAKSHMI	MUMBAI	MAHARASHTRA	00:55:00	ONLINE	POS	100	FEMALE	1	55	2	90	24	NO
5	AB528932	RIYA	BANGLORE	KARNATAKA	00:05	ONLINE	POS	18	FEMALE	3	100	12	85	54	YES
6	CF894532	KIYA	CHENNAI	TAMIL NADU	01:34:02	OFFLINE	CASH_WITHDR	1	FEMALE	1	18	21	200	28	YES
7	PL889966	SOMYA	KOLKATA	WEST BENGAL	13:34:00	ONLINE	NET_BANKING	60	FEMALE	2	200	10	90	32	YES
8	AW749623	AKSHITH	JAIPUR	RAJASTHAN	23:12:03	ONLINE	POS	12	FEMALE	1	50	5	185	34	NO
9	ED563217	AZZAR	KOCHI	KERALA	19:11:00	ONLINE	NET_BANKING	20	MALE	2	90	3	200	32	NO
10	ET789456	SHASHI	KOLKATA	WEST BENGAL	16:10:03	OFFLINE	CASH_WITHDR	28	MALE	2	70	5	100	29	YES
11	AZ123654	LIKESH	LUCKNOW	UTTAR PRADESH	02:54:00	ONLINE	NET_BANKING	34	MALE	1	78	4	100	27	YES
12	AF486248	SUMAN	PUNE	MAHARASHTRA	05:09:00	ONLINE	POS	36	MALE	1	28	6	150	26	NO
13	AR753951	ARVIND	VISHAKAPATANA	ANDHRA	09:22:00	OFFLINE	CASH_WITHDR	40	MALE	1	60	21	200	24	YES
14	UE159753	TARUN	BHOPAL	MADHYA PRADESH	11:33:35	ONLINE	NET_BANKING	26	MALE	3	36	8	250	31	NO
15	ER478532	BHASKER	INDORE	MADHYA PRADESH	08:06:54	ONLINE	NET_BANKING	36	MALE	2	120	19	300	30	YES
16	WQ964183	ASHISH	KANPUR	UTTAR PRADESH	00:30:00	OFFLINE	CHEQUE	16	MALE	1	85	17	150	20	YES
17	OO852147	BINDU	AGRA	UTTAR PRADESH	00:50	OFFLINE	CASH_WITHDR	2	FEMALE	3	12	20	100	24	NO
18	QS753698	KIRAN	VARANASI	UTTAR PRADESH	00:55:00	OFFLINE	CASH_WITHDR	9	MALE	3	60	20	300	41	YES

Figure 2: Credit card fraud detection

Figure 2 shows the credit card fraud detection analysis of various input.

Table 3: Experimental Results for Various ML Methods

	Random Forest [16]	Naïve Baiyes [21]	Logistic Regression [22]	kNN [23]	Decision Trees [24]	GBM [18]	**SVM**
Precision	93.998	91.201	92.8956	93.228	94.5891	90.998	95.9887
Recall	93.001	91.989	93.112	93.005	92.008	91.996	95.1234
F1-Score	93.998	91.7748	92.112	93.479	91.003	92.778	95.1102
Accuracy	94.001	91.8887	90.448	93.963	94.999	90.998	94.9991
Recall	93.556	91.0021	91.5456	92.789	91.998	91.7752	95.1023
FPR	4.665	4.7789	3.9785	3.889	3.998	4.665	3.9875

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

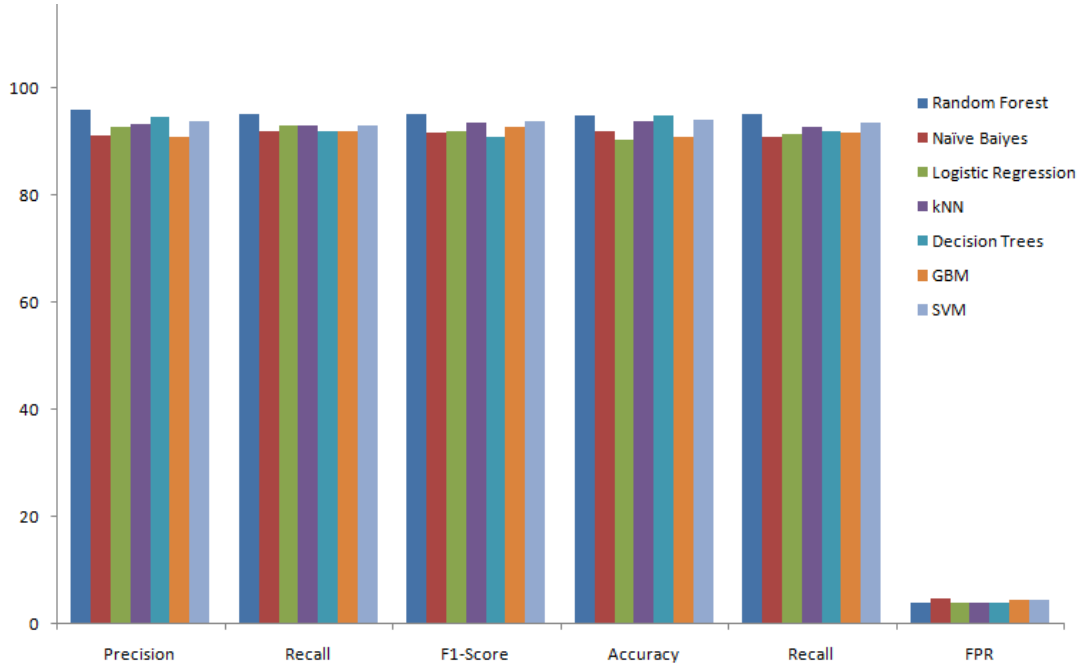


Figure 3: Experimental Results for Various ML Methods

Experimental results from Table 3 as well as Figure 3 demonstrate the percent of the different assessment parameters for just the credit card fraud dataset for distinct machine learning techniques. Findings indicate that SVM techniques demonstrate an accuracy percentage with 95.988 percent, although Random Forest 93.228 percent, LR 92.89 percent, NB 91.2 percent, Decision trees 90.9 percent as well as GBM 93.99 percent demonstrate a precision percentage of ULB machine learning credit card fraud identification. For any machine learning technique, greater values are shown to be accepted as just a higher performance method of precision, accuracy, recall, and F1-score. As we have seen, there are a few algorithms that have surpassed others as well quite significantly. Thus, selecting SVM over all other techniques could be a sensible approach in attaining a greater degree of completeness when decreasing quality just significantly.

5. CONCLUSION

It is certain that with the advent of deregulation liberalization, globalization and privatization new ways are opened for banks to enhance their revenues by diversifying their product portfolio and offerings. This paper investigates performance analysis of Support vector machine's Kernel methods are trained on transactional data and their performances are evaluated and compared based on accuracy, specificity and sensitivity performance metrics. The model is compared with existing classifiers like Naive Bayes, Decision Tree, KNN, and Logistic Regression and SVM. The highly skewed data is sampled where positive-class in down sampled and negative-class in up sampled to convert dataset into balanced dataset. The Results shows that SVM Kernel methods shows great performance for all three performance metrics like sensitivity, accuracy and specificity over traditional techniques. It is analyzed and observed that RBF kernel function outperforms and gives 96% accuracy and 96 % sensitivity compared with other techniques. Linear kernel function gives 90 % as highest sensitivity in comparison with other techniques. This study highlights the performance of SVM kernel classification of imbalanced and skewed data. In future scope multi-classifiers and meta

learning can be consider for highly imbalanced credit card fraud detection. For the evaluation of algorithms, a publically available credit card dataset was used. The accuracy and confusion matrix was adopted as metrics that can be used to evaluate algorithm efficiency. The current system for detecting credit card fraud was built with default parameters. In the future, it can also be designed in such a way that it would prevent overfitting by parameter tuning. Most machine learning models have hyper-parameters. The problem of choosing a set of suitable hyperparameters for a learning algorithm is hyperparameter optimization or tuning. A hyperparameter is a variable for which the learning process is regulated by its value.

REFERENCES

[1]. Awoyemi, J.O., Adetunmbi, A.O. and Oluwadare, S.A., 2017, October. Credit card fraud detection using machine learning techniques: A co parative analysis. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-9). IEEE.

[2]. Adewumi, A.O. and 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), pp.937-953

[3]. Fu, K., Cheng, D., Tu, Y. and Zhang, L., 2016, October. Credit card fraud detection using convolutional neural networks. In International Conference on Neural Information Processing (pp. 483-490). Springer, Cham.

[4]. Yee, O.S., Sagadevan, S. and Malim, N.H.A.H., 2018. Credit card fraud detection using machine learning as data mining technique. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(1-4), pp.23-27.

[5]. Khan, A.U.S., Akhtar, N. and Qureshi, M.N., 2014. Real-time credit-card fraud detection using artificial neural network tuned by simulated annealing algorithm. In Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC (pp. 113-121).

[6]. Carneiro, N., Figueira, G. and Costa, M., 2017. A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, pp.91-101.

[7]. Dhankhad, S., Mohammed, E. and Far, B., 2018, July. Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 122-125). IEEE.

[8]. Adewumi, A.O. and 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), pp.937-953.

[9]. Fiore, U., De Santis, A., Perla, F., Zanetti, P. and Palmieri, F., 2019. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479, pp.448-455.

[10]. Bahnsen, A.C., Stojanovic, A., Aouada, D., and Ottersten, B., 2014, April. Improving credit card fraud detection with calibrated probabilities. In Proceedings of the 2014 SIAM international conference on data mining (pp. 677-685). Society for Industrial and Applied Mathematics.

[11]. Popat, R.R. and Chaudhary, J., 2018, May. A survey on credit card fraud detection using machine learning. In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1120-1125). IEEE.

[12].] Patil, S., Nemade, V. and Soni, P.K., 2018. Predictive modelling for credit card fraud detection using data analytics. Procedia computer science, 132, pp.385-395.

[13]. [13] Malini, N. and Pushpa, M., 2017, February. Analysis on credit card fraud identification techniques based on KNN and outlier detection. In 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication, and Bio-Informatics (AEEICB) (pp. 255-258). IEEE.

[14]. Zareapoor, M. and Shamsolmoali, P., 2015. Application of credit card fraud detection: Based on bagging ensemble classifier. Procedia computer science, 48(2015), pp.679-685.