Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

Developing an Intelligent Credit Card Fraud Detection System with Machine Learning

Omkar Dabade[1], Aditya Admane[2], Deepak Shitole[3*], Vitthal Kamble[4]

[1,2,3,4]Computer Department, Cusrow Wadia Institute of Technology, Pune, India.*

Email: [1]204017.osdcwit@gmail.com, [2]204034.abacwit@gmail.com, [4]vitthalk13@gmail.com
Corresponding Email: [3]204057.dzscwit@gmail.com*

Received: 05 October 2021 **Accepted:** 22 December 2021 **Published:** 30 January 2022

***Abstract:** Credit card fraud is a significant issue in the economic services sector. Each year, billions of rupees are lost due to credit card fraud. Due to confidentiality concerns, there are an absence of studies examining actual credit card records. In this paper, machine learning algorithms are employed to detect credit card fraud. First, standard models are utilized. Then, hybrid techniques consisting of Random Forest, AdaBoost, XGBoost, and majority voting are implemented. To evaluate the effectiveness of the version, a set of publicly accessible credit card records is utilized. Then, credit card records from a real-world economic institution are analyzed. The experimental results suggest that Random Forest and majority voting accomplish precise accuracy estimates for detecting credit card fraud instances.*

***Keywords:** Ada Boost, Classification, Credit Card, Fraud Detection, Predictive Modeling, Majority Voting.*

1. INTRODUCTION

According to scholarly sources, social fraud refers to a deceptive or illicit act that is intended to result in financial or personal benefit [1]. To mitigate the impact of fraudulent activities, two strategies can be employed, namely fraud prevention and fraud detection. Fraud prevention is a proactive approach that aims to prevent fraudulent activities from occurring in the first instance. Fraud detection is a necessary measure in cases where a fraudulent transaction is attempted by an individual with fraudulent intent.

The unauthorized use of credit card information for the purpose of making purchases is an important issue commonly referred to as credit card fraud.Credit card transactions can be conducted through physical means or digital channels [2]. Throughout the course of a physical transaction, the credit card is consistently utilized. In the context of digital transactions, this phenomenon may manifest through either cellular devices or internet-based

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

platforms. Typically, individuals who possess a card provide the card number, expiration date, and card verification code through either a telephonic or online platform.

The utilization of credit cards has experienced a significant surge in recent years, coinciding with the rapid growth of e-commerce [3]. The volume of credit card transactions in India witnessed an increase from an estimated 100 million in 2017 to around 200 million in 2019. The incidence of fraud has been consistently increasing in accordance with the rise of credit card usage. Despite the implementation of multiple authorization strategies, instances of credit card fraud have persisted without significant hindrance. The internet is a preferred medium for fraudsters due to the concealment it provides for their identities and whereabouts. The increasing incidence of credit card fraud has major consequences for the economic sector. According to a report, the global credit card fraud in the year 2015 amounted to a staggering sum of USD 21.84 billion [4]. The incidence of credit card fraud results in financial losses for traders, who bear the burden of all associated costs, including fees charged by the card issuer, as well as administrative expenses [5]. Due to the necessity for traders to incur losses, certain products may be subject to higher pricing, while discounts and incentives may be reduced. Therefore, it is crucial to minimize losses by implementing an efficient fraud detection system that can effectively reduce or eliminate instances of fraudulent activities. Several research studies have been conducted on the topic of detecting credit card fraud. The utilization of machine learning and its related techniques is prevalent in various fields. These techniques include artificial neural networks, rule-induction techniques, decision trees, logistic regression, and support vector machines, as documented in reference [1]. These methodologies employ either individual or integrated approaches to construct hybrid models.

The present study has examined most effective machine learning algorithms for the purpose of detecting fraudulent credit card transactions. The selection of algorithms ranges from commonly favored neural networks to sophisticated deep learning models. The evaluation process involves the utilization of benchmark and real-world credit card data units. Moreover, hybrid models are formed by implementing XGBoost and majority voting techniques. In order to enhance the assessment of the models' durability and dependability, changes are introduced to the data. The principal contribution of the present study involves the evaluation of various machine learning models utilizing authentic credit card data for the purpose of detecting fraudulent activity. Simultaneously, diverse scholars have employed various methodologies on openly accessible datasets. The present study utilizes a dataset derived from authentic credit card transaction records spanning a period of three months.

Assosciated Research

This paper provides a comprehensive review of both single and hybrid machine learning algorithms utilized in economic applications. A range of economic applications, including credit card fraud and financial statement fraud, are subject to analysis.

A. Single Models

In a previous study [6], the performance of various algorithms, including Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LOR), was evaluated for the purpose of detecting instances of credit card fraud. The dataset comprised transactions

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

spanning a period of one year. The study utilized information below-sampling to evaluate the performances of algorithms, revealing that RF exhibited superior performance in comparison to SVM and LOR [6]. In previous research [7], a system utilizing artificial intelligence and rule-based techniques (AIRS) was suggested for the purpose of detecting instances of credit card fraud. The AIRS model represents a novel advancement from the conventional AIS approach, as it leverages the utilization of negative choice to achieve enhanced precision. As a result, there was a 25% increase in accuracy and a 40% reduction in system response time. [7]. In an earlier research project [8], a proposed system for detecting credit card fraud was presented. The system consists of a rule-based filter, a Dumpster-Shafer adder, a database of transaction records, and a Bayesian learner. The Dempster-Shafer principle was employed to integrate multiple sources of evidential data and generate an initial inference, which was subsequently utilized to classify a given transaction as either normal, suspicious, or anomalous. In cases where a transaction was deemed questionable, Bayesian learning was employed to further assess the situation by analyzing transaction records [8]. The simulation results demonstrated a genuine positive rate of 98% as reported in reference [8]. The detection of credit card fraud was performed utilizing a modified version of the Fisher Discriminant principle, as documented in reference [9]. The modification resulted in the conventional functions exhibiting increased sensitivity towards crucial instances. The calculation of variances was performed using a weighted average approach, facilitating the identification of valuable transactions. The results obtained from the modified function provide evidence that it has the potential to yield higher profits [9]. The utilization of association rules has been implemented to extract patterns of behavior in relation to instances of credit card fraud, as described in reference [10]. The dataset was focused on retail establishments located in Chile. The data samples were subjected to de-fuzzification and processing through the utilization of the fuzzy query 2+ data mining tool [10]. The output obtained led to a reduction in the excessive range of guidelines, thereby facilitating the task of fraud analysts [10]. A proposal has been put forth in reference [11] with the aim of improving the identification of occurrences of credit card fraud. The study utilized a dataset sourced from a financial institution in Turkey. Each transaction was classified as either fraudulent or non-fraudulent. The implementation of the Genetic Algorithm (GA) and Scatter Search has resulted in a reduction of misclassification quotes. The technique under consideration exhibited a twofold increase in performance as compared to the prior results cited in reference [11].

B. Hybrid Models

Hybrid models refer to a combination of multiple individual models. In a previous study [12], a hybrid approach was employed to detect instances of corporate tax evasion. This approach incorporated various techniques, including the Multilayer Perceptron (MLP) neural network, Support Vector Machines (SVM), Logistic Regression (LOR), and Harmony Search (HS) optimization. The utilization of HS proved advantageous in determining optimal parameters for the classification models. The MLP model with HS optimization achieved a peak accuracy rate of 90.07% by leveraging data from the food and fabric industries in Iran. [12]. In a previous study [13], a hybrid clustering system was employed to identify instances of fraudulent activity in the context of lottery and online games, which also included the capability of detecting outliers. The system utilized online algorithms and statistical analysis of input data to identify various types of fraudulent activities. The primary memory was

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

utilized to compress the training data set, while the stored information cubes will be updated incrementally with new data samples. According to reference [13], the system achieved a detection rate of 98% and a false alarm rate of 0.1%. Hybrid models have been developed in [14] utilizing clustering and classifier ensemble techniques to tackle financial distress. The clustering task has been accomplished through the utilization of the SOM and k-means algorithms, whereas classification has been performed by employing LOR, MLP, and DT. A set of techniques were employed to generate 21 hybrid models, each with unique combinations, which were subsequently assessed using the dataset. According to the findings presented in reference [14], the self-organizing map (SOM) in conjunction with the multilayer perceptron (MLP) classifier demonstrated superior performance by achieving the highest level of prediction accuracy. In a previous study [15], a fraud detection model for company financial statements was developed through the integration of multiple models, including RF, DR, Roush Set theory (RST), and the back-propagation neural network. The dataset utilized for the study comprises the financial statements of the company spanning the period from 1998 to 2008. The results indicated that the amalgamated approach of RF and RST yielded the greatest level of precision in classifying data [15]. The methodologies for detecting instances of fraudulent automobile insurance have been explicated in references [16] and [17]. In a previous study [16], a method was proposed that utilizes a principal component analysis (PCA) approach in conjunction with a potential nearest neighbor technique within a random forest (RF) framework. The traditional method of majority voting in Random Forest was replaced by the potential nearest neighbor approach. The experimental study utilized a total of 12 unique data sets. According to a study [16], the model based on PCA demonstrated superior classification accuracy and lower variance in comparison to those generated by RF and DT methodologies. In order to detect instances of automobile insurance fraud, a GA with Fuzzy C means (FCM) was introduced in a previous study (reference [17]). The test data was classified into distinct categories of genuine, malicious, or suspicious classes, determined by the clusters that were generated. The suspicious instances were subjected to further analysis through the utilization of DT, SVM, MLP, and GMDH techniques, following the removal of both genuine and fraudulent data. According to the findings [17], the Support Vector Machine (SVM) demonstrated superior rates of specificity and sensitivity.

Machine Learning Algorithms

The current research study employs variety of algorithms. These techniques are employed alongside with AdaBoost and majority voting. The data provided is presented as follows:

A. Algorithms

The Decision Stump (DS) algorithm produces a decision tree that contains only one split. The utilization of this method is applicable in the classification of non-uniform data sets. Classification models can be built with either strong or naïve independence assumptions. It is assumed in certain cases that certain capabilities of a class are no longer correlated with others. In order to calculate the means, it is necessary to use estimation techniques. A minimal amount of training data is sufficient, and classification based on variances is necessary. The utilization of a tree structure for data presentation is advantageous in terms of facilitating comprehension for clients. The Decision Tree (DT) is a collection of nodes that facilitates decision-making processes pertaining to specific classes. Each individual node

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

serves as a decision point that defines a partitioning criterion for a given function. Nodes are created until the specified criterion is satisfied. The assignment of a class label is reliant upon the amount of samples that relate to a given leaf. The Random Tree (RT) functions as a Decision Tree (DT) operator, but with the modification that during each split, only a subset of capabilities is randomly accessible. The algorithm is capable of acquiring knowledge from both nominal and numerical data sets. The parameter that characterizes the length of a subset is the subset ratio. The Random Forest (RF) algorithm generates a collection of decision trees through a collective learning approach, where each tree is constructed using a random subset of the available features. The range of trees is determined by the user. The ultimate classification outcomes are determined by the collective voting of all the constructed trees in the resultant model. The Gradient Boosted Tree (GBT) is an ensemble model that combines multiple classification or regression models. It employs a forward-learning approach to progressively improve its predictive accuracy by using increasingly sophisticated estimations. The utilization of boosting techniques facilitates the improvement of tree models' accuracy. The Decision Stump (DS) algorithm produces a decision tree that contains only one split. It has the potential to be employed in the classification of disparate data sets. The utilization of the backpropagation algorithm for training is also observed in the Feed-forward Neural network (NN). The absence of a directed cycle in the interconnections of the devices ensures that data transmission occurs unidirectionally, specifically from input nodes to output nodes, through the intermediary hidden nodes. The foundation of deep learning (DL) is predominantly rooted in a multilayer perceptron (MLP) network that is trained through the utilization of stochastic gradient descent with backpropagation. The architecture comprises a substantial quantity of concealed layers, incorporating neurons that possess activation characteristics such as tanh, rectifier, and max out. Every individual node acquires a copy of the global model parameters that pertain to the data in its vicinity and periodically makes a contribution towards the global model through the process of model averaging. Linear Regression (LIR) is a statistical method that models the relationship between scalar variables by fitting a linear equation to the observed data. Linear predictor features are utilized to model the relationships, whereby the model parameters, which are unknown, are predicted from the dataset. The Akaike criterion is a measure of the relative goodness of fit of a statistical model, which is utilized in the process of selecting the most appropriate model. Logistic Regression (LOR) is a statistical technique that is capable of handling data with both nominal and numerical attributes. The estimation of the likelihood of a binary reaction is based solely on one or more predictor functions. The Support Vector Machine (SVM) is capable of handling both classification and regression tasks. The Support Vector Machine (SVM) algorithm constructs a model by classifying new samples into one of two categories, thereby creating a non-probabilistic binary linear classifier. The technique involves the representation of data samples as factors in a mapped area, with the aim of maximizing the separation between data samples of different classes through the creation of a margin.

B. Majority Voting

The concept of majority voting is a method of decision-making in which a proposal is accepted or rejected based on the number of votes it receives. The employment of majority voting is a common practice in the realm of data classification, wherein a hybrid model incorporating minimal algorithms is employed. Each algorithm generates a unique prediction for each test pattern. The final output pertains to the option that receives the highest number

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

of votes, as indicated below. Let us consider a scenario where there are k target classes or labels denoted by Ci, where i belongs to the set Λ = {1, 2, ..., k}. Each Ci represents the i-th target class that is anticipated by a classifier. Upon receiving an input x, multiple classifiers generate predictions for the target class, resulting in k predictions denoted as P1 through PK. The objective of majority voting is to generate a blended forecast for a given input x, where P(x) = j and j belongs to the set Λ, by considering all k predictions, denoted as pk(x) = jk, where k ranges from 1 to k. A binary function can be employed to represent the votes [21]. Specifically, the votes from all k classifiers for each Ci are aggregated, and the label with the highest vote is determined as the final (combined) predicted class.

C. Adaboost

The AdaBoost algorithm is a popular machine learning technique used for classification and regression tasks. The AdaBoost algorithm is commonly employed together with diverse types of algorithms to boost their performance. The combined outputs are obtained through a weighted sum, which signifies the blended output of the boosted classifier [21]. The AdaBoost algorithm prioritizes the classification of mislabeled data samples over novice learners who may be more susceptible to error. Nevertheless, the sensitivity of the data to noise and outliers should be taken into consideration. Provided that the classifier's performance is not arbitrary, AdaBoost has the potential to improve the individual results obtained from different algorithms.

Experimental Work

In the dataset associated to credit card transactions, the frequency of fraudulent transactions is typically negligible in comparison to the total number of transactions. When a dataset is uneven, the accuracy obtained may not provide an accurate representation of the overall performance of the system. Incorrectly categorizing a legitimate transaction can result in adverse customer experiences, while the failure to identify instances of fraud can lead to financial losses for both the financial institution and its clients. The issue of information imbalance has been found to result in performance challenges for machine learning algorithms. The outcome is influenced by the class that has the highest number of samples. Bhattacharyya et al. [6], Duman et al. [18], and Phua et al. [19] have employed the technique of under sampling combined with semi-supervised learning and AutoEncoders to detect fraudulent credit card transactions. The authors in reference [21] provided a comparative summary of semi-supervised anomaly detection algorithms. Although there is no universally agreed upon method for characterizing true and false positives and negatives with a single metric, the Matthews Correlation Coefficient (MCC) [20] is widely regarded as the most reliable and widely accepted standard measure. The Matthews correlation coefficient (MCC) is a metric utilized to evaluate the performance of binary classification models. It accounts for both true and false positives and negatives in a two-class problem, thereby providing a comprehensive measure of satisfaction. The measure is equitable, even in cases where the classes exhibit dissimilar magnitudes.

Existing System

The existing systems utilize three distinct approaches for identifying instances of fraud. The grouping model is employed to classify legitimate and fraudulent transactions through data clustering of boundary value regions. The Gaussian blend model is employed to demonstrate

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

the probability density function of a credit card user's historical behavior, thereby enabling the estimation of the likelihood of current behavior and the identification of any deviations from past behavior. Bayesian systems are employed to represent the metrics of an individual user and the metrics of multiple fraud scenarios. The primary objective is to examine diverse viewpoints connected to a common matter and identify the benefits that can be derived from employing each distinct methodology.

Proposed System

The primary objective of this project is to conduct a thorough examination of various fraud detection methodologies and new machine learning approaches. The utilization of credit cards in financial transactions is progressively gaining popularity; however, simultaneously, the incidence of fraudulent activities is also on the rise. Traditional approaches rely on rule-based expert systems for identifying fraudulent behaviors, which may not account for various scenarios and significant imbalances between positive and negative samples. The present study introduces a fraud detection framework based on random forest methodology, which aims to capture the inherent patterns of fraudulent behaviors acquired from labeled data. The credit dataset is utilized as input for the machine learning algorithm to facilitate its training. The pre-processing of transactional data involves the removal of noise and redundant information. Subsequently, salient characteristics are extracted and compared with the dataset. Then, credit card transactions are categorized into fraudulent or legitimate ones through the utilization of a machine learning classifier.

Dataset:

The dataset under consideration involves to credit card transactions. In the context of machine learning, the process of training a model refers to the acquisition of optimal weight and bias values through exposure to labeled examples. Supervised learning involves the construction of a model by a machine learning algorithm through the analysis of numerous examples, with the aim of minimizing loss. This procedure is referred to as empirical risk minimization. The process of identifying and selecting relevant characteristics or attributes from raw data, with the goal of reducing the amount of The process of extracting and matching features relies on the utilization of these measures. In addition to the basic point feature, a more sophisticated type of feature is also introduced. The utilization of feature extraction techniques enables the extraction of pertinent features from a vast image dataset while retaining maximal information.

Classification:

Classification is a fundamental procedure that involves the organization of a given dataset into distinct categories. This process can be applied to data that is either structured or unstructured. The initial step involves making predictions regarding the classification of provided data points. The terms commonly used to refer to the classes includes target, label, or categories. The process of approximating the mapping function from input variables to discrete output variables is commonly referred to as classification predictive modeling. The primary objective is to figure out the classification or categorization of the incoming data. The process of categorization is executed through the utilization of machine learning algorithms.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

2. CONCLUSION

This paper presents a research investigation on the implementation of machine learning algorithms for the purpose of detecting credit card fraud in various systems. The empirical assessment utilized several standard models, including RF, NB, SVM, and DL. The evaluation of individual (standard) models and hybrid models utilizing AdaBoost and majority voting combination techniques has been conducted using a publicly accessible dataset of credit card information. The MCC (Matthews Correlation Coefficient) metric has been utilized as a comprehensive performance evaluation tool, as it accounts for both true and false positive and negative outcomes. In subsequent research endeavors, the methodologies examined in this the paper may be extended to incorporate online learning frameworks. Moreover, various models of online learning could be explored. The implementation of online learning has the potential to facilitate prompt identification of fraudulent incidents potentially in a real-time manner. Consequently, this will aid in the detection and prevention of fraudulent transactions prior to their occurrence, thereby mitigating the daily financial losses incurred.

3. REFERENCES

1. Y. Sahin, S. Bulkan, E. Duman, "A cost-sensitive decision tree approach for fraud detection", Expert Syst. Appl., vol. 40, pp. 5916-5923, 2013.
2. A. O. Adewumi, A. A. Akinyelu, "A survey of machine learning and nature-inspired based credit card fraud detection techniques", Int. J. Syst. Assurance Eng. Manage., vol. 8, no. 2, pp. 937-953, 2017.
3. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection using hidden Markov model", IEEE Trans. Depend. Sec. Comput., vol. 5, no. 1, pp. 37-48, Jan. 2008.
4. The Nilson Report, Oct. 2016, [online] Available: https://www.nilsonreport.com/upload/content_promo/The_Ni lson_Report_10-17-2016.pdf.
5. J. T. Quah, M. Sriganesh, "Real-time credit card fraud detection using computational intelligence", Expert Syst. Appl., vol. 35, no. 4, pp. 1721-1732, 2008.
6. S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, "Data mining for credit card fraud: A comparative study", Decision Support Syst., vol. 50, no. 3, pp. 602-613, 2011.
7. N. S. Halvaiee, M. K. Akbari, "A novel model for credit card fraud detection using artificial immune systems", Appl. Soft Comput., vol. 24, pp. 40-49, Nov. 2014.
8. S. Panigrahi, A. Kundu, S. Sural, A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster– Shafer theory and Bayesian learning", Inf. Fusion, vol. 10, no. 4, pp. 354-363, 2009.
9. N. Mahmoudi, E. Duman, "Detecting credit card fraud by modified fisher discriminant analysis", Expert Syst. Appl., vol. 42, no. 5, pp. 2510-2516, 2015.
10. D. Sánchez, M. A. Vila, L. Cerda, J. M. Serrano, "Association rules applied to credit card fraud detection", Expert Syst. Appl., vol. 36, no. 2, pp. 3630-3640, 2009.
11. E. Duman, M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search", Expert Syst. Appl., vol. 38, no. 10, pp. 13057-13063, 2011.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 02, No. 01, Dec 2021 - Jan 2022
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.21.45.53

12. I. T. Christou, M. Bakopoulos, T. Dimitriou, E. Amolochitis, S. Tsekeridou, C. Dimitriadis, "Detecting fraud in online games of chance and lotteries", Expert Syst. Appl., vol. 38, no. 10, pp. 13158-13169, 2011.
13. C.-F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress", Inf. Fusion, vol. 16, pp. 46-58, Mar. 2014.
14. F. H. Chen, D. J. Chi, J. Y. Zhu, "Application of random forest rough set theory decision tree and neural network to detect financial statement fraud—Taking corporate governance into consideration", Proc. Int. Conf. Intell. Comput., pp. 221-234, 2014.
15. Y. Li, C. Yan, W. Liu, M. Li, "A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification", Appl. Soft Comput..
16. S. Subudhi, S. Panigrahi, "Use of optimized Fuzzy C-Means clustering an supervised classifiers for automobile insurance fraud detection", J. King Saud Univ.-Comput. Inf. Sci..
17. M. Seera, C. P. Lim, K. S. Tan, W. S. Liew, "Classification of transcranial Doppler signals using individual and ensemble recurrent neural networks", Neurocomputing, vol. 249, pp. 337-344, Aug. 2017.
18. A. Phua, K. Smith-Miles, V. Lee, R. Gayler, "Resilient identity crime detection", IEEE Trans. Knowl. Data Eng., vol. 24, no. 3, pp. 533-546, Mar. 2012.
19. DZAKIYULLAH, Nur Rachman,PRAMUNTADI, Andri, et FAUZIYYAH, Anni Karimatul. Semi-Supervised Classification on Credit Card Fraud Detection using AutoEncoders. Journal of Applied Data Sciences, 2021, vol. 2, no 1, p. 01-07
20. Credit Card Fraud Detection, Nov. 2017, [online] Available: https://www.kaggle.com/dalpozz/credicardfraud.
21. VILLA-PÉREZ, Miryam Elizabeth, ÁLVAREZ- CARMONA, Miguel Á., LOYOLA-GONZÁLEZ, Octavio, et al. Semi-supervised anomaly detection algorithms: A comparative summary and future research directions. Knowledge-Based Systems, 2021, vol. 218, p. 106878