Credit Card Fraud Detection by Improved SVDD

Ayoub Mniai, Khalid Jebari

***Abstract*—The COVID-19 pandemic has brought dramatic changes in human beings' habits. One of these major changes is the increase use of credit card. Online shopping has become necessary to satisfy customers' needs during the pandemic. However, this kind of shopping opened a new way to hack information. Several research studies have focused on automatic and real-time online credit card fraud detection. In this context, machine learning (ML) techniques have played a considerable part in these studies, thanks to their characteristics that provide a model capable of detecting fraudulent transactions. This article aims to design a hybrid model for credit card fraud detection. Our hybrid solution combines the Support Vector Data Description (SVDD) and the Particle Swarm Optimization (PSO). For instance, SVDD is known by a random choice of two parameters, c and σ, which contribute to its efficiency. The proposed model uses the PSO algorithm, known by its speed, to find an optimal solution to optimize these two parameters to obtain better accuracy. Simulation results of real datasets indicate SVDD-PSO's performance compared to other machine learning techniques.**

***Index Terms*—Metaheuristics, Particle Swarm Optimization, Machine Learning, Support Vector Machine, Support Vector Data Description, Credit Card Fraud Detection.**

I. Introduction

GLOBAL e-commerce volumes have increased during the COVID-19 pandemic. Indeed, electronic credit card transactions have become a daily reality. However, the rapid growth in credit card transactions has led to an unfortunate increase in fraud cases [1]. As reported by Julie Conroy [2], a research director for Aite Group's fraud and anti-money laundering practice said, "Our estimate was that at the end of 2020, the US was seeing about 11 billion worth of losses due to credit card fraud". These fraudulent transactions committed by the third party can affect bank-customer relationships and result in financial losses for both parties. The purpose of credit card fraud is to obtain money or make payments without the owner's permission. This involves the illegal use of the card or card information without the owner's authorization.

Consequently, it has taken multiple steps toward preventing credit card fraud by different actors. On the other hand, ML techniques have offered practical algorithms for automatic credit card fraud detection. Various models have been provided, which divided into three categories: supervised, unsupervised, and semi-supervised techniques [3], [4], [5], [6], [7], [8], [9].

Supervised techniques focus on studying different past transactions, which are reported by the cardholder or credit card company, to predict whether any new transaction is fraud or not. This technique requires a labeled dataset as fraud and non-fraud observations [10], [11]. For instance, the authors in [13] provided a comparison of some established supervised learning algorithms to differentiate between genuine and fraudulent transactions. Another contribution evaluated the performance assessment of an imbalanced dataset by using supervised ML algorithms to identify the most delicate mechanism for the recognition of credit card scams [14]. In [15] the authors studied the comparison between different classifiers based on Random Forest, SVM, Decision Tree, logistics regression and oversampling by using SMOTE technique for fraud detection. Then, a comparative study on credit card fraud detection based on different SVM proposed in [16]. Finally, the authors in [17] presented Financial Fraud Detection using Deep SVDD.

Unsupervised techniques require an organization of unlabeled data into similarity groups called clusters. They rely on the assumption that outliers are fraud transactions. Clustering allows the identification of different data distributions for which different predictive models should be used [10],[18]. For example, the authors in [19] evaluated the performance of three unsupervised machine learning algorithms namely Local Outlier Factor, Isolation Forest Algorithm and K-means clustering on imbalanced credit card fraud data. Moreover, in [20] the authors presented a survey on unsupervised algorithms for Fraud Detection on the available sample of Bitcoin dataset. Finally, the authors in [21] evaluated the performance of the isolated forest algorithm for fraud detection in health care systems.

Semi-supervised ones combine the previous approaches to take advantage of learning past illegal transactions and applying unsupervised techniques to detect new transaction behavior. For example, in [23] the authors presented a hybrid technique that combines different machine learning techniques such as support vector machine (SVM), multilayer perceptron (MLP), random forest regression, autoencoder and isolation forest in order to detect fraudulent transactions in credit card. In [24] the authors combined semi-supervised learning and AutoEncoders to identify fraudulent credit card transactions. Then, the authors in [25] presented semi-supervised anomaly detection algorithms with a comparative summary.

Motivated by these contributions, this paper presents a hybrid machine learning model. In particular, we examine the benefits of combining the PSO and the SVDD for building a reliable credit card fraud detection model. The structure of the paper is as follows. In the next section, we will recall the algorithms used in the experiment. Then, the proposed method is outlined. In section 3, we evaluate our proposed model on real datasets. Finally, conclusions and future work directions are presented.

Manuscript received on March 30, 2022; revised on April 27, 2022. This work was supported by LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco.

A.Mniai is with LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco;(e-mail: ayoubm.m@gmail.com).

K.Jebari is with LMA, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco;(e-mail: khalid.jebari@gmail.com).

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

II. Methods

This section describes the algorithms used in this article: Support Vector machines (SVM) [26], SVDD [27] and PSO [28], [29].

A. Support Vector machines

Support Vector Machine (SVM) was first introduced by Boser, Guyon, and Vapnik to get a better solution to decision boundary [26]. In other words, SVM is able to separate data into two classes or groups. Also, it has been an active research area in many fields [41], [42], [43], [48]. So the purpose of this technique is to find the best linear classifier, called Hyperplane that separates two or more groups with a maximum margin between decision classes. This margin is defined by training examples called Support Vectors,as showed in Figure 1.

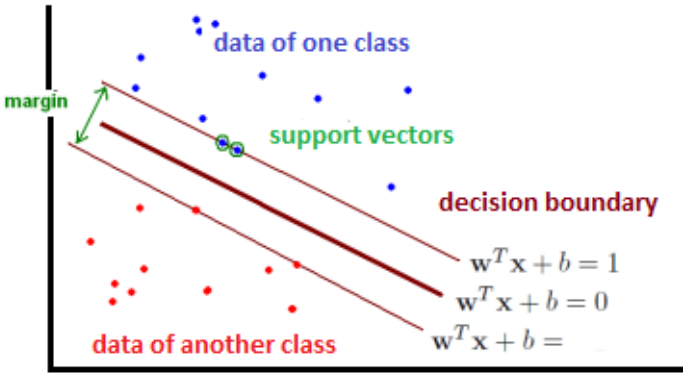

Figure 1: Principle of the SVM classifier.

For data represented as $\{x_i,\ i = 1, 2, ..., N\}$ The Hyperplane is defined using the following equation:

$$f(x) = w.x + b \tag{1}$$

Where w is the normal vector to the hyperplane. If $f(x) \geq 0$ then x is in class 1 otherwise it is in class -1. The optimal hyperplane that separates classes with a maximum margin is found by minimizing the following equation:

$$Min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i \tag{2}$$

Subject to $y_i(w^T x_i + b) \geq 1 - \xi_i\ ,\ \xi_i \geq 0$

Where C is a regularization parameter that determines the trade-off between maximum margin and the minimum classification error ξ_i is called slack variable. In some cases where the classes are not linearly separable they can be represented in a larger dimensional space using a technique call kernel trick, the kernel function is represented as follows:

$$K(x_i, x_j) \equiv \varphi(x_i)^T \varphi(x_j) \tag{3}$$

Different kernel functions are used as mentioned below. The equation of these Kernel functions are as follows:
Radial basis function:

$$K(x_i, x_j) = e^{-|x_i - x_j|^2/2\sigma^2} \tag{4}$$

where σ is the band-width of the Gaussian Radial basis kernel.
Polynomial function:

$$K(x_i, x_j) = (x_i.x_j + 1)^d \tag{5}$$

where, d is the degree of polynomial function.
Linear function:

$$K(x_i, x_j) = ({x_i}^d.x_j) \tag{6}$$

These parameters should be correctly and carefully chosen as it defines the structure of the high dimensional feature space $\varphi(x)$ and thus controls the complexity of the solution.

B. Support Vector Data Description

Support Vector Data Description has been introduced by Tax and Duin to address the problem of anomaly detection [27], which is inspired by the Support Vector Classifier[26]. The basic idea of SVDD is to determine the smallest sphere around a given data points to find the positive target inside the sphere in feature space as show in Figure 2. And, all data points those are outside the hyper-sphere are considered as outliers (negative target).

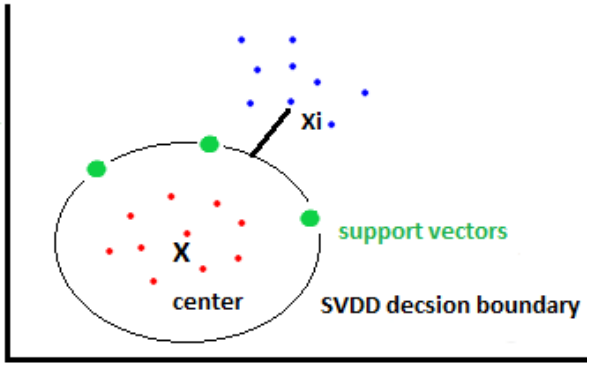

Figure 2: Principle of the SVDD classifier.

In other words, the distance from x_i to the center g should be strictly smaller than the minimum radius R otherwise should be penalized. So that, a slack variable ξ_i has been introduced and the formulation leads to the following optimization problem:

$$F(R, a) = R^2 + C\sum_i \xi_i \tag{7}$$

With:

$$\|x_i - a\|^2 \leq R^2 + \xi_i,\ i = 1, 2, ..., n \tag{8}$$

The parameter C controls the trade-off between the size of the hyper-sphere and the number of target points located outside the sphere. For example, a test sample z is considered as a positive target within the hyper-sphere when the distance is smaller than or equal to the radius R:

$$\|z - g\|^2 = (z.z) - 2\sum_i \alpha_i(z.x_i) + \sum_{i,j} \alpha_i\alpha_j(x_i.x_j) \leq R^2 \tag{9}$$

Also, the inner (x_i, x_j) can be replaced by a kernel function $K(x_i, x_j)$. This procedure defines the feature space and generates a decision boundary for input data points. Using the kernel reduces the complexity of the optimization problem that now only depends on the input space instead of the feature space. The current study aims to implement only three main SVDD kernel functions types which are Polynomial, Linear and Radial basis kernel functions.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

C. Particle Swarm Optimization

PSO is a nature inspired algorithm that optimizes iteratively a problem to improve a candidate solution with regard to a given measure of quality [28],[29]. It PSO is initially developed by Russell Eberhart and James Kennedy and in 1995 to describe the social behavior of birds and fish, it is easy to implement and is computationally inexpensive in terms of memory requirements and speed [28]. PSO is one of the most useful and famous metaheuristics and it is successfully applied to various optimization problems [44],[45]. The idea of the algorithm is to create a swarm of particles which move in the space around looking for their goal. Two optimization properties that are behind the PSO algorithm:

• A particle obtains its best position from the own experience, but also takes knowledge from the other particles movements, the evaluation of positions is done through a fitness function, the specification of this function depends on the problem being optimized.

• For a best exploration of the problem space, a stochastic factor in each particles velocity makes them move through unknown problem space regions. Each particle is characterized by a position and a velocity calculated as follows:

$$x_i(t+1) = x_i + v_i(t+1) \tag{10}$$

$$v_i\,(t+1) = v_i(t) + C1 * Rand_{(0,1)}\,[PBest_i(t) - x_i(t)]$$
$$+C2 * Rand_{(0,1)}\,[GBest_i(t) - x_i(t)] \tag{11}$$

Where :
xi = Position of the particle i vi= Velocity of the particle i PBesti = Personal best position for the particle i GBest = Global best position Rand(0,1)= Random value between 0 and 1 C1: acceleration constant corresponding to the Cognitive component C2: acceleration constant corresponding to the social component
An inertia Weight (w) can be used to control the velocity that affects the convergence, exploration, and the exploitation processes in the algorithm.

$$v_i\,(t+1) = w * v_i(t) + C1 * Rand_{(0,1)}\,[PBest_i(t) - x_i(t)]$$
$$+C2 * Rand_{(0,1)}\,[GBest_i(t) - x_i(t)] \tag{12}$$

The way positions and velocities are initialized can have an important impact on the performance of the algorithm.
Algorithm:

- ***Initialize Population:***
- ***while*** (condition=true)
- ***for*** i=1 to Population size
  - ***If*** $x_i < Pbest_i$ then $Pbest_i = x_i$
    $Gbest = mini\ Pbest_i$
  - ***End***
  - ***For*** d=1 to $Dimension$
    $v_{i,d}(t+1) = v_{i,d}(t) + C1 * R1\,(PBest_i(t) - x_{i,d}(t)]$
    $x_{i,d}(t+1) = x_{i,d}(t) + V_{i,d}(t+1)$
  - ***End***
- ***End***

III. The Proposed Model

All the techniques mentioned above have proposed solutions for fraud detection during the last decade. However, it still requires more contributions to provide a model capable to produce satisfied performance at the level of hyper-parameter initialization. Indeed, the effectiveness of SVDD depends on selecting an appropriate parameters c and σ [27]. The parameter c controls the trade-off between the size of the sphere and the number of negative target points assigned outside the sphere. In other words, increasing the value of c will allow a more positive target to fall outside the class boundary. At the same time, the width parameter σ regulates the number of support vectors, which works as a balance between the numbers of support vectors and the size of the sphere [27]. At this stage, a heuristic method should be integrated in the initialization phase in order to find an optimal solution for these parameters. Particle swarm optimization can be seen as a very promising solution for this task [44], [45]. And, it is now one of the most commonly used optimization technique.
To recap, the hybrid model is formed of three steps. First, a Particle Swarm Optimization algorithm is applied to select the optimal solution for c and σ. Then, a process of fraudulent transactions identification is made. Finally, an evaluation performance process is proposed using the confusion matrix to evaluate the model accuracy. The diagram of the proposed method is presented in Figure 3.

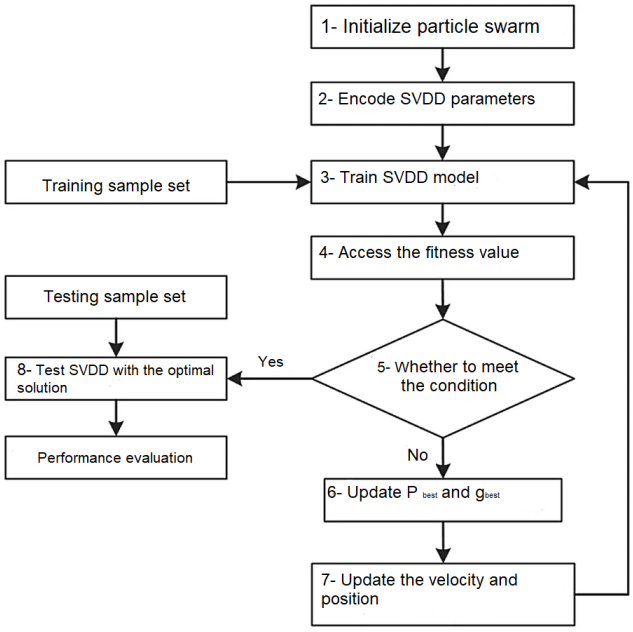

Figure 3: Flow Chart of the Proposed Model.

IV. Experimental results

A. Datasets

The performance of the proposal method was examined over three datasets. First, the dataset (DB1) is downloaded from [51], which contains fraudulent transactions generated by the European credit cardholders in September 2013. Secondly, the dataset (DB2), Synthetic Financial Datasets, is downloaded from [52]. It generated by the PaySim mobile

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

money simulator that is based on a sample of real transactions extracted from financial logs. Lastly, the dataset (DB3) is downloaded from [53], which includes ten million transactions characterized by seven attributes. A brief description of the datasets is presented in Table I.

Table I: Datasets description

Dataset	Attributes	Positive Target	Negative Target	Ratio of fraudulent cases
DB1	30	284315	492	0.17%
DB2	9	6354407	8213	0.12%
DB3	7	9403986	596014	0.15%

A preprocess step, cleaning and removing observations, is applied to training and testing data in the first place. Secondly, the datasets are normalized, which is a step that changes values to a standard scale without distorting the difference between the range of values. And to be precise, the training data is randomly selected from the whole dataset, while the remaining part is used for testing. Indeed, the three datasets are highly imbalanced, that means, the percentage of normal transaction is higher than the fraudulent ones [38]. In other words, the classification process tends to predict that most of the incoming data belongs to the majority class. Hence, it is important to engage a dataset resampling process [38],[39]. It consists of removing samples from the majority class and or adding more examples from the minority class to achieve better classifier performance. In our case, the proposed model used undersampling technique for the resampling process [54], [55], [56]. The Figure 4 and 5 explain the resampling process.

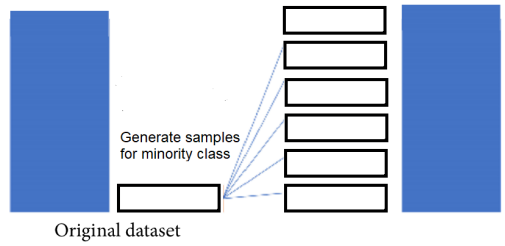

Figure 4: Oversampling technique

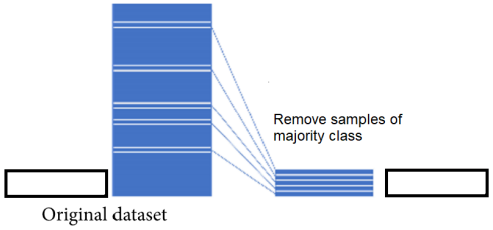

Figure 5: Undersampling technique

B. Experiments

All the experiments were evaluated using the Confusion Matrix (CM). There are four values obtained in a CM: True Positives, True Negatives, False Positive and False Negative through which the result is studied. The following tables explain the CM in details. Where TP is the number of fraud samples, TN is the number of non-fraud samples, FP is the number of non-fraud samples misrepresented as fraud samples and FN is the number of fraud sample misrepresented as non-fraud samples. To perform the analysis, we split each dataset into training (80%)and testing (20%) set. An 80% training set is employed to train the machine learning model, and the corresponding 20% testing set is employed for performance testing of the model. To evaluate the running model and the corresponding model performance in the testing set, we will use accuracy to determine the model's effectiveness. The details description of the datasets used in the experiment are presented in Table IV.

Table II: Confusion Matrix

	Predicted Negative	Predicted Positive
Actuel Negative	TN: true negative	FP: false positive
Actuel Positive	FN:false negative	TP: true positive

Table III: Classification Performance Measures

Measure	Definition
Recall	TP/(TP + FN)
Precision	TP/(TP + FP)
Accuracy	(TP+TN)/(TP+TN+FP+FN)
F1 Score	2*(Precision-Recall)/(Precision +Recall)

Table IV: The details of the datasets used in the experiments

Dataset	Attributes	Positive Target	Negative Target
DB1	30	284315	492
DB2	9	635441	411
DB3	7	188080	328

C. Results

The performance accuracy was calculated for SVM and PSO-SVDD with different kernels. The results show that the PSO-SVDD method obtained higher performance accuracy than SVM as shown in Table V.

Table V: Accuracy performance

Dataset	SVM	SVDD RBF	SVDD POLY	SVDD LINEAR
DB1	90%	94%	67%	79%
DB2	77%	82%	50%	50%
DB3	89%	97%	66%	70%

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

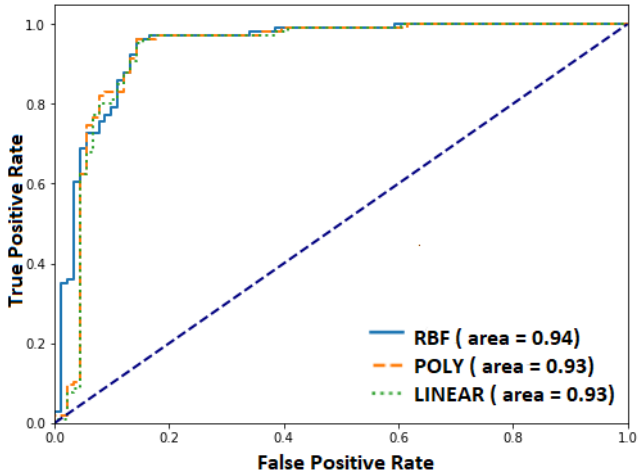

Figure 6: ROC Curve Analysis of SVDD on various kernel functions for DB1

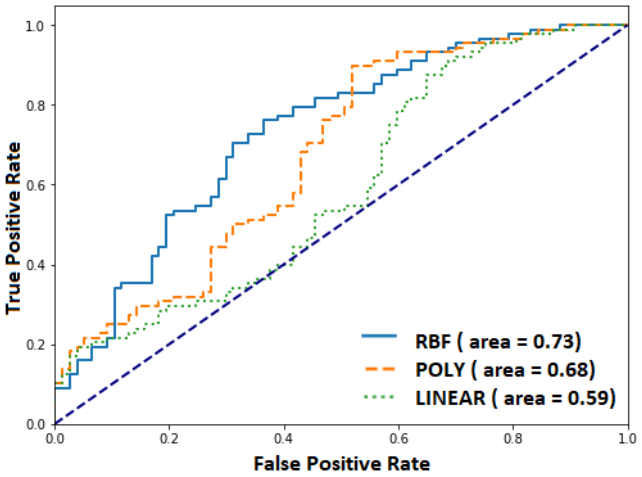

Figure 7: ROC Curve Analysis of SVDD on various kernel functions for DB2

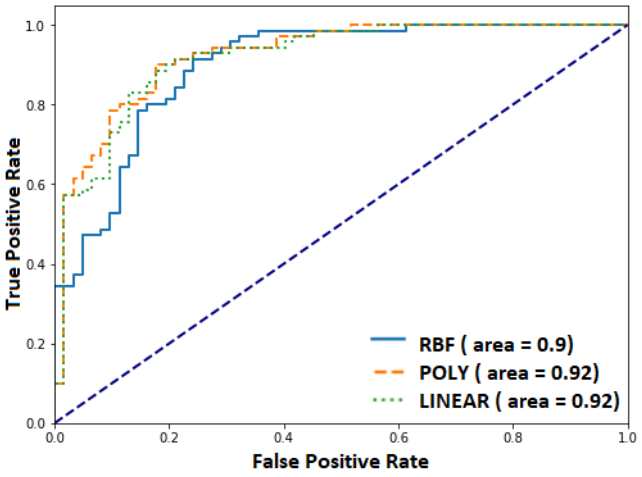

Figure 8: ROC Curve Analysis of SVDD on various kernel functions for DB3

V. Conclusion

In this paper, the present research investigated a new method of optimizing credit card fraud detection based on combining the PSO and the SVDD. The SVDD-PSO's performance was validated in terms of accuracy and learning speed and compared with the traditional SVM. The results show that the proposed solution achieved better accuracy with acceptable performance. It outperformed SVM in terms of accuracy. In future, we want to extend the work further with the integration of features selection methods in order to refine the proposed method.

References

[1] Ashby, M. (2020a). Initial evidence on the relationship between corona virus pandemic and crime in the United States. Crime Science, 2020.
[2] https://www.cnbc.com/2021/01/27/credit-card-fraud-is-on-the-rise-due-to-covid-pandemic.html, 2222.
[3] KOCHHAR, Heena. Analysis of Various Credit Card Fraud Detection Techniques. Items Page Number, 2021, vol. 5, no 2.
[4] SADGALI, Imane, SAEL, Nawal, et BENABBOU, Faouzia. Performance of machine learning techniques in the detection of financial frauds. Procedia computer science, 2019, vol. 148, p. 45-54.
[5] ALFAIZ, Noor Saleh et FATI, Suliman Mohamed. Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics, 2022, vol. 11, no 4, p. 662.
[6] SHARMA, Pratyush, BANERJEE, Souradeep, TIWARI, Devyanshi, et al. Machine Learning Model for Credit Card Fraud Detection-A Comparative Analysis. The International Arab Journal of Information Technology, 2021, vol. 18, no 6.
[7] Handa, Akansha, Yash Dhawan, and Prabhat Semwal. "Hybrid analysis on credit card fraud detection using machine learning techniques." Handbook of Big Data Analytics and Forensics. Springer, Cham, 2022. 223-238.
[8] Arafath, Yeasin, et al. "Developing a Framework for Credit Card Fraud Detection." Proceedings of the International Conference on Big Data, IoT, and Machine Learning. Springer, Singapore, 2022.
[9] Campus, K., 2018. Credit card fraud detection using machine learning models and collating machine learning models. International Journal of Pure and Applied Mathematics, 118(20), pp.825-838, 2018.
[10] Trivedi, N.K., Simaiya, S., Lilhore, U.K. and Sharma, S.K. An efficient credit card fraud detection model based on machine learning methods. International Journal of Advanced Science and Technology, 29(5), pp.3414-3424, 2020.
[11] MORE, Rashmi, AWATI, Chetan, SHIRGAVE, Suresh, et al. Credit Card Fraud Detection Using Supervised Learning Approach. International Journal of Scientific and Technology Research, 2021, vol. 9, p. 216-219.
[12] Dornadula, V.N. and Geetha, S. Credit card fraud detection using machine learning algorithms. Procedia Computer Science, 165, pp.631-641, 2019.
[13] Khatri, S., Arora, A. and Agrawal, A.P. January. Supervised machine learning algorithms for credit card fraud detection: a comparison. In 2020 10th International Conference on Cloud Computing, Data Science and Engineering (Confluence) (pp. 680-683.IEEE, 2020.
[14] BAINS, Kameron, FASANMADE, Adebamigbe, MORDEN, Jarrad, et al. Defeating the Credit Card Scams Through Machine Learning Algorithms. In : 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). IEEE, 2021. p. 193-198.
[15] TYAGI, Rishabh, RANJAN, Ravi, et PRIYA, S. Credit Card Fraud Detection Using Machine Learning Algorithms. In : 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2021. p. 334-341.
[16] LI, Chenglong, DING, Ning, ZHAI, Yiming, et al. Comparative study on credit card fraud detection based on different support vector machines. Intelligent Data Analysis, 2021, vol. 25, no 1, p. 105-119.
[17] ERFANI, Masoud, SHOELEH, Farzaneh, et GHORBANI, Ali A. Financial Fraud Detection using Deep Support Vector Data Description. In : 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020. p. 2274-2282.
[18] GAMINI, Prathima, YERRAMSETTI, Sai Tejasri, DARAPU, Gayathri Devi, et al. A Review on the Performance Analysis of Supervised and Unsupervised algorithms in Credit Card Fraud Detection. International Journal of Research in Engineering, Science and Management, 2021, vol. 4, no 8, p. 23-26.
[19] JOSHI, Abhishek, SONI, Shreedhar, et JAIN, Vaibhav. An Experimental Study using Unsupervised Machine Learning Techniques for Credit Card Fraud Detection.
[20] MANDAL, Amit Kumar et DINDA, Prosenjit. A Survey on Unsupervised Machine Learning Approach for Fraud Detection in Bitcoin. AIJR Abstracts, 2022, p. 33.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[21] BHASKAR, Aman, PANDE, Sagar, MALIK, Rahul, et al. An intelligent unsupervised technique for fraud detection in health care systems. Intelligent Decision Technologies, 2021, vol. 15, no 1, p. 127-139.
[22] BARRICKLOW, Austin. Unsupervised Machine Learning to Create Rule-Based Wire Fraud Detection. 2021. Thèse de doctorat. Utica College.
[23] PRUSTI, Debachudamani, KUMAR, Abhishek, PURUSOTTAM, Ingole Shubham, et al. A design methodology for web-based services to detect fraudulent transactions in credit card. In : 14th Innovations in Software Engineering Conference (formerly known as India Software Engineering Conference). 2021. p. 1-9.
[24] 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.
[25] 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.
[26] Jakkula, V. Tutorial on Support Vector Machine (SVM). Available online:http://www.ccs.neu.edu/course/cs5100f11/resources/jakkula.pdf, 2013.
[27] Tax, D.M., Duin, R.P.: Support vector data description. Machine learning 54(2004) 45–66.
[28] Chen, S.;Wang, J.-q.; Zhang, H.-y. A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. Technol. Soc. Chang. 2019, 146, 41–54. S. Kiranyaz, T. Ince, A. Yildirim, and M. Gabbouj, Evolutionary artificial neural networks by multi-
[29] dimensional particle swarm optimization, Neural Netw., vol. 22, no. 10, pp. 1448–1462, 2009.
[30] A. Dal Pozzolo, O. Caelen, and G. Bontempi, When is undersampling effective in unbalanced classification tasks? in Machine Learning and Knowledge Discovery in Databases. Cambridge, U.K.: Springer, 2015.
[31] N. Mahmoudi and E. Duman, Detecting credit card fraud by modified fisher discriminant analysis, Expert Syst. Appl., vol. 42, no. 5,pp. 2510–2516, 2015.
[32] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, 2018, pp. 1-6.
[33] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decis. Support Syst., vol. 50, no. 3, pp. 602-613, 2011.
[34] K. Chaudhary, J. Yadav, and B. Mallick, "A review of Fraud Detection Techniques: Credit Card," Int. J. Comput. Appl., vol. 45, no. 1, pp. 975-8887, 2012.
[35] F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System," vol. 6, no. 3, pp. 311-322, 2011.
[36] J. Awoyemi, A. Adetunmbi and S. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis", 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017.
[37] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357, 2002.
[38] Mohammed, Roweida, Jumanah Rawashdeh, and Malak Abdullah. "Machine learning with oversampling and undersampling techniques: overview study and experimental results." 2020 11th international conference on information and communication systems (ICICS). IEEE, 2020.
[39] ANIRUDH, G. et TALUKDAR, Upasana. An Experimental Analysis for Credit Card Fraud Detection with Imbalanced and Machine Learning Techniques. In : Edge Analytics. Springer, Singapore, 2022. p. 539-550.
[40] DU, Hongle et ZHANG, Yan. Network anomaly detection based on selective ensemble algorithm. The Journal of Supercomputing, 2021, vol. 77, no 3, p. 2875-2896.
[41] Biswas, Biswajit, Swarup Kr Ghosh, and Anupam Ghosh. "Automatic Image Segmentation by Ranking Based SVM in Convolutional Neural Network on Diabetic Fundus Image." Deep Learning in Data Analytics. Springer, Cham, 2022. 77-95.
[42] Fu, Weiqiong, Hanxiao Zhang, and Fu Huang. "Internet-based supply chain financing-oriented risk assessment using BP neural network and SVM." Plos one 17.1 (2022): e0262222.
[43] Hofmann, Lena A., Steffen Lau, and Johannes Kirchebner. "Advantages of Machine Learning in Forensic Psychiatric Research—Uncovering the Complexities of Aggressive Behavior in Schizophrenia." Applied Sciences 12.2 (2022): 819.
[44] Pradhan, Arabinda, Sukant Kishoro Bisoy, and Amardeep Das. "A survey on PSO based meta-heuristic scheduling mechanism in cloud computing environment." Journal of King Saud University-Computer and Information Sciences (2021).
[45] Yarat, Serhat, Sibel Senan, and Zeynep Orman. "A Comparative Study on PSO with Other Metaheuristic Methods." Applying Particle Swarm Optimization. Springer, Cham, 2021. 49-72.
[46] Sai Kiran, Jypti Guru, Rishabh Kumar, Naveen Kumar, Deepak Katariya,Credit card fraud detection using Naïve Bayes model based and KNN classifier, Int. Journal of Adv. Research , Ideas and Innovations in Technology,vol.4, 2018.
[47] Lakshmi, S.V.S.S. and Kavilla, S.D. Machine learning for credit card fraud detection system. International Journal of Applied Engineering Research, 13(24 Pt. 1), pp.16819-16824, 2018.
[48] Kumar, Sheo, et al. "Credit Card Fraud Detection Using Support Vector Machine." Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. Springer, Singapore, 2022.
[49] Alam, Md, et al. "Effective machine learning approaches for credit card fraud detection." International Conference on Innovations in Bio-Inspired Computing and Applications. Springer, Cham, 2020.
[50] Bains, Kameron, et al. "Defeating the Credit Card Scams Through Machine Learning Algorithms." 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). IEEE, 2021.
[51] SaiKumar, 2018, Credit Card Fraud Detection Dataset, Kaggle.
[52] Synthetic Financial Datasets For Fraud Detection, https://www.kaggle.com/kartik2112/fraud-detection-on-paysim-dataset.
[53] "Index of /datasets/." https://ackages.revolutionanalytics.com/datasets.
[54] HASANIN, Tawfiq et KHOSHGOFTAAR, Taghi. The effects of random undersampling with simulated class imbalance for big data. In : 2018 IEEE International Conference on Information Reuse and Integration (IRI). IEEE, 2018. p. 70-79.
[55] FDEZ-GLEZ, Jorge, RUANO-ORDÁS, David, FDEZ-RIVEROLA, Florentino, et al. Analyzing the impact of unbalanced data on web spam classification. In : Distributed Computing and Artificial Intelligence, 12th International Conference. Springer, Cham, 2015. p. 243-250.
[56] HANAFY, MOHAMED et MING, RUIXING. USING MACHINE LEARNING MODELS TO COMPARE VARIOUS RESAMPLING METHODS IN PREDICTING INSURANCE FRAUD. Journal of Theoretical and Applied Information Technology, 2021, vol. 99, no 12.
[57] KUBUS, Mariusz. Evaluation of resampling methods in the class unbalance problem. Ekonometria, 2020, vol. 24, no 1, p. 39-50.
[58] Sengupta, Saptarshi and Basak, Sanchita and Peters, Richard Alan "Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives " Machine Learning and Knowledge Extraction, vol 1, pp: 157–191, 2018.
[59] MISHRA, Anwesha. Fraud Detection: A Study of AdaBoost Classifier and K-Means Clustering. Available at SSRN 3789879, 2021.

ISBN: 978-988-14049-3-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)