A comparison of machine learning algorithms for credit card fraud detection

Rabab Cherkaoui*
Computer Science Département, Faculty of Sciences and Technologies of Tangier, Research Team DSAI2S, Abdelmalek Essaadi University of Tetouan, Morocco
rabab.cherkaoui@etu.uae.ac.ma

El Mokhtar En-Naimi
Computer Science Département, Faculty of Sciences and Technologies of Tangier, Research Team DSAI2S, Abdelmalek Essaadi University of Tetouan, Morocco
en-naimi@uae.ac.ma

ABSTRACT

With the increasing use of credit cards for online and offline transactions, the risk of fraudulent activities has also increased significantly. In this study, we propose a machine learning-based approach to predict credit card fraud. We used a public dataset with 284,807 transactions, of which 492 were fraudulent. We experimented with various machine learning algorithms such as k-nearest neighbor, random forests, and isolation forests to develop predictive models for credit card fraud. We also performed feature selection to identify the most important features that contribute to credit card fraud prediction. Our results suggest that the proposed machine learning approach can effectively detect fraudulent transactions and can be adopted by banks and financial institutions to reduce the risk of credit card fraud.

KEYWORDS

Machine learning, credit card, fraud detection, supervised learning, unsupervised learning

ACM Reference Format:
Rabab Cherkaoui and El Mokhtar En-Naimi. 2023. A comparison of machine learning algorithms for credit card fraud detection. In *The 6th International Conference on Networking, Intelligent Systems & Security (NISS 2023), May 24–26, 2023, Larache, Morocco.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3607720.3607759

1 INTRODUCTION

Machine Credit card fraud is a one of the biggest concerns for financial organizations, merchants and cardholders worldwide. Fraudulent transactions can result in significant financial losses and reputational damage to businesses and individuals. Fraudulent transactions can take many forms, including card data hacking, unauthorized transactions, counterfeit cards and identity theft.

The challenge of credit card fraud detection is to identify fraudulent transactions quickly and with high accuracy. Traditional approaches to detecting fraudulent transactions include rule-based systems that use predefined criteria to flag suspicious transactions. Unfortunately, such systems are limited in their ability to detect complex fraud patterns and can generate false positives, resulting in unnecessary investigations and operational costs.

With advances in machine learning and artificial intelligence, there has been renewed interest in using these techniques to improve the accuracy of fraud detection. Machine learning algorithms can identify patterns and anomalies in transaction data and learn from past fraud incidents to detect new ones. These algorithms can help financial institutions and merchants detect and prevent fraudulent transactions, enhancing user satisfaction and increasing profitability.

The problem of credit card fraud detection is an ongoing challenge, as hackers are constantly adapting and refining their techniques. As a result, businesses and merchants must continually update and refine their fraud detection systems to stay ahead of fraudulent activity.

The objective of this work is to determine which machine learning algorithm performs best among supervised and unsupervised learning algorithms for credit card fraud detection. To do this, we will use a dataset containing a large number of transactions, a small percentage of which are fraudulent. The goal is to train a machine learning model to identify transactions that are likely to be fraudulent based on the transaction details.

The dataset typically includes information such as transaction amount, transaction time, merchant category code, and other relevant information. Each transaction is classified as fraudulent or not, depending on whether it has been declared fraudulent by the credit card issuer.

2 RELATED STUDIES

Data scientists in credit card fraud detection have applied quite a few advanced algorithms and machine learning techniques. These techniques have yielded interesting results in improving the accuracy of fraud detection and limiting the number of false positives.

Anomaly detection techniques such as isolation forests, SVM and local outlier factors have also been applied to these types of problems. These techniques identify outliers or anomalies in the transaction data that do not conform to the typical behavior of legitimate transactions.

Deep learning approaches, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in detecting fraudulent transactions. CNNs are well suited for image-based fraud detection, such as signature verification, while RNNs are effective for sequence-based fraud detection, such as transaction sequences.

*Corresponding Author.

NISS 2023, May 24–26, 2023, Larache, Morocco

ACM ISBN 979-8-4007-0019-4/23/05…$15.00
https://doi.org/10.1145/3607720.3607759

Overall, the state of the art in credit card fraud detection involves using a combination of advanced machine learning techniques and approaches that address the class imbalance problem. However, fraudsters are constantly evolving their techniques and research in this area continues to stay ahead of emerging threats.

3 METHODOLOGY

For the comparison of supervised and unsupervised machine learning algorithms for predicting credit card frauds, we followed this methodology:

Dataset collection: collect the data that will be used for this study.

Dataset preparation: prepare the dataset that will be used to train and evaluate the machine learning models.

Supervised learning: Train and evaluate various supervised and unsupervised machine learning algorithms, such as random forest, K-nearest neighbors (KNN) and support vector machines (SVM).

Unsupervised learning: Train and evaluate various unsupervised machine learning algorithms, such as Principal Component Analysis (PCA), Local Outlier Factor (LOF) and Isolation Forest.

To do this work we used the approach:

Train-test split: in this approach, the dataset is split into two parts: the training set and the testing set. The model is trained on the training set, then evaluated on the testing set.

Model evaluation: Evaluate each model's performance by using appropriate metrics such as accuracy, precision, recall, and F1 score.

Comparison and analysis: Compare the performance of supervised and unsupervised machine learning algorithms based on their ability to detect fraudulent transactions accurately. Consider factors such as precision, recall, F1 score, and computational efficiency.

3.1 Dataset collection

The data that we used for the credit card fraud detection is collected from Kaggle. It contains the columns:

Time: The time elapsed between transactions.

Amount: The transaction amount.

V1, V2, V3, ... V28: These columns represent anonymous characteristics that capture different characteristics of the transaction, such as position and the transaction type.

Class: This column indicates whether the transaction is fraudulent or not. This is the target variable that the machine learning model is designed to predict.

3.2 Dataset preparation

First, we removed the "Time" column because it is not relevant for our analysis. Next, we removed the duplicate entries and check the class distribution to see if the dataset is balanced or not. Since the data set is unbalanced with more non-fraudulent transactions, we balance the data set by randomly sampling non-fraudulent transactions to correspond to the number of fraudulent transactions. Finally, we save the cleansed dataset in a CSV format for further evaluation.

3.3 Supervised learning

In the world of machine learning and artificial intelligence, supervised learning refers to a class of systems and algorithms that determine a predictive model using data points whose results are known. The name "supervised" learning derives from the idea that training this type of algorithm is like having a teacher supervise the entire process. The aim is to improve the algorithms of the targeted class by using the prediction data and the optimization routine to minimize a loss or the error function.

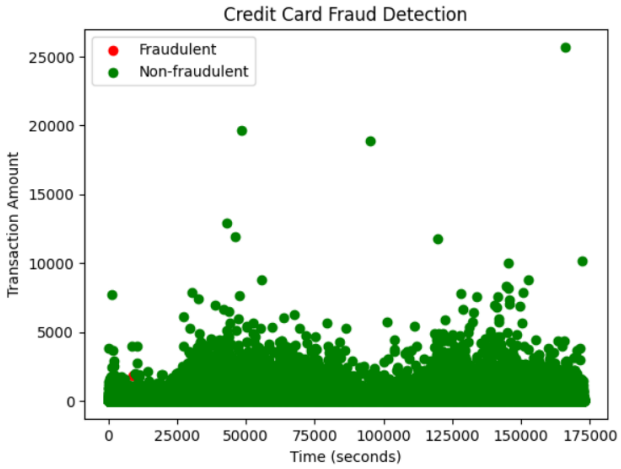

Figure 1: Dataset representation

Supervised learning can be categorized into two main types of problems:

Regression: the goal of regression is to learn a function that can accurately predict the output value based on the input data. Examples of regression problems include predicting the price of a house based on its characteristics or predicting the temperature based on the weather.

Classification: In classification, the output variable is a discrete value or category. The goal of classification is to learn a function that can accurately predict the category or label of a new input data point based on its characteristics. Examples of classification problems include predicting that an email is spam or not, or predicting that a credit card transaction is fraudulent or not.

3.3.1 Random Forest. The Random Forest algorithm is a commonly used machine learning algorithm in the category of supervised learning, it's used extensively in both classification and regression. It is widely used for classification and regression studies. It builds decision trees on different samples and then takes their majority score for classification and the average in the case of regression. One of the most important features of the random forest algorithm is that it can handle data sets containing both continuous variables, as in the case of regression, and discrete variables, as in the case of classification. It is very fast and provides a description and a presentation of the results without the need for parameterization. It also gives better results for both classification and regression tasks.

Hyperparameter:

Number of trees in the forest.

Maximum depth of each tree.

Number of features to consider when splitting each node.

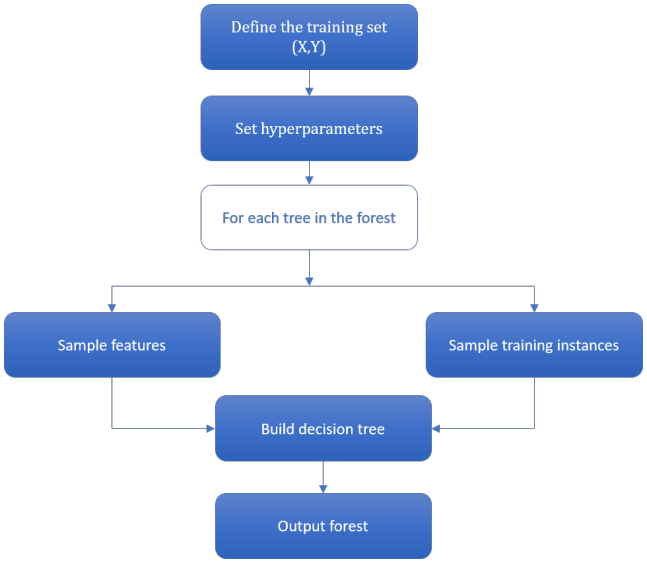

Figure 2: Schema of RF algorithm

3.3.2 K-nearest neighbors. KNN or k-NN (k-nearest neighbors) is a non-parametric supervised learning discriminant that uses proximity to make classifications or predictions on the clustering of a single data point. Although it can be used for regression or classification problems, it is generally used as a classification algorithm, based on the assumption that similar points will be found next to each other. It is ideal for multi-modal interfaces, as well as for situations where items may have multiple identifiers.

To make a prediction, the K-NN algorithm does not compute a predictive model from a training set, as is the case with linear regression. In fact, K-NN does not need to build a predictive model. Therefore, there is no learning phase as such for K-NN. To make a prediction, K-NN relies on the data set to produce a result.

3.3.3 Support Vector Machine. The Support Vector Machine, or SVM, is a well-known supervised learning algorithm. It is a difficult process, but with high accuracy. It is also used for classification and regression problems.

The goal of the SVM algorithm is to efficiently construct the best decision line or boundary that can divide the n-dimensional space into classes so that we can easily classify the new data point into the correct category in the future. This best decision line is defined as a hyperplane.

The SVM selects the extreme points/vectors that help create the hyperplane. These extremes are called support vectors, and the algorithm is therefore called a support vector machine.
(X, Y): X is a matrix of features and Y is a vector of labels.
w: the weight vector.
b: the bias.

3.4 Unsupervised learning

Unsupervised learning is the application of artificial intelligence algorithms to recognize patterns in data sets containing data points

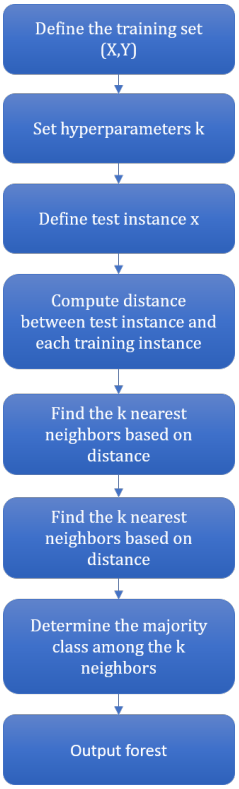

Figure 3: Schema of KNN algorithm

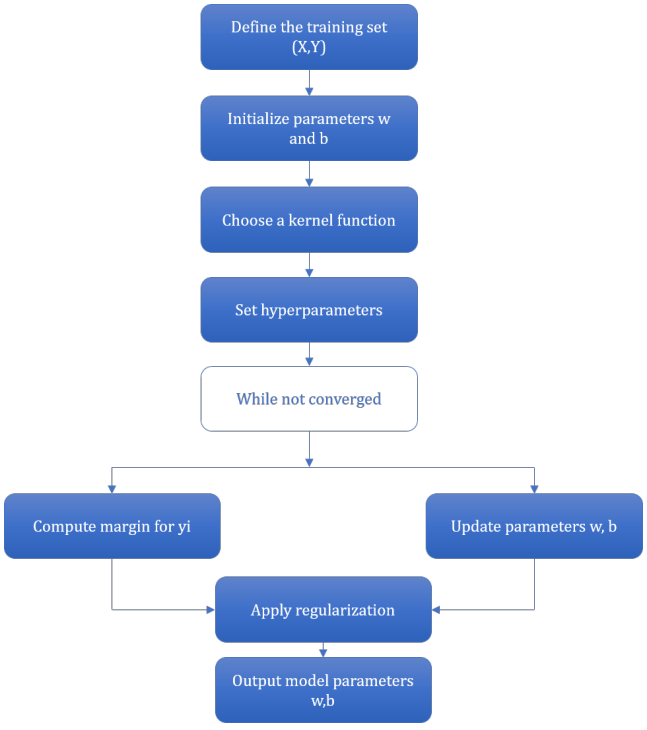

Figure 4: Schema of SVM algorithm

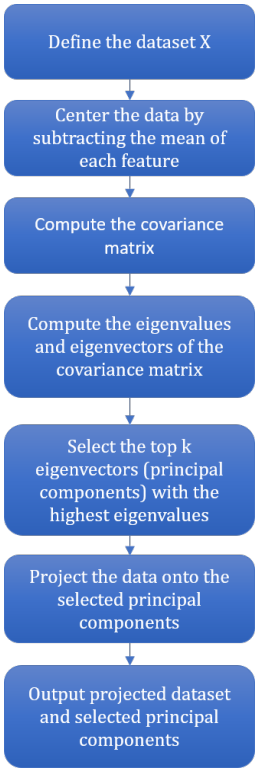

Figure 5: Schema of PCA algorithm

that have not been classified or labelled. The algorithms are therefore allowed to classify, label and/or cluster the data points contained in the datasets without any external assistance to perform this task. In other words, unsupervised learning allows the system to identify patterns in the data sets itself. In unsupervised learning, an artificial intelligence system combines unorganized information based on similarities and differences, even if no categories are provided. Unsupervised learning algorithms can handle more challenging data processing than supervised learning systems.

3.4.1 Principal Component Analysis. Principal Component Analysis (PCA) is a popular unsupervised learning technique for reducing the dimensionality of data. It increases interpretability but at the same time minimizes information loss. It helps to find the most significant features in a data set and facilitates plotting of data in 2D and 3D.

3.4.2 Local Outlier Factor. The Local Outlier Factor (LOF) is an algorithm that calculates the local density difference between a given data point and its neighbors.

What are the advantages of the LOF algorithm in anomaly detection?

One of the main difficulties in anomaly detection is the lack of training data providing examples of anomalies to train the models to predict. The LOF algorithm is an unsupervised technique that does not require prior examples.

Therefore, if the data has different densities in different regions, many simple distance-based anomaly detection methods fail, i.e.

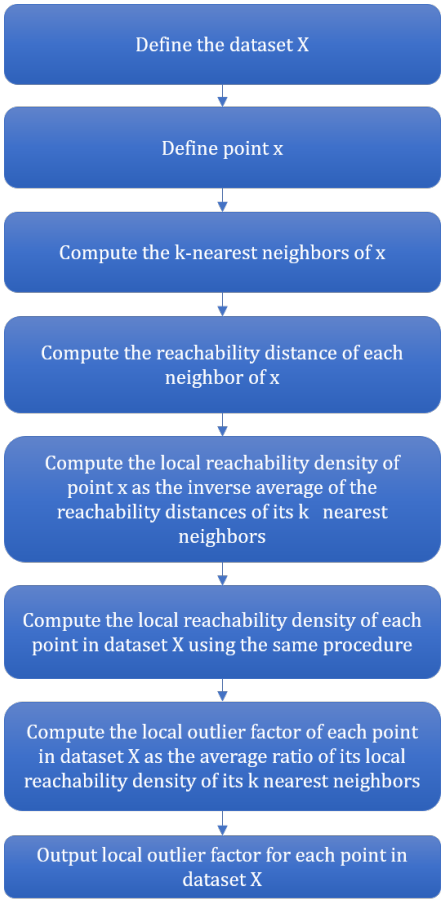

Figure 6: Schema of LOF algorithm

there are areas where the points are close together and areas where the points are scattered in the data. The LOF method can deal with these situations.

3.4.3 Isolation forest. The isolation forest is an unsupervised machine learning algorithm that detects anomalies by isolating outliers in data sets.

The algorithm generates an anomaly score for each data point in the dataset, which is a measure of the degree of atypicality of the data point. Isolation forests were designed on the basis that anomalies are "few and distinct" data points in a data set.

3.5 Model evaluation

3.5.1 Supervised learning.

3.5.2 Unsupervised learning. We defined a range of values for n_neighbors (the number of principal neighbors to keep) to test, and iterated through each value of n_neighbors. For each value of n_neighbors, we fit the model to the data, transform the data to lower dimensional space using the fit_transform method, calculate the Silhouette score using the silhouette_score function from scikit-learn, and print the Silhouette score for that value of n_neighbors.

Table 1: Evaluation results of supervised algorithms

	RF	KNN	SVM
Accuracy	94,36%	95,07%	95,07%
Precision	95%	93%	94%

Table 2: Evaluation results of unsupervised algorithms using precision, recall and F1-score

	PCA	LOF	IF
Precision	93%	69%	96%
Recall	13%	9%	23%
F1-score	23%	15%	37%

Table 3: Evaluation results of unsupervised algorithms using silhouette score

	PCA	LOF	IF
Silhouette score	3%	52%	75%
Neighbors range	10	10	10
Neighbor number	5	9	4

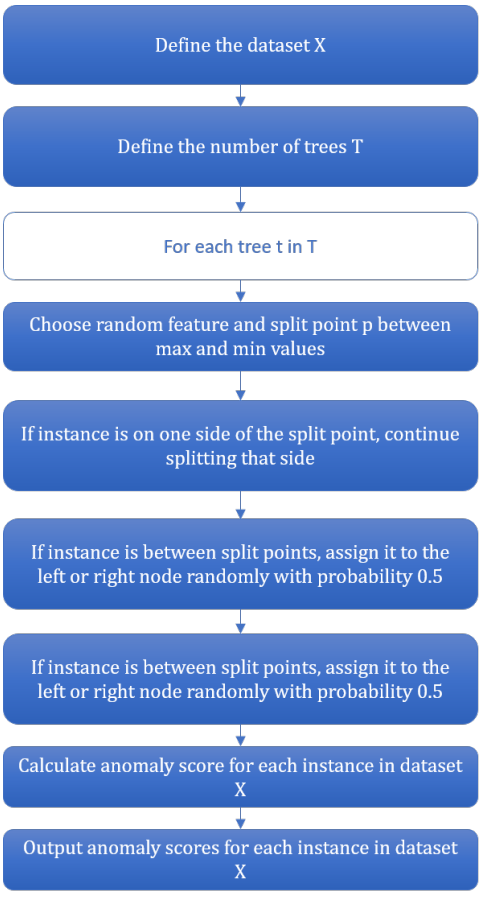

Figure 7: Schema of IF algorithm

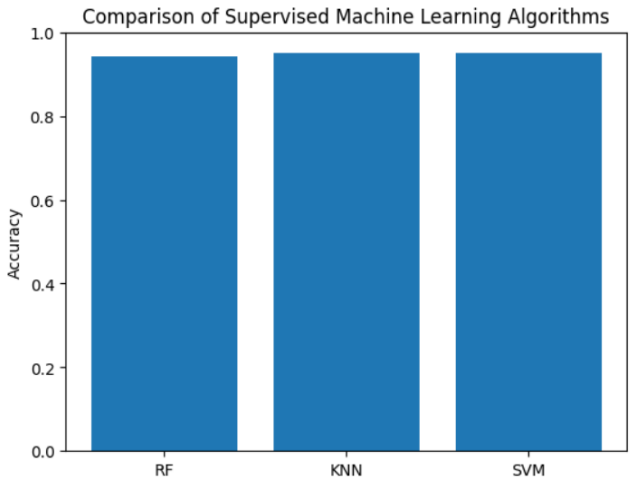

Figure 8: Supervised algorithms accuracy comparison results

We also kept track of the best silhouette score and the corresponding n_neighbours value.

4 DISCUSSION

The results obtained reveal that both the KNN and SVM algorithms achieve the same accuracy in the case of supervised learning, which was 95.07% with a very narrow precision difference, so we can say that the SVM algorithm is more efficient since it has the high precision value than the others, and this is due to the fact that it is characterized by:

- Limited overfitting: The SVM algorithm has a low tendency to overfit the data, which means that it can generalize well to new data. This is important in fraud detection because fraud patterns

are constantly evolving and the algorithm needs to be able to detect new fraud patterns.

- Robustness to outliers: Fraud detection data often contains outliers, which are data points that are significantly different from the rest of the data. The SVM algorithm is robust to outliers and its decision boundary is not significantly affected by outliers.

On the other hand, in the case of unsupervised learning, the Isolation Forest algorithm performs better than the LOF and PCA algorithms, with a precision of 96%, a recall of 23% and an F1 score of 37%, and a best silhouette score of 75%, indicating that the data points are well clustered, unlike the PCA and LOF algorithms, having a low score, which indicates that the clustering algorithm may have difficulty distinguishing between clusters or identifying outliers.

The high precision of the IF algorithm is due to the fact of:

- Robustness to noise: The IF algorithm is robust to noise, which is common in real-world data. Noise can be in the form of irrelevant features or corrupted data points. The algorithm can effectively separate the relevant features from the noise and identify fraudulent transactions.

Fast and scalable: The IF algorithm is computationally efficient and can handle large datasets with a high number of features. This is important in fraud detection where data can be large and complex.

5 CONCLUSION AND PERSPECTIVES

There have been many researches on credit card fraud detection using different machine learning techniques. The results of these studies vary depending on the engineering and feature selection methods, the machine learning algorithms and hyperparameters, and the evaluation methods used.

In conclusion, both supervised and unsupervised algorithms have provided relevant results in our study. In practice, a combination of both approaches may be employed to achieve the most pleasing results.

In fact, this work can be further improved by applying reinforcement learning algorithms, which can solve one of the main challenges, namely the imbalance between the number of fraudulent and non-fraudulent transactions. In addition, the investigation of the most appropriate method for processing a large number of transactions and the application of deep learning techniques could be the main focus of upcoming studies.

ACKNOWLEDGMENTS

We would like to thank Mr. El Mokhtar EN-NAIMI for his valuable guidance and support throughout this research project, which has significantly improved the paper.

We also thank the anonymous reviewers for their constructive feedback, which helped us to improve the quality of this article.

Finally, we would like to thank our families and loved ones for their unwavering support and encouragement throughout this project.

REFERENCES

[1] Tiwari. S. 2022. Supervised Machine Learning: A Brief Introduction. Conference: International Conference on Virtual Learning - VIRTUAL LEARNING - VIRTUAL REALITY (17th edition), 3-7. https://doi.org/10.58503/icvl-v17y202218

[2] Qu'est ce que l'algorithme KNN. Available at: KNN : Découvrez cet algorithme de Machine Learning (datascientest.com)

[3] Benzaki. Y. 2018. Introduction to the K Nearst Neighbors (K-NN) algorithm. Available at: https://mrmint.fr/introduction-k-nearest-neighbors

[4] Alizadeh. E. 2022. What K is in KNN and K-Means Get to know K-Nearest Neighbors and K-Means. Available at: https://ealizadeh.com/blog/knn-and-kmeans/#:$\sim$:text$=$KNN%20is%20a%20supervised%20learning%20algorithm%20mainly%20used%20for%20classification,on%20the%20chosen%20distance%20metric

[5] Brownlee. J. 2019. A Gentle Introduction to Expectation-Maximization (EM Algorithm). Available at: https://machinelearningmastery.com/expectation-maximization-em-algorithm

[6] Banoula. M. 2023. What Is Q-Learning: The Best Guide To Understand Q-Learning. Available at: https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-q-learning

[7] K.Pratt. M.2020. Unsupervised learning. Available at: https://www.techtarget.com/searchenterpriseai/definition/unsupervised-learning

[8] Biswal. A. 2023. Principal Component Analysis in Machine Learning: Complete Guide. Available at: https://www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis

[9] MLNerds. 2023. Local Outlier Factor for Anomaly Detection. Available at: https://machinelearninginterview.com/topics/machine-learning/local-outlier-factor-lof/

[10] alexsoft. 2022. Machine Learning Metrics: How to Measure the Performance of a Machine Learning Model. Available at: https://www.altexsoft.com/blog/machine-learning-metrics/

[11] Credit Card Fraud Detection. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud?resource$=$download