2026 Volume: 02 Issue: 03 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-18

COMPARATIVE ANALYSIS OF SUPPORT VECTOR MACHINE AND K-NEAREST NEIGHBOR ALGORITHMS ON FRAUD DETECTION OF PAYMENT CARD SYSTEMS

Okorie Kingsley Maduabuchi*[1] Ozioko Frank Ekene[2] Oko Gabriel Ota[3]

Computer Science Department, Enugu State University of Science and Technology.

Article Received: 29 January 2026
Article Revised: 18 February 2026
Published on: 10 March 2026

***Corresponding Author: Okorie Kingsley Maduabuchi**
Computer Science Department, Enugu State University of Science and Technology.

DOI: https://doi-doi.org/101555/ijrpa.8338

ABSTRACT

Credit card fraud refers to the physical loss of credit card or loss of sensitive card's information; the fraud could also in most cases be carried out virtually as when some essential information of the cards is made available to the adversaries. In order to detect the fraud, machine learning algorithms have been promising to that effect. However, different algorithms portray different accuracies. It is on that note that this research aims to carry out comparative analysis of SVM and KNN against NB, MLP, LR and RF on the same given datasets obtained from Kaggle. The dataset was highly imbalanced, so the SMOTE was used for oversampling. Similarly, feature selection was performed and dataset was split into two parts, training data and test data in the ratio of 70:30 percent respectively. The results of the analysis show the strength of various algorithms in terms of the three metrics considered in this. For instance, while on one hand, RF has 96.30% and 99.96% for precision and accuracy respectively; on the other hand, SVM has 1.89%, 71.43% and 93.60% for precision, recall and accuracy respectively. Therefore, this study has compared two algorithms included the hybrid against the existing four algorithms from the same dataset. It is therefore recommended that interested researchers can employ a robust deep learning algorithm with preprocessing technique for better performance.

INTRODUCTION

Overview. According to the Global Facts (2019), there are growing numbers of new companies all around the world. All of those companies are trying to provide best service quality for their customers. In order to succeed in that, companies are processing a lot of data on a daily basis. These data comes from vast number of resources and are in different

formats. Moreover, this data contains some of the key parts of the company's future business. Because of that, companies need to store that data, to process it and what is really important, to keep it safe. Without securing data, a lot of it can be used by other companies or even worse, it can be stolen. In most cases, financial information is stolen, which can harm whole company or individual.

There are several types of frauds according to the Legal Dictionary (2019). Check Fraud occurs when person forges a check or pays for something with check knowing that there is not enough money. Internet sales are fraud where fraudster sale fake items or counterfeit items, or taking payment without delivering the item. There are a couple more, such as charities fraud, identity theft, payment card fraud, debt elimination, Insurance fraud and others. Due to increasing popularity of cashless transactions, one of the most common frauds is payment card frauds. The European Central Bank, ECB, (2018) described payment card fraud as the situation where fraudster uses payment card for their needs while owner of that payment card is not aware of that. Fraudulent transactions conducted using payment cards acquired worldwide amounted to €1.8 billion in 2016. Although there is a tremendous volume increase in payment card transactions, the amounts of frauds is proportionally the same or have decreased due

to sophisticated fraud detection systems. However, fraudsters are constantly coming up with new ways to steal information as seen in Ignacio (2017).

There are two types of payment card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, CVV code, type of card and other. By stealing payment card information, a fraudster can broach a large amount of money or make a large amount of purchase before cardholder finds out. Because of that, companies use various machine learning methods to recognize which transactions are fraudulent and which are not.

Technological revolution has disrupted the finance industry during the last years. A lot of startups, the so-called fintechs, have brought innovation into banking. Payments have been simplified and mobile access to all our financial operations is now a reality. User friendly and more transparent financial services are being created, improving the way customers manage their finances. However, securing the increasing number of transactions can become a problem if the fintechs fail to scale up their data processes. According to Araujo, *et al* (2017) as payments increase, the number of frauds starts to be significant enough to translate into important losses for the companies. Bank transaction frauds cause over $13B annual losses,

which affect not only banks and fintechs, but also their clients. Regulations regarding fraud detection also need to be complied, so this becomes a very important task that has to be handled adequately. When the number of transactions is small, fraud detection can be worked around with some hand-crafted rules, but as the company grows, their complexity increases. Adding more rules and changing existing ones is cumbersome and validating the correctness of the new rules is usually hard. That is one of the reasons why expert systems based on rules are highly inefficient and pose scalability issues. One way to solve this is to combine these systems with automated approaches that leverage the data from previous frauds. By doing that, only basic rules which are easier to maintain need to be implemented. These automatic systems can detect difficult cases more accurately than complex rules and don't present scalability problems.

In order to implement an efficient Fraud Detection System (FDS), multi-disciplinary teams are needed. Data scientists, data engineers and domain experts are required to work together. Data engineers create data pipelines that fetch the information needed from production systems and transform it into a usable format. Then, data scientists can use that data to create fraud prediction models. Fraud investigators are also important, as they have extensive knowledge about the fraudsters' behavior, which can save a lot of time of data exploration to the data scientists. Finally, the data engineers need to put the models created by the data scientists into production. Small companies don't usually have the data pipeline processes already in place, increasing the complexity of implementing the FDS. University research s often focus on comparing different techniques and algorithms using toy datasets, so they don't have into account data level problems or deployment issues, which are usually present in real world cases. Dealing with financial data is sensible, so publicly available datasets in this study area are scarce. This slows down further advances because researchers use different datasets, which most of the times can't be made public, so comparing the results is hard.

According to Gaurav *et al* (2021), there are many algorithms that are and have been implemented for this credit card fraud detection. Mostly classification techniques are used such as K-means clustering, KNN, etc. Trees are also used for credit card fraud detection i.e., Decision Tree and Random Forest. But many a times Decision Trees led to overfitting and high computational cost, so to overcome this anomaly detection techniques are used like Isolation Forest.

In the literature of Roberto *et al* (2017), the common criterion used in almost all the state of-the-art approaches of fraud detection is substantially based on the comparison between the set

of previous legitimate transactions of a user and the new transactions under evaluation. This is a rather trivial criterion that in many cases, due to the high heterogeneity of data, leads toward misclassifications. In order to overcome this problem, a fraud detection approach should be able to use as much as possible information about the transactions during the evaluation process, but this is not always possible due to the inability of some approaches to manage some information (e.g., Random Forests, one of the most performing approaches, is not able to manage types of data that involve a large number of categories).

Another problem is that most of the academic work is not thought to be implemented and productionized in real environments. Conversely, companies aim to create models that can be deployed and integrated into their FDSs, but they rarely share their findings. During the last years, the amount of information that is being generated has increase data big pace. Although this is very promising or analytics, it might also turn into a problem because no insights can be extracted from them without the right infrastructures and technologies. Companies' data pipelines that correctly handle this volume of data require more technical complexity in order to scale well. One of the main reasons for this big complexity is the frequent bad quality of that data, which hinders analytics.

*Research Problem Statement.*Fraud is the act of swindling by some criminal individuals or fraudulent scheme. Payment card frauds are the types of fraud committed by cyber- criminals using stolen payment cards. Fraud can be avoided either by prevention or detection. Prevention avoids any type of attacks from fraudsters. In the case of an online mode of payment, the card may not present physically. In this type of case, the card's data is prone to attack by the hacker or cyber-criminal. Researchers in this field have been engaged by employing promising technologies such as machine learning algorithms to carry out detection of frauds on the online use of payment cards. For example, Dejan *et al* (2019), analyzed various machine learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection. However, there were further suggestions to exploit other techniques such as SVM. Similarly, in the work of Gaurav *et al* (2021), Isolation Forest machine learning algorithm was used to detect frauds in Credit Card transactions.

Therefore, this research work proposes to employ Support Vector Machine, K-Nearest Neighbor and possibly hybridized the two

techniques to analyze their strengths in payment cards fraud detection. This analysis would be carried out in comparison with the four techniques used by Dejan *et al* (2019), and on the same datasets.

Aim and Objectives

The aim of this research is to carry out comparative analysis on payment cards' fraud detection using Support Vector Machine and K-Nearest Neighbor algorithms. To actualize this aim, the following objectives are outlined.

1. To preprocess the dataset for both SVM and KNN algorithms.
2. To build the machine learning classification models.
3. To compare and evaluate the results of the classifications algorithms in objective i using precision, recall and accuracy.

Significance of the Study

This research work will solve the problem of detecting frauds in payment cards using machine learning algorithms such as Support Vector Machine and K-Nearest Neighbor. Detection of fraud in payment cards will reduce the loss of funds in billions of dollars by companies to fraudsters. A credit-card fraud leads to billions of dollars in losses for online merchants. With the development of machine learning algorithms, researchers have been finding increasingly sophisticated ways to detect fraud, but practical implementations are rarely reported. So, this research work can help researchers and practitioners to design and implement conceptual based systems for fraud detection or similar problems. This research has contributed not only with a conceptual base design, but also with insights to the fraud analysts for improving their manual revision process, which resulted in an overall superior performance. This work will take into account the definition of a hybrid approach of fraud detection that combines different machine learning and serve as a vital research material for data analysts, data scientists and other researchers in this field.

Scope and Limitation

This research work is based on detection of fraud in payment cards using Support Vector Machine and K- Nearest Neighbor. Supervised machine learning algorithms were used in the research . This research focuses on textual dataset. However, neither images nor audio nor video datasets are considered in the research work. Also, the algorithm considered does not include unsupervised, semi supervised or reinforcement machine learning algorithms.

Dealing with financial data is sensible, so publicly available datasets in this research area are scarce.

Literature Review

Overview of Fraud. The main task of a fraud detection system is the evaluation of a new financial transaction with the aim to classify it as legitimate or fraudulent, by using the information gathered in the past i.e. value of the features that compose each transaction and if it was a fraud or not.

According to Kuldeep et. al. (2018) worked on "Credit Card Fraud Detection Using AdaBoost and Majority Voting"; Credit card fraud is a serious problem in financial services. Billions of dollars are lost due to credit card fraud every year. There is a lack of research studies on analyzing real world credit card data owing to confidentiality issues. In this paper, machine learning algorithms are used to detect credit card fraud. Standard models are first used. Then, hybrid methods which use AdaBoost and majority voting methods are applied. To evaluate the model efficacy, a publicly available credit card data set is used. Then, a real-world credit card data set from a financial institution is analyzed. In addition, noise is added to the data samples to further assess the robustness of the algorithms. The experimental results positively indicate that the majority voting method achieves good accuracy rates in detecting fraud cases in credit cards.

In Malini et. al. (2017), "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection". An efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

According to Awoyemi et. al. (2017) in "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis". A hybrid technique of under- sampling and oversampling is carried out on the skewed data. The three techniques are applied on the raw and preprocessed data. The work implemented in Python. The performance of the techniques is evaluated based on accuracy, sensitivity, specificity, precision, Matthews's correlation coefficient and balanced classification rate. The comparative results show that k-nearest

neighbor performs better than naïve bayes and logistic regression techniques.

Harish (2017), worked on "Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review". The most common techniques used to make the fraud detection model. Incidentally, detection and prevention of credit card frauds are one of the vital problems in the digital world that need exact transactions analysis. One method for detecting fraud is to check for suspicious changes in user behavior. The purpose of this paper is to investigate Data mining techniques like Bayesian networks, Bayes Minimum Risk, Genetic algorithm, Hidden Markov Model (HMM) and Ontology for improve fraud detection in credit cards. This work primarily aims to improve current fraud detection processes by improving the prediction of fraudulent accounts.

You-Dai et. al. (2016) worked on," Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies". In this work he opined that researchers focused on designing an online credit card fraud detection framework with big data technologies, by which achieved three major goals: 1) the ability to fuse multiple detection models to improve accuracy; 2) the ability to process large amount of data and 3) the ability to do the detection in real time. To accomplish that, we propose a general workflow, which satisfies most design ideas of current credit card fraud detection systems.

According to Nuno et. al. (2017), who worked on "A data mining based system for credit-card fraud detection in e-tail", revealed that a credit-card fraud leads to billions of dollars in losses for online merchants. With the development of machine learning algorithms, researchers have been finding increasingly sophisticated ways to detect fraud, but practical implementations are rarely reported. The paper can thus help researchers and practitioners to design and implement data mining based systems for fraud detection or similar problems. This has contributed not only with an automatic system, but also with insights to the fraud analysts for improving their manual revision process, which resulted in an overall superior performance.

Offline Fraud: According to Khyati, et al (2018), offline fraud is committed by using a stolen physical card at call center or any other place. Offline fraud refers to obtaining goods/services and money by illegal way physically. Fraud deals with events which involve criminal motives that, mostly, are difficult to identify. Credit cards are one of the most popular objective of fraud but not the only one.

Online Fraud: Online fraud is a type of fraud which makes use of the internet. It is not a single fraud; there are numerous frauds under that. Internet fraudsters are everywhere and

they come up with innovative tricks to cheat people and wipe out money from their bank account. According Harish (2017), online customers are increasing day by day. The customers now want to purchase the goods by sitting at their homes because of different reasons. For example, purchasing goods online saves the time of the customers. The increasing number of online customers makes credit card fraud, a more challenging and important problem. Electronic payment has several issues but the major issue is the credit card fraud. The e-commerce stores also provide the facility of cash on delivery but the customers mostly purchase the goods through the credit card.

Payment Card Fraud: Payment Card Fraud is one of the biggest threats to business and commercial establishments today. Simply, Credit Card Fraud is defined as, "when an individual uses another individuals" credit card for personal use while the owner of the card as well as the card issuer are not aware of the thing that the card is being used." A number of systems/models, process and preventive measures will help to stop credit card fraud and reduce financial risks, Khyati, et al (2018).

Credit Card: Credit card is issued by financial institutions to its customers (cardholder); usually allowing them to purchase goods and services within payment limit or withdraw cash in advance. Credit card is a medium of selling goods or services without having cash in hand. A credit card is a simple way of offering credit to a consumer automatically. Today, almost every credit card carries an identifying number that helps in shopping transactions rapidity. According to Lakshmi (2018), Card fraud begins either with the theft of the physical card or with the important data associated with the account, including the card account number or other information that necessarily be available to a merchant during a permissible transaction. Card numbers generally the Primary Account Number (PAN) are often reprinted on the card, and a magnetic stripe on the back contains the data in machine-readable format. It contains the following Fields: Name of card holder, Card number, Expiration date, Verification/CVV code, Type of card.

Related Studies. Mirjana et al (2019) analyzed various machine learning algorithms, such as Logistic Regression, Random Forest, Naive Bayes, and Multilayer Perceptron to determine which is most suitable for credit card fraud detection using dataset which can be downloaded from Kaggle. Confusion matrix was used as a metrics for optimal result. Accuracy for LR, NB, RF, MLP are 97.46%, 99.23%, 99.96%, and 99.93%
respectively. Radom forest performed better than the rest. Further research should focus o different machine learning algorithms such as genetic algorithm and different type of stack

classifiers alongside with extensive feature selection to get better result.

According to Vaishnavi et al (2019) in Credit Card Fraud Detection using Machine Learning Algorithms, different classification algorithms were analyzed and the result shows that Logistic Regression, Decision tree and Random Forest are the algorithms that gave better results. The objective of the research work is to overcome the problem of concept drift to implement on real-world scenario. The future work will focus on solving the problem of using different machine learning algorithm to detect fraud transactions. The random forest algorithm should be improved.

In the literature by Elena-Adriana et al (2019), Meta-analysis was performed on various numbers of specialized articles (peer-reviewed journals, articles and conference papers) to analyze and classify machine learning techniques suitable to detect bank fraud in online environment taking high accuracy, high coverage and low cost into account. The result shows high accuracy and high coverage for supervised techniques with the disadvantage of high cost.

A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection carried out by Vipul et al (2020) using Machine learning process in credit card fraud detection, shows that different data mining and machine learning methods can be used for credit card fraud detection.

According to Lakshmi (2018) in Machine Learning For Credit Card Fraud Detection System, comparison are made for different machine learning algorithm such as logic regression, decision tree, random forest to determine which algorithm gives suits best and can be adapted by credit card merchants for identifying fraud transaction. The accuracy result show for logic regression, decision tree, and random forest is 92.7, 95.8, and 97.6 respectively. The result shows that the random forest performed better than the logistic regression and Decision tree techniques.

Harish (2017) worked o Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review to provide a complete literature review about credit card fraud using inclusion and exclusion, quality assessment, research question and search process. The research shows that credit card fraud is the major issue of financial sectors that is increasing with the passage of time and future researchers can work on fraud detection system for detection of credit card fraud

According to Khyati et al (2018) in A review of Fraud Detection Techniques: Credit Card, using modern techniques based on Data Mining, Machine Learning, Sequence Alignment,

Fuzzy Logic, Genetic Programming, and Artificial Intelligence for detecting credit card fraudulent transaction. The neural network based CARDWATCH shows much great accuracy in fraud detection and processing speed is also high but it is limited to one-network per customer.

In a literature by Ignacio (2017) in Fraud detection in online payments using Spark ML to create a fraud detection system to tackle fraudulent transactions made through online payment cards, the was structured using the CRISP-DM methodology and random forest performed well for this kind of problem because they can handle categories and they are not very sensitive to noisy data.

According to Roberto et al (2017), in Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach, the proposed methodologies are the Frequency Spectrum Evaluation and Random Forests Approach. The approach is able to reduce/overcome the data imbalance and cold-start issues, since only a class of transactions is used.

Batta (2020) in is literature, briefly reviewed the future prospect of the vast applications of machine learning algorithms. The paper surveys various machine learning algorithms and explained each in details.

According to Gaurav et al (2021) in Credit Card Fraud Detection Using Isolation Forest, KNN, Decision Tree, Logistic Regression and Random Forest models were used to illustrate the design of a data set using machine learning with Credit Card Fraud Detection. The study showed that Machine Learning can efficiently support fraud detection.

Iong-Zong et al (2021) worked on Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert to devise an advanced financial fraud detection systems that can actively detect the risks such as illegal transactions and irregular attacks using Deep Convolution Neural Network (DCNN) scheme based financial fraud detection scheme and an detection accuracy of 99% was been obtained.

Harshita et al (2021) carried out a research to predict the accuracy/precision of the fraud detection through different algorithms. The process involved the analysis and the pre-processing of data sets as well as the utilization of multiple Anomaly detection algorithms such as Local Outlier Factor, Super Vector Machine and many such relevant algorithms. The algorithm reaches over 99.6% exactness, its precision remains only at 28% when only a tenth of the data set is taken into consideration.

In Credit Card Fraud Detection using Machine Learning Algorithms by Parth et al (2020),

Logistic Regression, K-Nearest Neighbor, and Naïve Bayes are analyzed on extremely skewed Master card fraud information to vogue and develop a very distinctive fraud detection technique for Streaming dealing information, with an objective, to analyze the past dealing details of the purchasers and extract the behavioral patterns. Finally, it was determined that Logistic regression, decision tree and random forest area unit the algorithms that gave higher results.

According to Manoj et al (2021), Decision Tree, Logistic Regression, SVM and Naïve Bayes were used to develop a Credit Card Fraud Detection System to spot whether a new transaction is fraudulent or not with the knowledge of previous data. The research showed that Random Forest yielded the most satisfactory results with high accuracy and consistent F1-Score in both cases. Hence, it appears to be the most suitable for classifying large-volumes of data samples.

Research Gap

The researcher has reviewed many related literatures and results showed that different machine learning algorithms can be applied in credit card fraud detection. Therefore, this research was carried out to compare the result of Support Vector Machine and K-Nearest Neighbor algorithms on credit cards' fraud detection and hybridized SVM-KNN algorithm against the result by Dejan et *al* (2019), who analyzed various machine learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection.MLP.

Methodology

The research implores a result based evaluation methodology to carry out the comparative analysis of the three proposed machine learning algorithm, i.e. Support Vector Machine (SVM), K-nearest Neighbor (KNN) and combined KNN-SVM that are adopted against the four algorithms i.e. Naïve Bayesian, Logistic Regression, Multi-Layer Perceptron, and Random Forest as adopted by the related work in which serve as a benchmark work for this.

The conceptual framework of the proposed algorithms consists of the activities carried out to obtain the best result from the models.

Testing Buffer

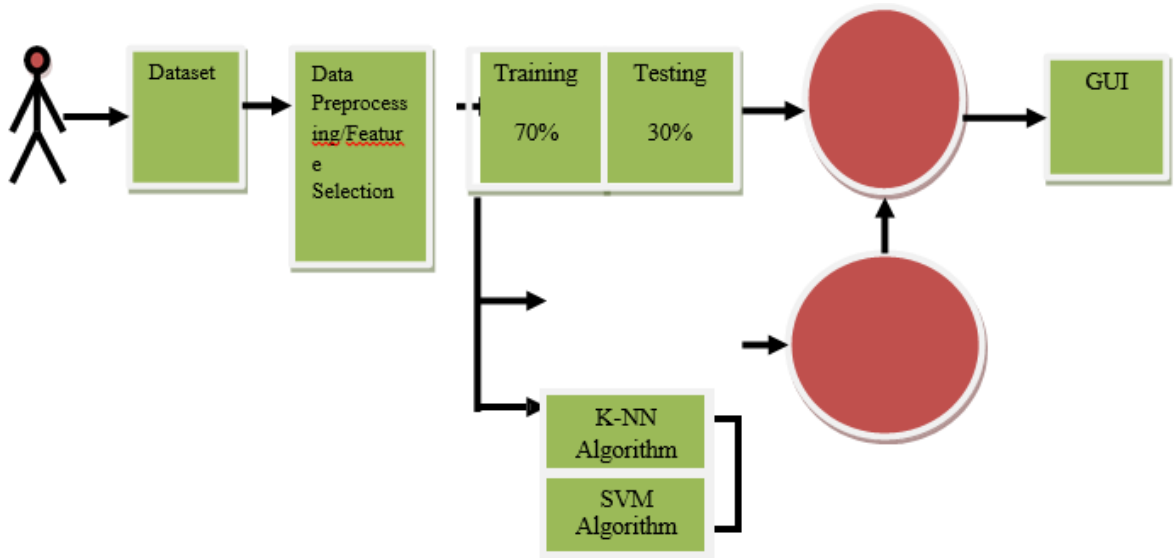

Figure 3. 1: Conceptual Framework of the proposed Algorithms.

Hybridized K- NN/SVM

It consists of dataset which can be downloaded from Kaggle. Feature selector tool by Will Koehrsen was used in this experiment for the feature selection. By using this tool it has been determined which features are the most important. Synthetic Minority Oversampling Technique (SMOTE) is a popular oversampling method that has proven useful when used on imbalanced dataset. The dataset was divided into training and testing dataset in the ratio of 70% and 30% respectively. 70% of the dataset was trained in the KNN, SVM, and combined KNN-SVM. 30% of the dataset was tested at the buffer section and the result was displayed on the Graphic User Interface. The dataset was divided into the training dataset and testing dataset. The training dataset was moved into KNN and SVM model and the result formed the combined KNN-SVM. The tested dataset was moved to the testing buffer and the result was displayed on the GUI.

This research proposed Result-based evaluation as technique to carry out the comparative analysis. The evaluation would be carried out by considering three metrics; these are precision, recall and accuracy.

Precision metric is determined by dividing the True Positive (TP) by the summation of TP and False Positive (FP).

Recall metric is computed by finding the summation of TP and True Negative (TN) and

divided it by the summation of TP and False Negative (FN).

Accuracy measure is the summation of TP and True Negative (TN) and divided it by the summation of TP, TN, FP and FN.

These metrics are mathematically expresses as follows:

1. Precision:

$$\frac{TP}{TP + FP}$$

2. Recall

$$\frac{TP + TN}{TP + FN}$$

3. Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN}$$

The performance of the adopted machine learning algorithms against the four algorithms in the related work would be outlined for researchers in this field to make decision.

RESULTS AND DISCUSSION

This chapter presents the proposed machine learning algorithms i.e. KNN, SVM and Hybridized KNN-SVM, the analysis of the dataset and outlining of the results against the models from related algorithms used as benchmark for this research.

The implementation and analysis of the proposed machine leaning algorithms i.e. K- Nearest Neighbor, Support Vector Machine and the hybridized K-Nearest Neighbor and Support Vector Machine was carried out using the python programming language. Python is an interpreted high-level general purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large scale . Python is mostly preferred when dealing with machine learning algorithms. The system requirements are divided into two categories namely: hardware and software requirements.

The minimum hardware requirement for the implementation and analyzing of the proposed machine learning model are as follows:

a. Processor: 1.5GHz CORE i3 Intel Inside b. RAM: 2G c. Display: 12”d.HDD: 64G

The implementation system environment is Windows 10 operating system, and the software

operating environment is Spyder, scientific python development environment, which is part of the Anaconda platform. Used libraries include: numpy, pandas, scikitplot, matplotlib, plotly, seaborn, jupyter notebook, and sklearn.

This section presents the results of the analysis in comparison with four other techniques (MLP, RF, LR and NB) which have been considered by the related researchers. From Table 4.1, the first four techniques are from the related researcher; while the last three are what this research considered based on the same datasets.

Summary of the Results

Table 1: Summary of the Results.

Techniques	Precision	Recall	Accuracy
MLP	79.21%	81.63%	99.93%
RF	96.38%	81.63%	99.96%
LR	16.17%	82.65%	99.23%
NB	58.82%	91.84%	97.46%
KNN	44.79%	79.59%	99.79%
SVM	1.89%	71.43%	93.60%
KNN-SVM	18.67%	74.49%	99.40%

Based on the results presented by the above table, the first four algorithms outperform the proposed two algorithms in this work.

LR against the proposed Algorithms

Table 2: Comparative Analysis of LR against the proposed Algorithms.

	Precision	Recall	Accuracy
LR	√	√	X
KNN-SVM	X	X	√
SVM	X	X	X
KNN	X	X	√

It is very obvious that LR technique outperform the proposed techniques in terms of precision and recall. However, the accuracy of the proposed techniques that is, KNN and the combination of SVM and KNN outweigh the LR.

NB against the proposed Algorithms

Table 3: Comparative analysis of NB against the proposed Algorithms.

	Precision	Recall	Accuracy
NB	√	√	X
KNN-	√	X	√

SVM			
SVM	X	X	X
KNN	√	X	√

From the comparative analysis on Table 3, NB performed better than the SVM in recall. KNN and KNN-SVM performed better than the LR in accuracy and precision.

RF and the proposed Algorithms

Table 4: Comparative analysis of RF against the proposed Algorithms.

	Precision	**Recall**	**Accuracy**
RF	√	√	√
KNN-SVM	X	X	X
SVM	X	X	X
KNN	X	X	X

From table 4, after the comparative analysis, it was deduced that RF algorithm performed better than KNN, SVM, and KNN-SVM in precision, recall and accuracy.

MLP and the proposed Algorithms

Table 5: Comparative analysis of MLP against the proposed Algorithms.

‘	**Precision**	**Recall**	**Accuracy**
MLP	√	√	√
KNN-SVM	X	X	X
SVM	X	X	X
NN	X	X	X

From table 5, after proper comparative analysis, MLP algorithm performed better than KNN, SVM, and KNN-SVM in precision, recall and accuracy.

1. SUMMARY, CONCLUSION AND RECOMMENDATION

In summary, the research shows that selected machine learning algorithms such as K- Nearest Neighbor, Support Vector Machine and the hybridized K-NN-SVM can be used to detect frauds in payment cards. The comparative analysis of the proposed machine learning algorithms against other related machine learning algorithms by researchers shows that Logistic Regression, Naïve Bayesian, Random Forest and Multilayer Perceptron can be perfectly used in detection of frauds in payment cards. In summary, according to accuracy for the proposed machine learning algorithms, KNN performed better than hybridized KNN-SVM and SVM with the value of 99.7%.

In conclusion, payment card frauds represent a very serious business problem. These frauds

can lead to huge losses, both business and personal. To detect fraud in payment cards, this research work has provided a reference and scientific base for researchers in the course of their research on this field of study. This research work has proven that many machine learning algorithms can be used to detect fraudulent payments using payment cards. Also, the comparative analysis has shown that some machine learning algorithm can outperform others. The research recommends that online fraud regarding payment cards has been a serious crime globally which has drawn the attention of researchers. Therefore, much research efforts are required to address the problems. In lieu of this, machine learning algorithms have been consistently employed to solve the gap. However, while some techniques are promising, others have serious challenge in meeting up with the expectations. Therefore, this research recommends that the application of RF and MLP algorithms are promising. Interested researchers in this field are advised to employ Deep Learning (DL) techniques for better and effective results. Improvements are also encouraged to be carried out in the techniques for preprocessing and features selection.

REFERENCES

1. Araujo M., M. Almeida, J. Ferreira, L. Silva, and P. Bizarro, "Breachradar: Automatic detection of points-of- compromise," in Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM, 2017, pp. 561– 569.
2. Awoyemi J. O, A. O. Adentumbi, S. A. Oluwadare, (2017) "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", Computing Networking and Informatics (ICCNI), 2017 International Conference on pp. 1-9. IEEE.
3. Batta Mahesh (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research* (IJSR) ISSN: 2319-7064 Research Gate Impact Factor (2018): 0.28 | SJIF (2018): 7.426 Volume 9 Issue 1, January 2020 www.ijsr.net
4. DejanVarmedja, Karanovic, SrdjanSladojevic, Marko Arsenovic, AndrasAnderla (2019). Credit Card Fraud Detection - Machine Learning methods. *18th International Symposium INFOTEH-JAHORINA, 20-22 March 2019*
5. Elena-Adriana MINASTIREANU, Gabriela MESNITA (2019). An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection. *InformaticaEconomică vol. 23, no. 1/2019. DOI: 10.12948/issn14531305/23.1.2019.01*
6. European Central Bank (2018). Fifth report on card fraud, September 2018. [online]. Available at: https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport2018 09.en.html#toc1 [Accessed21Jan. 2019].

7. Gaurav Kumar Singh, AkhileshBhayye, SanikaDhamnaskar, SandeepPatil, S. V. Phulari (2021). Credit Card Fraud Detection Using Isolation Forest. *International Journal of Recent Advances in Multidisciplinary Topics Volume* 2, Issue 6, June 2021 https://www.ijramt.com | ISSN (Online): 2582-7839
8. Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/st artups-worldwide/ [Accessed 10 Jan. 2019].
9. Harish Paruchuri (2017). Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review. ABC *Journal of Advanced Research*, Volume 6, No 2 (2017) ISSN 2304-2621(p); 2312-203X (e)
10. Harshita Anand, RichaGautam, Raman Chaudary (2021). Credit Card Fraud Detection using Machine Learning. School of Computer Science and Engineering, Galgotias University
11. Ignacio Amaya De La Peña (2017). Fraud detection in online payments using Spark ML. *Degree TechnologyAnd Learning*, Second Cycle, 30 Credits, Stockholm Sweden 2017
12. Joy Iong-Zong Chen, Kong-Long Lai (2021). Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert. *Journal of Artificial Intelligence and Capsule Networks* (2021) Vol.03/ No.02 Pages: 101-112
13. Kaggle.com. (2019). Credit Card Fraud Detection. [online] Available at: https://www.kaggle.com/mlg- ulb/creditcardfraud [Accessed 10 Jan. 2019].
14. Khyati Chaudhary, JyotiYadav, BhawnaMallick (2018). A review of Fraud Detection Techniques: Credit Card. *International Journal of Computer Applications* (0975 – 8887) Volume 45– No.1, May 2018
15. Kuldeep, Randhawa, et al (2018) "Credit Card Fraud Detection Using AdaBoost and Majority Voting."IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420.
16. Lakshmi S. V. S. S,SelvaniDeepthiKavila (2018). Machine Learning For Credit Card Fraud Detection System. *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 13, Number 24 (2018) pp. 16819-16824 © Research India Publications. *http://www.ripublication.com*
17. Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2019].
18. Malini N., Dr. M. Pushpa, "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection", Advances in Electrical, Electronics, Information,

Communication and BioInformatics (AEEICB), 2017 Third International Conference on pp. 255- 258. IEEE.

19. Manoj Kumar Reddy Mallidi, YeshwanthZagabathuni (2021). Analysis of Credit Card Fraud detection using Machine Learning models on balanced and imbalanced datasets. *International Journal of Emerging Trends in Engineering Research.* Volume 9, No 7. July 2021.
20. Nuno Carneiroa T., Gonc¸alo Figueiraa, Miguel Costab (2016) "A data mining based system for credit-card fraud detection in e-tail", Journal of DSS, Sep 2016, PP 1-11
21. Parth Parashar, PrashantBhati (2020). Credit Card Fraud Detection using Machine Learning Algorithms.International Research Journal In Advanced Science & Technology (IRJAST) Volume-1, Issue-1, December 2020
22. Roberto Saia and Salvatore Carta (2017). Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach .*DipartimentodiMatematicae Informatica, Universit`a di Cagliari, Italy*
23. VaishnaviNathDornadul, Geetha S. (2019). Credit Card Fraud Detection using Machine Learning Algorithms. *International Conference On Recent Trends In Advanced Computing 2019,* ICRTAC 2019
24. Vipul Patil, Dr. Umesh Kumar Lilhore (2018). A Survey on Different Data Mining & Machine Learning Methods for Credit Card Fraud Detection. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* © 2018 IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307
25. You Dai, Jin Yan, Xiaoxin Tang, Han Zhao and Minyi Guo," Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies", 2016 IEEE TrustCom/BigDataSE/ISPA, PP 1644-1653.