OPTIMIZING FRAUDULENT TRANSACTION DETECTION IN E-COMMERCE: A COMPARATIVE ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS WITH TIME AND CPU PERFORMANCE TRACKING

MSc Research Project
MSc Cybersecurity

CHIJIOKE FRANKLIN EMEJURU
STUDENT ID: X21114382

School of Computing
National College of Ireland

Supervisor: JOEL ALEBURU

National College of IrelandProject Submission Sheet School of Computing

Student Name:	CHIJIOKE FRANKLIN EMEJURU
Student ID:	X21114382
Programme:	MSc Cybersecurity
Year:	2024
Module:	MSc Research Project
Supervisor:	JOEL ALEBURU
Submission Due Date:	12/12/24
Project Title:	Optimizing Fraudulent Transaction Detection In E-Commerce: A Comparative Analysis of Machine Learning And Deep Learning Algorithms With Time And CPU Performance Tracking.
Word Count:	23
Page Count:	6394

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Chijioke Franklin Emejuru
Date:	12th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	Q
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	Q
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keepa copy on computer.	Q

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	

Date:	
Penalty Applied (if applicable):	

OPTIMIZING FRAUDULENT TRANSACTION DETECTION IN E-COMMERCE: A COMPARATIVE ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS WITH TIME AND CPU PERFORMANCE TRACKING

CHIJIOKE FRANKLIN EMEJURU

X21114382

ABSTRACT

This research presents a concise analysis on the application of machine learning and deep learning techniques for fraudulent detection in e-commerce. With the increasing number of cases of fraudulent activities, many institutions face challenges in detecting these practices in due time. This research evaluates some machine learning techniques such as logistic regression, random forest, support vector machine, decision trees, xgboost, gradient boosting and a deep learning multi-layer perceptron for their effectiveness in the detection. Key findings reveal that Random forest and the ensemble models, with their balance of accuracy and complexity, emerged as the best models with random forest being on top with an accuracy of 99.97% in the detection of fraudulent transactions.

Keywords: logistic regression, random forest, support vector machine, decision trees, xgboost, gradient boosting and a deep learning multi-layer perceptron, fraudulent detection, machine learning.

TABLE OF CONTENTS

1.0. INTRODUCTION

The financial sector's ongoing development has increased its vulnerability to fraudulent activity, making financial fraud detection and prevention a top priority for institutions all over the world (Khan & Malaika, 2021). Financial fraud is a wide range of illicit acts that can cause serious economic and reputational harm, compromise the stability of financial institutions (Sailio et al., 2020), and range from straightforward scams to complex white-collar crimes. As a result, one of the most important prerequisites for the security and integrity of financial systems is the capacity to quickly recognise and neutralize such threats (Shojaifar, 2023). With the rise of online financial transactions and the advent of digital banking, fraudsters now have more ways to commit their illicit crimes, and they are increasingly making use of technological improvements to do so (Bhasin, 2016). Due to the complexity and number of modern financial transactions, the old methods of fraud detection which frequently include manual checks and basic rule-based algorithms are showing themselves to be insufficient (Kulshrestha, 2022). In addition to requiring a lot of resources, these traditional methods have a significant time lag in fraud detection and a high proportion of false positives. Also, Since the advent of credit cards and online payments, a lot of fraudsters have figured out ways to take advantage of people and steal their credit card details so they may be used for unapproved transactions. This leads to a tremendous volume of fraudulent purchases every day. Before transactions are approved, financial institutions attempt to detect fraudulent activities using machine learning and deep learning techniques. In comparison to other payment methods including e-wallets and bank transfers, credit cards were the most widely used payment method worldwide in 2014, according to the Global Payments Report 2015 (Ul et al., 2017). Cybercriminals frequently target massive transactional services with the intention of utilizing credit card services to carry out fraudulent actions. Unauthorized use of a card, strange transaction patterns, or transactions on a deactivated card are all considered forms of credit card fraud. Credit card fraud often falls into three categories: traditional frauds which include stolen and counterfeit cards (Azam et al., 2023), internet frauds which include fraudulent merchant websites (Quah & Sriganesh, 2008), and merchant-related scams which include merchant collusion and triangulation (Gulati et al., 2017). The Nilson Report estimates that worldwide credit card fraud losses will surpass $35 billion by 2020 (Wu et al., 2019), having hit $16.31 billion in 2014. As a result, to counteract illicit activities, credit card fraud detection systems must be developed. Scholars have frequently adopted data mining and machine learning techniques to investigate and identify credit card fraud activities, given their widespread application in combating cybercrime (Mutemi & Bacao, 2024).

The process of extracting interesting, insightful, and perceptive patterns from massive data sets and identifying comprehensible, descriptive, and predictive models is called data mining. By distinguishing between the features of typical and questionable credit card transactions, data mining techniques can help detect credit card fraud by extracting valuable information from vast amounts of data using statistical and mathematical methods (Almarshad et al., 2023). Machine learning involves creating models based on learned features for tasks such as classification and regression (Arrieta et al., 2019), and other purposes, whereas data mining concentrates on finding valuable intelligence (Duan et al., 2019). In many computer science disciplines, including spam filtering, web searching, ad placement, recommender systems, credit scoring, medication design, fraud detection, stock trading, and many more, machine learning techniques are applied. Instead of rigidly following static program instructions, machine learning classifiers work by creating a model from sample inputs and utilizing that to generate predictions or choices (Shorten & Khoshgoftaar, 2019). Different types of machine learning algorithms are developed to address various problems. Classifying items using machine learning involves first identifying, comprehending, and then classifying them into preset categories (Razzak et al., 2019). Ideally, learning can be divided into several categories, such as supervised, unsupervised, semi-supervised, and reinforcement learning. (Mahdavinejad et al., 2017).

Rule-based systems are the foundation of conventional fraud detection methods (Hussain et al., 2020). These systems function in line with specified rules and guidelines developed by experts based on historical data and identified fraud trends. Despite their advantages, traditional rule-based systems have certain drawbacks that limit their effectiveness against developing forms of fraud. Because rule-based systems are fixed, they cannot adapt to new fraud techniques (Chami et al., 2009). The rules cannot be changed after they are created; they must be manually updated, which might take time and cause the rules to trail behind newly identified fraud patterns. Rules are usually too broad to identify as many fraud attempts as possible, which leads to a high number of false positives (Johnson & Khoshgoftaar, 2019). One limitation is that because they rely on existing patterns and past data, these solutions are useless against new or complex fraud strategies that don't follow predetermined standards. Updating and maintaining rule-based systems requires a significant amount of human labor and expert input, which can be expensive and resource-intensive (Alzubaidi et al., 2023). Given the limitations of traditional methods, fraud detection calls for more adaptable and dynamic solutions than in the past. Real-time learning, adaptability, and predictive capabilities are essential for systems due to the constantly changing fraud landscape and its increasingly complex tactics. These systems can identify patterns, identify anomalies, and analyze massive amounts of data without according to preset rules. Machine learning algorithms can continuously improve their accuracy and adaptability by utilizing historical transaction data. (Zhang et al., 2019).

Machine learning has become one of the most popular topics in the last decade. An increasing number of companies are seeking to improve their offerings by investing in machine learning. Machine learning integrates a variety of computer approaches with statistical modeling to allow the computer to perform jobs without the need for hard coding (Jumper et al., 2021). A machine learning algorithm with the right training would be able to spot distinct correlations throughout the whole dataset (Mehta et al., 2019).

Using one deep learning model, this study compares six popular machine learning classifiers to detect and classify fraudulent transactions. The classifiers include logistic regression, random forest classifier, extreme gradient boosting, gradient boosting, support vector machine, decision trees, and multilayer perceptron. By employing a range of specialized models to train the dataset provided in the study, every constraint in the data will be handled by the models utilized. Additionally, machine learning models may have biases due to the distribution of the data used in training the model. Therefore, by combining a few different approaches, it is possible to verify the many forms of fraudulent activity, reduce model biases, and enhance efficiency and scalability, all of which contribute to the creation of more effective solutions for the detection of fraudulent activity in e-commerce platforms. In this research, the computational efficiency and training time of the models will be measured as this would allow the best model with the least cost to be the most efficient model in the real world deployment.

1.1. SIGNIFICANCE OF THE STUDY

This study plays a crucial role in protecting financial assets by precisely detecting and identifying fraudulent transactions. By safeguarding individuals, businesses, and institutions from substantial risks and losses, it saves money for customers and boosts confidence in financial systems. Additionally, it reduces the risk of customer churn and operational costs for financial institutions, as automation decreases the need for manual reviews. Overall, the development of an effective fraud detection system will protect the image and reputation of financial institutions.

1.2. AIM AND OBJECTIVES

To evaluate and optimize the performance of machine learning and deep learning models for fraudulent transaction detection in e-commerce, focusing on both detection accuracy and computational efficiency (time and CPU usage) to enhance scalability and applicability with the following objectives:

- Implementing and comparing the performance of various machine learning algorithms and a deep learning multi-layer perceptron model for fraudulent transaction detection in e-commerce.
- To track and analyze time and CPU usage for each model to assess their computational efficiency.
- To identify the trade-offs between model accuracy and resource consumption to recommend the most efficient algorithm for real-world deployment.
- To provide insights into optimizing fraud detection systems by balancing detection performance with operational costs.

1.3. RESEARCH QUESTIONS

- Which machine learning and deep learning models offer the highest accuracy in the detection of fraudulent activities in e-commerce?
- How can computational efficiency be optimized without compromising the accuracy of fraud detection models?
- Which algorithm offers the best balance between accuracy, time efficiency and CPU performance for a scalable fraud detection system in practical e-commerce applications?

1.4. LIMITATIONS

When implementing a fraud detection system using data and machine learning techniques, there are several limitations which could be encountered when implementing the research process, these limitations can be seen in the following:

- Data availability and quality: as fraud detection models and systems depend solely on historical transaction data, a limited access to quality data and information can reduce the effectiveness and generalizability of the model.
- Scalability: it is common knowledge that most models perform well during building processes in a controlled environment but perform poorly at production level due to many reasons such as training on small data and computational complexity.
- Attacks from adversaries: as more novel detection techniques are developed; fraudsters and scammers continually adapt and evolve their methods to bypass the detection systems.

- The use of a single dataset which limited generalizability. There was no confirmation using cross validation on the results.

2.0. RELATED WORK

The detection of fraudulent transactions, especially in financial institutions, is pertinent for the reduction of crimes related to fraud. In this literature review, various techniques for detecting fraudulent transactions using machine learning will be explored in different sub-headings.

2.1. FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS

In the detection of fraudulent activities using machine learning, industries such as finance and telecommunications companies have become keen on developing protective softwares with respect to mitigating risks and vulnerabilities. This literature review examines several papers that applied different machine learning techniques to detect fraudulent activities with the strength and limitations of each being highlighted after each review. The development of algorithms capable of generating broad patterns and theories through externally given instances to forecast the fate of subsequent instances is known as supervised machine learning. The goal of supervised machine learning classification algorithms is to classify data based on previously acquired knowledge.

Dornadula et al. (2019) presented a predictive approach using supervised learning in the detection of fraudulent activity using machine learning in a research journal for computer science, in this research the authors utilized a dataset obtained known as credit card fraud dataset from kaggle which is an online database for machine learning. In this dataset, there were 284, 807 transactions in total and due to the nature of the research, most of the features were transformed using principal component analysis since providing private transaction details of customers would give rise to an issue. The dataset was highly imbalanced; hence the SMOTE sampling technique was done on the dataset to handle the data imbalance. Prior to the evaluation, the researchers trained the dataset on a few machine learning algorithms before and after using the SMOTE sampler. The machine learning algorithms utilized in this experiment are local outlier factor, isolation forest, support vector machine, logistic regression, decision tree and random forest. The results obtained before and after the data balancing showed that SMOTE boosts the performance of models like Logistic Regression, Decision Trees, and Random Forest by creating a more balanced dataset that allows for better boundary identification between classes. However, it disrupts anomaly detection methods such as LOF and Isolation Forest, which rely on the natural distribution of data. From the results obtained by the researchers, it was seen that the training process for a support vector machine with the SMOTE sampler resulted in a computation problem, so it was omitted from the results after the sampler was used. Before the sampler was used, the models suffered from a bias towards the majority class, as the accuracy was more, prior to the sampling.

Alarfaj et al. (2020) also used the same approach and the same dataset as the previous authors for the detection of fraudulent transactions. In this paper, the authors used a couple of machine learning algorithms as the main method for identifying fraudulent transactions with contrast to a deep learning model CNN as a model of comparison. The authors also took note of the class imbalance which the previous authors also took note of. The imbalance was treated by removing non fraudulent transactions from the dataset because in a real-world scenario, there can never be a balanced fraudulent and non- fraudulent transactions, only a few can be fraudulent. The machine learning algorithms utilized by the authors are decision trees, KNN algorithm, logistic regression, support vector machine, random forest and xgboost. The deep learning model utilized was the CNN and the accuracy performance was 96.34. While the obtained results look promising, it looks like the result from an overfitting which is not a good thing as it can result in poor generalization in real world situations. This can be seen in the contrast between the accuracy and the F1 scores in the machine learning algorithms where the SVM used alongside other machine learning algorithms had an accuracy of 99.93% and a F1 score of 77.71% . This problem can be solved by using proper dataset balancing techniques as stated as a limitation by the authors as the performance significantly decreases on unseen data.

Raghavan et al. (2019) took this approach further by utilizing the same dataset as the previous authors but this time he added two more datasets in his analysis to detect fraudulent transactions from the UCI machine learning repository, which are known as the Australian and German dataset. In this research, the authors utilized a combination of machine learning algorithms such as RBM, autoencoders, random forest, CNN, SVM, KNN, DBN and the ensemble of KNN, SVM and DBN, KNN, SVM and random forest, SVM, CNN and random forest. The results obtained from this research were divided according to the datasets in which the experiment was conducted and they showed that for the Australian dataset, the models show a range of effectiveness, with Restricted Boltzmann Machines (RBM) and autoencoders on the lower end of Area Under the Curve (AUC) scores, reflecting limited capability to capture the patterns within this dataset. As the models progress to more sophisticated algorithms like Random Forests, Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and K-Nearest Neighbors (KNN), the AUC scores increase, highlighting improved class separation. Deep Belief Networks (DBNs) show further enhancement, while ensemble models—especially those combining KNN, SVM, and Random Forest—achieve the highest AUC, demonstrating strong predictive power from the diversity of methods. For the German dataset, the trend is similar but with overall lower AUC scores, suggesting this dataset is more challenging for the models to generalize. While the AUC improves with advanced models and ensemble approaches, the SVM, CNN, and Random Forest ensemble produces the highest score, indicating that leveraging complementary strengths across models optimizes performance for this dataset. In this research, the SVM with combination with CNN performed more on a larger dataset and for the smaller dataset, the ensemble approaches of SVMs, KNN and random forests provided good results. The limitation in this research lies on the fact that this method of fraud detection is only for supervised learning as fraud patterns change periodically over time resulting in the need for a new dataset to be used to train a new machine learning model, this limitation can be mitigated by using unsupervised learning approaches to detect fraud in these real world scenarios.

2.2. FRAUD DETECTION IN BANKING SYSTEMS USING DEEP LEARNING NEURAL NETWORKS

Based on a clearly defined computational architecture, neural networks function as a potent computational tool for resolving problems related to diagnostics, detection, prediction, and decision-making. It has been effectively used in many different fields, including computer security, voice recognition, image and video identification, industrial problem detection, finance, and medical diagnostics. Deep artificial neural networks, including architectures like Convolutional Neural Network (CNN) and Transformer models, have been achieving state-of-the-art results in machine learning and pattern recognition competitions in recent years. Pumsirirat et al. (2018) proposed a system of detecting fraudulent transactions using autoencoders and restricted Boltzmann machines, both of which are deep learning algorithms. This study was mainly empirical, involving experiments using real datasets to evaluate the performance of deep learning models. The dataset utilized by the author was the Australian and German dataset, a Credit Card Fraud dataset from kaggle, which contains transactions made by European cardholders consisting of 383 normal and 307 fraud instances for the former and 700 normal and 300 fraud instances for the latter and added a third, which was the European dataset. The datasets were trained on the two algorithms, and the AUC was used as the main metric of performance to determine the strength of each algorithm. On the German dataset, the RBM performed better than the autoencoders with an auc score of 45.62, on the Australian dataset, the autoencoder performed better with an auc score of 54.83 and finally on the European dataset, the autoencoders performed better with an AUC score of 96.03. Based on the results of these calculations and experiments, the authors concluded that supervised learning techniques are better for the detection of fraudulent transactions. The authors concluded that it would be better to use a real dataset for fraudulent transactions to train the models for better inference which can also be concluded that the dataset used for the experiment is the limitation.

2.3. UNSUPERVISED LEARNING FOR FRAUD DETECTION IN TELECOMMUNICATIONS

Although artificial intelligence and machine learning have long been used in networking research, most of these studies have concentrated on supervised learning (Akyildiz et al., 2020). Utilizing unstructured raw network data, unsupervised machine learning has been more popular recently as a means of enhancing network performance and offering services like traffic engineering, anomaly detection, Internet traffic classification, fraudulent detection and quality of service optimization. Based on shared qualities, objects are grouped together. Partition clustering and hierarchical clustering are the two groups into which the clustering methods fall. Over the last ten years, many unsupervised learning approaches and algorithms have been developed, some of which are well-known and often used unsupervised learning algorithms (Shrestha & Mahmood, 2019). Unsupervised learning techniques have shown great promise in fields such as natural language processing, speech recognition, machine vision, and self-driving car development.

Bodepudi (2021) proposed a predictive approach for the detection of fraudulent activities using anomaly detection for unsupervised learning. This anomaly detection is also known as outlier detection which helps identify events and data points different from other normal events. The dataset utilized by the author was sourced from kaggle, which is an online repository for machine learning, credit card data. The author utilized three unsupervised machine learning techniques which are isolation forest, local outlier factor and one class SVM. The dataset was trained on the models using the data without any labels, and after the whole training process was conducted the accuracies of the three models; isolation forest, local outlier factor and one class SVM were 99.74%, 99.65% and 70.09% respectively. From this paper, the accuracy metric of performance was used because they were high but accuracy alone is not the best metric for unsupervised learning for model effectiveness and also due to dataset imbalance. The author didn’t state any limitation in the research, but the metric of performance would be a first step toward achieving the optimal outcome.

Mahesh et al. (2021) also used this approach in the international research journal for the detection of fraud using unsupervised machine learning techniques, the researchers utilized five (5) models which were unsupervised autoencoders neural network, isolation forest, local outlier factor and k means cluster. In this research, the input tested by the researchers was fed into the already trained machine learning algorithms hence, a good reason the accuracy was used as the metric of performance here. The accuracy obtained after testing, input given and prediction made for the neural network, auto encoders, isolation forest, local outlier factor and k-means cluster respectively were 99%, 97%, 98%, 98% and 99%. The authors didn’t state any limitations encountered during the research process in this paper but from the two papers above and the results obtained from the latter which shows signs of overfitting, as seen in the contrast between the accuracy, precision and recall of the models, unsupervised learning is not the best approach for fraudulent detection.

3.0. RESEARCH METHODOLOGY

In this section, the methodology used to develop a fraud detection system using machine learning techniques will be fully examined. The steps taken ensure the system's robustness while balancing performance and scalability will be discussed here.

3.1. DESIGN AND IMPLEMENTATION SPECIFICATIONS

The steps taken in this research to ensure its effectiveness are the problem definition, the collection of the appropriate data and its preprocessing, the model training and evaluation process and finally the analysis of the effective results.

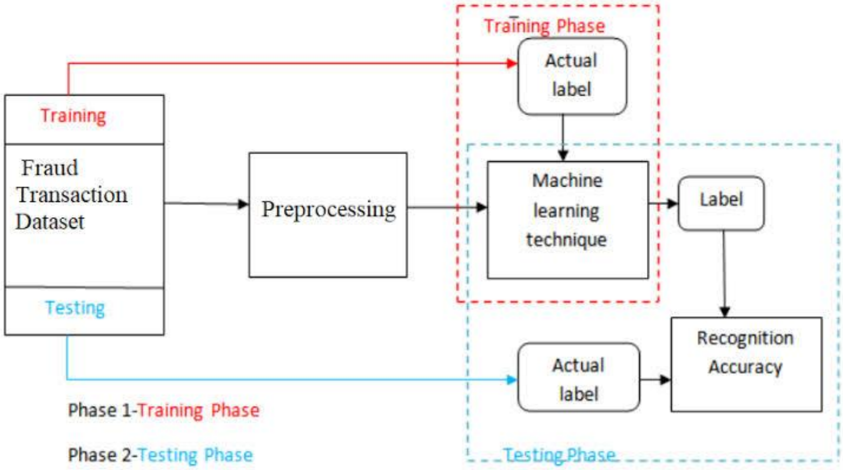

Figure 1: flowchart of design steps

3.1.1. THE DATASET

The Open Metaverse's blockchain financial transactions are included in this dataset obtained from kaggle, an online database for machine learning algorithms training, which aims to offer a realistic, varied, and rich set of data for fraud investigation, anomaly detection model development, and predictive analytics in virtual worlds. The dataset includes 78,600 records and 14 columns, each representing a metaverse transaction with the features of the dataset shown in the table below:

Table 1: dataset features description

S/N	COLUMN NAME	DESCRIPTION

1	Time stamp	Date and time of transaction
2	Hour of day	Hour part of the transaction timestamp
3	Sending address	Blockchain address of the sender
4	Receiving address	Blockchain address of the receiver
5	Amount	Transaction amount in a simulated currency
6	Transaction type	Categorization of the transaction
7	Location region	Simulated geographical region of the transaction
8	IP prefix	Simulated IP address prefix for the transaction
9	Login frequency	Frequency of login sessions by the user, varying by age group
10	Session duration	Duration of activity sessions in minutes
11	Purchase pattern	Behavioral pattern of purchases
12	Age group	Categorization of users into new, established and veteran based on their activity history
13	Risk score	Calculated risk score based on transaction characteristics and user behavior
14	Anomaly	Risk level assessment

3.2. MATERIALS AND TOOLS UTILIZED

- Hardware: NVIDIA RTX 4090 for deep learning
- Software and Libraries: Python v3.11 for system language. Pandas and numpy for data manipulation, matplotlib and seaborn for visualization, tensorflow, a deep learning framework for building neural networks.

3.3. DATA PREPROCESSING TECHNIQUES

In this research, the data preprocessing stage was done to ensure the data was clean and suitable for the development of machine learning and deep learning models in the research. The steps are outlined below:

- Missing data imputation: in this research, the dataset had no missing values but in scenarios where the information in some columns is missing, the continuous variable will be filled with the mean/median values while the categorical variables will be filled with the mode or dropped depending on the percentage of the missing data points.
- Normalization: the continuous variables were scaled using the normalizer as the StandardScaler and MinMaxScaler gave an overfitting result across all the models. The Normalizer is a normalization technique in machine learning that scales each data point independently by its L2 norm, making it suitable for where the magnitude of features of a dataset is important. Unlike the standard scaler and minmax scaler which adjusts the mean and standard deviation of features and scales features to a defined range, typically [0, 1] respectively, the Normalizer just focuses on transforming individual samples so that they each have a unit distribution, that is the length of the vector becomes 1. This scales the whole dataset so that an exploding value of, say, 1000 units can be represented within the range of 1 or 0 and 1.
- Encoding: the categorical variables in the research were encoded using the Label encoder.
- Feature selection: the redundant features in the dataset were filtered using the process of correlation analysis.
- Sampling: in this research, the sampling technique utilized to handle the data imbalance was the random over sampler which is a technique used to handle imbalanced datasets, where one class, mostly the minority class, is underrepresented compared to the other majority classes. It tackles this issue by randomly duplicating examples from the minority class until the number of examples matches or approaches that of the majority class enabling the algorithm to learn from a more representative distribution of the classes. The reason for not using SMOTE is that it introduces synthetic data samples which would not be good for model training.
- Data splitting: the dataset was split into training, validation and testing sets in the ratio 60:20:20 to ensure the model evaluation is unbiased.

3.4. STATISTICAL TECHNIQUES (MACHINE LEARNING MODELS)

A range of models will be discussed such as logistic regression, random forest, gradient boosting, xgboost, decision trees, support vector machine and the artificial neural network.

3.4.1 LOGISTIC REGRESSION

A logistic regression model is used to determine the likelihood that an event will occur given a collection of independent data components (Satterstrom et al., 2020). Logistic regression is typically used when a classification task has a categorical outcome, which is one of only two outcomes (Kowsari et al., 2019). This model was used because of its effectiveness in binary classification.

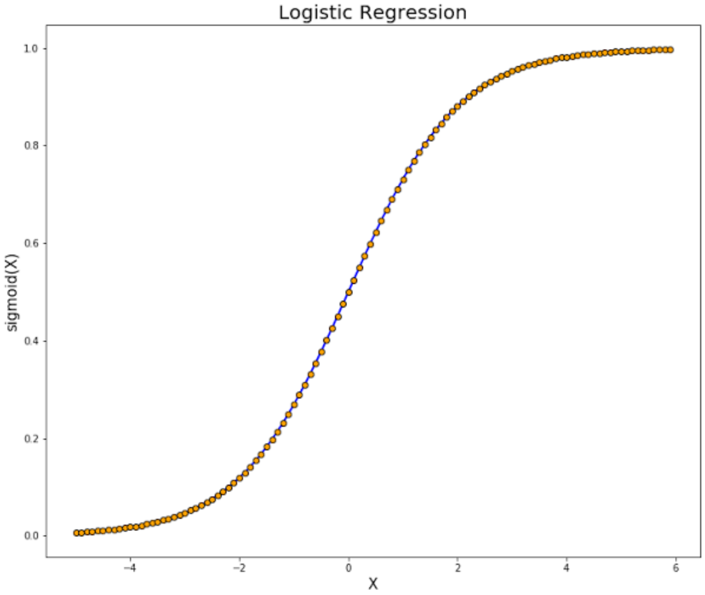

Figure 3: logistic regression (Thorn, 2022)

3.4.2 RANDOM FOREST

The Random Forest is a popular approach to machine learning that may be applied to both classification and regression (Maddikunta et al., 2020). Random forests are known for their scalability, durability, and ability to handle multidimensional data with complex relationships (Gharehchopogh et al., 2023). They also provide a feature relevance rating that makes feature selection easier. Among the model's advantages are its high accuracy, low overfitting, ability to manage missing data effectively, adaptability to outliers, and flexibility (Tsalikidis et al., 2023). While there are advantages and disadvantages to random forests, their primary advantages are their ability to handle irrelevant variables appropriately, handle large datasets, perform well in both regression and classification tasks, are robust to outliers, and effectively handle missing values (Razzak et al., 2019). In fraud detection cases, this model captures complex patterns by aggregating multiple decision patterns making it suitable for handling imbalances.

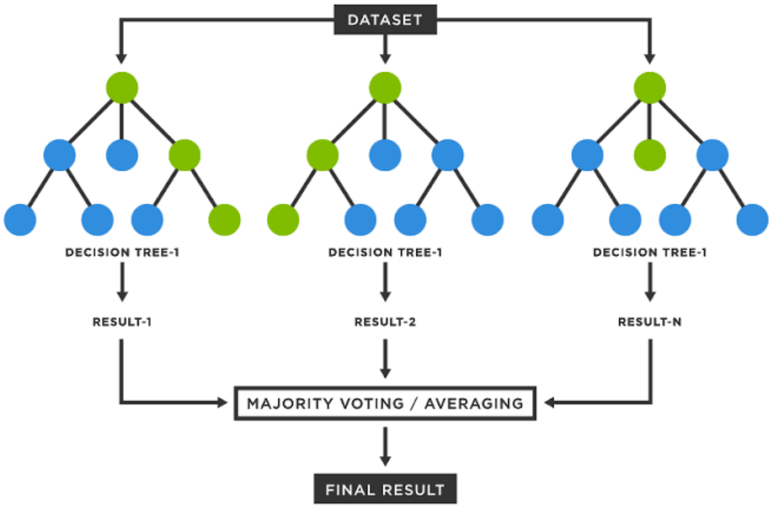

Figure 4: random forest (Kharkar, 2023)

3.4.3 GRADIENT BOOSTING

The gradient boosting works well for both classification and regression analysis because it constructs an ensemble of numerous decision trees sequentially (Çınar et al., 2020). In the research, this model was introduced from the sklearn library and the train set was fitted into the model using the fit function. Gradient boosting is a boosting strategy that iteratively learns from each weak learner to create a strong model. (Johnson & Khoshgoftaar, 2019). In fraud detection, it’s used to capture subtle patterns in imbalanced datasets. The reason for choosing this ensemble model over others is because of its popularity after the xgboost.

3.4.4 XGBOOST

The xgboost is a more advanced and portable variant of the previously stated gradient boosting classifier. This extreme boosting technique optimizes speed and performance using parallel processing and optimization techniques (Shorten & Khoshgoftaar, 2019). The xgboost library was installed as a stand alone library as it is not a model in the sklearn library. XGBoost is well renowned for its computational efficiency; it offers efficient processing, accurate feature importance analysis, and seamless handling of missing information. In fraudulent detection cases, this model reduces both bias and variance as it builds sequences of trees and optimizes errors in each step. The reason for choosing this ensemble model over others is because of its popularity.

3.4.5 DECISION TREES

A decision tree is a flowchart that resembles a branching tree where each node in the tree represents a test for a certain property, such as the result of flipping a coin, which could be heads or tails. Every branch shows the test result, and every leaf node stands for the class label. In a decision tree, the categorization scheme is represented by the path from the root to the leaf (Charbuty & Abdulazeez, 2021). As a result of this approach, a decision tree is a classifier that is represented as a recursive split. A rooted tree, which is a directed tree with a node known as the root that has no edges entering, is created when nodes from the decision tree join (Marjanović & Laurin, 2019). This model is easily interpretable which makes it easier to trace why certain transactions are classified as fraudulent or not.

3.4.6 SUPPORT VECTOR MACHINE

Primarily designed for classification purposes, the support vector machines are supervised max-margin models (Koroniotis et al., 2019). In the high-dimensional space, the support vector model separates data points of various classes into distinct groups by identifying the hyperplane that best divides them (Cervantes et al., 2020). This model, being a nonlinear model like fraudulent activity, maps data to a high dimension. Because of its boundary conditions, it has the ability to find the linear boundary between fraudulent and non fraudulent transactions.

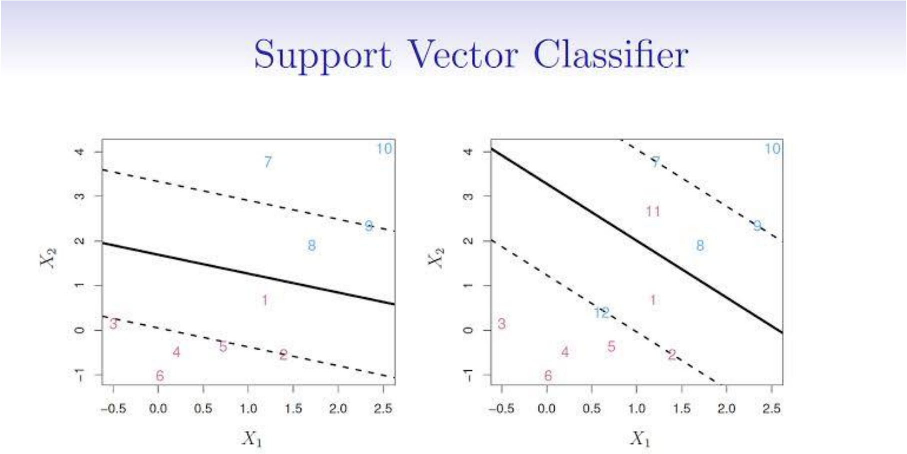

Figure 5: support vector machine (Tandel, 2018)

3.4.7 ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is a computational model modeled after the biological neural networks seen in the human brain (Chakraborty et al., 2021). By adjusting parameters according to input and output relationships, it allows systems to learn from data and is an essential part of machine learning and deep learning (A. Khan et al., 2020). This model is very suited for fraudulent transactions detection because of its ability to learn complex nonlinear relationships in a dataset. In this research, the ANN was implemented using the tensorflow framework.

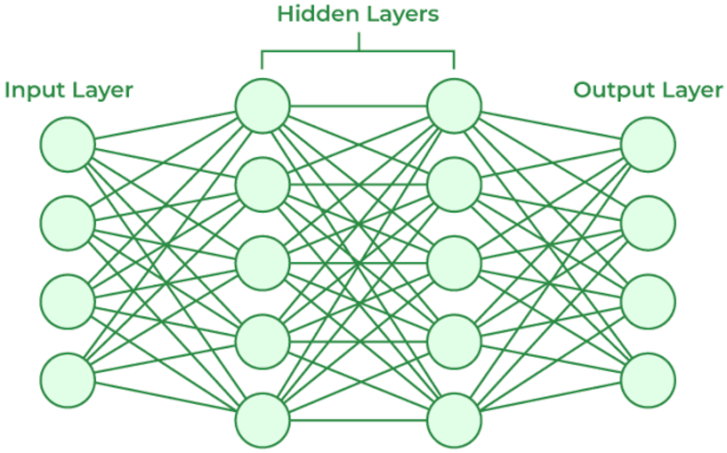

Figure 6: the artificial neural network (GeeksforGeeks, 2024)

3.5. CHALLENGES FACED DURING MODEL IMPLEMENTATION

- In this research, the dataset utilized was from kaggle and that being the case, it would have some statistically standardized figures which would disrupt some vital preprocessing steps such as data normalization as seen in this research. Standardizing the data using the standard scaler resulted in the overfitting of all the models which led to the usage of a milder scaling technique, the normalizer.
- As we know, in these kinds of problems, there are usually cases of data imbalance so utilizing balancing techniques usually introduces some false figures which may affect the model performance.
- Lack of robust dataset which shows the real time fraudulent activities. Because fraudulent activities are changing with time, it is important to use the current data from institutions which we didn't have access to.

4.0. EVALUATION

In this research, six different machine learning algorithms were used in contrast to a deep learning multi layered perceptron for the detection and classification of fraudulent activity online. In this research, accuracy was the main metric of performance because that's what the previous researchers used in the literature review, so it will be used as the yardstick of performance. The results of these models are displayed below:

MODEL	ACCURACY	PRECISION (0)	PRECISION (1)	PRECISION (2)
Logistic regression	79%	80%	96%	34%
Random forest	99.97%	100%	100%	100%
Gradient boosting	99.85%	100%	100%	99%
Xgboost	99.85%	99%	100%	100%
Decision trees	99.79%	100%	100%	99%
Support vector machine	85%	79%	97%	45%
Multi-layer perceptron	98.29%	98.9%	98.9%	98.9%

Table 4.1: Accuracy and Precision of the seven models used in the comparative analysis

MODEL	RECALL (0)	RECALL (1)	RECALL (2)	F1 SCORE (1)	F1 SCORE (2)	F1 SCORE (3)
Logistic	100%	78%	76%	89%	86%	47%

regression						
Random forest	100%	100%	100%	100%	100%	100%
Gradient boosting	100%	100%	100%	100%	100%	99%
Xgboost	100%	100%	100%	99%	100%	100%
Decision trees	99%	100%	99%	99%	100%	99%
Support vector machine	100%	84%	81%	88%	90%	58%
Multi-layer perceptron	95.6%	95.6%	95.6%	97%	97%	97%

Table 4.2: Recall and F1 score of the seven models used in the comparative analysis

MODEL	CPU USAGE (%)	TIME USAGE(sec)
Logistic regression	42.13	4
Random forest	38.18	56.49
Gradient boosting	33.18	123.4
Xgboost	100.0	13.82
Decision trees	41.23	0.86
Support vector machine	39.51	2888.11
Multi-layer perceptron	56.13	69.12

Table 4.3: Time and CPU usage of the seven models used in the comparative analysis

Table 4.1 shows the result from the comparative analysis of the different models for a classification task, focusing on their accuracy and precision across three classes (0, 1, and 2). From the table, Random Forest, Xgboost and Gradient Boosting are the best models, with nearly perfect accuracy and precision across all classes demonstrating excellent performance in handling this dataset. The Decision Tree model achieved high accuracy and similarly high precision across classes. The Multi-Layer Perceptron (MLP), with its accuracy and macro averaged precision, indicates a strong performance but not reaching the precision consistency of the top models. In contrast, Support Vector Machine (SVM) and Logistic Regression underperformed relative to the other models with inconsistent precisions, suggesting limitations in differentiating this class effectively.

In this research, the Random Forest and ensemble models are highly reliable for this classification task, while MLP provides a competitive alternative with slightly lower performance.

Table 4.2 also shows the result from the comparative analysis of the different models for the task, focusing on the recall and f1 score across three classes (0, 1, and 2) as fraud detection prioritizes these metrics. The SVM and Logistic Regression models show varying effectiveness across different recall and F1 scores. SVM performs well for all the recall classes which indicates its ability to correctly identify non-fraudulent and fraudulent cases to some extent, though it could miss more cases than other models like Random Forest or XGBoost. Its F1 scores, which balance precision and recall, show moderate performance and a relatively low score for the last class. Logistic Regression also achieves a perfect recall on the first class but drops for the next two classes, suggesting it might miss more fraudulent cases. Its F1 scores are fairly balanced but lower overall, making it potentially less effective for identifying all fraudulent cases compared to SVM and the ensemble methods listed.

Table 4.3 shows the CPU and time usage for the models. In this research, the time and CPU usage was calculated using the time and psutil libraries. From the result, the decision trees had the lowest training time while the gradient boosting had the least CPU usage. Now , the trade-off between the accuracy and the model training will be discussed. The tradeoff between accuracy and computational efficiency in machine learning refers to the balance between achieving high model performance and minimizing the computational resources required for training and inference. Heavy models, like random forest and neural networks often require extensive computational power, memory, and time to process large datasets and complex patterns. On the other hand, simpler models like linear regression or decision trees are computationally efficient but may lack the capacity to achieve high accuracy on intricate tasks.

4.1. DISCUSSION

In this study, the detection and classification of fraudulent activities online was proposed. In this section, the results obtained from the research will be compared with that of other existing literature from the reviewed section.

It can be seen from the previous literature that most of the models would have performed better if there was a more robust and balanced dataset, so in our research we tackled that problem using the random over sampler which made sure the dataset utilized was balanced by choosing features from the majority and minority classes mitigating accuracy bias. Also, Raghavan et al. (2019) observed that in their research, the random forest, SVM and CNN were the best performing models. We can also see that in our research, we improved on their accuracy and the random forest, and the ensembles were the best performing models. Dornadula et al. (2019) also took note of the SVM's computational issues while training with the SMOTE balanced data which was almost encountered in our research as we observed the SVM to be the machine learning model with the highest training time with lesser accuracy and performance. But being able to produce a result, we can conclude that we have improved on the computational inefficiency of the SVM by using a better sampling technique. Finally, we can see that for our deep learning model, we have better accuracy compared to the deep learning models proposed in the reviewed literature so we can conclude by saying we have achieved an improved system for the detection of fraudulent activities online. In a real world situation, models like the logistic regression with a lower training time might be effective because when deploying a model, model computation might be required to be efficient hence, a less time consuming model is required. Also, data balancing played an important role in the accuracy of the best models because there was an equal distribution of data points because accuracy is sensitive to data imbalance and this increases the model generalizability. In the SVM, the model training time was so prolonged but yielded lower accuracy because the model may not be suitable for the task, as we have seen other models perform well meaning the data is viable.

5.0. CONCLUSION AND FUTURE WORK

In this research, seven machine learning models were developed based on a comparative analysis approach with a dataset obtained from kaggle to detect fraudulent transactions and activities in e-commerce. After the model training and testing, through a comparative analysis of various algorithms, it is evident that the random forest model which is an ensemble model can significantly enhance the accuracy and efficiency of fraud detection being the best model and the support vector machine (SVM) was observed to be the worst performing model.

Prior to the research, the multi layer perceptron, which is a deep learning model was thought to perform well being a deep learning model but due to data distribution, the random forest outperformed the deep learning. The limitation of this research can be seen in the context of the dataset, as to be more precise in the detection of real time fraud, data privacy concerns pose an issue.

Future work should consider incorporating more sophisticated techniques and real time scraped data, such as a more complex deep learning system to further improve detection rates. Additionally, implementing explainable AI techniques can increase transparency, helping stakeholders understand the decision-making process and build trust in automated fraud detection systems. In a real world scenario, institutions can apply this fraud detection system by gathering data or information from customers, say, a banking system, which is further cleaned and feature engineered to scale the data, then a machine learning model is chosen and the data is trained on. This final model is then deployed and a threshold is set for flagging suspicious transactions.

REFERENCES

Akyildiz, I. F., Kak, A., & Nie, S. (2020). 6G and Beyond: The Future of Wireless Communications Systems. *IEEE Access*, *8*, 133995–134030.

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, 39700-39715.

Almarshad, F. A., Gashgari, G. A., & Alzahrani, A. I. A. (2023). Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset. *IEEE Access*, *11*, 107348–107368.

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-Dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, *10*(1).

Azam, H., Dulloo, M. I., Majeed, M. H., Wan, J. P. H., Xin, L. T., & Sindiramutty, S. R. (2023). Cybercrime Unmasked: Investigating cases and digital evidence. *International Journal of Emerging Multidisciplinaries Computer Science & Artificial Intelligence*, *2*(1).

Bhasin, M. L. (2016). Challenge of mitigating bank frauds by judicious mix of technology: Experience of a developing country. *Economics Management and Sustainability*, *1*(1), 23–41.

Bodepudi, H. (2021). Credit card fraud detection using unsupervised machine learning algorithms. Int J Comput Trends Technol, 69, 1-13.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215.

Chakraborty, A., Alam, M., Dey, V., Chattopadhyay, A., & Mukhopadhyay, D. (2021). A survey on adversarial attacks and defences. *CAAI Transactions on Intelligence Technology*, *6*(1), 25–45.

Chami, R., Sharma, S., & Fullenkamp, C. (2009). A framework for financial market development. *IMF Working Paper*, *09*(156), 1.

Charbuty, B., & Abdulazeez, A. (2021b). Classification based on decision tree algorithms for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28.

Çınar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability*, *12*(19), 8211.

Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. Procedia computer science, 165, 631-641.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.

GeeksforGeeks. (2024, August 7). *Artificial Neural Networks and its Applications*. GeeksforGeeks.

Gharehchopogh, F. S., Ucan, A., Ibrikci, T., Arasteh, B., & Isik, G. (2023). Slime Mould Algorithm: A comprehensive survey of its variants and applications. *Archives of Computational Methods in Engineering*, *30*(4), 2683–2723.

Gulati, A., Dubey, P., MdFuzail, C., Norman, J., & Mangayarkarasi, R. (2017). Credit card fraud detection using neural network and geolocation. *IOP Conference Series Materials Science and Engineering*, *263*, 042039.

Hussain, F., Hussain, R., Hassan, S. A., & Hossain, E. (2020). Machine learning in IoT Security: current solutions and future challenges. *IEEE Communications Surveys & Tutorials*, *22*(3), 1686–1721.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1).

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. a. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.

Khan, A., & Malaika, M. (2021). Central bank risk management, fintech, and cybersecurity. *SSRN Electronic Journal*.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020b). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*(8), 5455–5516.

Kharkar, D. (2023, July 9). About Random Forest Algorithms. - Dishant kharkar - Medium. *Medium*.

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems*, *100*, 779–796.

Kulshrestha, P. (2022). *Cyber Crime, Regulations and Security - contemporary issues and challenges*.

Maddikunta, P. K. R., Srivastava, G., Gadekallu, T. R., Deepa, N., & Boopathy, P. (2020). Predictive model for battery life in IoT networks. *IET Intelligent Transport Systems*, *14*(11), 1388–1395.

Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2017). Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks*, *4*(3), 161–175.

Mahesh, V. U., Reddy, S. K., Reddy, L., Krishna, S., Dilleshwar, J., & Kaur, S. (2021). Credit Card Fraud Detection using Unsupervised Machine Learning. International Research Journal of Engineering and Technology (IRJET) Vol, 8.

Marjanović, D., & Laurin, M. (2019). Phylogeny of Paleozoic limbed vertebrates reassessed through revision and expansion of the largest published relevant data matrix. *PeerJ*, *6*, e5565.

Mehta, P., Bukov, M., Wang, C., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to Machine Learning for physicists. *Physics Reports*, *810*, 1–124.

Mutemi, A., & Bacao, F. (2024). E-Commerce fraud detection Based on Machine Learning Techniques: Systematic Literature review. *Big Data Mining and Analytics*, *7*(2), 419–444.

Pirandola, S., Andersen, U. L., Banchi, L., Berta, M., Bunandar, D., Colbeck, R., Englund, D., Gehring, T., Lupo, C., Ottaviani, C., Pereira, J. L., Razavi, M., Shaari, J. S., Tomamichel, M., Usenko, V. C., Vallone, G., Villoresi, P., & Wallden, P. (2020). Advances in quantum cryptography. *Advances in Optics and Photonics*, *12*(4), 1012.

Pumsirirat, A., & Liu, Y. (2018). Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. International Journal of advanced computer science and applications, 9(1).

Quah, J. T. S., & Sriganesh, M. (2008). Real-time credit card fraud detection using computational intelligence. *Expert Systems With Applications*, *35*(4), 1721–1732.

Raghavan, P., & El Gayar, N. (2019, December). Fraud detection using machine learning and deep learning. In 2019 international conference on computational intelligence and knowledge economy (ICCIKE) (pp. 334-339). IEEE.

Razzak, M. I., Imran, M., & Xu, G. (2019). Big data analytics for preventive medicine. *Neural*

Computing and Applications, *32*(9), 4417–4451.

Sailio, M., Latvala, O., & Szanto, A. (2020). Cyber threat actors for the factory of the future. *Applied Sciences*, *10*(12), 4334.

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., . . . Minshew, N. (2020). Large-Scale Exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, *180*(3), 568-584.e23.

Shojaifar, A. (2023). *Volitional cybersecurity*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1).

Shrestha, A., & Mahmood, A. (2019). Review of Deep learning Algorithms and Architectures. *IEEE Access*, *7*, 53040–53065.

Tandel, A. (2018, June 13). Support vector machines — a brief overview - towards data science. *Medium*.

Thorn, J. (2022, August 30). Logistic regression explained - towards data science. *Medium*.

Tsalikidis, N., Mystakidis, A., Tjortjis, C., Koukaras, P., & Ioannidis, D. (2023). Energy load forecasting: one-step ahead hybrid model utilizing ensembling. *Computing*, *106*(1), 241–273.

Ul, B., F, R., Mehraj, A., Ahmad, A., & Assad, S. (2017). A compendious study of online payment systems: past developments, present impact, and future considerations. *International Journal of Advanced Computer Science and Applications*, *8*(5).

Wu, Y., Xu, Y., & Li, J. (2019). Feature construction for fraudulent credit card cash-out detection. *Decision Support Systems*, *127*, 113155.