Financial Fraud Predictions in E-Commerce using Machine Learning and Deep Learning Models

A Thesis Submitted
In Partial Fulfillment of the Requirements for the degree of

MASTER OF TECHNOLOGY
IN
Data Science

Submitted by

Naman Jain

(2K23/DSC/22)

Under the Supervision of

Ms. Shweta Meena

Assistant Professor, Department of Software Engineering

Delhi Technological University

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana Road, Delhi
110042

MAY, 2025

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana Road, Delhi
110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to Ms Shweta Meena for his continuous guidance and mentorship that he provided us during the Thesis. She showed us the path to achieve our targets by explaining all the tasks to be done and ex- plained to us the importance of this project as well as its industrial relevance. She was always ready to help us and clear our doubts regarding any hurdles in this project. Without her constant support and motivation, this thesis would not have been successful.

Place: Delhi
Date: 20.05.2025

Naman Jain
(2K23/DSC/22)

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering) Bawana Road, Delhi
110042

CANDIDATE'S DECLARATION

I, Naman Jain, Roll No. – 2K23/DSC/22 students of M.Tech (Data Science), hereby certify that the work which is being presented in the thesis entitled "Financial Fraud Predictions in E-Commerce using Machine Learning and Deep Learning Models" in partial fulfilment of the requirements for the award of degree of Master of Technology, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from Jan 2025 to May 2025 under the supervision of Ms. Shweta Meena. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor (s)

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering) Bawana Road, Delhi 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Financial Fraud Predictions in E-Commerce using Machine Learning and Deep Learning Models" which is submitted by Naman Jain, Roll No – 2K23/DSC/22, Department of Software Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the thesis work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Ms. Shweta Meena
Assistant Professor
Department of Software Engineering, DTU

Date: 20.05.2025

ABSTRACT

Nowadays it is often said that so many money related transactions are done using online services, day to day fraud has become a major issue for everyone using E-commerce services. Online fraudsters now focus on financial transactions, as old security measures cannot detect the more advanced ways they commit fraud. This paper examines the evolving issues related to E-commerce fraud through machine learning (ML) and deep learning (DL). Today, many shopping-related transactions are taken care of by online sites which has led to an increase in daily fraud faced by people making such purchases. Intelligent fraudsters now aim for financial transactions, as the older ways of detecting fraud cannot identify them. With ML and DL, this paper studies the developing trends in E-commerce fraud. The goal is to introduce flexible approaches to better detect financial frauds in real time. Over 20,000 transactions in E-commerce were used for the research because they appeared both imbalanced and unreliable. I made the training successful by first oversampling (SMOTE), undersampling the data and analyzing it using box plots to remove any outliers. For training and testing the models, six frameworks chosen are Random Forest, AdaBoost, CatBoost, XGBoost, Long Short-Term Memory and Gated Recurrent Unit. They were picked because they have managed to identify frauds in the past, mostly thanks to their ability to be decisive and observe the data accurately. It covers each phase of a modeling project, mainly focusing on handling dirty data, selecting the best features, selecting an appropriate model and measuring its accuracy, precision, recall and F1-score and area under the AUC-ROC curve. Even though Random Forest outperformed the other models regarding classifying, I find that the others are just as trustworthy. For this thesis study, we finish by addressing the main problems such as ensuring data balance, improving frameworks and introducing multiple ways to boost AI by making detection of fraudulent activities easier, letting AI explain itself and creating mixing models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In the forthcoming digital era, online E-transactions have become a magnificently essential part of day-to-day life, with credit and debit cards transactions often called "Plastic Money". An unauthorized individual can take your important data and make a purchase for their benefit through E-commerce [1]. Growth in Online Shopping is creating more E-transactions on the web and this has resulted in more frauds occurring in online shopping. Those who engage in fraud make the most of all the minor shortcomings within E-transactions to let their E-commerce deals be fraudulent. Every year, a considerable amount of sellers' and producers' money is used to cover payment card fraud [2]. Therefore, this paper focuses on how fraud can be detected by combining large scale frameworks with machine learning and deep learning technologies, using these, like random forest and LSTM, will determine if E-transactions in future will result in a fraud or are safe and can be used legitimately [3].

In this topic and research of E-commerce frauds which are related to hacking payments services provided by multiple vendors, it's has become crucial to know that criminals who do cyber fraud prepares out multiple attacks, including Botnets that lead to distributed denial-of-service attacks, Bricking things, sending excessive spam, cryptocurrency mining, cryptojacking, phishing and others. Furthermore, these specific cyber-attacks are carried out to steal data, mine cryptocurrency, deliver spam messages and grant the attacker access to secret data stored on the user's device and in its entire network of devices [4].

Data mining is widely applied to handle frauds that occur in E-transactions for E-commerce. A system should use Card or Plastic Money related fraud detection to identify a real or a fake transaction. A person's actions and spending through his card assist in detecting fraud cases. The issue here is that fraud-related transactions can appear like the real thing and finding the card databases is a tough task. If the data we have comes from a dataset, it could be highly imbalanced. Because performance in fraud detection services depends on the available variables, methods such as over and under sampling, also SMOTE are used [5].

With each going day, there is a noticeable rise in E-transactions and this will at some point make it hard to identify E-frauds. Machine learning and deep learning function on the basis of data in a certain area to label information that will come in the future. The purpose was to work on the issue of imbalanced classes. Instead of overcoming the problem, we can use machine learning platforms to properly sort through the current data. Cybercriminals handle this by cloning real cards with card proxy and reproducing fraudulent websites to fool trusted users [6].

1.1 Problem Statement of Dissertation

As we mentioned in the start of this work, the rise of e-commerce has made it simpler and cheaper for consumers everywhere to handle their money [7]. As more people use digital methods to pay, there has been an increase in cyber fraud when paying online. The small losses that result from these kinds of fraud have also weakened customers' confidence in using online financial services. Due to the quick growth of fraudulent acts, the original rule-based systems are slow in catching them, so predictive notices are being developed using new techniques. They are focused on coming up with a system that can efficiently detect cyber frauds by using machine learning and deep learning methods[8].

The research study which we are talking about in this paper applies on a 20,000 e-transaction dataset with atmost 31 features, from which the target variable "Class" column is binary which means it holds only 2 values which are '0' and '1'. Here, '0' denotes a legit or right transaction which does not lead to any fraud and '1' for a fraud related transaction. To begin with, our dataset was very unbalanced, meaning that it favored some frameworks over others, so it was useless for our study. Later on, the training data was both under-sampled and over-sampled. Once that was done, several predictive frameworks were applied to control the high learning rate and prevent biased classification of classes. For higher accuracy in finding fraud, you can use multiple AI frameworks found in ensemble learning, including Random Forest, XGBoost, CatBoost and AdaBoost. Long Short-Term Memory and Gated Recurrent Units are examples of deep learning methods that are used to analyze long transactions and see any patterns within them. In this research paper, both methodologies will be evaluated for precision, recall, F1-score and in general how well they predict. Insight from machine learning methods will be used to build a system that accurately detects and identifies fraudulent e-transactions as they happen [9].

1.2 Overview of the research objectives of the Dissertation

The primary goal of this research study is to develop a highly accuracy oriented and resilient fraud detection system for e-commerce related e-transactions based on machine learning and deep learning frameworks. In this research, the following goals are fulfilled:

Initially, we will detect fraudulent e-transactions by using strong ensemble techniques such as Random Forest, XGBoost, CatBoost and AdaBoost and afterward, we will employ learning models such as LSTM and GRU deep learning. Next, After collecting a dataset of 20,000 records and 31 features, I preprocessed and explored the data to balance and maintain its quality.

Later on this topic we will definitely try to discuss multiple ways to enhance and improve the model's efficiency and performance by mutually handling this extra ordinary issue of data being not equal and finding the absolutely right corresponding significant values for hyperparameters which will give good accuracy. Currently, people who are studying how to be valuable it would be used as to provide the usage of deep learning frameworks at place of ensemble machine learning to detect frauds.

The research will later concentrate on enhancing existing e-commerce security to manage fraud using a scalable and reliable model.

1.3 Types of Frauds in E-Commerce

The media and the press often report on a wide variety of banking frauds. Below are a few known examples:

1. **Cheque frauds:** In this year 2025, cheque related frauds has become very common, usually occur when a unknown person knowingly gives a fraud check to you to cash in without having required cash amount in his bank account or if somebody unknown steals another person's identity theft money. As per the article dated July 30, 2019, people in Lucknow tricked several entrepreneurs through false claims about their startup. They bought laptops by putting in fake check payments, but once their money in the accounts was checked, the bank rejected them.

2. **Online shopping scams:** From the past one decade, these online shopping scams involve money Because of these scams throughout the past decade, people swindled online who paid for cell phones or similar items that turned out not real and the seller failed to send the delivered products because it was a fake website. As an example, a doctor was taken to court for defrauding ₹2.62 lakh online on August 29, 2019. Before receiving the call about his laptop, he had spent ₹399 on a laundry bag. After giving his confidential details and making a GST payment of ₹5,580, he was scammed again.

3. **Insurance fraud:** Insurance fraud these days have been building massively in metro cities, which involve in submitting of false claims to insurance for losses in property, treatment costs for illness, or damage done to the vehicles. In metro cities, insurance fraud is increasing. Example: In a briefing made public on July 9th, 2019, it

was revealed that ten individuals, one a lawyer and one a doctor, were detained because they helped set up a scheme to secure insurance for patients with serious illnesses for when they could have needed regular care. Patients who died during treatment were always told by the doctor that it was an accident to control things with the insurance company.

4. **Work from home scams:** Criminals offer home work opportunities to people and tell them to pay a large sum of money. Fraudsters make it appear easy for the victims by advertising different courses to help them earn large sums of money.

5. **Credit card or debit card frauds:** The card related frauds these days has been on the highest highs, Card related frauds includes the illegal and unwanted usage of an person's card or card's details which is from india to make out forged or fake transactions or buys which are of very high amount that too out of india, so that they can not be caught. For example: The Delhi Police senior officer, Atul Katiyar, became a victim of credit card fraud on August 9, 2019. At this point, his device received a text message from the bank. It informed him that he has won digital points on his debit card and to claim them, he needs to provide certain info such as his account number, cvv code and so on. Later, he faced the challenge of losing ₹28,000 [9].

CHAPTER 2

RELATED WORK

In very recent fraud publications, credit card related essential fraud detection in e-commerce buisness has been studied in detail using multiple ML and DL approaches for detection. These multiple ways have ultimate focused on overcoming key and various obstacles in fraud detection, including too few examples of certain types of fraud, spotting distinctive fraud activities and making the process work on time. Even with improvements, it is still a challenge to identify credit card fraud since new scams appear regularly, models have to be easy to interpret and production systems must be accurate and not produce many false positives.

They introduced and later suggested a neural network classifier based on blending ensemble learning and combining it with several data resampling approaches. The main goal of this sampling process in this thesis was to address an important situation where the large number of true transactions which are legit is much more than the number of false ones which were resultant as fraud[13]. This method highly depended on using an algorithm called the LSTM neural network as the foundation for an AdaBoost model. For that reason, LSTM models work well with data flowing over time such as in the analysis of financial transactions. Because of AdaBoost, the LSTM model was able to learn from the mistakes made on misclassified data. Authors who have written previous papers have found that the ensemble models have performed better on classification problems as accuracy compared to traditional models such as Decision Trees, SVM and MLP [11]. Based on the experiments were played a significant role, the work revealed how the process of mixing recurrent and thundering ensemble networks of large neurons helps achieve better results of work and more robustness of the models.

Likewise, Mienye and Sun proposed a smart system that combined LSTM and GRU models to help improve how fraud is detected [15]. Although GRU has fewer gates than LSTM, it still manages to recognize long-term connections in a sequence with less calculation. By combining LSTM and GRU models, the ensemble model gained the advantages of each architecture. To correct the unequal numbers in the data, they decided to use SMOTE to manufacture samples of the rare class and include these in the training set. Because of their method, the rate of detecting frauds increased while the rate of false alarms dropped, both essential for fraud detection. Researchers observed that these AI models greatly outperformed Random Forests and XGBoost when the possible consequences of error are huge.

In another paper, Khalid et al. pointed out that ensemble learning works well, especially when you combine different classifiers such as SVM, KNN, Random Forests and major boosting algorithms [13]. Using many different classifiers allowed them to detect a wider variety of objects. By using under-sampling and SMOTE, they were able to balance the data used for training the classifiers. Not this was one only study which was the basis learning significantly improved, yes but good this however method developed also a brought corresponding the optimum risk welcoming of overfitting counterfeit which would is a significant in financial fraud detection, down. Collaborative model strategies were proved to be effective because they achieved better accuracy, precision and recall compared to using one classifier alone.

Randhawa and his co-authors took part by looking into hybrid ensemble models. They compared how ensemble forests work compared to Decision Trees, Logistic Regression and Naïve Bayes [12]. According to them, methods that use majority voting proved to be more successful in detecting fraud in online shopping. This is consistent with the usual belief that learning by ensembles help stabilize predictions and lower the risk of making errors. Pulling data from a mix of learners, majority voting reduces the biases in every model which leads to better results. The authors demonstrated that ensemble methods are effective on large financial datasets.

Another study by Zhang et al. added advanced feature engineering tools to deep learning for better detection of fraud [14]. Featuring data is crucial in machine learning since raw financial data may not immediately expose signs of fraud. Zhang et al.'s approach focused on extracting temporal and behavioral features from transaction logs and integrating them into deep learning models such as convolutional neural networks (CNNs) and LSTMs. Then, the models were taught to spot certain patterns that indicate fraud. The system performed very well and was reliable because of its impressive features and advanced DL models.

The current study improves on what has come before by suggesting a combination of ensemble learning and deep learning. To meet our objective, the framework uses models such as Random Forests, XGBoost, CatBoost, AdaBoost, LSTM and GRU. Blending different algorithms in an ensemble approach helps with learning more accurately and consistently. In addition, deep learning models work with temporal and sequential data to help identify potential frauds in payment records.

The current study improves on the previous models by proposing to combine deep learning with ensemble learning. We plan to achieve the right balance between accuracy, speed and applying the model fast by making use of Random Forests, XGBoost, CatBoost, AdaBoost, LSTM and GRU. The goal of using ensemble methods is to increase accuracy, whereas deep learning models help find critical links between transactions to prevent fraud from occurring.

In summary, the related work in this domain has laid a strong foundation for the development of advanced fraud detection systems. The use of ensemble learning and deep learning techniques has proven to be highly effective in addressing key challenges such as class imbalance, detection precision, and model robustness. However, the need for real-time detection, model interpretability, and adaptability to concept drift continues to drive research in this area. Our study contributes to this evolving field by integrating and expanding upon these techniques to develop a more accurate, responsive, and reliable fraud detection system tailored for the complexities of e-commerce transactions.

CHAPTER 3

RESEARCH METHODOLOGY

In the domain of financial fraud detection, especially within the fast-paced and high-volume environment of e-commerce, traditional rule-based systems often fall short due to their rigidity and inability to adapt to evolving fraud patterns. This necessitates the use of more intelligent, adaptable, and scalable approaches—hence the integration of Machine Learning (ML) and Deep Learning (DL) algorithms in the current study. ML algorithms are capable of learning from historical transaction data and detecting underlying patterns that distinguish fraudulent behavior from legitimate ones, even in highly imbalanced datasets where genuine transactions vastly outnumber fraudulent ones. Techniques such as Random Forest, XGBoost, CatBoost, and AdaBoost offer ensemble-based learning, which improves prediction accuracy by combining the strengths of multiple base learners and mitigating individual model biases. On the other hand, Deep Learning models, particularly Recurrent Neural Networks (RNNs) like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), excel in capturing sequential and temporal patterns in transaction data—an essential characteristic when dealing with time-stamped financial records that exhibit evolving user behavior and transaction flows.

Moreover, ML and DL models offer the significant advantage of adaptability. Unlike static rule-based systems, these algorithms can be retrained and updated as new types of fraud emerge, making them highly relevant in a dynamic threat landscape. The use of synthetic sampling techniques like SMOTE alongside these models also helps in addressing class imbalance issues effectively, thereby improving the model's ability to detect rare fraudulent instances without being overwhelmed by the majority class. Additionally, deep learning frameworks are highly effective in automatically extracting complex features from raw data, reducing the need for extensive manual feature engineering and enabling end-to-end learning. In the context of this research, the combination of ML and DL allows us to leverage both structured patterns and hidden nonlinear relationships in transaction data, resulting in more robust, scalable, and accurate fraud prediction systems. This methodology not only enhances detection rates but also minimizes false positives—crucial in maintaining customer trust and reducing operational costs in real-world e-commerce platforms. Therefore, the use of machine learning and deep learning algorithms is not just beneficial but imperative for developing a fraud detection system that is responsive, intelligent, and capable of evolving alongside emerging threats in the digital financial ecosystem…

3.1 Overview of the Models

Here we present the ML and DL algorithms used in our study of cyber security fraud prediction Advance Machine Learning and Deep Learning frameworks-based predictions for E-commerce Transactions.

1. **Random Forest:** Random Forest is a kind of ensemble learning that trains a huge number of trees at the start and combines their estimations to ensure accuracy. Although one decision tree will be susceptible to high variance, Random Forest generates a "forest" of trees, where every tree is trained on a random subset of the data. Overall, it relies on the most common solution for GP problems and can successfully address noisy data, working well with lots of dimensions.

2. **AdaBoost (Ensemble Learning):** is another ensemble learning variant aimed at enhancing weak learners, typically decision trees (or "stumps" in case of depth one). It increases the importance of mistakes so that following learners are encouraged to learn the cases which others find more difficult. With iterative training, Ada-Boost modifies the model's emphasis on hard-to-classify in-stances, and it does very good work in bias reduction with no trade-off to good generalization. Yet, AdaBoost can be affected by atypical and noisy data, as it tries harder to fix the errors made on these points which sometimes leads to overfitting.

3. **CatBoost (Ensemble Learning):** deals easily with categorical data by automatically boosting its performance on these features. CatBoost which is developed by Yandex, increases the performance of XGBoost and LightGBM by using ordered boosting and performing light preprocessing. Unlike traditional gradient boosting, CatBoost manages categorical variables directly and enhances performance by preventing data leakage. It further uses a light tree-splitting method to improve speed and protect against overfitting, making it effective in handling problems such as fraud detection, recommendation systems and financial risk rating.

4. **Long Short-Term Memory (LSTM):** LSTM extends the Recurrent Neural Network (RNN) by being able to remember information for an extended period and it is successfully used to overcome a problem called the vanishing gradient. This is achieved by adding just one memory cell, along with the forget gate, the input gate and the output gate. To begin, the forget gate sorts out which elements from the cell state should be removed. Secondly, the network team decides which new details should be retained in the cell and the output gate

determines the amount of cell memory to send on to the next step. When using an explicit memory framework, LSTMs perform well on activities that need handling sequences such as speech, text and time series situations.

5. **Grated Recurrent Unit (GRU):** GRU stands for Gated Recurrent Unit and is also a type of RNN, offering the same capability of processing long-range connections as LSTM at a simpler level. Unlike LSTM, GRU includes only reset gate and update gate. With reset gate, information from the past is discarded using an argument and the update gate lets you modify the current state with input using another argument. Unlike LSTMs, GRUs group the hidden state and the cell state into the same representation.

3.2 Dataset Profile

The e-transaction dataset for fraud detection holds the important eventful information of electronic money transaction which takes place between e-commerce websites using debit or credit card methods. All these transaction in the dataset takes place from various sources and in limited time period. From the total of 20,000 transactions, 7534 e-transactions are found out to be related to fraud. The Positive class aggregates to 37.7% of all the transactions. The entire dataset has been divided into 2 classes; on one side it is fraud happened transactions which are kept in fraudulent class which is represented by '1' and on the other hand the non fraud transactions which are kept in non fraudulent class which is represented by 0.

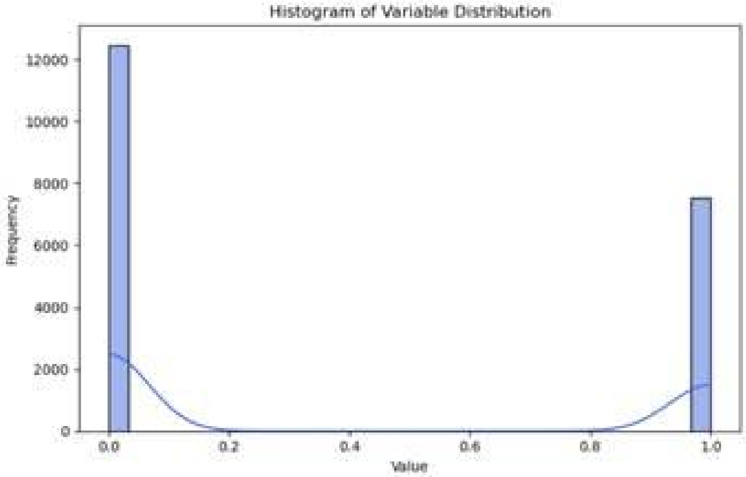


Fig. 1. Histogram for Class Distribution

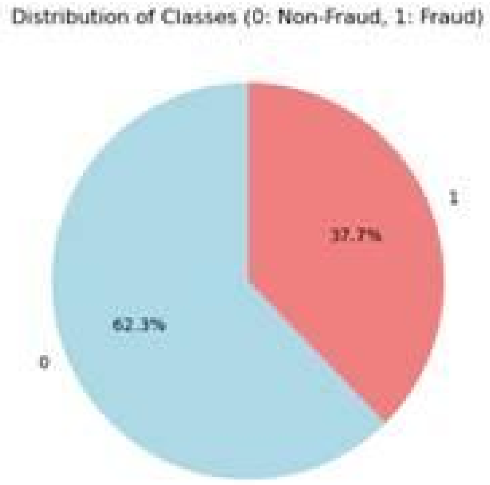
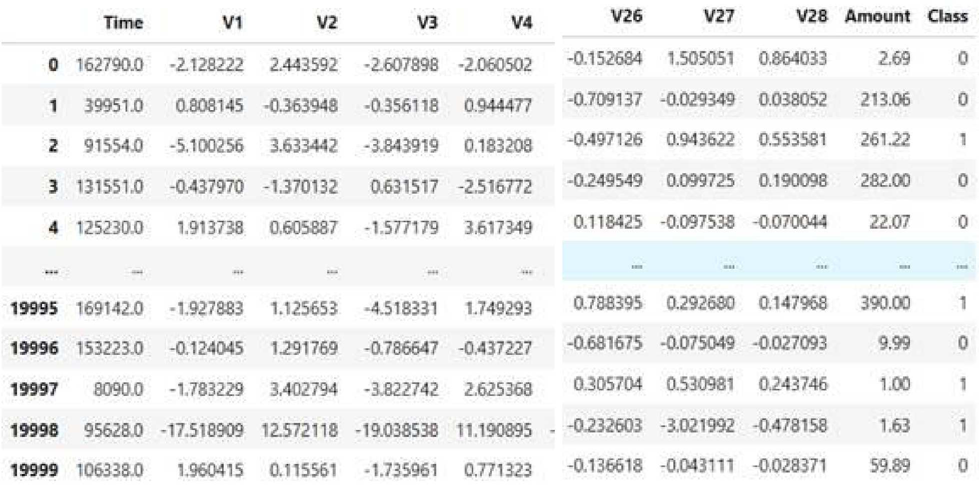


Fig.2. Pie Distribution for Class Variable

Here Figure 1. Is representing the Histogram distribution of class variable where we can see that 12,466 entries from 12,00 entries are not fraudulent while remaining 7534 entries are of class 1 which is the related to fraud. On the other hand, Figure 2. It shows variable division of e-transactions which are either + ve and – ve which represents 37.7% cases are fraudulent.

Table 1. Dataset Overview



	Time	V1	V2	V3	V4	V26	V27	V28	Amount	Class
0	162790.0	-2.128222	2.443592	-2.607898	-2.060502	-0.152684	1.505051	0.864033	2.69	0
1	39951.0	0.808145	-0.363948	-0.356118	0.944477	-0.709137	-0.029349	0.038052	213.06	0
2	91554.0	-5.100256	3.633442	-3.843919	0.183208	-0.497126	0.943622	0.553581	261.22	1
3	131551.0	-0.437970	-1.370132	0.631517	-2.516772	-0.249549	0.099725	0.190098	282.00	0
4	125230.0	1.913738	0.605887	-1.577179	3.617349	0.118425	-0.097538	-0.070044	22.07	0
...	...	...	...	...	...	...	...	...	...	...
19995	169142.0	-1.927883	1.125653	-4.518331	1.749293	0.788395	0.292680	0.147968	390.00	1
19996	153223.0	-0.124045	1.291769	-0.786647	-0.437227	-0.681675	-0.075049	-0.027093	9.99	0
19997	8090.0	-1.783229	3.402794	-3.822742	2.625368	0.305704	0.530981	0.243746	1.00	1
19998	95628.0	-17.518909	12.572118	-19.038538	11.190895	-0.232603	-3.021992	-0.478158	1.63	1
19999	106338.0	1.960415	0.115561	-1.735961	0.771323	-0.136618	-0.043111	-0.028371	59.89	0

The dataset which we are using in this study is the raw dataset of e-transaction which is later changed into number formed input features using exploratory data analysis (EDA), moreover with that we have also focused on using other methods such as reduction of the dimensions, compression of information and filtering of the noise. Also, the variables which are independent from V1 to V28 are later changed into features which are delivered using Principle Component Analysis (PCA).

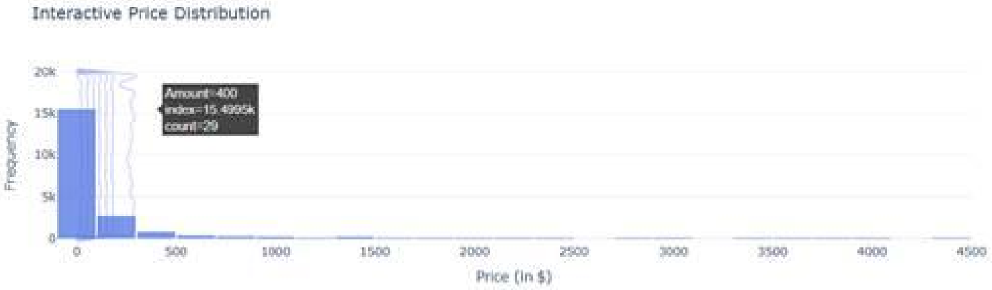


Fig. 3. Distribution of Transaction Amount

The Fig. 4 The figure 4 represents us the word Amount which simply means the value of fraud in the e-transaction data set. Amount's value can be different in every row of the dataset. On the other hand the Class feature is served as one and only target variable which we will be predicting. The value of the Class feature will be binary; 0 or 1, where 0 will indicated a right transaction and 1 will indicated a transaction related to fraud.

CHAPTER 4

RESEARCH IMPLEMENTATION

The implementation of this paper involves in multiple procedures where the work starts from gathering the dataset and ends by evaluating the performance of the 5 different machine and deep leering models which we are comparing for the prediction of E-transaction fraud detection in E-commerce platforms. Let's understand the multiple procedures which we were talking about:

4.1 Implementation Steps

Step 1: In the very first step we focus on collecting and cleaning the e-transaction dataset of a e-commerce website for the purpose of its exploratory data analysis phase. The data here is collected from multiple sites which had null values, error values and similar related problems which were required to be removed before any further word could be performed on them.

Step 2: Among multiple variables in the dataset best variables/features are selected which have larger impact on the output value from the cleaned dataset. The best features are selected on the basis of the correlation and covariance of the features with each other and output variable. If the values of relations are higher that means they impact the outcome, so they are kept and ones with low relation value are discarded and thus this step is called as feature extraction or selection.

Step 3: In this step we receive the data set which is cleaned and have the required features which will contribute towards predicting the output value. This dataset is not split in two groups which are train split group and another one is test split group. Train split group is used to train the data over the 5 different machine learning frameworks, while the Test split group is used to test the trained the model to give out the performance evaluation.

Step 4: In this most crucial step pre-processed data from above steps are used to train the 5 different models along with tunning the hyper parameters. Mostly our models involve ensemble learning frameworks or deep learning frameworks. Random Forest, AdaBoost, CatBoost, Long Short Term Memory and Gated Recurrent Unit are used for training of the models.

Step 5: In this last step the performance of the trained frameworks are compared to find which one is best for this particular dataset and use case. The best selected model then is deployed to detect the presence of fraudulent transaction which is leading to theft of money.

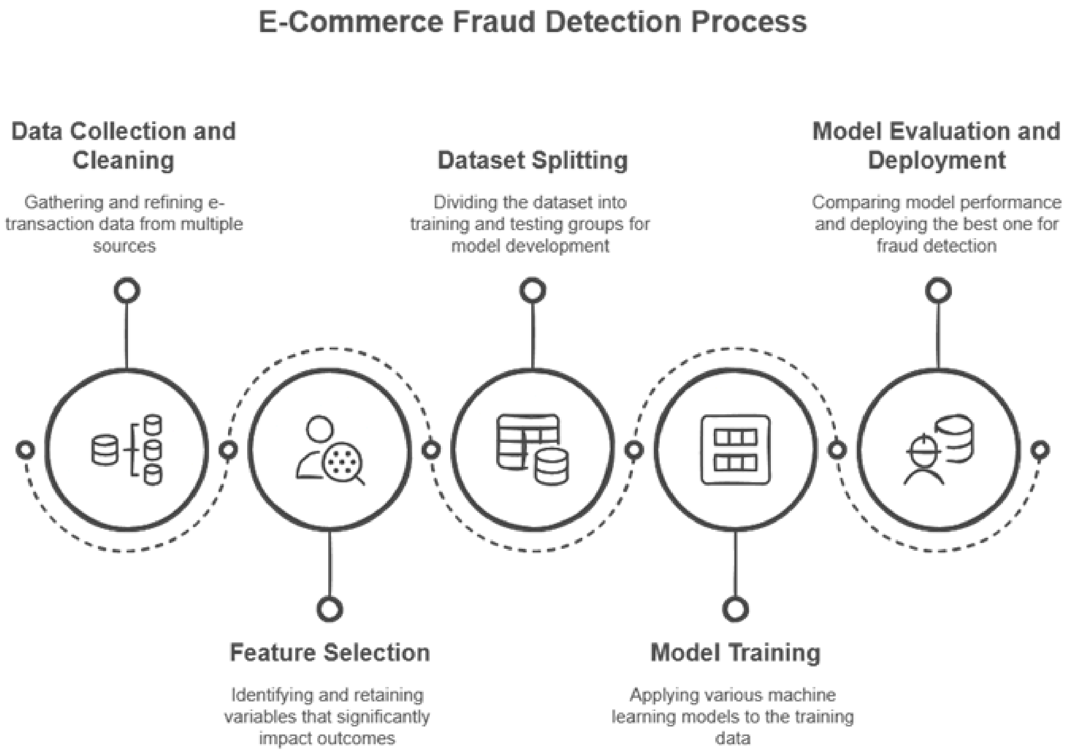

E-Commerce Fraud Detection Process
Data Collection and Cleaning
Gathering and refining e-transaction data from multiple sources
Dataset Splitting
Dividing the dataset into training and testing groups for model development
Model Evaluation and Deployment
Comparing model performance and deploying the best one for fraud detection
Feature Selection
Identifying and retaining variables that significantly impact outcomes
Model Training
Applying various machine learning models to the training data

Fig. 4. E-Commerce Fraud Detection Process

4.2 Implementation Methodology

The crucial steps which are involved in the detection of Fraud e-transaction on e-commerce websites which are done using unauthorized usage of cards are:

Step 1: First of all, the data is collected from multiple e-commerce websites by using web scrapping. For web scrapping python's very useful library beautiful soup is used. Then this data is converted in tabular format which is later saved in .csv format which is an essential way of storing data for data cleaning and data pre-processing. Once data is converted into .csv file format, proper names to the columns are given, null values were been removed and encoding of categorical variables had been conducted.

Step 2: Second step is generally related to feature extraction and selection, but due to necessity of every column holding an importance in prediction of outcome variable, none of the feature column is removed but rather to maintain the privacy of data, 28 out of 31 features had been encrypted in a way that their name and value has been encoded so that no one knows what the details are but along with it they can still contribute to prediction of target variable.

Step 3: Generally, when we collect the data, it is usually not arranged in systematic order and specifically when you have to use it for prediction it usually not balanced. By balanced it means the outcome variable should contain somehow equal distribution of the classes. In our case there are two classes; Class 1 if there is a fraud and Class 2 if the transaction not ended up in fraud. To make our data balanced I down sampled the non-fraudulent class and on the other side I over sampled the fraudulent class. This is how we were able to have equivalent entries in both the classes.

Step 4: In this particular phase the given e-transaction dataset is broken down in 2 different sets, on one side it is the training dataset containing e-transactions for training and on other side it is the testing dataset containing e-transactions for testing. The training set is majorly implied to develop and train the framework. Then the testing dataset is later used to find out the framework's accuracy. As we know while breaking the dataset we keep the data split ratio as 90:10 where 90% is given in training dataset and 10% in testing dataset.

Step 5: Last and the most important step in the methodology of predicting the e-commerce transaction fraud is to train the different models over your data. The following 5 different ML and DL frameworks were implied to perform e-commerce fraud detection:

a) **Random Forest:** Random Forest is a kind of ensemble learning that trains a huge number of trees at the start and combines their estimations to ensure accuracy. Although one decision tree will be susceptible to high variance, Random Forest generates a "forest" of trees, where every tree is trained on a random subset of the data. Overall, it relies on the most common solution for GP problems and can successfully address noisy data, working well with lots of dimensions.

b) **AdaBoost:** is another ensemble learning variant aimed at enhancing weak learners, typically decision trees (or "stumps" in case of depth one). It increases the importance of mistakes so that following learners are encouraged to learn the cases which others find more difficult. With iterative training, Ada-Boost modifies the model's emphasis on hard-to-classify in-stances, and it does very good work in bias reduction with no trade-off to good generalization. Yet, AdaBoost can be affected by atypical and noisy data, as it tries harder to fix the errors made on these points which sometimes leads to overfitting.

c) **CatBoost:** deals easily with categorical data by automatically boosting its performance on these features. CatBoost which is developed by Yandex, increases the performance of XGBoost and LightGBM by using ordered boosting and performing light preprocessing. Unlike traditional gradient boosting, CatBoost manages categorical variables directly and enhances performance by preventing data leakage. It further uses a light tree-splitting method to improve speed and protect against overfitting, making it effective in handling problems such as fraud detection, recommendation systems and financial risk rating.

d) **LSTM (Long Short Term Memory):** LSTM extends the Recurrent Neural Network (RNN) by being able to remember information for an extended period and it is successfully used to overcome a problem called the vanishing gradient. This is achieved by adding just one memory cell, along with the forget gate, the input gate and the output gate. To begin, the forget gate sorts out which elements from the cell state should be removed. Secondly, the network team decides which new details should be retained in the cell and the output gate determines the amount of cell memory to send on to the next step. When using an explicit memory framework, LSTMs perform well on activities that need handling sequences such as speech, text and time series situations.

e) **GRU (Gated Recurrent Unit):** GRU stands for Gated Recurrent Unit and is also a type of RNN, offering the same capability of processing long-range connections as LSTM at a simpler level. Unlike LSTM, GRU includes only reset gate and update gate. With reset gate, information from the past is discarded using an argument and the update gate lets you modify the current state with input using another argument. Unlike LSTMs, GRUs group the hidden state and the cell state into the same representation.

4.3 Performance Evaluation

In the entire development of predicting an target outcome which is the target class outcome on basis of previous data, the most important aspect is to find out how well is your predicted outcome working. Weather the predicted outcome is just a luck guess or is it an proper estimated outcome. To measure the correctness of the outcome several performance evaluation techniques are used in the field of machine learning, let’s look at those evaluation techniques:

One of the earliest statistical methods of correctness and precision testing of a framework is through Cross-Validation (CV) Technique. Cross-validation estimates a framework's performance by testing its ability to generalize to an independent data set. Cross-validation splits data into subsets and uses one of them as the training data set for a model and another as the validation or test data set. There are different cross-validation methods, and we have used K fold-cross validation. Mechanism and formula of K fold-cross validation are as follows:

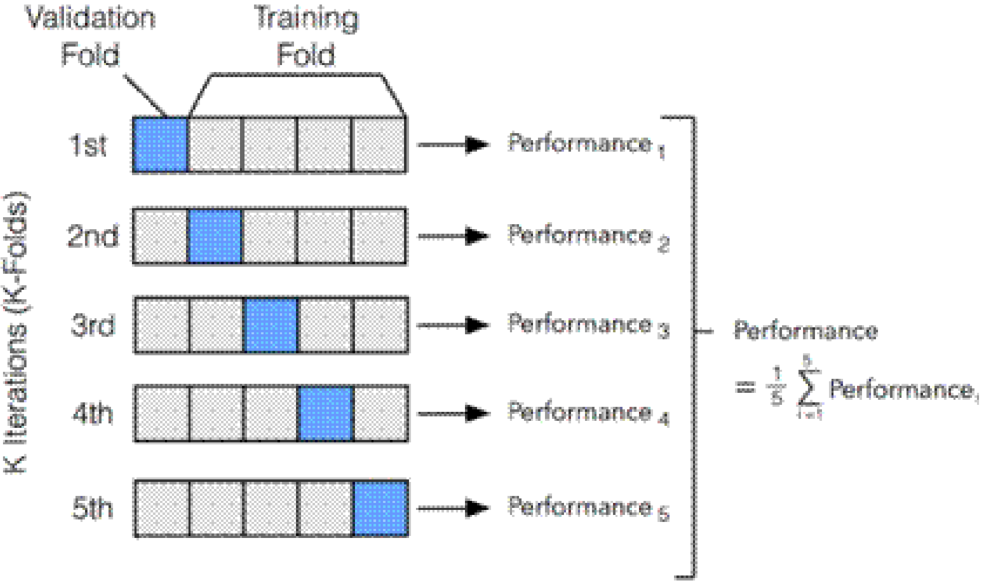


Fig. 5. K-Fold Cross Validation Process

We tend use cross validation to examine the performance of a framework due to several reasons which look like, CV reduces bias compared to simple splits. Also, it provides a best way to find of framework's accuracy which helps in detecting overfitting and underfitting issues.

In the field of data science one method which is widely used for finding out how well a binary classification framework works us AUC - ROC curve/score. The AUC-ROC score well stands for the Area Under the Curve - Receiver Operating Characteristics, with the use of ROC component of this score we can showcase a graphical representation which will tell us the trade off which will take place between false positive and true positive rate at multiple hyper tuned settings. On the other hand, the AUC component of this score will show us the possibility if that the framework will rank a randomly chosen + ve class which is above than an another randomly chosen -ve one. At last, we can say that if AUC score is 1.0 it tells that the absolutely perfect accuracy where on the other side it the AUC score is 0.0 it tells us that the frameworks predictions are completely wrong.

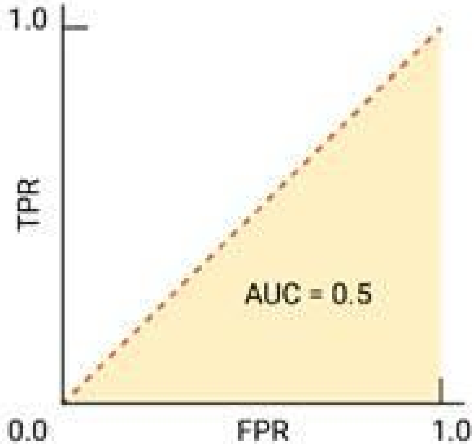
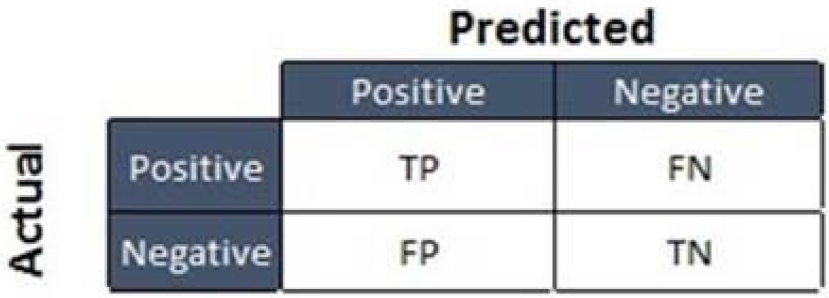


Fig. 6. AUC – ROC Score Curve

We can also implement confusion matrix to display how does this classification framework is confused while making an new prediction. The confusion matrix holds 4 different values which are True Positive, False Positive, False Negative, True Negative. Let's look at the confusion matrix;



		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Fig. 7. Confusion Matrix

As an output True Positive (TP) justify that model predicts the positive class, which means the both prediction and actual value are in positive co-ordinate. On other hand False Positive (FP) is an outcome which comes when the framework makes a wrong prediction of the negative type class which means actually it was negative but the prediction which was made was positive, this is called Type I error.

Similarly, if an output is True Negative (TN) that means the model has rightly predicted the negative class, in this case predicted value and the actual value both are negative. Along with it, if outcome is False Negative (FN) that means the model has wrongly predicted the positive class which implies that predicted value is negative but the actual value is positive, this is called Type II error.

Lastly, another way to analyze classification problem's performance is using threshold-oriented evaluation techniques which are Precision, Recall and F1 score.

Here, Precision shows the percentage of truly positive predictions out of all those that were predicted to be positive. It explains how many of the predicted fraud transactions are actually fraud. If the precision score is high, it means the model makes only a few errors in thinking a transaction is fraud when it is not. This metric becomes essential when allowing a false alarm can be expensive, as in the case of not letting a legit customer check out. Combining Precision with Recall is beneficial, especially when there is an imbalance in the data. We can see this in the formula that is given.

$$Precision = \frac{TP}{TP+FP} \tag{4.1}$$

The term recall or sensitivity, measures the performance of a classification model. It represents the percentage of positive instances that the model finds out of all the positive instances. To put it simply, recall assesses how well a model locates every important case in the dataset. When there is a high recall, most actual positive cases are not missed and this is very useful in serious fields like diagnosis or crime detection. In fraud detection, if a company does not identify a fraud transaction, this can cause serious losses. But, remembering must go hand in hand with being precise to make certain the model does not label everything as positive. Considering how well a model recalls and how precisely it predicts, through F1-score, gives a clearer impression of its performance. More precisely, we can state it this way:

$$Recall = \frac{TP}{TP+FN} \tag{4.2}$$

F1 Score is the name given to the HM of Precision and Recall. Here, measurement is about the accuracy of predicting a positive case. As opposed to precision, recall determines how well the model sees all positives in the data. As a result, the F1 Score is better able to balance these two characteristics since the harmonic mean punishes big outliers. differential employment is used for The harmonic mean yields a high F1 Score, only if both precision and recall are high.

Hence, F1 Score is widely used when dealing with imbalanced data, like in fraud detection and illness diagnosis. It helps avoid an unrealistic high score if one of the performance indicators is quite low. When recall is high and precision is low (or the other way around), the results may not be trustworthy — this is why F1 takes both values into account. Furthermore, F1 Score values can vary between 0 and 1 and a value of 1 means both precision and recall are perfect. You can find it being used in situations where the outcome of a false positive is more serious than that of a false negative.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (4.3)$$

Above mentioned six different ways to evaluate the performance of machine learning framework once trained over training related data and then tested over testing related data. One with the highest value in them, is chosen as the best mod-el to predict the fraud in e-transaction on e-commerce platform.

CHAPTER 5

RESULTS AND DISCUSSION

Our training data was used to train each of the five frameworks and afterward, we tested each on the testing data to determine which works best. The results showed visible differences in accuracy, precision, recall, F1-score and AUC-ROC. Based on the preprocessed and balanced data, the models exhibited their strengths as well as what they were not good at. Random Forest and XGBoost ensemble models showed good generalization and reliability, but LSTM and GRU outperformed them at noticing patterns in the transactions. Analyzing the results allowed me to better understand the performance of each model. We have analyzed all the models we mentioned above and stated which one delivered the best and most accurate performance in handling fraud tasks.

A) Cross Validation:

By using K-Cross Validation on our frameworks, it is clear that Random Forest scored 99.95% accuracy, AdaBoost achieved 95.63%, CatBoost managed 99.91%, LSTM stood at 98.64% and GRU finished with 98.32% accuracy. It is obvious from these results that handling complex e-commerce fraud data is easier for Random Forest and CatBoost models than for others. While LSTM and GRU showed weaker results compared to CNN and RNN, they excellently captured sequences in the data, displaying great results. Therefore, time-series models play a significant role in detecting fraud when dealing with transactions. Furthermore, the K-Fold Cross Validation approach ensured that the models did not overfit. Furthermore, the success of our experiments on new and independent data proved that our model works reliably for practical use.

B) AUC – ROC Score:

While using the AUC – ROC curve, we noticed that the frameworks could successfully and reliably predict the outcome. Rach model gave 100% on the AUC, meaning the framework separated both sets of classes perfectly and did not produce any errors. Compared to AdaBoost with a score of 0.9951, CatBoost preferably scored 0.999, demonstrating its powerful abilities. LSTM managed to capture the temporal aspects of transactions with an AUC score of 0.9991. In the same way, GRU had a good performance with a score of 0.9989, meaning it too is a dependable model for preventing fraud.

The findings suggest that ensembles and sequential models can effectively deal with fraudulent transactions in banking. Significantly, LSTM and GRU almost matched Random Forest in accuracy and play an important role in identifying patterns in groups of e-commerce transactions. AUC – ROC is useful for checking the performance of models, especially when the data is not balanced and accuracy might give incorrect results. By this point, we can tell that these models efficiently detect fraud in data.

C) Precision, Recall and F1 Score:

After putting all the models into action, we picked Precision, Recall and F1 Score to assess the accuracy and correctness of what we built. Such metrics were used since they let us evaluate both the true positive and true negative results in the model. No model was subject to overfitting because we tested them with a separate set of data from what was used for training. Using these evaluation metrics, we found out how each system responds to real-time threats of fraud. Here are the outcomes of what we saw.

Table 2. Performance comparison of models over different evaluation metrics

	Precision Score	Recall Score	F1 Score
Random Forest	0.9993	1.0000	0.9997
Ada Boost	0.9611	0.9214	0.9480
Cat Boost	0.9980	1.0000	0.9990
LSTM	0.9695	0.9681	0.9688
GRU	0.9702	0.9485	0.9592

Based on the above scores, CatBoost and Random Forest achieved the highest performance with near-perfect precision, recall, and F1 scores. LSTM and GRU also performed well, demonstrating the strength of deep learning models on sequential data. AdaBoost, while effective, showed comparatively lower recall and F1, indicating it may miss more positive cases.

D) Confusion Matrix:

Using the confusion matrix we have gathered the outcomes of all the different models respectively, Fig. 4 to Fig. 8 represents different confusion matrix for different frameworks;

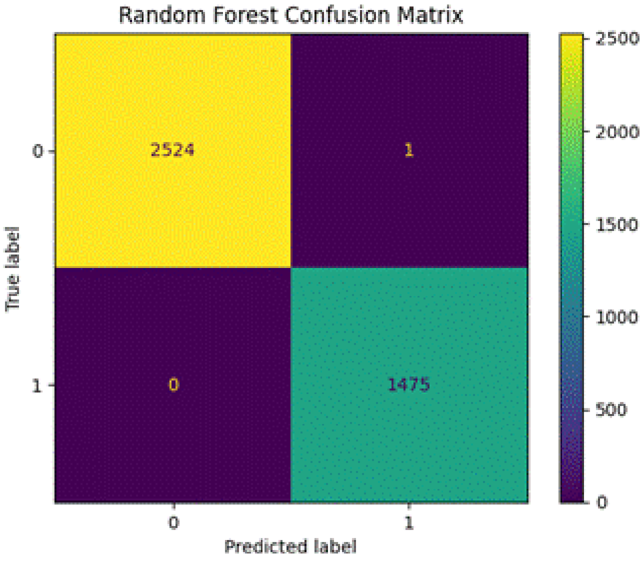


Fig. 8. Confusion Matrix for Random Forest Framework

Fig.8. In the fig. 8 the random forest model does correctly classifies 2524 e-transactions as fraud and remaining 1475 e-transactions are classified as non-fraud. However, it also misclassifies 1 fraudulent e-transaction as a non-fraudulent e-transaction.

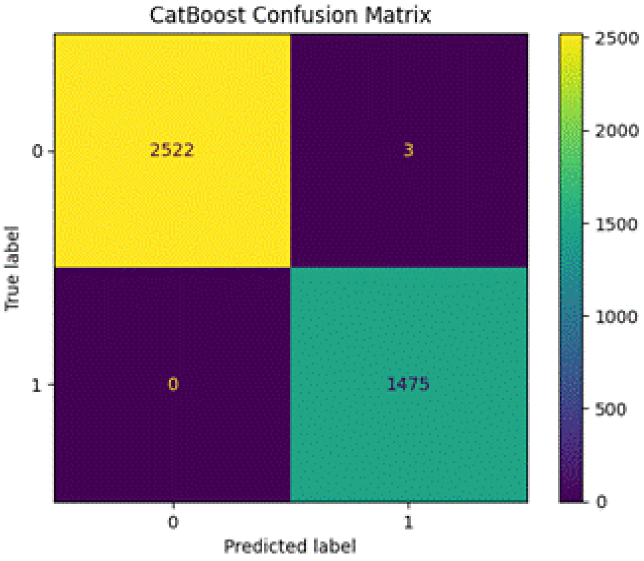


Fig. 9. Confusion Matrix for CatBoost Framework

Fig.9. In the fig. 9 the Catboost model does correctly classifies 2522 e-transactions as fraud and remaining 1475 e-transactions are classified as non-fraud. However, it also misclassifies 3 fraudulent e-transaction as a non-fraudulent e-transaction.

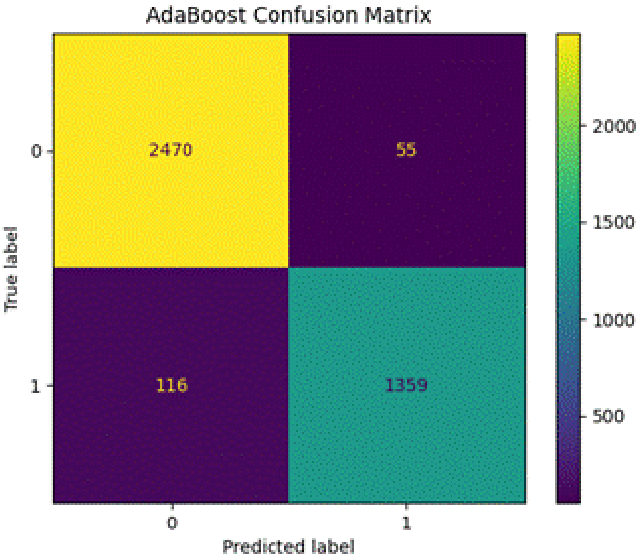


Fig. 10. Confusion Matrix for AdaBoost Framework

Fig.10. In the fig. 10 the Adaboost model does correctly classifies 2470 e-transactions as fraud and remaining 1359 e-transactions are classified as non-fraud. However, it also misclassifies 55 fraudulent e-transaction as a non-fraudulent e-transaction and 116 non fraud e-transaction as fraud e-transactions.

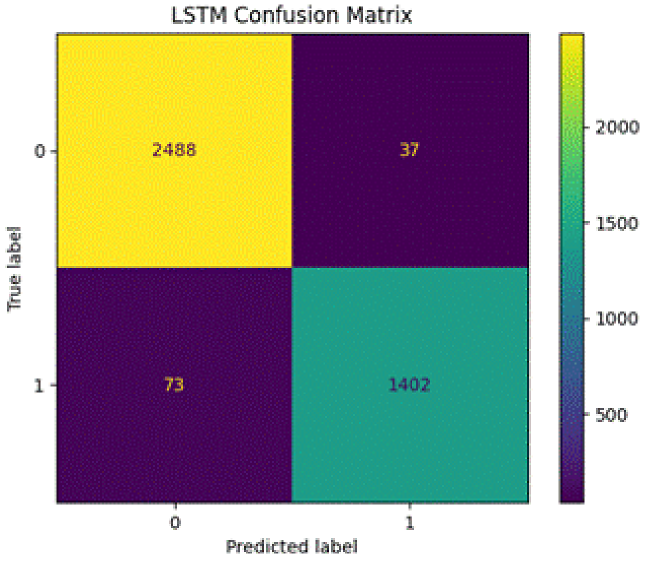


Fig. 11. Confusion Matrix for LSTM Framework

Fig.11. In the fig. 11 the LSTM model does correctly classifies 2488 e-transactions as fraud and remaining 1402 e-transactions are classified as non-fraud. However, it also misclassifies 37 fraudulent e-transaction as a non-fraudulent e-transaction and 73 non fraud e-transaction as fraud e-transactions.

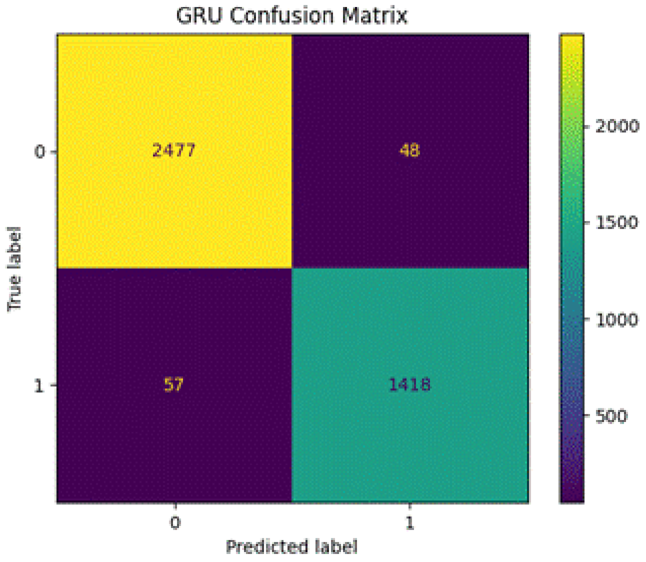


Fig. 12. Confusion Matrix for Random Forest Framework

Fig.12. In the fig. 12 the GRU model does correctly classifies 2477 e-transactions as fraud and remaining 1418 e-transactions are classified as non-fraud. However, it also misclassifies 48 fraudulent e-transaction as a non-fraudulent e-transaction and 57 non fraud e-transaction as fraud e-transactions.

CHAPTER 6

CONCLUSION

Accessing credit cards without permission or paying for items using them in risky or dangerous spots is a major problem in the digital economy today. As a result, users and banks can suffer not only monetary damage but other more severe cybercrimes too. Here, we look at how the use of ML and DL can improve the performance and dependability of fraud detectors used in e-commerce. Since cyber threats are constantly changing and becoming more advanced, just running predefined rules cannot protect a company anymore. For this reason, scientists sought to study models that can pick up data patterns and change over time.

The major focus of the research was using ML and DL to analyze older transaction records and examine them for warning signs of fraud. Using a supervised learning approach with labeled data, we built models and compared their results with a number of metrics for classification. First, we used exploratory data analysis (EDA) to look at the data and created graphs and charts to highlight any patterns and unusual behavior. Understanding this allowed us to pick and design effective features for our solution.

It was difficult to detect fraud since the activities in the dataset showed that there were significantly fewer instances of fraud compared to other (legitimate) transactions. Since classification happens without balance, the algorithms end up ignoring uncommon symptoms of fraud and learning to spot the most usual cases. Therefore, we applied techniques that randomize data by boosting the number of instances in the minority group and decreasing them in the majority group. Some models also used SMOTE to make new synthetic samples and help balance the data. These techniques ensured that the models were not biased and could effectively learn to distinguish between fraudulent and non-fraudulent transactions.

Even though there were some difficulties, the final models that we chose performed incredibly well, with the most accurate prediction of cybersecurity fraud in e-commerce reaching 99.95%. The Random Forest classifier was the most reliable and best performing in all the evaluations using precision, recall, F1-score and specificity. Handling many datasets, limiting overfitting problems and order of feature importance made this machine learning technique very useful here. The fact that Random Forest combines outcomes from many trees helped make it both reliable and accurate.

It should also be noted that XGBoost, CatBoost, AdaBoost, LSTM and GRU each scored highly successful results as well. While certain algorithms focused on catching patterns in our data, others worked better on improving the performance through tactics that help with the certain categorical features. Ensemble models that use these various models in reading tasks were found to be a helpful approach as well.

Overall, this research asserts that using machine and deep learning is essential for building fraud detection systems of today. They improve the security of e-commerce businesses and also give useful advice to prevent fraud as it happens. In the future, more work could use real-time streaming data, add explainability to the models using SHAP or LIME and boost their performance for live usage. Thanks to this research, it will be easier to design and improve cybersecurity systems as fraud moves through different stages.

CHAPTER 7

FUTURE WORK

In relation to our project on cybersecurity fraud prediction for e-commerce, this paper paves the way for using the very advanced machine learning and deep learning tools in detection of frauds. Still, this subject is wide-reaching and many scientists have a lot more ground to explore and improve on. New studies will try to discover ways to strengthen and enhance the trustworthiness, effectiveness and ease of use of systems for fraud detection. With new tricks being developed every day, the best systems can adapt instantaneously and work at a high level across all kinds of transactions and in any environment.

Futuristic AI development should put a strong emphasis on innovating and implementing hybrid and ensemble frameworks for deep learning. Unlike individual learning models, hybrid models can make use of a variety of learning techniques to overcome their unique problems. Development of LSTM-based custom CNNs combines neural network learning of time sequences with the understanding of spatial patterns. Blending GRUs with Transformers allows both the memory of each segment of the data and the system's ability to focus on specific parts to discover small and detailed fraud patterns in transactions.

Since these types of models combine different algorithms, their results are more accurate and better at detecting fraud. The vanishing gradients are often a problem for the RNN models, unlike transformer models that have performed very good at processing long-range relationships found in e-commerce log data. When partnered with advanced technologies like GRU or LSTM networks, transformers make it possible for the system to notice that both the short-term oddities and longer-term activities are likely related to fraud.

Likewise, optimization can be carried out using automated tuning of hyperparameters, neural architecture search and evolutionary algorithms that target multiple objectives. Using these techniques can ensure the systems provide good detection accuracy while also using fewer computing resources.

It may be useful to add the ability for the model to analyze data from ongoing transaction flows in real time. Online courses or gradual learning can enable these systems to detect e-commerce fraud quicker by adapting to newly found patterns in them rather than using outdated data. If these algorithms are included in a semi-supervised approach, they might detect the fraud that does not exist in the current labelled data.

Additionally, enhancing models for fraud detection should include explainability through tools such as SHAP and LIME. Hi-tech tools allow for finding out why a transaction was considered fraudulent which helps developers and financial firms accept and depend on the results given by the model. During the same process, ensuring privacy and adhering to regulations is important when dealing with financial data.

To sum up, success in predicting cyber fraud in e-commerce depends on building flexible and scalable systems that are highly efficient and able to change as new threats emerge. Achieving this will depend on using new deep learning methods, advanced group methods, continuous adaptation and ability to be understood by people.

BIBLIOGRAPHY

[1] R. Aggarwal, P. K. Sarangi and A. K. Sahoo, "Credit Card Fraud Detection: Analyzing the Performance of Four Machine Learning Models," In: 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, pp. 650-654 (2023).

[2] Chhabra, R., Goswami, S. & Ranjan, R.K, “A voting ensemble machine learning based credit card fraud detection using highly imbalance data.” In: Multimed Tools Appl **83**, 54729–54753 (2024).

[3] Asha RB, Suresh Kumar KR, “Credit card fraud detection using artificial neural network.” In: Global Transitions Proceedings, Volume 2, Issue 1, Pages 35-41 (2021).

[4] K. B. Aswathi, S. Jayadev, N. Krishna, R. Krishnan and G. Sarath, "Botnet Detection using Machine Learning," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-7 (2021).

[5] Siddhant Bagga, Anish Goyal, Namita Gupta, Arvind Goyal, Credit Card Fraud Detection using Pipeling and Ensemble Learning, Procedia Computer Science, Volume 173, Pages 104-112 (2020).

[6] Devikar, M., Khadke, A., Lad, A., Sapkal, R., & Nikalje, S. Credit card fraud detection using ensemble learning. International Research Journal of Engineering and Technology (IRJET), 7(05), 7402 (2020).

[7] D. Prusti and S. K. Rath, "Web service based credit card fraud detection by applying machine learning techniques," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, pp. 492-497 (2019).

[8] Sahin, Y., & Duman, E. In: Detecting credit card fraud by decision trees and support vector machines. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 1, pp. 1-6) (2011).

[9] S. K. Babu, S. Vasavi and K. Nagarjuna, "Framework for Predictive Analytics as a Service Using Ensemble Model," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, pp. 121-128 (2017).

[10] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods,"18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina.” (2019)

[11] Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., & Singh, A. K. Credit card fraud detection using machine learning: a study. arXiv preprint arXiv:2108.10005 (2021).

[12] I. D. Mienye and N. Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," in IEEE Access, vol. 12, pp. 96893-96910 (2024).

[13] Khalid AR, Owoh N, Uthmani O, Ashawa M, Osamor J, Adejoh J. Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. Big Data and Cognitive Computing (2024).

[14] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407 (2022).

[15] I. D. Mienye and Y. Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection," in IEEE Access, vol. 11, pp. 30628-30638 (2023).

[16] I. Vejalla, S. P. Battula, K. Kalluri and H. K. Kalluri, "Credit Card Fraud Detection Using Machine Learning Techniques," 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), Nagpur, India, pp. 1-4 (2023).