Machine learning model for credit card fraud detection using Ensemble methods

Chelsy Fernandes and Megha Manoj
Department of Computer Science,
Birla Institute of Technology and Science, Pilani,
Dubai International Academic City,
Dubai, United Arab Emirates

Abstract

The internet is a huge open community filled with different kinds of activities from searching to sharing information that can be accessed all over the world. This ease of access to the internet is its advantage and disadvantage as it puts sensitive information at risk of being misused. Despite the placement of security measures, internet frauds have been taking place for a long period. Internet fraud refers to a cybercrime where the financial resources of an individual are stolen. The American internet crime report issued by the FBI states that about 10.3 billion USD was lost to internet fraud in the year 2022. Furthermore, an analysis conducted by The Federal Trade Commission's (FTC) Consumer Sentinel Network revealed that 46% of the analyzed 5.1 million reports were credit card frauds. There was a rise of 65% in the amount of identified or prevented credit card frauds in 2022 according to the National Hunter Fraud Prevention Service, that is, nearly 2 in every 3 cardholders have experienced this type of fraud at least once. The use of advanced technology and the anonymity of the fraudsters make it more difficult to identify a card theft. To counter the hike in card fraud, machine learning methods have been applied to transaction-related datasets collected from financial institutions such as banks, investment companies, etc. The various machine learning models researched and implemented include classifiers as well as regression models such as Logistic Regression, Supervised Vector Machine (SVM), Random Forest, etc due to the continuous nature of the transactions. The main objective of the applied models is to detect anomalies within the transactions taking place by observing the expenditure, time, and location details of the completed transactions. However, the irregularity of the transactions and lack of detailed information on them downplay the accuracy levels of the output produced. It is also to be noted that there is a shortage of data as most institutions are reluctant to provide client-related information to avoid security breaches. To counter this problem, we propose building and comparing machine learning models for anomaly detection capable of detecting any unusual behavior related to the duration and delay between the transactions. The models built will use supervised vector machines, logistic regression, Xgboost, and ensemble learning techniques which combine the SVM and logistic regression models. The results of each model will be compared against each other in terms of accuracy to observe which one shows better performance.

Introduction

Businesses have advanced in the twenty-first century because of online banking and e-payments. E-payments are crucial factors affecting today's society. The purchase of goods and services have become very easy. Nowadays banks provide credit cards to their holders. The use of credit cards has made life easy. Customers can shop without cash. Customers need to confirm the transaction by an OTP (one time password). And after the transactions are done the credit card owner gets a message of the transaction.

Although credit cards have many benefits associated with it it also has some problems like fraud and security issues. Credit cards are considered one of the challenging frauds to identify. Information about credit cards is very sensitive and needs to be protected and handled with utmost care. Without a lot of risks a huge amount of money can be withdrawn by the fraudsters. It is an issue that affects a large number of people across the globe. About $5.5 billion is lost to credit card fraud annually that is 40% of the total monetary frauds committed. Identity theft is also a part of credit card fraud which affects about 10.7 million people yearly. Fraud mostly happens when a card is stolen or the information is hacked .

Credit card frauds can happen in many ways. One such way is phishing. In this type of crime, the culprits usually send messages or emails that trick the people in such a way that the person might give their personal details and credit card details to them. These fraudsters pretend to be legitimate institutions or companies and ask for sensitive details of the credit card like the security code, expiry date etc. Another fraud technique is skimming where the fraudsters use a device with a magnetic strip to read all the information of the credit card. These fraudsters often use the identity to make payments for purchases which the card holder doesnt make. Another way could be to withdraw money from the victim's account. The fraudster could possibly apply for a new credit card or even open a new bank account. Scammers may use very simple methods like information from papers in trash cans or just peeking in when a person is entering PIN at the atm.

Card not present frauds also happen where the fraudster may use a stolen credit card information to make purchases online. Cards that are stolen or lost also pose a threat to the owner of the card. The fraudsters can use it to withdraw large amounts of money. And so advised that the card holder should immediately block the card if it is lost or they feel that the transactions were not made by them. ATM frauds are also another problem. The fraudsters can fix a device that would capture all the details of your card. We must be very careful while buying something online. A fraudster might create a fake online website which might store all your credit card information. Therefore, we need to regularly keep a watch on statements of credit cards. Never give your secret information like security code to anybody. Make sure to use strong pins and passwords that fraudsters cannot crack.

Financial institutions also need to play a role in preserving the card holder's information by devising credit card fraud detecting mechanisms. The detection of credit card fraud is a heavily explored domain. Often this detection relies on auto analyzed transactions that are recorded. About 77% of the credit card holders examine their bank account every week to look for any suspicious activity. Fraudsters may use the details of the card holders to make online purchases. About 62% of people believe that they are at a higher risk of fraud nowadays than about two years ago. 38% of card holders have canceled your credit or debit cards completely because of online frauds and illegitimate transactions. Sometimes transactions are declined by the banks by mistake. This is because banks suspect fraudulent activity. The issue of credit card fraud is very significant today and impacts banks and companies that issue the cards. In this huge system of transactions credit card frauds are an extremely serious problem for banks. They have extensive mechanisms to detect fraudulent transactions as soon as possible. A trustworthy banking system would have two step authentication and verification mechanisms which allow legitimate transactions to take place smoothly. But detect any transaction that is under suspicion.

Anomaly detection tools use ML and these models collect information continuously and keep analyzing the new data that comes into the model. Even the advanced methods or algorithms analyze the users transaction history and behavioral pattern.

There are many challenges faced in detecting credit card frauds. The data sets are usually very imbalanced because the number of legitimate transactions are more than the fraud transactions. And so the traditional classifier methods often fail to detect the fraud transactions from the data sets that are skewed. The classification algorithms identify the transaction that might be highly probable of being fraud. The basis of this classification are the details of the transaction like location and time of transaction, amount etc. These details often do not have enough information of the customer to detect the frauds accurately as these models do not keep track of the spending pattern of the customer.

A card holder must often keep a watch on their credit card transactions. The user must inspect their monthly bank statements to identify illegitimate transactions. Keep a watch on mobile messages for any OTP or alert messages from the banks to warn about transactions that seem suspicious. In case of suspicious transactions, block the card as soon as possible and contact the credit card company for the information of the transaction and file a complaint for investigation.

In this paper we build an anomaly detection model to predict credit card frauds from online sites or physical illegitimate transactions. Our main aim is to use various machine learning algorithms to compare the models built and bring out the most efficient model in detecting fraudulent transactions.

OBJECTIVES

For this project we have proposed the following agenda:

- Select a transaction dataset related to credit card frauds and preprocess it using relevant data mining techniques.
- Build classifier models for anomaly detection using SVM, Logistic Regression, Xgboost and ensemble learning models
- Use evaluation metrics to analyze the accuracy levels of the model outcomes.

PROBLEM STATEMENTS

- Build anomaly detection models to predict credit card transactions**.**
- Solve the imbalance in dataset
- Classify the transactions to be genuine or fraudulent and by evaluating different performance measures and find an efficient model for detecting credit card fraud out of various algorithms used.

DATASET DESCRIPTION

The dataset used in this project is in csv format and was derived from Kaggle. The dataset has transaction details of European card holders and it took 2 days for its completion. The dataset contains a total of 284,807 transactions out of which 492 transactions are labeled as fraud, which accounts for 0.172% of the entire data. The dataset has the presence of imbalanced records for the legitimate class. The features of the dataset comprise of the time which the transactions took place, 28 attributes(v1-v28) depicting the transactions and the amount transferred which have been encrypted into numeric values using principal component analysis (PCA) to compress the large amount of data present within them and a feature called class, which contains the records' label. The column 'class' contains binary values 0 and 1 where 1 indicates the transaction is fraudulent and vice versa. To resolve the imbalance present within the dataset, Area Under the Precision-Recall Curve (AUPRC) has been applied. To determine which features will provide most efficient results, a correlation function and heatmap is used which displays numeric values as an estimate of how related the features are to each other. For extracting the exact duration of the transactions, a pandas function 'time_delta' is applied and a derived column Time_Hour is created.

LITERATURE REVIEW

A. Bansal et. al.[1] have researched various efficient and effective technologies which can detect frauds using machine learning and anomaly-detection methodologies. According to them, machine learning was applied for categorization of sensitive and suspicious transactions amongst internal and external credit card frauds. They have defined internal credit card fraud as transactions which take place by faking the approvals of the card owner as well as the bank using false identities. On the other hand, eternal credit card frauds refer to fraudulent transactions conducted through information theft. Credit card frauds are divided into the following types- application fraud,card not present(CNP), card id theft, stolen or lost card, Electronic Card, Phishing, Skimming, Account Takeover. The evolving technology makes retrieval of sensitive card and profile information easy for fraudsters where the commonly targeted information is the CVV number and expiry date. For research purposes, a large imbalance Kaggle dataset was chosen upon which Principal Component Analysis(PCA),due to its low computational and memory constraints, and Anomaly Detection Algorithm was performed for data analysis. Numerous machine learning algorithms, namely- Artificial Neural Networks(ANNs), Random Forest (RF), Support Vector Machine (SVM), Genetic Algorithm(GA), K-Nearest Neighbor (KNN), Bayesian Network, Decision Tree (DT) and Logistic Regression (LR) were trained and tested using the data for fraud classification, out of which SVM classifier gives the best results with a high accuracy of 91%.

An efficient method to detect frauds is to analyze the card holder's spending pattern to classify whether the transaction is fraudulent. The anomalies present within the dataset may hinder the model's predictive accuracy , which can be resolved by one-hot-encoding the categorical attributes and dates. [2]A. Singh has tried to utilize features such as the average and total amount of money spent by the card holders to track their spending pattern in the past 24 hours and has also kept track of the transaction time as most fraud transactions take place during night time. He has compared several classifier models to group transactions into 14 purchase categories, yielding a high result of 99.87% by using catboost.

V. B. Nguyen [3] states the main goal of fraud detection systems is real time fraud detection during payment processing and posterior detection to prevent further breach of security by blocking the card in use. According to the authors the constant evolution of technology makes it difficult to have a fixed strategy for prevention of frauds. The definition of a fraudulent transaction has been stated as an unusual change in behavior from the user's normal spending. To perform an in depth analysis based on the bandwidth and duration of transactions, a posterior fraud detection system is used. The model created consists of a real time detection system and a posterior detection system. The real time detection disables the card during the transaction when a fraud is detected while posterior detection ensures that only genuine transactions are completed after processing by using detailed algorithms which can detect frauds with higher accuracy rates and obtaining verification from human authorities in case a fraud is detected. The data used in the research was obtained from Kaggle, consisting of about 300000 anonymized transactions, information on real transactions from renowned personnel in the payment industry and a private dataset from the time range January to April, 2017. Statistical information related to the time delay between transactions is retrieved from the dataset and an estimate of the effect of the delay on transactions is calculated. A bidirectional LSTM classifier is applied to the dataset to classify the transactions with respect to 2 adjacent transactions committed by the card user most recently along with other features in the dataset. The model built is later on compared with other basic machine learning models like random forest and LSTM. Cross-entropy and Adam optimizer have been used for training the models.

There is reluctance amongst banks and other financial institutions to share information to protect their clients' information and other legal reasons. The authors [4] have used a dataset consisting of 2 day transactions from a bank with data related to closed cards to respect the clients' security. They have built and compared 2 different machine learning models for anomaly detection, namely- Local Outlier Factor and Isolation Forest Algorithm. The local outlier factor is an unsupervised algorithm used to measure local deviation based on the anomaly scores of the data samples. The local values of each sample has been calculated via K nearest neighbors algorithm.Through observation of the output visualized data, the samples with lower local values have been considered as anomalies in comparison to its neighboring values. The isolation forest algorithm implemented selects split values from a selected feature's range. Recursive partitioning through random forest trees is used for isolating a sample to retrieve a solution path. From the paths generated the shorter paths are considered as anomalies. For faster processing of the data, only a small fraction of the entire dataset was implemented for testing purposes, which gave a high accuracy of 99.6% and above.

Large datasets classified with domain expertise can be used for enhanced classification of fraud detection. Though various classifiers have been implemented for fraud detection, the use of imbalance datasets have always hindered the performance of the models.The authors have used Convolutional Neural Networks with Adaptive Synthetic (ADASYN) sampling technique (ADASYN) as an extension on a previously done research on credit card fraud detection using 3 evaluation metrics, for forming a balanced dataset by oversampling the minority class, in this case, the fraudulent class. CNN is then applied to the preprocessed and balanced dataset as well as the unbalanced dataset to compare the results. The evaluation metrics applied on the results were confusion matrix, accuracy, precision and recall, which helped conclude that the model using a balanced dataset displayed higher performance in terms of accuracy.[5]

The use of Recurrent Neural Networks (RNN) in combination with GRU, proposed by B. Branco et. al. [6] for real time detection of fraudulent transactions using the payment history of the card in the form of unrestrained uneven sub-sequences. They have attempted to deploy a machine learning model capable of classifying card frauds without the use of expensive techniques such as feature engineering. The focus of the research was to address the delay in occurring in between transactions. The model comprises data preprocessing, training of the dataset as well as batch prediction, which are conducted offline with z-scoring and percentile bucketing utilized for calculation of numeric parameters and prediction of incoming transactions, conditioned with time constraints and high workload done online. The concept of Bayesian probability is taken into account to assume that the ruling of an event can be dependent on the encryption of past events. For updating the status of an event, GRUs were chosen as recurrent units. As GRUs consider time intervals between transactions as constant units, entity grouping is done based on the transactions' chronological order, and any repetitive patterns within these time intervals over a period of time will be deemed as fraudulent with its irregularity taken into account. The chosen dataset includes transactions which have both card present and card not present data which is trained via binary-cross entropy. It was observed that the model showed better performance when trained with transactions where the card was present, followed by the latter. The final product of the research was a streaming engine called Pulse which is capable of online fraud detection which took a maximum duration of 200ms to complete a transaction.

J. R. D. Kho et al. [7] states that the most frequent form of fraud occurrences are counterfeit credit cards, theft, no-card fraud, mail phishing and identity theft fraud. He built an anomaly detection model which identifies credit card transactions by monitoring the usage of credit cards by their owners. The research made use of a dataset based on previous credit card frauds and the transaction log files from these cases. Several algorithms were applied on the preprocessed dataset out of which the Random Tree classifier gave a high accuracy rate of 94.32%.

In e-commerce websites the most predominant problem is credit card fraud. This causes money loss and identity theft. O. Adepoju et. al. [8] in their paper states that extracting information for detection of frauds is very easy but actually analyzing a fraud requires proper dataset and fraud detecting techniques. There credit card fraud is divided into two types: card present and not present. Card present means that the card is stolen. Credit card fraud can take place even without the presence of the card, only the CVV and the card number are sufficient to commit the fraud. For this purpose, a data set from Kaggle contains data of 3075 transactions with 12 features. Naive Bayes, Support vector machine, K-nearest neighbor and Logistic regression are used. Logistic regression has the highest accuracy.

P. Tomar ,S. Shrivastava and U. Thakar [9] in their paper say that ensemble learning is a better approach than fraud detecting systems. Credit card frauds can be detected by using the card holder's transaction history. The previously used rule based technique doesn't work with today's advanced credit card frauds. In solution to this hybrid techniques are used. And so, ensemble technique is used. With it, Ensemble learning along with hard voting, naïve bayes,logistic regression and decision tree are used. Ensemble learning along with hard voting shows highest accuracy. The data set was obtained from Kaggle. Out of 284,807 transactions 492 were fraud.

Detecting group spam is more difficult than detecting individual fraud because of spam data and the variations in the group dynamics. S. Dhawan, S. C.Reddy Gangireddy , S.Kumar and T. Chakraborty [10] in their paper propose a DeFrauder method. This is an unsupervised technique for detecting frauds online from review groups. First, it detects the fraudulent groups by using a product review graph and using many behavioral signals that will model complicated collaborations amongst reviewers. It assigns a high score to the spam reviews by mapping all the reviews in an embedding space. 4 datasets are used in this paper. The DeFeaudes baseline with a 17.11% greater NDCG@50 amongst the datasets.

The number of fraudulent transactions are increasing on a large scale today and so financial institutions have a vital role to play in detecting the frauds. Fraud tendencies are detected from behavior sequencing recently. W. Lin, L.Sun, Q. Zhong , C.Liu, J. Feng , X. Ao , H.Yang [11] in their paper use a multi-scale sequencing behavior to ignore the native structure of the web pages and model called SAH-RNN that consumes the sequential behavior for fraud detection of payments online. A dual attention is developed to record the information in sequential form from web pages granularities. The dataset used is used from Alibaba and the model outshines state-of-the-art approaches.

I.Benchaji , S. Douzi, B. E. Ouahidi and J. Jaafari [12] in their paper state that they make a model for detecting credit card frauds on the basis of modeling data in sequential manner. For this LSTM and attention mechanism were used. A deep recurrent neural network was used as well. The method proposed predicts high rates of fraudulent transactions from a sequence of input transactions that use a classifier to identify such transactions. The model is made robust by the usage of three approaches : uniform manifold approximation (UMAP) and projection UMAP for the selection of features that are predictive, Long Short Term Memory(LSTM) that uses sequencing of transactions and attention mechanisms. The model is efficient and effective.

V. N. Dornadula, S Geetha in their paper [13] detect frauds by solving the problem of concept drift and imbalance in data sets. The paper aims to build a fraud detection model for transactional data with an intention of analyzing the transactions of the past and extracting the behavioral patterns from it. The method of clustering is used to divide the cardholders into various clusters. The transactions are aggregated by using the method of slide window. SMOTE (Synthetic Minority Over-Sampling Technique) is applied to the dataset. A feedback mechanism is used as a solution for concept drift problems. Various classification methods are applied to get good accuracy. Local Outlier factor, Decision tree, Logistic regression, Isolation forest, and random forest are the classification algorithms used to find the best suited algorithm.

A.D.Pozzolo, G. Boracchi, O.Caelen, C.Alippi [14] in their paper talk about detecting credit card frauds. It states the issues about how the fraudsters have evolved over time and the very little transactions are investigated on time. The paper suggests building a model to analyze a large number of frauds on a day-to-day basis. The dataset used is of ecommerce websites from European countries with 75 million transactions or more. The performance measure used is Mann–Whitney statistic and average precision. Several measures were used to alert the system of fraud based on the details of the transaction based on the pattern of customers transactions, whether the transactions were genuine or not and those investigated on time.

Another approach towards detecting frauds in credit card transactions include a hybrid data analysis technique proposed by the M. Zanin et. al [15]. The idea involves the use of parenclitic networks, a method of dataset representation in numerical form, on complex attributes which represent credit card transactions. The authors have highlighted the differences between using supervised and unsupervised machine learning methods for fraud detection stating that a supervised approach is more accurate in comparison with an unsupervised technique which follows the patterns provided by the user and is unable to detect new anomalies based on the provided data. They have compared the performances of traditional data mining techniques to that of a model using feature extraction through parenclitic networks and used artificial neural networks(ANN) for classification. The application of evaluation metrics revealed that the application of parenclitic networks with ANN showed higher accuracy than those of traditional methods.

IMPLEMENTATION

Experimental setup

The execution of the models has been carried out using Jupyter Notebook. The required packages and libraries were installed via the pip command. For deploying the full-stack application created, a virtual environment was set up in Visual Studio using the Anaconda library. The back-end file consisted of Python scripts run using the command-line in the Powershell terminal. The website is deployed in google chrome.

Data Preprocessing

The dataset contains 284,807 transactions out of which 492 are fraudulent transactions and rest are legitimate. The dataset is imbalanced. In order to make it balanced we have taken 492 legitimate data samples and concatenated with fraudulent transactions and a balance dataset is created.

Streamlit

The user interface of the model has been created using a Python library named streamlit which makes use of Python functions to take input and print output onto the web application. The execution of the program to run the web application is done using Powershell by writing the relevant execution command.

Implemented GUI

The GUI of the proposed system is a web application that takes the time of the transaction, 31 transaction features and the amount of money sent in total as comma-separated inputs from a user and returns the predicted class the transaction belongs to upon clicking the submit button.

ARCHITECTURE DIAGRAM

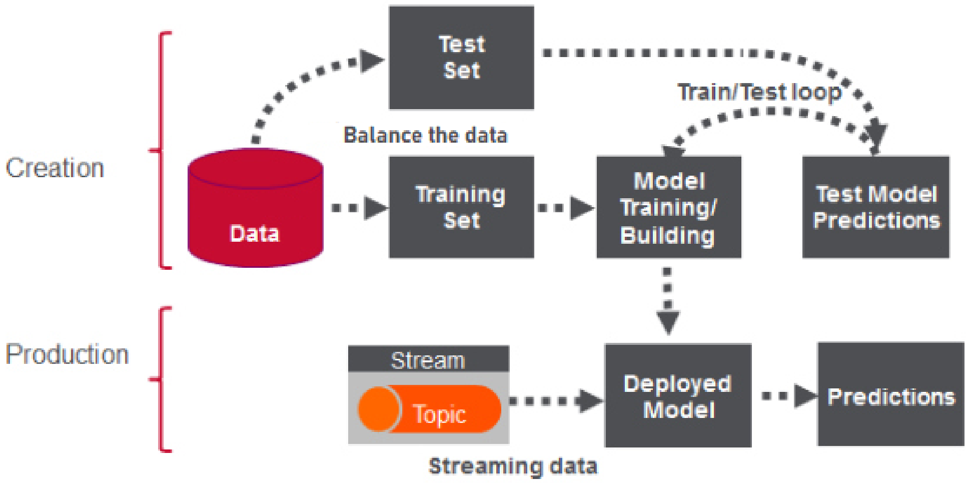


Figure 1. Architecture diagram of the proposed system

ALGORITHM

The class label is taken as ‘y’, the dependent variable whereas the rest of the labels are taken as the dependent variable,x. A train-test split of the ratio 8:2 is done. The training dataset is then used for training Logistic regression, Supervised Vector Machine (SVM), and XGBoost classifiers, after which the testing data is passed through each of them separately. Out of the created training models, a separate ensemble model is also built which consists of 2 layers, where the first layer is made of Logistic Regression and SVM classifier models and the second layer consists of an XGBoost classifier model to form a 2 layer stacking ensemble model which combines the results of the first layer models’ accuracies and passed them through the final layer of the model.

LOGISTIC REGRESSION

Logistic regression is a method for the prediction of labels of continuous data by computing the total numerical values of the input features to give a numerical value between 0 and 1 as the predictive value using logistics.

The dataset is split into train and test data and the function used is,
lr = LogisticRegression(solver='lbfgs', max_iter=1000) used for logistic regression.

SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) classifier is a supervised learning algorithm for classification of data into 2 labels. It makes use of data points plotted using the features from the derived dataset. A hyperplane acts as a decision boundary between the data points to segregate them into 2 groups.

The dataset is split into train and test data and the function
svm = SVC(gamma = 'auto', kernel = 'linear', decision_function_shape='ovo') for svm.

XGBoost

XGBoost or eXtreme Gradient Boosting is a classification algorithm used for supervised learning. It utilizes concepts derived from gradient boosting and decision trees. XGBoost makes use of a stack of decision trees where the output of one decision tree is passed as input to another until there are no new outputs formed. Although the process resembles the methodology of Random Forests, it does not make use of any assigned weight like Random Forest does.

The dataset is split into train and test data and the function for fitting and testing is,
xg = xgb.XGBClassifier(objective= 'binary:logistic', n_estimators = 10, seed= 123)
for xgboost.

ENSEMBLE LEARNING

Ensemble learning refers to combining various predictive models to create an enhanced learning model. For this project, we have created an ensemble model of 2 levels where the first layer is made of Logistic Regression and SVM classifier models and the second layer consists of an XGBoost classifier model to form a 2 layer stacking ensemble model which combines the results of the first layer models’ accuracies and passed them through the final layer of the model..

The dataset is split into train and test data and the functions:
estimators= [('svm', svm), ('logreg', lr)]
stack = StackingClassifier(estimators=estimators, final_estimator=xg)

RESULT AND DISCUSSION

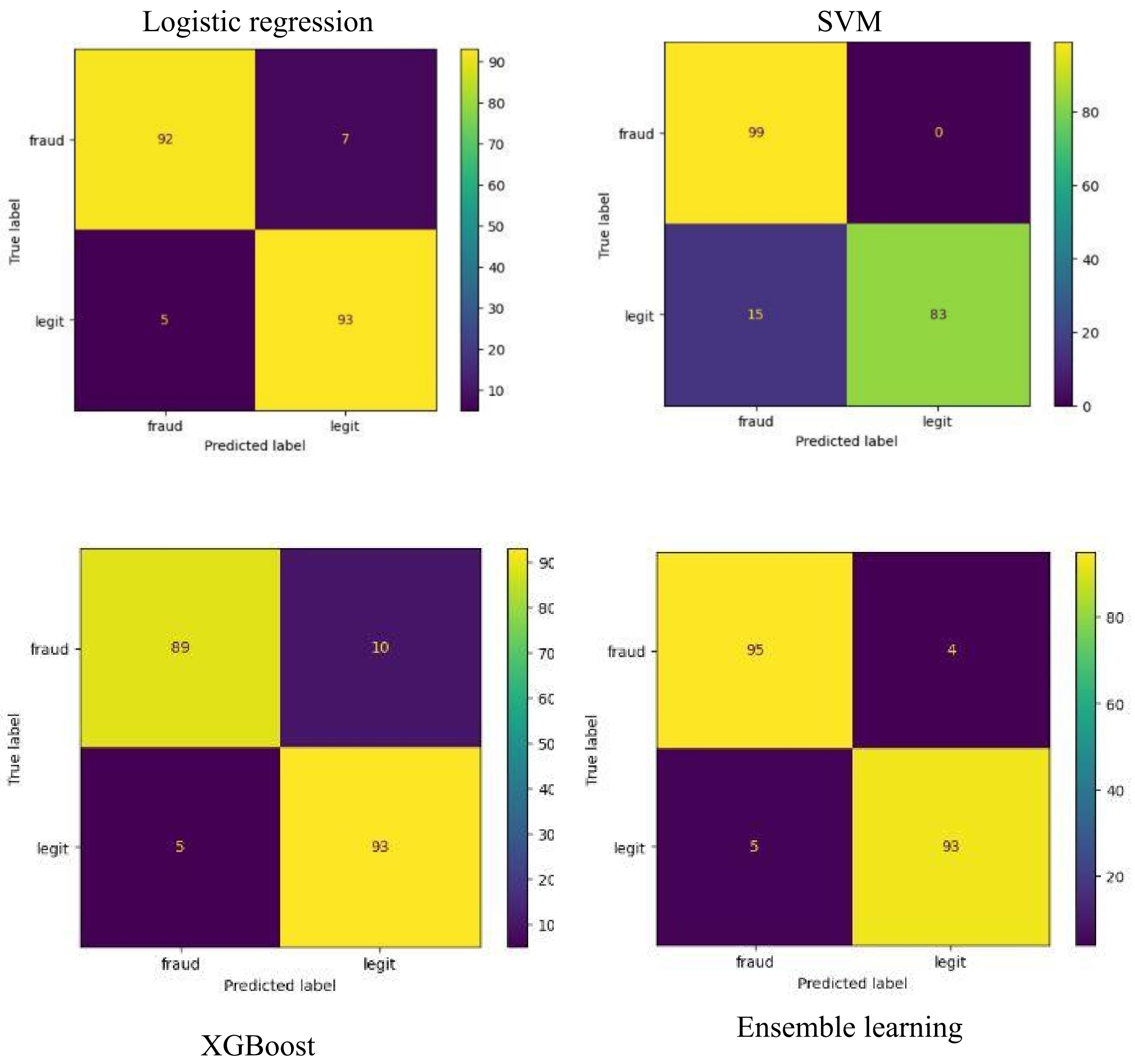
Evaluation

For measuring the performance of the built models, a classification report, as well as a matrix is created for each one of them. The considered metrics for comparing the performance of the models are accuracy and f1-score of each class.

	Model	SVM	Logistic Regression	XGBoost classifier	Stacking ensemble model
Precision	Legit	0.87	0.95	0.95	0.95
	Fraud	1.00	0.93	0.90	0.96
Recall	Legit	0.92	0.93	0.92	0.95
	Fraud	0.92	0.95	0.92	0.95
F1-score	Legit	0.93	0.94	0.92	0.95
	Fraud	0.92	0.94	0.93	0.95

Table 1. Accumulated classification report scores for the constructed models.

The precision of the fraudulent transactions is 1 for SVM indicating that the model predicted all fraudulent transactions correctly as per the goal. Logistic regression, XGBoost, and the ensemble model gave similar precision values while SVM gave 0.87 thereby exhibiting better performance in the classification of the transactions. The recall score for the ensemble model is highest for both fraud and legitimate transactions, hence, the majority of the transactions were predicted correctly concerning their true labels. In terms of the F1-score which combines the precision and recall of the fraudulent transactions to give accuracy, ensemble learning displayed the best results too.


Logistic regression

SVM

XGBoost

Ensemble learning

Figure 2. Classification matrix for each model.

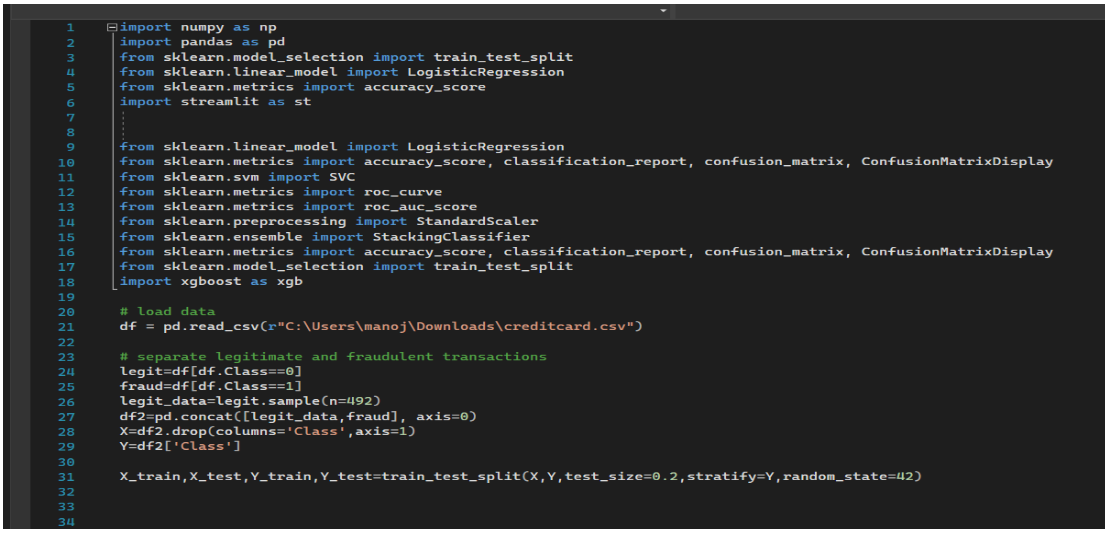
Model	SVM	Logistic Regression	XGBoost classifier	Stacking ensemble model
Accuracy	92%	94%	92%	95%

Table 2. Accuracies of each model

Stacking ensemble learning has highest accuracy compared to the other models.

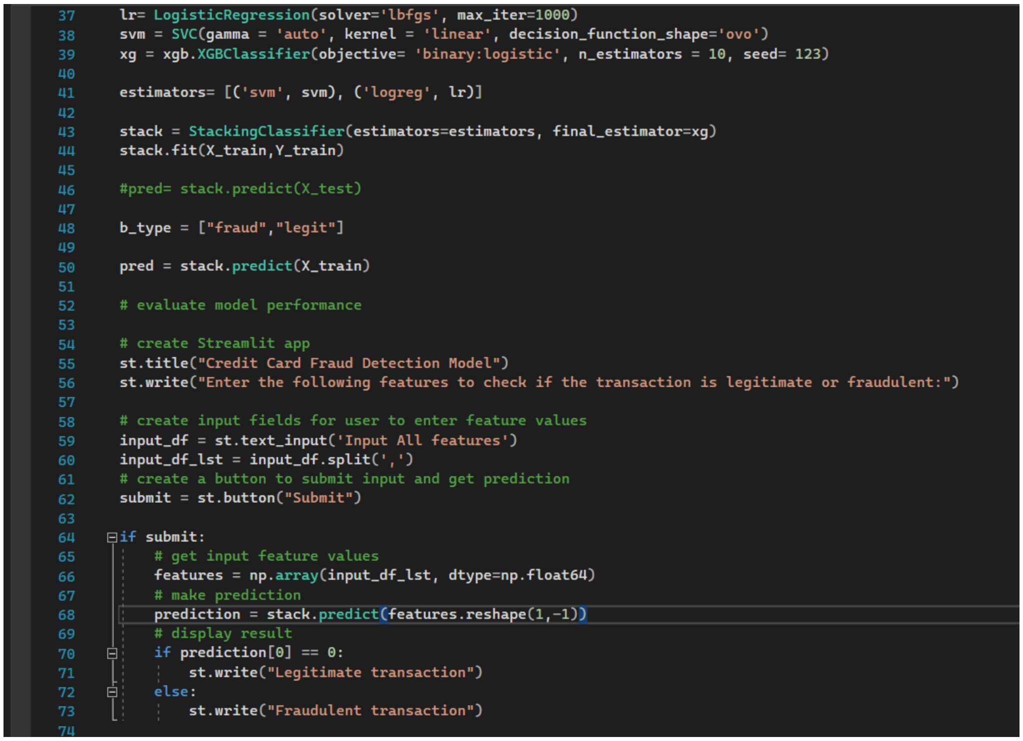
Comparing our model with the already existing model in paper [9] where the model uses ensemble learning with hard voting. The accuracy of that model is 92%. The model proposed in this paper shows an accuracy of 95% which is more than the model in paper [9]. Their model uses ensemble learning with hard voting. The method of ensemble learning with stacking improves the prediction of the models. The method of ensemble voting involves combining predictions of many models by voting.

SCREENSHOTS OF THE CODE AND WORKING OF GUI



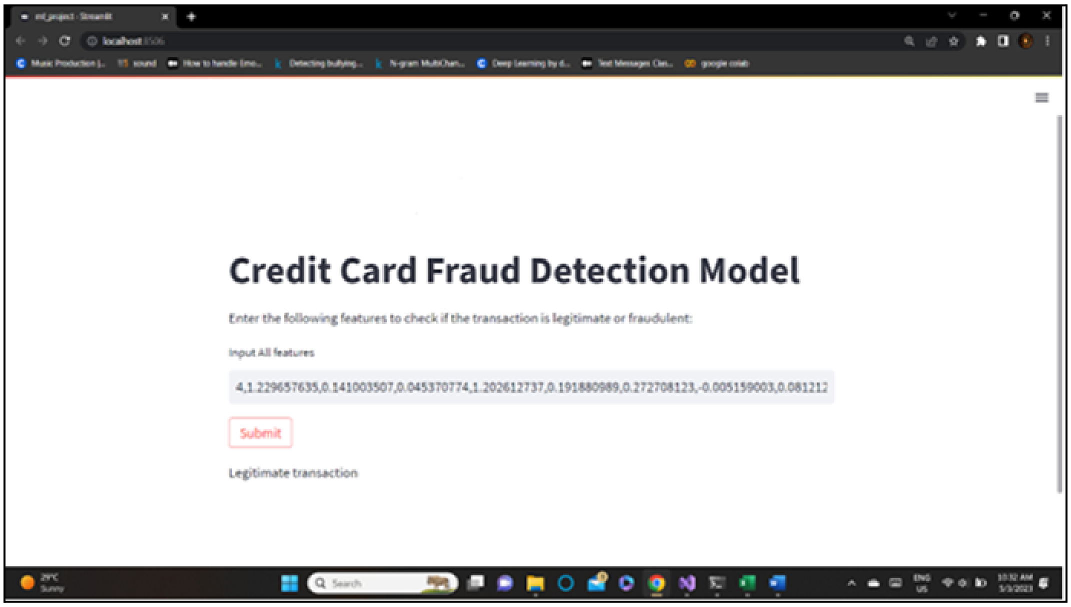
```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import streamlit as st


from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
from sklearn.svm import SVC
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import StackingClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
from sklearn.model_selection import train_test_split
import xgboost as xgb

# load data
df = pd.read_csv(r"C:\Users\manoj\Downloads\creditcard.csv")

# separate legitimate and fraudulent transactions
legit=df[df.Class==0]
fraud=df[df.Class==1]
legit_data=legit.sample(n=492)
df2=pd.concat([legit_data,fraud], axis=0)
X=df2.drop(columns='Class',axis=1)
Y=df2['Class']

X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,stratify=Y,random_state=42)
```

Figure 3. Screenshot of the code 1



```python
lr= LogisticRegression(solver='lbfgs', max_iter=1000)
svm = SVC(gamma = 'auto', kernel = 'linear', decision_function_shape='ovo')
xg = xgb.XGBClassifier(objective= 'binary:logistic', n_estimators = 10, seed= 123)

estimators= [('svm', svm), ('logreg', lr)]

stack = StackingClassifier(estimators=estimators, final_estimator=xg)
stack.fit(X_train,Y_train)

#pred= stack.predict(X_test)

b_type = ["fraud","legit"]

pred = stack.predict(X_train)

# evaluate model performance

# create Streamlit app
st.title("Credit Card Fraud Detection Model")
st.write("Enter the following features to check if the transaction is legitimate or fraudulent:")

# create input fields for user to enter feature values
input_df = st.text_input('Input All features')
input_df_lst = input_df.split(',')
# create a button to submit input and get prediction
submit = st.button("Submit")

if submit:
    # get input feature values
    features = np.array(input_df_lst, dtype=np.float64)
    # make prediction
    prediction = stack.predict(features.reshape(1,-1))
    # display result
    if prediction[0] == 0:
        st.write("Legitimate transaction")
    else:
        st.write("Fraudulent transaction")
```

Figure 4. Screenshot of the code 2.



Credit Card Fraud Detection Model

Enter the following features to check if the transaction is legitimate or fraudulent:

Input All features

4,1.229657635,0.141003507,0.045370774,1.202612737,0.191880989,0.272708123,-0.005159003,0.081212

Legitimate transaction

Figure 5. Screenshot of the deployed webapp

FLOWCHART

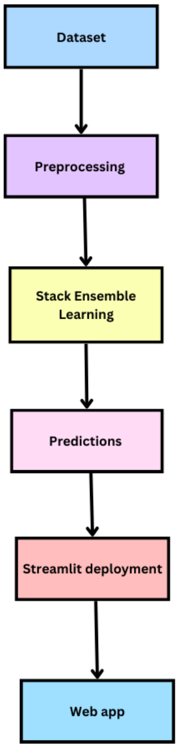


Figure 6. Flowchart of the project model.

CONCLUSION

A web application has been developed for classifying transactions as fraud and legitimate. To understand the application of machine learning in fraud detection, literature review of 10 to 20 papers was conducted. A dataset consisting of bank transactions was chosen for the project. The classifier used in the web application was chosen by building and comparing the performance of different machine learning models. The application is deployed on an internet browser where the users can interact and verify their transactions. The resultant web application displays outstanding performance with an accuracy of 95%. However, the runtime for training the model is high. For the future, model runtime can be enhanced by experimenting with deep learning models concepts such as hybrid learning.

CONTRIBUTIONS

Megha Manoj

- Abstract
- Literature Review
- Dataset description
- Objectives
- Methodology
- Results and discussion
- Flow chart
- Screenshots of code and GUI
- Conclusion
- Paper Summary

Chelsy Fernandes

- Introduction
- Literature review
- Problem statements
- Methodology
- Paper summary
- Architecture Diagram
- Algorithm
- Results and discussion

REFERENCES

[1]A. Bansal and H. Garg, "An Efficient Techniques for Fraudulent Detection in Credit Card Dataset: A Comprehensive Study," IOP Conference Series: Materials Science and Engineering, vol. 1116, no. 1. IOP Publishing, p. 012181, Apr. 01, 2021. doi: 10.1088/1757-899x/1116/1/012181.

[2]A. Singh, A. Singh, A. Aggarwal, and A. Chauhan, "Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, Nov. 16, 2022. doi: 10.1109/iceccme55909.2022.9988588.

[3]V. B. Nguyen, K. G. Dastidar, M. Granitzer, and W. Siblini, "The Importance of Future Information in Credit Card Fraud Detection." arXiv, 2022. doi: 10.48550/ARXIV.2204.05265.

[4]S P Maniraj, Aditya Saini, Shadab Ahmed, and Swarna Deep Sarkar, “Credit Card Fraud Detection using Machine Learning and Data Science,” International Journal of Engineering Research and, vol. 08, no. 09. ESRSA Publications Pvt. Ltd., Sep. 13, 2019. doi: 10.17577/ijertv8is090031.
[5]M. L. Gambo, A. Zainal, and M. N. Kassim, “A Convolutional Neural Network Model for Credit Card Fraud Detection,” 2022 International Conference on Data Science and Its Applications (ICoDSA). IEEE, Jul. 06, 2022. doi: 10.1109/icodsa55874.2022.9862930.

[6]B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, “Interleaved Sequence RNNs for Fraud Detection,” Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &amp; Data Mining. ACM, Aug. 20, 2020. doi: 10.1145/3394486.3403361.

[7]J. R. D. Kho and L. A. Vea, “Credit card fraud detection based on transaction behavior,” TENCON 2017 - 2017 IEEE Region 10 Conference. IEEE, Nov. 2017. doi: 10.1109/tencon.2017.8228165.

[8]O. Adepoju, J. Wosowei, S. lawte and H. Jaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques," *2019 Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2019, pp. 1-6, doi: 10.1109/GCAT47503.2019.8978372.

[9] P. Tomar, S. Shrivastava and U. Thakar, "Ensemble Learning based Credit Card Fraud Detection System," 2021 5th Conference on Information and Communication Technology (CICT), Kurnool, India, 2021, pp. 1-5, doi: 10.1109/CICT53865.2020.9672426

[10] S. Dhawan , S. C.Reddy Gangireddy , S.Kumar and T. Chakraborty, “Spotting Collective Behaviour of Online Frauds in Customer Reviews”, (July 2019)

[11] W. Lin, L.Sun, Q. Zhong , C.Liu, J. Feng , X. Ao , H.Yang, "Online Credit Payment Fraud Detection via Structure-Aware Hierarchical Recurrent Neural Networks." July 2021. doi:10.24963/ijcai.2021/505.

[12] I.Benchaji , S. Douzi, B. E. Ouahidi and J. Jaafari, “Enhanced credit card fraud detection based on attention mechanism and LSTM deep model.” December 2021 doi:10.1186/s40537-021-00541-8

[13] V. N. Dornadula, S Geetha, “Credit Card Fraud Detection using Machine Learning Algorithms”, Volume 165, 2019, Pages 631-641, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.01.057.

[14] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784-3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643

[15] M. Zanin, M. Romance, S. Moral, and R. Criado, “Credit Card Fraud Detection through Parenclitic Network Analysis,” Complexity, vol. 2018. Hindawi Limited, pp. 1–9, 2018. doi: 10.1155/2018/5764370.

[16] https://docs.streamlit.io/
[17] https://machinelearningmastery.com/extreme-gradient-boosting-ensemble-in-python/
[18] https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

PAPER SUMMARY

Sl no	Reference	Objective	Problem Statement	Methodology	Dataset	Algorithm	Advantage	Disadvantage	Performance Measure value
1.	A. Bansal and H. Garg, "An Efficient Techniques for Fraudulent Detection in Credit Card Dataset: A Comprehensive Study," IOP Conference Series: Materials Science and Engineering, vol. 1116, no. 1. IOP Publishing, p. 012181, Apr. 01, 2021. doi: 10.1088/1757-899x/1116/1/012181.	Conduct a research on credit card fraud detection and identify the various techniques used.	Create a synopsis of distinct techniques used for credit card fraud detection to alert the card owner.	Build machine learning models using Anomaly Detection Algorithm(ADA), Artificial Neural Networks(ANNs), Genetic ,Random Forest (RF), Bayesian Network, Algorithm(GA), Logistic Regression (LR), K-Nearest Neighbor (KNN), Principal Component Analysis(PCA), Anomaly Detection Algorithm(ADA), Support Vector Machine (SVM), Decision Tree (DT), Artificial	A larger, imbalanced Kaggle dataset is used with the aim of analyzing various features present in the dataset.	Principal Component Analysis(PCA), Anomaly Detection Algorithm(ADA), Artificial Neural Networks(ANNs), Genetic Algorithm(GA), Random Forest (RF), Bayesian Network, Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR),	The Supervised Vector Machine classifier was able to attain a high accuracy of 91%.	The dataset is not balanced.	Logistic Regression (LR), Artificial Neural Networks(ANNs), Random Forest (RF), Support Vector Machine (SVM), Genetic Algorithm(GA), Bayesian Network, Decision Tree (DT), K-Nearest Neighbor

				Neural Networks(ANNs) , Genetic Algorithm(GA),Random Forest (RF), Support Vector Machine (SVM), Bayesian Network, Decision Tree (DT), Logistic Regression (LR), K-Nearest Neighbor (KNN).		K-Nearest Neighbor (KNN).			(KNN) were trained and tested using the data for fraud classification, out of which SVM classifier gives the best results with a high accuracy of 91%.
2.	A. Singh, A. Singh, A. Aggarwal, and A. Chauhan, “Design and Implementation of Different Machine Learning Algorithms for Credit Card	Build and compare various machine learning models for detecting credit card fraud.	Discover an efficient model for detecting credit card fraud.	The dataset is preprocessed ,analyzed and resampled to resolve imbalances in the categorical values. The average amount spent, by the user along with the time of transaction are used as features	A classified dataset from kaggle for credit card fraud detection is used. There are 14 unique purchasing classes.	Logistic regression, decision tree, random forest, Catboost .	Catboost showed high accuracy of 99.87 inc comparison to all the other algorithms.	Hard to predict if user has an irregular pattern of spending.	Accuracy was taken as a performance measure metric on logistic regression (93.70%), , random forest(99.60%) , Catboost classifier(

	Fraud Detection,” 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, Nov. 16, 2022. doi: 10.1109/iceccme55909.2022.9988588.			to categorize test records using classifier models.					99.87%), Decision tree(99.40%).
3.	V. B. Nguyen, K. G. Dastidar, M. Granitzer, and W. Siblini, “The Importance of Future	Construct a posterior detection model for fraud detection	Build and analyze the performance of a posterior detection system with that of other basic anomaly	Build a real time detection system model and pair it with a posterior detection system based on transaction delays. Three models are built	The data used in the research was obtained from Kaggle, consisting of about 300000	Bi-LSTM, LSTM, Random forest.	The Bi-LSTM model displayed an improvement in fraudulent detection at an early stage in	The results obtained are difficult to compare due to differences in the context of transactions utilized for testing.	Area Under Precision-Recall curve (AUPRC) was used to evaluate all 3 models.AUPRC is

	Information in Credit Card Fraud Detection.” arXiv, 2022. doi: 10.48550/ARXIV.2204.05265.		detection models for fraudulent transaction detection.	using Bi-LSTM, LSTM and Random forest, and their performances are compared using a sequence of transactions.	anonymized transactions, information on real transactions from renowned personnel in the payment industry and a private dataset from the time range January to April, 2017. The entire dataset is labeled by human experts within the field and has a small imbalance of 0.0263% .		transactions.		calculated and maximized during validation by performing a random search over the hyperparameters present in the validation dataset. The evaluation of the models using 2 months worth transaction data through a sequence method of past-present-future digits, where each token represented the number of

									transactions within that time period, revealed a better performance in the LSTM models in comparison to that of Random Forest.
4.	S P Maniraj, Aditya Saini, Shadab Ahmed, and Swarna Deep Sarkar, “Credit Card Fraud Detection using Machine Learning and Data Science,”	Detecting fraudulent transactions with a minimum accuracy error.	Build and compare the performance of isolation forest and Local outlier factor algorithm for credit card fraud	Various visualization techniques are used to analyze the data for preprocessing inconsistencies. Local outlier factor is applied to obtain the	The dataset used is based on 2 days transaction details and is obtained from kaggle and consists of 31 columns out	Isolation, forest Local outlier factor algorithm.	The algorithm achieves high accuracy of over 99.6%.	The dataset used cannot be used on a commercial scale as it only contains 2 days worth transaction data.	The dataset was able to achieve an increment in precision by 33% and an accuracy over 99.6%.

	International Journal of Engineering Research and, vol. 08, no. 09. ESRSA Publications Pvt. Ltd., Sep. 13, 2019. doi: 10.17577/ijertv8is090031.		detection.	anomaly score of each sample. Isolation forest is also implemented on the dataset to detect anomalies by finding the short paths for anomalies via random forest trees.	of which 28 (v1-v28) are used for masking sensitive data. The remaining columns contain time, amount spent and the class-valid and fraudulent transactions.				
5.	M. L. Gambo, A. Zainal, and M. N. Kassim, “A Convolutional Neural Network Model for Credit Card	Improve financial losses and customer confidence through automated fraud detection.	Build a convolutional neural network with ADASYN for detecting credit card fraud.	Dataset is preprocessed and Adaptive Synthetic sampling technique (ADASYN) is used to correct the imbalance present within the	European CardHolders dataset from kaggle consisting of 30 features describing transactions , time and amount. A	Convolutional neural network (CNN) applied to the balanced dataset as well as the imbalance dataset for	The balanced dataset showed major improvement in prediction.	Accuracy of the model using balanced dataset was lower during training and testing.	The model achieves an accuracy of 99.99% n prediction of fraudulent transactions.

	Fraud Detection," 2022 International Conference on Data Science and Its Applications (ICoDSA). IEEE, Jul. 06, 2022. doi: 10.1109/icodsa55874.2022.9862930.			dataset by oversampling the minority class.	categorical column represents whether the record consists of a fraudulent transaction through binary classification.	classification.			
6	B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, "Interleaved Sequence RNNs for Fraud Detection,"	Build a machine learning model for fraud detection which is suitable for systems with	Construct a fraud detection model capable of offline training and online streaming classification.	A multi-sequence RNN is constructed using GRU via batch training techniques, a streaming inference technique and evaluation of the	Two datasets used which consist of real data from 2 European financial institutions. It has records present	Sequence RNN,GRU,Z-scoring, Bayes Theorem, binary cross-entropy.	The cyclic nature of RNN detects patterns eliminating the need for feature engineering.	The processing time of a transaction is very strict due to direct application during a transaction.	Efficiency of the model was tested by holding a transaction using TensorFlow Serving host(TCX). It was

	Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &amp; Data Mining. ACM, Aug. 20, 2020. doi: 10.1145/3394486.3403361.	millisecond latency requirements to address the irregular intervals between the usage of the credit card.		model built using real life cases to compare the model evaluations against other classification models.	where the credit in use was physically present and not present.				observed that the latency was high in percentile but the inference latency of the transactions was deducted 5 times by compilation using TCX.
7	J. R. D. Kho and L. A. Vea, “Credit card fraud detection based on transaction behavior,” TENCON 2017	Build a detection model to predict anomalous credit card transactions.	Use credit card transaction logs and case files to differentiate legitimate and illegitimate	The dataset is preprocessed and additional records are added to account for its imbalance. A feature selection is performed to understand which	The dataset used consists of files on previously recorded credit card fraud incidents and the	Several classifiers such as Naive Bayes,, Random Tree, Bayes Net, J48 libSVM, and	Reduce expenditure of banks in communication by sending alert messages when an anomalous	Details from the confidentiality agreement between the bank and client.	An accuracy rate of 94.67% was observed from the application of Random Tree.

	- 2017 IEEE Region 10 Conference. IEEE, Nov. 2017. doi: 10.1109/tencon. 2017.8228165.		transfers.	attributes are most significant. Several classifiers are tested on the dataset and evaluation metrics are applied.	transaction logs related to these cases.	MOLEM Were applied to the preprocessed dataset.	transaction is detected.		
8	O. Adepoju, J. Wosowei, S. lawte and H. Jaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques," *2019 Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2019, pp. 1-6, doi: 10.1109/GCAT	Does a comparison to identify credit card fraud by using logistic regression based on biased information, naïve bayes, k-nearest neighbor. Classification methods are used to classify whether the transactions are fraudulent	Classify the transactions- genuine or fraudulent and by evaluating different performance measures.	The dataset is preprocessed by converting the categorical data into integers. The data is then split into train and test data and machine learning algorithms like naïve bayes,logistic regression k-nearest neighbor,. The learning phase is to classify the transactions, and the test set is evaluated by using a confusion matrix.	data set from Kaggle contains data of 3075 transactions with 12 features.	Logistic Regression, K-Nearest Neighbor , Support Vector, And Machine, Naive Bayes	Logistic regression is very accurate.	The data keeps changing over time which causes authentic transactions to be considered fraudulent as compared to the actual frauds.	Support Vector Machine (97.53%), Naive Bayes (96.93%), Logistic Regression (99.07%) and K-Nearest Neighbor (96.91%).

	47503.2019.8978372.	or not.							
9	P. Tomar, S. Shrivastava and U. Thakar, "Ensemble Learning based Credit Card Fraud Detection System," 2021 5th Conference on Information and Communication Technology (CICT), Kurnool, India, 2021, pp. 1-5, doi: 10.1109/CICT53865.2020.9672426.	Detect fraudulent transactions by using ensemble learning along with hard voting.	1. Classify the transactions of the dataset to check their authenticity. 2. Use the best suited model to get high accuracy.	The dataset is first preprocessed and then split into train and test data. Ensemble learning is applied to it and hard voting is applied. Then various performance measures are applied on the dataset.	The data set is taken from Kaggle. Out of 284,807 transactions 492 were fraud.	naïve bayes,Decision tree , logistic regression and Ensemble learning along with hard voting .	The various machine learning algorithms increase the model's discriminative ability.	Class imbalance, Low prediction for a large dataset.	Decision tree (91%), naïve bayes (91%),logistic regression (92%) and Ensemble learning along with hard voting (92%).

10	S. Dhawan , S. C.Reddy Gangireddy , S.Kumar and T. Chakraborty, “Spotting Collective Behaviour of Online Frauds in Customer Reviews”, (July 2019).	1. A DeFrauder method is proposed which detects candidate fraudulent groups. 2. Measurement of fraud indicators from the candidate groups. 3. Giving spam score to the groups.	To label the reviews of customers as genuine or fraudulent.	Detection the frauds using the DeFrauder framework.	4 datasets are used: 1. From reviews of the play store. 2.Review of musical instruments from Amazon. 3. YelpNYC reviews from restaurants in New York. 4. YelpNYZip : aggregated reviews of restaurants in the same area.	NDCG@k , Defrauder.	High accuracy of defrauder model.	Generalization ability is limited.	Defrauder has 17.11% more accuracy than baseline NDCG@k.

11	W. Lin, L.Sun, Q. Zhong , C.Liu, J. Feng , X. Ao , H.Yang, "Online Credit Payment Fraud Detection via Structure-Aware Hierarchical Recurrent Neural Networks." July 2021. doi:10.24963/ijcai.2021/505.	Structure Aware hierarchical RNN (SAH-RNN) is used for multi-layer sequencing. A dual attention mechanism is designed for learning multiscale behavior of the users.	The credit card frauds through online payment is expressed as a binary sequencing problem.	Structure Aware Factor Sequence Extraction is applied along with SAH-RNN and the class labels are assigned to classify genuine and fraudulent transactions. Dual attention mechanisms are applied.	The data set is from e-commerce website Alibaba.	LSTM, GRU, TextCNN, AUC and R@P0.1, ON-LSTM, CW-LSTM.	Has better generalization capability as compared to other behavior sequences.	Baseline method performance as compared to other methods will deteriorate as in when the length of the sequence increases.	SAH-RNN outshines GRU by 8.42% increase, SAH-LSTM gets 3.24% greater AUC and 6.04% higher R@P0.1 than 3 layer stack LSTM. ON-LSTM outshines CW-LSTM.
12	I.Benchaji , S. Douzi, B. E. Ouahidi and J. Jaafari, “Enhanced credit card fraud detection based on attention	LSTM sequencing models and attention mechanism for finding correlation between the event which	Identify hidden relations between the transactions to classify it as a fraudulent transaction.	On the two given data sets feature selection and dimensionality reductions techniques are applied to preprocess the data and LSTM with attention mechanism is	The first data set was taken from Kaggle. Out of 284,807 transactions 492 were fraud.	Long short memory Term (LSTM) network.	LSTM gives better accuracy than other classification methods.	recurrent networks are used for processing sequences.	LSTM attention (96.7%) on dataset 1, LSTM attention (97.4%) on dataset 2.

	mechanism and LSTM deep model.” December 2021 doi:10.1186/s40537-021-00541-8.	could be far from the sequence which is given as input. To improve the classification tasks efficiency and to increase the detection of frauds from transactions.		applied to predict the fraudulent transactions.	Second data set has 594 thousand transactions made over 180 days.				
13	V. N. Dornadula, S Geetha, “Credit Card Fraud Detection using Machine Learning Algorithms”, Volume 165,	Extraction of behavioral patterns of past transactions to predict the frauds in the future.	1. Solve concept drift by feedback mechanism. 2. Extraction of behavioral patterns of past transactions.	PCA is applied for preprocessing. The method of clustering is used to divide the cardholders into various clusters. The transactions are aggregated by using the method	Dataset is from kaggle which contains 492 frauds out of 284,807 transactions.	Local Outlier factor Logistic regression, Isolation forest, Decision tree and random forest	Solves concept drift and imbalance of data.	NIL	Local Outlier factor (45.8%), Decision tree (97.08%), Isolation forest (58.8%), Logistic regression

	2019, Pages 631-641, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.01.057.			of slide window. Train and test the data set. SMOTE (Synthetic Minority Over-Sampling Technique) is applied to the dataset. Various classification methods are applied to get good accuracy.					(97.18%) and random forest (99.99%).
14	A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," in	Developing a classification model to address the Fraud detecting system by describing an alert feedback system with an appropriate	To reduce the credit card frauds by reducing class imbalance, Concept Drift and Alert–Feedback Interaction and Sample Selection Bias.	Data preprocessing, Feature Engineering, Training and testing of the model. Evaluation of the model by using confusion.	The dataset used is an ecommerce website of European cardholders.	AUC, Confusion matrix.	Performance of the model is improved by preprocessin te data.	NIL	The performance measure used is AUC.

	IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784-3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643.	performance measure.							
15	M. Zanin, M. Romance, S. Moral, and R. Criado, “Credit Card Fraud Detection through Parenclitic Network Analysis,” Complexity, vol. 2018. Hindawi Limited, pp.	Improve anomaly detection of fraud transactions with parenclitic networks.	To compare and analyze the performance of parenclitic networks against traditional machine learning techniques in detecting credit card	Two different methodologies are applied on the dataset-Feature extraction is performed on the dataset using parenclitic networks and traditional data mining techniques.	The dataset used is derived from the Spanish bank BBV between a time period 2 years from 20011 to 2012. It contains credit and debit card transactions classified into legal and illegal on the	The algorithms used as parenclitic networks, multilayer perceptrons in artificial neural networks (ANN).	The model is efficient at detecting certain kinds of fraudulent transactions such as online and medium sized transactions with improved scores.	Feature extraction through parenclitic networks does not provide a low classification error for large transactions. The model also provides no information on how effective it would be in a	The inclusion of parenclitic features into the raw dataset reduced the error rate from 19.2% to 12.23%. A false positive rate of below 4% was observed along with a Z-Score of

	1–9, 2018. doi: 10.1155/2018/5764370.				basis of customer complaints and an automated algorithm which rated the suspicion of transaction fraud on a scale of 1-100.			commercial system.	+1.723 from the Areas Under the ROC Curve (AUC) graph.