Government Polytechnic For Women

Sheshadri Road, Bengaluru-560001

Report on

Online payment Fraud Detection

Team Mates:

Varsha Reddy BV(119CS23705)

Madhu CM(119CS21028)

Kannika GV(119CS22023)

Final year in Computer Science Engineering

Under the guidance of:

Yatish.

Abstract:

The rise of the internet and e-commerce appears to entail the usage of online payment transaction. The increased usage of online payments is leading to a rise in fraud. However, as the number of online transactions increases, so does the number of fraud instances. Fraud detection is an important component of online payment systems since it serves to protect both customers and merchants from financial damages. In this project, we propose a fraud detection system for online payments that uses machine learning techniques to identify and prevent fraudulent transactions. Using Machine Learning algorithms, we can find unique data patterns or uncommon data patterns that will be useful in detecting any fraudulent transactions. The random Forest Classifier will be utilized to get the best results. Our approach strives to improve fraud detection accuracy while reducing the amount of false positives, resulting in a more efficient and effective method for identifying and combating fraud.

When it comes to the simplicity of making a payment while sitting anywhere in the world, online payments have been a source of attractiveness. Over the past few decades, there has been an increase in online payments. E-payments enable business earn a lot of money in addition to consumers. However, because electronic payments or so simple, there is also a risk of fraud associated with them. A consumer must ensure that the payment he is paying is going exclusively to the appropriate service provider. Online fraud exposes users to the possibility of their data being compromised, as well as the inconvenience of having to report the fraud, block their payment method, and other things. When business are involved, it causes some issues; occasionally, they must issue refunds in order to keep customers. Therefore, it is crucial that both consumers and businesses are aware of these internet scams. A model to determine if an online payment is fraudulent or not is put forth in this study. To determine if a certain online payment is fraudulent or not, some features like the type of payment, the recipient's identity, etc. Would be taken into account.

By leveraging machine learning algorithms and data analytics techniques, this study aims to identify the whether an online payment is fraud or not. The report explores the underlying mechanisms of online fraud, identifies key vulnerabilities in current payment systems, and discusses the role of technology in fraud detection. It highlights how significant it is to have safe and dependable online transactions in order to identify theft, find theft, illegal transactions, and other fraudulent entities.

Problem Statement:

Online payment fraud has become a significant issue as digital transactions continue to grow rapidly. The challenge lies in detecting fraudulent transactions accurately and efficiently amidst a vast number of legitimate ones. Traditional rule-based systems are often ineffective due to the evolving nature of fraud tactics, making it difficult to keep pace with emerging threats. The problem, therefore, is to design a machine learning-based system that can detect and prevent fraud in real-time. This system needs to address several key challenges: handling highly imbalanced datasets, as fraudulent transactions are rare compared to legitimate ones; adapting to new fraud patterns as they emerge; minimizing false positives to avoid inconveniencing genuine users; and scaling effectively to manage large transaction volumes across global payment platforms. Machine learning models offer a promising solution by learning from historical data to identify patterns and anomalies associated with fraudulent behaviour, improving detection accuracy and reducing manual intervention. However, developing a reliable ML-based fraud detection system requires careful consideration of model choice, data preprocessing, and real-time performance to ensure it is both effective and adaptable in the dynamic landscape of online payments.

Online payment fraud detection is a complex and critical issue as digital transactions become increasingly prevalent. The challenge lies in distinguishing between legitimate and fraudulent transactions in real-time while dealing with a variety of sophisticated and evolving fraud tactics. Traditional rule-based methods, while effective initially, are unable to keep pace with the ever-changing nature of online fraud, which includes tactics like identity theft, account takeovers, and transaction spoofing. Fraudsters continuously adapt their methods to exploit vulnerabilities, making static systems inefficient.

The problem, therefore, is to create a machine learning-driven system that can intelligently detect and prevent fraud by identifying suspicious patterns and anomalies in transaction data. This system must address the core challenge of dealing with highly imbalanced data, where legitimate transactions far outnumber fraudulent ones. If not managed properly, this imbalance can result in models that perform poorly in detecting fraud, leading to both false positives (flagging legitimate transactions as fraudulent) and false negatives (allowing fraud to go unnoticed). Minimizing false positives is crucial because high rates of legitimate transaction rejections can frustrate users, leading to a poor customer experience and potential loss of business.

Solution:

In the context of online transactions, fraudulent activities pose a severe risk. The challenge is to build a model that can accurately distinguish between legitimate transactions and fraudulent once. this project aims to create a predictive model that classifies transactions as either “fraudulent” or “non fraudulent” based on various transactions attributes.

To address the challenge of online payment fraud, a comprehensive machine learning (ML)-based solution can be developed, focusing on several key aspects: data collection, model development, and deployment. The solution begins with gathering extensive transaction data, which includes information such as transaction amounts, timestamps, locations, payment methods, and user details. This data is crucial for training an effective fraud detection model.

In summary, a machine learning-based approach to online payment fraud detection involves meticulous data handling, thoughtful model selection, and ongoing system maintenance. By leveraging advanced ML techniques and adhering to ethical standards, organizations can effectively reduce fraud and enhance the security of online transactions.

Work Flow

The process includes the following steps:

Preparing the dataset:

The procedure is to acquire raw data, here we have loaded the data with infinite set which consists of online payment transactions of users from Kaggle.

Data preprocessing:

In order to improve the efficacy of fraud detection techniques, it involves preparing the raw transactional data for analysis through the application of several strategies. The dataset includes the following entities:'step', 'type', 'amount', 'nameOrig', 'oldbalanceOrg', 'nameDest', 'oldbalanceDest', 'newbalanceDest', 'isFraud', 'isFlaggedFraud'. Additionally, the target variable 'isFraud' indicated either a fraudulent transaction or not.

Data Splitting:

It refers to the process of dividing up the available data into two groups. In this project, the "train_test_split" function from the scikit-learn library is used to split the dataset into training and testing subsets. The split sets aside 20% of the data for performance evaluation and uses the remaining 80% of the data for training the model, which guarantees a sizable sample size for pattern recognition.

Model training:

A procedure wherein enough training data is provided to a machine learning (ML) algorithm so that it can gain knowledge from it. Using the fit technique, the training data(X_train,Y_train) are fed into the Decision tree classifier during the training phase. By doing this, the classifier establishes connections between the input features and the target variable by examining patterns and dependencies in the data.

Model Evaluation:

It is the essential to the process of developing a model. Determining which model best fits our data and how well it will perform going forward is helpful. The model is evaluated using testing data(X_test,Y_test) after training.

Code:

Importing the necessary libraries and loading the data:

```
import pandas as pd
import numpy as np
data = pd.read_csv("C:/Users/Admin.DESKTOP-IAELC5G.000/Downloads/archive/onlinefraud.csv")
```

```
data.head((5))
```

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

Checking for missing values:

```
print(data.isnull().sum())
```

```
step              0
type              0
amount            0
nameOrig          0
oldbalanceOrg     0
newbalanceOrig    0
nameDest          0
oldbalanceDest    0
newbalanceDest    0
isFraud           0
isFlaggedFraud    0
dtype: int64
```

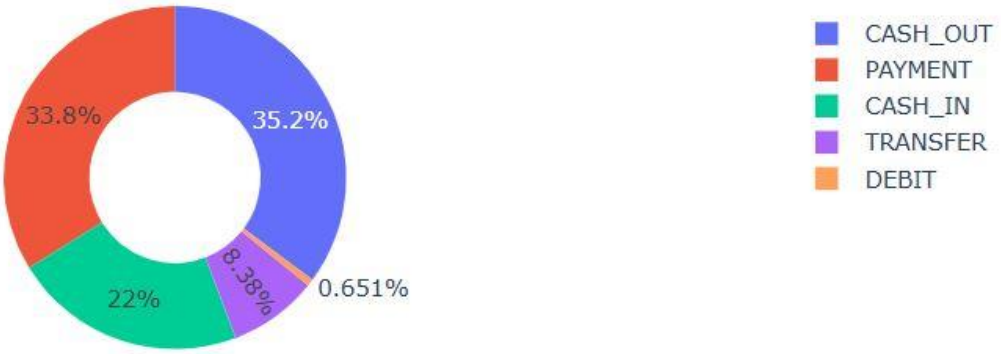
Exploring transaction type:

```
type=data["type"].value_counts()
print(type)
```

```
type
CASH_OUT    2237500
PAYMENT     2151495
CASH_IN     1399284
TRANSFER     532909
DEBIT         41432
Name: count, dtype: int64
```

```
transaction=type.index
quantity=type.values
```

```
import plotly.express as px
figure=px.pie(data,values=quantity,names=transaction,hole=0.5,title="distor of transaction type")
figure.show()
```



Correlation Analysis:

```
numeric_cols=data.select_dtypes(include=['float64','int64'])
correlation=numeric_cols.corr()
print(correlation)
```

```
                    step    amount  oldbalanceOrg  newbalanceOrig  \
step            1.000000  0.022373      -0.010058       -0.010299
amount          0.022373  1.000000      -0.002762       -0.007861
oldbalanceOrg  -0.010058 -0.002762       1.000000        0.998803
newbalanceOrig -0.010299 -0.007861       0.998803        1.000000
oldbalanceDest  0.027665  0.294137       0.066243        0.067812
newbalanceDest  0.025888  0.459304       0.042029        0.041837
isFraud         0.031578  0.076688       0.010154       -0.008148
isFlaggedFraud  0.003277  0.012295       0.003835        0.003776

                oldbalanceDest  newbalanceDest   isFraud  isFlaggedFraud
step                  0.027665        0.025888  0.031578        0.003277
amount                0.294137        0.459304  0.076688        0.012295
oldbalanceOrg         0.066243        0.042029  0.010154        0.003835
newbalanceOrig        0.067812        0.041837 -0.008148        0.003776
oldbalanceDest        1.000000        0.976569 -0.005885       -0.000513
newbalanceDest        0.976569        1.000000  0.000535       -0.000529
isFraud              -0.005885        0.000535  1.000000        0.044109
isFlaggedFraud       -0.000513       -0.000529  0.044109        1.000000
```

Data preprocessing:

```
correlation["isFraud"].sort_values(ascending=False)
```

```
isFraud           1.000000
amount            0.076688
isFlaggedFraud    0.044109
step              0.031578
oldbalanceOrg     0.010154
newbalanceDest    0.000535
oldbalanceDest   -0.005885
newbalanceOrig   -0.008148
Name: isFraud, dtype: float64
```

Prepare data for modeling:

```
data["type"]=data["type"].map({"CASH_OUT":1,"PAYMENT":2,"CASH_IN":3,"TRANSFER":4,"DEBIT":5})
```

```
data["isFraud"]=data["isFraud"].map({0:"no fraud",1:"fraud"})
data.head(5)
```

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	2	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	no fraud	0
1	1	2	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	no fraud	0
2	1	4	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	fraud	0
3	1	1	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	fraud	0
4	1	2	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	no fraud	0

Training, testing, splitting the model:

```
#train the model
from sklearn.model_selection import train_test_split
x=np.array(data[["type","amount","oldbalanceOrg","newbalanceOrig"]])
y=np.array(data[["isFraud"]])
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.10,random_state=42)
model=DecisionTreeClassifier()
model.fit(xtrain,ytrain)
print(model.score(xtest,ytest))
```

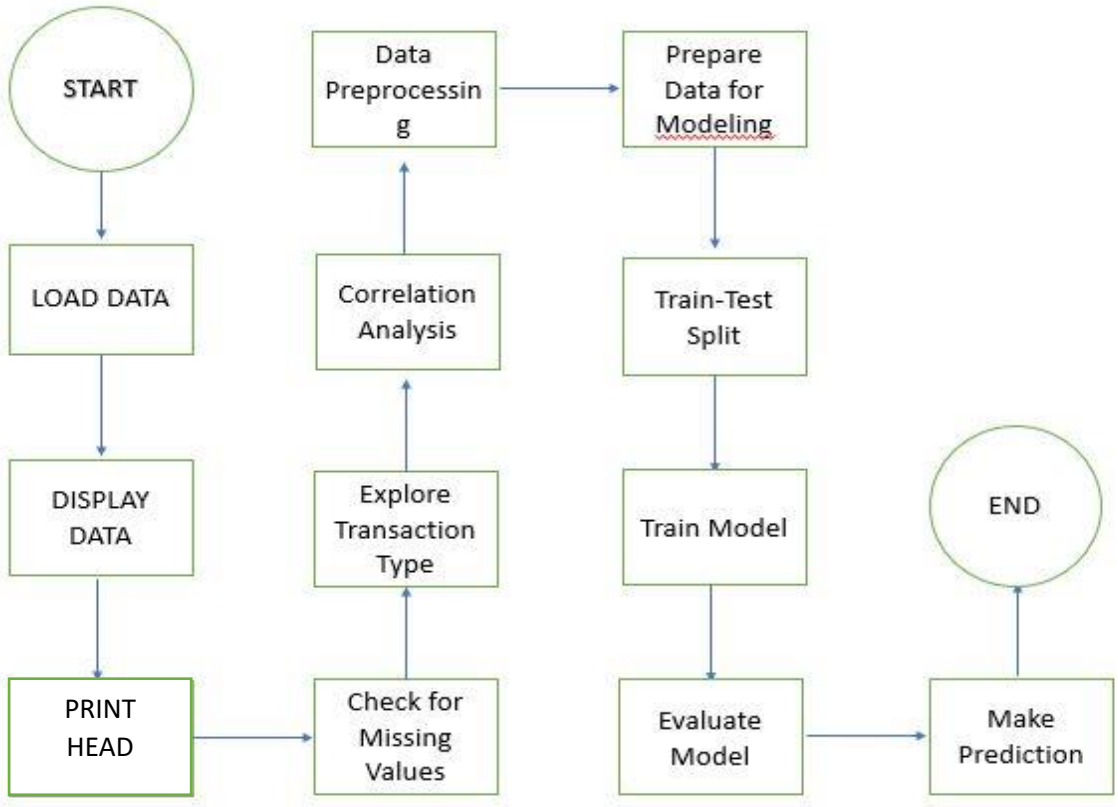
```
0.9997343861491021
```

Prediction:

```
 #prediction
features=np.array([[4,9000.0,9000.0,0.0]])
print(model.predict(features))
```

```
['fraud']
```

Flow chart



Key metrics(Benefits):

1.Reduced Financial Losses: Effective fraud detection helps minimize direct financial losses from fraudulent transactions. By identifying and preventing fraudulent activities before they are completed, businesses can avoid losing revenue and incurring chargeback fees.

2.Enhanced Security: Advanced fraud detection systems protect sensitive financial information and personal data from unauthorized access and theft. This enhances overall security and helps safeguard against data breaches.

3.Improved Customer Trust: When customers see that a business is taking steps to protect their financial transactions, they are more likely to trust the company. This trust can lead to increased customer loyalty and retention.

4.Decreased Operational Costs: By reducing the number of fraudulent transactions and chargebacks, businesses can lower operational costs associated with fraud investigation, customer service, and dispute resolution.

5.Fewer Chargebacks: Fraud detection systems can help reduce the frequency of chargebacks, which can be costly for merchants. Fewer chargebacks mean lower fees and less administrative work related to resolving disputes.

6.Compliance with Regulations: Many industries and regions have regulatory requirements for data protection and fraud prevention. A strong fraud detection system helps ensure compliance with these regulations, avoiding potential fines and legal issues.

7.Optimized Fraud Prevention Strategies: By analysing data from fraud detection systems, businesses can gain insights into fraud patterns and trends. This allows them to continuously improve and adapt their fraud prevention strategies to stay ahead of evolving threats.

8.Enhanced Customer Experience: Effective fraud detection minimizes the likelihood of false positives (legitimate transactions flagged as fraudulent), reducing customer frustration and ensuring smoother transactions.

9. Fraud Prevention Innovation: Investing in fraud detection technology often drives innovation in other areas of security and payments, leading to improved overall systems and processes.

Future Scope

Online payment fraud detection using ML offers a comprehensive solution to combat the rising threat of illicit activities in the digital realm. By leveraging machine learning algorithms, businesses can effectively analyse vast datasets of transaction data to identify patterns and anomalies that indicate fraudulent behaviour. These algorithms can learn to distinguish between legitimate and fraudulent transactions, enabling proactive fraud prevention and detection.

Through techniques such as supervised and unsupervised learning, as well as deep learning, ML models can effectively identify and mitigate various types of online payment fraud, including identity theft, credit card fraud, chargeback fraud, and more. By implementing robust ML-based fraud detection systems, businesses can protect their customers and financial assets while ensuring a secure and reliable online payment experience.

The future of online payment fraud detection using machine learning seems promising, with various developing trends and advances on the horizon. Here are some important areas of future opportunity for online payment fraud detection:

A. Advanced Machine Learning Models:

Continued development of increasingly advanced machine learning models, including deep learning and reinforcement learning, to increase the accuracy and flexibility of fraud detection systems.

B. Explainable AI (XAI):

Addressing the interpretability of machine learning models to make them more transparent and intelligible, enabling for better decision-making and regulatory compliance.

C. Defence Against Adversarial Machine Learning:

Defending against adversarial assaults on machine learning models used in fraud detection.

D. Continuous monitoring and updates:

Implementing systems capable of constantly updating and adapting machine learning models when new fraud tendencies arise.

Conclusion:

The landscape of online payment fraud detection is evolving rapidly, driven by the dynamic nature of cyber threats and the increasing complexity of fraudulent activities. Our study delves into the multifaceted realm of online payment fraud detection, elucidating the significance of robust detection mechanisms in safeguarding financial transactions in today's digital era. By exploring methodologies such as anomaly detection, machine learning, and pattern recognition, we underscore the pivotal role of advanced algorithms and techniques in mitigating financial losses and protecting businesses and consumers alike.

Our investigation highlights the effectiveness of anomaly detection not only in identifying a significant proportion of fraudulent activities but also in minimizing false alarms. By integrating anomaly detection with statistical risk management methods, we demonstrate a substantial reduction in overall risk exposure. Our proposed end-to-end risk management framework, comprising fraud detection, fraud detection optimization, and risk modelling, offers a comprehensive approach to addressing online payment fraud, thereby enhancing the resilience of financial institutions against fraudulent activities.

Furthermore, our study emphasizes the importance of leveraging large datasets for evaluating fraud detection methods accurately. Through the utilization of real-world transaction data from a private bank, we demonstrate the efficacy of our approach in capturing deviations from normal customer behavior and encoding them as essential features for fraud detection. Looking ahead, our framework opens avenues for future research, particularly in the realm of reinforcement learning for implementing feedback loops between risk modelling and fraud detection. Additionally, extending the optimization of fraud detection by considering transaction dependent loss risks and customer segmentation presents promising directions for enhancing fraud detection capabilities further. In conclusion, our study contributes to the ongoing discourse on online payment fraud detection by providing insights into effective methodologies and frameworks for combating fraudulent activities. By fostering a deeper understanding of the challenges and opportunities in this domain, we aim to catalyze the development of robust and adaptive fraud detection mechanisms crucial for ensuring secure and trustworthy online transactions.

Reference: Internet