LOVELY
PROFESSIONAL
UNIVERSITY

ARTIFICIAL INTELLIGENCE PROJECT

AI in finance (Fraud Detection)

Artificial Intelligence (INT 404)

Submitted by:

Sr. No	Registration No	Name of students	Section	Group
1	12105265	AADI PADAMWAR	K21GP	1
2	12213334	SURAJ KUMAR	K21GP	2
3	12105630	UDAY BHARDWAJ	K21GP	1

School of computer Science and Engineering

Submitted To:

Lovely Professional University

Jalandhar, Punjab, India-144411

INDEX

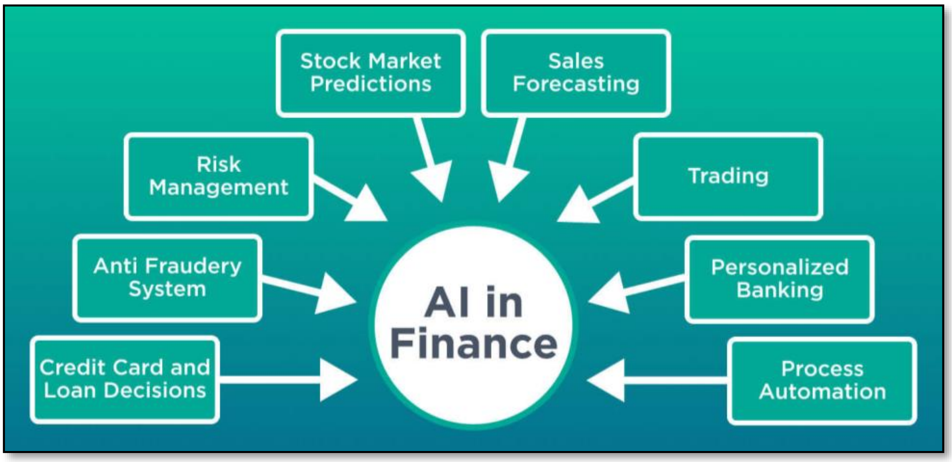
Artificial Intelligence in Finance (Fraud Detection)

Aadi Padamwar, Suraj Kumar, Uday Bhardwaj

Department of Computer Science and Engineering, Lovely Professional University, Phagwara, India

Abstract

By altering the way businesses process and analyze massive volumes of data, artificial intelligence (AI) has the potential to completely change the finance sector. Compared to conventional approaches, AI algorithms are more accurate in analyzing complex financial data, finding patterns and anomalies, and making forecasts. For a variety of uses, including fraud detection, investment analysis, risk management, and customer service, AI-based financial solutions have been developed. To make sure that these systems are open and impartial, ethical issues related to algorithmic bias and privacy invasion are also raised using AI in finance. Despite these obstacles, it is anticipated that the use of AI in finance will increase, offering major advantages to businesses that can successfully tap into its power.



One of the primary areas in which AI has made significant advancements in finance is in fraud detection. AI algorithms can analyze vast amounts of transactional data in real-time, identify patterns that are indicative of fraudulent activity, and alert financial institutions to potential fraudulent transactions. The use of AI in fraud detection has significantly improved detection accuracy while reducing the number of false positives, leading to a more effective and efficient fraud detection process.

Introduction

Losses resulting from financial fraud are thought to be in the billions of dollars yearly, and it is an issue that is chronic and getting worse for organizations of all kinds. Traditional techniques of detecting fraud rely on rule-based systems, which have difficulty identifying novel forms of fraud or adjusting to shifting fraud patterns. But because of the quick development of artificial intelligence (AI), businesses now have a potent weapon at their disposal to fight financial crime. Artificial intelligence (AI)-based fraud detection solutions have the potential to revolutionize how businesses identify and stop financial fraud, resulting in more precise and efficient fraud detection.

Utilizing AI for fraud detection has several benefits, one of which is its capacity to spot patterns and anomalies that weren't previously recognized. AI systems can examine enormous volumes of data and find minute trends and anomalies that might be signs of fraud. This is especially crucial because financial fraud is a dynamic crime that calls for agile and adaptable detection techniques.

By lowering the frequency of false positives, AI algorithms can also increase the accuracy of fraud detection. When genuine transactions are mistakenly labelled as fraudulent, false positives happen, resulting in pointless investigations and potential client displeasure. Artificial intelligence (AI) systems can analyze transaction data in real-time and spot trends that point to valid transactions, lowering the incidence of false positives and lessening the impact on honest clients.

The application of AI to fraud detection also opens new possibilities for cost reduction, enhanced customer service, and boosted productivity. Organizations can lessen the stress on their fraud detection teams and eliminate the need for manual intervention by automating the fraud detection process. By reducing costs and increasing efficiency, this can free up resources for other parts of an organization’s operations.

The application of AI in fraud detection, however, also poses considerable difficulties, particularly in terms of ethics and regulations. The possibility of algorithmic bias, in which AI algorithms might discriminate against groups based on criteria like race, gender, or age, is one of the main ethical worries. Financial institutions need to make sure that their AI systems are fair and transparent, and that they don't reinforce data biases that already exist.

To ensure that these systems are open, equitable, and responsible, strong ethical frameworks and regulatory monitoring must also be implemented along with the use of AI in fraud detection. Businesses must put strong data protection procedures in place and make sure that client information is only used to detect fraud. Additionally, they must be open and honest about the information they get, how they use it, and who has access to it.

Data and Knowledge Sources

Some common data and knowledge sources used in AI-based fraud detection include:

1. Historical transactional data: This data is collected from past transactions and is used to identify patterns and anomalies that may be indicative of fraudulent activity.
2. External data sources: External data sources such as social media, public records, and online marketplaces can provide additional insights into potential fraudulent activity. For example, social media data can be used to identify connections between individuals involved in fraudulent activity.
3. Industry expertise: Industry experts with a deep understanding of fraud patterns and techniques can provide valuable knowledge to help develop more effective AI-based fraud detection systems.
4. Rule-based systems: Rule-based systems use a predefined set of rules to identify potential fraud. AI algorithms can learn from these rules and identify new patterns that may not have been previously identified.
5. User behaviour data: User behaviour data can be used to build a profile of a customer's normal behaviour. Any deviation from this behaviour can be flagged as potentially fraudulent activity.
6. Human expert review: AI algorithms can be trained to learn from human expert reviews of potentially fraudulent activity. This knowledge can be used to develop more accurate and effective fraud detection algorithms.
7. Regulatory data: Regulatory data such as watch lists and blacklists can be used to identify potentially fraudulent activity involving individuals or entities on these lists.

Type of fraud in Finance Sector

The finance industry encompasses a broad range of activities, including banking, insurance, investment, and accounting. Each of these activities is vulnerable to different types of fraud, which can result in significant financial losses for individuals and businesses. In this article, we will discuss some of the most common types of fraud in the finance industry.

a) **Identity theft**:

This occurs when someone steals another person's personal information, such as their name, address, Social Security number, or credit card information, and uses it to commit fraud. For example, they may open a credit card account in the victim's name, make purchases, and then leave the victim with the debt. Identity theft can also involve stealing passwords or other login information to gain access to financial accounts.

b) **Credit card fraud:**

This occurs when someone uses another person's credit card information to make unauthorized purchases. Credit card fraud can involve physical theft of the credit card, as well as online or telephone purchases made with stolen credit card information.

c) **Mortgage fraud:**

This occurs when someone misrepresents information on a mortgage application to obtain a loan. For example, they may lie about their income, employment status, or assets to qualify for a mortgage they would not otherwise be eligible for.

d) **Investment fraud:**

This occurs when someone misrepresents or withholds information about an investment to convince someone to invest. This can include Ponzi schemes, where investors are promised high returns but are actually paid with the money of new investors rather than actual profits from the investment.

e) **Insurance fraud**:

This occurs when someone makes a false insurance claim in order to receive payment. This can include staged accidents, where someone intentionally causes an accident to file a claim, or filing claims for damages that were not actually suffered.

f) Money laundering:

This occurs when someone takes money obtained through illegal means and makes it appear to be legitimate. This can involve transferring money through a complex network of accounts and companies in order to hide its source.

g) Accounting fraud:

This occurs when a company or individual misrepresents financial information to deceive investors or lenders. This can include overstating revenues or assets, understating expenses or liabilities, or manipulating financial statements in order to make a company's financial performance look better than it actually is.

h) Cybercrime:

This refers to any type of criminal activity that is carried out online. This can include hacking into financial systems to steal information or money, phishing scams where people are tricked into giving up their login information, and ransomware attacks where criminals demand payment in exchange for restoring access to data.

i) Insider trading:

This occurs when someone uses non-public information to make a profit on a stock trade. For example, an employee of a company may trade on information about a new product or earnings report before it is made public, giving them an unfair advantage in the market.

j) Check fraud:

This occurs when someone alters or forges a check to obtain money. This can include stealing checks and altering the payee or amount or creating fake checks from scratch.

Working of AI in Detecting Frauds

Financial institutions are now able to spot patterns and abnormalities that can point to fraudulent conduct because to AI algorithms' speedy and accurate analysis of vast amounts of data. We will examine how AI is used to detect fraud in this post, from data preprocessing to model training and deployment.

a) Data Processing

Preprocessing of the data is necessary before AI may be used to identify fraud. This entails preparing the data for analysis by cleaning it. Several sources, including transaction logs, customer databases, and outside sources, may provide the data. It might also be free form, like in text messages or posts on social media.

Cleaning the data is the first stage in the preprocessing of data. This entails getting rid of any duplication, missing values, or contradictions. The data is then changed into a format that the AI system can utilize. Unstructured data may need to be transformed into structured data to achieve this, or feature engineering may be used to extract additional features from the existing data.

Data preprocessing is critical to the success of the AI algorithm. If the data is not cleaned and prepared properly, the algorithm may produce inaccurate or unreliable results.

b) AI Algorithm Selection

The next stage is to choose an AI algorithm that is suitable for the task at hand after the data has been preprocessed. There are several AI algorithms available, and each has advantages and disadvantages.

The following list includes some of the most popular AI fraud detection algorithms:

1) Supervised learning algorithms – They are trained on a dataset of transactions that have been classified as either valid or fraudulent. The program then learns patterns and anomalies that can point to fraud using this labelled dataset. Decision trees, logistic regression, and random forests are examples of common supervised learning algorithms used in fraud detection.

2) Unsupervised learning algorithms - These algorithms are used to detect anomalies in the data, without the need for labelled data. Unsupervised

learning algorithms can be useful in detecting new or previously unknown types of fraud. Common unsupervised learning algorithms used in fraud detection include clustering and anomaly detection.

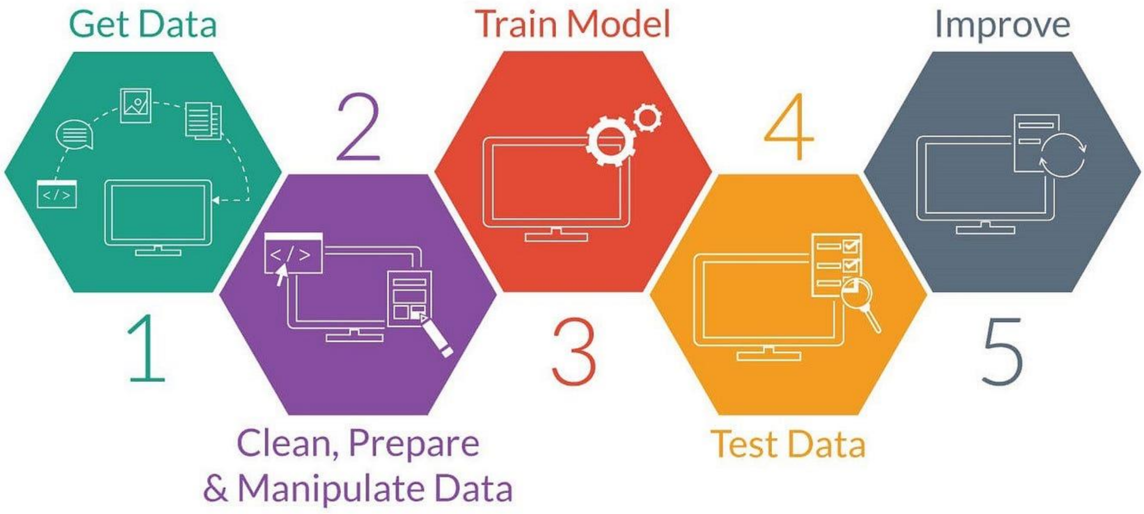
3) Deep learning algorithms - These algorithms are designed to work with complex data, such as images or audio. Deep learning algorithms can be used in fraud detection to analyse patterns in visual data, such as security camera footage or social media posts.

c) Model Training and Deployment

Once an AI algorithm has been chosen, the preprocessed data is used to train it. This entails dividing the data into training and testing sets, with the training set being used to evaluate the algorithm and the testing set being used to train it. Finding an algorithm that can correctly forecast fraudulent behaviour is the aim of model training.

The algorithm learns to recognise patterns and abnormalities in the data that are connected to fraudulent activity during model training. The system is then implemented in a real-time setting where it can monitor transactions in real-time and alert users to possibly fraudulent activities.

The effectiveness of an AI algorithm in detecting fraud depends on many factors, including the quality of the data, the choice of algorithm, and the accuracy of the model. AI algorithms must also be regularly updated and retrained to ensure that they remain effective in detecting new types of fraud.



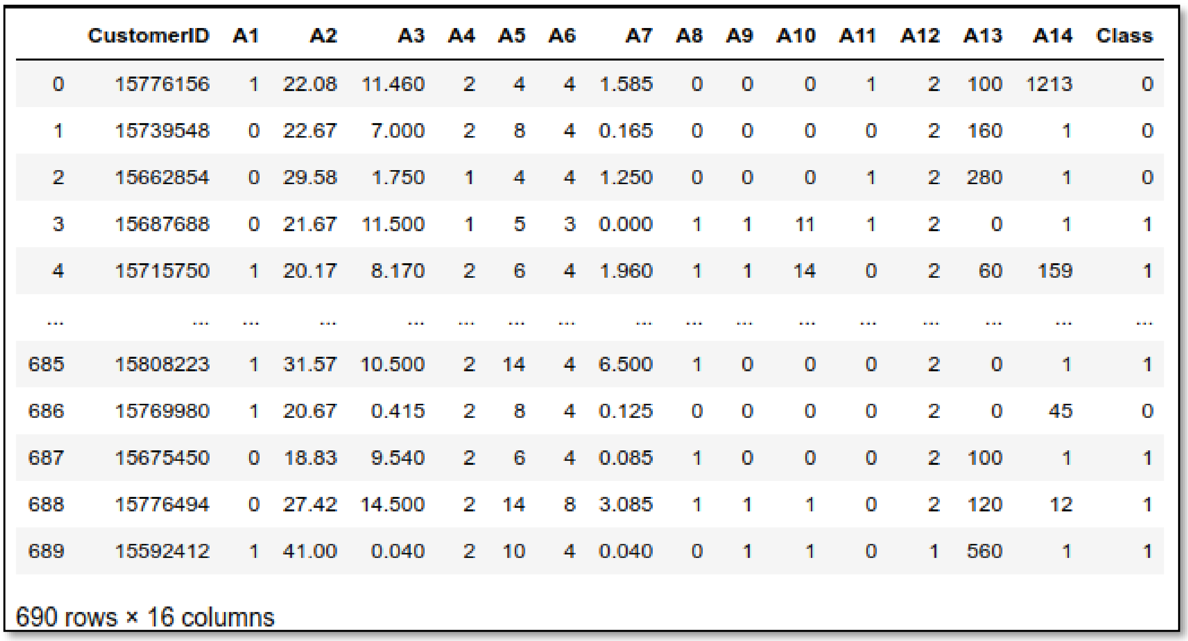
In the case of fraud detection using AI the data sets can credit card transactions of millions of users of 10 to 15 years or it can the no of transactions performed by a certain used in a certain time. This data is then cleaned and manipulated to extract the data we need to train our model.

Example of Data set used to train the Model

https://data.world/datagov-uk/9a39b56b-1cd5-446e-b7fc-99358cd4edc7

https://data.world/datagov-uk/c0f64d78-4ed2-4ece-8b7f-55e582a1350a

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud



	CustomerID	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	Class
0	15776156	1	22.08	11.460	2	4	4	1.585	0	0	0	1	2	100	1213	0
1	15739548	0	22.67	7.000	2	8	4	0.165	0	0	0	0	2	160	1	0
2	15662854	0	29.58	1.750	1	4	4	1.250	0	0	0	1	2	280	1	0
3	15687688	0	21.67	11.500	1	5	3	0.000	1	1	11	1	2	0	1	1
4	15715750	1	20.17	8.170	2	6	4	1.960	1	1	14	0	2	60	159	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
685	15808223	1	31.57	10.500	2	14	4	6.500	1	0	0	0	2	0	1	1
686	15769980	1	20.67	0.415	2	8	4	0.125	0	0	0	0	2	0	45	0
687	15675450	0	18.83	9.540	2	6	4	0.085	1	0	0	0	2	100	1	1
688	15776494	0	27.42	14.500	2	14	8	3.085	1	1	1	0	2	120	12	1
689	15592412	1	41.00	0.040	2	10	4	0.040	0	1	1	0	1	560	1	1

690 rows × 16 columns

Methods used to monitor Frauds using AI

(a) IP Analysis:

The first way will analyse the IP address of the user who wants to purchase. Artificial Intelligence enables businesses to find out a person's area. Top of that, it allows trades to match the location with the billing address.

(b) Device Analysis:

Device analysis identifies the device type, operating system, browser, and other prominent parameters. AI development solutions can help companies recognize devices if it is used in the past to engage in fraudulent payments. Fraudsters may use multiple devices to commit fraud, like a laptop, mobile or tablet that could be stolen or even shoplifted. The ability to identify different models and make these devices through AI can help businesses understand whether a device is new or used before.

(c) Phone analysis:

Artificial Intelligence-based solutions can help businesses to authenticate a customer's phone number in real-time. It is crucial because a fraudster may use a Voice over Internet Protocol (VoIP) numeral or other sophisticated methods to commit fraud. With AI, businesses can understand whether the number is a VoIP or an authentication.

In addition, artificial Intelligence can assist businesses with analysing call data records, as well as analysing all the incoming/outgoing calls and identifying patterns. For example, if a fraudster is using the same fake number to log into different e-commerce sites and proceeding to commit fraud.

(d) Email analysis:

AI implementation can help businesses automatically analyse email addresses to detect and prevent fraudulent activities. By analysing the email addresses, enterprises can understand whether the mail address is real or fake, including its location and other crucial details.

(e) Billing address analysis:

Fraudsters often target e-commerce businesses by using fake invoices to collect payments. However, deploying artificial intelligence (AI)–powered fraud detection can prevent fraud from occurring in the first place. AI can carefully examine customer information, payment details, invoice details, and other relevant data before a payment. It thoroughly analyses historical data of both valid and fraudulent invoices and tries to map out any repeated patterns that indicate possible fraud.

(f) Credit Card analysis:

In today's world, there are multiple uses of Artificial Intelligence, and one of those ways is in the credit card industry. AI is used to determine the type of credit card, the issuing bank, and the country of origin. It is done by automatically reviewing a customer's credit card details. With AI technologies, companies can identify whether the credit card is lost, real or fake. Moreover, AI can detect if the credit card is from a high-risk country or location where fraud is frequent.

(g) Social Media analysis:

It is a way of analysing the customer's social media profiles to understand their identity. AI-powered services can help businesses automatically scrutinize their social media profiles. With AI, companies can understand the users' names, ages, gender, interests, and other vital details. Moreover, AI can help businesses to understand consumers' social media behaviour.

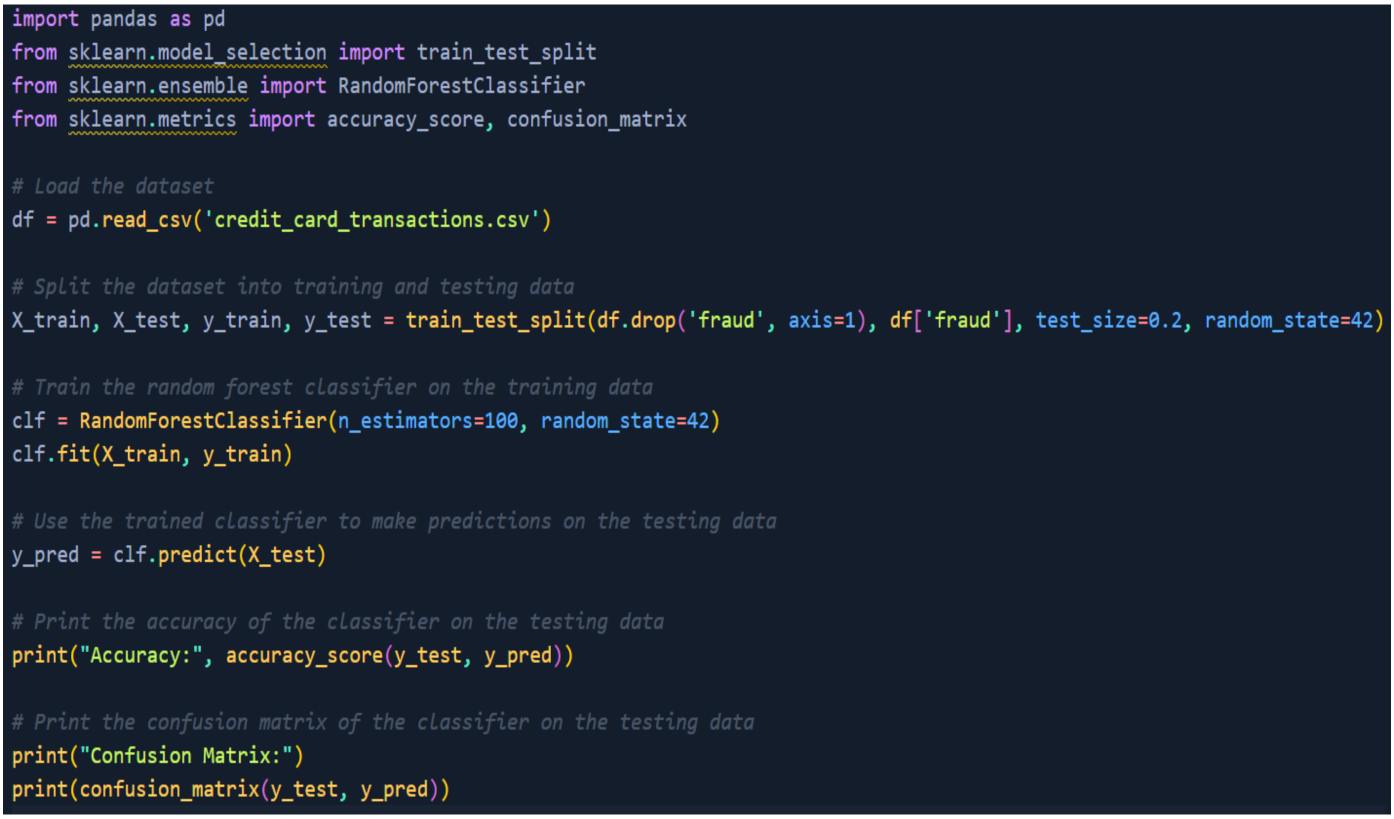
Conclusion

In conclusion, AI has proven to be a valuable tool in the fight against fraud. By analyzing large amounts of data and identifying patterns and anomalies, AI algorithms can detect fraudulent activity more quickly and accurately than traditional methods. Machine learning algorithms can also adapt and improve over time, making them more effective at identifying new types of fraud as they emerge.

While AI has great potential in fraud detection, it is important to remember that it is not a panacea. AI algorithms must be carefully designed and trained to avoid bias and false positives, which can lead to unnecessary investigations and damage to innocent individuals. Additionally, AI should be used in conjunction with other fraud detection methods, such as human analysis and audits, to ensure the highest level of accuracy and to provide a comprehensive approach to fraud prevention.

Overall, AI has the potential to revolutionize the way we detect and prevent fraud in the financial industry. As technology continues to advance and AI algorithms become more sophisticated, we can expect to see even greater improvements in fraud detection and prevention in the years to come.

Code to demonstrate the working of AI



```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix

# Load the dataset
df = pd.read_csv('credit_card_transactions.csv')

# Split the dataset into training and testing data
X_train, X_test, y_train, y_test = train_test_split(df.drop('fraud', axis=1), df['fraud'], test_size=0.2, random_state=42)

# Train the random forest classifier on the training data
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

# Use the trained classifier to make predictions on the testing data
y_pred = clf.predict(X_test)

# Print the accuracy of the classifier on the testing data
print("Accuracy:", accuracy_score(y_test, y_pred))

# Print the confusion matrix of the classifier on the testing data
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))
```

In this program, we first load a dataset of credit card transactions into a Pandas DataFrame. The dataset contains a column called "fraud" which indicates whether a transaction is fraudulent or not. We then split the dataset into training and testing data using the *‘train_test_split’* function from scikit-learn.

Next, we train a random forest classifier on the training data using the *‘RandomForestClassifier’* class from scikit-learn. This classifier is a type of machine learning algorithm that is well-suited for detecting fraud in credit card transactions because it can handle many input features and is robust to overfitting.

Once the classifier is trained, we use it to make predictions on the testing data using the ‘predict’ method. We then print the accuracy of the classifier on the testing data using the *'accuracy_score’* function from scikit-learn, as well as the confusion matrix using the *‘confusion_matrix’* function. The confusion matrix tells us how many true positives, false positives, true negatives, and false negatives the classifier produced.

This is just a simple example of how AI can be used for fraud detection, but there are many other machine learning algorithms and techniques that can be used depending on the specific requirements of the problem.

References

Dhieb, Najmeddine, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud. "*A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement*." IEEE Access 8 (2020): 58546-58558.

Alhaddad, Musaab Mohammad. "*Artificial Intelligence in Banking Industry: A Review on Fraud Detection, Credit Management, and Document Processing*." ResearchBerg Review of Science and Technology 2, no. 3 (2018): 25-46.

Choi, Dahee, and Kyungho Lee. "*An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation.*" Security and Communication Networks 2018 (2018).

Alzahrani, Reem A., and Malak Aljabri. "AI-based Techniques for Ad Click Fraud Detection and Prevention: Review and Research Directions." Journal of Sensor and Actuator Networks 12, no. 1 (2022): 4.

Bao, Yang, Gilles Hilary, and Bin Ke. "Artificial intelligence and fraud detection." *Innovative Technology at the Interface of Finance and Operations: Volume I* (2022): 223-247.

Jensen, David. "Prospective assessment of ai technologies for fraud detection: A case study." In *AAAI Workshop on AI Approaches to Fraud Detection and Risk Management*, pp. 34-38. 1997.