

Kshitij Jadhav (Team Lead)

Atif Shaik

GITHUB

ABSTRACT

The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of artificial intelligence is machine learning, which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Fraud detection can be used in various industries, including banking and finance, insurance, healthcare, and e commerce. It is essential for mitigating financial losses and protecting against reputational damage resulting from fraudulent activities. Fraud detection identifying and preventing fraudulent activities or transactions, which are illegal or unauthorized actions carried out to gain a financial benefit or cause harm to others. Fortunately, we can use machine learning to identify employees who may be committing fraud through the use of financial and email data by using the machine learning. Therefore, my model is designed in such a way that it would help the company as well as customers from different kind of fraud that occur to them.

PROBLEM STATEMENT

In today's digital landscape, the rise of online transactions and financial activities has led to an alarming increase in fraudulent activities, posing significant threats to businesses and individuals alike. Existing fraud detection methods often fall short in effectively identifying and preventing these malicious activities, resulting in substantial financial losses, damaged reputations, and compromised user trust. There is a pressing need for a robust and adaptive fraud detection application that leverages advanced technologies such as machine learning, artificial intelligence, and data analytics to proactively detect, mitigate, and deter fraudulent transactions in real-time. My application addresses the evolving tactics of fraudsters, minimize false positives, and provide businesses across various sectors with a reliable and comprehensive solution to safeguard their assets, customer data, and operational integrity. By filling this critical gap in the market, the proposed fraud detection application has the potential to revolutionize fraud prevention strategies, foster secure online environments, and empower businesses to stay ahead in the ongoing battle against fraud.

MARKET/CUSTOMER/BUSINESS NEED ASSESSMENT

1. Market need assessment -

The market need for fraud detection in machine learning is growing rapidly, as businesses of all sizes are increasingly targeted by fraudsters. In 2022, the global fraud detection and prevention market was valued at \$30.8 billion and is projected to reach \$53.5 billion by 2027, growing at a CAGR of 10.7%. There are a number of factors driving the growth of the fraud detection and prevention market, including:

The increasing sophistication of fraudsters, who are using increasingly complex techniques to commit fraud. The growth of e-commerce, which has created new opportunities for fraudsters to target businesses and consumers. The increasing amount of data that businesses are collecting, which can be used to train machine learning models to identify fraudulent activity. The growing regulatory requirements for businesses to have robust fraud detection and prevention systems in place. Machine learning is playing an increasingly important role in fraud detection, as it can be used to analyse large amounts of data to identify patterns and anomalies that may indicate fraud. Machine learning models can also be updated in real time to reflect changes in fraud patterns, which can help businesses to stay ahead of fraudsters.

The following are some of the benefits of using machine learning for fraud detection:

Accuracy: Machine learning models can be trained to identify fraudulent activity with a high degree of accuracy.

Speed: Machine learning models can process large amounts of data quickly, which allows businesses to identify and respond to fraud in real time.

Scalability: Machine learning models can be scaled to meet the needs of businesses of all sizes.

Cost-effectiveness: Machine learning models can be more cost-effective than traditional fraud detection methods.

As the market need for fraud detection in machine learning continues to grow, businesses are increasingly adopting machine learning solutions to protect themselves from fraud. Machine learning is a powerful tool that can help businesses to identify and prevent fraud, and it is expected to play an increasingly important role in the fraud detection and prevention market in the years to come.

2. Customer needs -

Accuracy: Businesses need a fraud detection solution that can identify fraudulent activity with a high degree of accuracy. This is critical to minimizing losses and protecting customers from harm.

Speed: Businesses need a fraud detection solution that can process large amounts of data quickly. This is essential for identifying and responding to fraud in real time.

Scalability: Businesses need a fraud detection solution that can be scaled to meet the needs of their business. This is important for businesses that are growing rapidly or that have a large customer base. **Cost-effectiveness:** Businesses need a fraud detection solution that is cost-effective. This is important for businesses that are on a budget or that are looking to save money on fraud prevention.

Ease of use: Businesses need a fraud detection solution that is easy to use. This is important for businesses that do not have a lot of technical expertise or that do not have the time to learn a complex system. In addition to these general customer needs, there are also some specific needs that businesses may have in terms of fraud detection machine learning. For example, businesses that operate in regulated industries may need a fraud detection solution that can meet specific regulatory requirements. Businesses that operate in high-risk industries, such as banking and finance, may need a fraud detection solution that is more sophisticated than what is needed by businesses in lower-risk industries. The best fraud detection machine learning solution for a particular business will depend on the specific needs of that business. However, all businesses should consider the factors listed above when evaluating fraud detection machine learning solutions.

Here are some additional benefits of using machine learning for fraud detection: *Machine learning models can be updated in real time to reflect changes in fraud patterns. This allows businesses to stay ahead of fraudsters and protect themselves from new and emerging threats.*

Machine learning models can be used to identify fraud patterns that are difficult or impossible to detect with traditional methods. This can help businesses to identify and prevent fraud that would otherwise go undetected.

Machine learning models can be used to automate fraud detection tasks, which can save businesses time and money. This can free up human resources to focus on other important tasks, such as investigating fraudulent activity and developing new fraud prevention strategies.

Overall, machine learning is a powerful tool that can help businesses to identify and prevent fraud. As the market need for fraud detection continues to grow, businesses are increasingly adopting machine learning solutions to protect themselves from fraud.

3. Business needs -

The business need in fraud detection is clear: fraud is a costly problem that can damage a business's reputation, bottom line, and even its ability to operate. According to the Association of Certified Fraud Examiners (ACFE), organizations lose an average of 5% of their annual revenue to fraud. That's a staggering number, and it's only getting worse. There are a number of reasons why fraud is becoming more common. One reason is the increasing sophistication of fraudsters. They are using more complex techniques to commit fraud, and they are always looking for new ways to exploit vulnerabilities. Another reason is the growth of e-commerce. E commerce transactions are more difficult to monitor than traditional brick and mortar transactions, which makes them more attractive to fraudsters. The good news is that there are a number of fraud detection solutions available that can help businesses to protect themselves. These solutions use a variety of techniques to identify and prevent fraud, including:

Machine learning: Machine learning algorithms can be used to analyse large amounts of data and identify patterns that may indicate fraud.

Rule-based systems: Rule-based systems use a set of pre-defined rules to identify suspicious activity.

Data analytics: Data analytics can be used to identify trends and patterns that may indicate fraud.

Human review: Human review is still an important part of fraud detection, as it can help to identify fraud that is not detected by automated systems.

By using a combination of these techniques, businesses can significantly reduce their risk of fraud. However, it's important to remember that no fraud detection solution is perfect. Fraudsters are constantly evolving their techniques, so businesses need to be vigilant and update their fraud detection solutions as needed. In addition to the financial losses that fraud can cause, it can also damage a business's reputation. When customers are defrauded, they are likely to lose trust in the business and take their business elsewhere. This can lead to a

decline in sales and profits. Fraud can also lead to legal problems. If a business is found to have been negligent in preventing fraud, it may be held liable for the losses that customers suffer. This can be a costly and time-consuming process. For all of these reasons, it's clear that there is a strong business need for fraud detection. By investing in fraud detection solutions, businesses can protect themselves from financial losses, reputational damage, and legal problems.

TARGET SPECIFICATION

In fraud detection, target specification is the process of defining the specific types of fraud that a business wants to detect. This is important because it helps to focus the fraud detection efforts and to ensure that the right data is collected and analysed.

There are a number of factors to consider when defining target specifications for fraud detection, including:

The type of business: Different types of businesses are more susceptible to different types of fraud. For example, banks are more likely to be targeted by fraudsters who want to steal money, while e-commerce businesses are more likely to be targeted by fraudsters who want to steal credit card information.

The size of the business: larger businesses are more likely to be targeted by fraudsters because they have more money and resources.

The industry: Some industries are more prone to fraud than others. For example, the financial services industry is a prime target for fraudsters.

The location: Some regions are more prone to fraud than others. For example, countries with weak regulations are more likely to be targeted by fraudsters.

Once the target specifications have been defined, the next step is to collect the data that will be used to detect fraud. This data can include customer information, transaction data, and device data. The data is then analysed using a variety of techniques, such as machine learning and rule-based systems, to identify patterns that may indicate fraud.

Here are some examples of target specifications for fraud detection:

Credit card fraud: This type of fraud involves the unauthorized use of a credit card to make purchases. The target specification for credit card fraud might include transactions that are made in a different country than the cardholder's billing address, or transactions that are made for a large amount of money.

Identity theft: This type of fraud involves the use of someone else's personal information to commit fraud. The target specification for identity theft might include transactions that are made with a stolen Social Security number, or transactions that are made in the name of someone who has not authorized the transaction.

Money laundering: This type of fraud involves the attempt to conceal the source of illegally obtained money. The target specification for money laundering might include transactions that are made between shell companies, or transactions that are made in large amounts of cash. By carefully defining the target specifications for fraud detection, businesses can improve their chances of detecting fraud and protecting themselves from financial losses.

APPLICABLE PATENT

This model would detect fraud detection using anomaly detection. The system involves collecting data from a variety of sources, such as transaction data, customer data, and device data. The data is then analyzed using anomaly detection algorithms to identify patterns that are unusual or unexpected. If a pattern is identified that is unusual or unexpected, then the transaction or account may be flagged for further investigation. There is no such technologies in the market. Hence it can be created an applied for a Patent.

APPLICABLE REGULATIONS

Fraud detection is subject to various regulations, depending on the jurisdiction and industry in which it operates. Below are some of the key regulations that may be applicable to fraud detection efforts:

1. Payment Card Industry Data Security Standard (PCI DSS):

Relevant for organizations that handle credit card transactions, PCI DSS outlines security standards to protect cardholder data and prevent fraud. Compliance involves implementing measures like encryption, access controls, and regular security assessments.

2. General Data Protection Regulation (GDPR):

Applicable to organizations that handle personal data of EU citizens, GDPR mandates data protection and privacy safeguards. Fraud detection efforts must ensure that data processing complies with GDPR requirements, including obtaining valid consent, ensuring data subject rights, and secure data handling.

3. Bank Secrecy Act (BSA) and Anti-Money Laundering (AML) Regulations:

Financial institutions are required to implement AML programs to detect and prevent money laundering and fraud. These programs involve customer due diligence, transaction monitoring, and reporting suspicious activities to relevant authorities.

4. Sarbanes-Oxley Act (SOX):

Public companies in the United States must comply with SOX, which includes internal controls and reporting requirements. Fraud detection systems can play a role in ensuring accurate financial reporting and preventing fraudulent practices.

5. Consumer Financial Protection Bureau (CFPB) Regulations:

Financial institutions and service providers must adhere to CFPB regulations to protect consumers from unfair, deceptive, or abusive practices. Effective fraud detection helps prevent such practices and maintains consumer trust.

6. Health Insurance Portability and Accountability Act (HIPAA):

In the healthcare sector, HIPAA requires safeguarding patient data. Fraud detection measures must be designed to protect sensitive health information and prevent unauthorized access.

7. Securities and Exchange Commission (SEC) Regulations:

Organizations trading securities or operating in the financial markets need to adhere to SEC regulations to ensure fair and transparent operations. Fraud detection mechanisms can help maintain market integrity.

8. Cyber security Frameworks (e.g., NIST Cyber security Framework):

Various industries may follow cyber security frameworks to protect against fraud and cyber threats. These frameworks provide guidelines for risk assessment, mitigation, and incident response.

APPLICABLE CONSTRAINTS

Data Quality and Quantity: The accuracy and reliability of the FRAUD DETECTION advice provider model heavily rely on the availability historical data. It is very crucial for accurate predictions. Low-quality or insufficient data can lead to unreliable results.

Class imbalance: In many cases, there are far fewer fraudulent transactions than legitimate transactions. This can make it difficult for machine learning models to learn to identify fraud.

False positives: Machine learning models can sometimes generate false positives, which means that they incorrectly flag a legitimate transaction as fraudulent. This can lead to inconvenience and frustration for customers.

False negatives: Machine learning models can also generate false negatives, which means that they incorrectly fail to flag a fraudulent transaction. This can lead to financial losses for businesses.

Imbalanced Data: Fraudulent activities are often rare compared to legitimate transactions, leading to class imbalance. This can result in models that are biased towards the majority class and have difficulty detecting fraud instances.

Feature Engineering: Extracting relevant features from raw data is crucial for model performance. However, identifying meaningful features for fraud detection can be complex and domain-specific.

Model Interpretability: Many machine learning algorithms, especially complex ones like deep neural networks, lack interpretability. In fraud detection, it's important to understand why a model made a particular decision to meet regulatory and ethical requirement.

BUSINESS MODEL

1)Collaboration with other banks and financial services-

Collaborate with other banks, financial institutions, and cyber security experts to share information and best practices for fraud detection. It would charge a yearly fee for providing the service.

Establish a secure data-sharing mechanism with partner institutions. This may involve sharing anonymize transaction data for training and refining the model.

Ensure compliance with data protection and privacy regulations (e.g., GDPR, HIPAA) during data sharing and integration.

2)Value-added Services:

Consider offering additional services such as fraud risk assessment, fraud prevention consultancy, or analytics.

3)Monetization

Generate revenue by offering fraud detection and prevention services to other banks, financial institutions, or businesses looking to enhance their security measures.

4)Subscription Model-

This is the most common way for fraud detection software to generate revenue. Customers pay a monthly or annual fee to use the software. The fee covers the cost of developing and maintaining the software, as well as the cost of customer support.

Per-transaction fees: Some fraud detection software charges customers a fee for each transaction that is processed through the software. This fee is typically a small percentage of the transaction amount.

Freemium model: This model offers a basic version of the software for free, and charges customers for additional features or services. For example, a fraud detection software might offer a basic version that only scans for credit card fraud, and charge customers for a premium version that also scans for identity theft and other types of fraud.

Pay-per-use: This model charges customers for the amount of data that they process through the software. This model is typically used for fraud detection software that is used by large businesses or organizations that process a lot of data.

White labeling: This model allows businesses to rebrand the fraud detection software as their own. Businesses pay a fee to the fraud detection software company, and then they can sell the software to their customers as their own product.

5)Advertising and Sponsorship:

Display advertising: This is the most common type of advertising in fraud detection software. Display ads are typically images or text ads that are displayed on the user interface of the software. Businesses can pay to have their ads displayed on the software, and the fraud detection app can earn a commission on each click or conversion.

Native advertising: Native ads are designed to blend in with the rest of the content on the fraud detection software. They are typically articles, videos, or infographics that are sponsored by a business. Businesses can pay to have their native ads displayed on the software, and the fraud detection app can earn a commission on each click or conversion.

In-app purchases: My fraud detection app will offer an in-app purchases, such as additional features or tools. Businesses can pay to have their products or services promoted as in-app purchases, and the fraud detection app can earn a commission on each sale.

6) API Access for Developers:

Offer an API that developers can integrate into their own fraud detection apps or platforms. Charge a licensing fee for access to the API, expanding your app's reach and potential revenue streams.

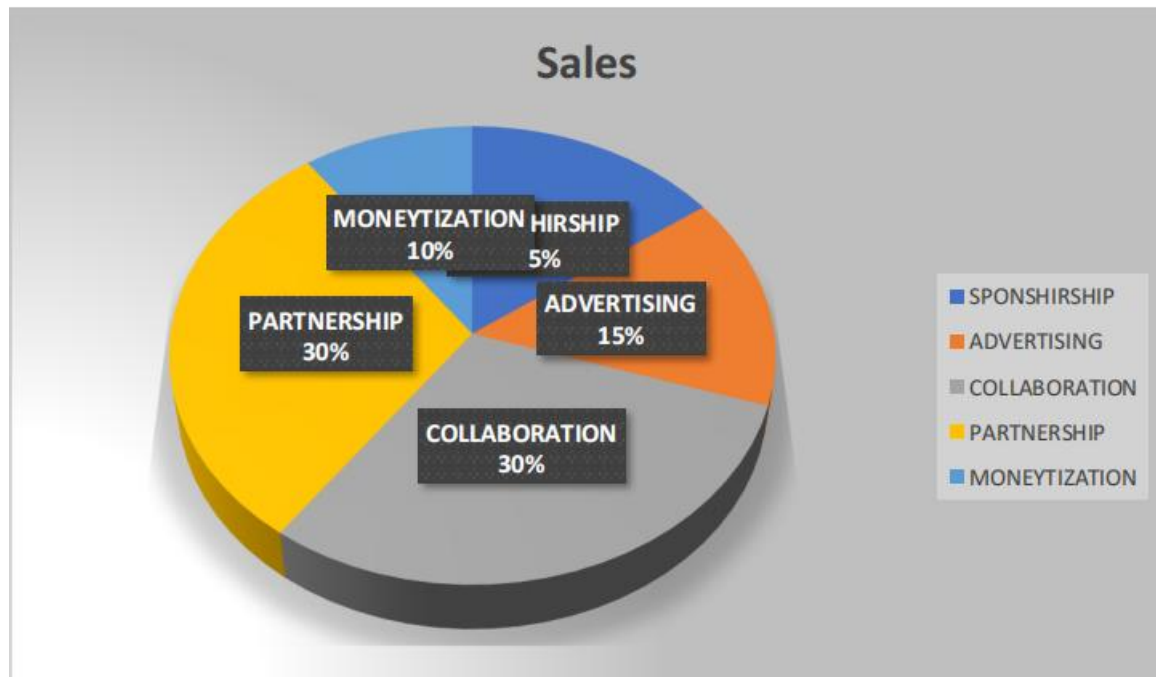
7)Partnerships with Banks and Customers:

A fraud detection app can collaborate with banks in a number of ways to make money. Here are a few examples:

Data sharing: The fraud detection app can share data with banks to help them identify and prevent fraud. This data can include customer transactions, financial records, and social media activity. By sharing data, the fraud detection app can help banks to improve their fraud detection capabilities and reduce their fraud losses.

Fraud prevention services: The fraud detection app can provide banks with fraud prevention services, such as real-time fraud monitoring and customer risk scoring. These services can help banks to identify and prevent fraud before it occurs.

Fraud training: The fraud detection app can provide banks with fraud training for their employees. This training can help employees to identify and prevent fraud, and it can also help to build a culture of fraud awareness within the bank.



CONCEPT GENERATION

Fraud patterns: This could involve identifying patterns in the timing of transactions, the amounts of transactions, or the types of transactions. For example, an organization might identify a pattern of fraudulent transactions that all occur on the same day of the week or that all involve large amounts of money.

Fraud rules: This could involve identifying rules that can be used to flag suspicious transactions. For example, an organization might create a rule that flags any transaction that exceeds a certain amount or that occurs from an IP address that has been previously associated with fraud.

Fraud models: This could involve creating models that can predict the likelihood of a transaction being fraudulent. These models can be used to flag transactions that are more likely to be fraudulent, which can help organizations to focus their resources on the most suspicious transactions.

CONCEPT DEVELOPMENT

1. Implement advanced anomaly detection algorithms to identify unusual transaction patterns.
2. Combine statistical methods, clustering, and neural networks to detect outliers and potential fraud.

Graph Analysis:

1. Create a graph database to model relationships between customers, accounts, and transactions.
2. Apply graph analysis techniques to identify hidden connections and networks involved in fraudulent activities.

Geolocation and Device Tracking:

1. Monitor geolocation data and device information to detect unusual login or transaction locations.
2. Implement device fingerprinting to recognize unauthorized devices and prevent account takeover.

Risk Scoring and Decision-making:

1. Assign risk scores to transactions based on various parameters, such as transaction amount, recipient, and location.
2. Employ rule-based systems and machine learning models to make realtime decisions on transaction approval, denial, or further investigation.

Customer Communication and Education:

Develop proactive communication strategies to inform customers about potential fraud risks and preventive measures.

PRODUCT DETAILS

The product works with the following steps -

1.User Registration and Profile Setup:

- First Users can create an account with the app.
- They provide their basic information and preferences, such as what type of services they want, like credit card, banking transactions, loans etc.

2.Data Collection

- The app gathers a wide range of data sources, including historical transaction prices, market trends, economic indicators etc.

3.Data Analysis:

Once the data is collected, it is analyzed to identify patterns and anomalies that may be indicative of fraud. This analysis can be done manually or using automated fraud detection software.

Statistical analysis: This method uses statistical techniques to identify patterns in data that may be indicative of fraud. For example, a business may use statistical analysis to identify customers who are making unusually large purchases or who are making purchases in a suspicious pattern.

Machine learning: This method uses machine learning algorithms to identify patterns in data that may be indicative of fraud. Machine learning algorithms can learn to identify patterns that are too complex for humans to identify.

Rules-based detection: This method uses a set of rules to identify fraudulent activity. For example, a business may have a rule that any transaction over \$10,000 must be manually approved.

Fraud scoring: This method assigns a score to each transaction or activity based on the likelihood of it being fraudulent. Transactions or activities with a high score are then flagged for further investigation.

SUBSCRIPTION MODEL

Recurring revenue: Subscription models generate recurring revenue, which can be more predictable and stable than one-time sales. This can be important for businesses that need to budget for fraud detection costs on a regular basis.

Scalability: Subscription models are scalable, which means that businesses can easily add or remove users as needed. This is important for businesses that are growing or that experience seasonal fluctuations in fraud activity.

Upsell opportunities: Subscription models can be used to upsell businesses on additional features and services. For example, a fraud detection app could offer businesses a premium plan that includes additional features, such as real-time fraud detection and customer support.

CHARGE AN ANNUAL FEE FROM BANKS

Data sharing: The fraud detection app can share data with banks to help them identify and prevent fraud. This data can include customer transactions, financial records, and social media activity. By sharing data, the fraud detection app can help banks to improve their fraud detection capabilities and reduce their fraud losses.

Fraud prevention services: The fraud detection app can provide banks with fraud prevention services, such as real-time fraud monitoring and customer risk scoring. These services can help banks to identify and prevent fraud before it occurs.

Fraud training: The fraud detection app can provide banks with fraud training for their employees. This training can help employees to identify and prevent fraud, and it can also help to build a culture of fraud awareness within the bank.

REQUIRED TECHNOLOGIES

- o Machine Learning linear regression algorithm
- o Sklearn
- o Seaborn
- o Pandas
- o Flask
- o HTML
- o CSS
- o JavaScript
- o Docker
- o SQL Database
- o Android development
- o IOS development
- o Git
- o API
- o AWS

TEAM REQUIRED TO DEVELOP THE MODEL

- o Machine Learning Engineer
- o Full-stack developer
- o Data Scientist
- o Business Analyst
- o Dev Ops Engineer
- o Software developer
- o App developer

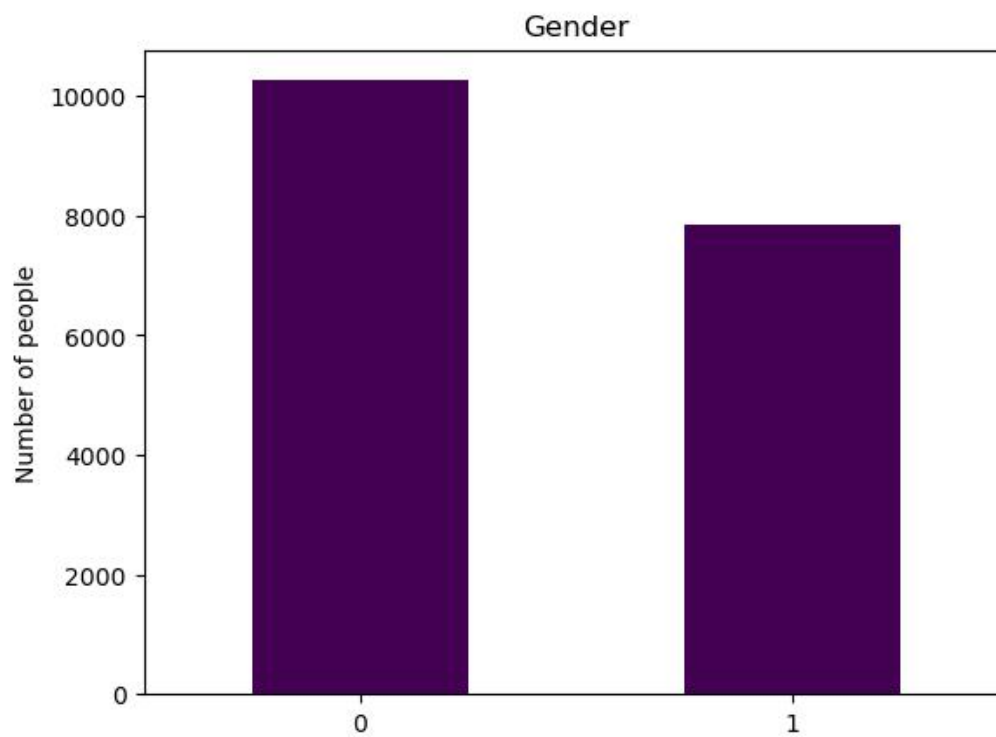
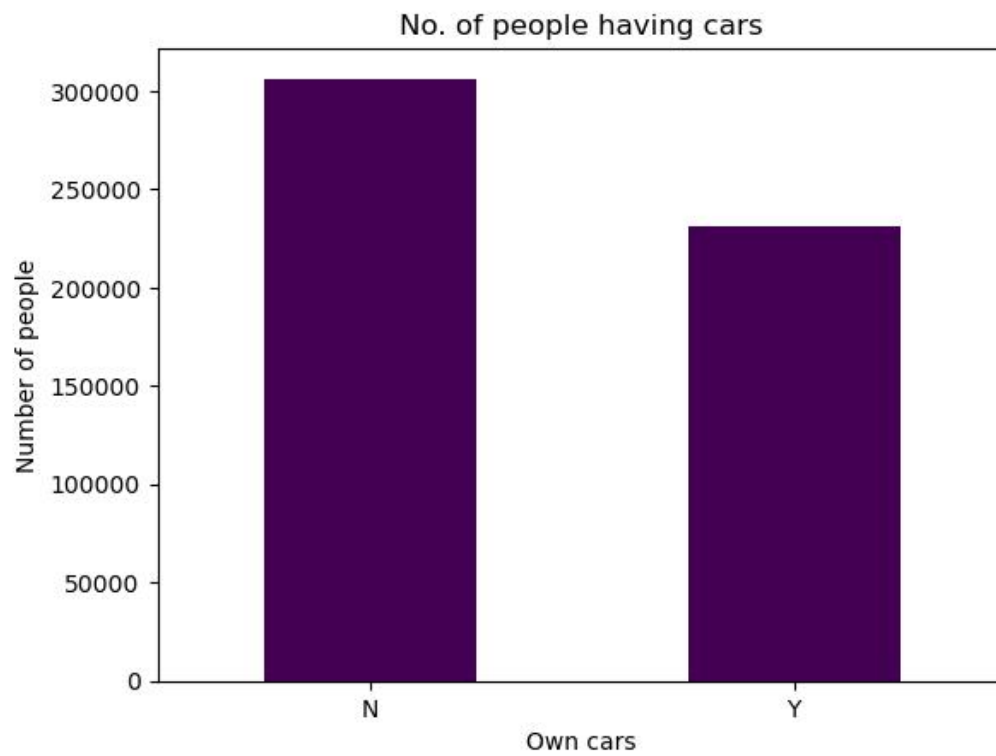
PYTHON LIBRARIES

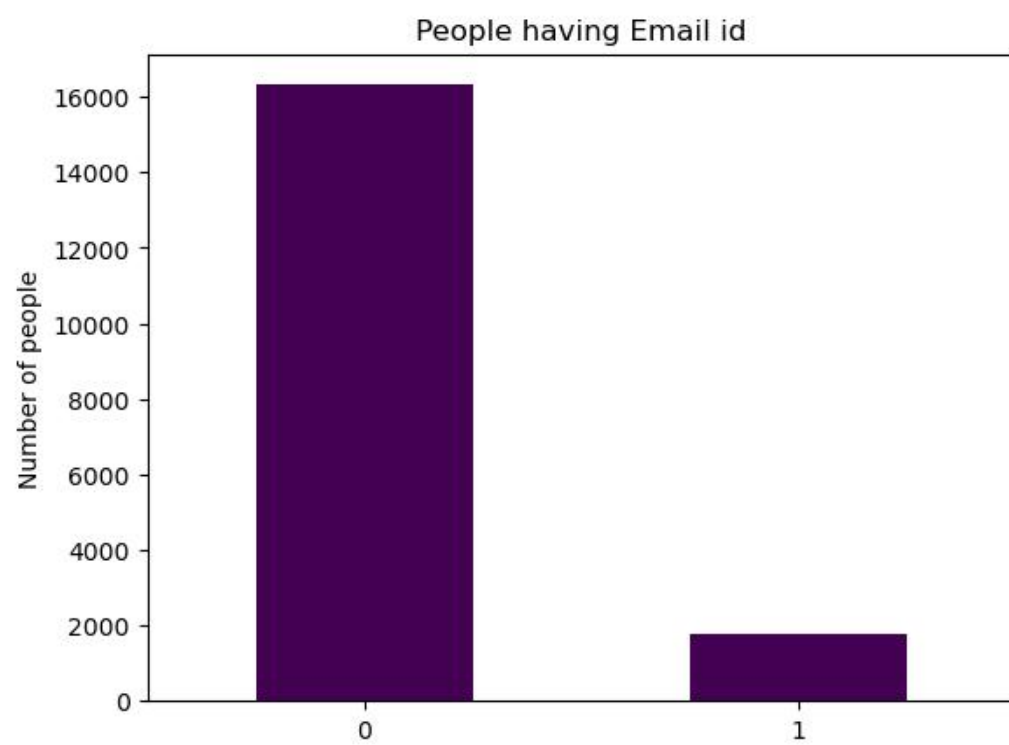
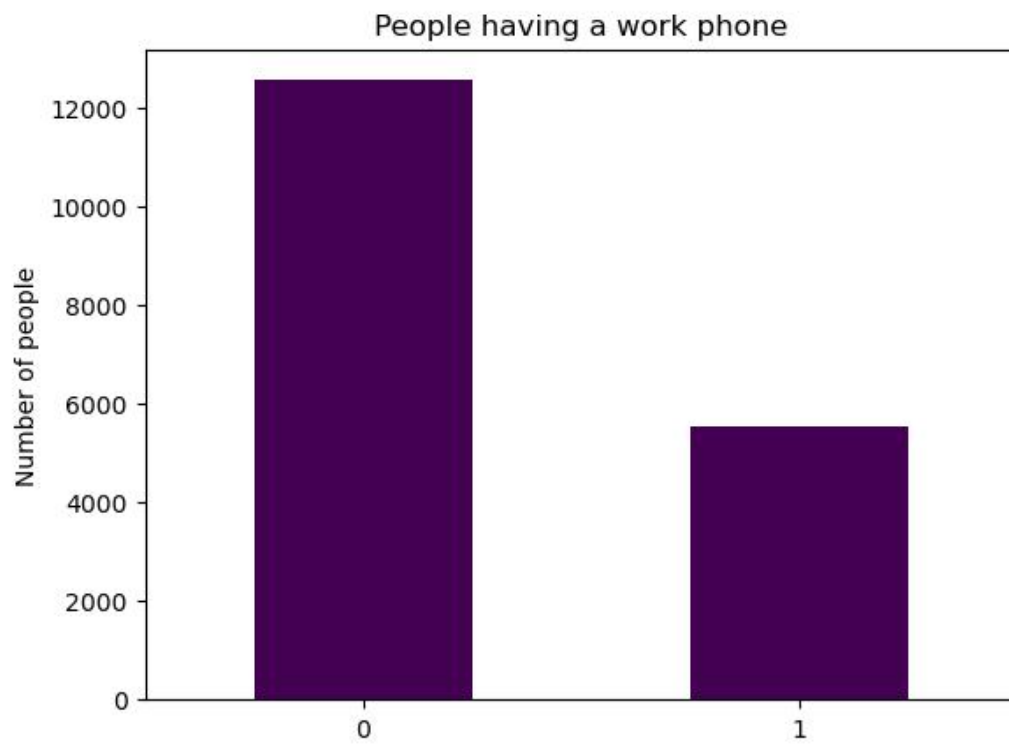
o Pandas: Pandas is defined as an open-source library that provides high performance data manipulation in Python. Data analysis becomes easier with this python library.

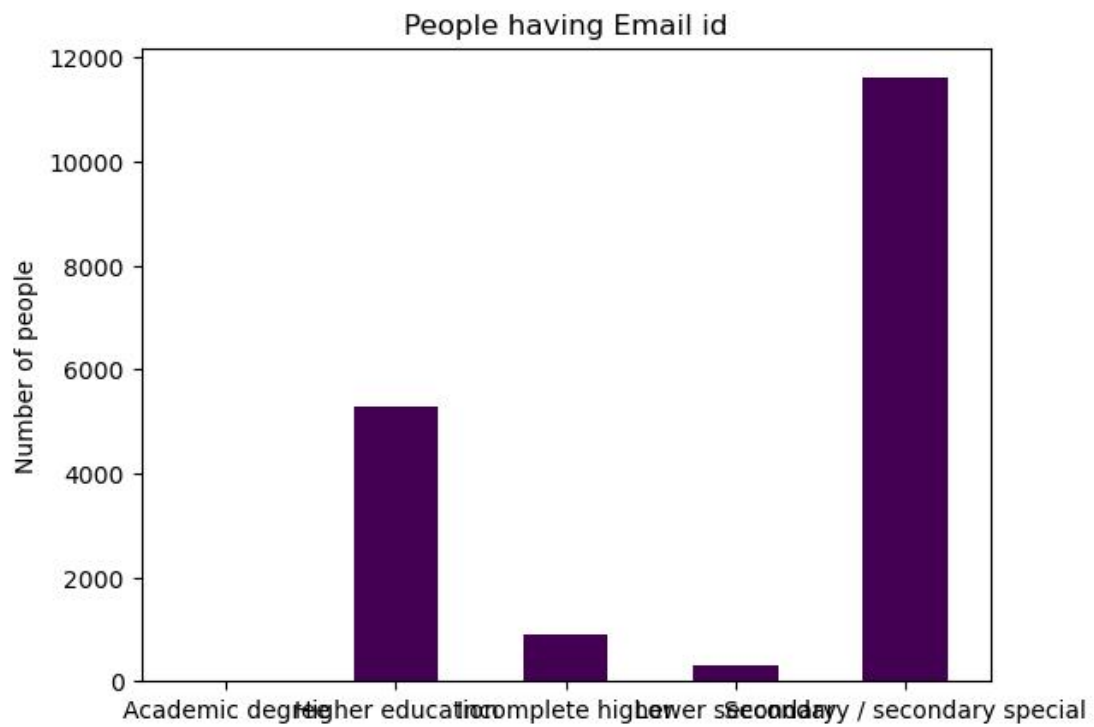
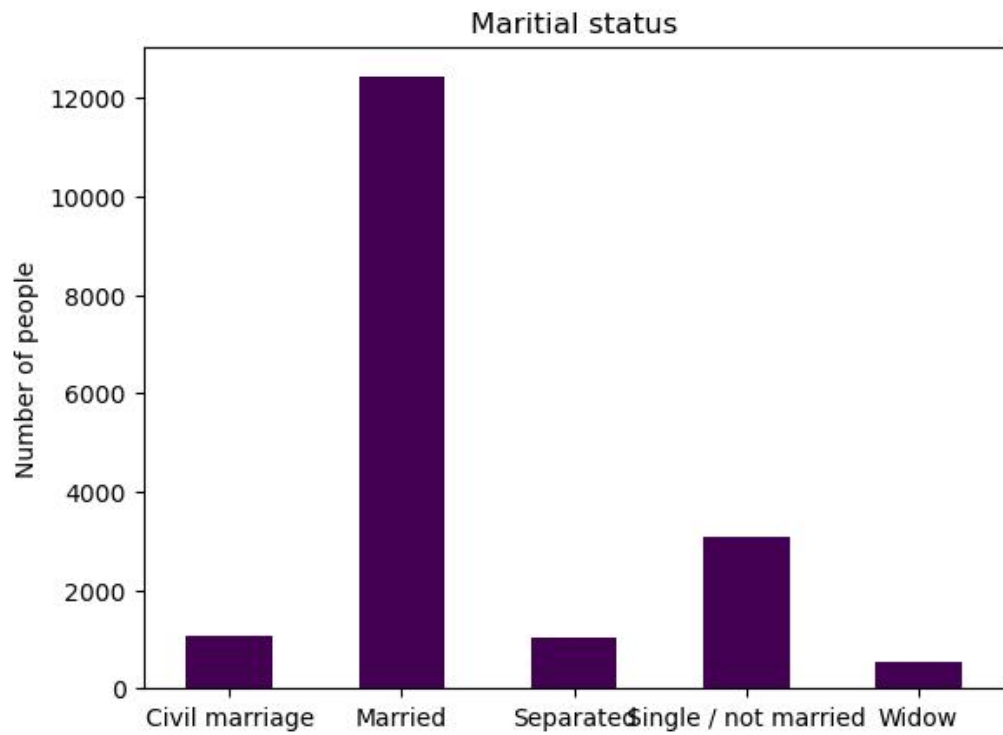
o Matplotlib: Matplotlib is a visualization library of python. We can graphically represent data and their relations through Matplotlib.

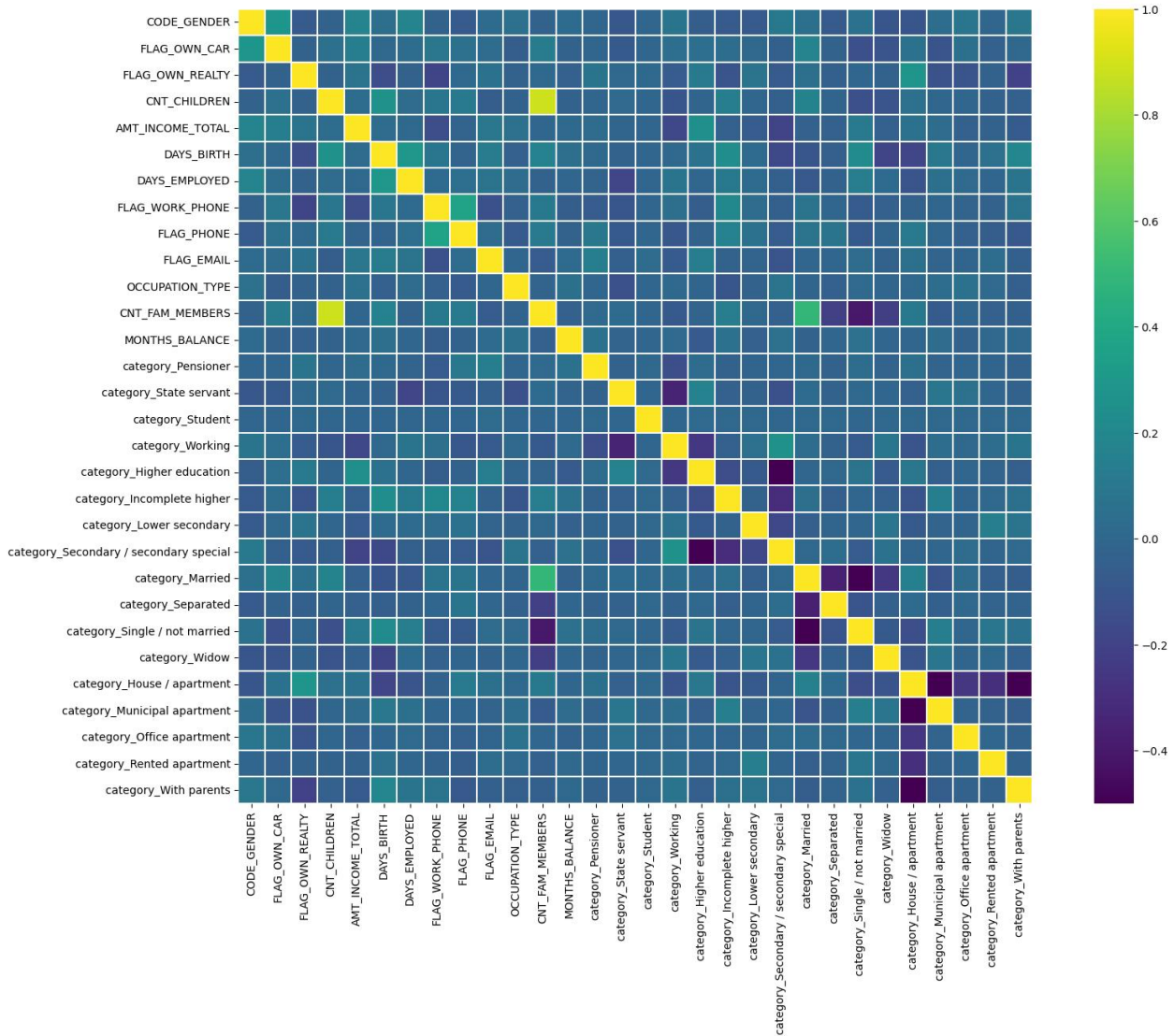
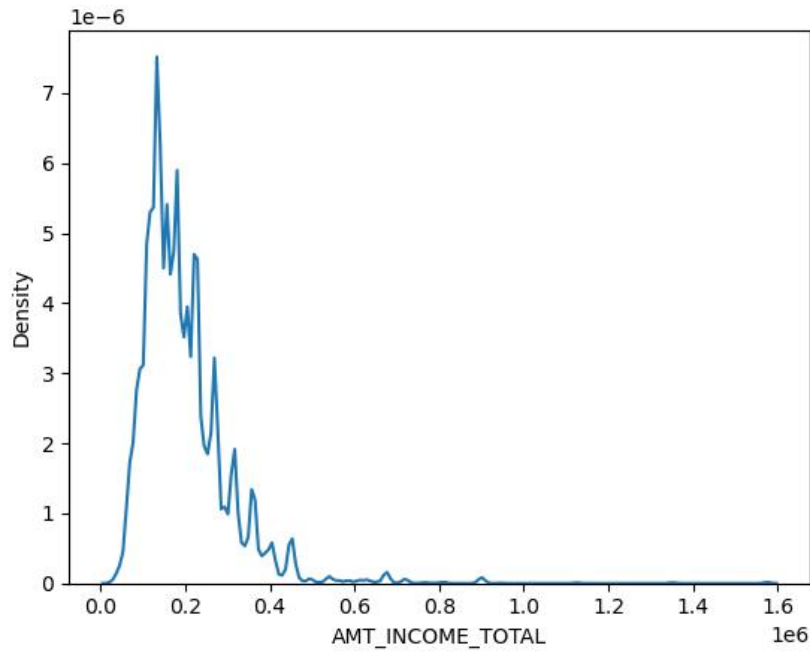
o Numpy: NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

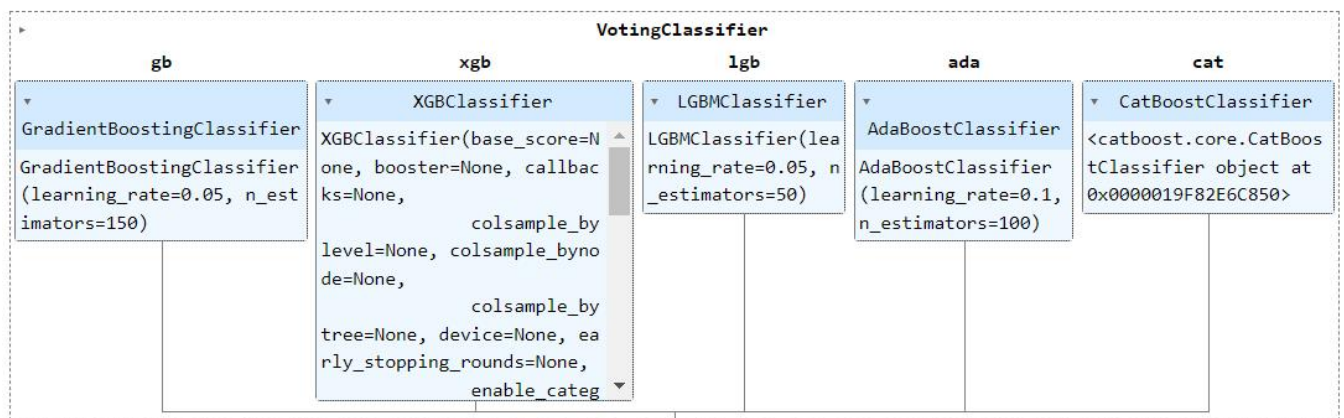
EXPLORATORY DATA ANALYSIS



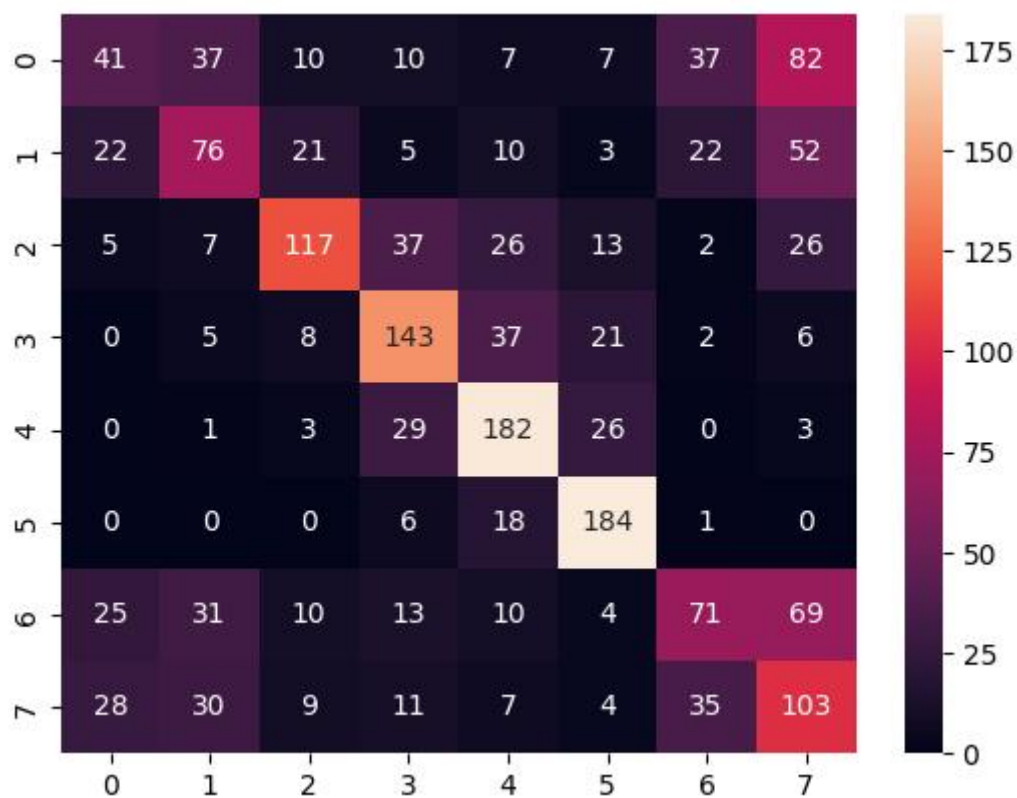








ROC score - 0.8625305132685666



WEB PAGE USER INTERFACE

Banking Fraud Detection

Select your Gender

☒ Male

☐ Female

Do you own any property

☒ Yes

☐ No

Enter Your Per Annum Income

0.00

-

+

Do you have a car

☒ Yes

☐ No

How many childers do you have

0.00

-

+

How many days passed since your Birthday

0.00

-

+

How many days you got employed

0.00

-

+

Do you have a mobile phone

☒ Yes

☐ No

Do you have a work phone

☒ Yes

☐ No

Do you have a another phone

☒ Yes

☐ No

Do you have a email

☒ Yes

☐ No

Select your Occupation

Laborers

▼

Enter your family members count

0.00

-

+

How Many years back you applied for loan through credit card

0.00

-

+

☒ Do you earn

☐ Are u Educated

☐ Are u over 18 years?

☐ Do you have a house

Calculate Fraudness

Streamlit Link : <https://feynn-internship-final-phase-aruj6hgbwprtjuu2xfemyw.streamlit.app/>

FINANCIAL EQUATION

Linear Growth Model: If the market is growing linearly, the financial equation would be:

$$y = (m * x(t)) - c$$

Where:

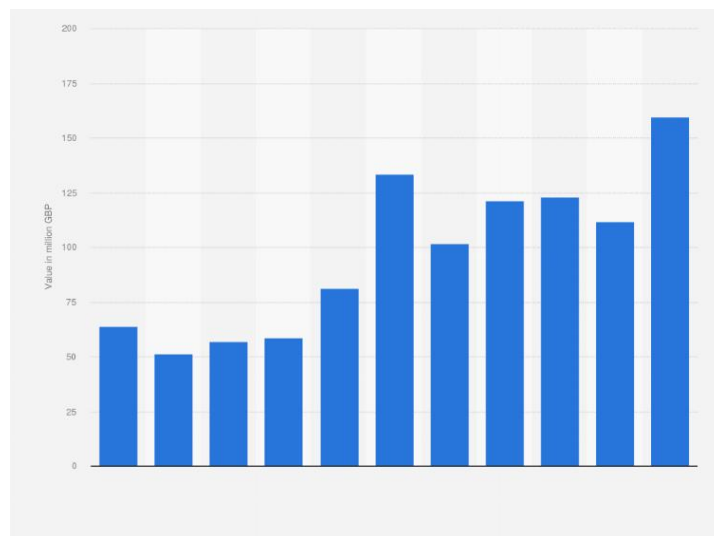
y = Total profit

m = Pricing of your product (subscription price)

x(t) = Total sales (market as a function of time)

c = Production, maintenance, and other fixed costs.

The graph presented illustrates the occurrence of banking fraud over the past decade. If this pattern persists, we anticipate an escalation in such fraudulent activities in the upcoming years.



The banking fraud detection app's annual subscription is priced at approximately 1000 INR. The combined production and maintenance expenses total around 150,000 INR. Hence, the Financial Equation would be -

$$y = 1000 * x(t) - 150000$$

CONCLUSION

The Fraud detection business model presents a promising and valuable opportunity for the banking, fin tech industry. By leveraging advanced data analytics, machine learning, and artificial intelligence techniques, the model aims to enhance and reduce the fraud detection in resulting in numerous benefits for both banks and their customers. If we look to our surroundings we can see various type of cyber crime, fraud in banking industry this thing is directly connected to the thousands of lives of our country, because banks and money play an important role in the economy of the country. Apart from this common people also suffer during fraud happening to a bank as their hardearned money comes to risk. All of these problems can be solved in a large scale by using our business model. In conclusion, the banking fraud detection business model holds the potential to revolutionize the banking industry, enabling improved quality, security, customer satisfaction, and a competitive advantage for companies that embrace this technology-driven approach. By aligning with sustainability goals and capitalizing on data insights, the model can foster long-term success and growth for the stakeholders involved. In a Single phrase we can say that our model will be a boom to the banking industry in near future.