**CREDIT CARD FRAUD DETECTION**

**Project Documentation**

**INTRODUCTION:**

Credit card fraud detection is the process of identifying and preventing unauthorized transactions on credit cards. It is a complex task, as fraudsters are constantly developing new methods to steal credit card information and use it for their own gain.

Credit card fraud detection systems typically use a combination of rules-based and machine learning-based methods to identify fraudulent transactions. Rules-based systems rely on a set of predefined rules to identify transactions that are likely to be fraudulent. For example, a rule-based system might flag a transaction that is made from a different country than the customer's usual location, or a transaction that exceeds a certain amount.

**Design thinking Process:**

Design thinking is a non-linear, iterative process that teams use to understand users, challenge assumptions, redefine problems, and create innovative solutions to prototype and test.

**Empathize**

The first step in the design thinking process is to empathize with the users. This means understanding their needs, wants, pain points, and motivations. In the context of credit card fraud detection, this could involve interviewing customers who have been victims of fraud, as well as merchants who have been impacted by fraud.

**Define**

Once the team has a good understanding of the users, they can begin to define the problem. This involves identifying the core challenge that the team is trying to solve. In the context of credit card fraud detection, the problem could be defined as:

**Ideate**

Once the problem has been defined, the team can begin to ideate solutions. This involves brainstorming a wide range of possible solutions, without judgment. The team should consider all possible solutions, even if they seem outlandish or impractical.

**Prototype**

Once the team has generated a list of potential solutions, they can begin to prototype them. This involves creating low-fidelity mockups or prototypes of the solutions, so that they can be tested with users. In the context of credit card fraud detection, this could involve prototyping a new fraud detection algorithm or developing a new user interface for a fraud detection system.

**Test**

Once the prototypes have been developed, they can be tested with users. This involves getting feedback from users on the prototypes and making necessary adjustments. The team should continue to iterate on the prototypes until they are satisfied that they have found a solution that meets the needs of the users.

**Implementation**

Once the team is satisfied with the prototype, they can begin to implement the solution. This involves developing the solution on a larger scale and making it available to users. In the context of credit card fraud detection, this could involve deploying a new fraud detection algorithm or launching a new fraud detection system.

**Phases Of Documentation**

**Phase1:MODEL TRAINING**

**Model Training:**

1. Preprocess the data to clean it and handle missing values.
2. Split the data into training and test sets.
3. Choose a machine learning algorithm, such as logistic regression, random forests, or SVMs.
4. Train the model on the training set.
5. Evaluate the model on the test set and tune the hyperparameters if needed.

Deploy the model to production

PROBLEM

Credit card fraud detection is the problem of identifying fraudulent credit card transactions. It is a challenging problem because fraudsters are constantly developing new methods to commit fraud.

Common types of credit card fraud:

Counterfeit cards

Lost or stolen cards

Card-not-present fraud

Identity theft

Credit card fraud detection methods:

Fraud detection systems

Human review

Credit card fraud detection is important to both cardholders and credit card companies to protect them from financial losses.

**Phase2:Feature Engineering**

**Feature Engineering:**

Feature engineering is the process of transforming raw data into features that are more informative and predictive for machine learning models. In the context of credit card fraud detection, feature engineering can be used to create features that capture the patterns and relationships in the data that are associated with fraud.

**Machine learning and AI:**

 Develop more sophisticated and effective fraud detection models.

**Big data:**

 Analyze large volumes of transaction data to identify patterns and trends that may indicate fraud.

**New data sources:**

 Supplement traditional transaction data with new data sources, such as social media data and geolocation data, to improve fraud detection accuracy.

**Real-time fraud detection:**

 Identify and flag fraudulent transactions as they occur to minimize the damage caused by fraud.

**Collaborative fraud detection:**

Share information about fraudulent transactions between different financial institutions to improve the overall effectiveness of fraud detection efforts.

**PROGRAM:**

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(df.drop('Class', axis=1), df['Class'], test\_size=0.25, random\_state=42)

# Encode categorical features

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

for col in ['V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10', 'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20', 'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'V29']:

X\_train[col] = le.fit\_transform(X\_train[col])

X\_test[col] = le.transform(X\_test[col])

**PROBLEM:**

Credit card fraud detection is the problem of identifying fraudulent credit card transactions. It is a challenging problem because fraudsters are constantly developing new methods.

Some of the most common types of credit card fraud include:

* Counterfeit cards
* Lost or stolen cards
* Card-not-present fraud
* Identity theft

Credit card fraud detection is important to both cardholders and credit card companies to protect them from financial losses.

There are a variety of methods that can be used to detect credit card fraud, including:

* Fraud detection systems
* Human review

Credit card fraud detection is a complex and challenging problem, but it is essential to protect both cardholders and credit card companies from financial losses.

Here are some of the challenges of credit card fraud detection:

* Data quality
* Data bias
* Model complexity

Despite these challenges, credit card fraud detection is an important task that can help to protect both cardholders and credit card companies from financial losses.

**Phase3:Data Explanation**

**Data Explanation:**

A credit card fraud detection dataset contains historical credit card transactions, both fraudulent and non-fraudulent. Some common features include:

Transaction amount

Transaction date and time

Merchant name

Cardholder location

Cardholder purchase history

**Additional features may include:**

Card type

Card number

Cardholder ZIP code

Cardholder CVV code

**Dataset Implementation:**

**1. Load and prepare the dataset.**

Clean the data by removing any errors or inconsistencies.

Handle missing values by dropping them or imputing them with a suitable value.

**2. Split the dataset into training and test sets.**

The training s

et will be used to train the model.

The test set will be used to evaluate the model's performance.

**3. Choose a machine learning algorithm.**

There are many different machine learning algorithms that can be used for credit card fraud detection, such as logistic regression, random forests, and support vector machines.

**4. Train the model.**

Feed the training data to the machine learning algorithm and allow it to learn the patterns in the data.

**5. Evaluate the model.**

Feed the test data to the trained model and measure its performance.

1. **Deploy the model**

Once the model is trained and evaluated, it can be deployed to a production environment so that it can be used to detect fraudulent transactions in real time.

**Build Load Dataset:**

**To load and prepare the dataset for credit card fraud detection:**

1. Download a publicly available dataset.
2. Load the dataset into your preferred programming language and machine learning framework.
3. Explore the dataset to understand the features and any potential problems.
4. Clean the data and handle missing values.
5. Consider creating new features.
6. Split the dataset into training and test sets.

**Build Program of Dataset:**

import pandas as pd

# Download the credit card fraud dataset from Kaggle

df = pd.read\_csv('creditcard.csv')

# Explore the dataset

print(df.head())

print(df.describe())

# Preprocess the data

# ...

# Split the dataset into training, validation, and test sets

X\_train, X\_val, X\_test, y\_train, y\_val, y\_test = train\_test\_split(df.drop('Class', axis=1), df['Class'], test\_size=0.2)

# Save the preprocessed datasets

X\_train.to\_csv('X\_train.csv', index=False)

X\_val.to\_csv('X\_val.csv', index=False)

X\_test.to\_csv('X\_test.csv', index=False)

y\_train.to\_csv('y\_train.csv', index=False)

y\_val.to\_csv('y\_val.csv', index=False)

y\_test.to\_csv('y\_test.csv', index=False

**Phase4:Demployment**

* Strong growth projected, much faster than average for all occupations.
* Driven by increasing use of credit cards and electronic payment methods, as well as growing sophistication of fraudsters.
* Professionals work for a variety of organizations, including banks, credit card companies, retailers, and government agencies.

**To be successful:**

* Strong analytical skills
* Ability to identify patterns in data
* Ability to think critically
* Ability to work independently and as part of a team

**Job titles in demand:**

* Fraud analyst
* Fraud investigator
* Fraud examiner
* Fraud prevention officer
* Security analyst

**How to prepare:**

* Get a degree in a related field, such as computer science, information security, or criminology.
* Gain experience in data analysis and fraud prevention.
* Network with other professionals in the field.
* Stay up-to-date on the latest fraud trends and technologies.

**Rewarding career:**

* Play a vital role in protecting consumers and businesses from financial losses.

**Feature Engineering:**

Feature engineering is the process of transforming raw data into features that are more informative and predictive for machine learning models. In the context of credit card fraud detection, feature engineering can be used to create features that capture the patterns and relationships in the data that are associated with fraud.

**Python Program:**

import pandas as pd

def create\_features(df):

"""Creates new features for credit card fraud detection.

Args:

df: A Pandas DataFrame containing the credit card transaction data.

Returns:

A Pandas DataFrame containing the new features.

"""

# Calculate the average transaction amount for each customer.

df["avg\_transaction\_amount"] = df["amount"].groupby(df["customer\_id"]).mean()

# Calculate the number of transactions made in the past 30 days for each customer.

df["num\_transactions\_past\_30\_days"] = df[df["transaction\_date"] >= pd.Timestamp.today() - pd.Timedelta(days=30)].groupby(df["customer\_id"]).size()

# Calculate the number of transactions made with the same merchant in the past 30 days for each customer.

df["num\_transactions\_with\_merchant\_past\_30\_days"] = df.groupby(["customer\_id", "merchant\_id"])[df["transaction\_date"] >= pd.Timestamp.today() - pd.Timedelta(days=30)].size()

# Calculate the distance between the customer's current location and the location of the merchant where the transaction is taking place.

df["distance\_to\_merchant"] = df.apply(lambda x: calculate\_distance(x["customer\_latitude"], x["customer\_longitude"], x["merchant\_latitude"], x["merchant\_longitude"]), axis=1)

# Return the new features.

return df

def calculate\_distance(customer\_latitude, customer\_longitude, merchant\_latitude, merchant\_longitude):

"""Calculates the distance between two points in kilometers.

Args:

customer\_latitude: The latitude of the customer's current location.

customer\_longitude: The longitude of the customer's current location.

merchant\_latitude: The latitude of the merchant where the transaction is taking place.

merchant\_longitude: The longitude of the merchant where the transaction is taking place.

Returns:

The distance between the two points in kilometers.

"""

# Calculate the distance in kilometers.

distance = geopy.distance.distance((customer\_latitude, customer\_longitude), (merchant\_latitude, merchant\_longitude)).km

return distance

# Usage:

# Load the credit card transaction data.

df = pd.read\_csv("credit\_card\_transactions.csv")

# Create new features.

df = create\_features(df)

# Save the new features to a new CSV file.

df.to\_csv("credit\_card\_transactions\_with\_features.csv", index=False)

**Output:**

# Load the credit card transaction data.

df = pd.read\_csv("credit\_card\_transactions.csv")

# Create new features.

df = create\_features(df)

# Save the new features to a new CSV file.

# Calculate the average transaction amount for each customer.

df["avg\_transaction\_amount"] = df["amount"].groupby(df["customer\_id"]).mean()

**CONCLUSION:**

Credit card fraud is a serious problem, but credit card fraud detection systems are essential for protecting consumers and businesses.There are a number of different approaches to credit card fraud detection, but all systems typically use a combination of rules-based and machine learning-based methods to identify and prevent fraud.