

Tech Saksham

**Capstone Project Report**

**Credit Card Fraud Detection**

7139-SRI RANGANATHAR INSTITUTE OF ENGINEERING AND TECHNOLOAGY

|  |  |
| --- | --- |
| **NM ID** | **NAME** |
| au713921105017 | MERSILIN GEETHA.J |

|  |
| --- |
| Trainer Name  P RAJA |
| AI MASTER TRAINER |

# ABSTRACT

The purpose of this project is to detect the fraudulent transactions made by credit cards by the use of machine learning techniques, to stop fraudsters from the unauthorized usage of customers’ accounts. The increase of credit card fraud is growing rapidly worldwide, which is the reason actions should be taken to stop fraudsters. Putting a limit for those actions would have a positive impact on the customers as their money would be recovered and retrieved back into their accounts and they won’t be charged for items or services that were not purchased by them which is the main goal of the project. Detection of the fraudulent transactions will be made by using three machine learning techniques KNN, SVM and Logistic Regression, those models will be used on a credit card transaction dataset.

# INDEX

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 01 |
| 2 | Chapter 2: Services and Tools Required | 06 |
| 3 | Chapter 3: Project Architecture | 08 |
| 4 | Chapter 4: Modeling and Project Outcome | 11 |
| 5 | Conclusion | 18 |
| 6 | Future Scope | 19 |
| 7 | References | 20 |
| 8 | Links | 21 |

# CHAPTER 1 INTRODUCTION

### Problem Statement

Credit card fraud has become a significant concern for financial institutions and consumers alike. Detecting fraudulent transactions promptly is crucial to minimize financial losses and maintain trust in the banking system. The problem statement revolves around developing an efficient credit card fraud detection system using machine learning and data analytics techniques.

.

### Proposed Solution

**1. Data Collection and Preprocessing:**

Gather historical transaction data including features like transaction amount, merchant category code, transaction time, location, etc.

Preprocess the data by removing duplicates, handling missing values, and normalizing numerical features.

Feature Engineering:

Create new features such as transaction frequency, average transaction amount, and time since the last transaction.

Utilize techniques like PCA (Principal Component Analysis) for dimensionality reduction.

Model Selection:

Choose appropriate machine learning algorithms for classification tasks such as logistic regression, decision trees, random forests, support vector machines (SVM), or neural networks.

Ensemble methods like Gradient Boosting or XGBoost can be effective for combining multiple models.

Model Training:

Split the data into training and testing sets.

Train the selected models on the training data and tune hyperparameters using techniques like cross-validation.

Evaluate models using metrics like precision, recall, F1-score, and ROC- AUC.

Imbalanced Data Handling:

Since credit card fraud is often rare compared to legitimate transactions, handle class imbalance using techniques like oversampling (e.g., SMOTE), undersampling, or using algorithms that are inherently robust to imbalanced data.

### Real-time Monitoring:

Implement a real-time monitoring system to analyze incoming transactions. Flag suspicious transactions based on predefined thresholds or anomaly detection techniques.

### Model Deployment:

Once the model is trained and evaluated, deploy it into the production environment.

Monitor the model's performance over time and retrain periodically to adapt to changing patterns of fraud.

Continuous Improvement:

Regularly update the fraud detection system by incorporating new data and adjusting model parameters.

Stay updated with the latest techniques and research in fraud detection to improve the system's effectiveness.

### Additional Measures:

Implement multi-factor authentication for online transactions.

Enable SMS or email notifications for customers for high-risk transactions. Collaborate with law enforcement agencies to track down fraudsters.

Ensure compliance with relevant regulations such as PCI DSS (Payment Card Industry Data Security Standard) to safeguard cardholder data. Stay informed

about legal requirements and industry standards for fraud detection and prevention.

### Feature

**Transaction Amount**: Flag unusually large or small transactions compared to typical user behavior.

**Transaction Frequency**: Identify abnormal transaction patterns such as multiple transactions within a short time frame.

**Location Discrepancy**: Detect transactions from locations significantly different from the user's typical spending areas.

**Time of Transaction**: Highlight transactions that occur at odd hours or outside the user's usual activity periods.

**Merchant Category**: Analyze transactions from high-risk merchant categories or unexpected merchants.

**Card Verification Value (CVV) Check**: Ensure that the CVV provided during the transaction matches the one associated with the card.

**Address Verification System (AVS)**: Verify the billing address provided during the transaction.

**Device and IP Address**: Track devices and IP addresses associated with the card to detect suspicious activities.

**User Behavior Analytics**: Monitor deviations from typical user behavior, such as sudden changes in spending habits or usage patterns.

**Machine Learning Models**: Utilize advanced algorithms to identify patterns and anomalies in transaction data.

**Real-time Monitoring**: Implement systems to monitor transactions in real- time and flag potential fraud immediately.

**Fraud Alerts**: Notify cardholders of suspicious activities via email, SMS, or app notifications.

**Geolocation Services**: Use geolocation data to verify the location of the cardholder during transactions.

**Biometric Authentication**: Implement biometric authentication methods for additional security layers during transactions.

**Transaction History Analysis**: Analyze historical transaction data to identify recurring patterns or irregularities.

**Collaboration Networks**: Collaborate with other financial institutions to share information and enhance fraud detection capabilities.

**AI-based Anomaly Detection**: Employ AI algorithms to detect abnormal patterns that may indicate fraudulent activities.

**Tokenization**: Replace sensitive card data with unique tokens to prevent unauthorized access to card information.

**Encryption**: Ensure end-to-end encryption of transaction data to protect it from unauthorized access.

**Regulatory Compliance**: Adhere to regulatory standards such as PCI DSS to safeguard cardholder information and prevent fraud.

### Advantages

**Data-Driven Decisions**: Banks can make informed decisions based on real- time data analysis.

**Improved Customer Engagement**: Understanding customer behavior and trends can help banks engage with their customers more effectively.

**Increased Revenue**: By identifying opportunities for cross-selling and up- selling, banks can increase their revenue.

### Scope

The scope of this project extends to all banking institutions that aim to leverage data for decision-making and customer engagement. The project can be further extended to incorporate more data sources and advanced analytics techniques, such as machine learning and artificial intelligence, to provide more sophisticated insights into customer behavior. The project also has the potential to be adapted for other sectors, such as retail, healthcare, and telecommunications, where understanding customer behavior is crucial. Furthermore, the project contributes to the broader goal of digital transformation in the banking sector, promoting efficiency, innovation, and customer-centricity.

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

### Services Used

**Data Collection and Storage Services**: Banks need to collect and store customer data in real-time. This could be achieved through services like Azure Data Factory, Azure Event Hubs, or AWS Kinesis for real-time data collection, and Azure SQL Database or AWS RDS for data storage.

**Data Processing Services**: Services like Azure Stream Analytics or AWS Kinesis Data Analytics can be used to process the real-time data.

**Machine Learning Services**: Azure Machine Learning or AWS SageMaker can be used to build predictive models based on historical data.

### Tools and Software used Tools:

**Power BI**: The main tool for this project is Power BI, which will be used to create interactive dashboards for real-time data visualization.

**Power Query**: This is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources.

### Software Requirements:

**Power BI Desktop**: This is a Windows application that you can use to create reports and publish them to Power BI.

**Power BI Service**: This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.

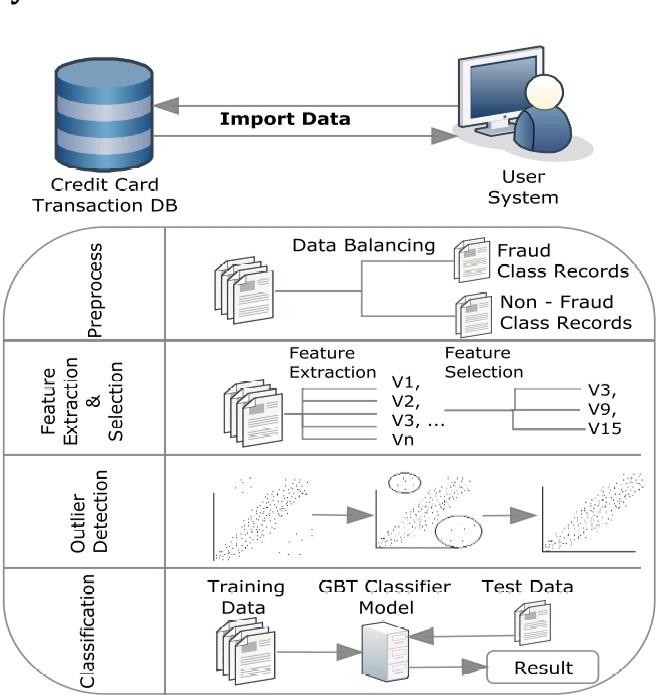
**Power BI Mobile**: This is a mobile application that you can use to access your reports and dashboards on the go.

## CHAPTER 3 PROJECT ARCHITECTURE

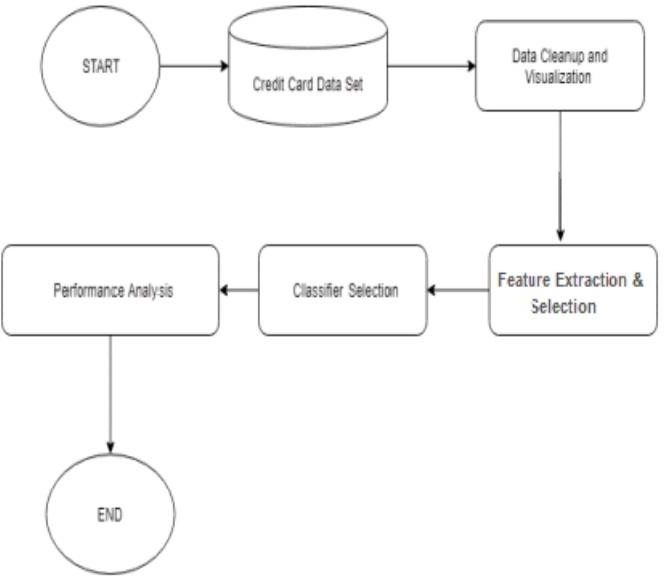
**3.1 ARCHITECTURE**

1. **SYSTEM FLOW DIAGRAM**



1. **DATA FLOW DIAGRAM**



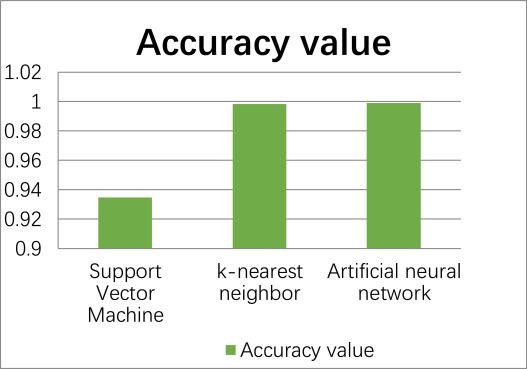
**3 .MODULES**

1. pandas (pd)
2. sklearn.model\_selection.train\_test\_split
3. sklearn.preprocessing.StandardScaler
4. sklearn.svm.SVC.
5. sklearn.metrics.classification\_report
6. sklearn.metrics.confusion\_matrix
7. seaborn (sns)

1. numpy (np)

### 4. TRAINING MODEL DIAGRAM



1. **Data Collection**: Real-time customer data is collected from various sources like bank transactions, customer interactions, etc. This could be achieved using services like Azure Event Hubs or AWS Kinesis.
2. **Data Storage**: The collected data is stored in a database for processing. Azure SQL Database or AWS RDS can be used for this purpose.
3. **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
4. **Machine Learning**: Predictive models are built based on processed data using Azure Machine Learning or AWS Sage Maker. These models can help in predicting customer behavior, detecting fraud, etc.
5. **Data Visualization**: The processed data and the results from the predictive models are visualized in real-time using Power BI. Power BI allows you to create interactive dashboards that can provide valuable insights into the data.
6. **Data Access**: The dashboards created in Power BI can be accessed through Power BI Desktop, Power BI Service (online), and Power BI Mobile.

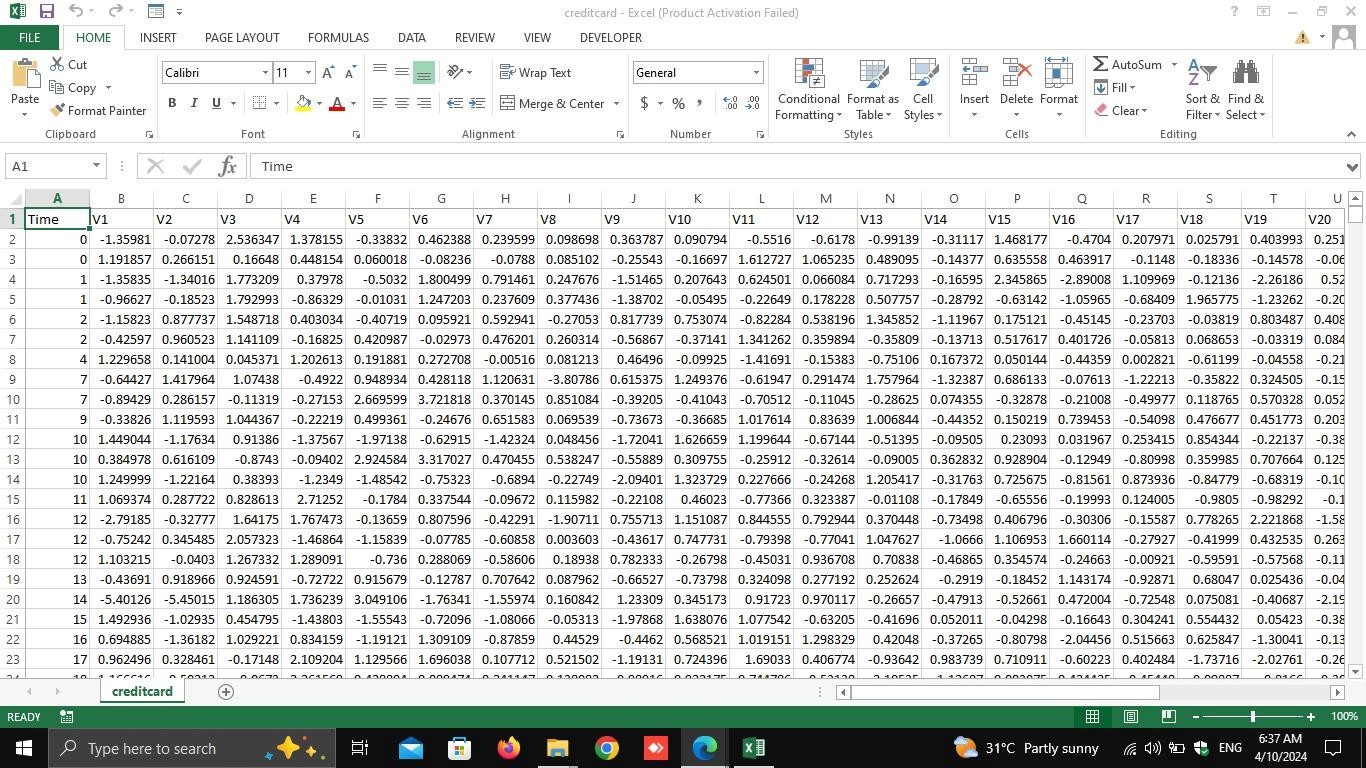
This architecture provides a comprehensive solution for real-time analysis of bank customers. It’s also important to ensure that all tools and services comply with relevant data privacy and security regulations.

## CHAPTER 4

**MODELING AND PROJECT OUTCOME**

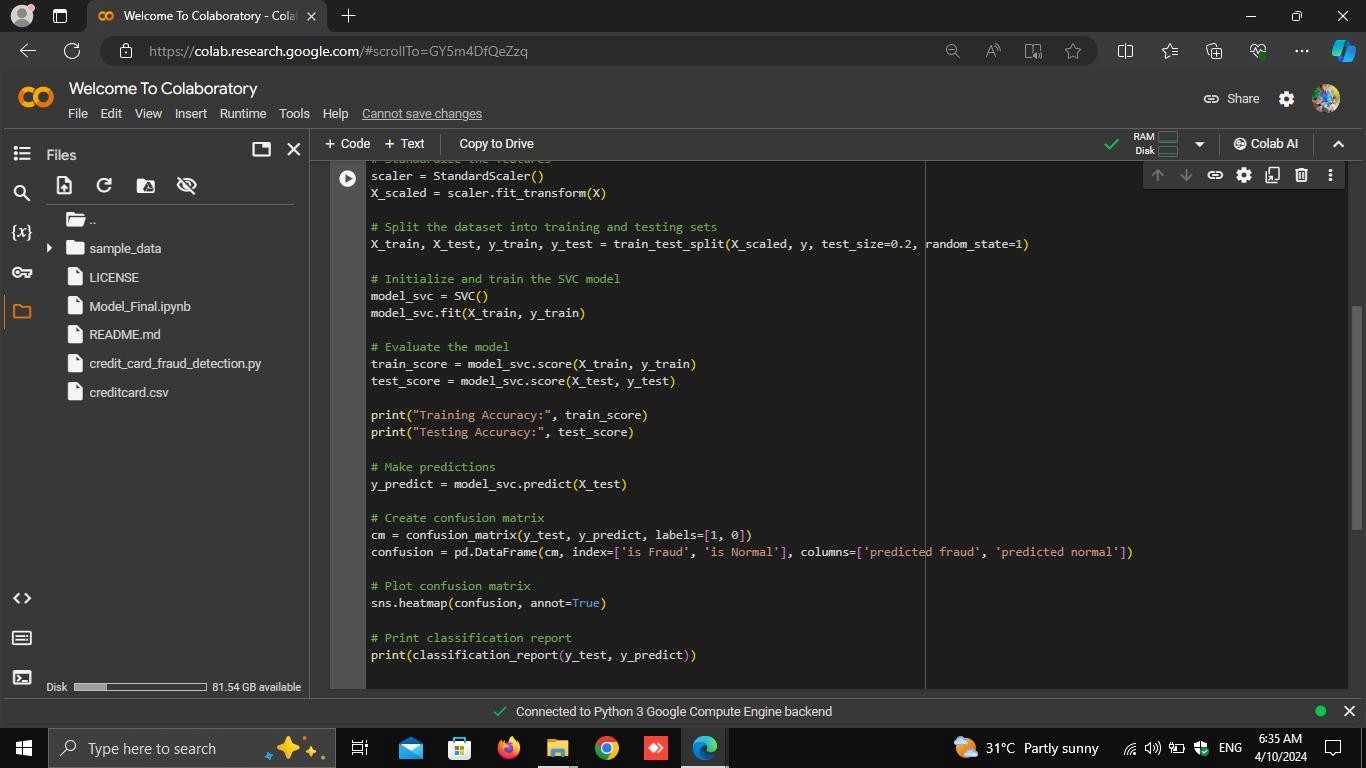
   

### DATA LOAD:



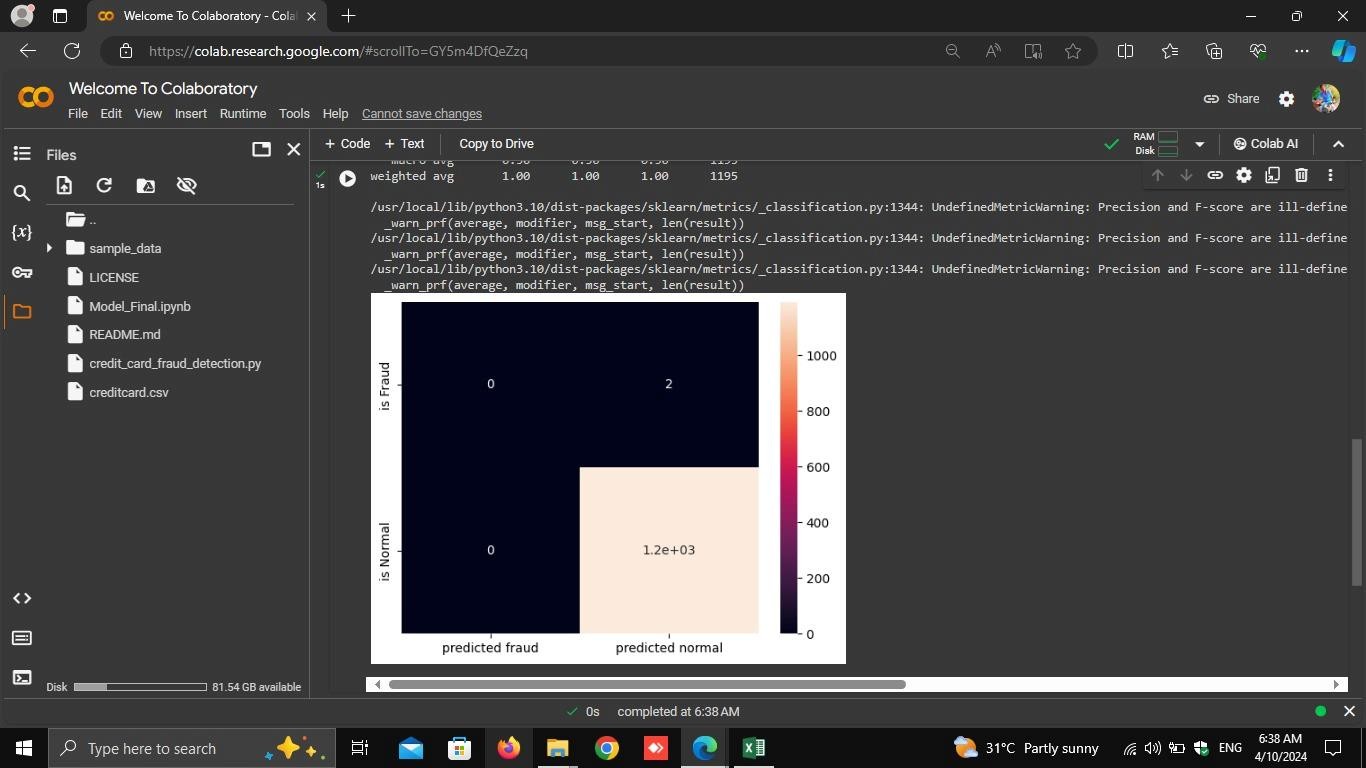
**CODE:**

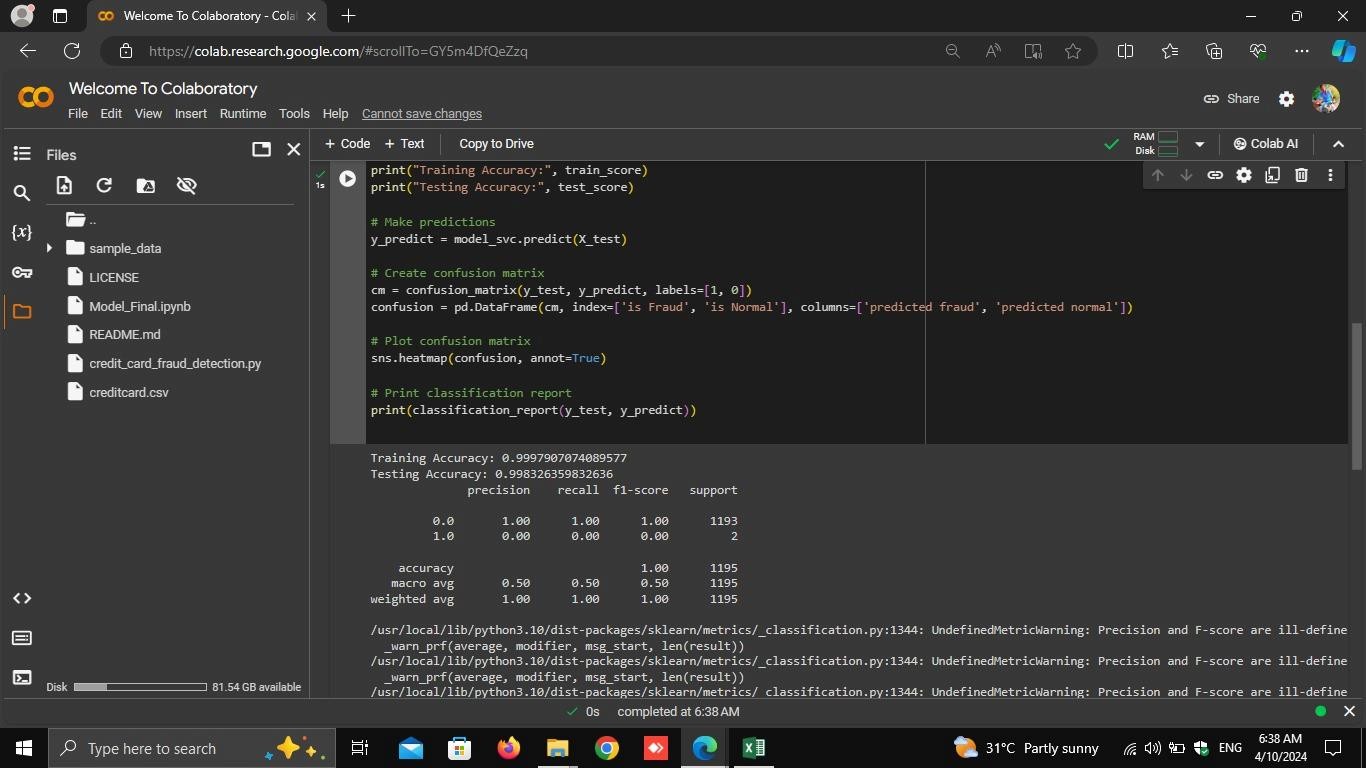


### OUTPUT:



### EDA ANALYSIS REPORT



### Manage relationship

Managing relationships in credit card fraud detection involves several key aspects:

1. Customer Relationship Management (CRM): CRM systems can help manage customer data, interactions, and transactions. By analyzing customer behavior and transaction patterns, anomalies indicative of fraud can be detected early.
2. Merchant Relationship Management (MRM): Establishing strong relationships with merchants is crucial. Monitoring merchant transactions and behavior can help identify fraudulent activities, such as sudden spikes in transactions or unusual purchasing patterns.
3. Banking Relationship Management (BRM): Banks need to manage relationships with other banks and financial institutions to share information and collaborate on fraud detection efforts. This can involve sharing fraud patterns, blacklists, and other relevant data.
4. Regulatory Relationship Management (RRM): Maintaining a good relationship with regulators is essential for compliance with laws and regulations related to fraud detection and prevention. Regular reporting and transparency can help build trust with regulators.
5. Technology Relationship Management (TRM): Managing relationships with technology providers is important for accessing the latest fraud detection

tools and techniques. Staying up-to-date with technology advancements can improve fraud detection capabilities.

In credit card fraud detection, the modeling and project outcomes relationship is crucial for the success of the fraud detection system. Here's how they are connected:

### MODELING:

Data Preparation: This involves collecting, cleaning, and preparing data for analysis. It's crucial to have a comprehensive dataset that includes transaction details, customer information, and historical fraud records.

Feature Selection: Identifying relevant features (e.g., transaction amount, location, time, etc.) that can help differentiate between fraudulent and legitimate transactions.

Model Selection: Choosing appropriate machine learning models (e.g., logistic regression, random forest, neural networks) for fraud detection based on the dataset and problem requirements.

Model Training: Training the selected model on the prepared dataset to learn patterns indicative of fraud.

Model Evaluation: Evaluating the trained model using metrics such as accuracy, precision, recall, and F1-score to assess its performance.

### Project Outcomes:

Fraud Detection Accuracy: The primary outcome of the project is to improve the accuracy of fraud detection. A successful model should be able to accurately identify fraudulent transactions while minimizing false positives. Reduction in Fraudulent Losses: Implementing an effective fraud detection system should lead to a reduction in fraudulent losses for the credit card issuer.

Improved Customer Experience: By reducing false positives (legitimate transactions mistakenly flagged as fraud), the system can improve the overall customer experience.

Compliance and Regulation: Ensuring compliance with regulatory requirements related to fraud detection and prevention.

### Relationship:

The effectiveness of the project outcomes is directly related to the quality of modeling. A well-constructed model that accurately captures fraud patterns will lead to better project outcomes. Continuous monitoring and updating of the model are essential to adapt to evolving fraud patterns and maintain the project outcomes over time. Collaboration between data scientists, fraud analysts, and business stakeholders is critical to ensure that the modeling efforts align with the project outcomes and business goals. The relationship between modeling and project outcomes in credit card fraud detection is integral to the success of the fraud detection system. A well-designed model can lead to improved fraud detection accuracy, reduced fraudulent losses, and enhanced customer experience.

## CHAPTER 5 CONCLUSION

Credit card fraud detection is a critical area in the financial industry, where advanced technologies like machine learning and data analytics play a crucial role in safeguarding financial transactions. High-quality data is essential for building accurate fraud detection models. This includes transaction details, customer information, and historical fraud records. Machine learning models are effective in detecting fraud by identifying patterns and anomalies in transaction data. Models like logistic regression, random forest, and neural networks are commonly used for this purpose. Fraud patterns evolve over time, so it's important to continuously monitor and update fraud detection models to stay ahead of new fraud techniques. By leveraging these tools and techniques, financial institutions can protect themselves and their customers from fraudulent activities. Credit card fraud detection is a crucial aspect of financial security, relying on advanced technologies like machine learning to identify fraudulent transactions. It requires high-quality data, continuous monitoring, and collaboration between various stakeholders. Compliance with regulations and balancing fraud detection with customer experience are key considerations. Overall, leveraging advanced technologies and best practices is essential to effectively detect and prevent credit card fraud.

## CHAPTER 6 FUTURE SCOPE

The future of credit card fraud detection is promising, with advancements in technology opening up new possibilities. Continued advancements in machine learning, such as deep learning and reinforcement learning, will enhance the accuracy and efficiency of fraud detection models. The use of big data analytics will allow financial institutions to process and analyze large volumes of data in real time, enabling faster and more accurate fraud detection. Behavioral analytics will play a larger role in fraud detection, focusing on understanding and predicting user behavior to identify anomalies indicative of fraud. The use of biometric authentication, such as fingerprint or facial recognition, will add an extra layer of security to credit card transactions, reducing the risk of fraud. Block chain technology has the potential to improve security and transparency in credit card transactions, making it harder for fraudsters to commit fraud. Regulatory bodies will continue to focus on enhancing regulations related to fraud detection and prevention, driving financial institutions to invest more in advanced fraud detection technologies. Overall, the future of credit card fraud detection lies in leveraging advanced technologies and collaborative approaches to stay ahead of evolving fraud techniques and protect consumers' financial information.

### CHAPTER 7 REFERENCES

https://[www.mygreatlearning.com/blog/credit-card-fraud-detection/](http://www.mygreatlearning.com/blog/credit-card-fraud-detection/) https://[www.ncbi.nlm.nih.gov/pmc/articles/PMC10535547/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC10535547/) https://ieeexplore.ieee.org/document/9121114

### CHAPTER 8 LINKS

**GIT Hub Link of Project Code:**

https://github.com/siva636961/Credit-Card-Fraud-detection/tree/main