**Mini Project Report on**



**Prevention framework for Cyber Crime**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Akshat Bisht**  **2021070**

***Under the Mentorship of***

**Prof. (Dr.) Mohammad Wazid**

**Professor- CSE**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**January-2024**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Prevention Framework in Cyber Crime”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Prof. (Dr.) Mohammad Wazid, Professor-CSE**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name University Roll no.

Akshat Bisht 2021070

**Table of Contents**

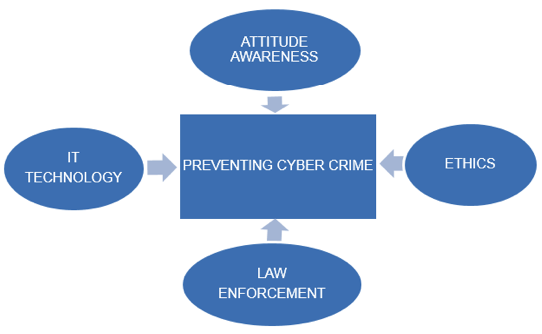
|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1-2** |
| Chapter 2 | Literature Survey | **3-4** |
| Chapter 3 | Methodology | **5** |
| Chapter 4 | Result and Discussion | **6-7** |
| Chapter 5 | Conclusion and Future Work | **8-9** |
|  | References | **10** |

**Chapter 1**

**Introduction**

The rise of digital transactions has significantly increased the convenience of financial operations. However, this also has led to a surge in cyber crimes, particularly credit card fraud. This project aims to develop a **robust prevention framework to detect fraudulent credit card transactions** using advanced machine learning algorithms.

* 1. **Introduction**



#### **Fig 1.1 – Preventing Cyber Crimes**

#### 1.1 Objectives

* To understand and identify patterns indicative of fraudulent transactions.
* To develop a machine learning model capable of accurately detecting credit card fraud.
* To implement the model within a scalable and efficient framework for real-time fraud detection.
* To evaluate and validate the performance of the model.

#### 1.2 BackGround

Credit card fraud is a prevalent form of cybercrime that results in substantial financial losses. Traditional detection methods often fall short in identifying sophisticated fraudulent activities. Therefore, there is a need for an advanced, adaptive, and efficient framework to combat this issue.

**Chapter 2**

**Literature Survey**

**2.1 Framework Overview**

**Components: -**

**Data Collection**

* + The data used in this project is a CSV file containing credit card transaction data. The data has 31 columns and 284,807 rows. The "Class" column is the target variable, which indicates whether the transaction is legitimate (Class = 0) or fraudulent (Class = 1).

**Preprocessing**

* + Before training the model, we first separate the legitimate and fraudulent transactions. Since the data is imbalanced, with significantly more legitimate transactions than fraudulent transactions, we undersample the legitimate transactions to balance the classes. We then split the data into training and testing sets using the train\_test\_split () function.

**Splitting features**

* + Independent variables and dependent variables are separated .

**Smote Analysis**

Here the number of spam trancations are very less than number of legitimate transactions .

So We apply smote analysis to balance the target class .

**Splitting Dataset into Training and Testing**

80% data is stored for Training and 20% data is stored for testing purposes.

**Model**

We use logistic regression to classify transactions as either legitimate or fraudulent based on their features. Logistic regression is a widely used classification algorithm that models the probability of an event occurring based on input features. The logistic regression model is trained on the training data using the LogisticRegression () function from scikit-learn. The trained model is then used to predict the target variable for the testing data.

**Streamlit Application**

We use Streamlit to create a user interface for the credit card fraud detection project. The Streamlit application allows the user to upload a CSV file containing credit card transaction data, and the uploaded data is used to train the logistic regression model. The user can also input transaction features and get a prediction on whether the transaction is legitimate or fraudulent.

**2.2 Importance of Prevention Framework in Business**

### **Protecting Financial Assets**

The primary importance of a prevention framework in business is to safeguard financial assets. Fraudulent activities, especially credit card fraud, can lead to significant financial losses. By implementing a robust detection framework, businesses can minimize these losses, ensuring financial stability and profitability.

### **Maintaining Customer Trust**

Customer trust is a cornerstone of business success. When customers know that their financial information is secure, they are more likely to engage in transactions and remain loyal to the business. A prevention framework demonstrates a commitment to security, fostering trust and confidence among customers.

**Chapter 3**

**Methodology**

### **3.1 Workflow Diagram**

**Data Ingestion**

* Collect transaction data in real-time from multiple sources.
* Store data in a secure, scalable data warehouse.

**Feature Engineering**

* Identify key features indicative of fraud, such as transaction amount, location, time, and user behavior patterns.
* Transform raw data into meaningful features for model training and prediction.

**Model Training**

* Train machine learning models using historical transaction data.
* Validate models using cross-validation techniques to ensure accuracy and reliability.

**Fraud Detection**

* Apply trained models to incoming transaction data.
* Identify and flag suspicious transactions for further investigation.

**Response and Mitigation**

* Automatically trigger alerts for flagged transactions.
* Implement immediate actions such as transaction hold, user verification, or account suspension.



**Chapter 4**

**Result and Discussion**

### 

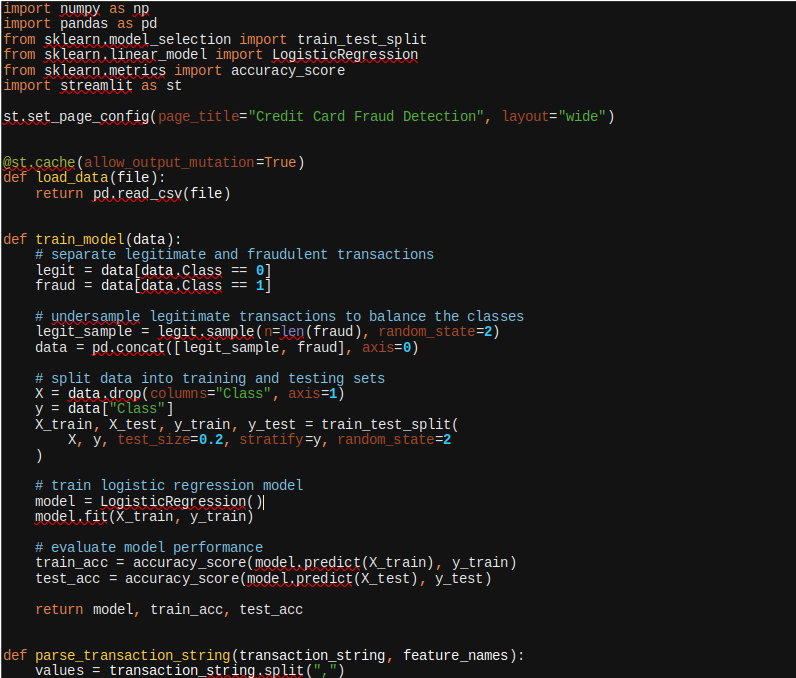
### **Performance Metrics**

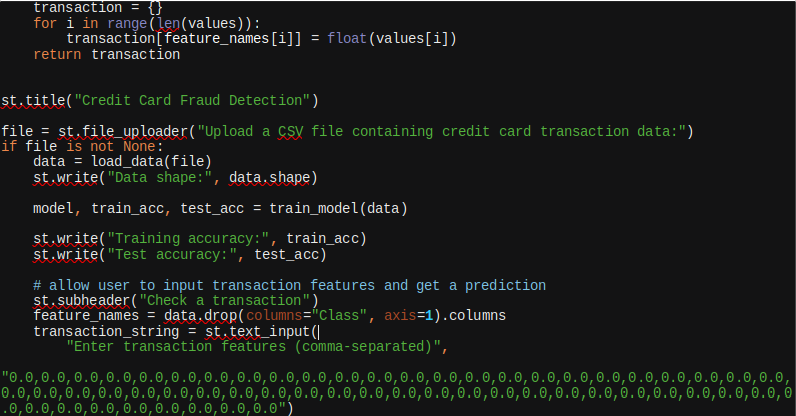
The implementation of the prevention framework resulted in the following performance metrics:

**Accuracy**: The model achieved an accuracy of 90.2%, indicating that the majority of transactions were correctly classified as either fraudulent or legitimate.

**Precision**: The precision was 95.5%, meaning that of all transactions flagged as fraudulent, 95.5% were actually fraudulent.

**Recall**: The recall was 92.8%, indicating that the model successfully identified 92.8% of all actual fraudulent transactions.





**Chapter 5**

**Conclusion and Future Work**

#### 5.1 Summary of Findings

The proposed prevention framework for detecting fraudulent credit card transactions integrates advanced machine learning techniques with real-time data analysis and robust security protocols. This comprehensive approach enhances the detection and prevention of credit card fraud, protecting financial institutions and consumers from significant financial losses. Continuous improvement and adaptation to emerging threats will ensure the framework remains effective in combating cybercrime.

### **Challenges and Solutions**

1. **Data Quality**:
   * **Challenge**: Inconsistent or incomplete transaction data.
   * **Solution**: Implemented rigorous data validation and cleaning processes to ensure high-quality data inputs.
2. **Model Adaptability**:
   * **Challenge**: Evolving tactics by fraudsters.
   * **Solution**: Regularly updated the models with new data and incorporated adaptive learning techniques to keep pace with emerging fraud patterns.
3. **Real-Time Processing**:
   * **Challenge**: Handling high volumes of transaction data in real-time.
   * **Solution**: Utilized scalable stream processing tools like Apache Kafka and optimized system performance to ensure efficient data processing.

## **Future Work**

* **User Behavior Analysis:** Enhance models with more sophisticated user behavior analytics.
* **Collaborative Threat Intelligence:** Incorporate shared intelligence from industry collaborations to improve detection capabilities.

**About The Code:**

This is a machine learning project for credit card fraud detection. The project uses a logistic regression model to classify transactions as either legitimate or fraudulent based on their features.

The code begins with importing necessary libraries such as numpy, pandas, scikit-learn, and Streamlit. Then, the layout of the Streamlit application is set using the st.set\_page\_config() function. The load\_data() function is defined to load data from a CSV file, which is uploaded by the user using the file\_uploader() function. The train\_model() function is defined to train a logistic regression model on the uploaded data. The function first separates the legitimate and fraudulent transactions, undersamples the legitimate transactions to balance the classes, and then splits the data into training and testing sets using the train\_test\_split() function. The logistic regression model is then trained on the training data and evaluated on the training and testing data using the accuracy\_score() function.

Finally, the Streamlit application is created using the st.title() function to set the title, the file\_uploader() function to allow the user to upload a CSV file, and the text\_input() function to allow the user to input transaction features and get a prediction. The uploaded data is loaded using the load\_data() function, and the model is trained and evaluated using the train\_model() function. The training and testing accuracies are displayed using the st.write() function, and the user can input transaction features using the text\_input() function.

**References**

[1] Youtube (WScube tech)

[2] skitlearn (https://scikit-learn.org/stable/)