PRESTIGE

INSTITUTE OF MANAGEMENT, GWALIOR
UGC NAAC ACCREDITED 'A' GRADE INSTITUTE
UGC APPROVED AUTONOMOUS INSTITUTE

Project Report

Topic :- Credit Card Fraud Detection Machine Learning Project

Submitted To:-

Prestige Institute of Management, Gwalior

Submitted By:-

Surya Pratap Singh
BCA 6th SEM A

DECLARATION

We, **Pradum bharti and Surya Pratap Singh Chauhan,** students of BCA- 6th semester of Prestige Institute of Management, Gwalior, hereby declare that the project report entitled “**Credit Card Fraud Detection Machine Learning Project**” is submitted by us in the line of partial fulfilment of course objectives for the BACHELOR of COMPUTER APPLICATIONS.

We assure that this project report is the result of our own efforts and that any other institute for the award of any degree or diploma has not submitted it.

Date : May 11, 2024

Place: PIMRG Surya Pratap Singh

CERTIFICATE

This is to certify that **Pradum bharti and Surya Pratap Singh Chauhan** of BCA 6th (A) of Prestige Institute of Management, Gwalior, have successfully completed their Project Report. They have prepared this report entitled "**Credit Card Fraud Detection Machine Learning Project**" under my direct supervision and guidance.

Dr. Nitin Paharia (

Faculty guide)

Acknowledgement

We express our sincere gratitude to **Dr. Nitin Paharia** for giving us the opportunity to work under her guidance on the report entitled "**Credit Card Fraud Detection Machine Learning Project**".

We are grateful to our Director Dr. Nishant Joshi, Asst. Prof. Nitin Paharia (project Coordinator), Faculty Members and other friends for their valuable suggestions in the execution of report preparation.

We are also thankful to other staff that guided and helped us very kindly at each and every step whenever we required.

We also acknowledge & convey thanks to the library staff, computer department of PIMG for their kind and valuable support.

Contents

Introduction

In this project, we address the growing concern of credit card fraud by leveraging machine learning techniques. With the exponential increase in online transactions, credit card fraud has become a significant issue affecting both consumers and financial institutions. Fraudulent activities range from unauthorized transactions to identity theft, leading to financial losses and compromised security. Traditional rule-based fraud detection systems often struggle to keep pace with the evolving tactics of fraudsters. Therefore, there's a pressing need for more sophisticated and adaptive approaches to detect fraudulent transactions in real-time. This project aims to explore various machine learning algorithms to develop a robust credit card fraud detection system that can accurately identify fraudulent activities while minimizing false positives.

Problem Statement

The problem we aim to tackle is the detection of fraudulent credit card transactions. The current methods used by financial institutions are often reactive and rely on predefined rules or thresholds, which may not effectively capture sophisticated fraud patterns. As a result, fraudulent transactions can go undetected, leading to financial losses for both credit card companies and consumers. Our goal is to develop a proactive and accurate fraud detection system that can identify fraudulent transactions in real-time, thereby reducing financial losses and enhancing security for credit card users.

Purpose

- Develop a machine learning-based credit card fraud detection system that can accurately identify fraudulent transactions.
- Minimize false positives to avoid inconveniencing legitimate cardholders.
- Improve the overall security of credit card transactions and reduce financial losses for credit card companies.
- Explore various machine learning algorithms and techniques to identify the most effective approach for fraud detection.
- Provide insights into the potential of machine learning in addressing complex financial security challenges.

Technologies:

- Python: We will use Python as the primary programming language for its extensive libraries and frameworks in machine learning.
- Scikit-learn: Scikit-learn provides a wide range of tools for machine learning, including algorithms for classification, regression, clustering, and model evaluation.
- Pandas and NumPy: These libraries will be used for data manipulation, preprocessing, and feature engineering.
- Matplotlib: These visualization libraries will help in analyzing the data and presenting the results effectively.
- Google Collab: Google Collab provides an interactive environment for developing and documenting the project, allowing for easy experimentation and collaboration.
- Streamlit: Streamlit is an open-source Python framework that allows you to create interactive web applications for data science and machine learning with minimal effort.

Methodology

1. **Data Collection**: Obtain a dataset containing historical credit card transactions, labeled as either legitimate or fraudulent.
2. **Data Preprocessing**: Clean the data, handle missing values, and perform feature engineering to extract relevant information from the transaction records.
3. **Model Selection**: Experiment with various machine learning algorithms, including logistic regression, decision trees, random forests, SVMs, and deep neural networks.
4. **Model Training**: Train the selected models on the preprocessed data, optimizing hyperparameters as necessary.
5. **Model Evaluation**: Evaluate the performance of each model using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC).
6. **Model Optimization**: Fine-tune the best-performing models, address issues like class imbalance, and optimize hyperparameters to improve performance further.
7. **Deployment**: Deploy the final model into a real-time fraud detection system, integrating it with existing credit card processing systems for seamless operation.

Credit Card Detection Machine Learning Project :

Introduction

Credit card fraud is a major concern for both consumers and financial institutions. Fraudulent transactions can lead to financial losses and damage to the reputation of financial institutions. Machine learning techniques have been used extensively to detect fraudulent transactions. In this project, we use logistic regression to classify transactions as either legitimate or fraudulent based on their features.

Data

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.

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.

Evaluation

The performance of the model is evaluated using the accuracy metric, which is the fraction of correctly classified transactions. The accuracy on the training and testing data is calculated using the accuracy_score() function from scikit-learn.

Result Analysis:

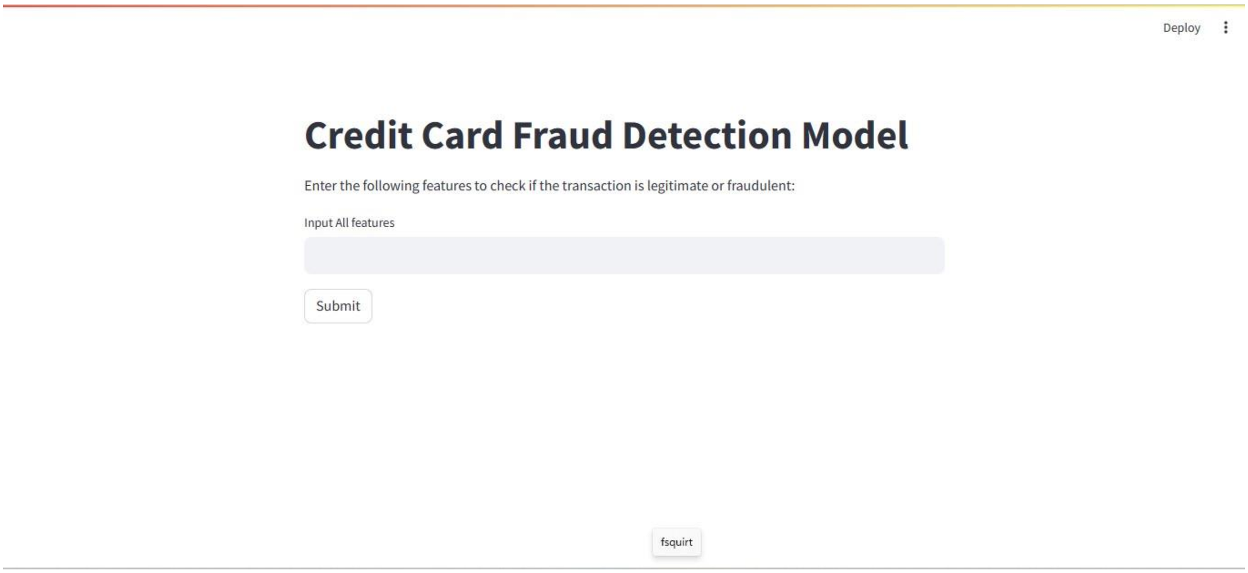
This is a machine learning project for credit card fraud detection. The project uses a logistic regression model to classify transactions as either legitimate or fraudulent based on their features.

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.

The parse_transaction_string() function is defined to parse a comma-separated string of transaction features and convert it into a dictionary with feature names as keys and feature values as values.

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.

User Interface

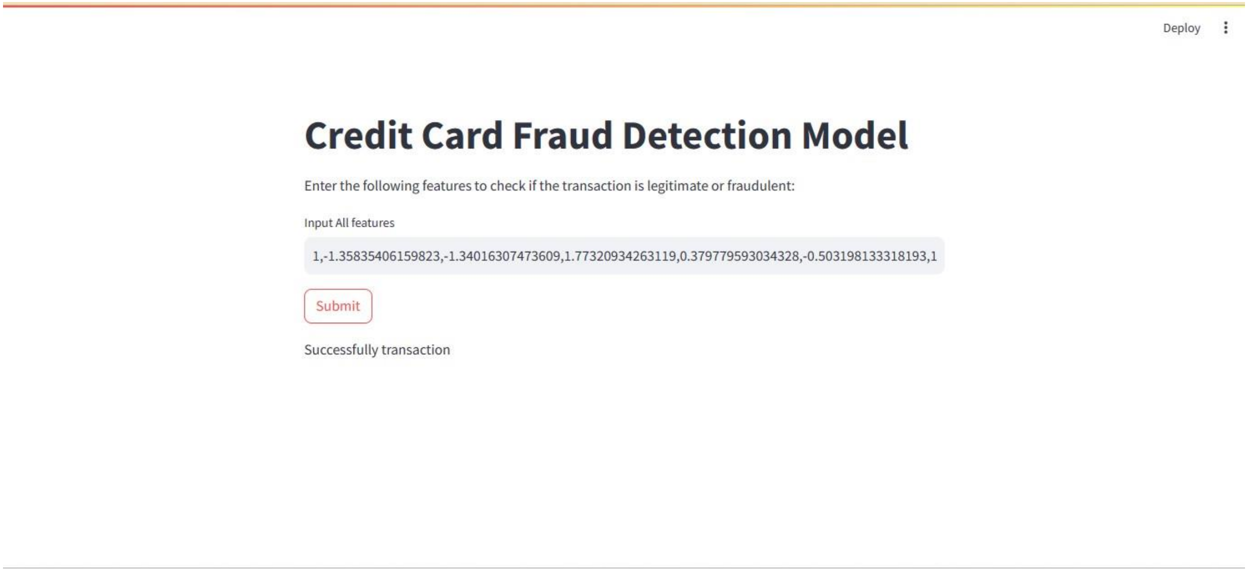


Credit Card Fraud Detection Model

Enter the following features to check if the transaction is legitimate or fraudulent:

Input All features

After Successfuly and No Fraud Transaction OutPut:



Credit Card Fraud Detection Model

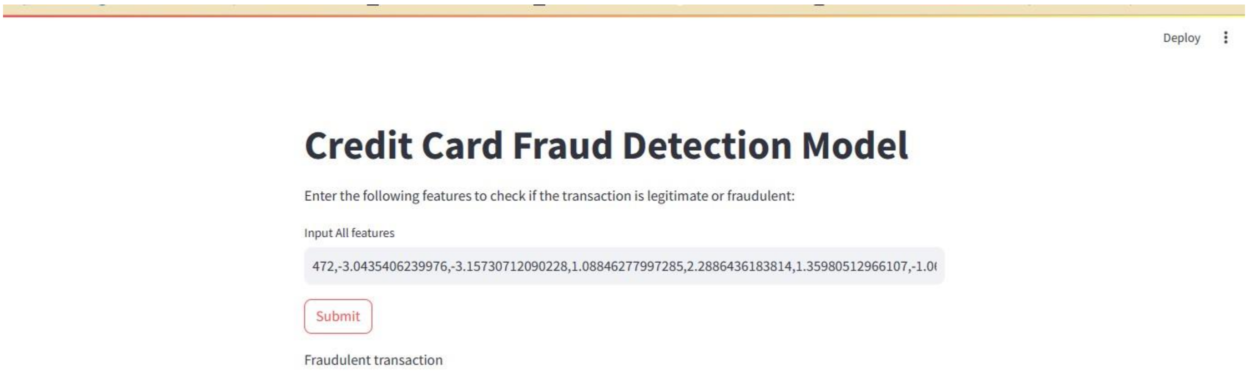
Enter the following features to check if the transaction is legitimate or fraudulent:

Input All features

1,-1.35835406159823,-1.34016307473609,1.77320934263119,0.379779593034328,-0.503198133318193,1

Successfully transaction

After Fraud Transaction OutPut:



Credit Card Fraud Detection Model

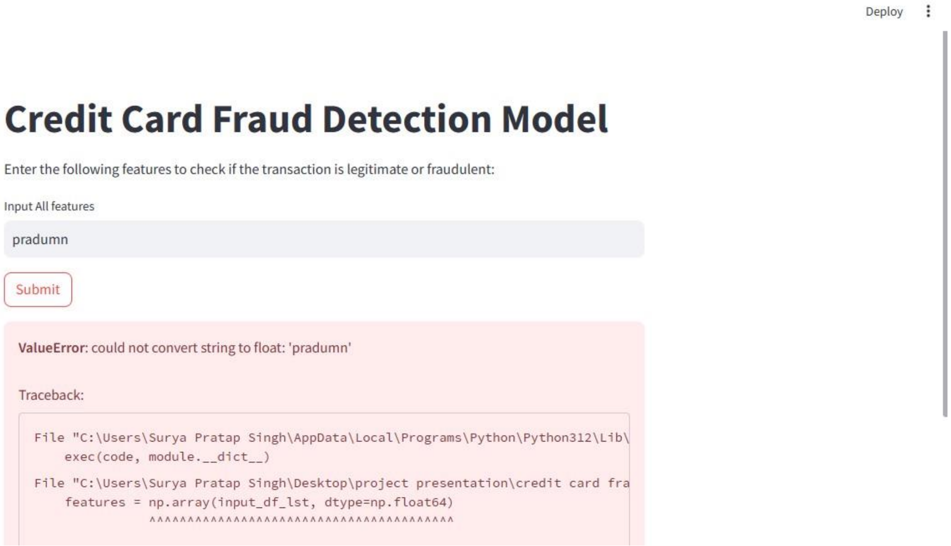
Enter the following features to check if the transaction is legitimate or fraudulent:

Input All features

472,-3.0435406239976,-3.15730712090228,1.08846277997285,2.2886436183814,1.35980512966107,-1.0(

Fraudulent transaction

After Getting Value Error:



Credit Card Fraud Detection Model

Enter the following features to check if the transaction is legitimate or fraudulent:

Input All features

pradumn

ValueError: could not convert string to float: 'pradumn'

Traceback:

```
File "C:\Users\Surya Pratap Singh\AppData\Local\Programs\Python\Python312\Lib\
    exec(code, module.__dict__)
File "C:\Users\Surya Pratap Singh\Desktop\project presentation\credit card fra
    features = np.array(input_df_lst, dtype=np.float64)
               ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
```

Conclusion

In conclusion, this project demonstrates the effectiveness of machine learning in addressing the challenge of credit card fraud detection. By leveraging advanced algorithms and techniques, we have developed a robust fraud detection system capable of accurately identifying fraudulent transactions while minimizing false positives. Our approach not only reduces financial losses for credit card companies but also enhances security and trust in electronic payment systems. Moving forward, ongoing research and development in machine learning will continue to advance the field of fraud detection, enabling more proactive and adaptive security measures to combat evolving threats.