Credit Card Fraud Detection System

Presented by:

- Cecille Jatulan
- David Higuera
- Diana Reyes
- Maria Melencio
- Michael Montanez

Table of contents

01

Introduction

02

Data collection and preparation

03

Methodology

04

Analysis and Results

05

Discussion

06

Conclusion

07

UI Demo

01

Introduction

Problem Statement

Project develops advanced Credit Card Fraud Detection Model to proactively identify and prevent fraud in real-time, addressing increasing sophistication in transactions

Significance and Relevance of the Project

Fraud Response

Project addresses sophisticated credit card fraud with effective detection methods to safeguard financial transactions

Security Boost

Using tech to boost fraud detection accuracy for secure financial transactions

Scalability and Adaptability

Designing scalable model for adaptable fraud detection to accommodate growing credit card transactions

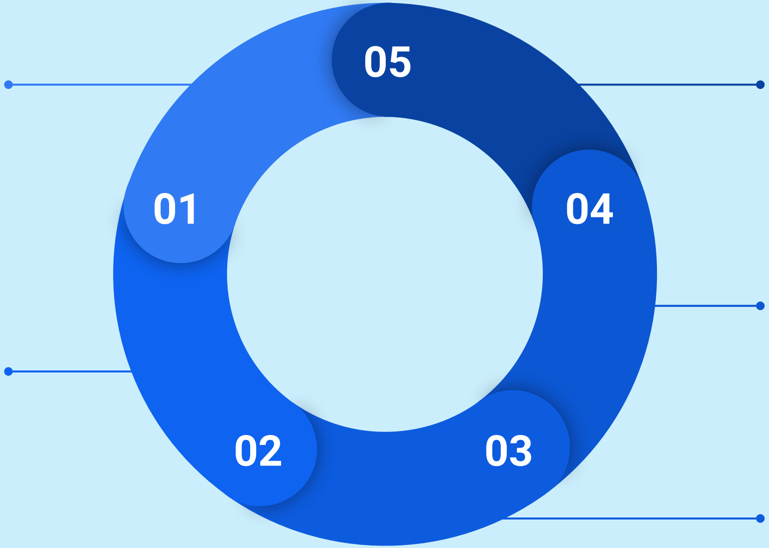
02

Data collection and preparation

Data collection

Credit Card Fraud Detection Dataset
https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud?resource=download

Numerical input variables which are the result of a PCA transformation, due to confidentiality.

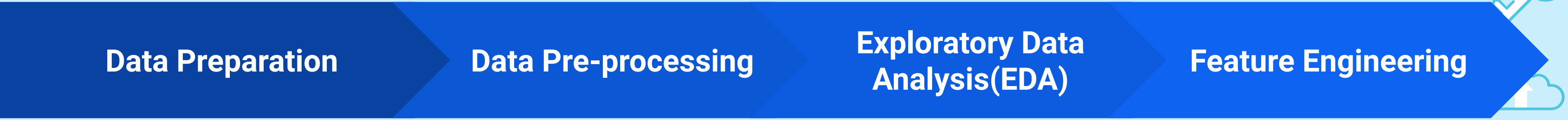


Feature Class is the response variable
value 1 = Fraud
Value 0 = No Fraud.

The only features which have not been transformed with PCA are Time and Amount.

Features V1, V2, ... V28 are the principal components obtained with PCA;

Data preparation



Data Preparation

1. Acquiring Essential Libraries
2. Establishing Connectivity
3. Data Preparation Process

Data Pre-processing

1. Check for missing values
2. Check data unbalance
3. Identify data types of the features

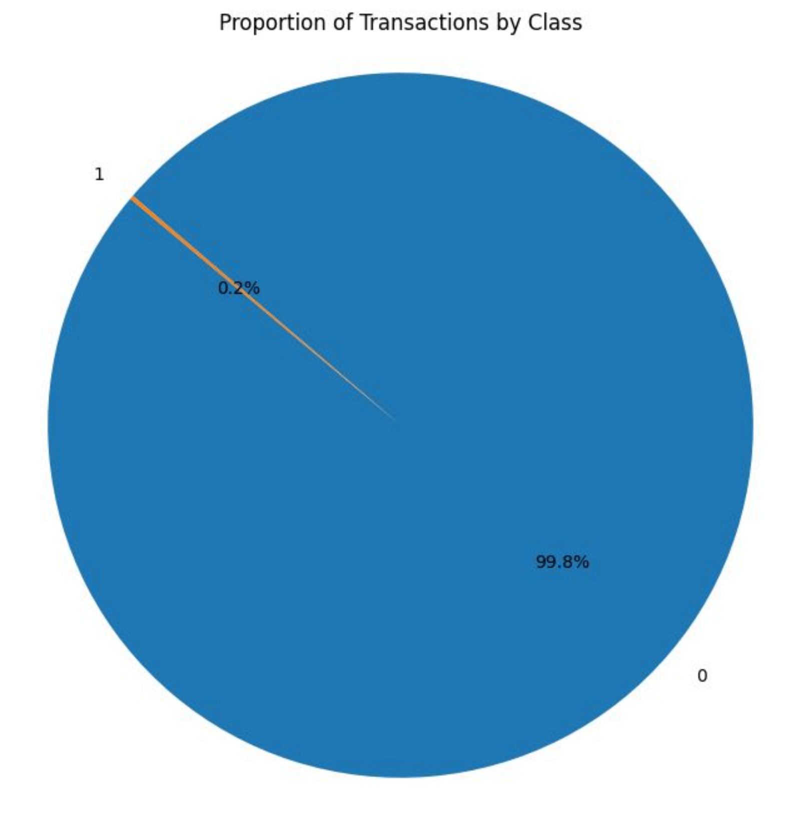
Exploratory Data Analysis(EDA)

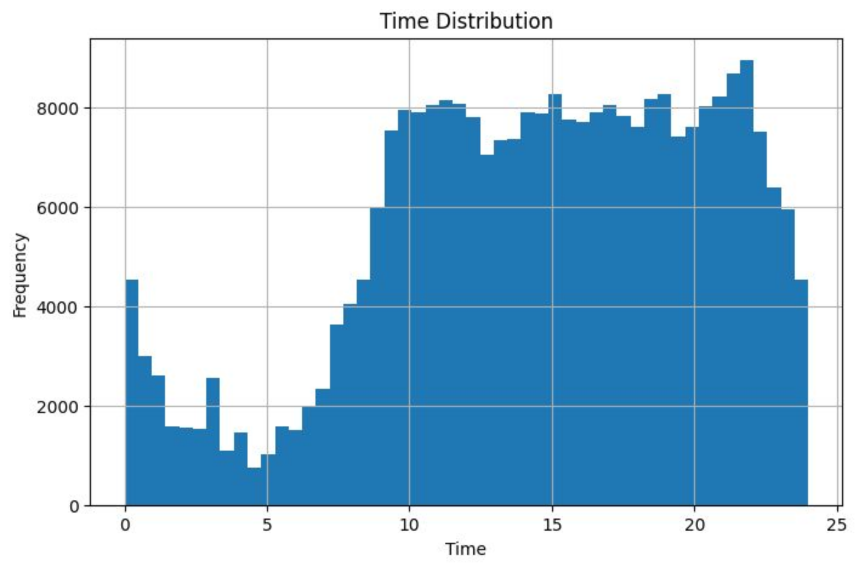
1. Summary Statistics
2. Distribution Analysis
3. Exploratory Visualizations

Feature Engineering

1. Time Feature Day Representation
2. Correlation Matrix

Exploratory Data Analysis (EDA)





03

Methodology

Model Selection

Random Forest Classifier

- Robustness
- Ability to handle complex data
- Handling imbalanced datasets with ensemble learning

XGBoost Classifier

Efficiency and effectiveness in sequential boosting.

Resampling Techniques

Utilize RandomUnderSampler, Balanced Random Forest Classifier, and SMOTE to address class imbalance

Model Training, Validation, and Evaluation

1 **Training**

Train each model on the training set using default hyperparameters or after hyperparameter tuning.

2 **Validation**

Validate model performance on the testing set using the chosen evaluation metric, primarily focusing on recall to minimize false negative

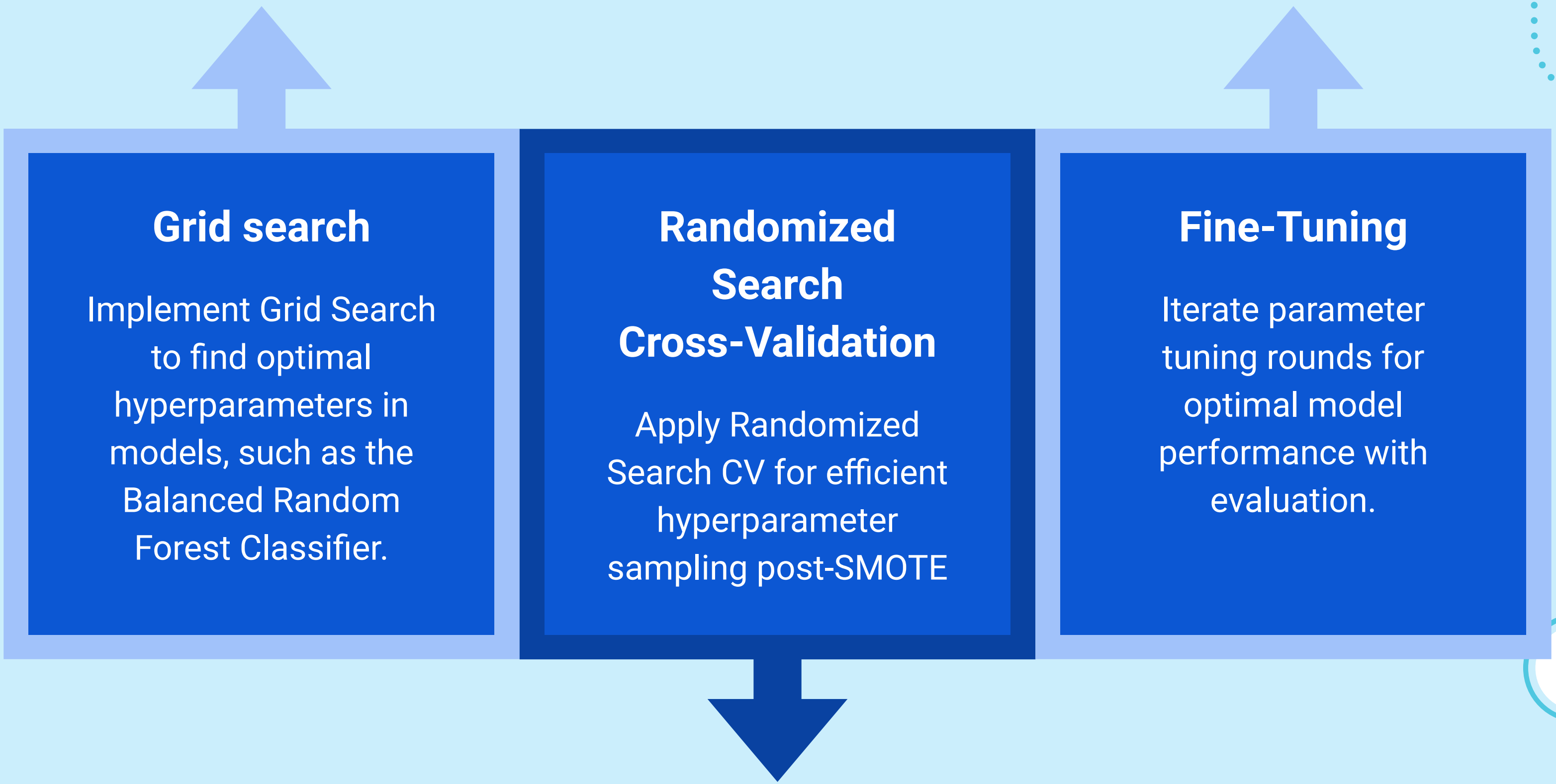
3 **Evaluation Metric**

Use evaluation metrics such as precision, recall, F1-score, to assess model performance comprehensively.

4 **Model Comparison**

Compare the performance of different models to identify the most effective approach for fraud detection.

Hyperparameter tuning and modifications



Grid search

Implement Grid Search to find optimal hyperparameters in models, such as the Balanced Random Forest Classifier.

Randomized Search Cross-Validation

Apply Randomized Search CV for efficient hyperparameter sampling post-SMOTE

Fine-Tuning

Iterate parameter tuning rounds for optimal model performance with evaluation.

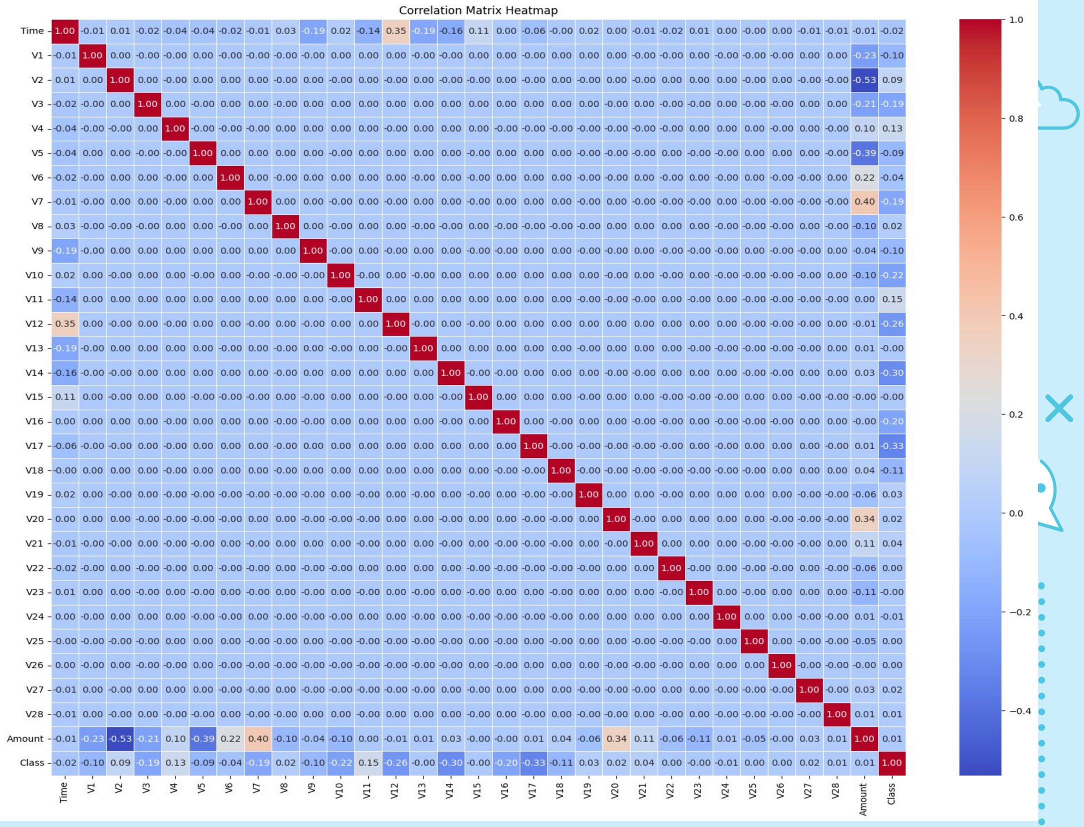
Feature Engineering

Time Feature Modification

Modify 'Time' feature to capture time-of-day patterns in fraud.

Correlation Matrix

Eliminated around 50% of the features with correlation close to 0



Correlation Matrix Heatmap

	Time	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	Amount	Class
Time	1.00	-0.01	0.01	-0.02	-0.04	-0.04	-0.02	-0.01	0.03	-0.19	0.02	-0.14	0.35	-0.19	-0.16	0.11	0.00	-0.06	-0.00	0.02	0.00	-0.01	-0.02	0.01	0.00	-0.00	0.00	-0.01	-0.01	-0.01	-0.02
V1	-0.01	1.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.23	-0.10
V2	0.01	0.00	1.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.53	0.09
V3	-0.02	-0.00	0.00	1.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	-0.21	-0.19
V4	-0.04	-0.00	-0.00	0.00	1.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.10	0.13
V5	-0.04	0.00	0.00	-0.00	-0.00	1.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.39	-0.09
V6	-0.02	-0.00	0.00	0.00	-0.00	0.00	1.00	0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	0.22	-0.04
V7	-0.01	-0.00	0.00	0.00	-0.00	0.00	0.00	1.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.40	-0.19
V8	0.03	-0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	1.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.10	0.02
V9	-0.19	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.04	-0.10
V10	0.02	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	1.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.10	-0.22
V11	-0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.15
V12	0.35	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	1.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.01	-0.26
V13	-0.19	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	1.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.01	-0.00
V14	-0.16	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	1.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.03	-0.30
V15	0.11	0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	1.00	0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
V16	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.00	1.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.20
V17	-0.06	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.01	-0.33
V18	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	1.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.04	-0.11
V19	0.02	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	1.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.06	0.03
V20	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	1.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.34	0.02
V21	-0.01	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	1.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.11	0.04
V22	-0.02	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	1.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.06	0.00
V23	0.01	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	1.00	0.00	-0.00	0.00	0.00	0.00	-0.11	-0.00
V24	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	-0.00	-0.00	0.01	-0.01
V25	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	1.00	0.00	-0.00	-0.00	-0.05	0.00
V26	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	1.00	-0.00	-0.00	-0.00	0.00
V27	-0.01	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	1.00	-0.00	0.03	0.02
V28	-0.01	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	1.00	0.01	0.01
Amount	-0.01	-0.23	-0.53	-0.21	0.10	-0.39	0.22	0.40	-0.10	-0.04	-0.10	0.00	-0.01	0.01	0.03	-0.00	-0.00	0.01	0.04	-0.06	0.34	0.11	-0.06	-0.11	0.01	-0.05	-0.00	0.03	0.01	1.00	0.01
Class	-0.02	-0.10	0.09	-0.19	0.13	-0.09	-0.04	-0.19	0.02	-0.10	-0.22	0.15	-0.26	-0.00	-0.30	-0.00	-0.20	-0.33	-0.11	0.03	0.02	0.04	0.00	-0.00	-0.01	0.00	0.00	0.02	0.01	0.01	1.00

04

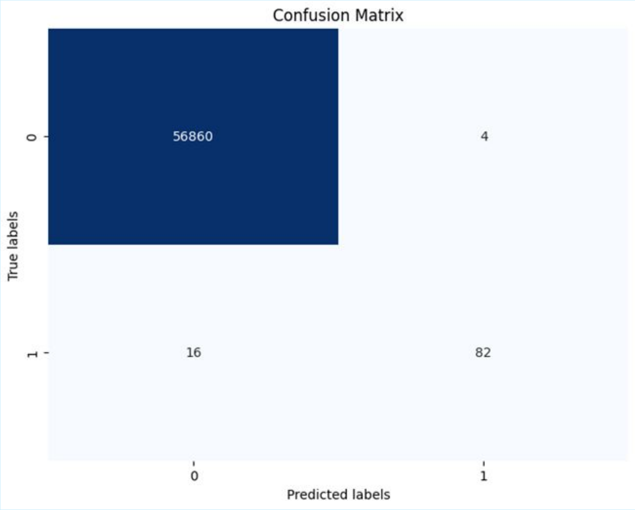
Analysis and Results

Results of Models

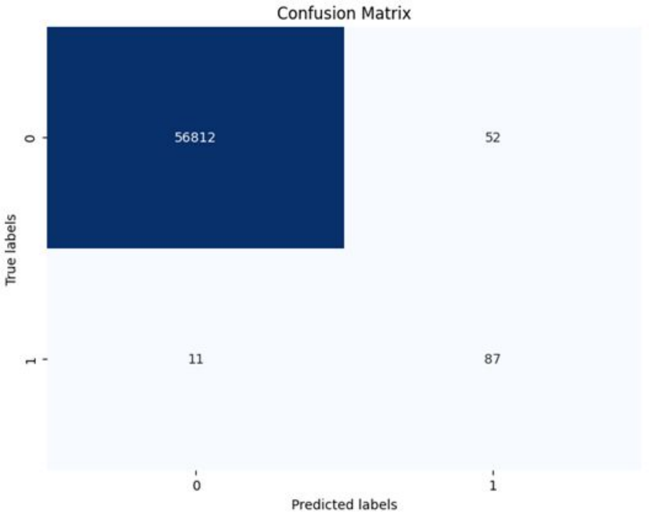
Models		Precision	Recall
1	Random Forest Classifier	0.953488	0.836735
2	XGB classifier	0.864583	0.846939
3	Random Forest Classifier after RandomUnderSampler	0.039190	0.908163
4	Balanced Random Forest Classifier	0.048276	0.928571
5	Balanced Random Forest Classifier After SMOTE	0.625899	0.887755

Results of Models

Random Forest Classifier



Balanced Random Forest + SMOTE



05

Discussion

Conclusions

- Random Forest and XGB Classifiers: High precision; recall varied for fraud.
- Balanced RF and SMOTE models: Improved recall, lowered precision.
- Trade-off: Recall vs. precision; higher recall may affect precision.
- RandomUnderSampler and SMOTE enhanced recall, penalizing precision.
- Future: Explore advanced ensemble models or deep learning algorithms for better balance. Also going deep into the current models checking documentation to increase as much as possible the performance.

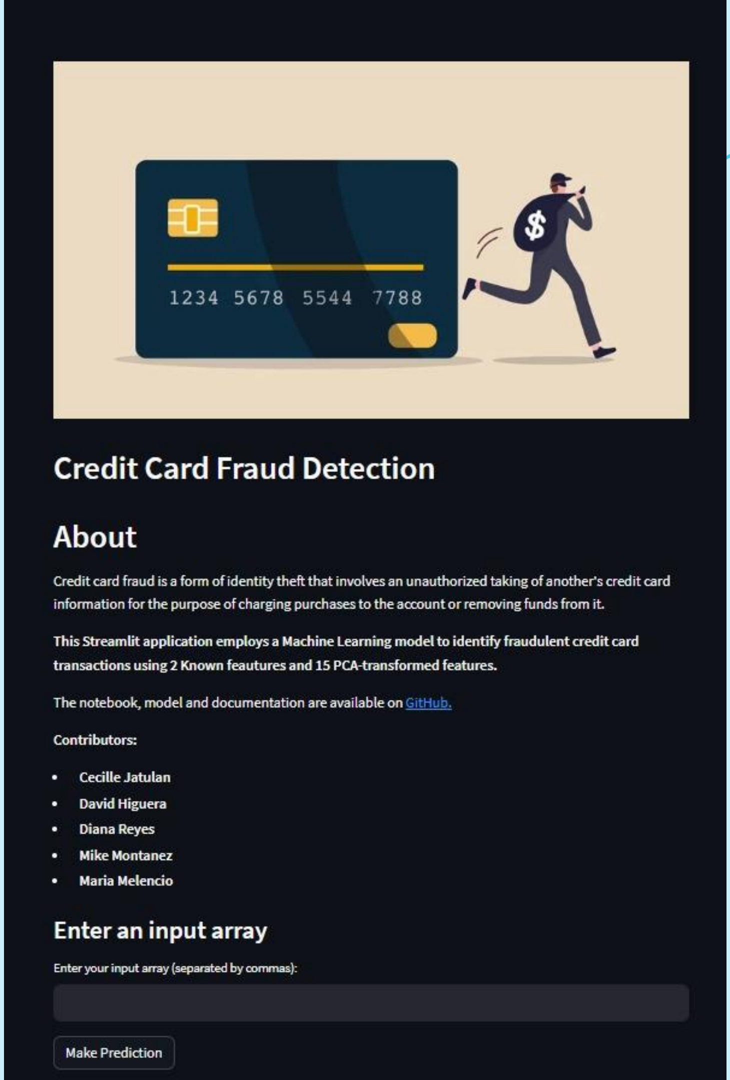
References

- Mrozek, P., Panneerselvam, J., & Bagdasar, O. (2020). Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets. 2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC). https://doi.org/10.1109/ucc48980.2020.00067.

- Windari, L. (2024, April 16). Credit Card Fraud Detection. Medium. https://medium.com/analytics-vidhya/credit-card-fraud-detection-how-to-handle-imbalanced-dataset-1f18b6f881

- OpenAI. (2023). ChatGPT [Computer software]. https://www.openai.com/chatgpt

UI DEMO



Credit Card Fraud Detection

About

Credit card fraud is a form of identity theft that involves an unauthorized taking of another's credit card information for the purpose of charging purchases to the account or removing funds from it.

This Streamlit application employs a Machine Learning model to identify fraudulent credit card transactions using 2 Known feautures and 15 PCA-transformed features.

The notebook, model and documentation are available on GitHub.

Contributors:

- Cecille Jatulan
- David Higuera
- Diana Reyes
- Mike Montanez
- Maria Melencio

Enter an input array

Enter your input array (separated by commas):

Make Prediction

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**