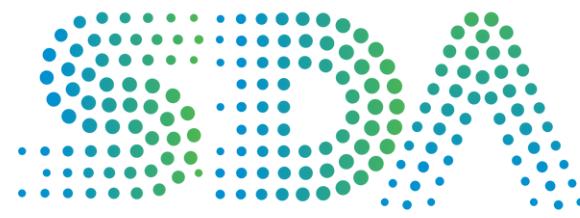


Fraud Detection Project

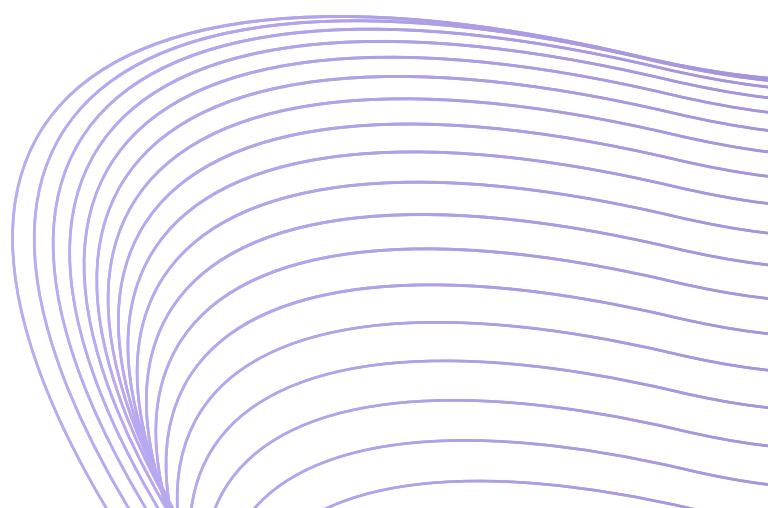
By: G-1 Team

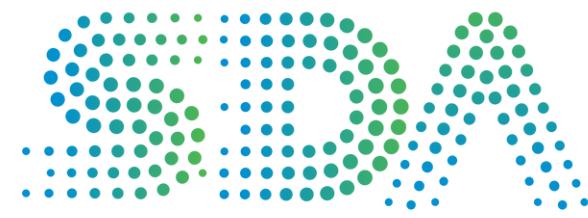




الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

 CLARUSWAY
WAY TO REINVENT YOURSELF

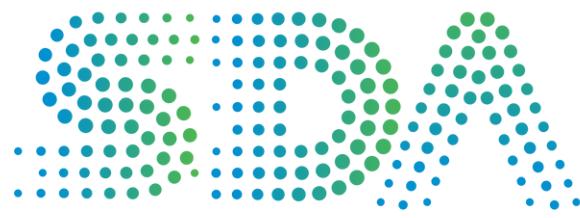




G-1 Team Members

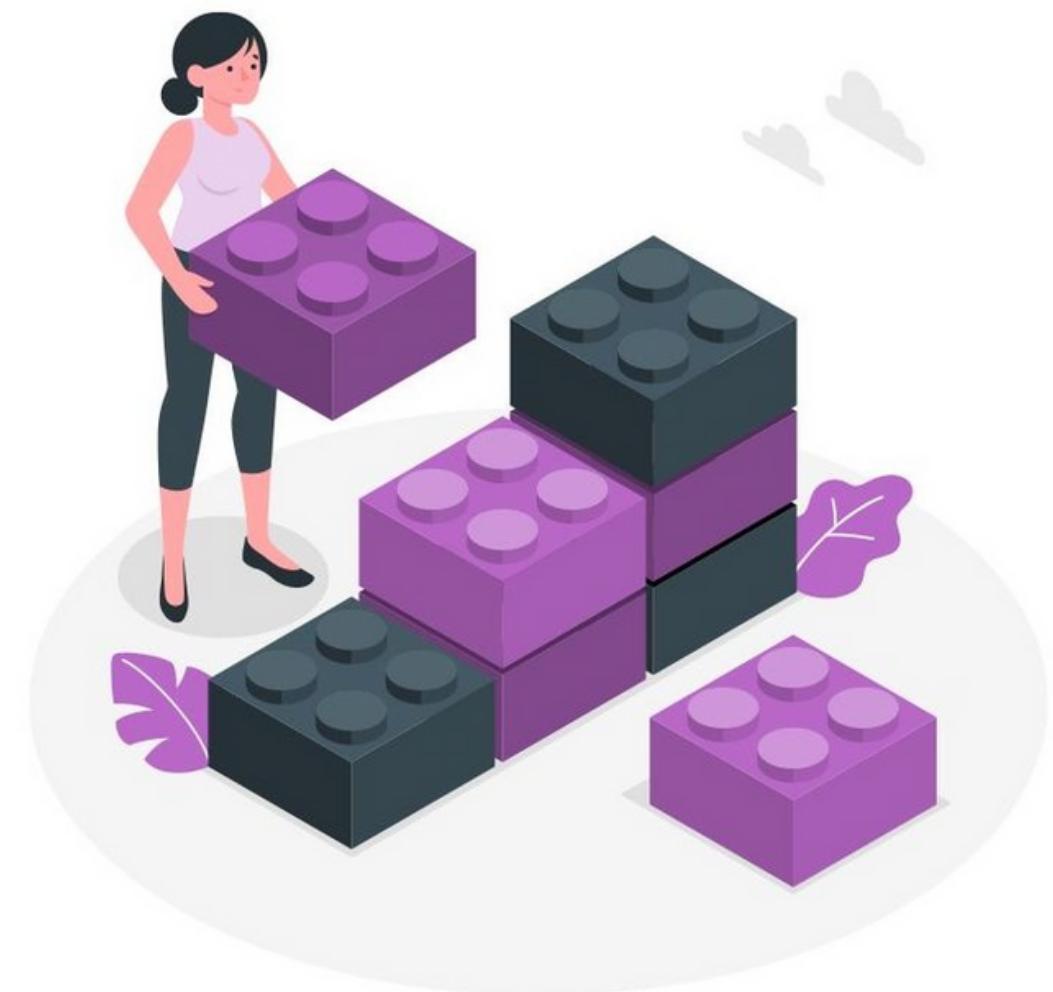
- Ammar Alhawashem
- Maram Alzahrani
- Noof Alsafi
- Taif Alzahr
- Sara
- Noura Alajmi
- Hibah Sind
- Shahad Ali
- Reema Almeshal
- Raya

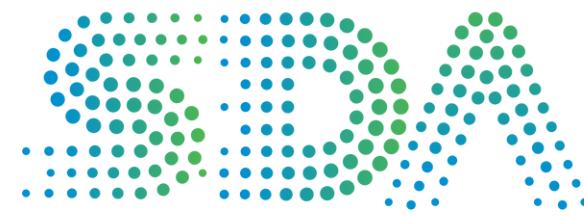




Outlines:

- Project Overview
- Data Analysis and Visualization
- Predictive Model
 - Logistic Regression model
 - RF model
 - XG Boost model
 - Neural Network
- Model Deployment
- Conclusion

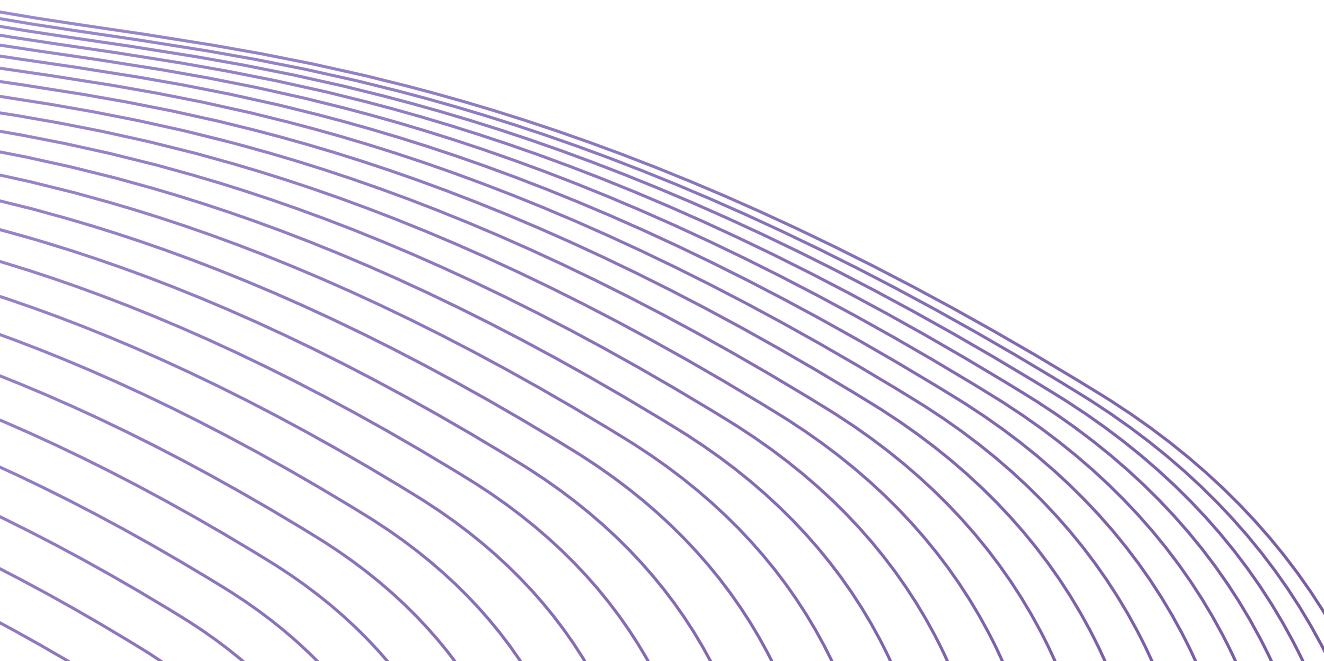


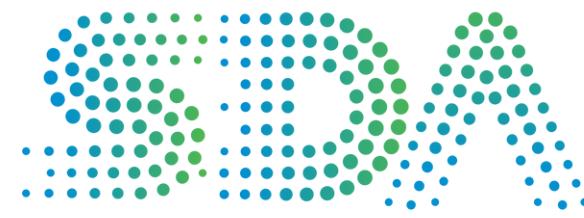


الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY



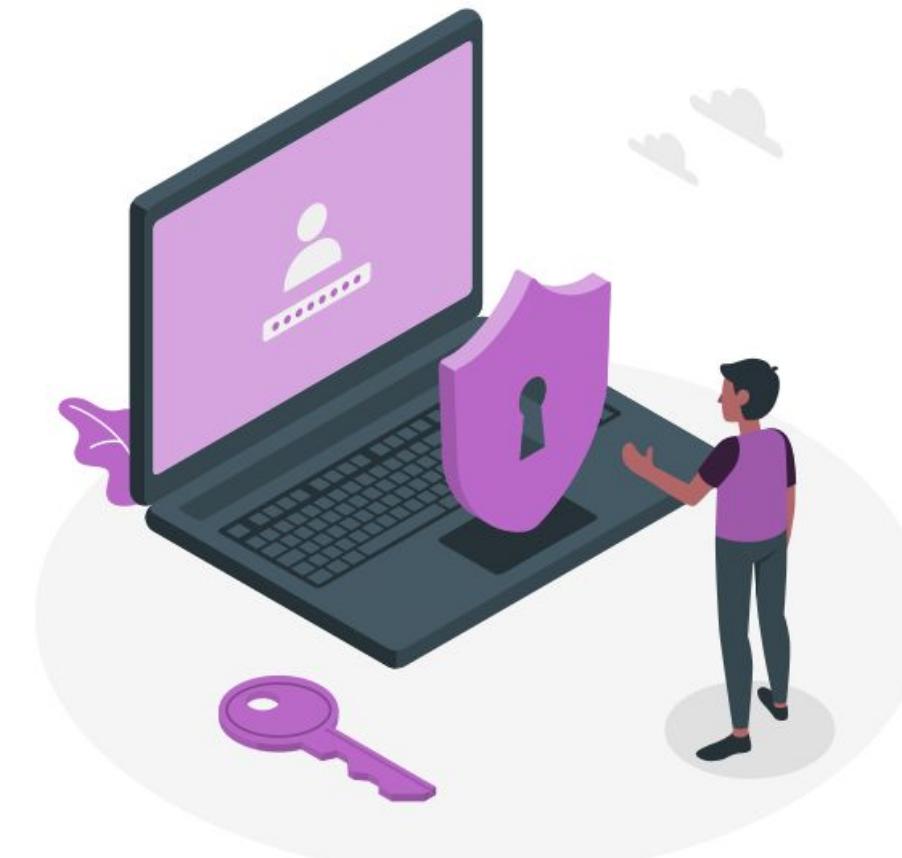
Project Overview

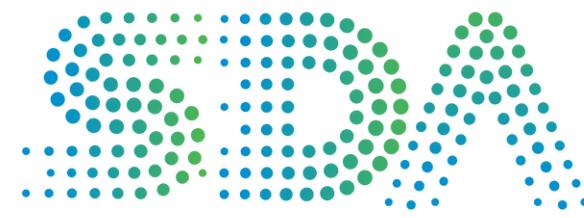




Problem Statement

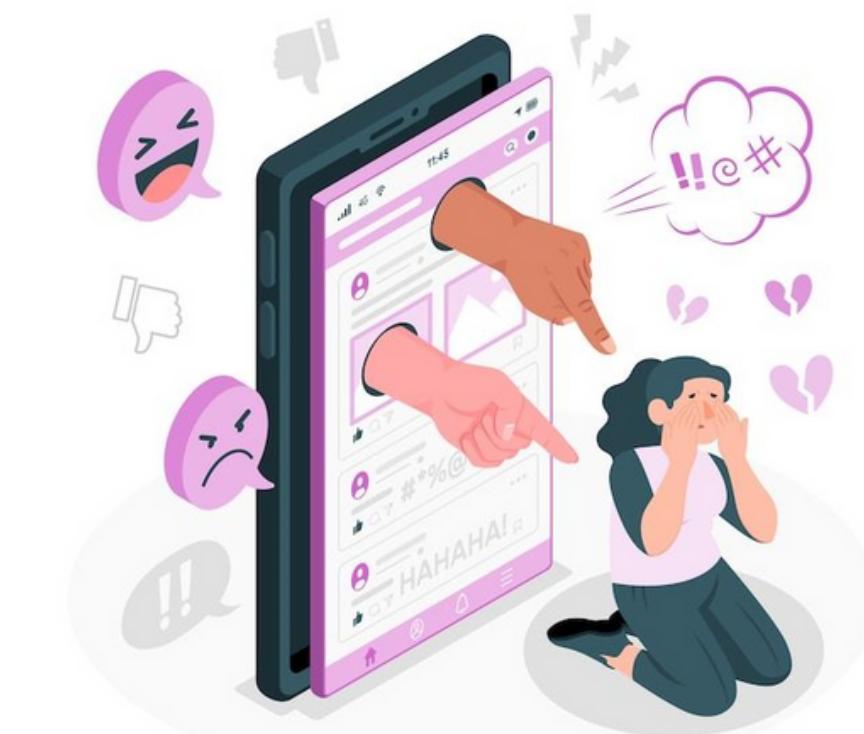
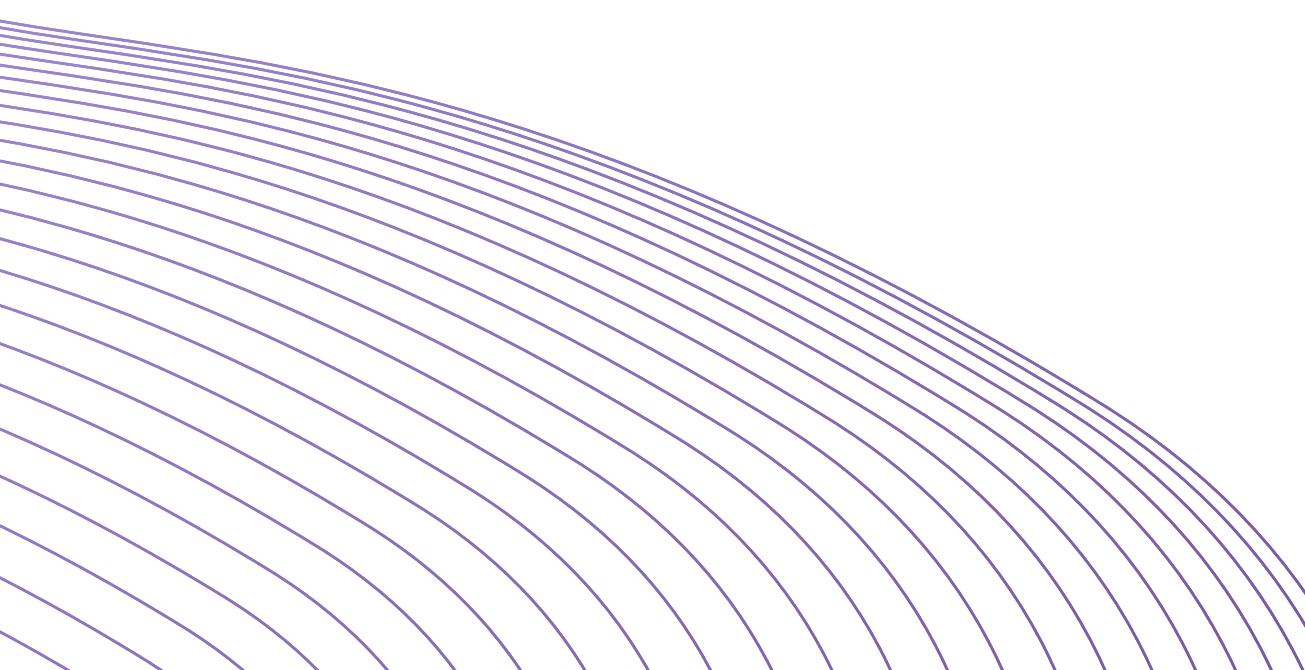
Detect fraudulent credit card transactions using a dataset containing information about transactions made by European cardholders over two days.

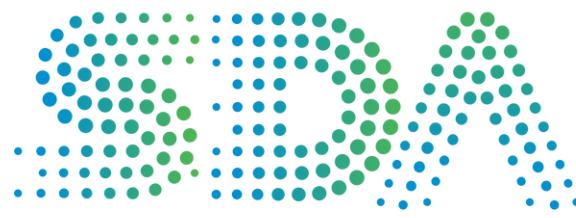




Goal

The primary goal of this project is to design and implement a robust machine learning model to reduce false alarms while effectively identifying instances of fraud. The ultimate objective is to provide financial institutions with a dependable tool to enhance fraud prevention strategies and safeguard customers' financial security.



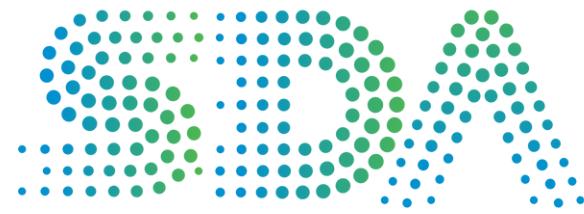


Dataset

The dataset contains credit card transaction records from September 2013, featuring 284,807 transactions. Within this dataset, only 492 instances (0.172%) represent fraudulent activities, making it highly imbalanced.

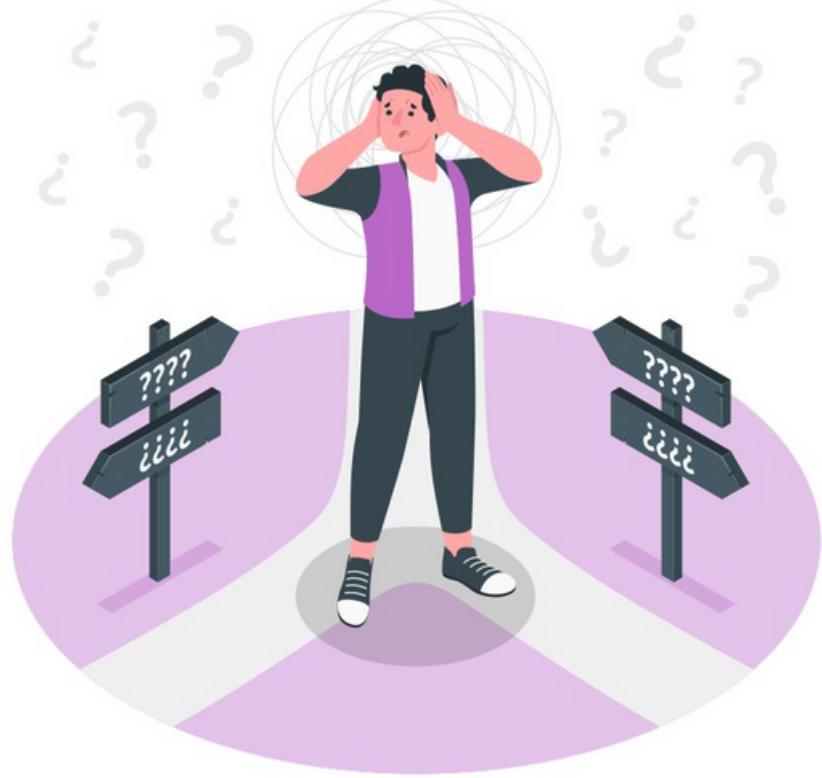
	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458
...
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.918215	7.305334	1.914428	...	0.213454	0.111864	1.014480
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.024330	0.294869	0.584800	...	0.214205	0.924384	0.012463
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.296827	0.708417	0.432454	...	0.232045	0.578229	-0.037501
284805	172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.686180	0.679145	0.392087	...	0.265245	0.800049	-0.163298
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271	-0.012546	-0.649617	1.577006	-0.414650	0.486180	...	0.261057	0.643078	0.376777



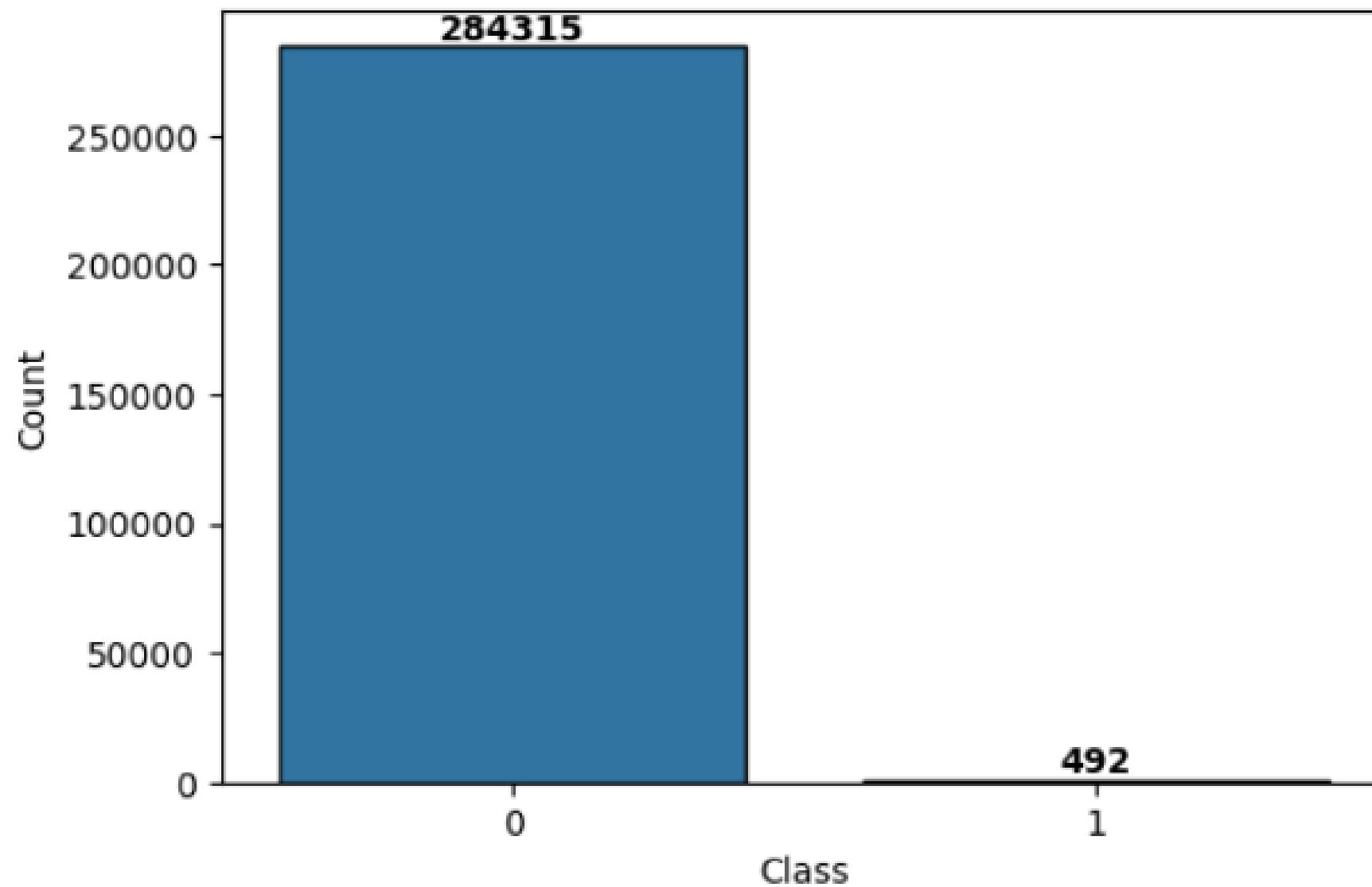


Data Analysis and Visualization

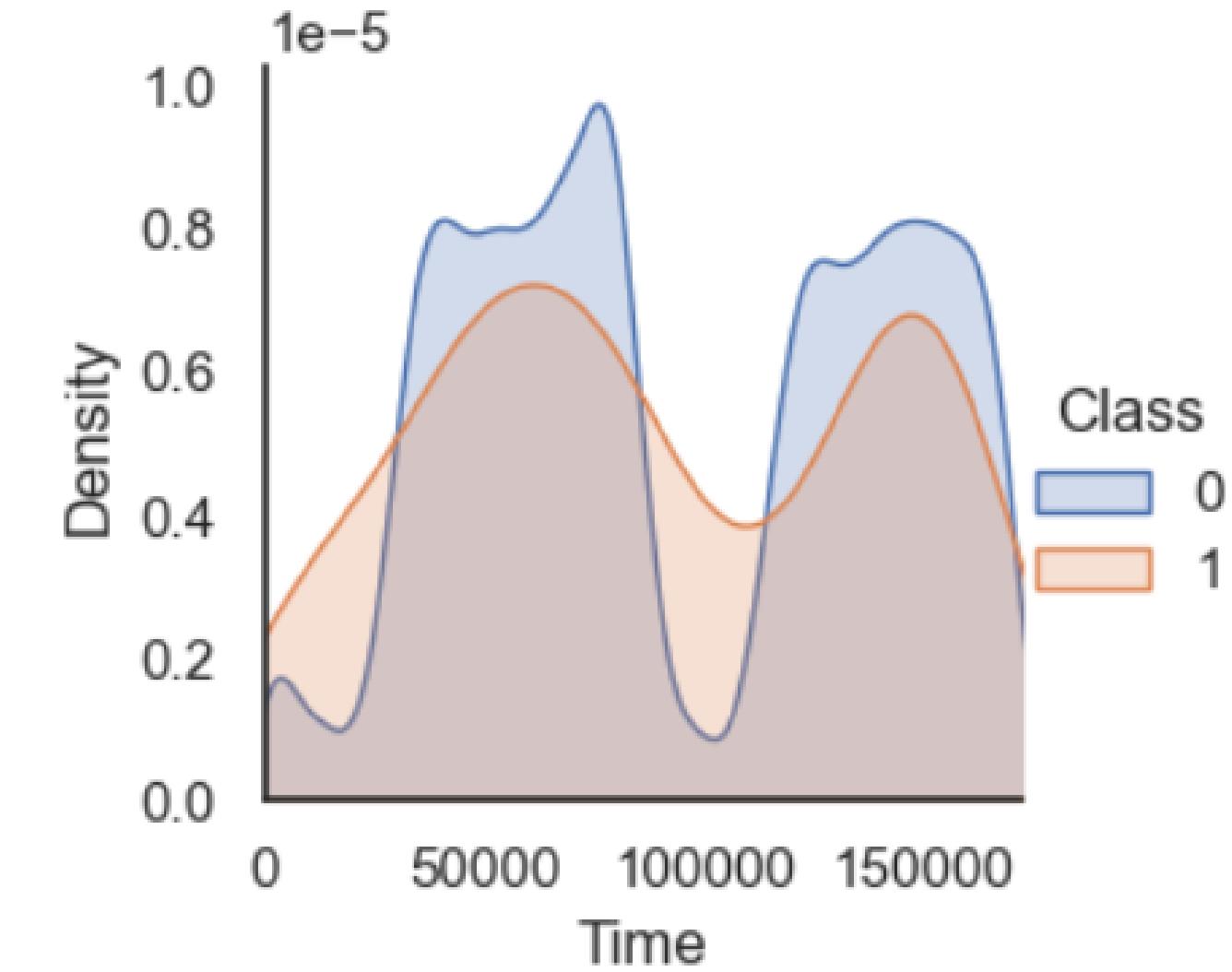
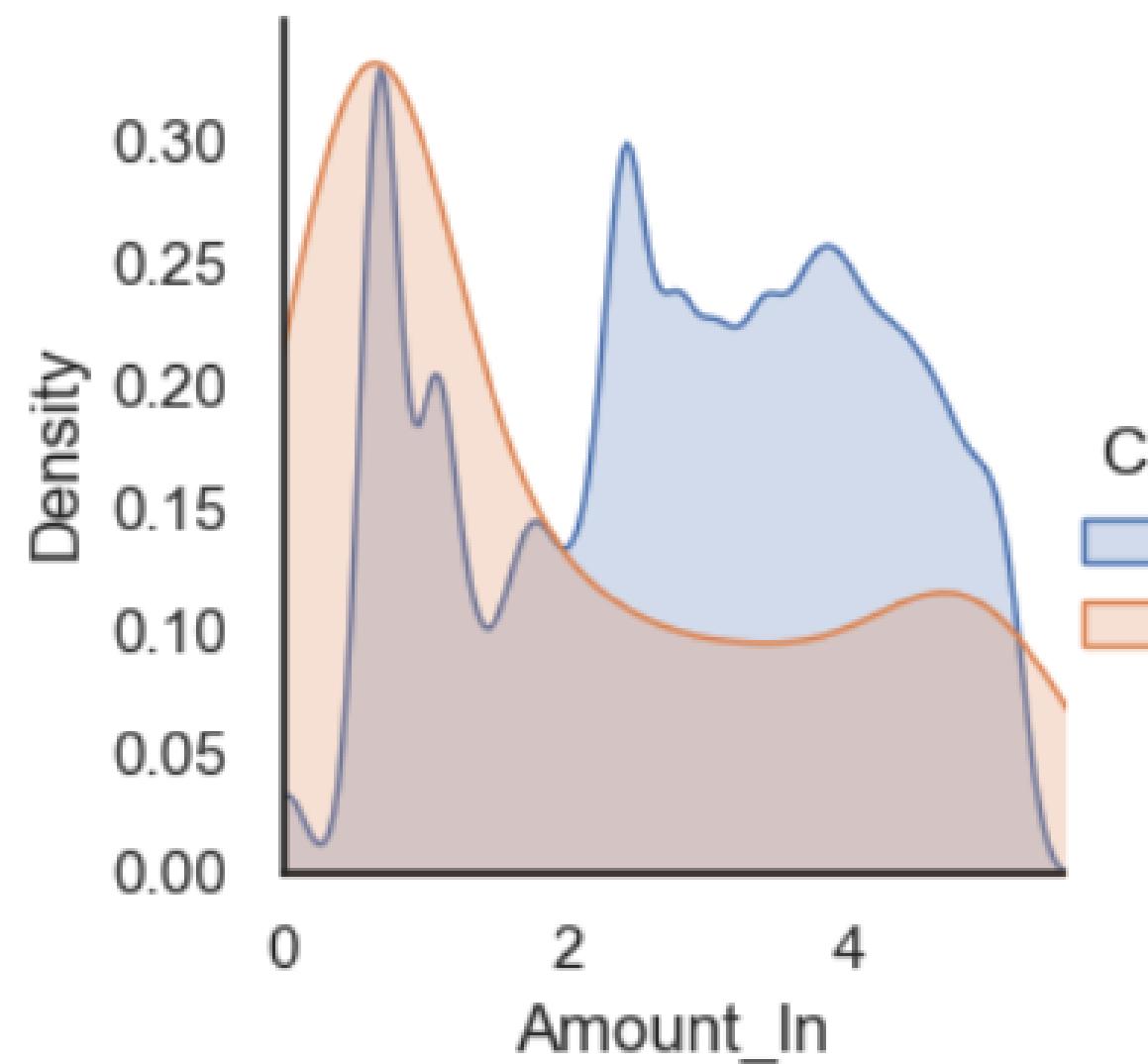


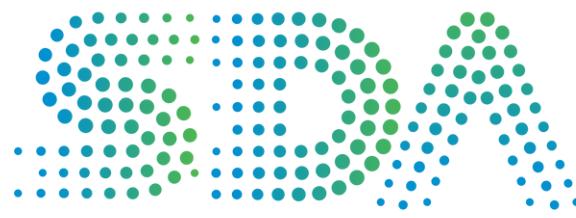


Fraud distribution

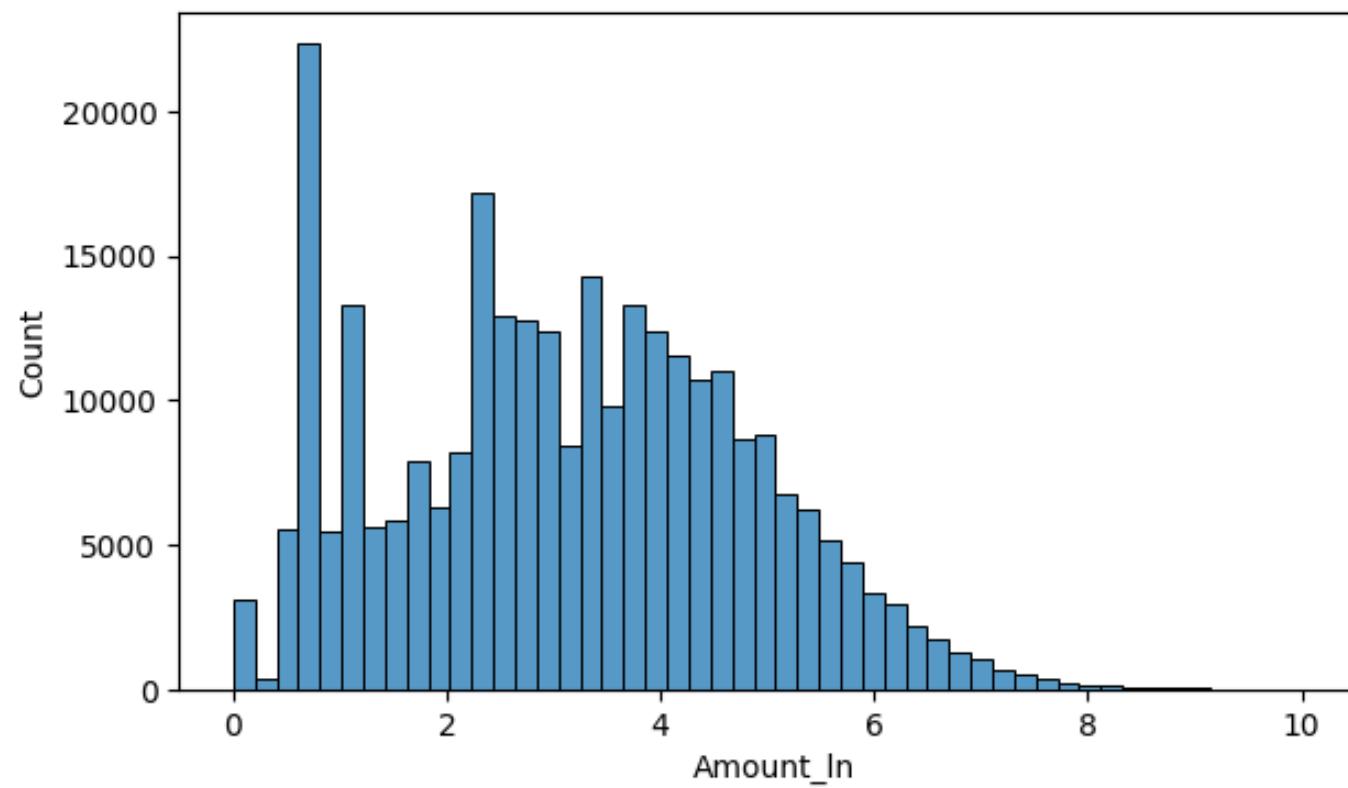
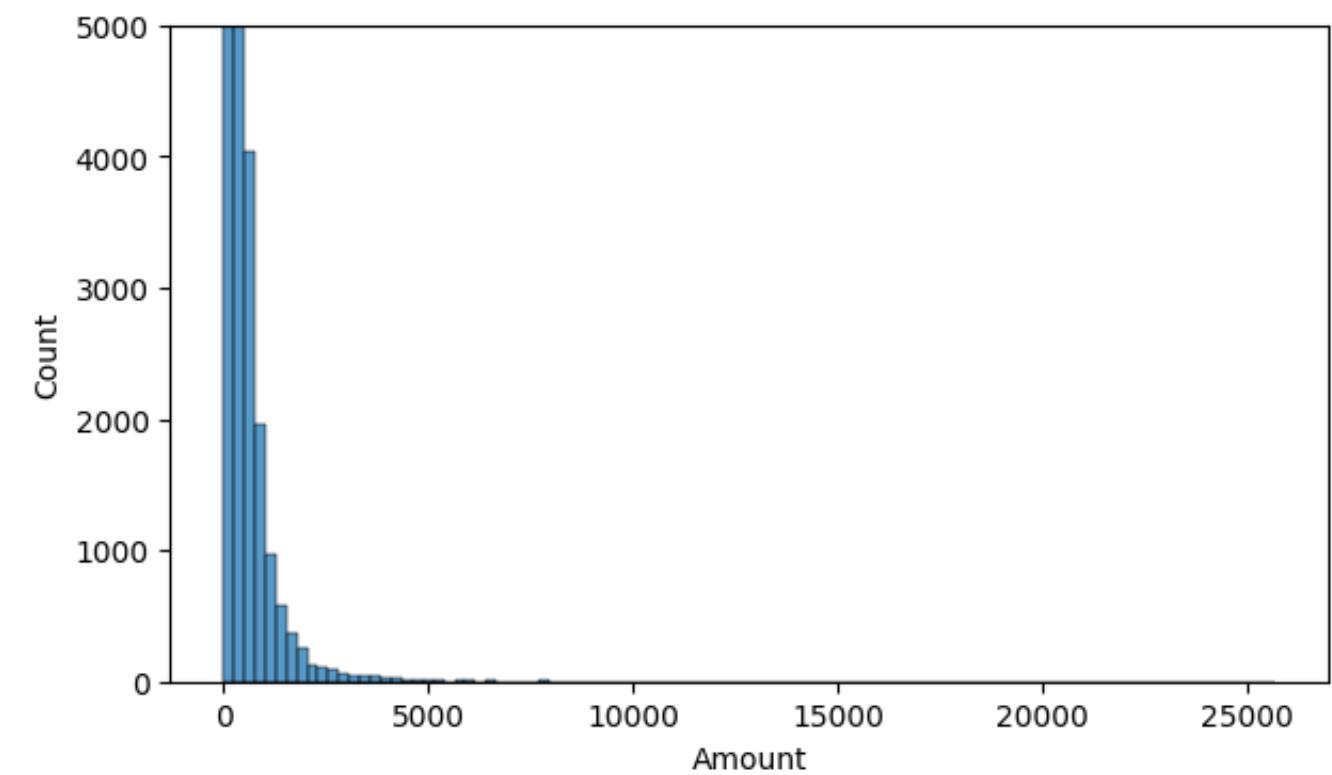


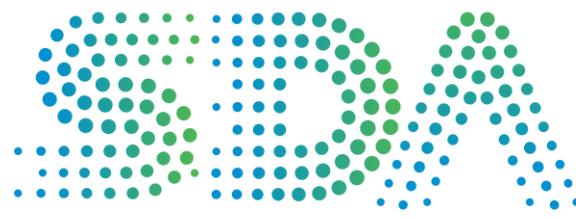
Kernel Density Estimation of Transformed Features (Amount and Time)



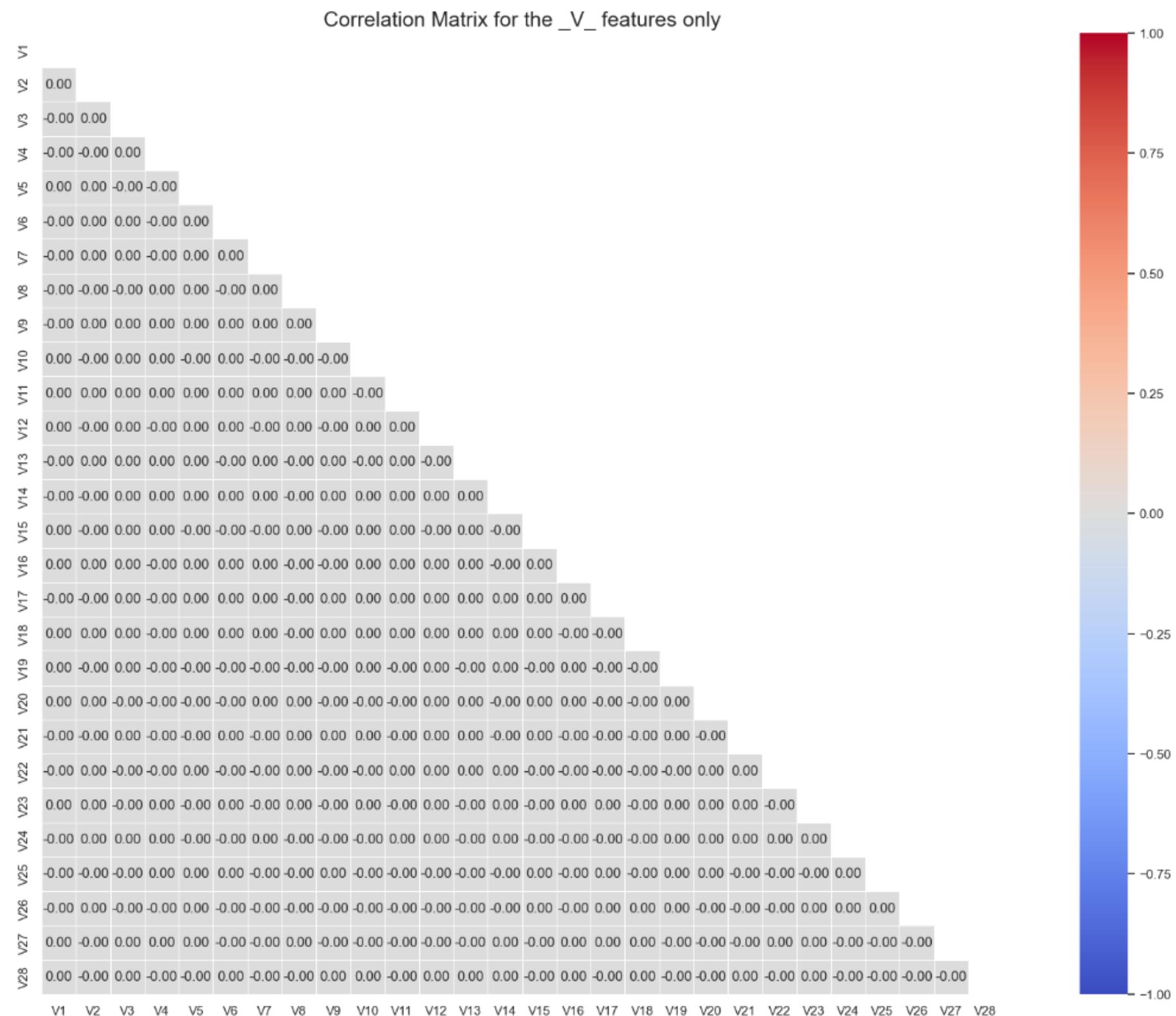


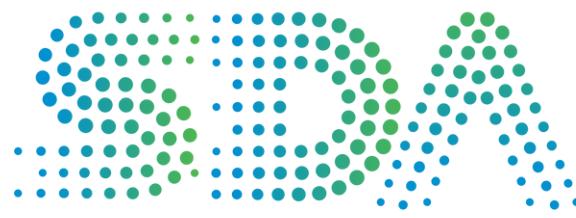
Amount before and after transformation



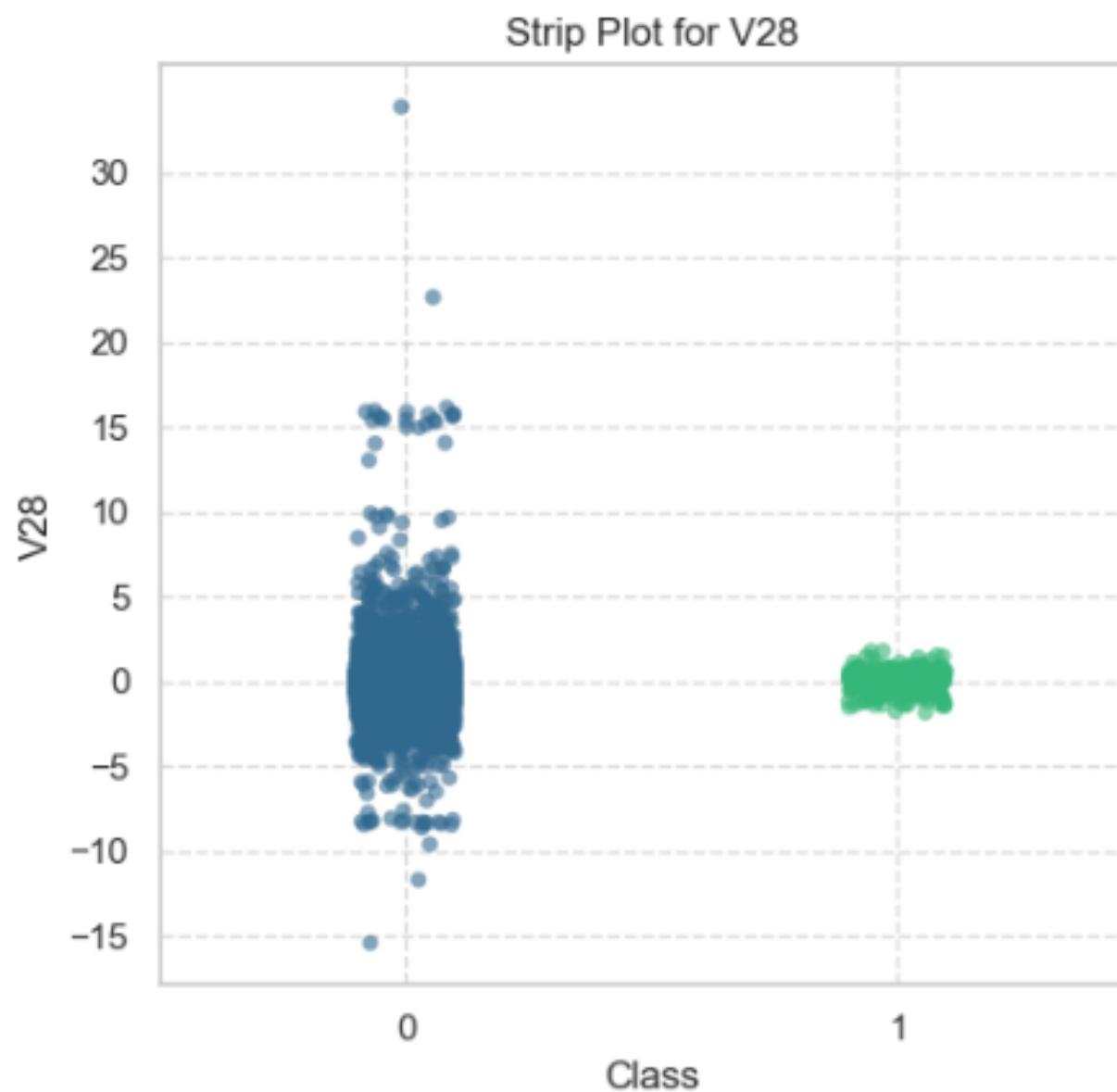
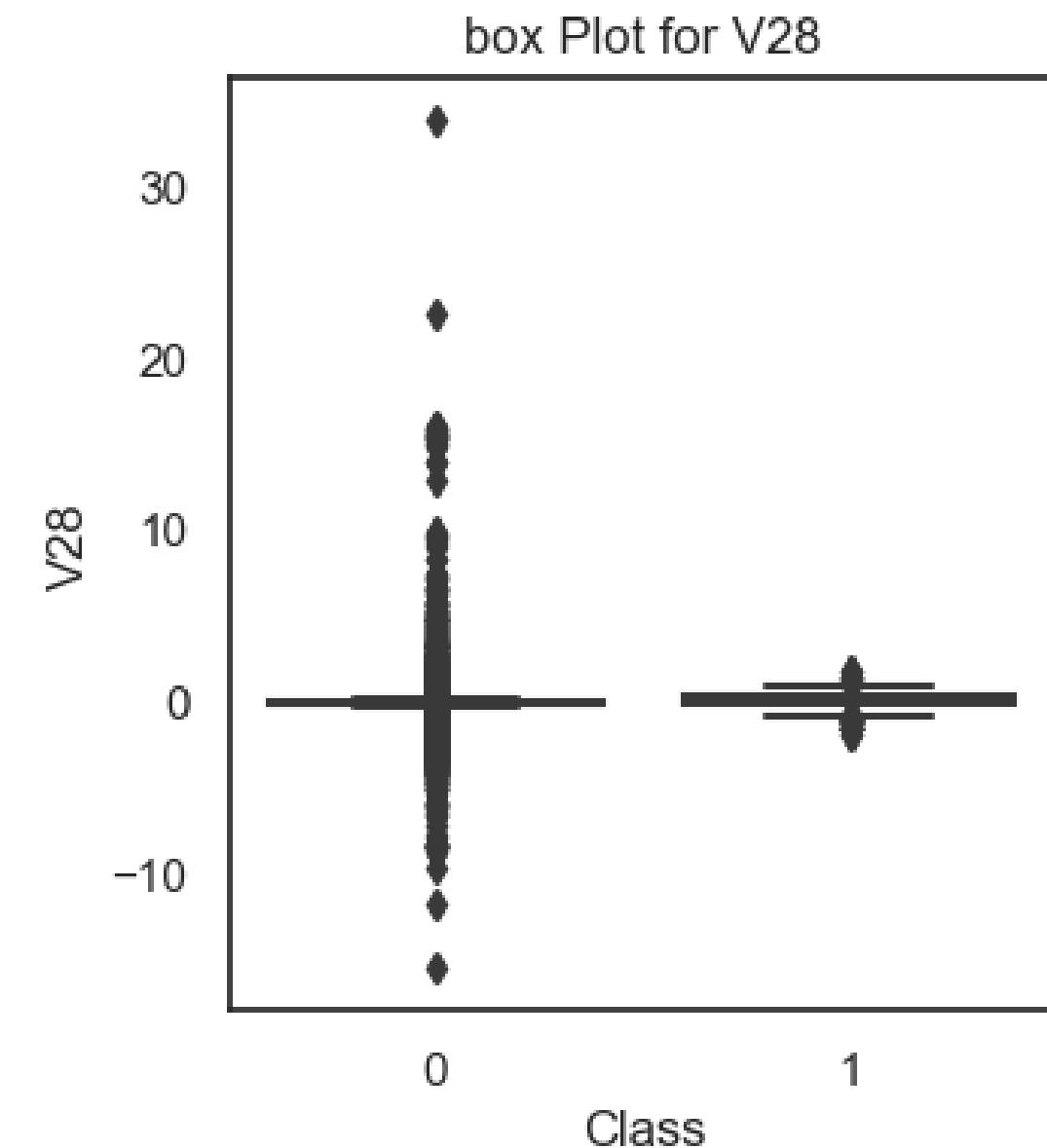


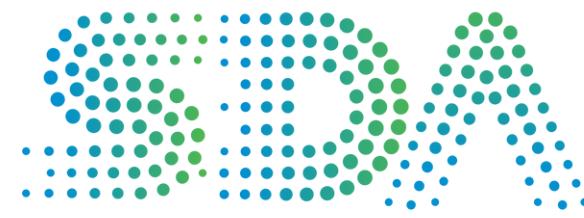
Features Correlation





Are they outliers?

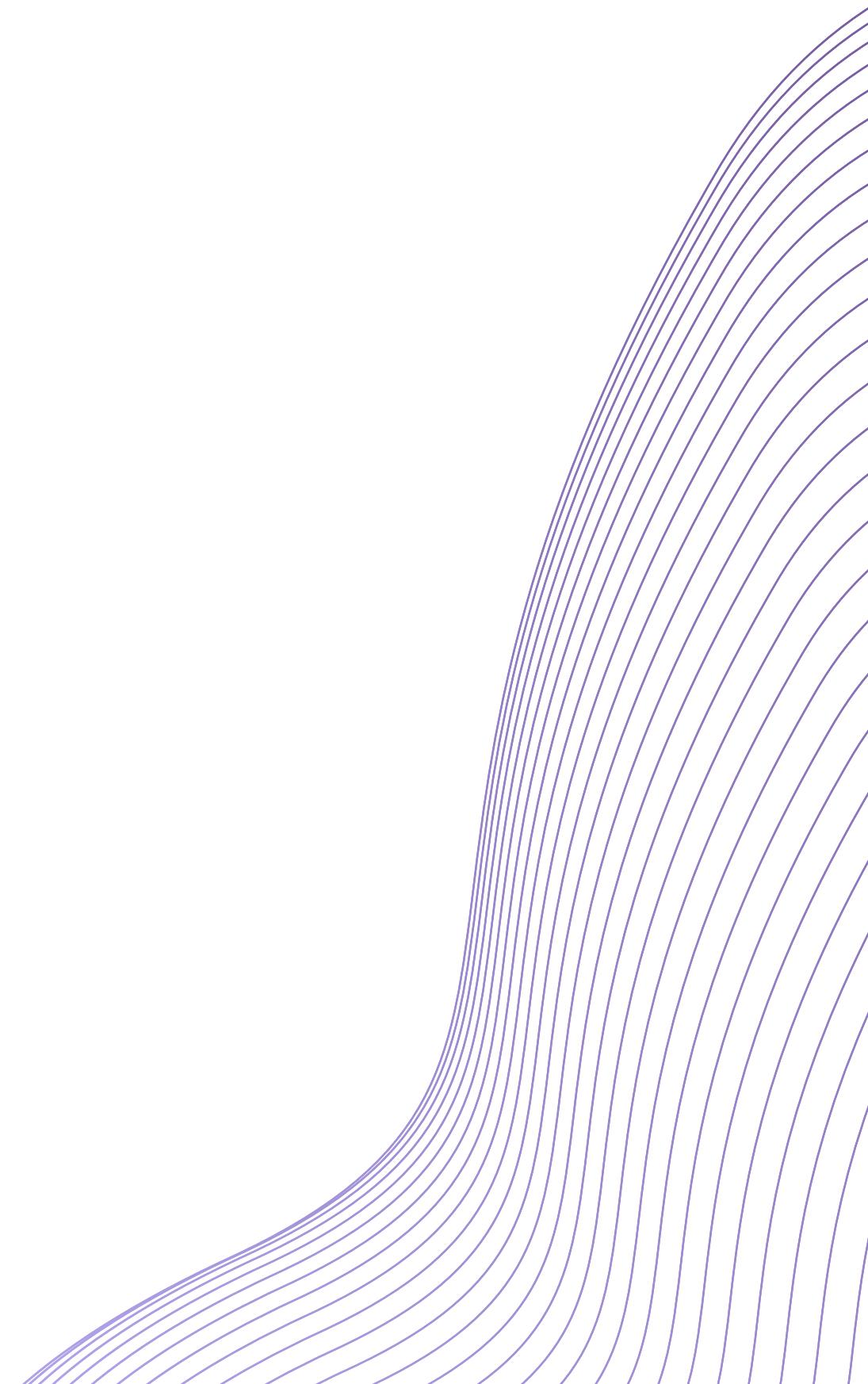


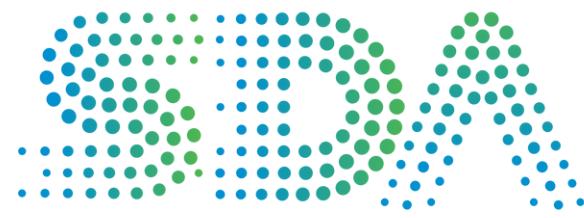


الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

 CLARUSWAY
WAY TO REINVENT YOURSELF

Predictive Model



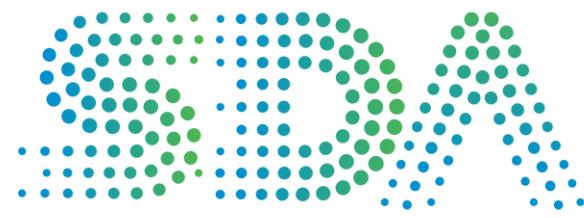


Over/Under Sampling

```
# For resampling

under_class0 = int(y_train.value_counts()[0] - (0.2 * y.value_counts()[0]))
over_class1 = int(y_train.value_counts()[1] + (0.2 * y.value_counts()[1]))
```

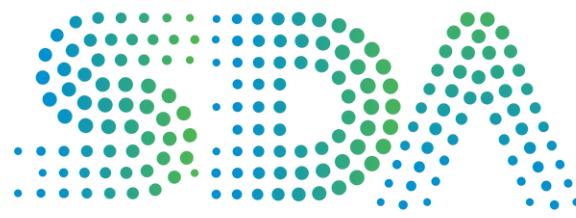
```
Original X_train distribution: Class
 0    220079
 1     378
Name: count, dtype: int64
-----
Resampled X_train distribution: Class
 0    165059
 1     472
Name: count, dtype: int64
```



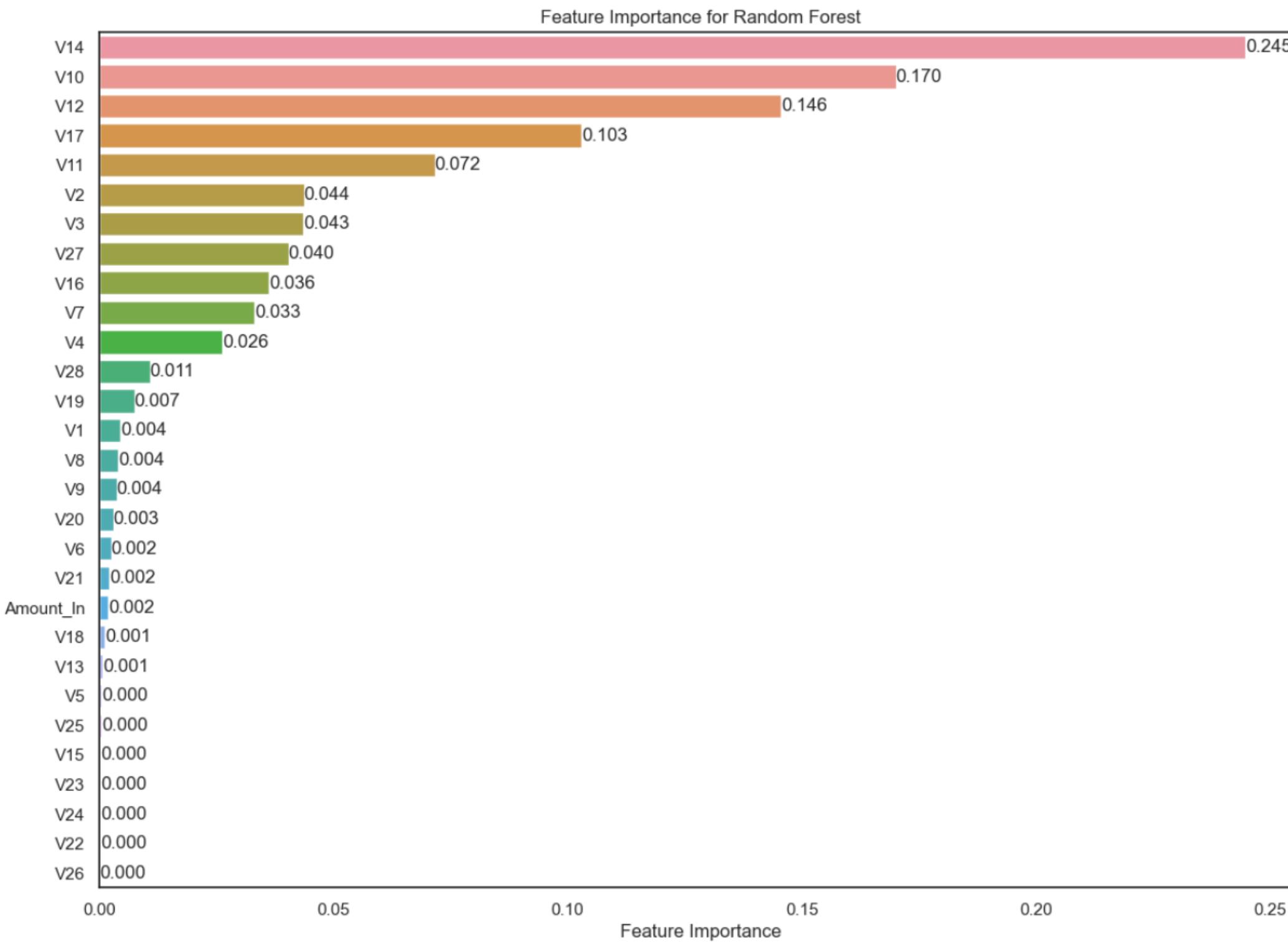
Predictive Model Building

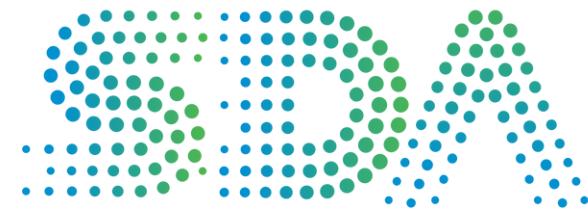
Classification algorithms we applied:

1. Logistic Regression
2. Random Forest Classifier
3. XGBoost Classifier
4. Neural Network



Features Importance for the Best Model



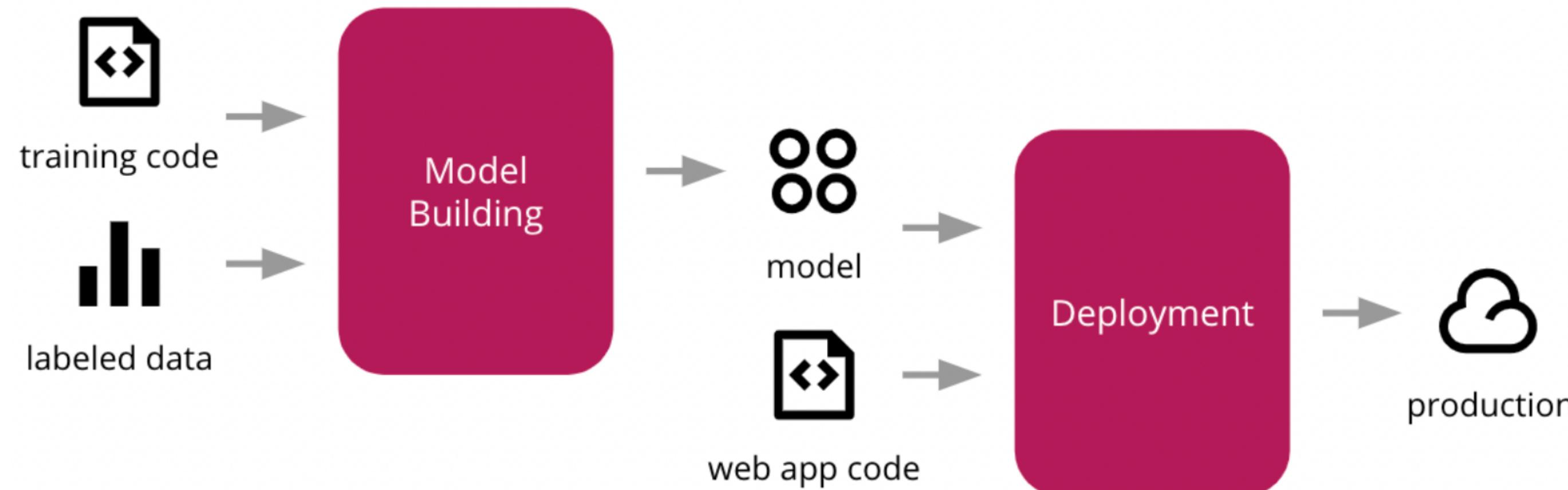


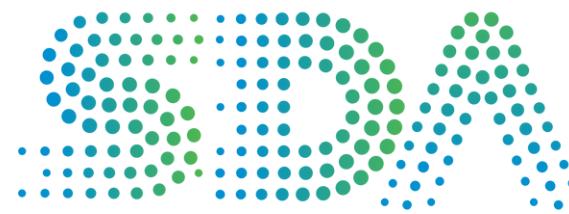
Model Deployment



What is Model Deployment?

Deployment means putting a machine learning model into action in a real business setting to make decisions based on data. It's one of the last steps in the machine learning process. In this project we deployed our ML model with Streamlit on AWS.





Fraud Detection App

my app
prediction page

Main page 

Welcome to the Fraud Detection App 

Fraud Detection App



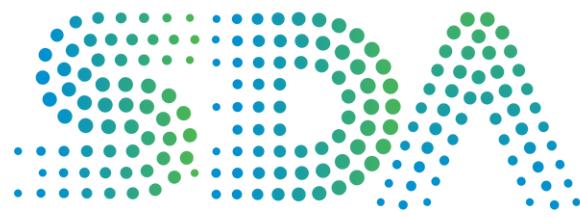
About the app

Fraud Detection App is a cutting-edge fraud detection application that empowers businesses to proactively safeguard their operations against the ever-evolving threat of fraudulent activities. With its powerful artificial intelligence and advanced machine learning algorithms, Fraud Detection App offers unparalleled protection and peace of mind.

Stay one step ahead of fraudsters

Powered by state-of-the-art technology, Fraud Detection App is designed to detect and prevent fraudulent activities in real-time. Its robust algorithms analyze vast amounts of data, quickly identifying patterns, anomalies, and suspicious behaviors that may indicate fraud attempts. By staying one step ahead, Fraud Detection App ensures that you can protect your business and customers from financial loss and reputational damage.

Features



Fraud Detection App

my app

prediction page

Prediction page 🎉

Fraud Detection

Select the value of V2

-5.00

Select the value of V3

-10.00

Select the value of V4

-5.00

Select the value of V7

-10.00

Select the value of V10

-10.00

Select the value of V11

-5.00

Select the value of V12

-11.00

Select the value of V14

-15.00

Select the value of V16

-10.00

Select the value of V17

-15.00

Select the value of V27

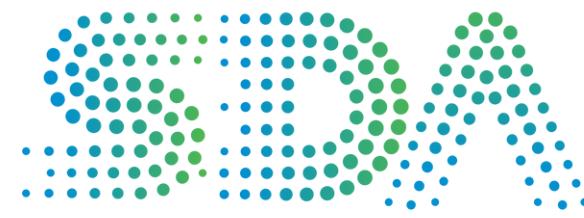
-1.00

Customer Information

	-----Customer Information-----
V2	-5.00
V3	-10.00
V4	-5.00
V7	-10.00
V10	-10.00
V11	-5.00
V12	-11.00
V14	-15.00

Predicting with ML 🎉





الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

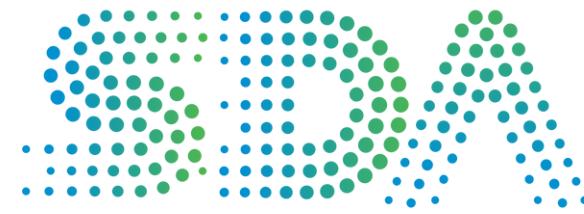
 CLARUSWAY
WAY TO REINVENT YOURSELF

Visit Our App



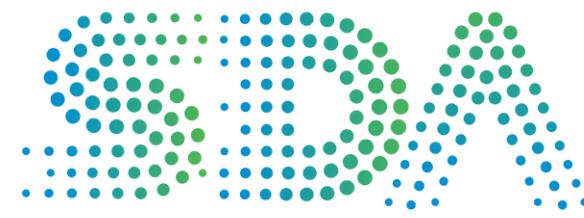
URL





Conclusion





In this project, we deploy a Streamlit app that detects transaction Fraud based on user inputs. The app loads a trained Random Forest classifier model and applies it to the user input data to detect whether or not the transactions is likely to Fraud.





Thanks for Listening