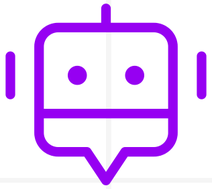
NEW BANK

Bana Marco Amir Jean Philippe

Detection of

Payment Fraud

A Machine Learning Model



PROJECT SOPHIA

Identify and prevent payment fraud.

Model

Introduction

*As **New Bank**, we aim to provide secure payment systems to protect our clients from fraud that can lead to significant financial loss and erode customer trust. Online payment fraud is a significant issue in the banking industry, causing financial losses and reduced customer trust.*

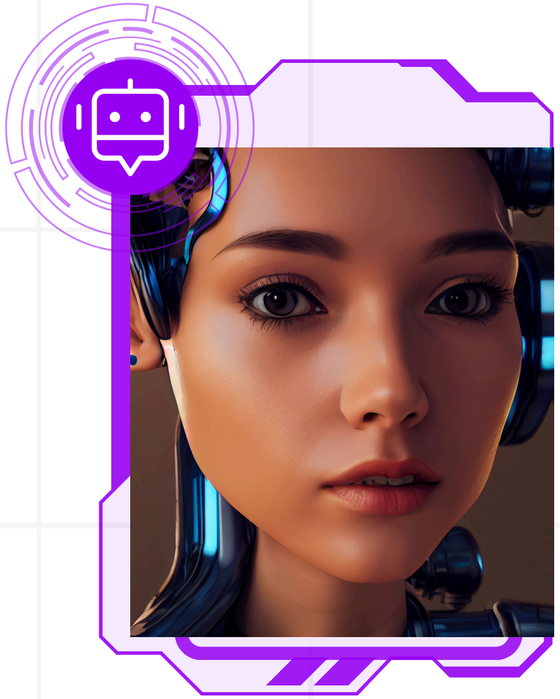
Our goal is to detect fraud early and accurately to improve transaction safety for our customers.

How can we efficiently identify fraudulent transactions while minimizing false alarms?

NEW BANK

Bana Marco Amir Jean Philippe

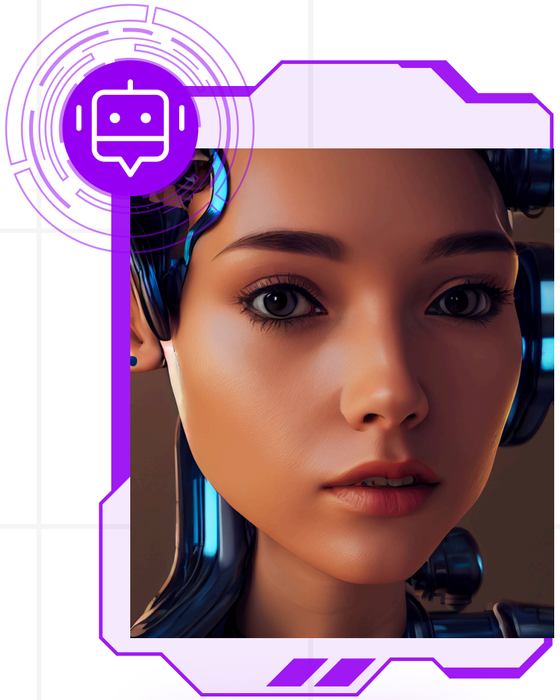
Security Operations for Payment Handling & Intelligent Analysis



NEW BANK

Bana Marco Amir Jean Philippe

Security Operations for Payment Handling & Intelligent Analysis

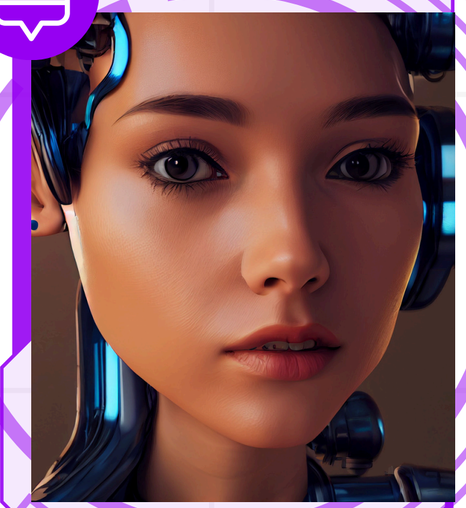


SOPHIA V1

NEW BANK

Bana Marco Amir Jean Philippe

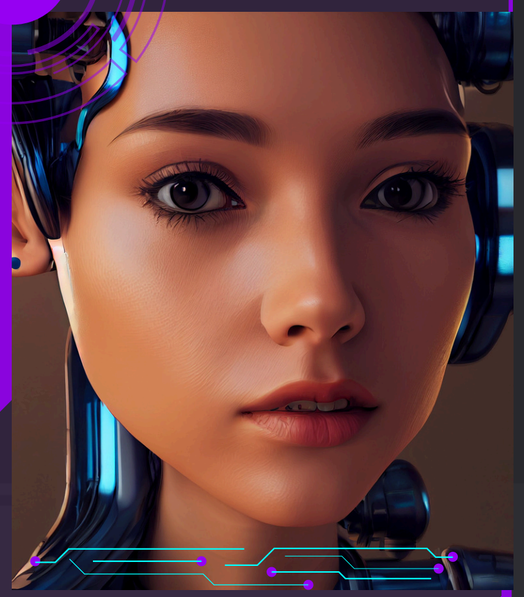
Security Operations for Payment Handling & Intelligent Analysis



Designed for our client's safety.

Accurate & Optimised to flag Fraudulent payment transaction as they appear.

SOPHIA V1



SOPHIA V1

MY MISSION

1. My mission will consist of analysing New bank’s daily payment transactions and identify fraudulent activities for safer banking transactions.

2. Improve New bank’s internal processes and security protocols regarding online payment.

List of data points included:

step: unit of time (1 hour)
type: type of online transaction
amount: the amount
nameOrig: customer
oldbalanceOrg: customer balance before transaction
newbalanceOrig: customer balance after transaction
nameDest: recipient
oldbalanceDest: recipient balance before transaction
newbalanceDest: recipient balance after the transaction
isFraud: fraud marker

The dataset contains information about online payment fraud, to understand what type of transactions lead to fraud.

# step	type	# amount	nameOrig	# oldbalanc...	# newbalan...	nameDest	# oldbalanc...	# newbalan...	# isFraud
1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0
1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0
1	TRANSFER	181.0	C1305486145	181.0	0.0	C553264065	0.0	0.0	1
1	CASH_OUT	181.0	C840083671	181.0	0.0	C38997010	21182.0	0.0	1
1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0

Payment type

type
PAYMENT
PAYMENT
TRANSFER
CASH_OUT
PAYMENT

Sender's Account Balance

# amount	nameOrig	# oldbalanc...	# newba
9839.64	C1231006815	170136.0	160296.36
1864.28	C1666544295	21249.0	19384.72
181.0	C1305486145	181.0	0.0
181.0	C840083671	181.0	0.0
11668.14	C2048537720	41554.0	29885.86

# oldbalanc...	# newbalan...
0.0	0.0
0.0	0.0
0.0	0.0
21182.0	0.0
0.0	0.0
0.0	0.0
M1230701703	0.0	0.0

Receiver's Account Balance

is Fraud?

isFraud
0
0
1
1
0
0

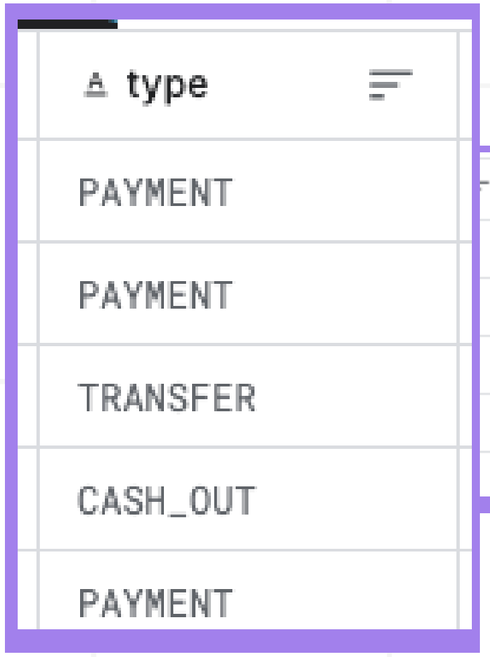
Data Imbalance

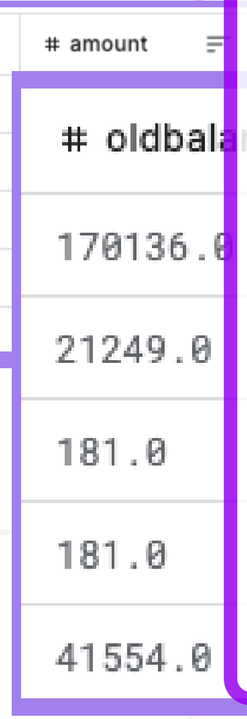
NON FRAUD PAYMENT

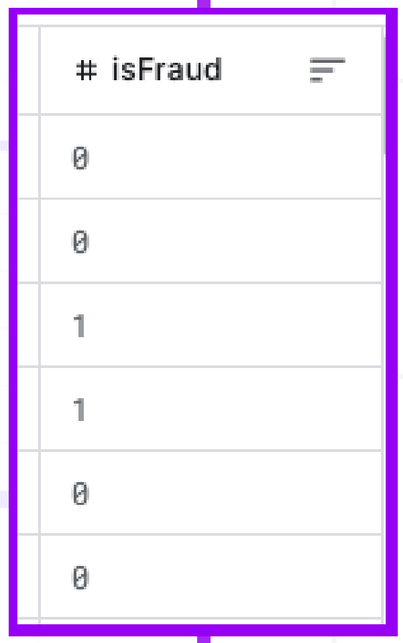
6,354,407

FRAUD PAYMENT

8,213







SMOTE Sampling Technique Used

Our Method

Data Collection

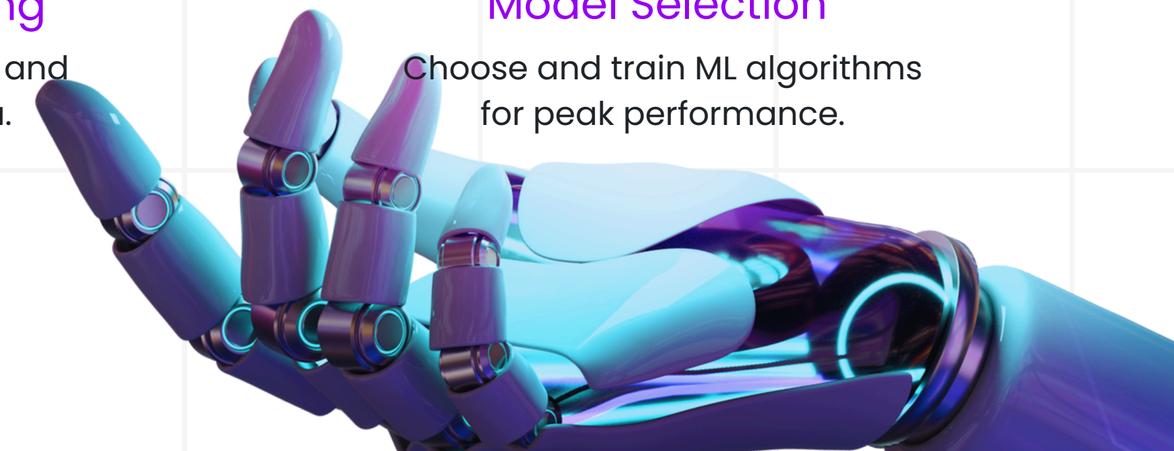
Gather relevant data from diverse sources.

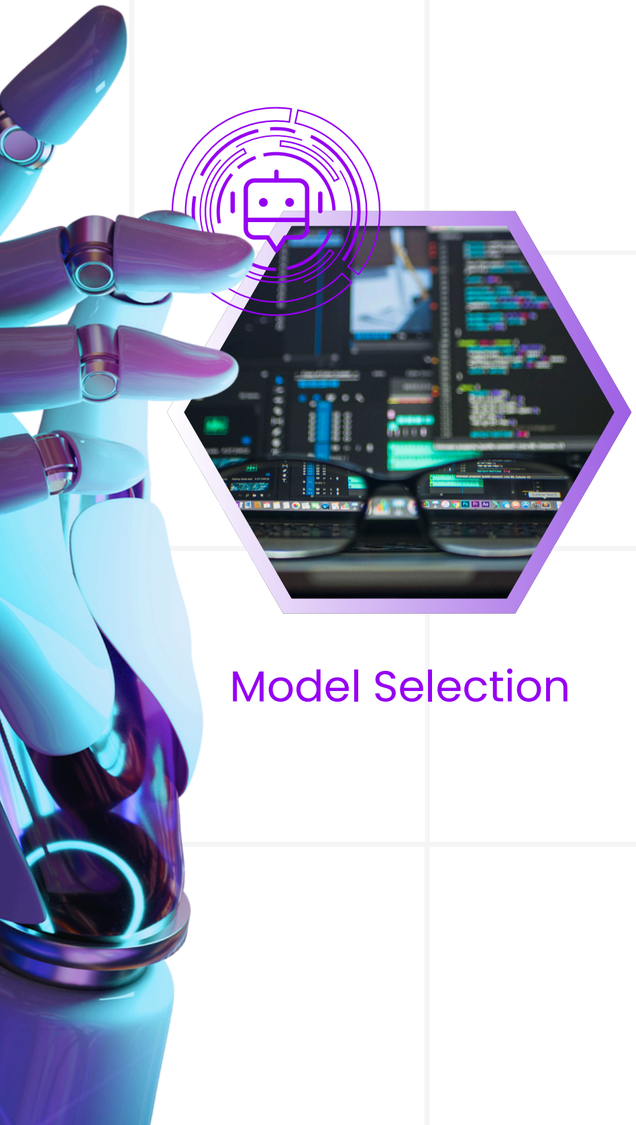
Preprocessing

Clean, normalize, and engineer data.

Model Selection

Choose and train ML algorithms for peak performance.





Model Selection

KNN Model

Simple & Effective: *KNN is intuitive, making predictions based on the proximity of similar data points without needing complex training.*

Pattern Adaptation: *Works well in scenarios where fraudulent patterns change, as it relies on real-time similarities in the data.*

Descision Tree Model

High Accuracy & Interpretability: *Achieved 100% accuracy and is easy to interpret, with clear, step-by-step decision paths.*

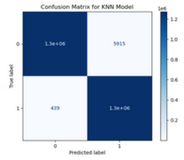
Handles Imbalanced Data: *Decision Trees manage large, skewed datasets like fraud detection efficiently, making them robust for financial use cases.*

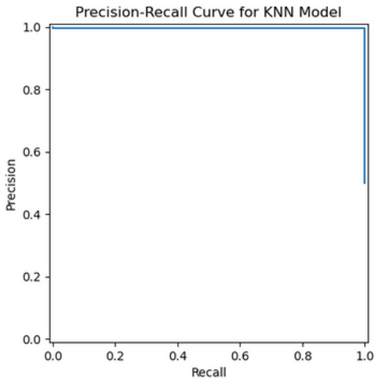
KNN Model

***Accuracy Test:** 1*

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1269605
1	1.00	1.00	1.00	1272158
accuracy			1.00	2541763
macro avg	1.00	1.00	1.00	2541763
weighted avg	1.00	1.00	1.00	2541763

KNN Confusion Matrix:
[[1263598 6007]
[401 1271757]]





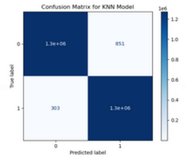
MAE: 0.00
RMSE: 0.02
R2 score: 0.62

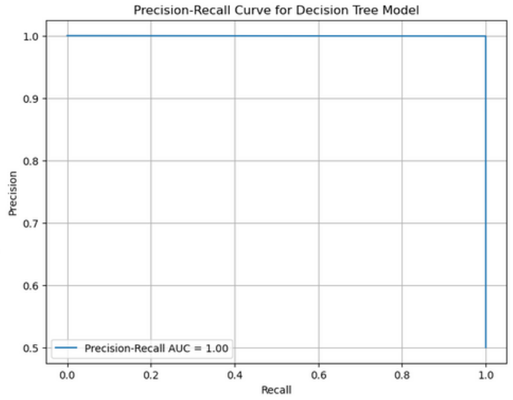
Descision Tree Model

***Accuracy Test:** 1*

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1269605
1	1.00	1.00	1.00	1272158
accuracy			1.00	2541763
macro avg	1.00	1.00	1.00	2541763
weighted avg	1.00	1.00	1.00	2541763

Decision Tree Confusion Matrix:
[[1268754 851]
[303 1271855]]





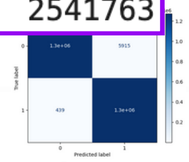
MAE: 0.02
RMSE: 0.10
R2 score: 0.96

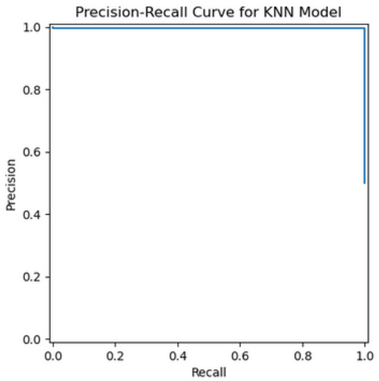
KNN Model

Accuracy Test: *1*

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1269605
1	1.00	1.00	1.00	1272158
accuracy			1.00	2541763
macro avg	1.00	1.00	1.00	2541763
weighted avg	1.00	1.00	1.00	2541763

```
[[1263598    6007]
 [    401 1271757]]
```





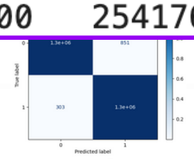
MAE: 0.00
RMSE: 0.02
R2 score: 0.62

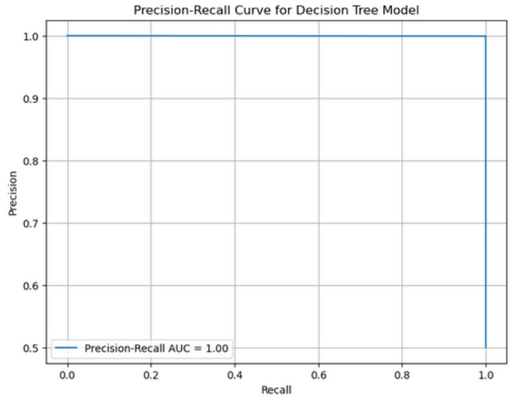
Descision Tree Model

Accuracy Test: *1*

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1269605
1	1.00	1.00	1.00	1272158
accuracy			1.00	2541763
macro avg	1.00	1.00	1.00	2541763
weighted avg	1.00	1.00	1.00	2541763

```
[[1268754     851]
 [    303 1271855]]
```



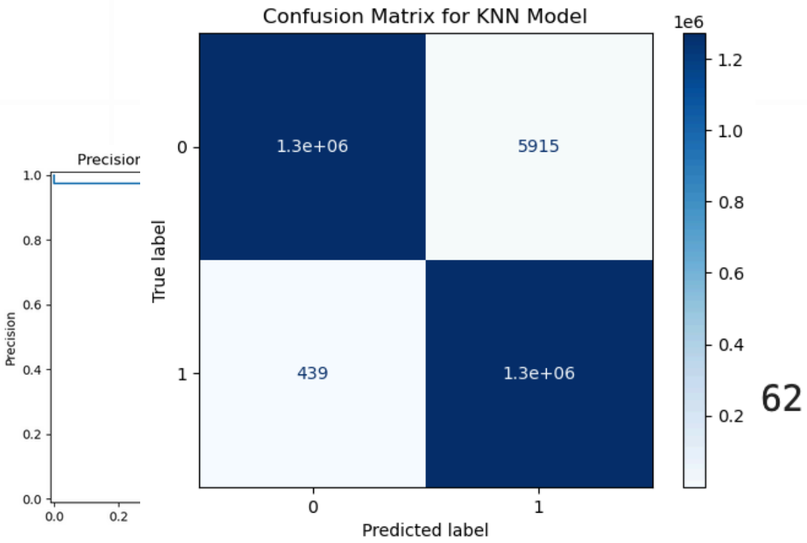


MAE: 0.02
RMSE: 0.10
R2 score: 0.96

KNN Model

***Accuracy Test:** 1*

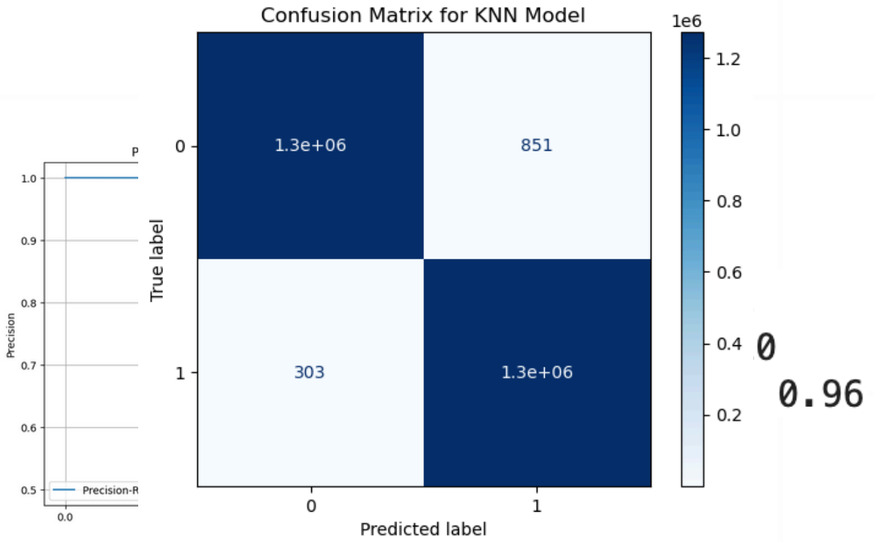
```
KNN Confusion Matrix:
[[1263598    6007]
 [    401 1271757]]
```



Descision Tree Model

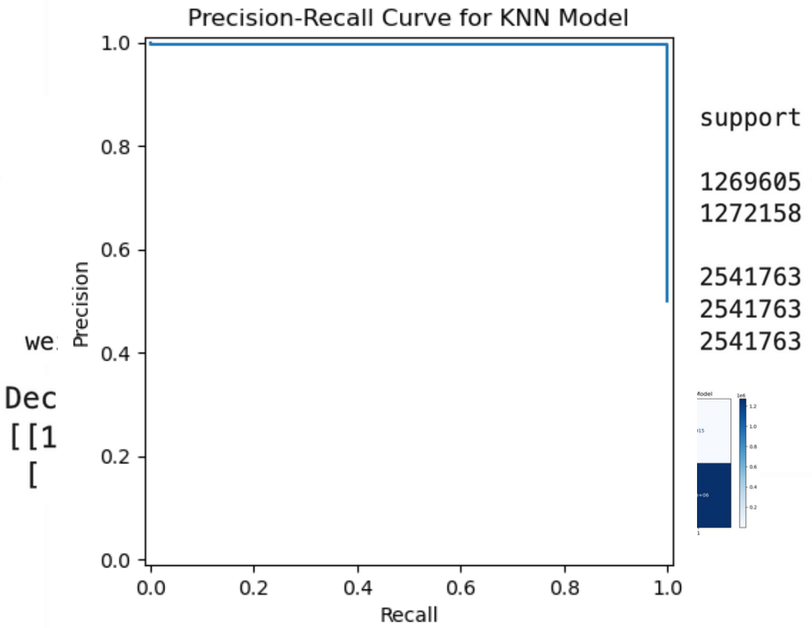
***Accuracy Test:** 1*

```
Decision Tree Confusion Matrix:
[[1268754     851]
 [    303 1271855]]
```



KNN Model

***Accuracy Test:** 1*

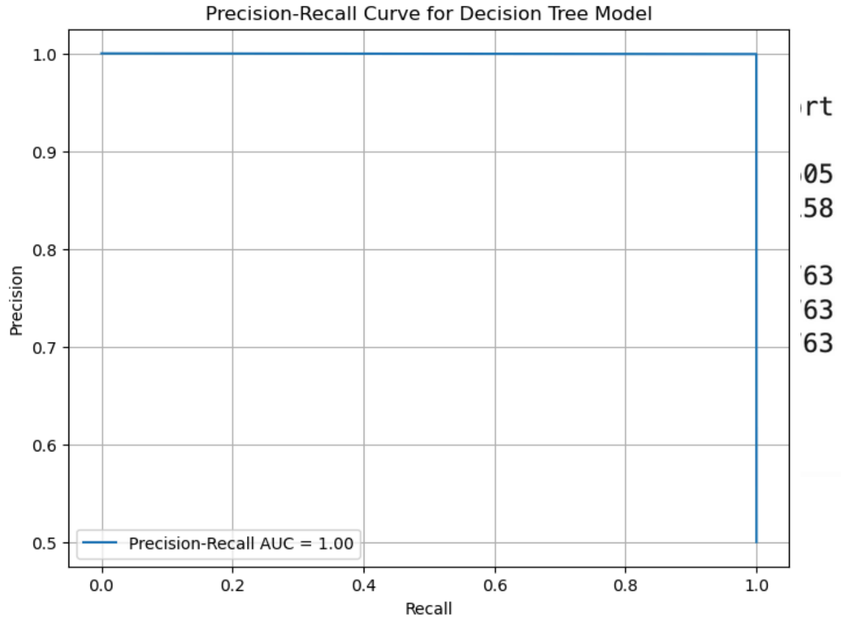



```
                 support

1269605
1272158

2541763
2541763
2541763
```

```
MAE: 0.00
RMSE: 0.02
R2 score: 0.62
```

Descision Tree Model

***Accuracy Test:** 1*




```
MAE: 0.02
RMSE: 0.10
R2 score: 0.96
```

Descision Tree Model

From the Classification Report:

the performance across classes is balanced, and class imbalance (if present) is well handled. Precision, Recall and F1-Score all equals to 1.00

From the confusion matrix: => high precision and recall

The model predicts almost all instances correctly, but there are a small number of misclassifications (851 false positives and 303 false negatives). However, given the scale of the dataset, these numbers are relatively small.

From the MAE, RMSE, and R^2 Score:

on average error prediction is quite small (0.02) and difference between real and actual values are also small (0.10) indicating high accuracy with very good predictions.

Descision Tree Model

We then decided to test those numbers:

Cross validation of the Training Set vs Test Set

MAE: *0.23*

RMSE: *0.42*

R^2 score: *0.08*

Test set MAE: *0.02*

Test set RMSE: *0.10*

Test set R^2 score: *0.94*

Training set MAE: *0.00*

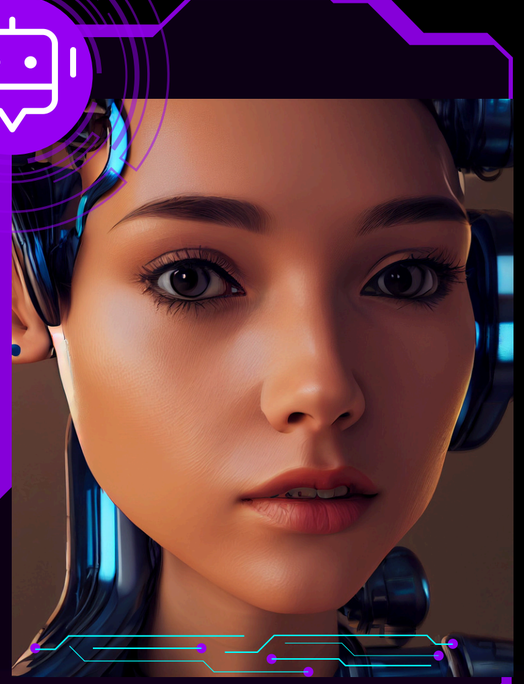
Training set RMSE: *0.00*

Training set R^2 score: *1.00*

Conclusion

The **SOPHIA v1 model**, powered by a **decision tree algorithm**, is highly effective in identifying both fraudulent and non-fraudulent transactions with minimal errors on its trained data. However, to enhance its practical application, further refinement is needed to make it as efficient on new test data.

Overall, SOPHIA proves to be a robust solution for payment fraud detection in the bank's operations.



SOPHIA

THANK YOU

I will be happy to answer all your questions and hope this presentation on my machine learning model was insightful and up to your expectations.