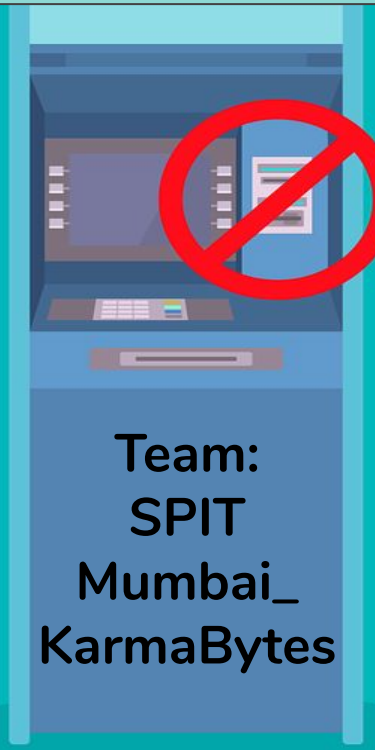
TechE

FRAUD DETECTION SYSTEM

Deloitte.



Team:
SPIT
Mumbai_
KarmaBytes

Team Members :
Pulin Prabhu
Tania Rajabally
Anisha Gharat

College: Sardar Patel Institute of Technology
Theme: Data Science / Machine Learning

TECHNOUTSAV 3.0

Round 2

Phase 3 – Proof Of Concept - Prototype

Defined Deliverables	Sample
What is the qualitative and quantitative impact of your proposed solution	Provided in presentation  Issue2Impact
Detailed Solution Design and Architecture (includes Data Model Design)	 Solution Design & Architecture  DataModel
Solution/Code to be made available on cloud	Development is done on cloud (Details of code components is available in execution guide)
Test Data to be made available on cloud	The test data is available in the form of as per designed Data Model (Details are available in execution guide)
Executed Test cases and Results	Test case is attached with the submission  Test_Case.xlsx
Execution Guide (including cloud login credentials)	Step by step guide to install application with screen shots are attached

Business Problem

Fraud Rates increasing alarmingly in the finance sector ⊕ **Data Collected from various points that are not consolidated enable this** ➡ **A Fraud Detection System to tackle this issue**

Reeling Under the Pressure

Bank Group/ Institution	2017-18		2018-19	
	Number of Frauds (₹ cr)	Amount involved (₹ cr)	Number of Frauds (₹ cr)	Amount involved (₹ cr)
Public Sector Banks	2,885	38,260.87	3,766	64,509.90
	-48.8	-92.9	-55.4	-90.2
Private Sector Banks	1,975	2,478.52	2,090	5,515.10
	-33.4	-6	-30.7	-7.7
Foreign Banks	974	256.00	762	955.30
	-16.5	-0.6	-11.2	-1.3
Financial Institutions	12	164.70	28	553.40
	-0.2	-0.4	-0.4	-0.8
Total	5,916	41,167.00	6,801	71,542.90

Source: RBI

Solution

Extract the Data + Provide Training Sets

- We have used two datasets- Online financial transactions frauds and Credit card transaction frauds.The data is split into three different segments – training, testing and cross - validation.
- The **train** data will be trained for prediction.
- The **test** data will be used for predicting the frauds.
- The high performing models will be cross - validated to ensure consistency in results.

Training & Building Models

- The data is trained for certain input parameters to predict boolean outputs.
- Models will be built for prediction based on previous examples of data.
- After experimentation, the model with the best accuracy is found.

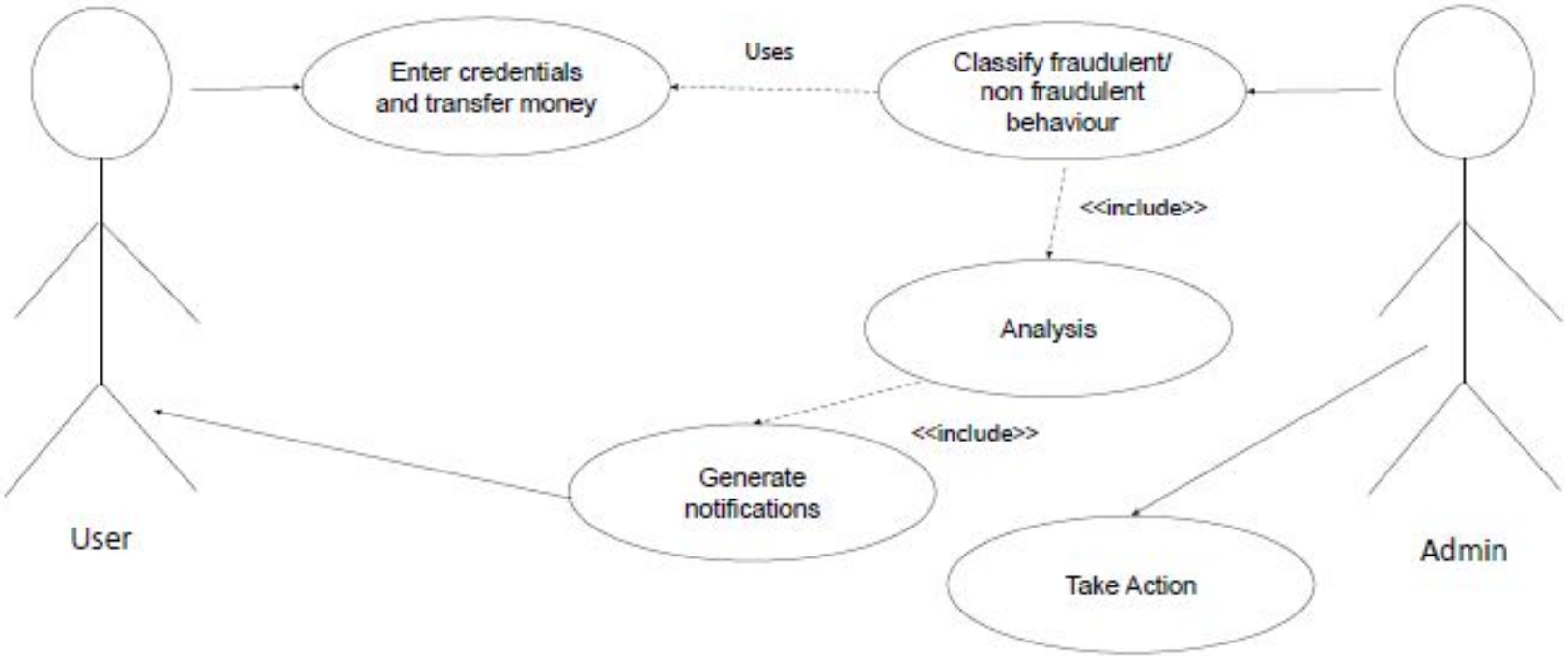
Detect/Predict Frauds

- The first model will predict whether a new transaction is **fraudulent** or not. The second model will predict a credit card transaction fraud.
- The model will give a probability score of fraud based on earlier scenarios.
- Any fraud by individual or companies will be detected on a single platform.

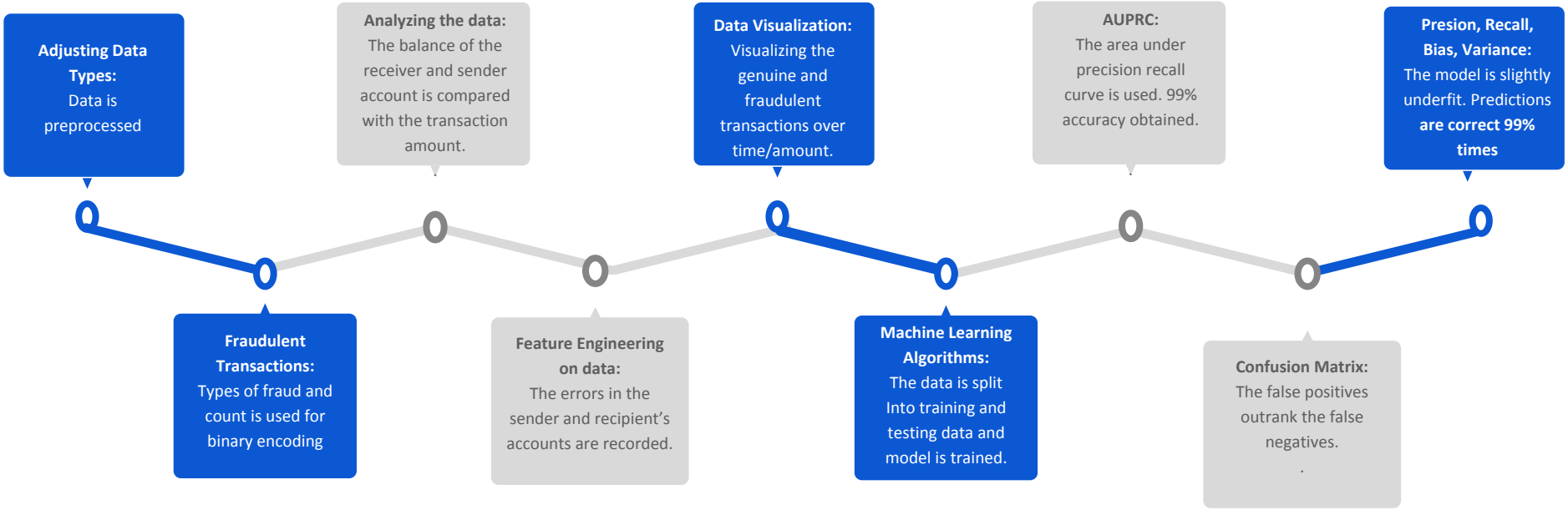
Notify/Take actions

- The risk of frauds will be estimated.
- Based on the predictive analysis and probabilities of frauds the company can either **Accept** or **Reject** transactions and credit card frauds can be stopped..

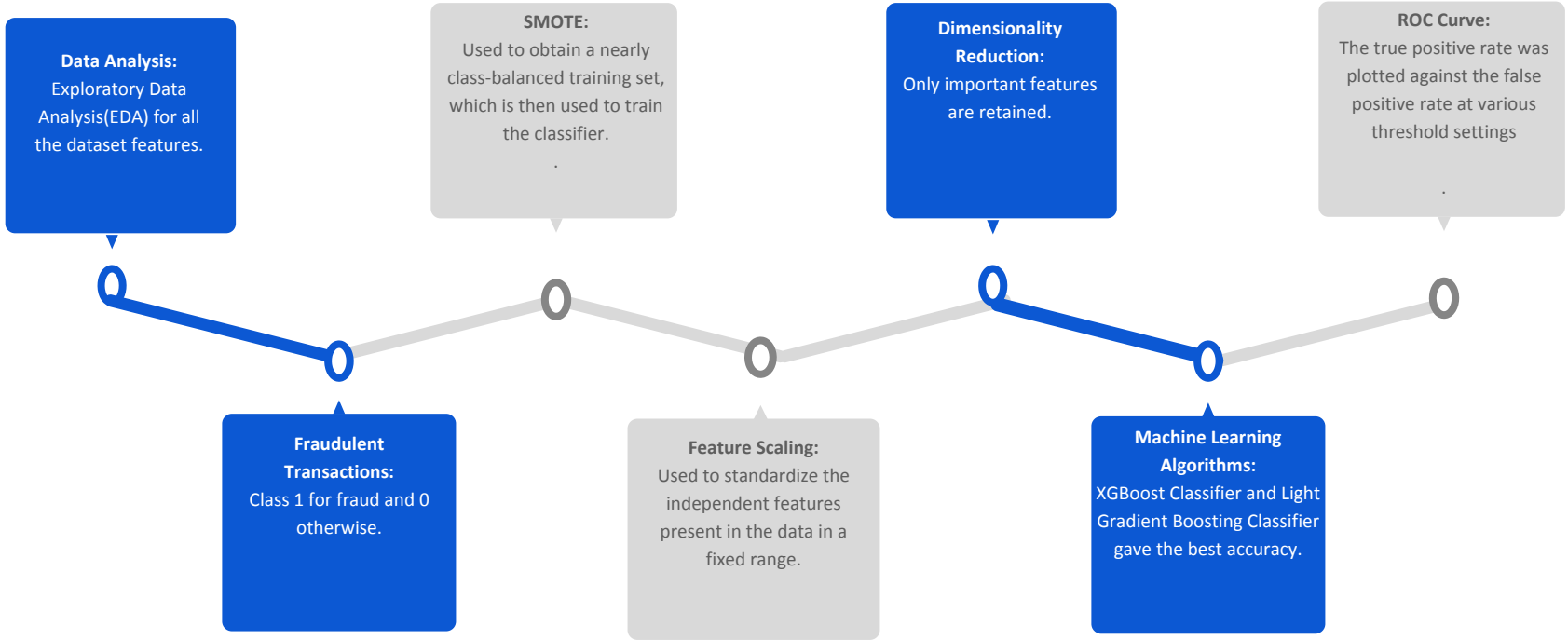
Use Case


Enter credentials and transfer money
Uses
Classify fraudulent/ non fraudulent behaviour
<<include>>
Analysis
<<include>>
Generate notifications
Take Action
User
Admin

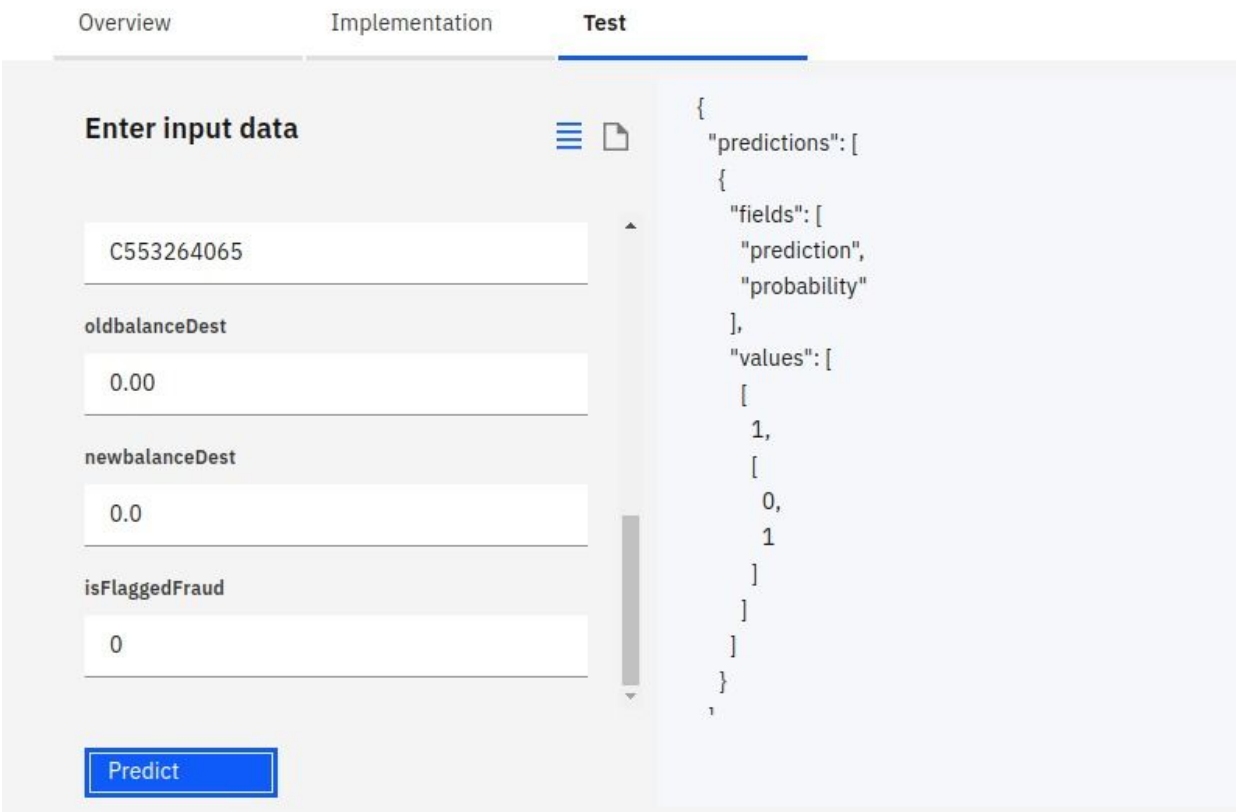
Detailed Solution Design and Architecture (Online Payments Transaction Frauds)



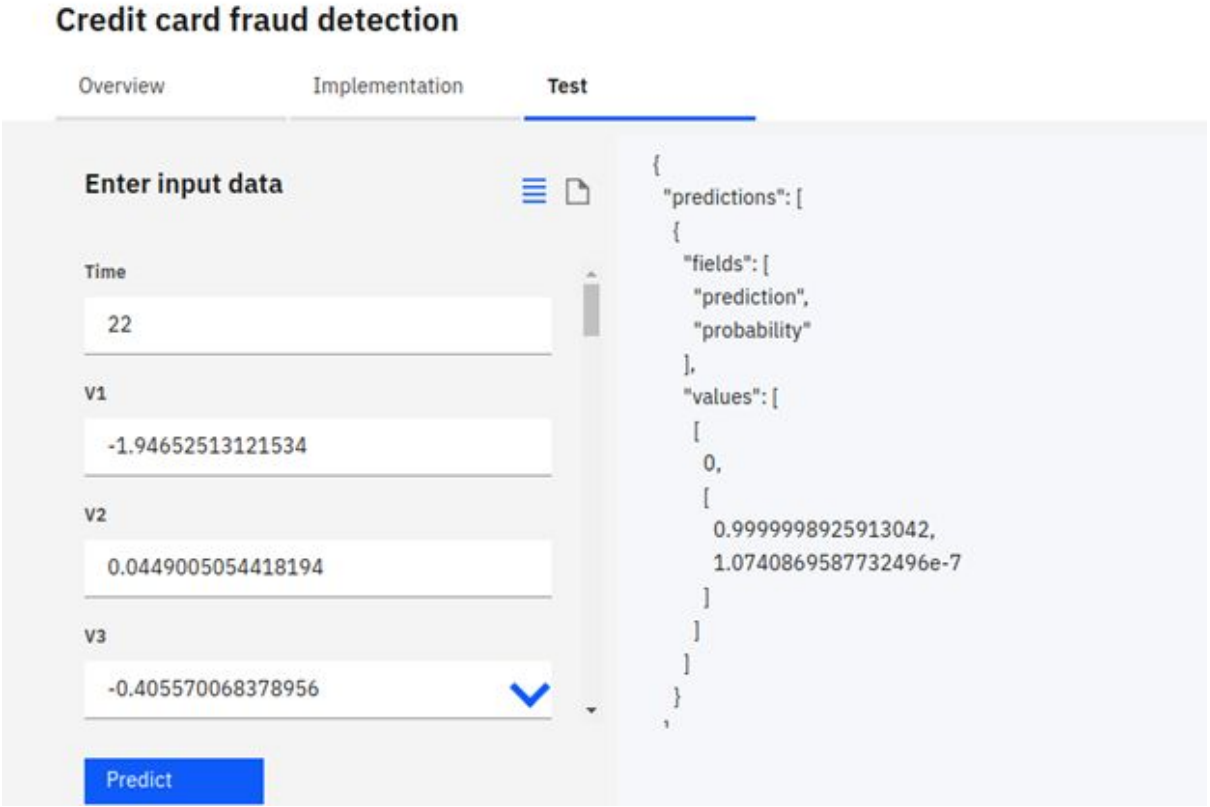
Detailed Solution Design and Architecture (Credit Card Transaction Frauds)



Executed Test Cases (Online Payment Transaction Fraud)



Executed Test Cases (Credit Card Transaction Fraud)



Results (Online Payment Transaction Frauds)

- It was found that Decision tree and XGB Classifier gave the best results
- It gave an accuracy of 99.96% with Decision Tree
- It gave an accuracy of 99.99% with XGB Classifier
- Since the data is highly skewed, the area under the precision-recall curve (AUPRC) is used.

```
AUPRC = 0.996797
```

```
Decision Tree Classifier Accuracy: 0.9996361561746576
```

```
XGBClassifier Accuracy: 0.9999837569168463
```

Results (Credit Card Transaction Frauds)

- It was found that Light Gradient Boosting Classifier and XGBoost Classifier gave the best results
- It gave an accuracy of 97.63% with Light Gradient Boosting Classifier
- It gave an accuracy of 97.16% with XGBoost Classifier

LGB Classifier Accuracy: 0.9733333333333334

LGB Classifier Classification report:

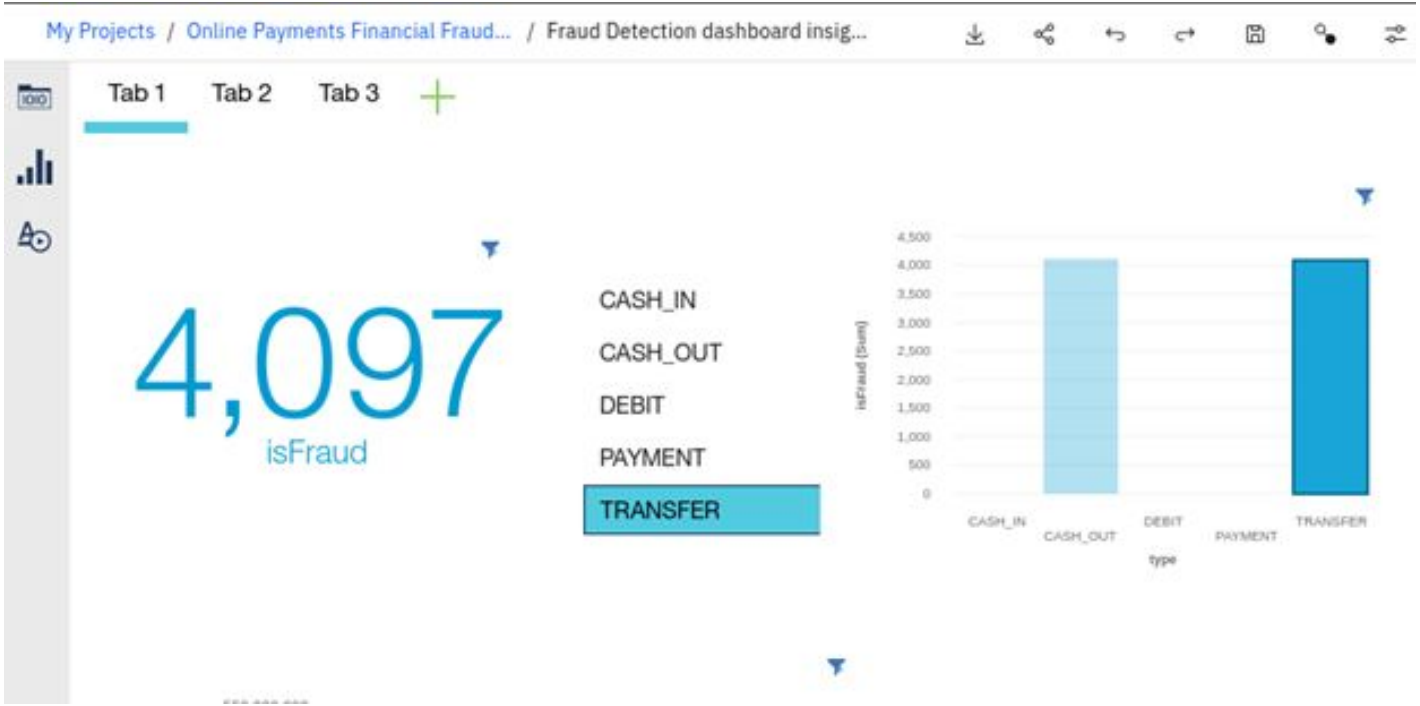
	precision	recall	f1-score	support
0.0	0.97	0.98	0.97	620
1.0	0.98	0.97	0.97	580
accuracy			0.97	1200
macro avg	0.97	0.97	0.97	1200
weighted avg	0.97	0.97	0.97	1200

XGBoost Classifier Accuracy: 0.9725

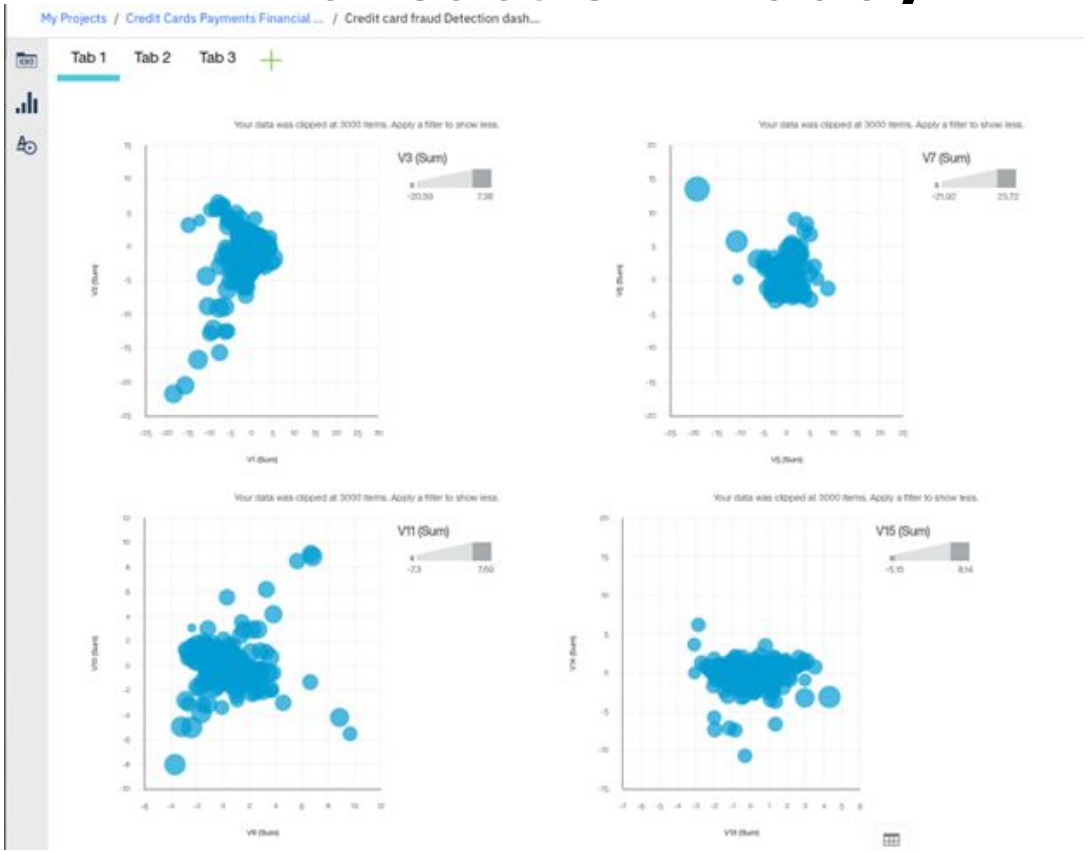
XGBoost Classifier Classification report:

	precision	recall	f1-score	support
0.0	0.97	0.98	0.97	620
1.0	0.98	0.97	0.97	580
accuracy			0.97	1200
macro avg	0.97	0.97	0.97	1200
weighted avg	0.97	0.97	0.97	1200

Dashboard for Visualization (Online Payment Transaction Fraud)



Dashboard for Visualization (Credit Card Transaction Fraud)



Business Model