

PITCH DECK

Detection of Fraudulent Bank Transactions

George Shameih, BOUCHABCHOUB Badr, ROUGEMONT Luca,
BEN RHOUMA Seifeddine, CHARYYEV Serdar, BERNHART Eric

TABLE OF CONTENTS



Introduction & Problem Statement

Project Objective

Dataset Overview

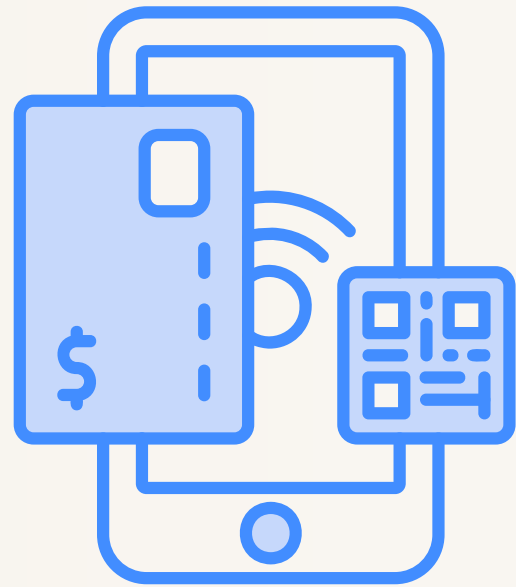
Methodology & Approach

Results & Model Performance

Dashboard & Visualizations

Conclusion

THE PROBLEM



Fraud is the act of deceiving a financial institution or user to gain unauthorized monetary benefit.

Global card fraud losses exceeded \$12.5 billion in 2024
(Federal trade commission)

Bank fraud continues to grow year after year.
A 24% rise between 2023 and 2024.



PROJECT OBJECTIVE



Develop a Machine Learning model capable of predicting whether a banking transaction is fraudulent (1) or legitimate (0) using historical transaction, user, and card data.



DATASET OVERVIEW

210 000 transactions

File Name	Description
transactions_train.csv	Training data containing transaction features (amounts, dates, merchants, etc.)
train_fraud_labels.json	Labels for training transactions (1 = fraud, 0 = non-fraud)
cards_data.csv	Information about payment cards (e.g., card type, issuer)
users_data.csv	User profile and demographic data
mcc_codes.json	Merchant Category Codes and descriptions
evaluation_features.csv	Evaluation transactions (no labels, used only for prediction)

METHODOLOGY & APPROACH

The screenshot shows a data science tool interface. On the left, a sidebar titled 'Étapes (5)' lists the workflow steps: 1. Convertir le type de colonne (with a description of automatic conversion), 2. Convertir le type de colonne (with a description of manual conversion for 4 columns), and 3. Jointure (with a description of a full join operation). The main area displays a table of transaction data with columns: transacti..., date, client_id_x, card_id, and amount. The table contains 16 rows of data. At the bottom, there is a 'Nouvelle étape' button and a status bar showing 'Affichage : 10000 ligne(s), 38 colonne(s)' and 'Ensemble de données échantillonné : 8978 ligne(s), 12 colonne(s)'.

	transacti...	date	client_id_x	card_id	amount
1	19765990	2017-07-07 09:53:...	1581	2519	\$12.35
2	22160255	2018-11-26 17:44:...	1862	4049	\$58.29
3	17566794	2016-03-26 12:42:...	1967	3367	\$11.03
4	17318690	2016-02-01 08:30:...	921	3457	\$85.74
5	20994060	2018-03-24 14:42:...	456	2800	\$13.43
6	20501200	2017-12-09 22:26:...	0	4639	\$53.12
7	18454176	2016-10-01 21:40:...	408	4960	\$52.86
8	21998064	2018-10-23 12:50:...	1432	4929	\$35.33
9	18977227	2017-01-21 09:30:...	177	2681	\$188.77
10	17244867	2016-01-16 13:40:...	1895	1090	\$27.40
11	17266754	2016-01-21 06:30:...	846	3443	\$4.43
12	20085114	2017-09-12 20:08:...	1904	1004	\$54.00
13	20446347	2017-11-28 11:58:...	696	39	\$100.41
14	20741827	2018-01-29 22:37:...	1439	6094	\$31.98
15	19788807	2017-07-12 08:07:...	762	1080	\$63.54
16	17292719	2016-01-26 15:30:...	1358	3426	\$26.96

1

Data preparation

type conversion, table
joins

2

Models tested
with auto AI

Logistic Regres..sion,
Random Forest, XGBoost...

3

Evaluation metrics

ROC-AUC,
precision, recall,
F1-score

4

Validation strategy

ensuring generalization
to new (unseen) data

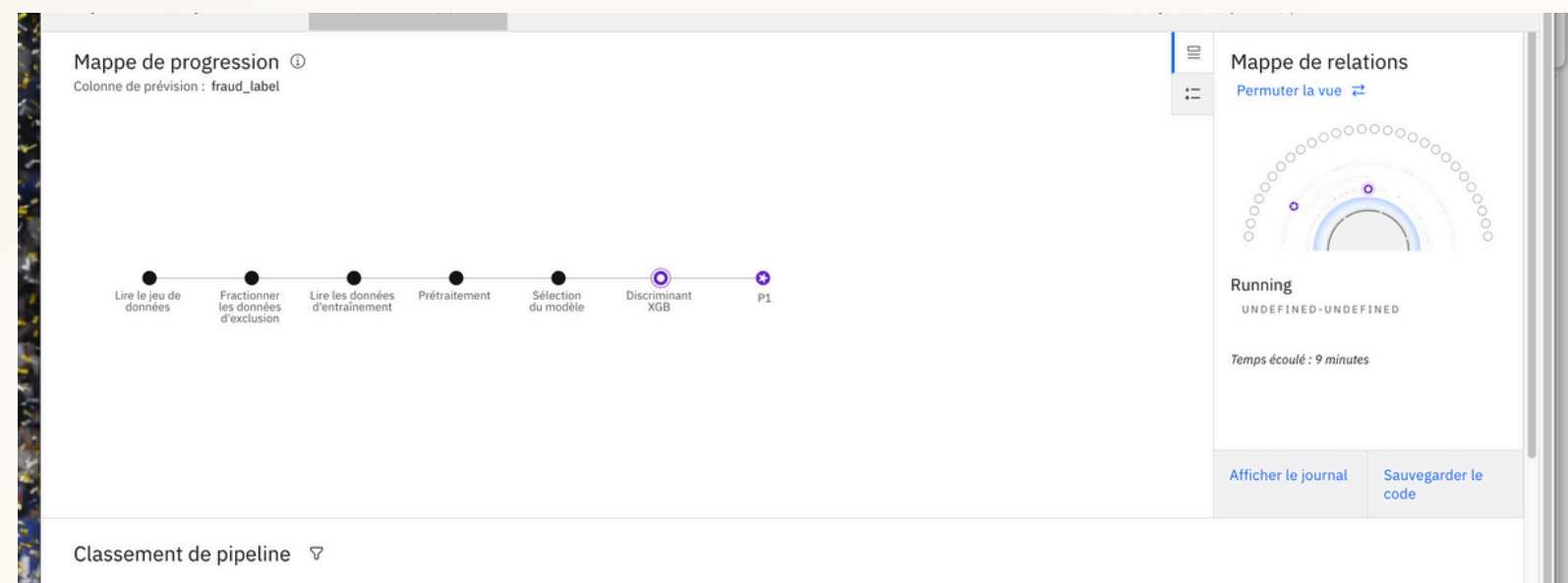
RESULTS & MODEL PERFORMANCE



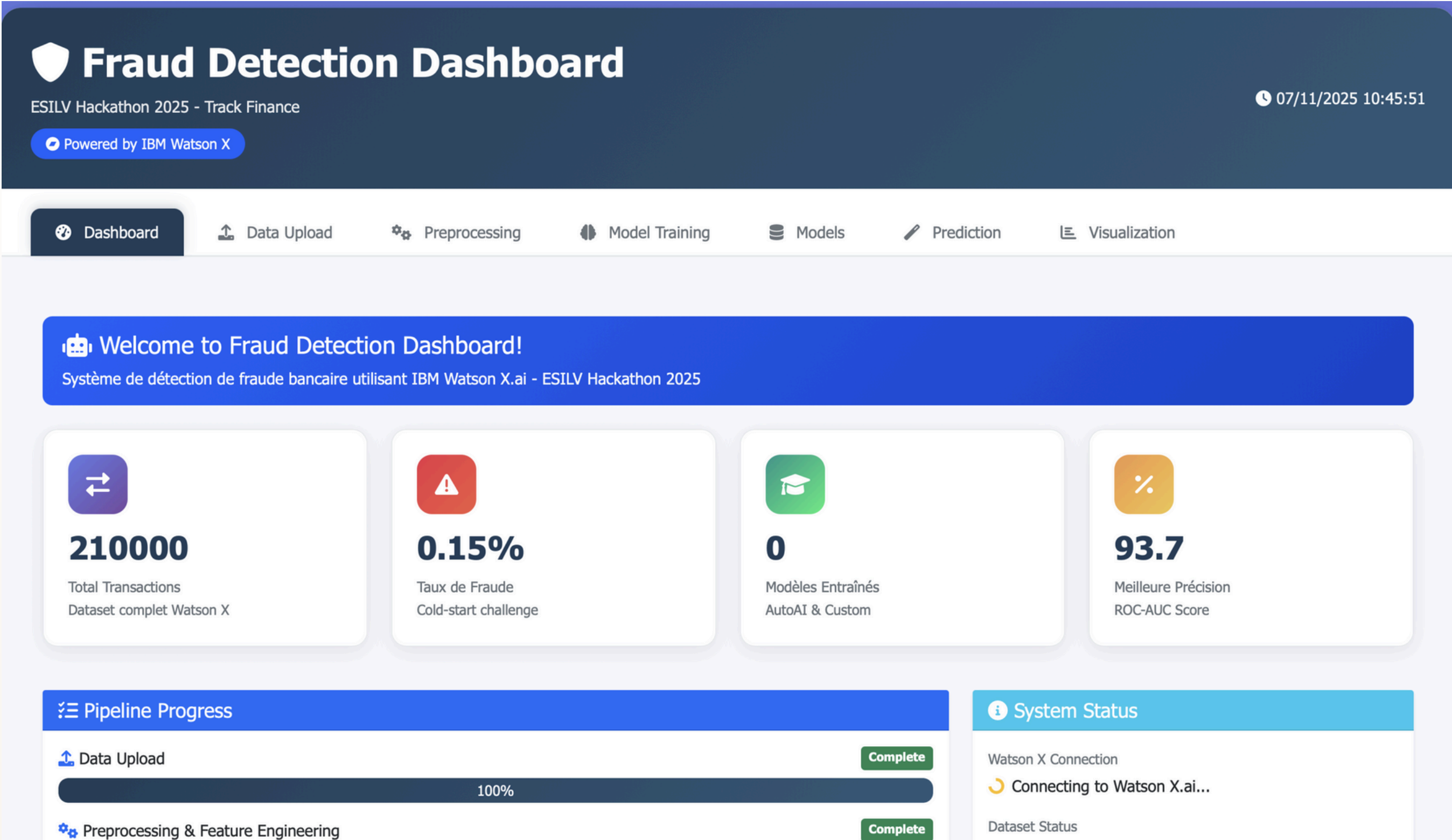
XGboost



Acuracy 0.99
F1 0.71



DASHBOARD & VISUALIZATIONS



THE END

Thank Your For Listening