



PasDeFraude Company

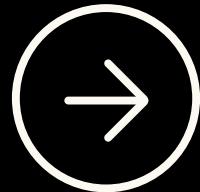
PITCH DECK



Presented by Team 2

Raphael Roux, Léonard Seidlitz, Adrien Servas, Alexandre Morel, Alexandre Francony, Romain Requena

PITCH SUMMARY



- 1 Introduction
- 2 Fraud Definition
- 3 Problem statements
- 4 Data analysis
- 5 Borderline Smote
- 6 Model Training
- 7 Model Training
- 8 Tool comparaison



INTRODUCTION

We are redefining financial security through advanced fraud detection. In an era where financial institutions face increasingly sophisticated card and online payment fraud, our mission is to provide intelligent, adaptive, and scalable protection.

Our technology focuses on developing models that not only detect known fraud patterns but also generalize to unseen clients—addressing the critical cold start challenge. By combining state-of-the-art machine learning with real-world data insights, we enable banks and payment providers to stay one step ahead of fraudsters, ensuring trust and safety in every transaction.



FRAUD DEFINITION

Key Forms

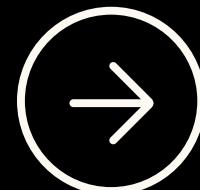
- Card Theft: Physical stealing of a credit or debit card.
- Card Skimming: Copying card details using malicious devices at ATMs or terminals.
- Online Fraud: Using stolen card information for online transactions.
- Identity Theft: Using personal information to open or manipulate credit accounts.

Impacts

1. Significant financial losses for banks and customers.
2. Erosion of trust in digital payment systems.
3. Increased need for advanced fraud detection systems leveraging AI and machine learning.



PROBLEM STATEMENT



Data Preprocessing

The challenge lies in effectively preprocessing raw financial transaction data to handle imbalance, missing values, and inconsistencies while preserving patterns essential for accurate fraud detection.

Dataset enhancement

The main challenge is to enhance the dataset with meaningful features and transformations that improve model performance without introducing bias or data leakage.

Model Training

The challenge is to develop a generalized model capable of adapting to new, unseen clients and evolving fraud patterns without compromising detection accuracy.

Dashboard

The goal is to design an interactive and intuitive dashboard that clearly visualizes fraud detection results and key performance metrics for effective decision-making.

DATA ANALYSIS



General information

Valeurs manquantes (top 10) :

	Missing	Percentage
errors	206727	98.44
zip	26797	12.76
merchant_state	25431	12.11
transaction_id	0	0.00
date	0	0.00
client_id	0	0.00
card_id	0	0.00

Labels normalisés :

Total : 210,000

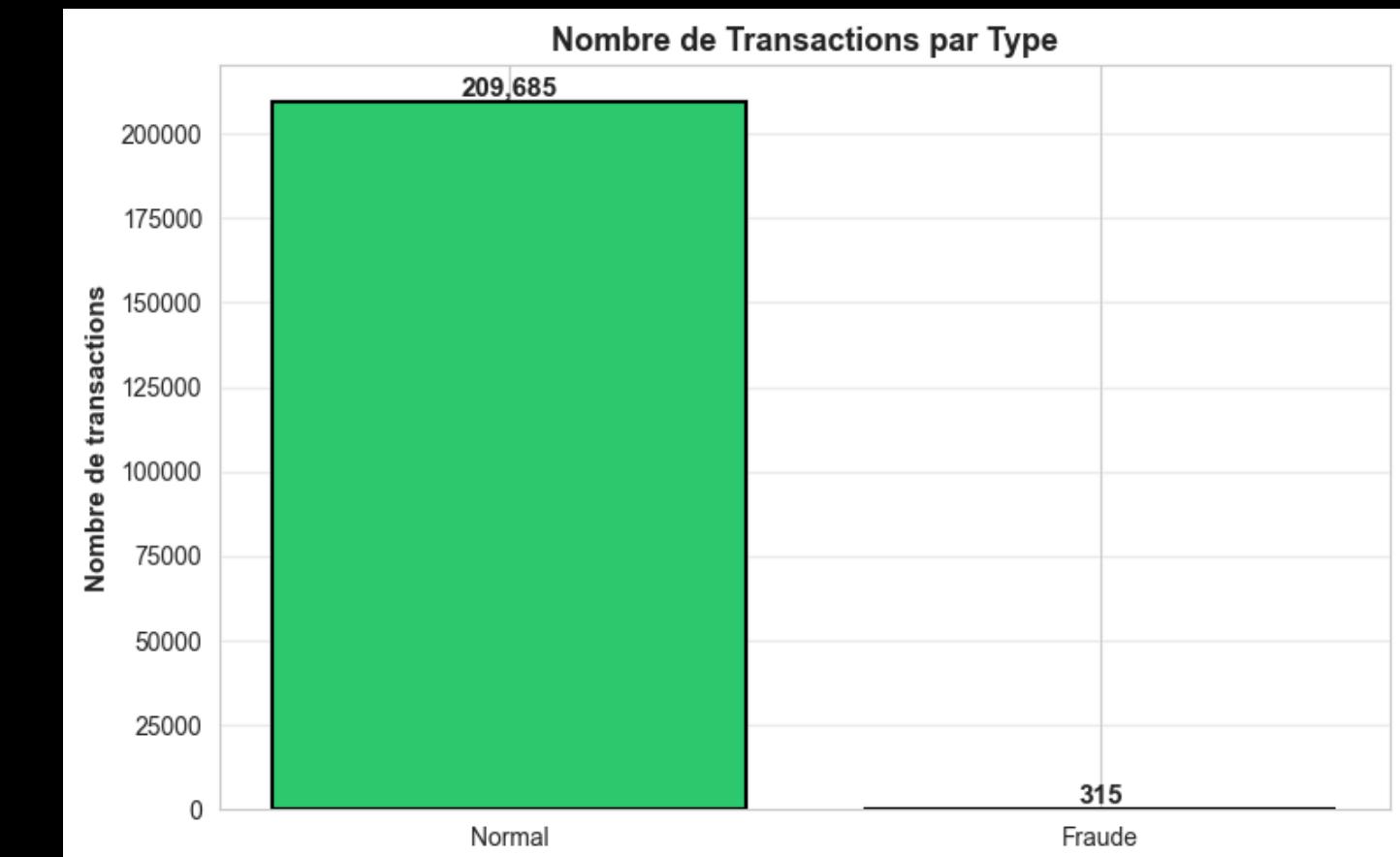
Fraude (1) : 315 (0.15%)

Normal (0) : 209,685 (99.85%)

Inconnu (None) : 0

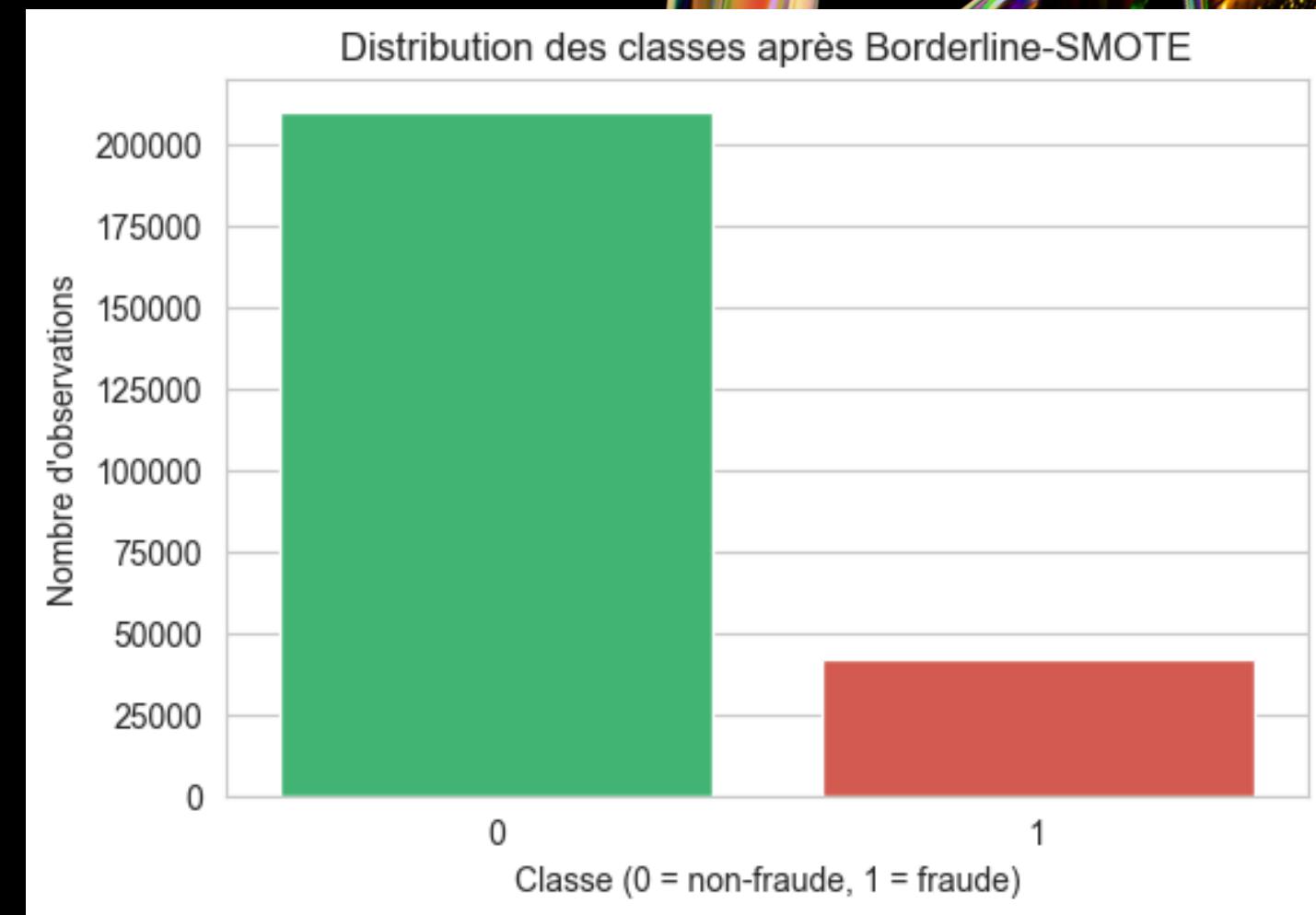


Status distribution



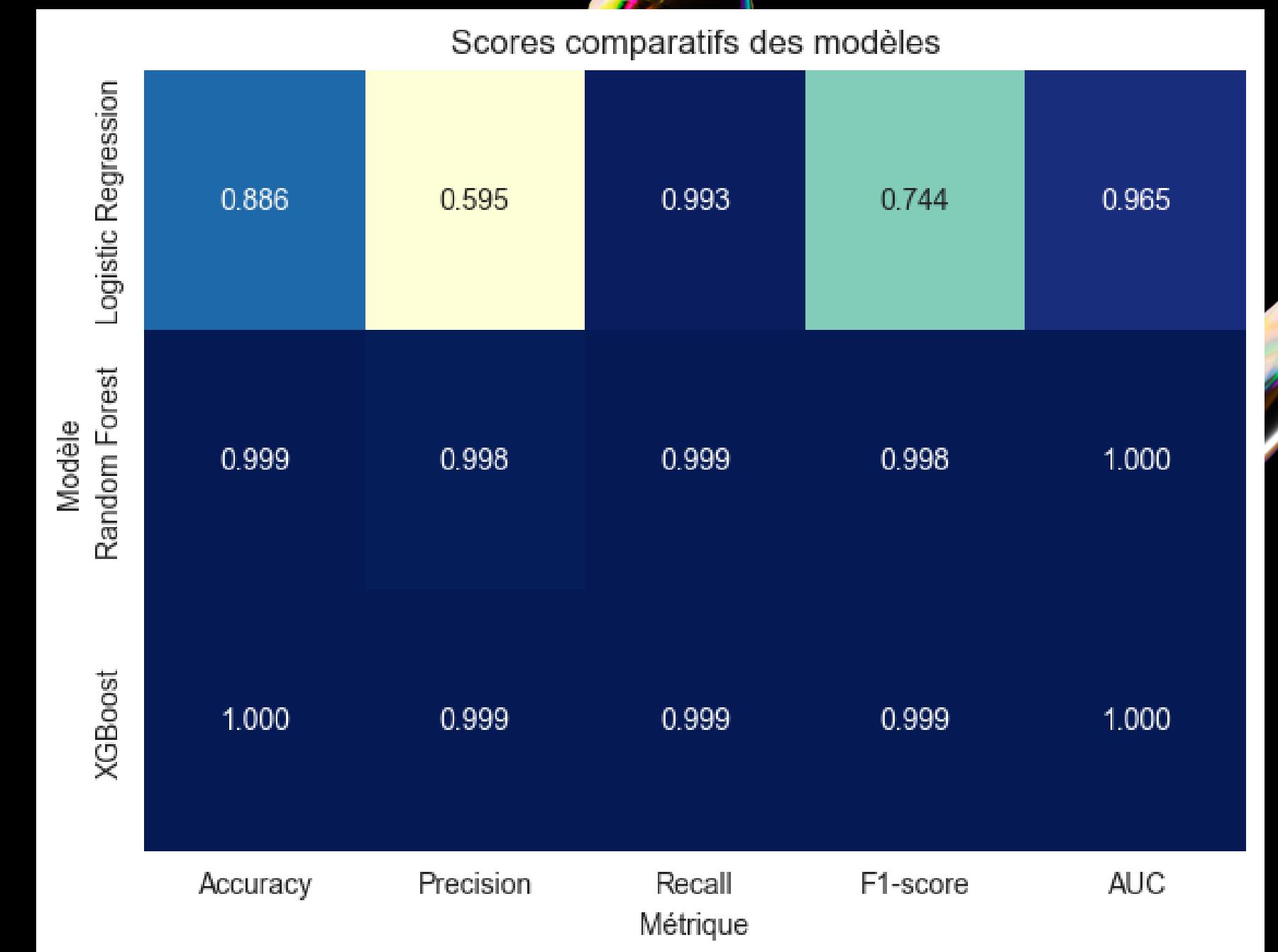
BORDERLINE SMOTE

- Synthetic Minority Over-sampling Technique focused on generating synthetic fraud samples near decision boundaries.
- Improves model sensitivity to fraud patterns.
- Reduces bias toward majority class, enabling better generalization and higher recall for fraud detection.



MODEL TRAINING

Model comparison



Evaluation de l'entraînement

Métrique	Données de formation	Données d'exclusion
Accuracy	0.9990197	0.9991654
Average precision	0.9951445	0.9957309
Balanced accuracy	0.9982884	0.9982594
F1	0.99705946	0.99749494
Log loss	0.031857736	0.0289588
Precision	0.99692744	0.9980902
Recall	0.99719155	0.9969003
Roc auc	0.99849707	0.99827963

MODEL TRAINING

Snap Decision Tree Classifier
(WatsonX)

TOOL COMPARAISON

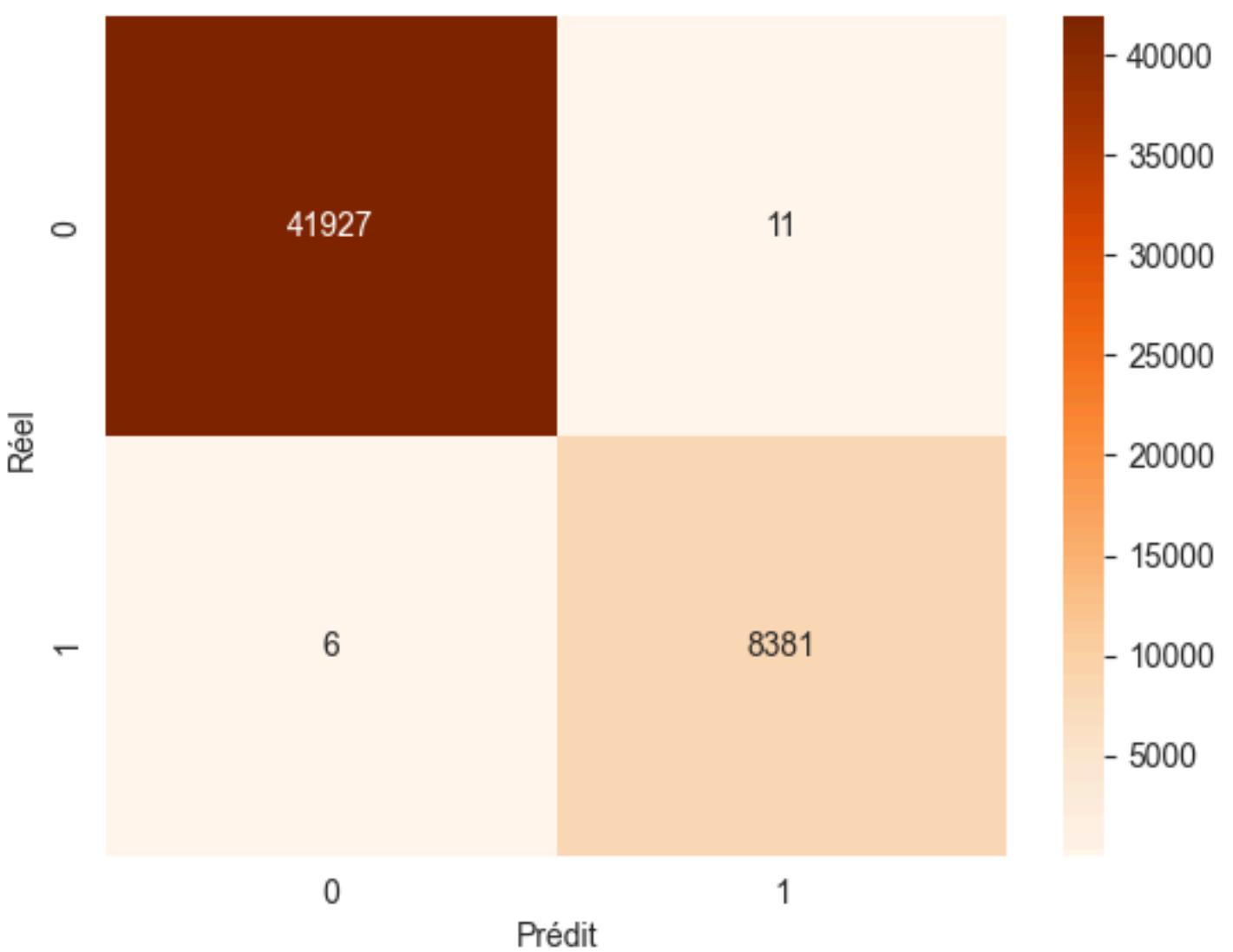
Matrice de confusion ⓘ

Observé	Prédit		Pourcentage correct
	1	0	
1	4181	13	99.7%
0	8	20960	100.0%
Pourcentage correct	99.8%	99.9%	99.9%

Le moins correct

Le plus correct

Matrice de confusion - XGBoost





PasDeFraude Company

THANK YOU

for your time and attention

