



**KHÓA LUẬN TỐT NGHIỆP**

**PHÁT HIỆN GIAN LẬN TÀI CHÍNH**

**DÙNG CÁC KỸ THUẬT**

**PHÂN TÍCH BẤT THƯỜNG**

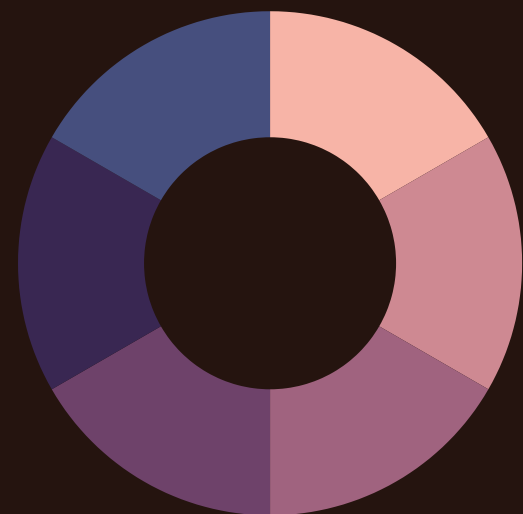
Lại Minh Phú (20127593)  
Hồ Minh Thanh Tài (20127068)

# Giới thiệu

---

- Động lực
- Mục tiêu
- Phạm vi

# Nội Dung Trình Bày



1. Tổng quan về đề tài
2. Công trình liên quan
3. Phương pháp đề xuất
4. Kết quả thực nghiệm
5. Kết luận
6. Hướng phát triển

# Tổng Quan

---

- Gian lận tài chính
- Phát hiện bất thường
- Kỹ thuật học máy

# Gian Lận Tài Chính

---

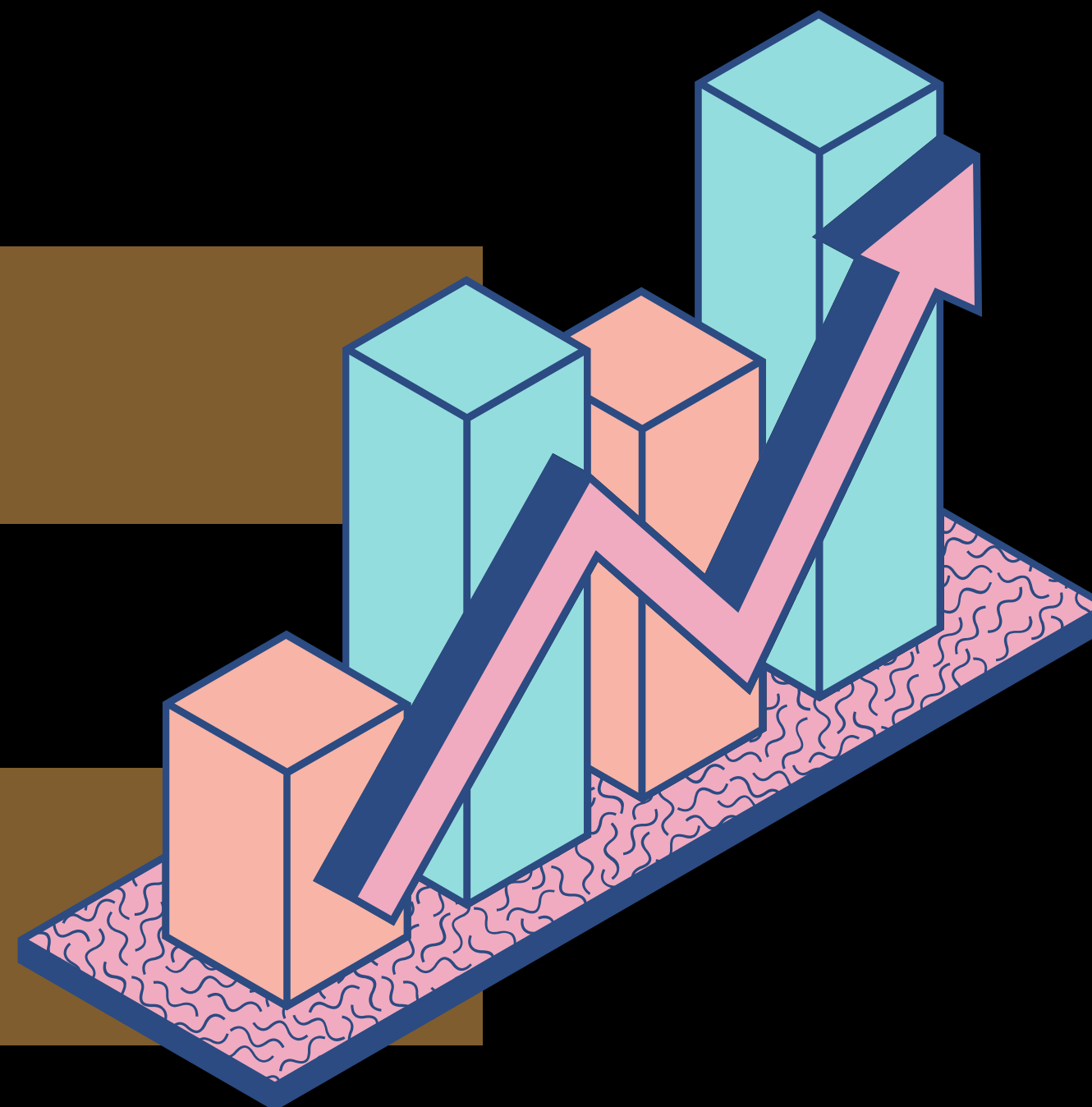
Trang 04 / 16

## ○ Thực trạng Hiện tại

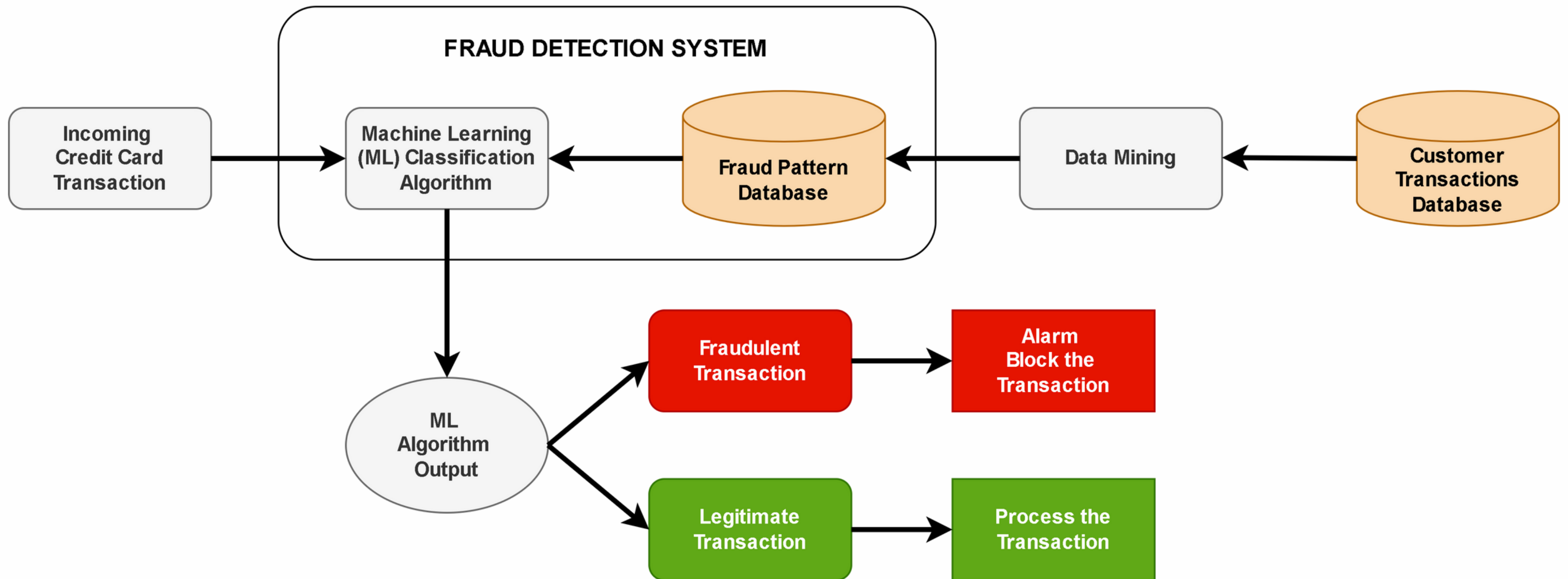
Tổn thất to lớn

## ○ Nguy cơ Tương lai

Mối đe dọa nguy hiểm

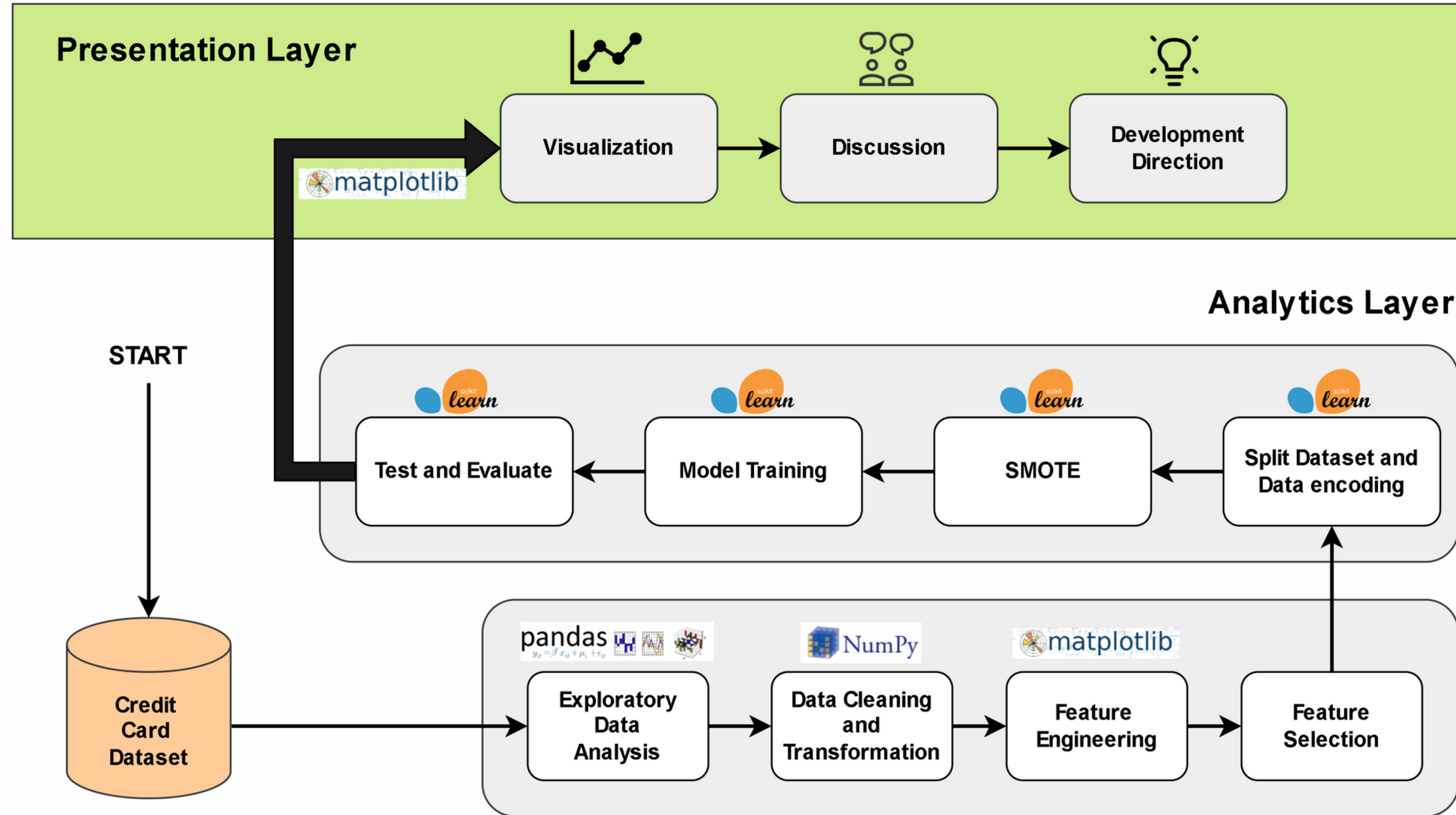


# Hệ Thống Phát Hiện Bất Thường



# Thiết Kế Chi Tiết Của Mô Hình

Trang  
06 / 16



# Công Trình Liên Quan

---

**01**

V. T. Gowda

Credit Card Fraud Detection using  
Supervised and Unsupervised Learning

Wichita State University, USA, 2021.

**03**

D. Varmedja, M. Karanovic, S. Sladojevic, M.

Credit Card Fraud Detection -  
Machine Learning methods

INFOTEH, Republic of Srpska, Bosnia and Herzegovina, 2019

**02**

A. Rosenbaum

Detecting Credit Card Fraud with  
Machine Learning

Stanford University, USA, 2019.

**04**

S. S. B. K. C. Taneja

Application of balancing techniques  
with ensemble approach for credit  
card fraud detection

GUCON, New Delhi, India, 2019.





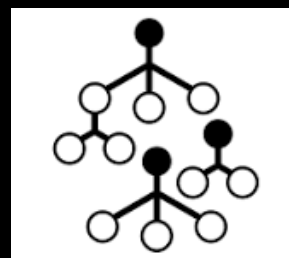
# Định Hướng

- SMOTE
- Thuật toán kết hợp
- Độ đo PR-AUC

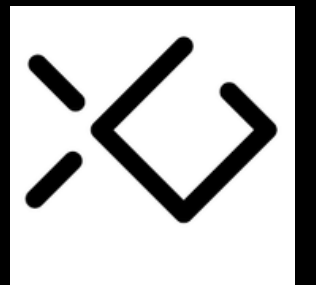
# Phương Pháp Đề Xuất

---

**Random Forest**



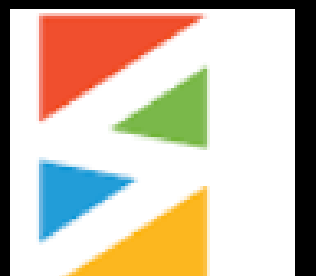
**XGBoost**



**CatBoost**



**LightGBM**



# Kết Quả Thực Nghiệm

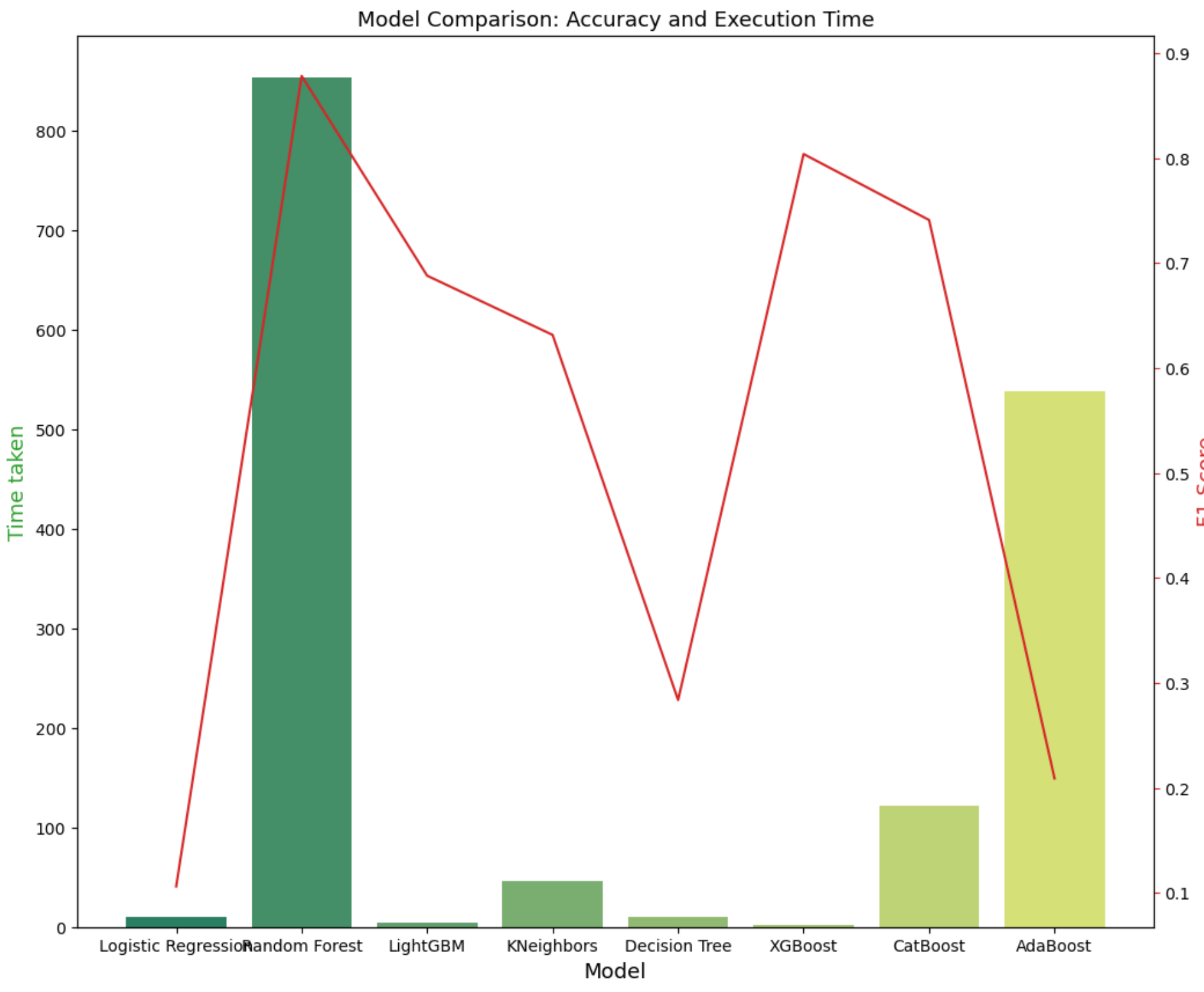
---

- Độ đo các mô hình
- Đánh giá và so sánh
- Mô hình tối ưu nhất

# Độ Đo

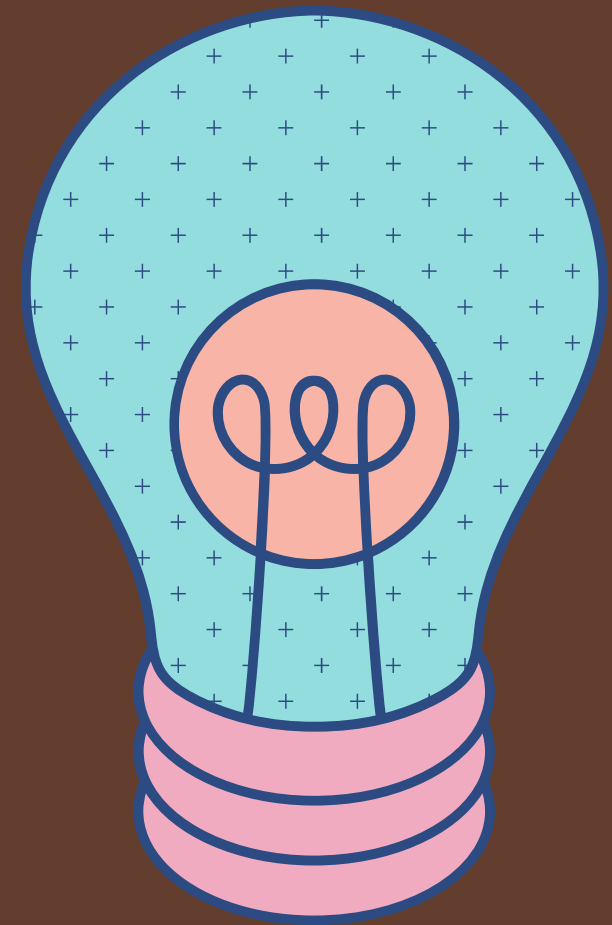
Mô hình	Thời gian	Độ chuẩn xác	Độ phủ	F1-Score	AUC-PR
LightGBM	4.53	0.56	0.87	0.68	0.72
Rừng ngẫu nhiên	853.42	0.91	0.84	0.87	0.87
XGBoost	2.15	0.77	0.83	0.80	0.8
CatBoost	121.83	0.65	0.84	0.74	0.75
AdaBoost	538.67	0.11	0.87	0.20	0.49
Cây quyết định	10.21	0.17	0.81	0.28	0.49
Hồi quy Logistic	10.16	0.05	0.91	0.10	0.48
KNN	45.82	0.5	0.85	0.63	0.6
LightGBM sau tinh chỉnh	4.04	0.63	0.89	0.74	0.76

# Biểu Đồ



# LIGHTGBM

- Thời gian
- Độ phủ
- Tiềm năng



# Kết Luận

---

- Hạn chế
- Hướng phát triển

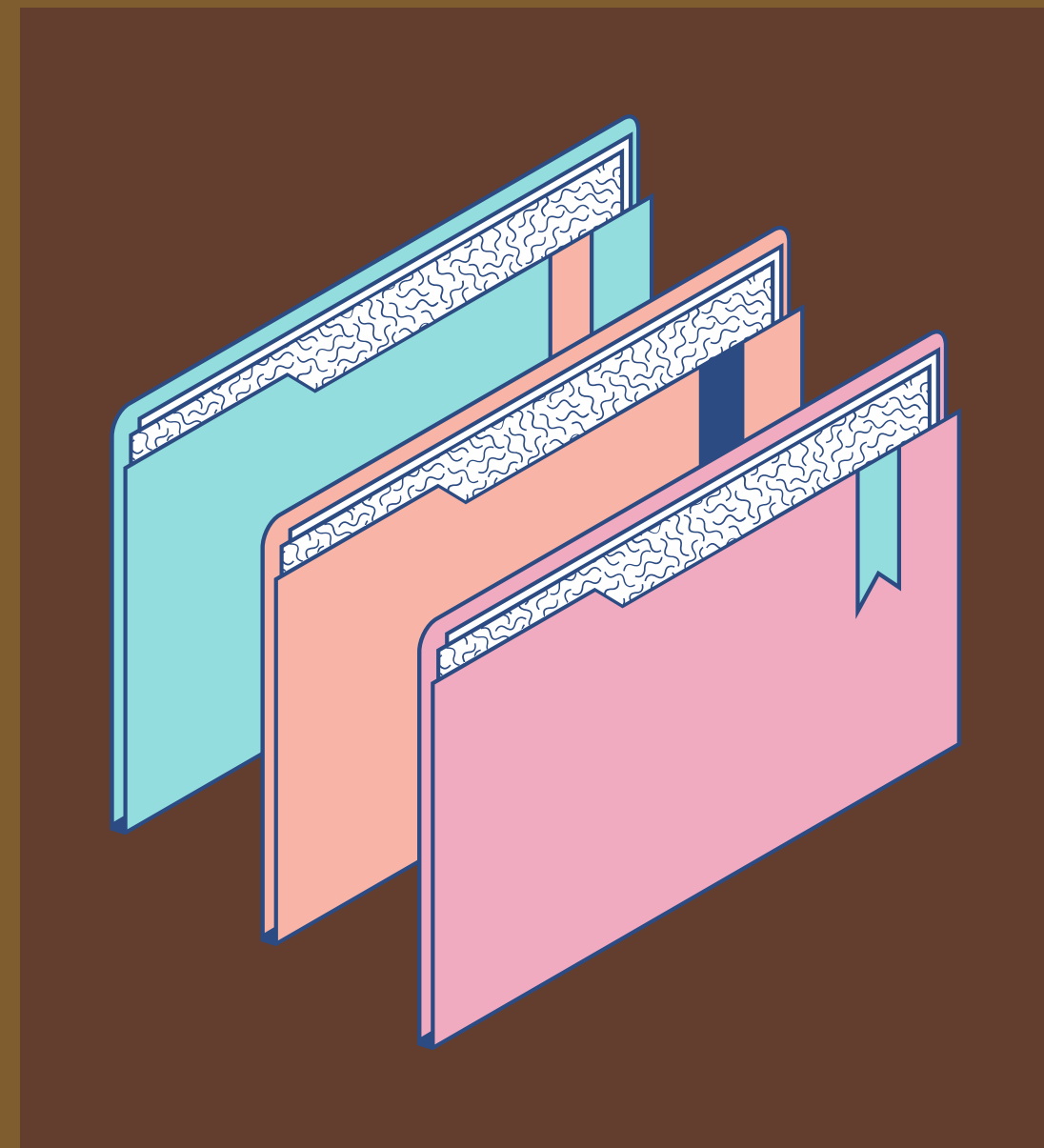
# Hạn chế

## ○ Dữ liệu

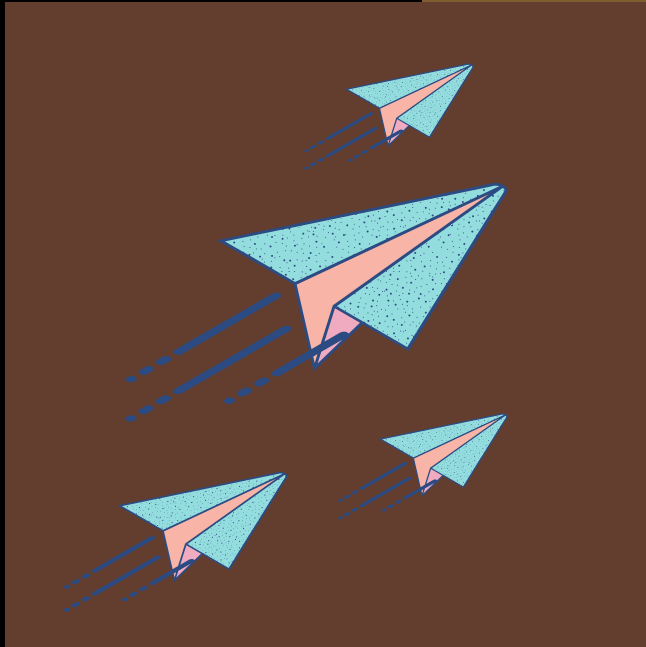
- Mất cân bằng
- Lựa chọn đặc trưng

## ○ Kỹ thuật

- Giới hạn trong sự lựa chọn
- Độ đo chưa cân bằng



# Hướng Phát Triển



- Tìm kiếm lưới, tìm kiếm ngẫu nhiên
- Borderline-SMOTE, ADASYN
- Độ đo MCC, G-Mean
- Học bán giám sát



**Cảm Ơn Thầy Cô và Các Bạn  
Đã Lắng Nghe**