

ĐƠN VỊ TỔ CHỨC



ĐỐI TÁC CHIẾN LƯỢC



CUỘC THI DATA SCIENCE TALENT COMPETITION 2024 VÒNG 3

Đội thi: Vô Gia Cư

Ngày thực hiện: 14 – 17/10/2024



Giới thiệu thành viên

Nguyễn Đức Dũng (leader) - 106042

Nguyễn Văn Hoàng - 106127

Nguyễn Bảo Phong - 106001

Mục lục

1. Giới thiệu
2. Xây dựng giải pháp
3. Đánh giá hiệu suất giải pháp
4. Kết luận về giải pháp
5. Tính ứng dụng thực tế

1. Giới thiệu

Tổng quan

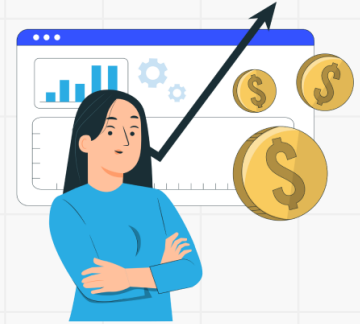
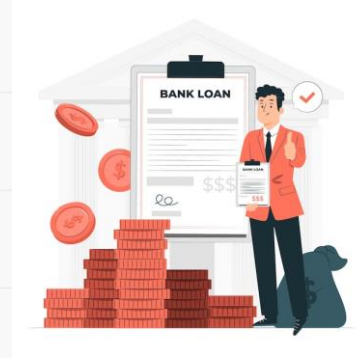
- Trong lĩnh vực tài chính và ngân hàng, **dữ liệu lịch sử tín dụng** đóng vai trò quan trọng trong việc đánh giá khả năng tín dụng của khách hàng.
- Phân tích và xử lý dữ liệu lịch sử tín dụng của khách hàng

Bộ dữ liệu

- ID cá nhân của khách hàng.
- Các biến lịch sử tín dụng.
- **Label:** Label: 1 - khách hàng bị quá hạn (default), 0 - khách hàng trả nợ đúng hạn.

Nền tảng mô hình

- Ngôn ngữ sử dụng: Python.
- Nền tảng sử dụng: Google Colab.



2. Xây dựng giải pháp

SHORT_TERM_COUNT	1	-0.06	0.73	0.2	0.55	0.45	0.58	0.63	0.64	0.02	-0	0.02	0.02	-0.31
MID_TERM_COUNT	-0.06	1	0.63	0.25	0.53	0.12	0.2	0.28	0.35	0.03	0.06	0.1	0.05	-0.15
NUMBER_OF_LOANS	0.73	0.63	1	0.34	0.79	0.43	0.58	0.68	0.74	0.05	0.05	0.1	0.06	-0.35
NUMBER_OF_CREDIT_CARDS	0.2	0.25	0.34	1	0.84	0.1	0.14	0.2	0.25	0.07	0.09	0.09	0.06	-0.2
NUMBER_OF_RELATIONSHIP	0.55	0.53	0.75	0.64	1	0.31	0.45	0.52	0.55	0.07	0.05	0.12	0.07	-0.33
NUM_NEW_LOAN_TAKEN_3M	0.45	0.12	0.43	0.1	0.31	1	0.65	0.57	0.53	0.01	-0	0.02	0.01	-0.15
NUM_NEW_LOAN_TAKEN_6M	0.58	0.2	0.58	0.14	0.43	0.65	1	0.73	0.68	0.02	0.01	0.03	0.01	-0.23
NUM_NEW_LOAN_TAKEN_9M	0.63	0.28	0.68	0.2	0.52	0.57	0.73	1	0.76	0.03	0.02	0.04	0.02	-0.27
NUM_NEW_LOAN_TAKEN_12M	0.64	0.35	0.74	0.25	0.58	0.53	0.68	0.76	1	0.04	0.02	0.04	0.02	-0.29
OUTSTANDING_BAL_ALL_3M_6M	0.02	0.03	0.05	0.07	0.07	0.01	0.02	0.03	0.04	1	0.04	0.01	0.24	-0.03
OUTSTANDING_BAL_ALL_6M_9M	-0	0.06	0.05	0.09	0.09	-0	0.01	0.02	0.02	0.04	1	0.04	0.02	-0.03
OUTSTANDING_BAL_ALL_9M_12M	0.02	0.1	0.1	0.09	0.12	0.02	0.03	0.04	0.04	0.01	0.04	1	0.22	-0.02
OUTSTANDING_BAL_ALL_6M_12M	0.02	0.05	0.06	0.06	0.07	0.01	0.01	0.02	0.02	0.24	0.02	0.22	1	-0.02
label	-0.31	-0.15	-0.35	-0.2	-0.33	-0.15	-0.23	-0.27	-0.29	-0.03	-0.03	-0.02	-0.02	1

13 đặc trưng
được đội thi
lựa chọn

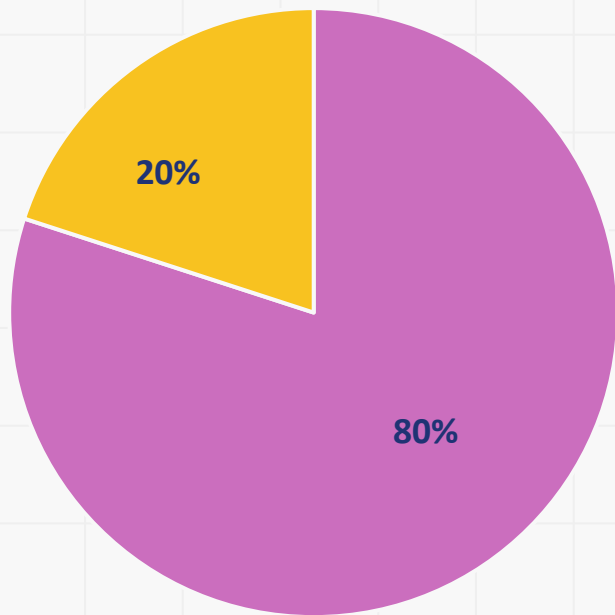
Các đặc trưng
có tương quan
cao

Các đặc trưng tuy
tương quan
không cao nhưng
quan trọng khi
xét trên thực tế

OUTSTANDING_
BAL_ALL_3M_
12M

2. Xây dựng giải pháp

DATASET



■ Train ■ Test

STANDARD SCALER

$$z = \frac{x - \mu}{\sigma}$$

$$\mu = \frac{1}{N} \sum_{i=1}^N (x_i)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$



3. Đánh giá hiệu suất giải pháp

Đánh giá hiệu suất các mô hình

	Model	Accuracy	Recall (macro avg)	Precision (macro avg)	F1-Score (macro avg)
0	Logistic Regression	0.882666	0.683003	0.929449	0.733912
1	Decision Tree	0.812938	0.688137	0.688137	0.688137
2	Random Forest	0.847942	0.706034	0.748730	0.723436
3	XGBoost	0.870624	0.705161	0.817664	0.740854
4	Naive Bayes	0.660599	0.688743	0.615812	0.599270



Đề cao chỉ số đánh giá *Accuracy* và *Recall*



3. Đánh giá hiệu suất giải pháp

Lựa chọn 3 mô hình hiệu quả

	Model	Accuracy	Recall (macro avg)	Precision (macro avg)	F1-Score (macro avg)
0	Logistic Regression	0.882666	0.683003	0.929449	0.733912
1	Decision Tree	0.812938	0.688137	0.688137	0.688137
2	Random Forest	0.847942	0.706034	0.748730	0.723436
3	XGBoost	0.870624	0.705161	0.817664	0.740854
4	Naive Bayes	0.660599	0.688743	0.615812	0.599270

Decision Tree, Random Forest, XGBoost



3. Đánh giá hiệu suất giải pháp

Đánh giá chi tiết hiệu suất mô hình Random Forest

Random Forest Accuracy: 0.8479417530103612

precision recall f1-score support

0	0.89	0.93	0.91	2915
1	0.61	0.48	0.54	656

accuracy			0.85	3571
macro avg	0.75	0.71	0.72	3571
weighted avg	0.84	0.85	0.84	3571



4. Ưu & nhược điểm mô hình

ƯU ĐIỂM

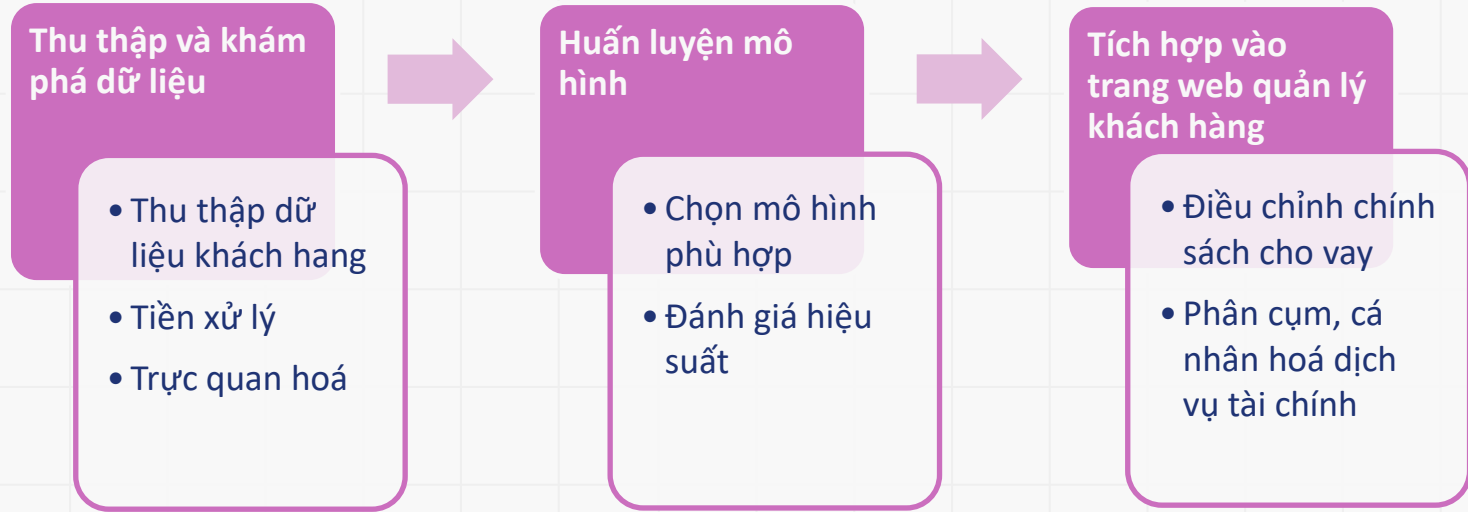
- Độ chính xác tổng thể của mô hình khá cao.
- Dự đoán được triệt để khách hàng trả nợ đúng hạn.
- Đưa ra cảnh báo về khách hàng trả nợ đúng hạn với độ tin cậy cao.

NHƯỢC ĐIỂM

- Khả năng dự đoán khách hàng trả nợ không đúng hạn còn khiêm tốn, cần thu thập thêm dữ liệu khách hàng, từ đó huấn luyện và kiểm thử.



5. Cách triển khai mô hình trong doanh nghiệp



Luôn cập nhật dữ liệu thường xuyên và đánh giá hiệu suất định kì

Thay đổi biến nếu có thể cải thiện hiệu suất mô hình

Thank You!