

PRUDENTIAL **LIFE INSURANCE** **ASSESSEMENT**

Group Happy Tet



OUR TEAM



A42865

Tạ Quang Dũng



A43855

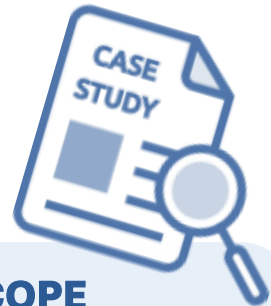
Lê Thu Lan



A42730

Vũ Thị Anh Phương

BUSINESS CASE



BUSINESS OBJECTIVE

- Dự đoán rủi ro bảo hiểm ?
- Giảm thiểu chi phí kinh doanh
- Giảm thiểu thời gian vận hành

STAKEHOLDERS

- Người tư vấn bảo hiểm
- Phòng thẩm định rủi ro
- Cá nhân được bảo hiểm

SCOPE

- Train: 59381 hồ sơ
- Test: 19765 hồ sơ

MODELS

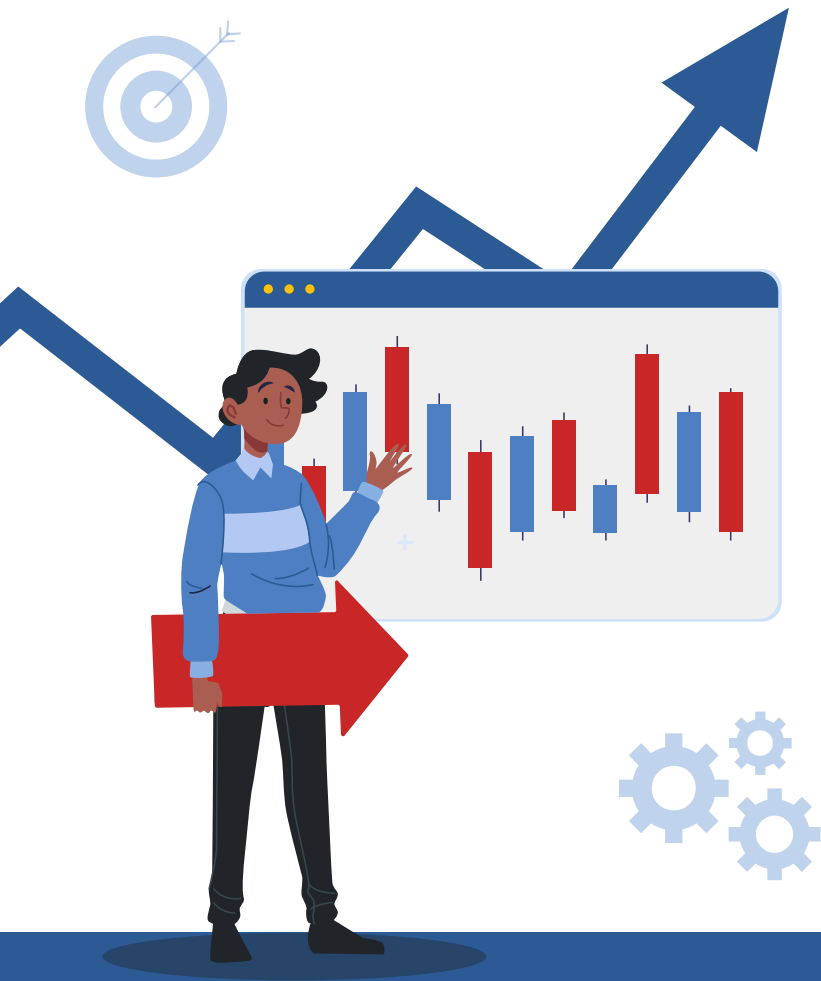
- Xây dựng mô hình
- Đánh giá và lựa chọn mô hình tốt nhất

ASSUMPTIONS

- Bộ dữ liệu đáng tin cậy
- Thời gian dán giá cho mỗi hồ sơ là bằng nhau

RESULT

- **Accuracy: 51 %**
- **F1-Score: 49 %**



INTRODUCTION



Prudential

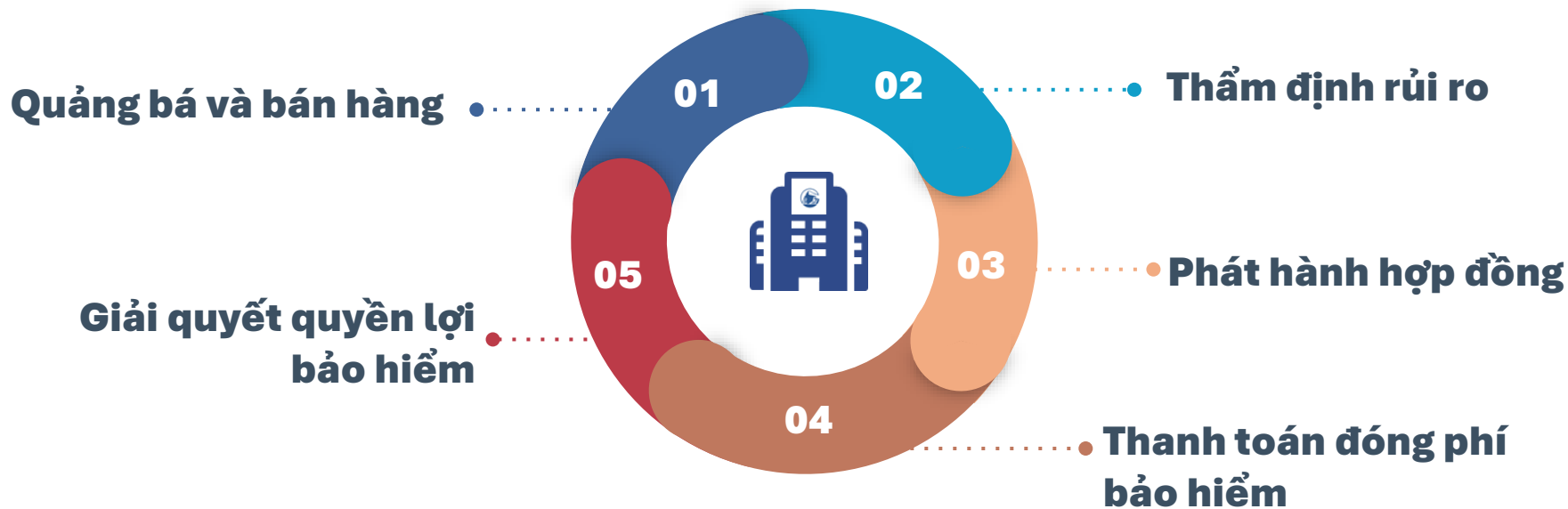
- Prudential là một trong những tập đoàn tài chính hàng đầu thế giới cung cấp các giải pháp bảo hiểm nhân thọ, sức khỏe và quản lý tài sản lớn nhất của Hoa Kỳ thành lập vào năm 1848 tại Luân Đôn.
- Prudential hiện có mặt tại hơn 24 thị trường trên khắp Châu Á và Châu Phi, phục vụ hơn 20 triệu khách hàng



- Việt Nam
- Thái Lan
- Indonesia
- Kenya
- Nam Phi

**Hình ảnh chỉ mang tính chất minh họa*

Prudential's Insurance Sales Process



DATA MINING GOALS



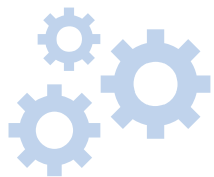
Đánh giá mức độ rủi ro của khách hàng mua bảo hiểm



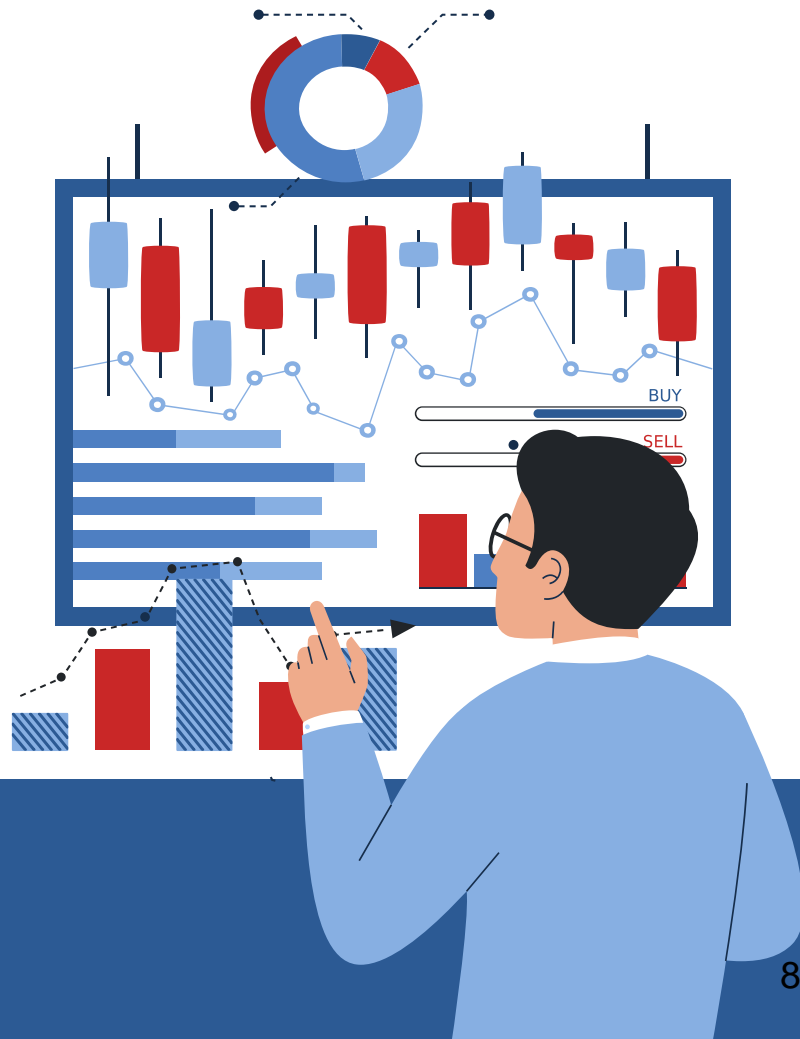
Hồ sơ khách hàng



Accuracy: 51%
F1-Score: 49%



DATA PREPARATION



DATASET OVERVIEW



Train: có 59381 dòng và 128 cột

Test : có 19765 dòng và 127 cột

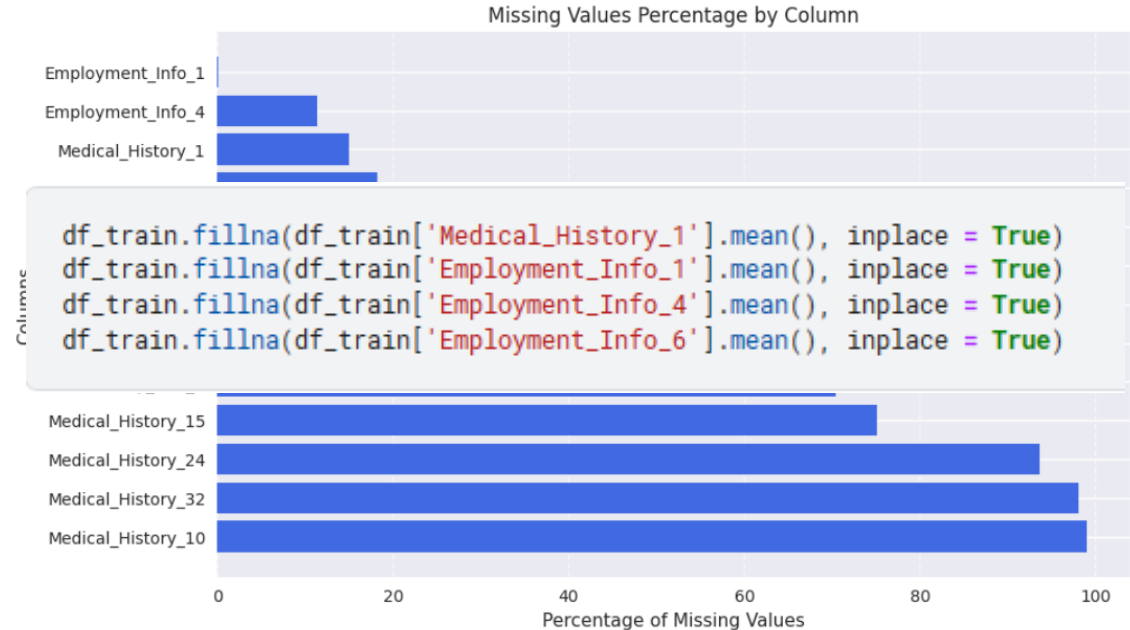
Mã định danh khách hàng	ID
Thông tin cá nhân	Family_Hist_1-5, InsuredInfo_1-6, Ins_Age
Chỉ số cơ thể	Ht, Wt, BMI
Lịch sử khám chữa bệnh	Medical_History_1-41, Medical_Keyword_1-48, Insurance_History_1-9
Lịch sử công việc	Employment_Info_1-6
Loại bảo hiểm	Product_Info_1-7
Biến mục tiêu	Responses



DATA PROCESSING

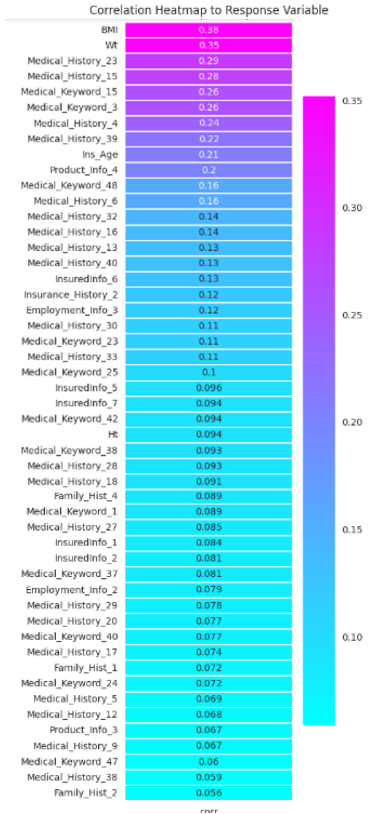
**Loại bỏ cột có
dữ liệu trống (>30%)**

Điền dữ liệu trống (mean)



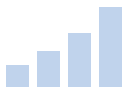


FEATURE SELECTION



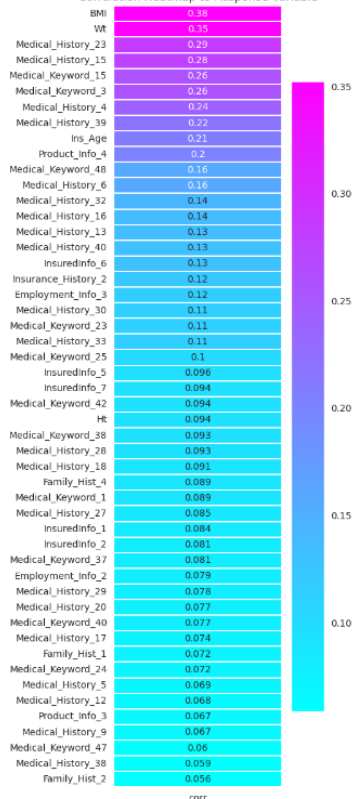
Dựa vào tương quan của
các biến với Response



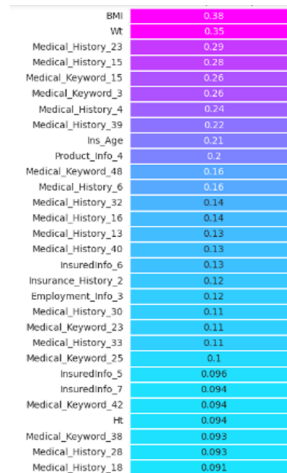


FEATURE SELECTION

Correlation Heatmap to Response Variable



Dựa vào tương quan của
các biến với Response





MODEL SELECTION

MODEL SELECTION

KNN

Random
Forest

Logistic
Regression

MLP

Khả năng giải thích



Phù hợp với bài toán



Hiệu suất



Chi phí



MODEL SELECTION

KNN

**Random
Forest**

Logistic
Regression

MLP

Khả năng giải thích



Phù hợp với bài toán

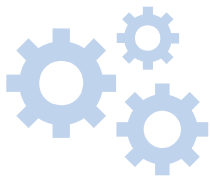


Hiệu suất

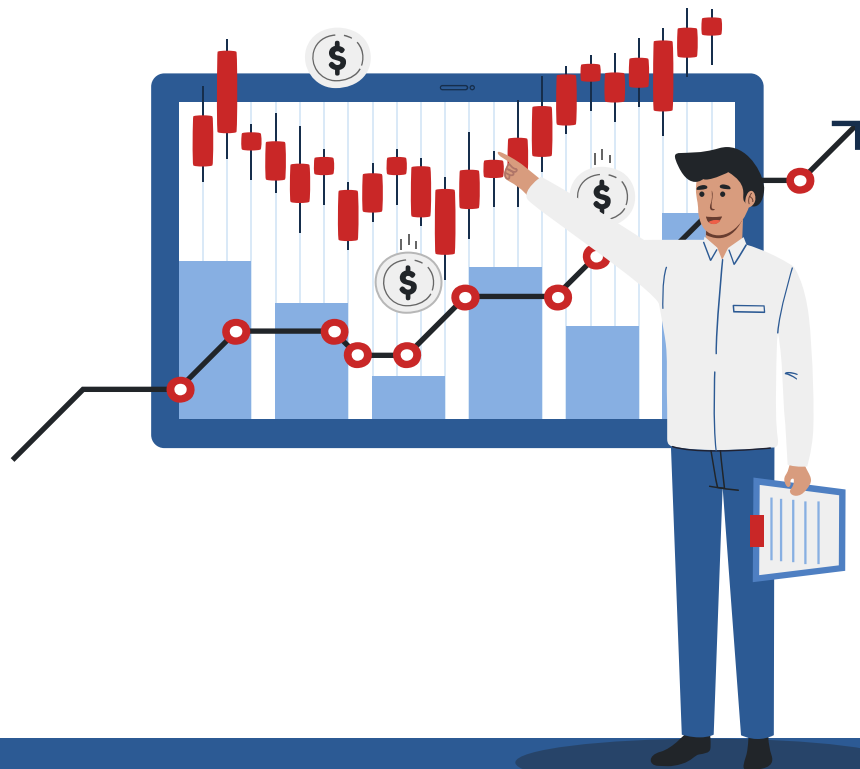


Chi phí

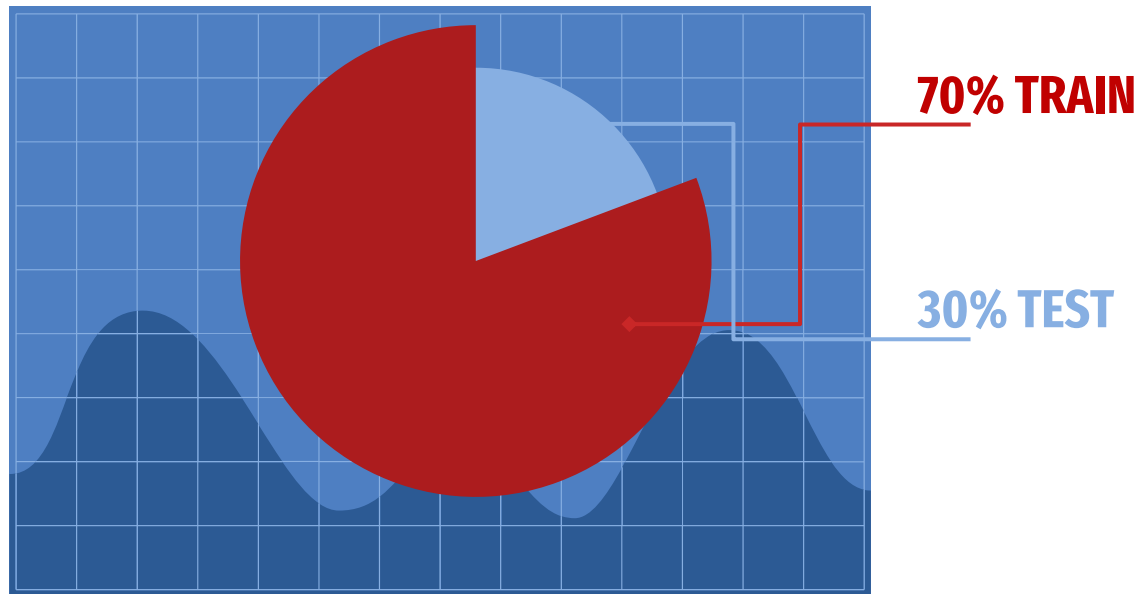




BUILD MODEL

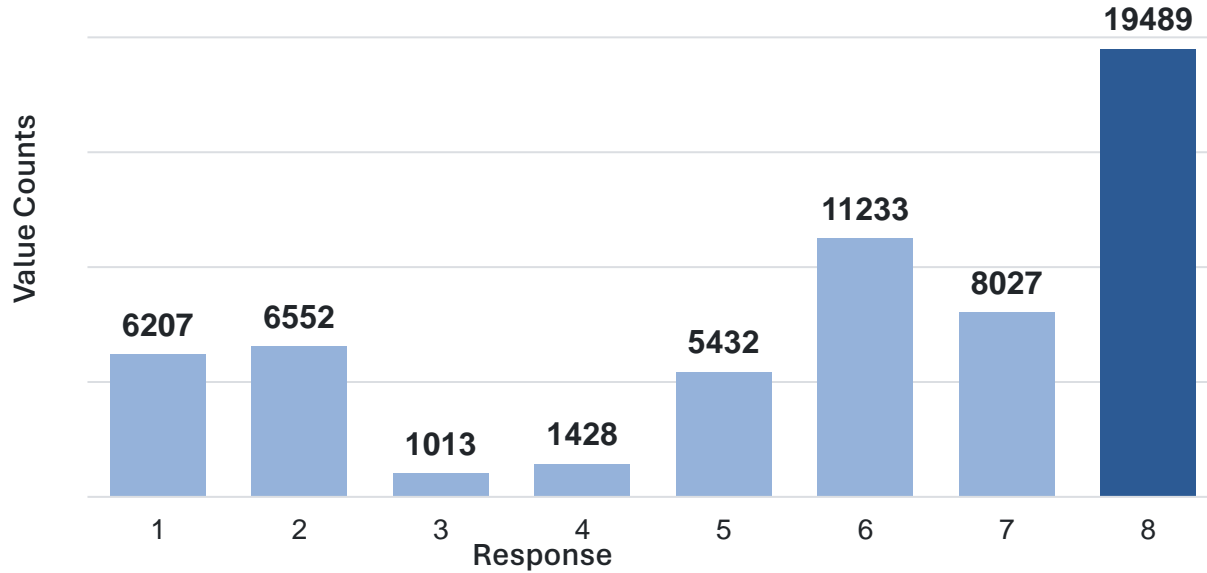


SPLIT DATA

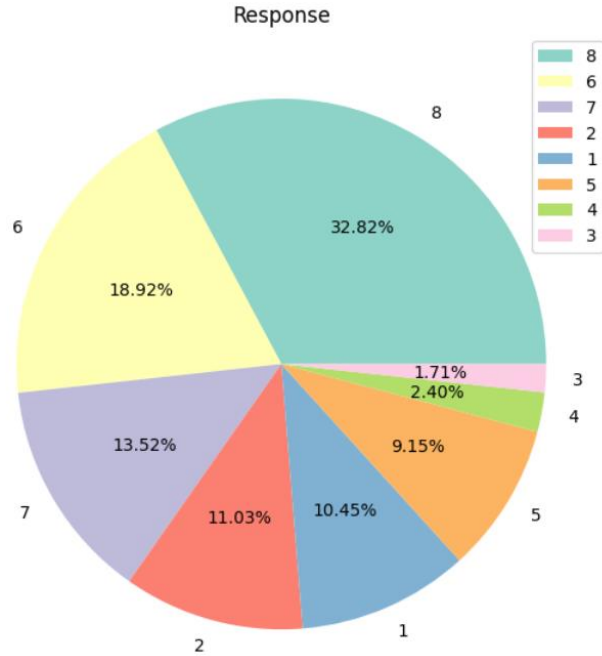


DATA IMBALANCE

Response Number Distribution Chart

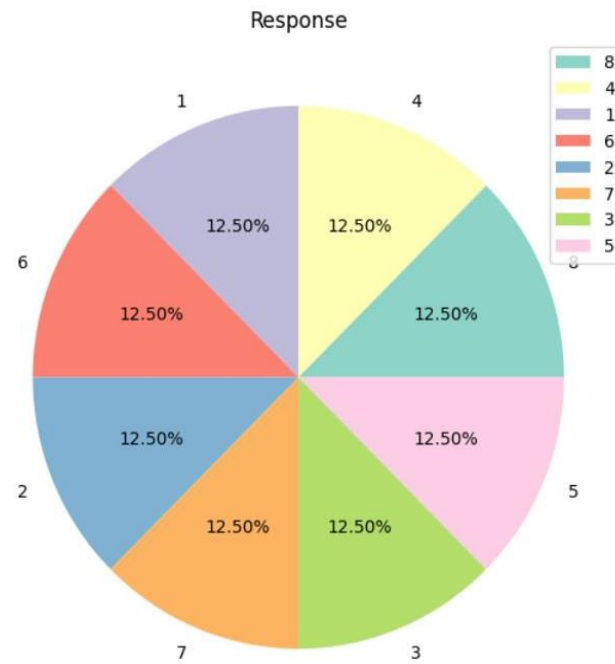
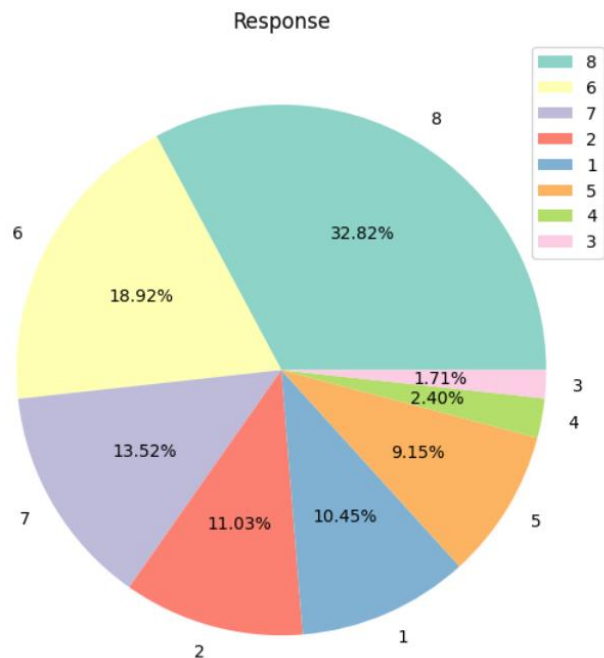


DATA IMBALANCE

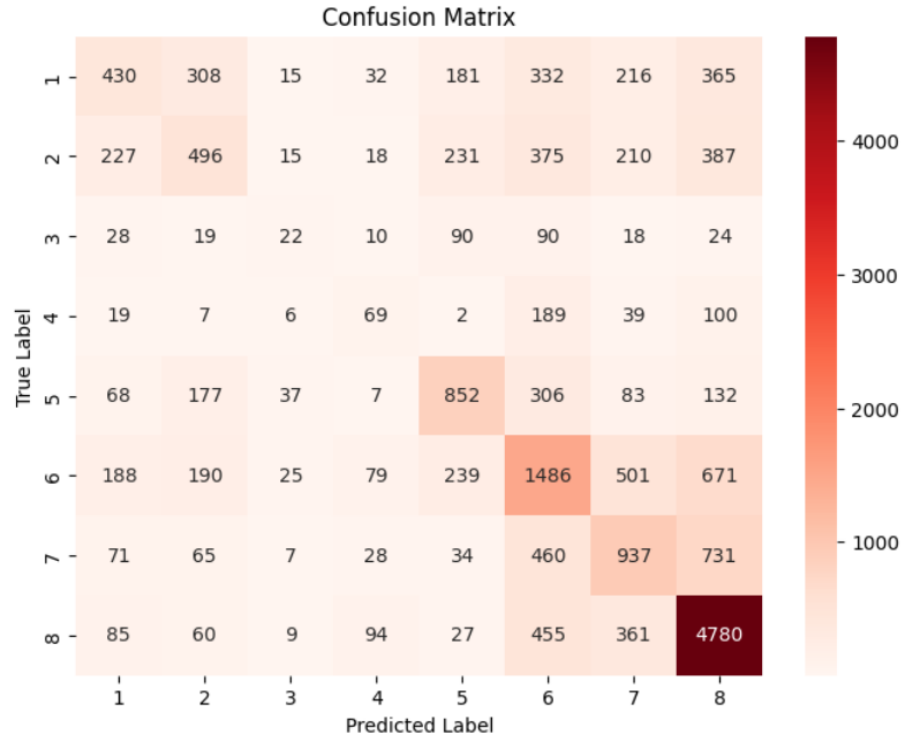


Dùng RandomOverSampler

Dùng RandomOverSampler



MODEL EVALUATION



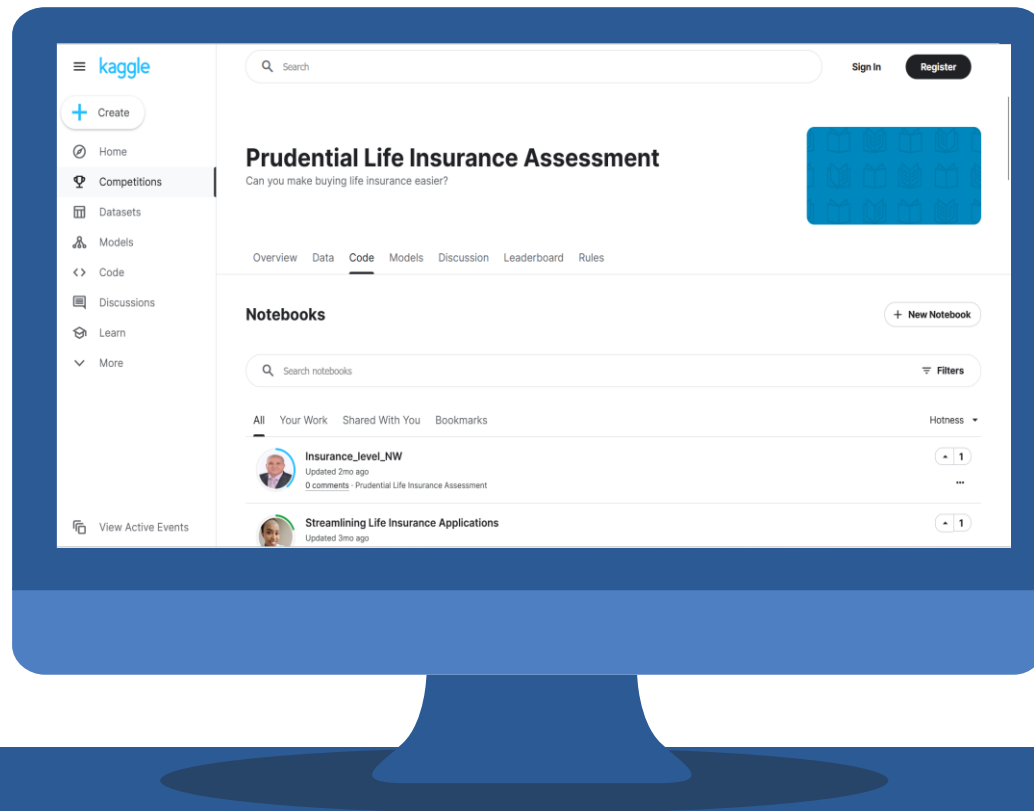
MODEL EVALUATION



	precision	recall	f1-score	support
1	0.39	0.23	0.29	1879
2	0.38	0.25	0.30	1959
3	0.16	0.07	0.10	301
4	0.20	0.16	0.18	431
5	0.51	0.51	0.51	1662
6	0.40	0.44	0.42	3379
7	0.40	0.40	0.40	2333
8	0.66	0.81	0.73	5871
accuracy			0.51	17815
macro avg	0.39	0.36	0.37	17815
weighted avg	0.48	0.51	0.49	17815

References

<https://www.kaggle.com/competitions/prudential-life-insurance-assessment/code>





**THANK
YOU!**

