

# Cross Sell Prediction

Mi AI

# WARNING

Mọi thông tin trong bài chia sẻ là quan điểm cá nhân của một người không chuyên về AI, DA, DE, DS...

Có nhiều cách giải quyết khác cho bài toán này và đây chỉ là 01 trong N cách đó.

Bài toán không dùng thuật toán mới, chỉ là một góc nhìn mới về áp dụng vào đời sống.

Xin cảm ơn!

# Problem Statement



# Problem Statement



Car Insurance



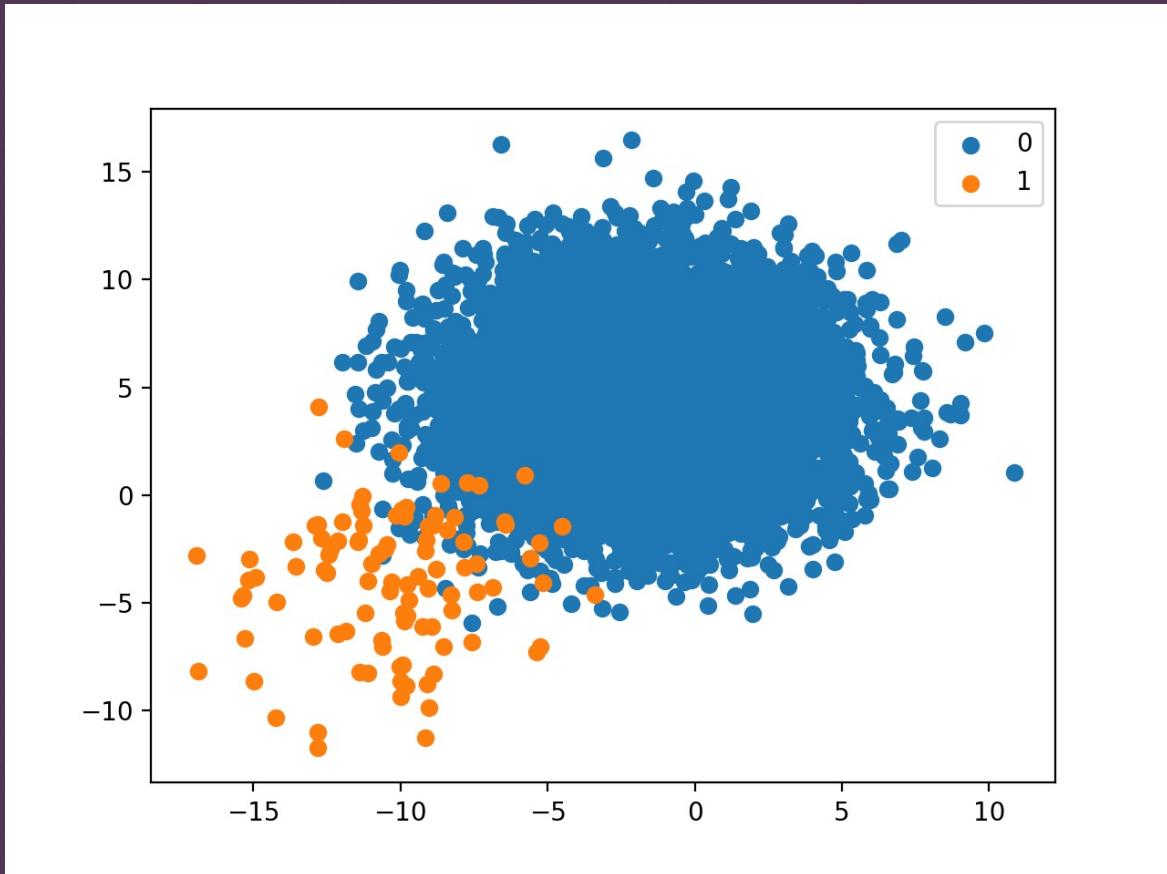
Health Insurance

# Input & Output Data

```
train = pd.read_csv("train.csv")
train.head()
```

	<b>id</b>	<b>Gender</b>	<b>Age</b>	<b>Driving_License</b>	<b>Region_Code</b>	<b>Previously_Insured</b>	<b>Vehicle_Age</b>	<b>Vehicle_Damage</b>	<b>Annual_Premium</b>	<b>Policy_Sales_Channel</b>	<b>Vintage</b>	<b>Response</b>
<b>0</b>	1	Male	44	1	28.0	0	> 2 Years	Yes	40454.0	26.0	217	1
<b>1</b>	2	Male	76	1	3.0	0	1-2 Year	No	33536.0	26.0	183	0
<b>2</b>	3	Male	47	1	28.0	0	> 2 Years	Yes	38294.0	26.0	27	1
<b>3</b>	4	Male	21	1	11.0	1	< 1 Year	No	28619.0	152.0	203	0
<b>4</b>	5	Female	29	1	41.0	1	< 1 Year	No	27496.0	152.0	39	0

# Difficulty



Imbalanced Data

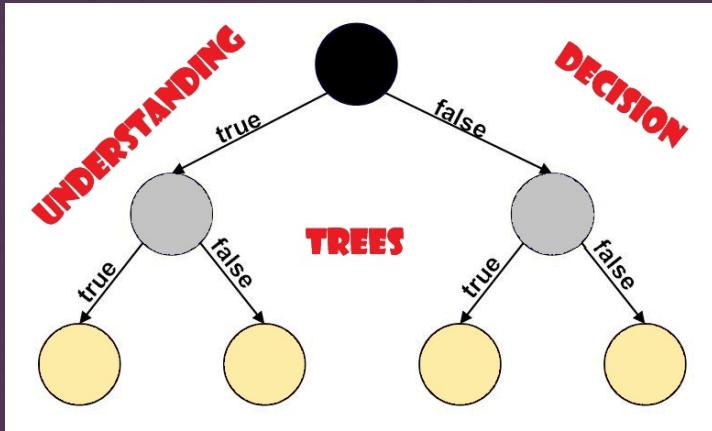
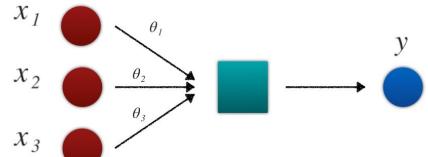
# Experiments

- Thử nhiều loại thuật toán
- Thử nhiều kiến trúc model
- Thử nhiều lựa chọn:
  - Có drop các cột có Correlation lớn không?
  - Correlation như nào là lớn?
  - Size của input, output như nào cho hợp lý...
- Thử nhiều hàm loss khác nhau

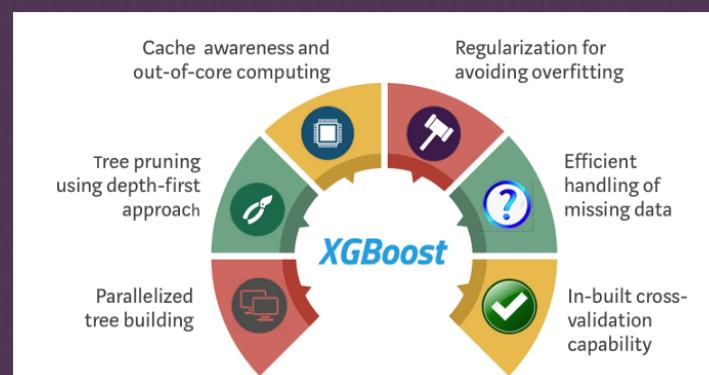
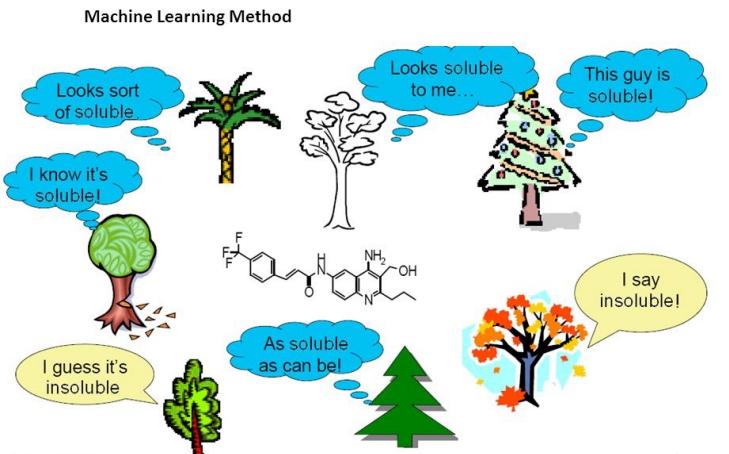
# Experiments

# Methods

## Logistic regression model



## Random Forest



# Hands-on