



# **Target Sales Campaign**

---

## **Predicting Potential Subscribing Customers**

---

**Presented by: Yishu Dai**



# Background

- Nessus Vitamin (a made-up company) is a company that sells vitamins online.
- One-time purchase of the vitamins (no discount) or subscribe to the monthly delivery plan (little bit discount).
- Not many customers are subscribing to the service
- The company wants the data analytics team to build a classification model to identify which customers are more likely to subscribe based on the current clients' information.

# Goal

**Build a classification model to predict if a customer will subscribe to the monthly service**



# Table of Content

**DATA & Tools**



**DESIGN**



**METHODOL  
OGY**



**RESULT**



**FUTURE  
WORK**



**APPENDIX**

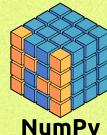


# DATA & TOOLS

kaggle

## Data Source

A nebulous data from Kaggle (41188 rows, 17 columns)



## Data Manipulation

Pandas, NumPy



## Visualization

Tableau, Matplotlib, Seaborn

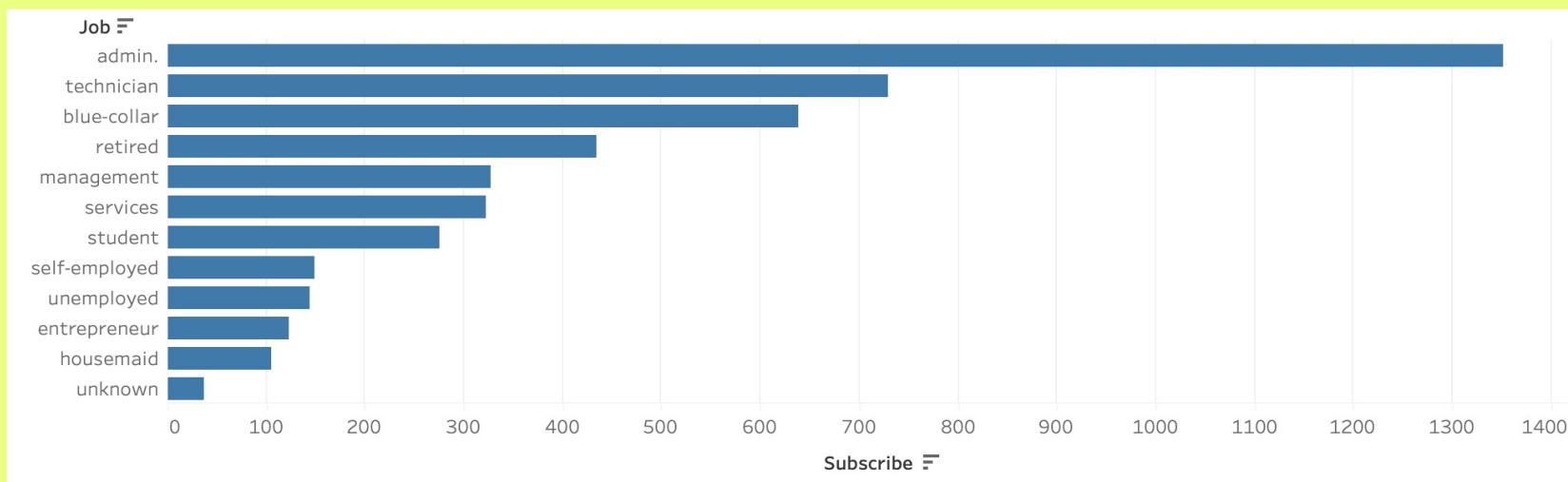


## Modeling

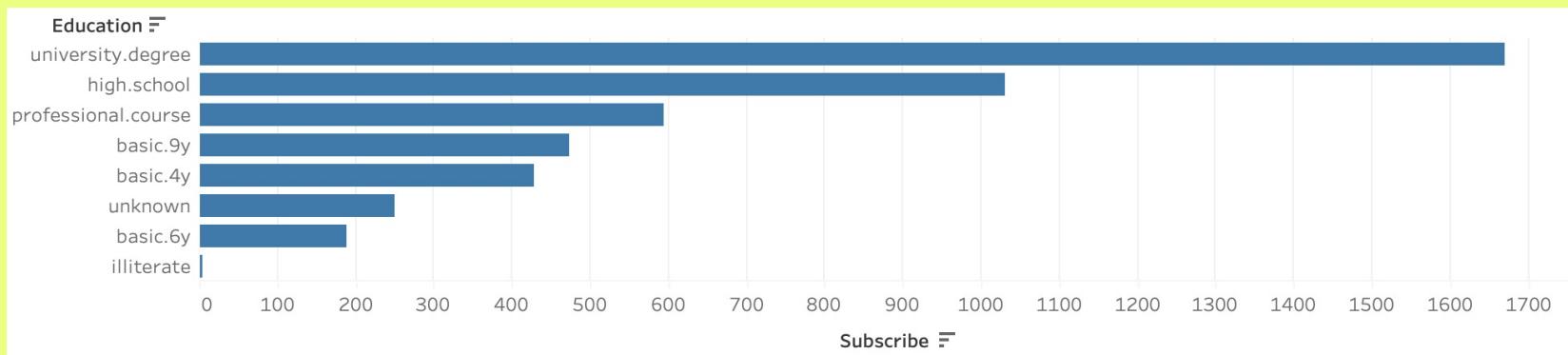
Scikitlearn



## Job and Subscription



## Education and Subscription



## Marriage and Housing Loan



# Design & Methodology



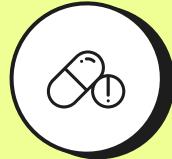
## Feature Engineering

- SMOTE: solve imbalanced data
- Build baseline model, iterate the process by adding features



## Model Selection

- Cross-validation: Find best scored models
- GridSearch: Find hyperparameters



## Model Validation

- ROC/AUC curve: model comparison
- Recall: prediction performance

# Methodology

## Class Imbalance

90%

10%



90%

10%

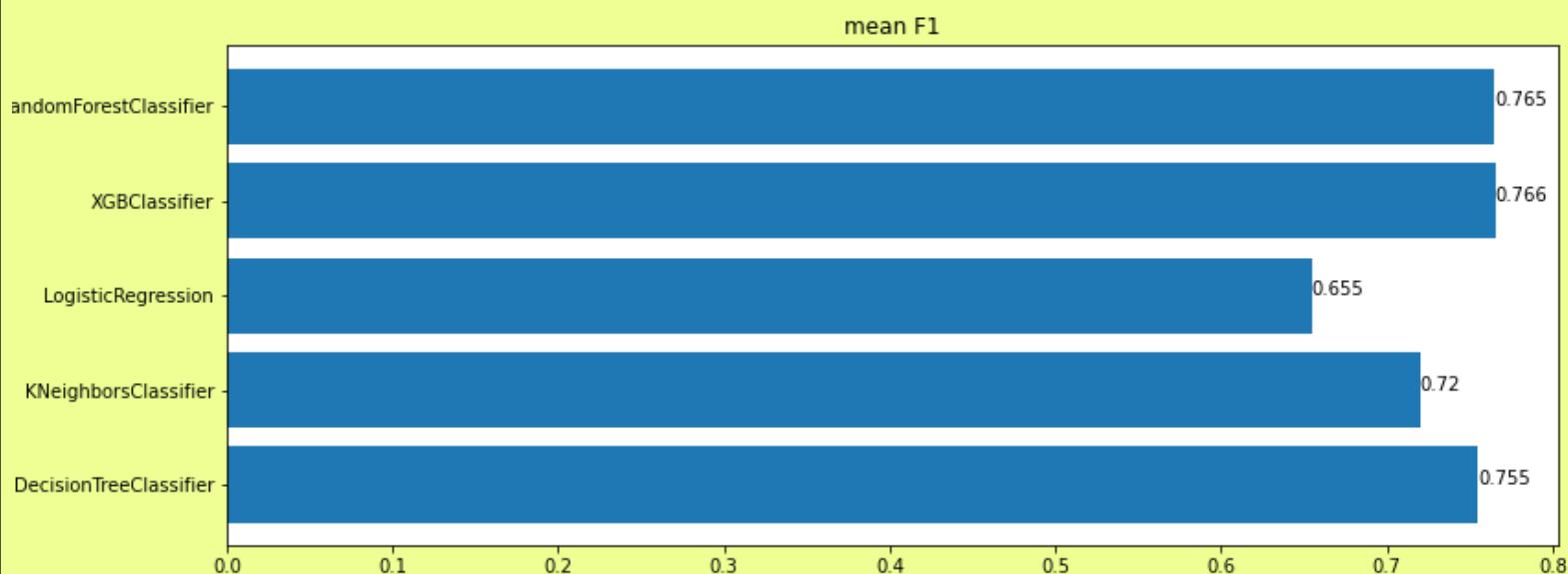


## Categorical → Numerical

	age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration
0	56	1	2	1	0	0	0	1	5	1	261
1	57	2	2	2	1	0	0	1	5	1	149
2	37	2	2	2	0	2	0	1	5	1	226
3	40	3	2	1	0	0	0	1	5	1	151
4	56	2	2	2	0	0	2	1	5	1	307



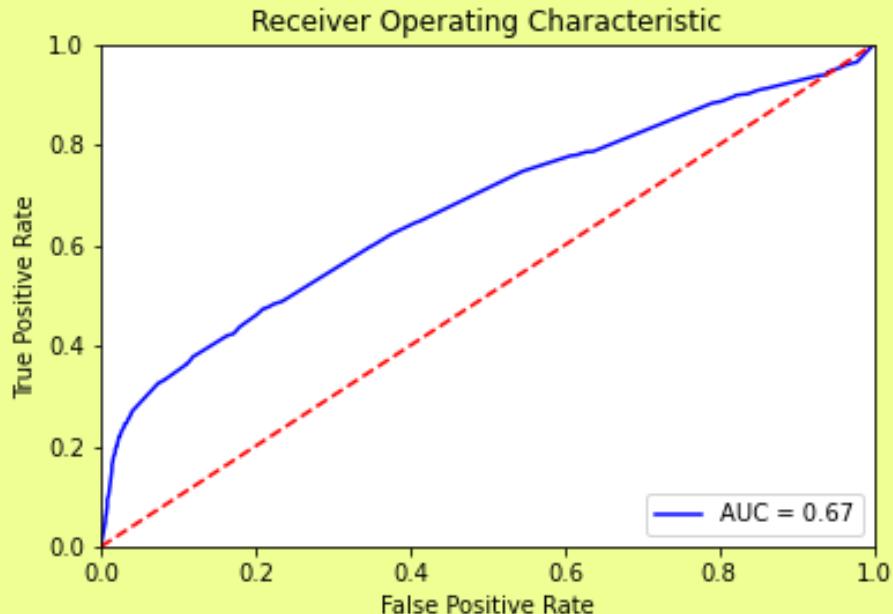
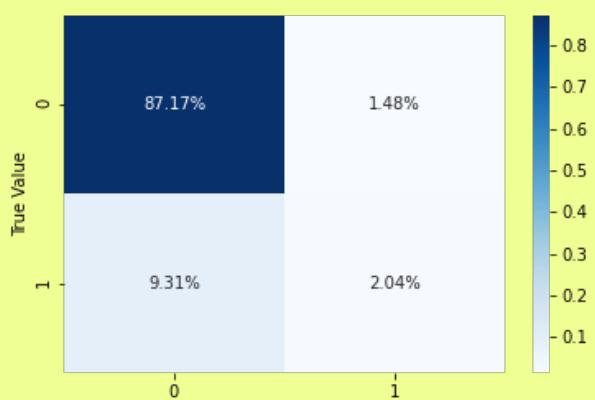
# Cross-Validation



# GridSearch: Hyperparameters

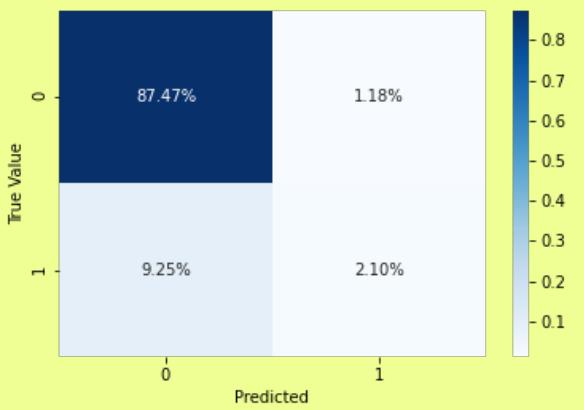
	Decision Tree	Random Forest	XGB
best score	<b>0.779</b>	<b>0.789</b>	<b>0.788</b>
max_depth	<b>10</b>	<b>15</b>	<b>9</b>
min_sample_leaf/ min_child_weight	<b>1</b>	<b>1</b>	<b>1</b>

# Decision Tree

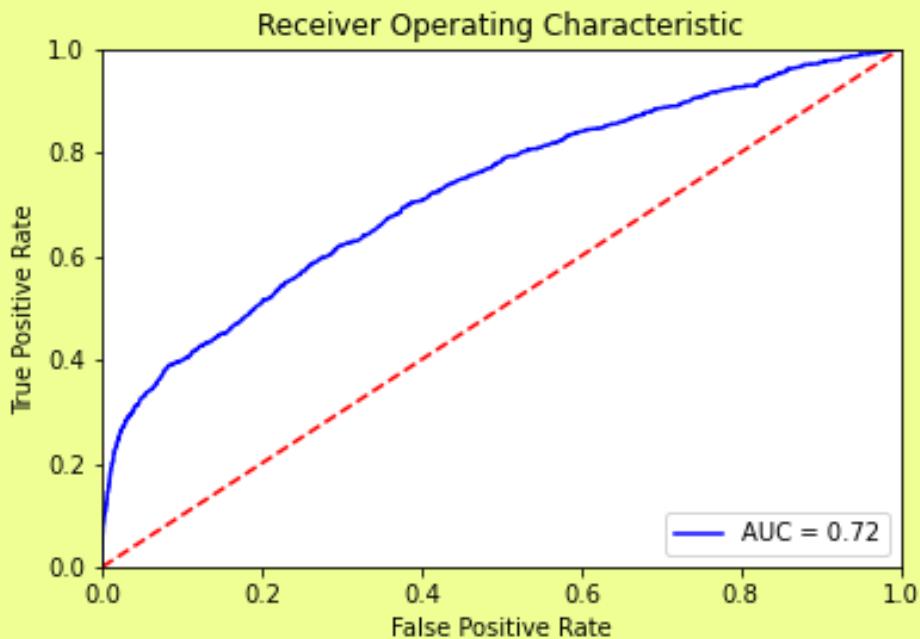


	Precision	recall	f1-score
0	0.90	0.98	0.94
1	0.58	0.18	0.27

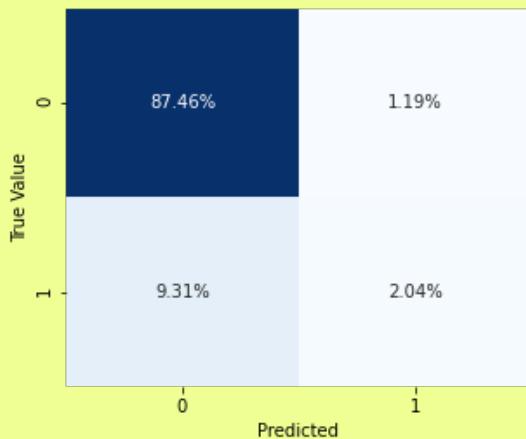
# XGBClassifier



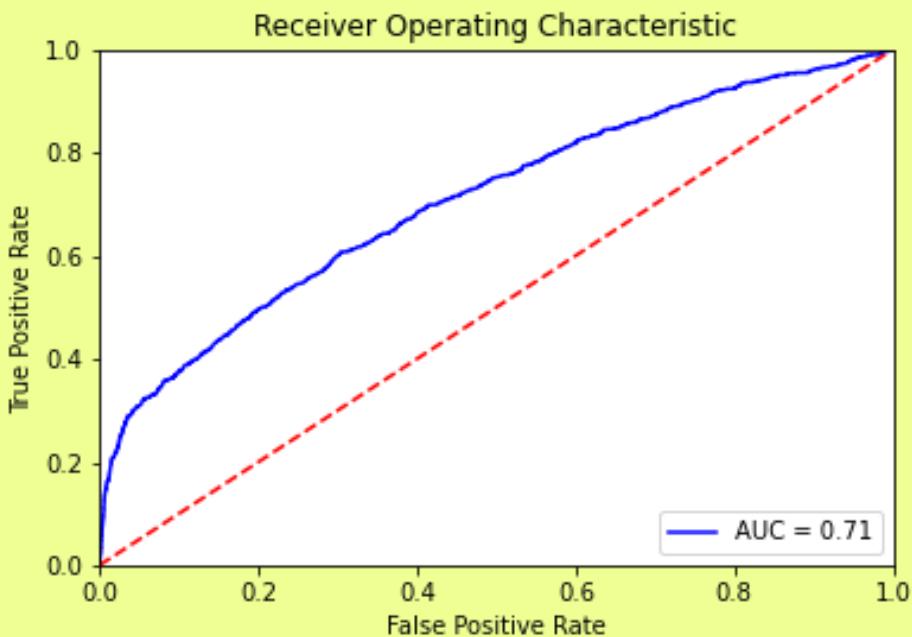
	Precision	recall	f1-score
0	0.90	0.99	0.94
1	0.64	0.19	0.29



# GridSearch: Hyperparameters

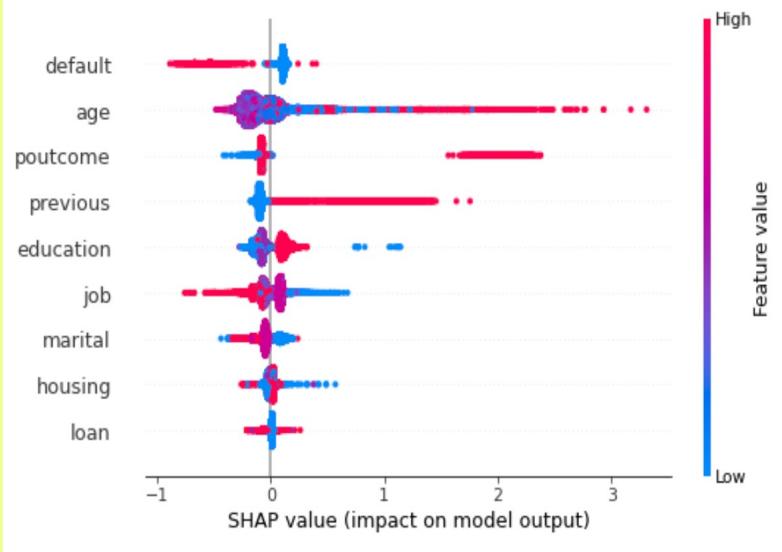


	Precision	recall	f1-score
0	0.90	0.99	0.94
1	0.63	0.18	0.28



# Results

- Demographic features don't have a huge impact on the sales outcome, except age.
- Sales techniques like cold call, active outreach might be most helpful for increasing subscriptions.



# Results Analysis



## Precision

3 out 5 of predicted sales can convert



## Recall

Overall subscription conversion rate 1:5



All predicted potential subscribers should be contacted by phone. But in general, customers who have credit in their account are more likely to be converted.

# Future Work

## Tailored dataset

This dataset originally is not for vitamin online sales, true dataset might serve better

## More Feature Engineering

Find demographic features that are more relevant

## Manual Tuning

Manual XGBClassifier tuning might increase the model performance

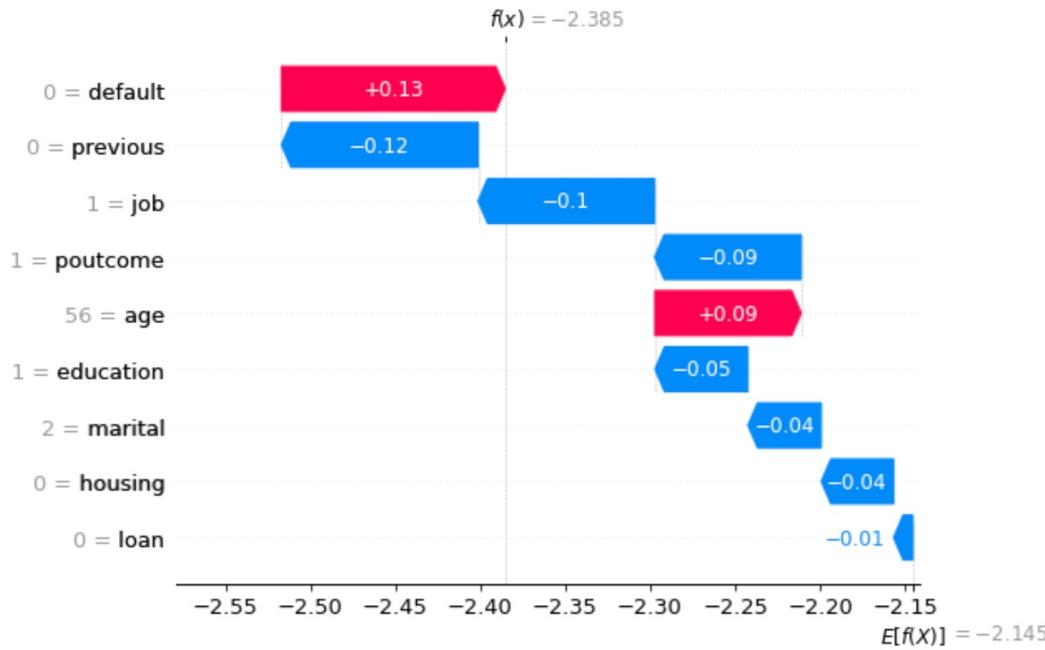
## Predict Customer Churn

Same method can be used to prevent customer churn.

# Appendix - Encoding

```
cleanup = {"job": {"housemaid":1, "services":2, "admin.":3, "blue-collar":4, "technician":2,
"retired":3, "management":4, "unemployed":0, "self-employed":3, "unknown":1,
"entrepreneur":4, "student":2},
"marital": {"married":2,"divorced":3,"single":0,"unknown":1},
"education": {"basic.4y":1, "high.school":2, "basic.6y":1, "basic.9y":1,
"professional.course":3, "unknown":1, "university.degree":3,
"illiterate":0},
"default": {"no":0, "unknown":1, "yes":2},
"housing": {"no":0, "unknown":1, "yes":2},
"loan": {"no":0, "unknown":1, "yes":2},
"contact": {"telephone":1, "cellular":0},
"poutcome": {"nonexistent":1, "failure":0, "success":2},
"month": {"may":5, "jun":6, "jul":7, "aug":8, "oct":10, "nov":11, "dec":12, "mar":3, "apr":4,
"sep":9},
"day_of_week": {"mon":1, "tue":2, "wed":3, "thu":4, "fri":5}
}
```

# Appendix - Feature Importance



# THANKS!

**DO YOU HAVE ANY QUESTIONS?**

**CREDITS:** This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik

Please keep this slide for attribution