

# Predicting Successful Bank Campaign

Anastasia Kharina  
Sisay Filate

October 2019



# Objective:

Will a customer subscribe to a  
term deposit offered via phone  
calls?

# Data & Context

**41,176 Observation**

**14 features**

Customer Data:

- Job
- Marital Status
- Education
- Age
- Whether they have loans (mortgage, etc)

Call Data:

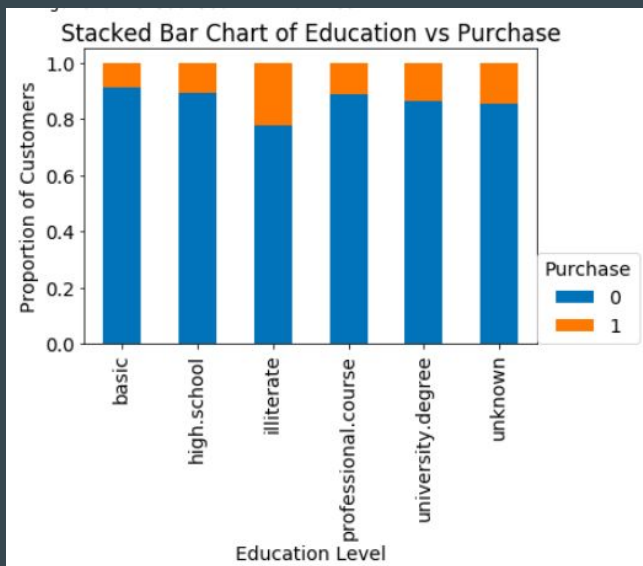
- Day of the Week
- Month of the Year
- Days after last call
- Number of calls for campaign

S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

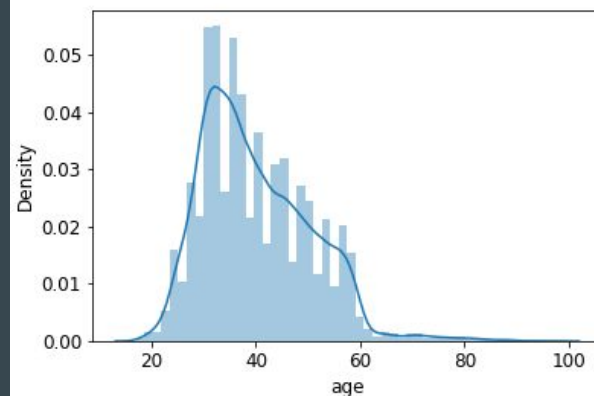
(UCI Machine Learning Repository - Bank Marketing Data Set)

# Data Exploration

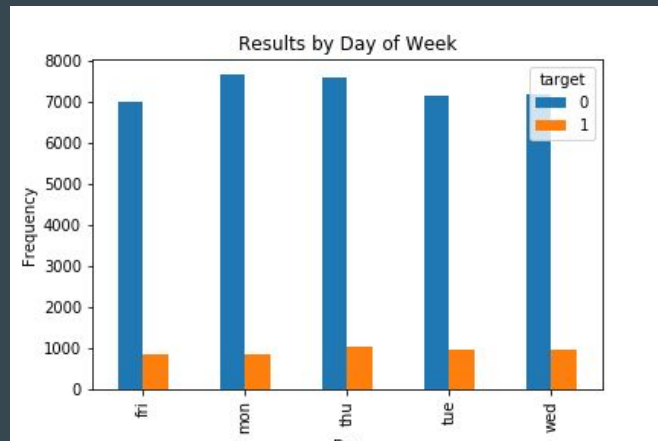
## Subscription by education level



## Age distribution

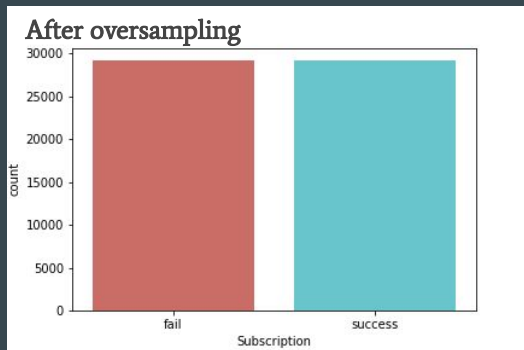
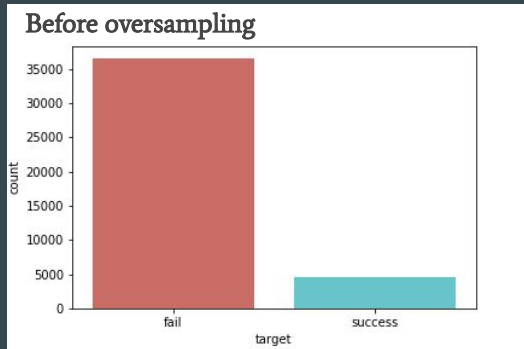


## Subscription by Day of the Week



# Logistic Regression Process

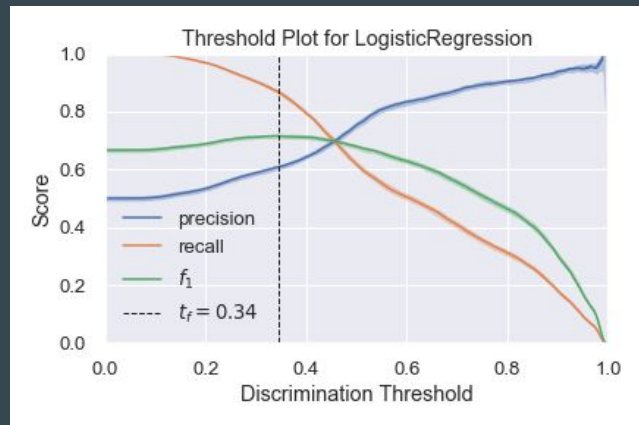
## 1. Synthetic Minority Oversampling Technique (SMOTE)



## 2. Choosing Parameters with K-Fold CV



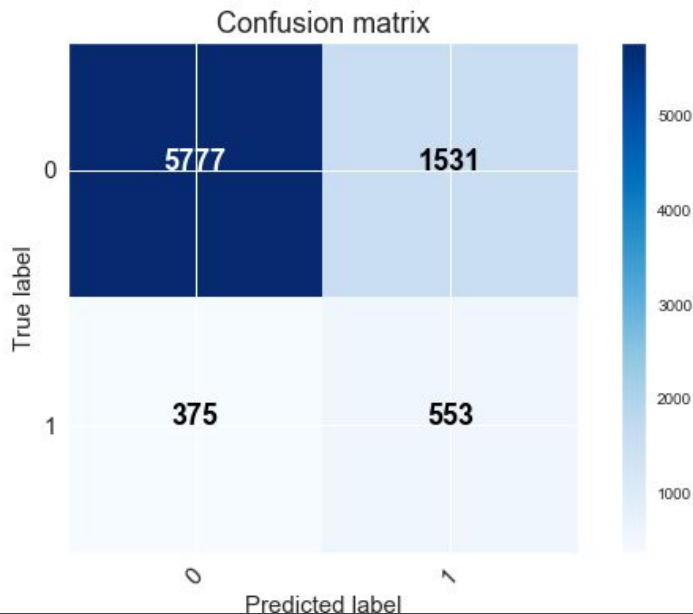
## 3. Optimizing Threshold



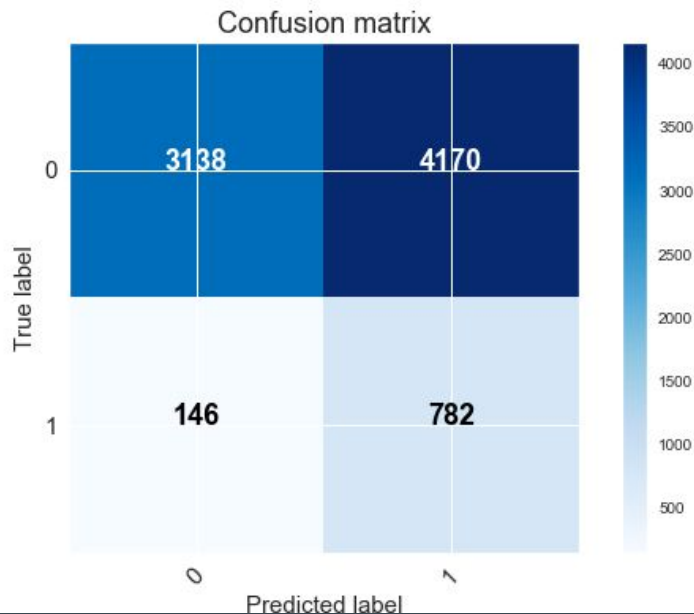
# Modeling Results -1

	Threshold = 0.5	Threshold = 0.34
Score	0.769	0.769
Precision	0.265	0.158
Recall	0.596	0.843
Accuracy	0.769	0.476
F1-Score	0.367	0.266

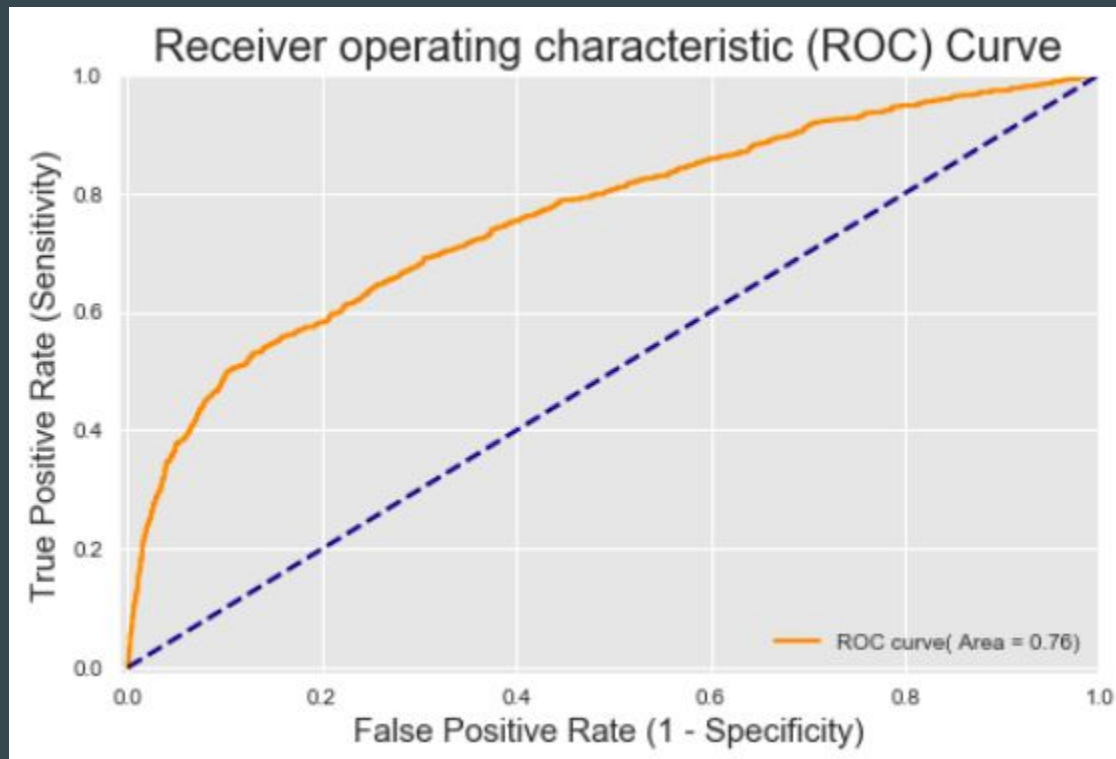
Threshold = 0.5



Threshold = 0.34



# Modeling Result -2 : ROC Curve



# Future Work

1. Investigate driving predictors
2. Try other prediction models (e.g. KNN, Random Forest, etc)



**Additional Slides**

# Metric Scores

	Not-oversampled	Oversampled (SMOTE)	Undersampled	RFE-data
Score	0.803	0.708	0.701	0.694
Precision	0.302	0.754	0.767	0.745
Recall	0.572	0.617	0.578	0.591
Accuracy	0.302	0.754	0.767	0.745
F1-Score	0.395	0.679	0.659	0.569

# k-Fold Cross Validation

Logistic regression with the three different regularizations

	Mean [train, validation]	Standard Deviation [train, validation]
Vanilla	0.70825, 0.70741	0.00044, 0.00296
Ridge (l2)	0.70826, 0.70748	0.00046, 0.00290
Lasso (l1)	0.70825, 0.70741	0.00044, 0.00296

# CM with Threshold = 0.55

	Threshold = 0.5	Threshold = 0.55
Score	0.769	0.769
Precision	0.265	0.331
Recall	0.596	0.533
Accuracy	0.265	0.331
F1-Score	0.367	0.409

