



Classification Project - Predicting the Possibility of someone subscribing term deposit

Calvin Yu



Problem/Need

- Portuguese banking institution is doing a new telemarketing campaign on term deposit
- Time consuming and not Cost efficient (Last) /
- Low Successful rate

Implement a more efficient way



Impact Hypothesis

- Create a classification model that gives you the probability of someone subscribing term the deposit
- Prioritize who to call first (High to Low)
- Less time consuming



Solution Path

**EDA:
visualizations
,heat map**

**Pick a
classification
metric
(ROC_AUC)**

**Test different
classification
algorithms**

**Pick the best
one !**



Scoping considerations

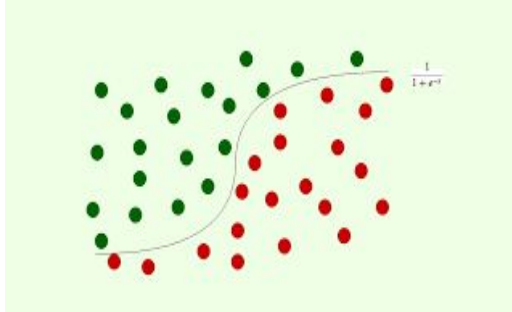
- Assumption:
 - People who subscribe term deposit share similar features
- Measure of Success :
 - How well does the model do
 - ROC AUC score
- Risk/Difficulty:
 - Some of the algorithms are hard to interpret
 - The period of the dataset is not long enough

Data

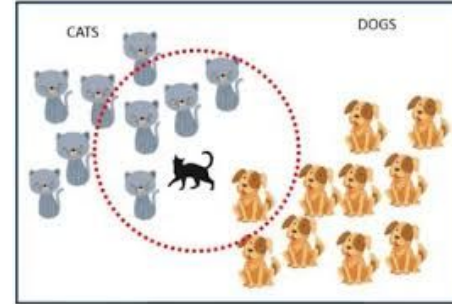
- Bank Marketing Data Set (UCI data base)
- 41188 rows with 19 features and 1 target variable
- May 2008 to Nov 2010



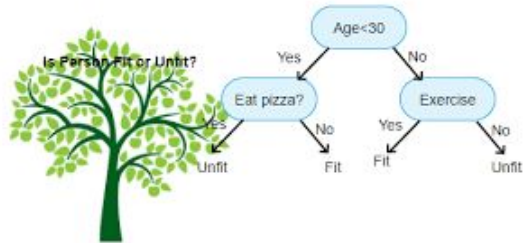
Algorithms



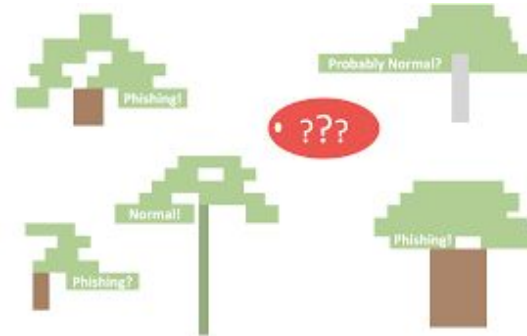
Logistic Regression



K-Nearest Neighbor

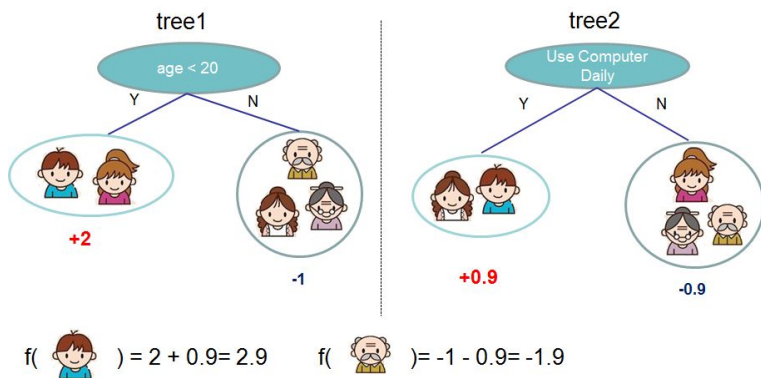


Decision Tree



Random Forest

Algorithms con.

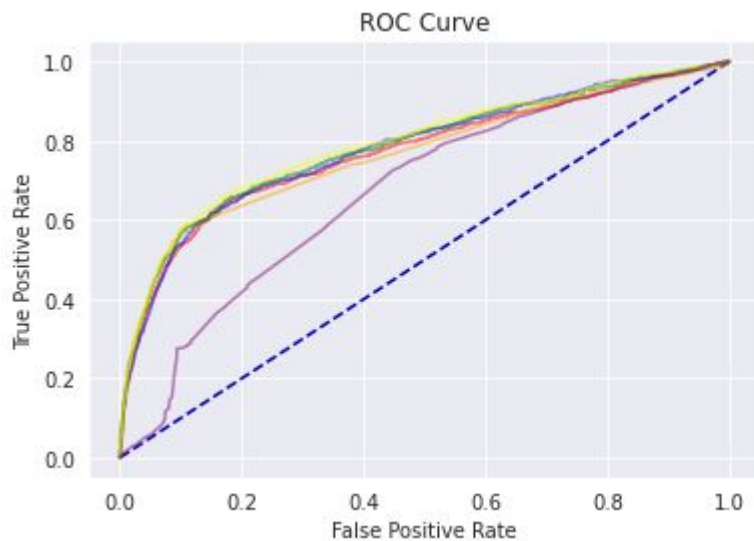


Gradient Boost / XGBoost



Naive Bayes

Results



Logistic Regression = 0.771

KNN = 0.780

Decision Tree = 0.766

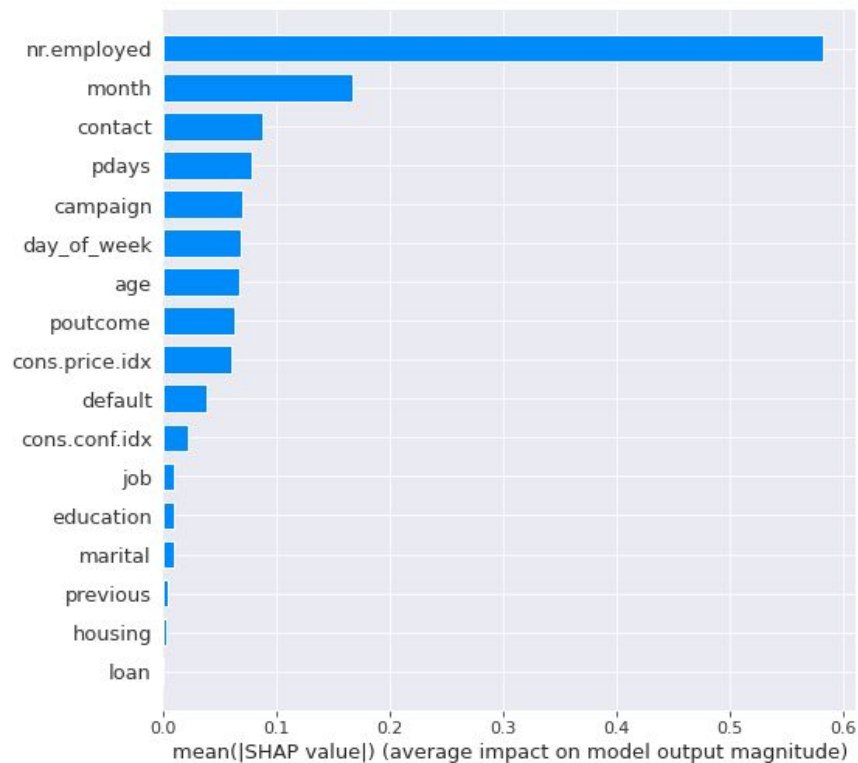
Random Forest = 0.784

XGboost/Gradient Boosting = 0.792

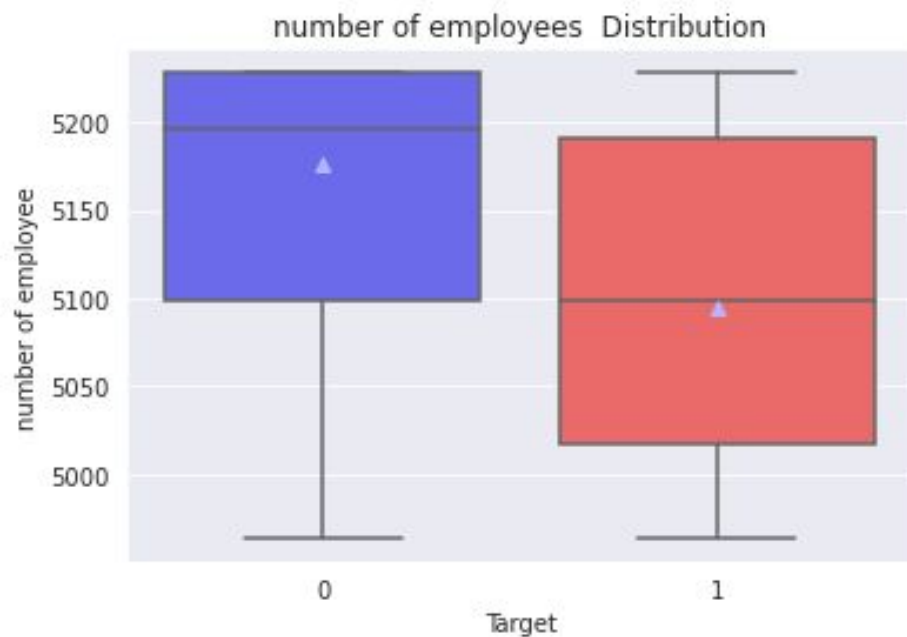
Naive Bayes = 0.672

Results con.

	Actual	Predict
1405	1	0.911900
6262	1	0.909961
1255	1	0.909126
743	0	0.899270
3298	1	0.891696
6862	0	0.891157
4636	1	0.879987
5597	1	0.879507
353	1	0.878762
4736	1	0.874104



Feature Importance



Further step

- Investigate on the feature importance
- Feature engineering
- Try more hyperparameters on some of the algorithms