

Project4-classification_MVP

Metrics:

I use F-beta as my main metrics(beta = 2, favors recall) as I want to lower the FN(true customer but predicted as not going to subscribe).

Imbalance handling:

Oversampling: ratio: 7:1

Models:

Logistic regression(feature engineering)

F-beta: 0.8222

ROC: 0.8841

5-NN:

F-beta: 0.8187

ROC: 0.6874

Random Forest:

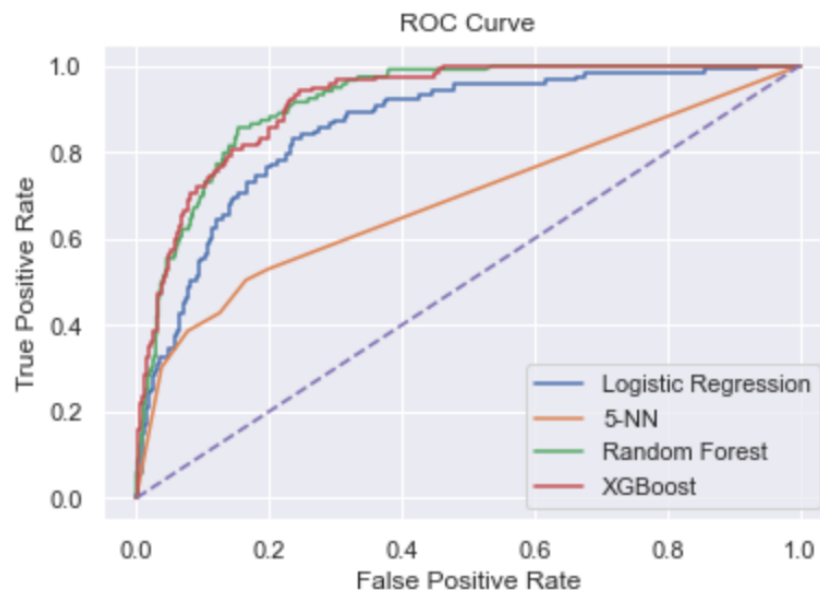
F-beta: 0.8853

ROC: 0.9171

I will also try xgboost & naive bayes to select the best model. If RF still performs the best on both F-beta score and ROC score, I will tune hyperparameters and threshold for Random.

Question:

when I choose F-beta as my priority metric(hard prediction), can I also compare ROC AUC as a secondary metric(soft prediction) just as a reference?



(ROC curve comparison)