

Assignment

To help & select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

Target Lead Conversion
rate=80%

Present conversion rate=38%

Methodology

By building a logistic regression model we assigned a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

I have indicated probability of conversion as yes and no. 1 indicated Yes and 0 indicated No.

Top 3 categorical variables impacting lead conversion

1.Tags_Will revert after reading the email

2.What matters most to you in choosing a course_Better Career Prospects

3.Lead Quality_Might be

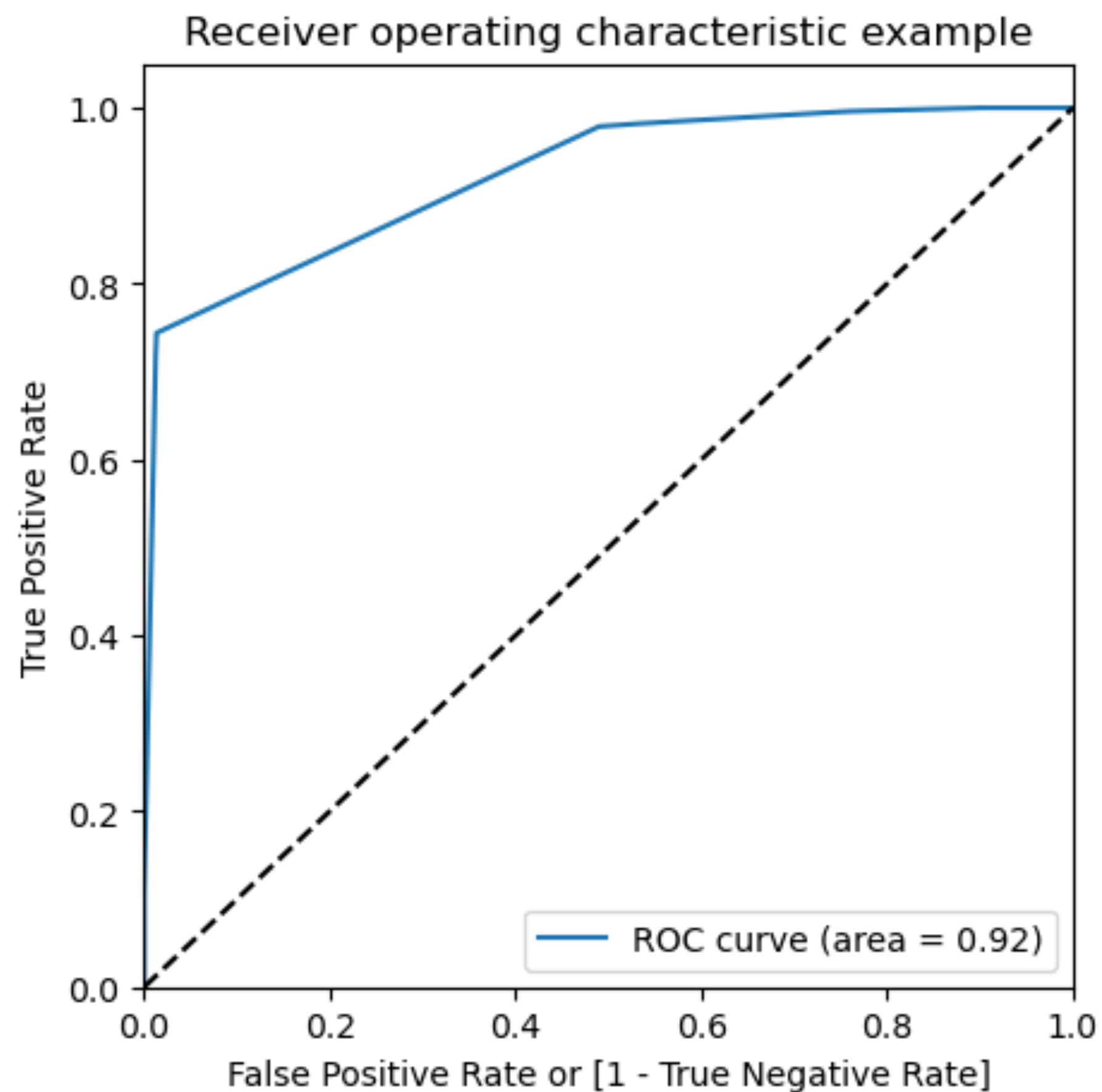
Top 3 variables contributing to lead conversion:

Total Time Spent on Website

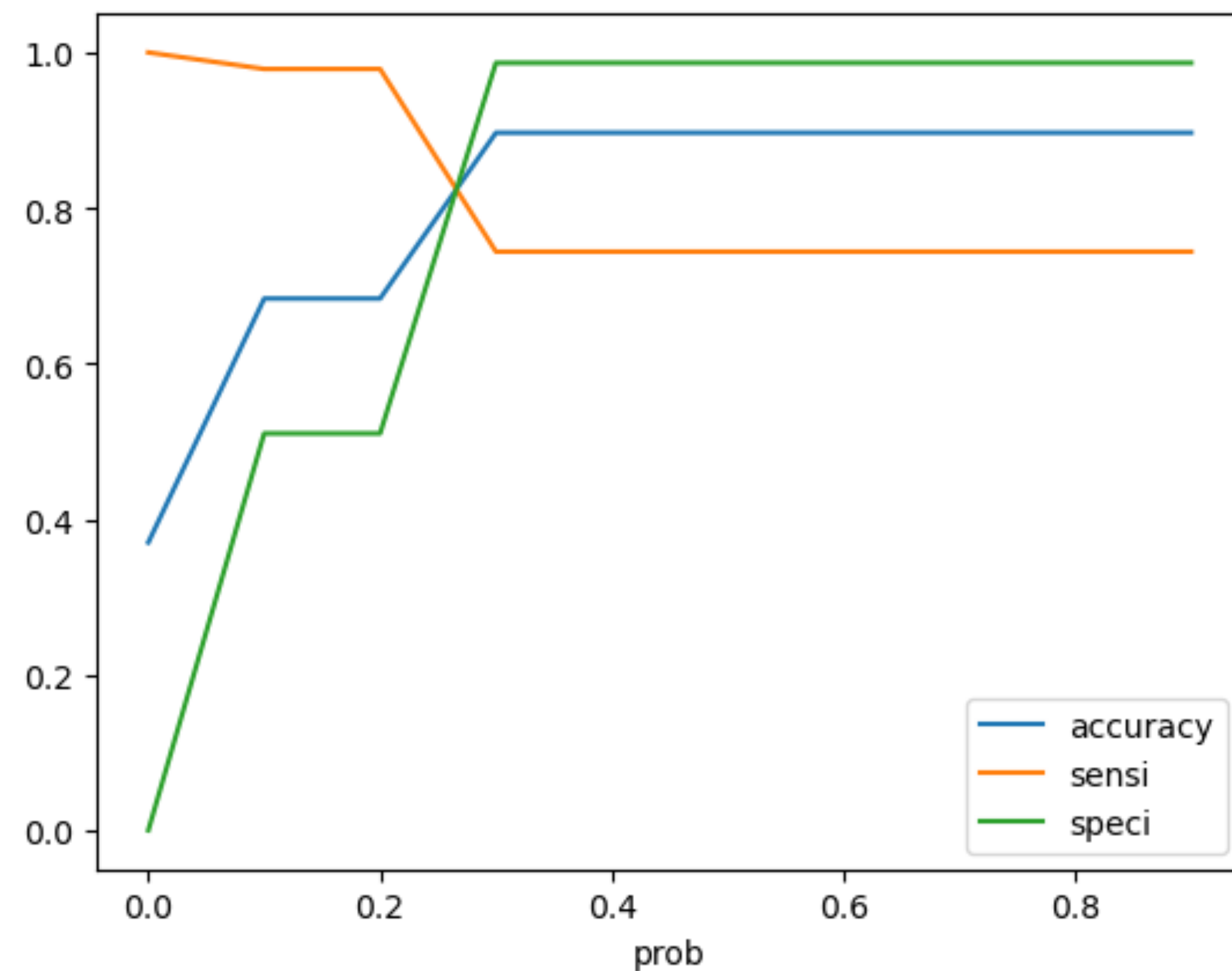
Tags_Will revert after reading the email

What matters most to you in choosing a course_Better Career Prospects

ROC Curve

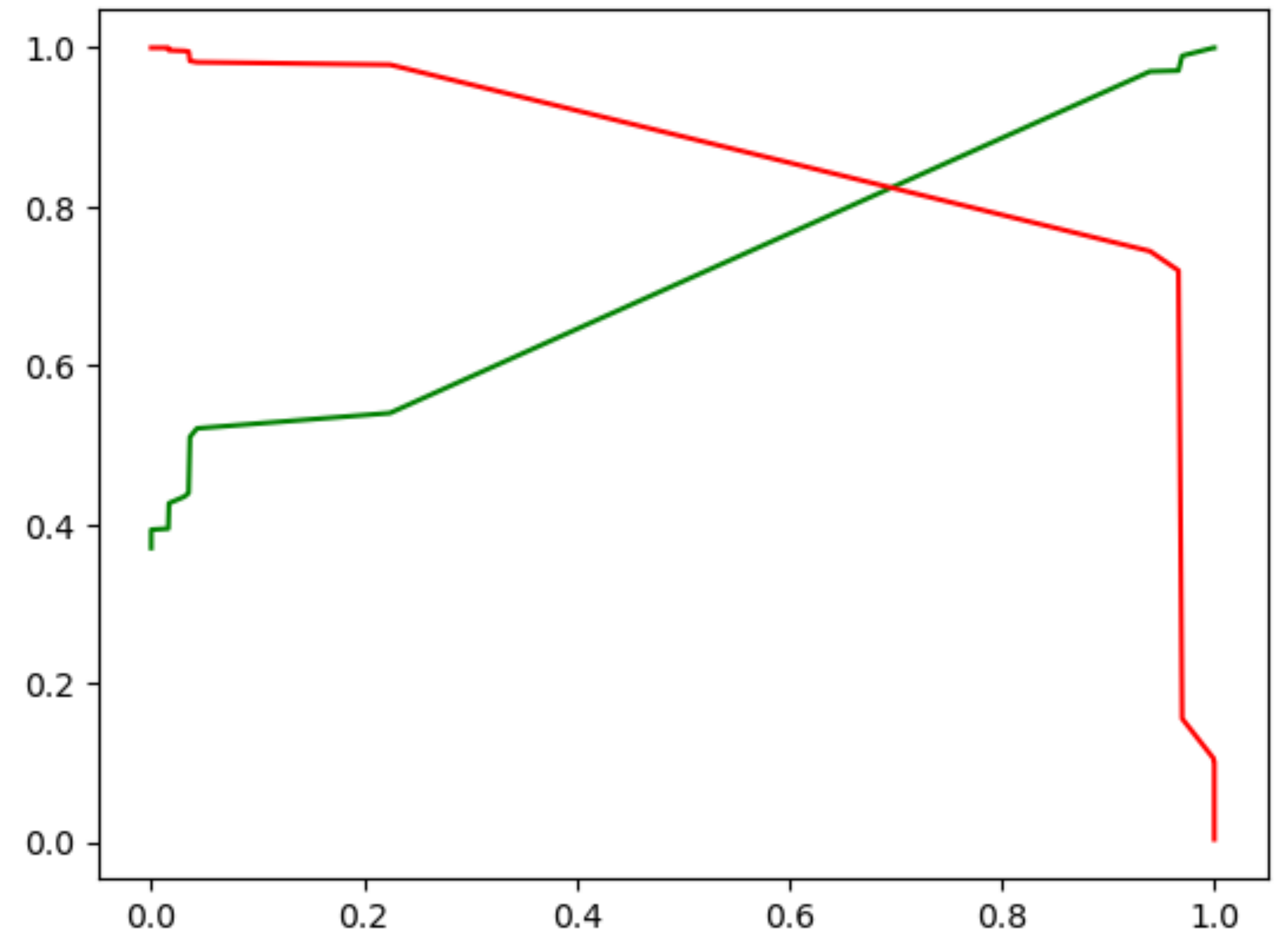


Probability Cutoff



Precision-Recall Curve

To choose a threshold where recall in Red is high and precision in Green is still acceptable.



Recommendation

1. If we have manpower (e.g new interns) we can decrease probability to include more prospects
2. In reverse case, if company has achieved target before time, it may increase probability to increase precision.