

Lead Scoring case study

Rucha Shelke
Sneha Velankar

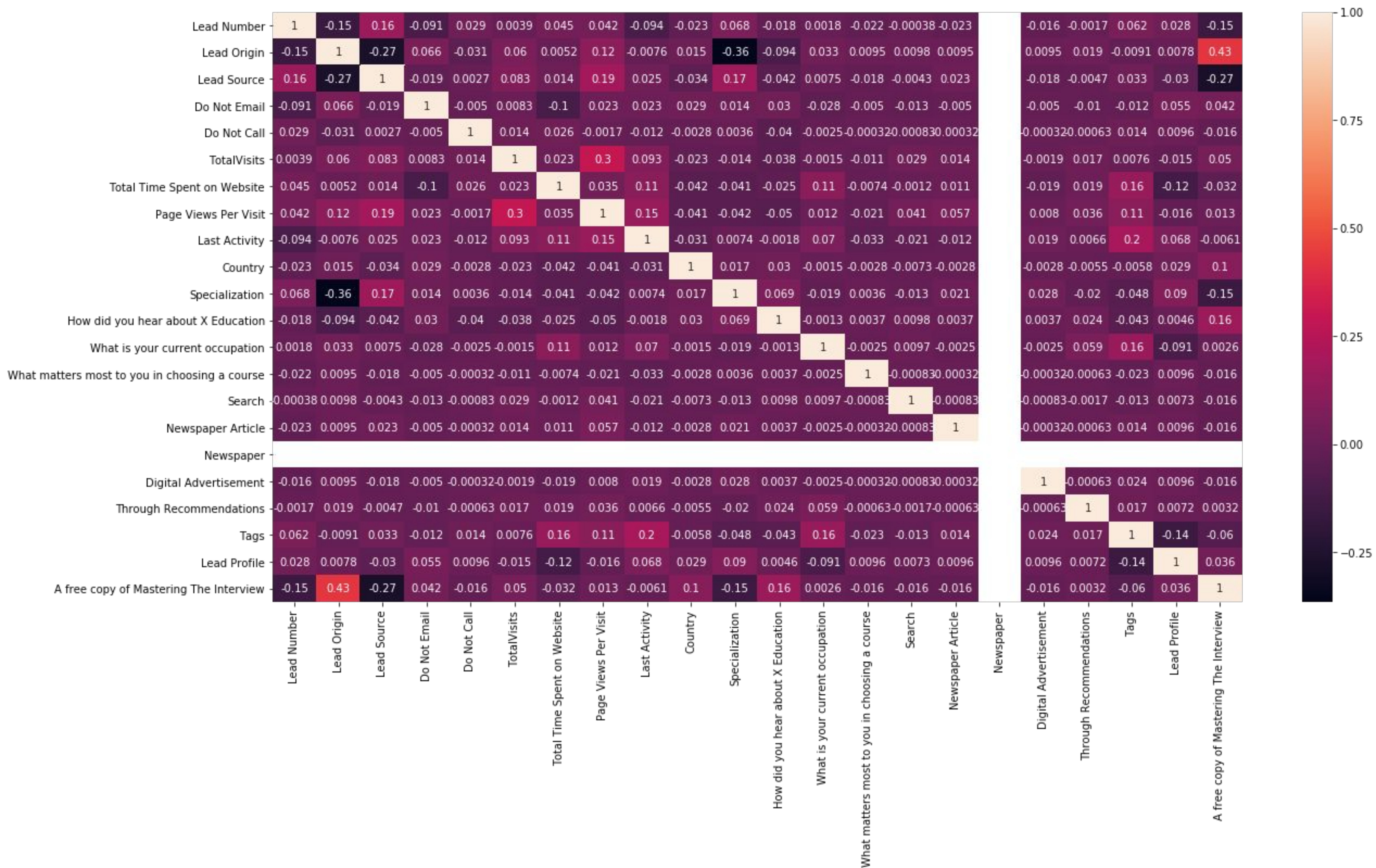
Problem statement

An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. The typical lead conversion rate at X education is around 30%. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. Objective is to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Analysis approach

1. Data reading and understanding
2. Data preparation
3. Test - Train split
4. Feature scaling
5. Model building
6. Feature selection using RFE
7. Manual feature elimination
8. Confusion metrics and ROC curve
9. Precision and recall
10. Predictions on test set

Correlation matrix of Final variables

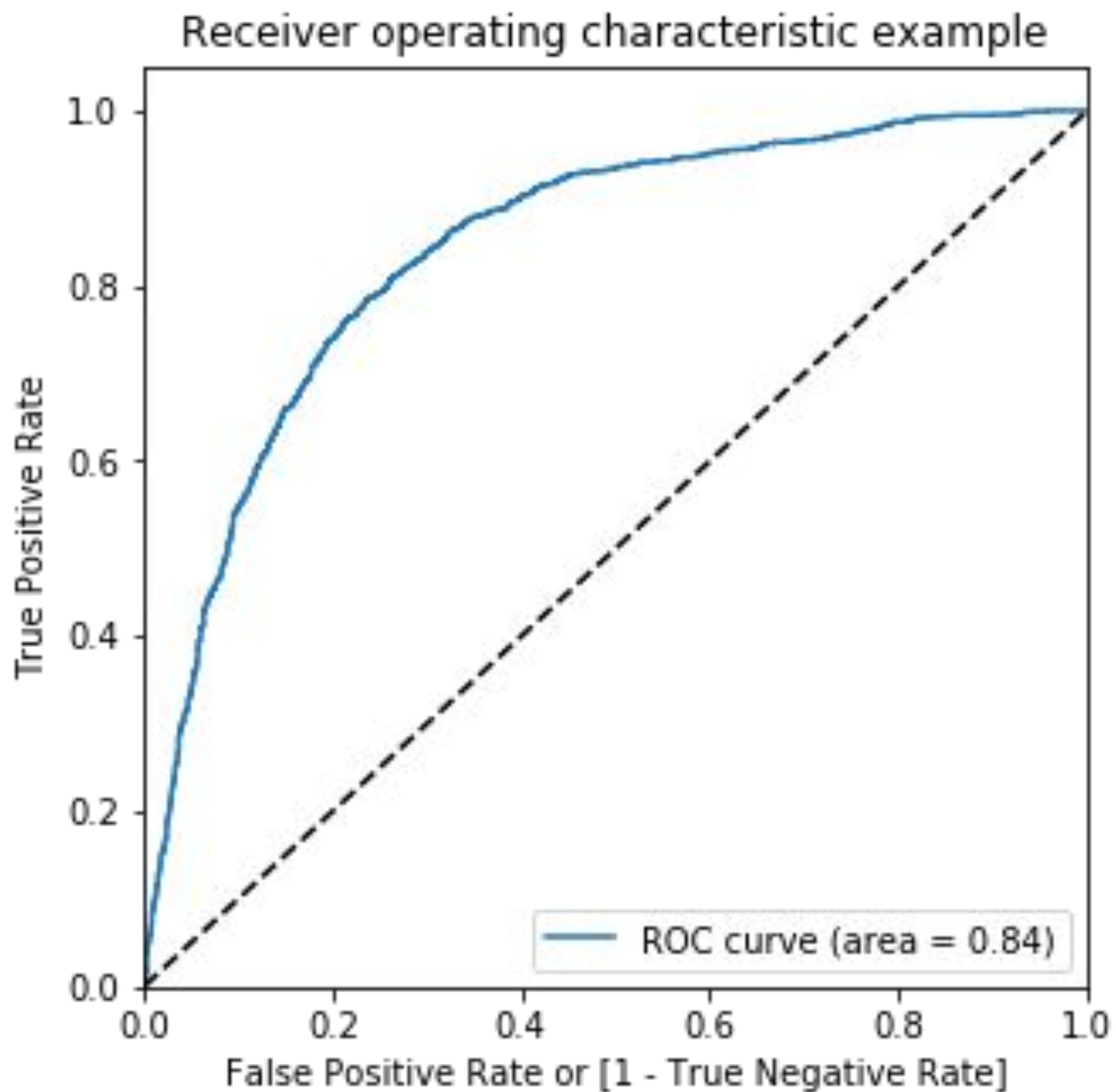


Confusion matrix

Actual\ Predicted	Not converted	converted
Not converted	1527	313
converted	416	917

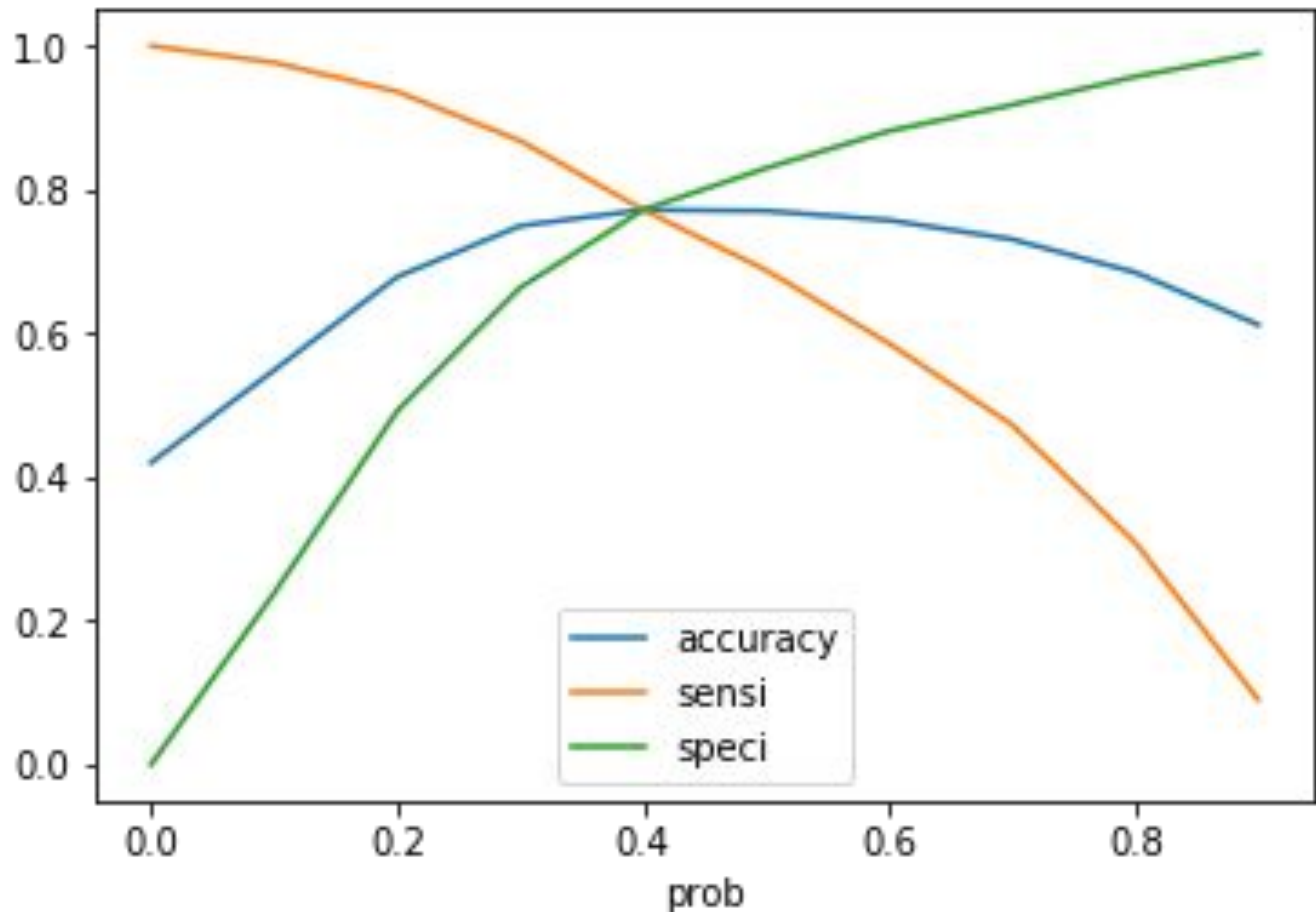
- Accuracy: 77%
- Sensitivity: 69%
- Specificity: 83%
- False positive rate: 17%
- Positive predictive value: 74%
- Negative predictive value: 78%
- Precision: 74%
- Recall: 69%

ROC Curve



From above ROC curve, we can depict that the model is accurate.

Optimal cut off



From the curve above, 0.4 is the optimum point to take it as a cutoff probability i.e. in general terms, leads with probability > 0.4 are considered to be getting converted.

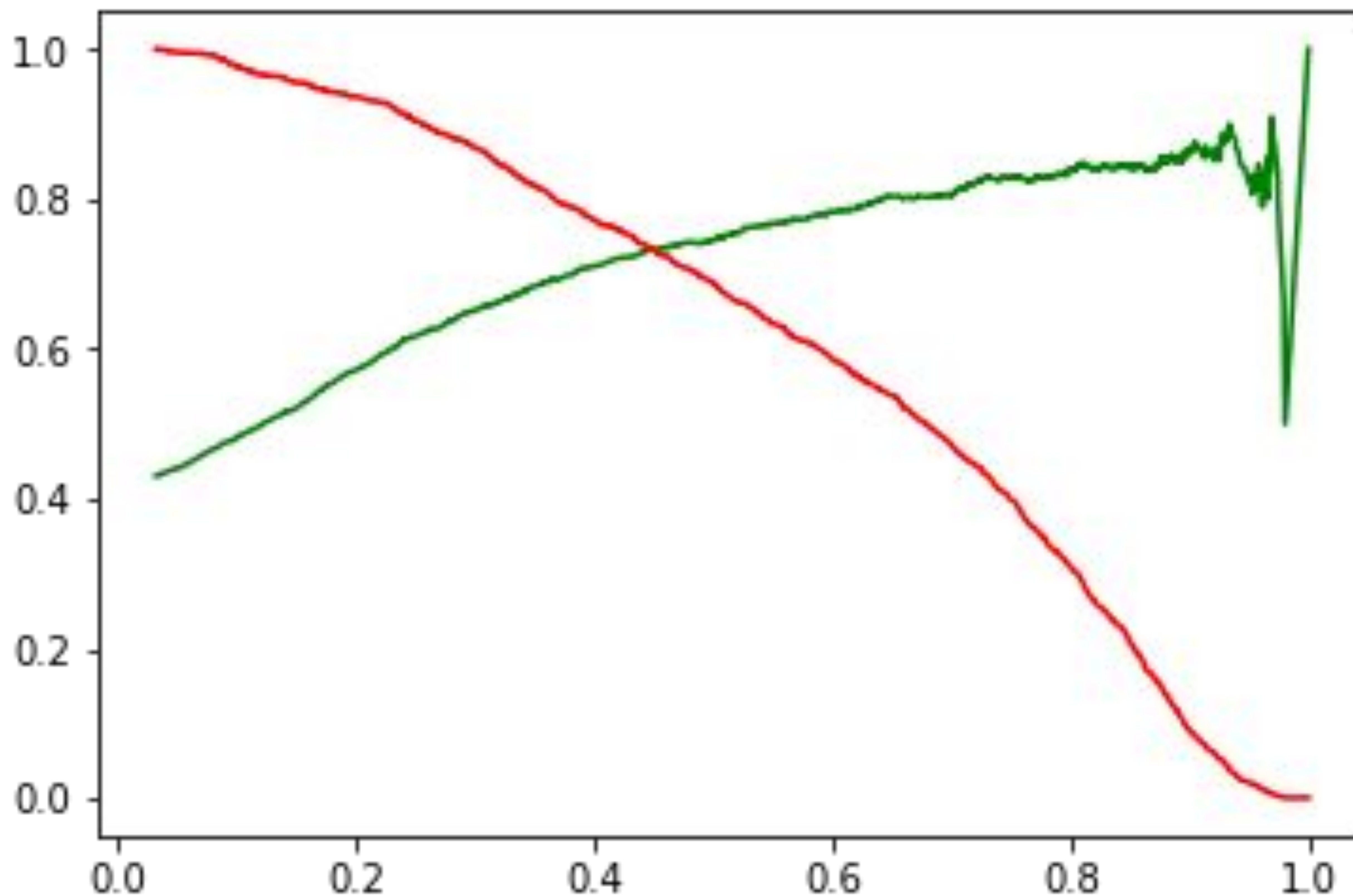
Confusion matrix and other metrics

Actual\ Predicted	Not converted	converted
Not converted	1420	420
converted	304	1029

Based on cut off point 0.4 below are the metrics

- Accuracy: 77%
- Sensitivity: 77%
- Specificity: 77%
- False positive rate: 23%
- Positive predictive value: 71%
- Negative predictive value: 82%

Precision and Recall tradeoff



From the curve above, 0.42 is the optimum point to take it as a cutoff probability i.e. in general terms, leads with probability > 0.42 are considered to be getting converted.

Observations on Test dataset

Actual\ Predicted	Not converted	converted
Not converted	610	167
converted	151	433

- Sensitivity: 74%
- Specificity: 78%

Recommendations

To increase lead conversion probability, company should focus most on below fields:

1. Lead origin
2. Specialization
3. Last Activity

As now lead score (between 0 and 100) is applied to each lead, it's obvious that hot leads would get converted. Hence, it is recommended to target them first.