

Lead Scoring Case Study

By – Subhransu Pradhan
Suman Gandhi

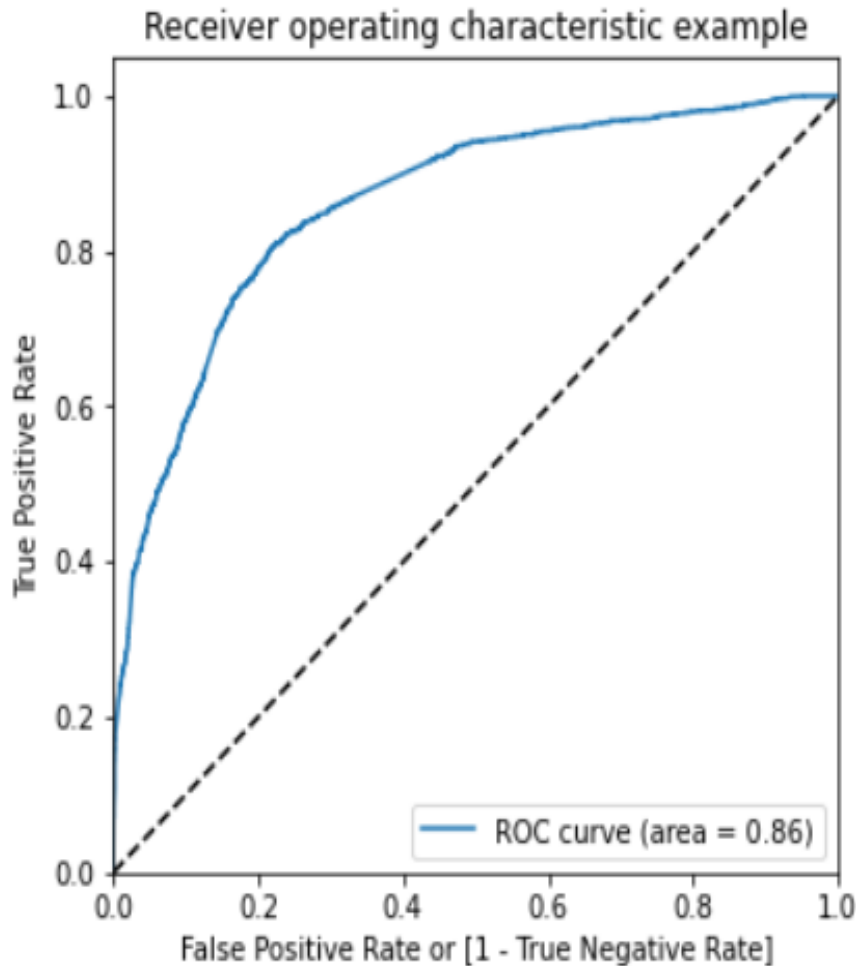
Problem Statement

- The X Education company requires you to build a logistic regression model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

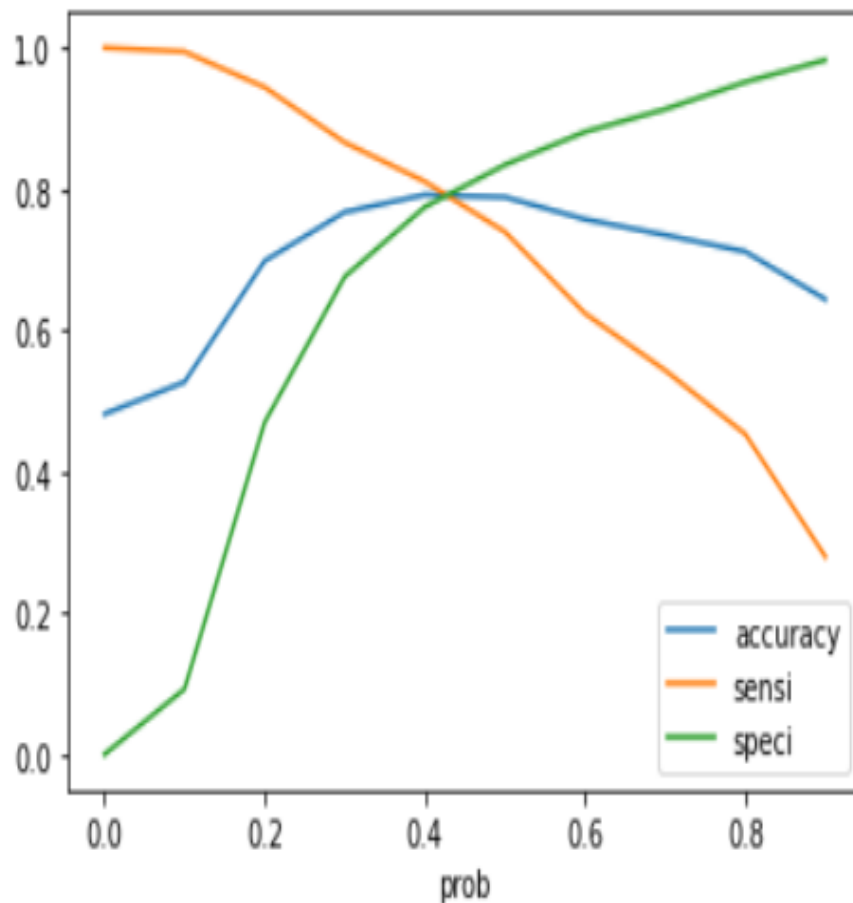
Roadmap

- Created train and test set by splitting the original cleaned data set after treating missing values.
- Selected 15 features using Recursive Feature Elimination (RFE) after creating dummy variables and scaling the data.
- Applied Logistic Regression algorithm to build a model and more than 79% accuracy and 81% sensitivity.
- Identified the optimal probability cutoff from the accuracy, sensitivity and specificity.
- Applied the model on the test data to identify the conversion probability.
- Based on the calculated predicted probability, and optimal probability cutoff, all the leads are assigned with a lead score value (lead score = predicted probability x 100)

ROC Curve



- The ROC curve shows that the 86% of the area is under the curve.
- The classification probability of lead conversion (1/0) is very high by the model.



- Optimal probability cutoff is identified as 0.4 for better accuracy of the classification of lead conversion.
- With 0.4 cutoff the model has
 - Accuracy : 79%
 - Sensitivity : 81%
 - Specificity : 77%

Confusion matrix on Test data

Actual/Predicted	Not Converted	Converted
Not Converted	1823	489
Converted	444	1705

Accuracy : 79% | Sensitivity : 81% | Specificity : 77%

The model can predict if a lead can be converted or not with 79% accuracy on unseen data. This will help the company to predict the probability of 'hot' leads with 79% accuracy.

Also, the model can predict the probability of a lead which are actually converted over total converted lead with 81% chances.

The model's prediction of a lead not getting converted is also very high (77% over unseen data). This means that the X education company will save lot of time and resources by discarding low scoring leads.

Recommendation

- The leads which have high score can be treated as “hot” leads and sales team need to follow up as there is high possibility to convert those leads.
- Leads who have applied for ‘Do Not Email’ already does not needs to be attended again.
- Based on the previous chat conversations if the lead is classified as ‘Might be’ or ‘Worst’ then those leads can be ignored.

Thank You