

Lead Score Case Study



Prepared By
Arpita Nandwani

01. Business Problem
02. Business Goal
03. Data Cleaning and Preparation
04. Model Building
05. Model Evaluation
06. Conclusion

CONTENTS



01

Business Problem

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%. Now, although X Education gets a lot of leads, its lead conversion rate is very poor.





02

Business Goal

The company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build 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. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.





03

Data Cleaning and Preparation

Steps Involved:

- Importing libraries
- Reading the data
- Basic data check
- Checking for null values
- Dropping columns with more than 40% null values
- Checking for duplicates
- Imputing missing values with median or mode for numerical and categorical Variables respectively
- Creating Dummy Variables





04

Model Building

Steps Involved

- Splitting the data into Train and Test Sets
- Scaling the data using Min Max Scaler
- Feature Selection using RFE
- Fitting the model
- Dropping the variables with high p-value
- Checking VIF for selected features
- Dropping the variable with high VIF
- Repeating the process until best fit model is obtained



05

Model Evaluation

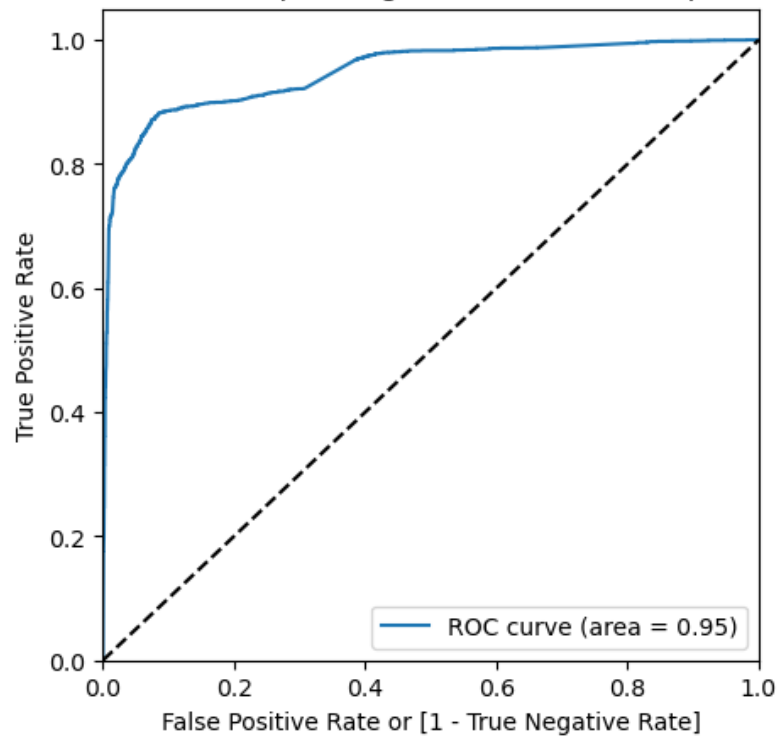
Steps Involved

- Checking Confusion matrix and accuracy on Train data
- Plotting the ROC Curve
- Finding the optimal threshold
- Checking for Precision and Recall values on Train Data
- Making Predictions using trained model on Test Data
- Checking Confusion matrix and accuracy on Test data
- Checking for Precision and Recall values on Test Data



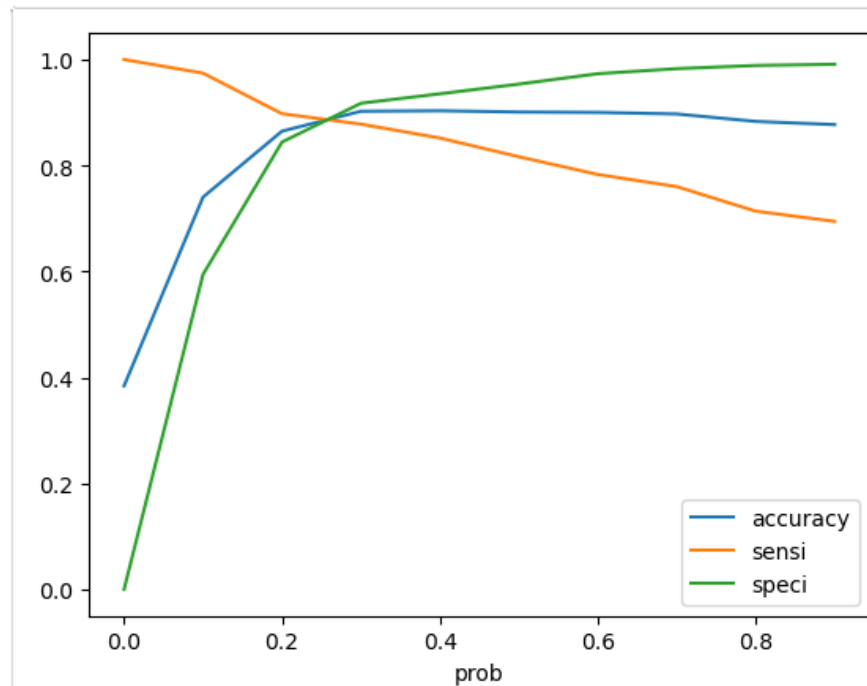


Receiver operating characteristic example



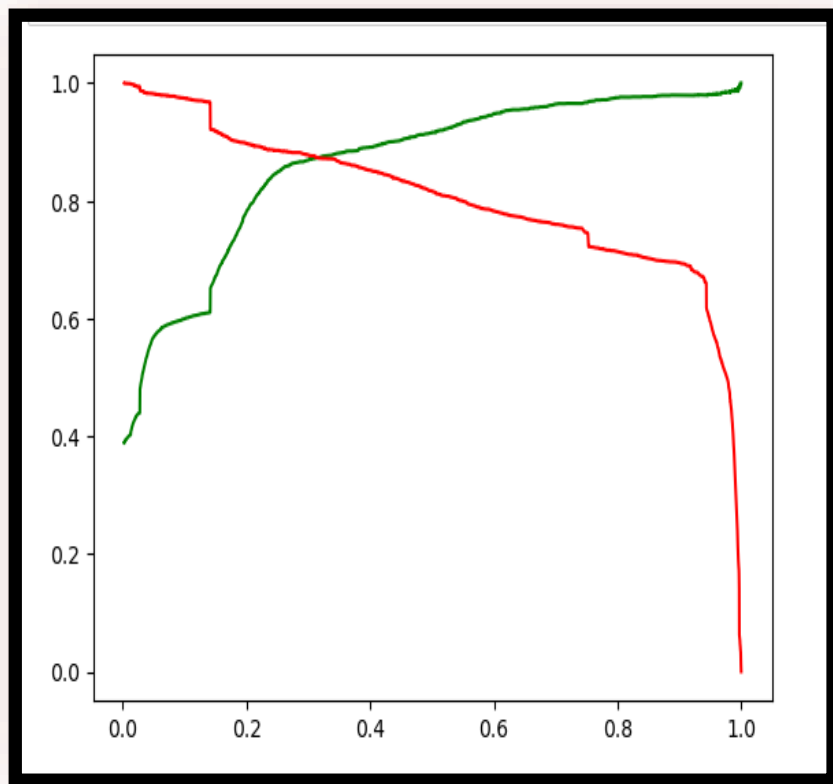
ROC Curve

Since ROC curve area is 0.95, this indicates the predictive model is good.



Optimal Threshold

From this curve, 0.3 is the optimum point to take it as a cutoff probability



Precision Recall Curve

Precision Score: 86.89%

Recall Score: 87.80%



06

Conclusion

1. Train Data

Accuracy : 90.23% Sensitivity : 81.80% Specificity : 91.74%

2. Test Data

Accuracy : 90.76% Sensitivity : 88.21% Specificity : 92.39%

While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.

Accuracy, Sensitivity and Specificity values of test set are around 90%, 88% and 92% which are approximately closer to the respective values calculated using trained set.

Hence overall this model seems to be good.



A large, solid orange circle is positioned on the left side of the slide, partially cut off by the edge. It has a soft shadow effect.

THANK YOU

