Lead Scoring case study

Name:

Prabha Muthusamy

Lead Score Case Study for X Education

Problem Statement :

- 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.
- 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%.

Business Goal:

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned 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.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

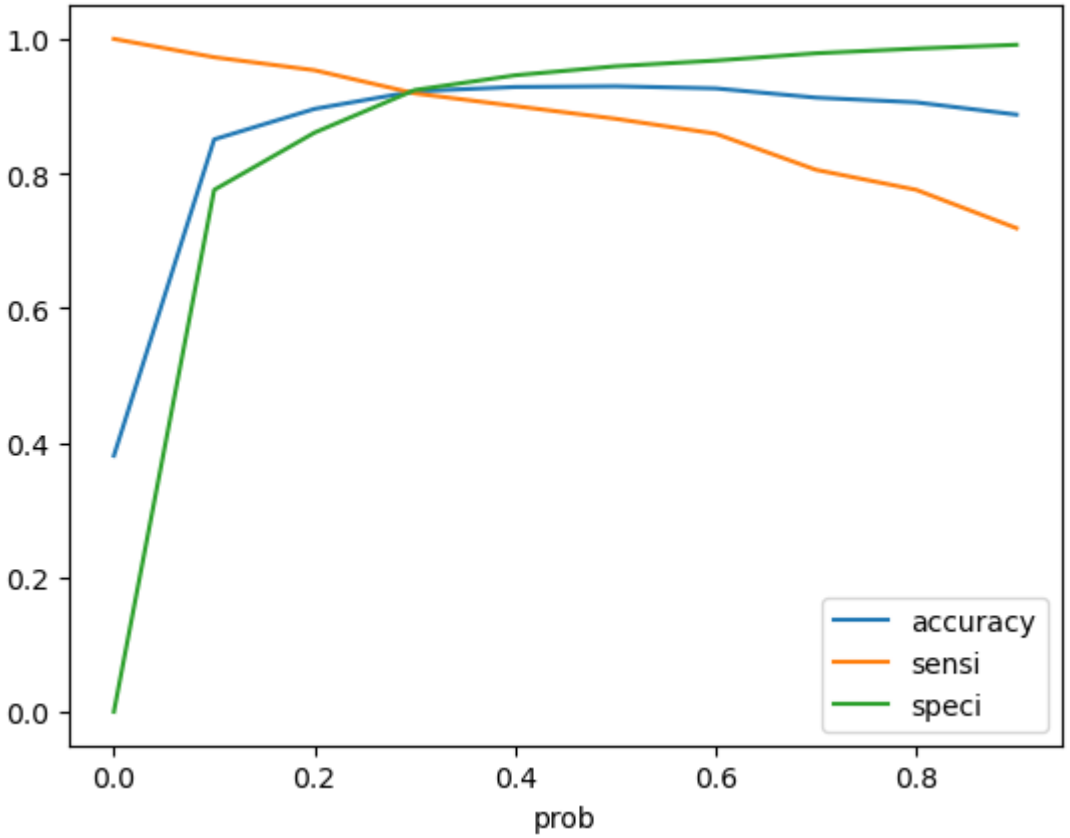
Steps to be followed

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics

Variables Impacting the Conversion Rate

- Lead Origin
- Lead Source
- Do Not Email
- Do Not Call
- Converted
- TotalVisits
- Total Time Spent on Website
- Page Views Per Visit
- Last Activity
- Specialization
- What is your current occupation
- Search
- Magazine
- Newspaper Article
- X Education Forums
- Newspaper
- Digital Advertisement
- Through Recommendations
- Receive More Updates About Our Courses
- Tags
- Update me on Supply Chain Content
- Get updates on DM Content
- City
- I agree to pay the amount through cheque
- A free copy of Mastering The Interview

Model Evaluation - Sensitivity and Specificity on Train Data Set



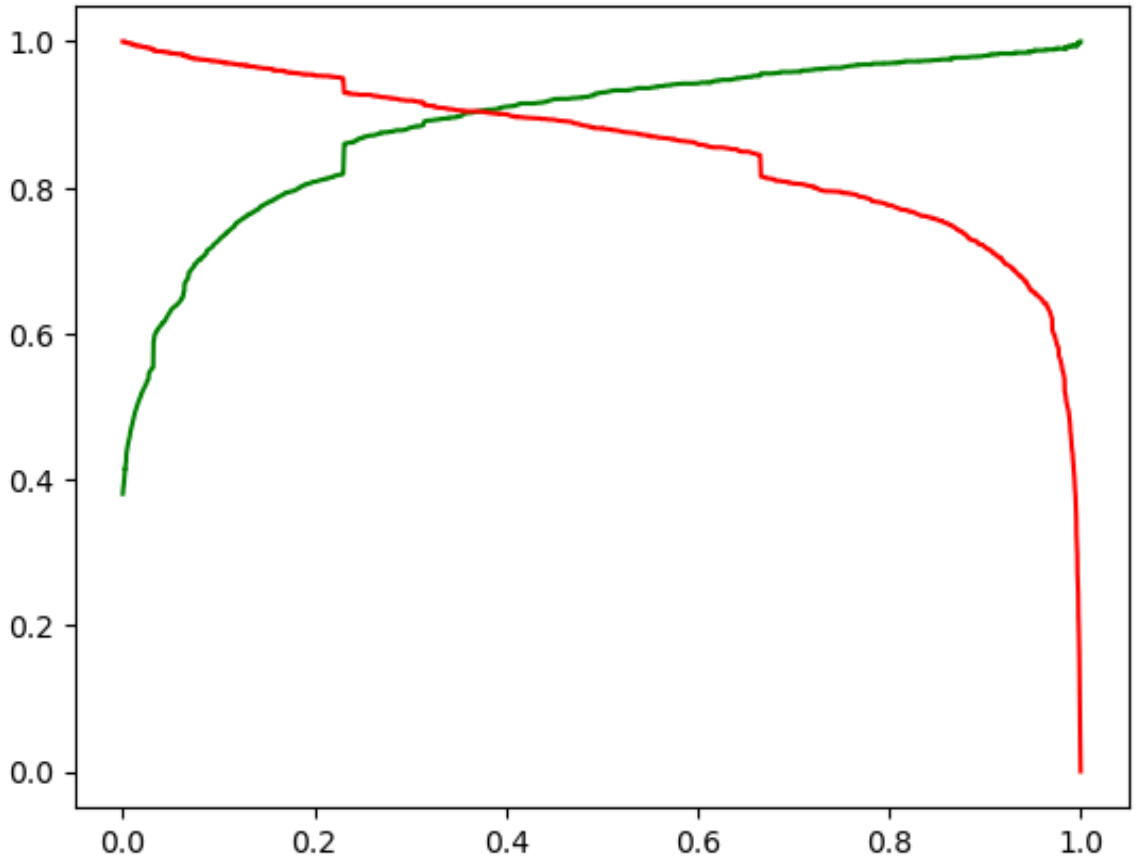
Confusion Matrix

3724 158

283 2102

- Accuracy- 92.21%
- Sensitivity- 91.9 %
- Specificity- 92.4 %

Computing the Precision and Recall metrics



Confusion Matrix

3587 295

193 2192

- Precision- 88.13 %
- Recall- 91.2 %

Conclusion

- Both the 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.
- The value for ‘accuracy=92.21%, ‘sensitivity=91.9%’, ’specificity=92.4%’ are found. which are approximately closer to the respective values.
- The top 3 variables that contribute for lead getting converted in the model are
  - Total time spent on website
  - Lead Add Form from Lead Origin
  - Had a Phone Conversation from Last Notable Activity
- Hence overall this model seems to be good.