

Lead Scoring Case Study

Omprakash kumar
sharma

Sunita Sharma

chiranjeev vyas

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.

Step of model creation

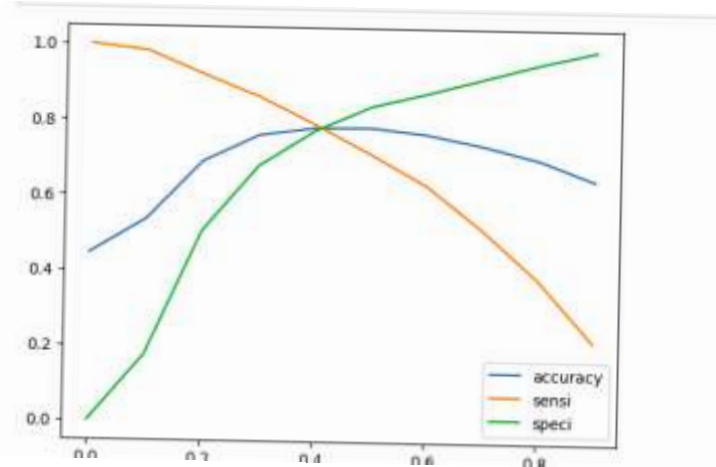
- › Source the data for analysis
- › Clean and prepare the data
- › Exploratory Data Analysis.
- › Feature Scaling
- › Splitting the data into Test and Train dataset.
- › Building a logistic Regression model and calculate Lead Score.
- › Evaluating the model by using different metrics - Specificity and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

Variable impacting the conversion rate

- ›Do Not Email
- ›Total Visits
- ›Total Time Spent On Website
- ›Lead Origin – Lead Page Submission
- ›Lead Origin – Lead Add Form
- ›Lead Source - Olark Chat
- ›Last Source – Welingak Website
- ›Last Activity – Email Bounced
- ›Last Activity – Not Sure
- ›Last Activity – Olark Chat Conversation
- ›Last Activity – SMS Sent
- ›Current Occupation – No Information
- ›Current Occupation – Working Professional

Model Evaluation - Sensitivity and Specificity on Train Data Set

- The graph depicts an optimal cut off of 0.37 based on Accuracy,
- Sensitivity and Specificity
- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%



Model Evaluation – Sensitivity and Specificity on Test Dataset

- Accuracy - 81 %
- Sensitivity - 79 %
- Specificity - 82 %
- confusion matrix:

```
0.797952541001333
```

```
confusion2 = metrics.confusion_matrix(y_pred_final['Converted'], y_pred_final.final_predicted )  
confusion2
```

```
array([[682, 154],  
       [149, 493]])
```

```
TP = confusion2[1,1] # true positive
```

Conclusion

- 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 81%, 79% and 82% which are approximately closer to the respective values
 - calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- 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