Lead Score Case Study

Presented By :

Vishal Rathi

Vishal Jha

Agenda

- 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%.

Strategy

- 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

- 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
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

Problem Solving Methodology

Data Sourcing, Cleaning and Preparation

- Read the data from source
- Convert data into clean format suitable for analysis.
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization

Feature Scaling and Splitting train and Test Sets

- Feature Scaling of Numeric Data
- Splitting data into train and test set.

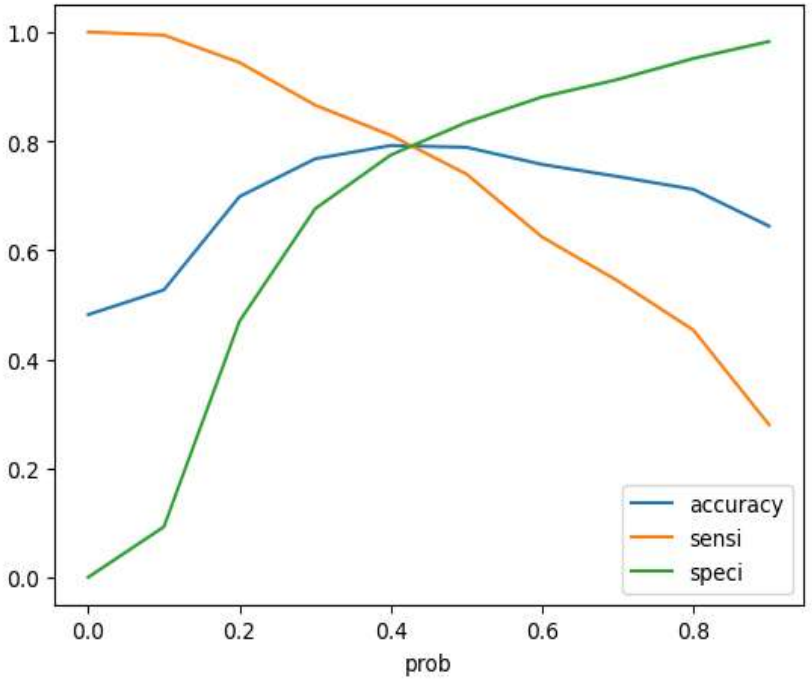
Model Building

- Feature Selection using RFE
- Determine the optimal model using Logistic Regression.
- Calculate various metrics like accuracy, senstivity, specificity, precision and recall and evaluate the model.
-

Results

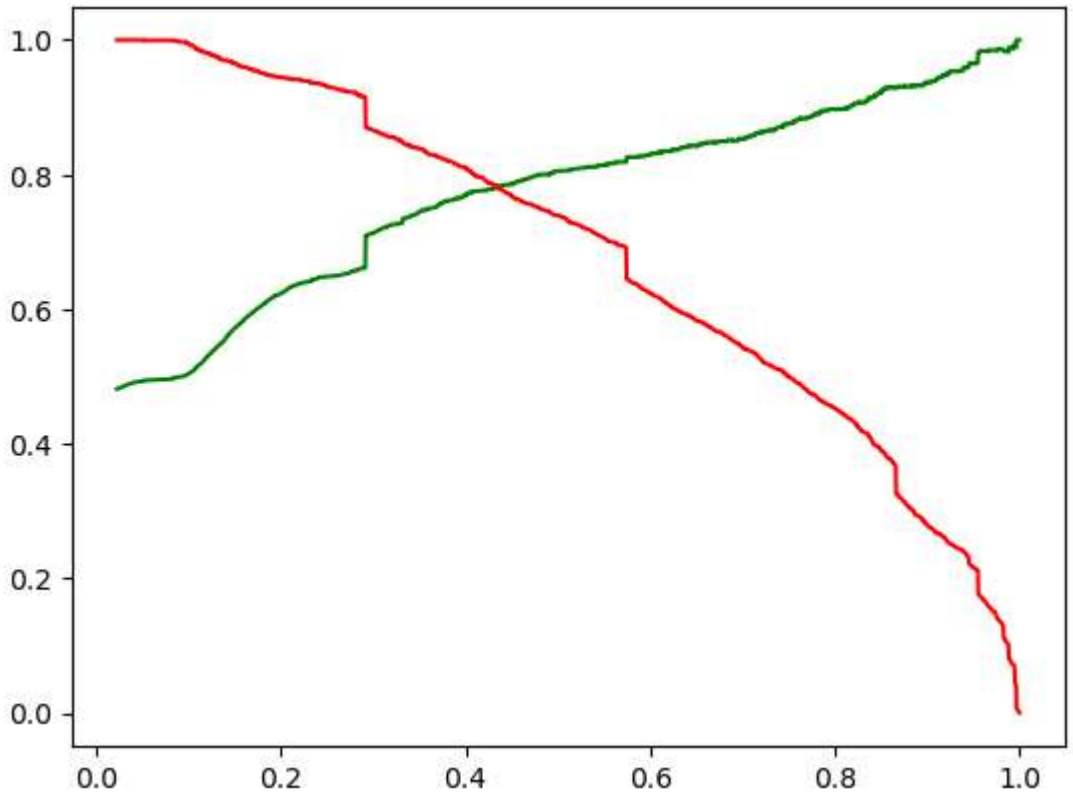
- Determine the lead score and check if target final prediction amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from senstivity and specficity metrics.

Model Evaluation - Sensitivity and Specificity on Train Data Set



- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset



- Precision - 79 %
- Recall - 71 %

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
- Had a Phone Conversation from Last Notable Activity
- Hence overall this model seems to be good.