



LEAD SCORE CASE STUDY

By David Lee & Satish Singh



PROBLEM STATEMENT

X Education sells online course to industry professionals and markets its courses on several websites and search engines.

Once users are on the X Education website, they browse courses, watch videos, or fill up a form for a course. When they fill up a form and provide their address or phone number, they are classified as a lead. The company also gets leads through past referrals. Once the leads are acquired, the sales teams reaches out to the leads for conversion. Currently, the typical lead conversion rate is around 30%.

X Education needs help in selecting the most promising leads (hot leads) increase the conversion rate.



OBJECTIVE

X Education wants to make the lead identification process more efficient by assigning a lead score to each customer such that the customer with a higher lead score have a higher conversion.

1. Build a machine learning model to assign a lead score between 0 and 100 to each of the leads, which can then be used by the company to target potential leads. A higher score would mean that the lead has a higher chance of conversion.
2. Using the model, obtain a target lead conversion rate of around 80%
3. The model should be able to adjust to new problems presented if the company's requirements change in the future.

ANALYSIS APPROACH

- EDA
 - Initial exploration of data
 - Clean data and handle outliers
 - Convert category values to numerical representation
- Model building
- Model validation
- Final analysis and recommendations

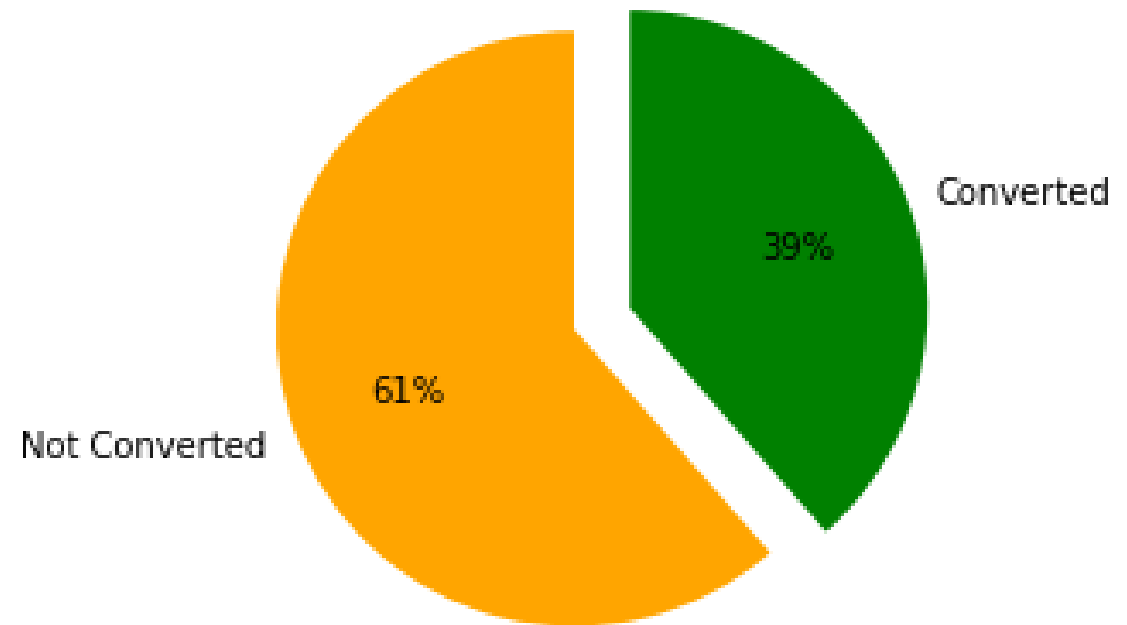


MODEL SUMMARY

HISTORIC CONVERSION RATE

- Based on provided data of leads, the historic conversion rate is 39%

Converted vs Not Converted



LOGISTIC REGRESSION MODEL

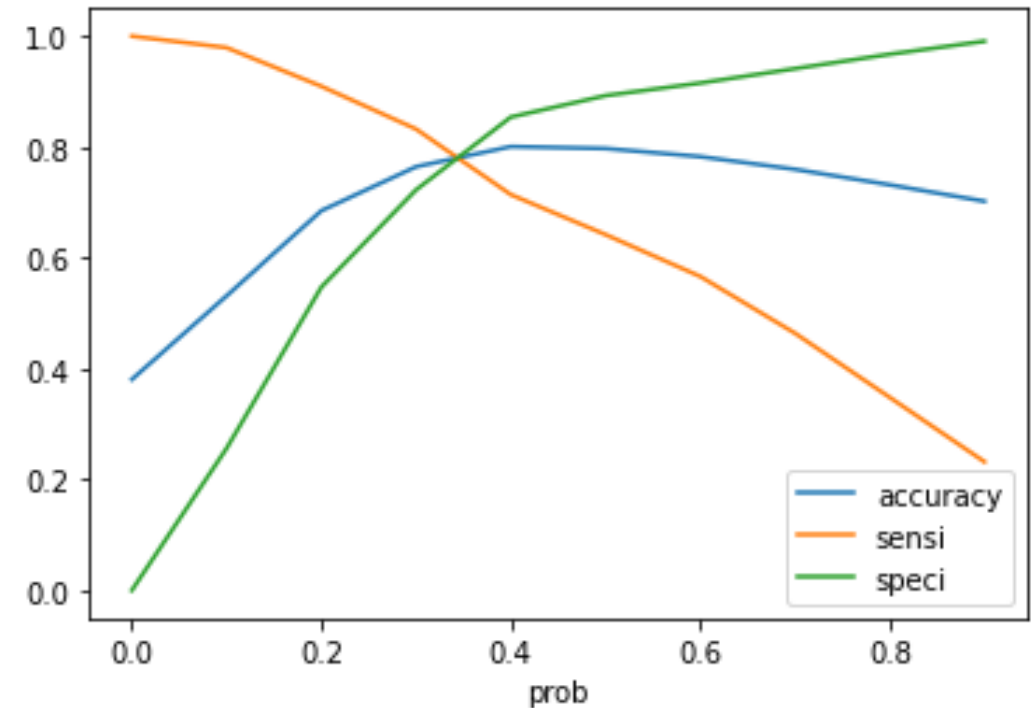
- Using logistic regression (binomial) model, historic data will be used to predict the conversion rate, based on several lead customer features, to around 80%.
- Recursive Feature Elimination (RFE) used to reduce features to 15. And manual feature elimination was done based on P values and VIF.

ACCURACY, SENSITIVITY, SPECIFICITY SELECTION

Prediction for the data is based on finding the optimum boundary of accuracy, sensitivity, and specificity

- Accuracy is the percentage of correct prediction the model made
- Sensitivity measures how well the model predicts those customers that converted
- Specificity measures how well the model predicts those customers that were not converted

Based on the graph, 0.35 is the optimum point to take a cutoff of probability

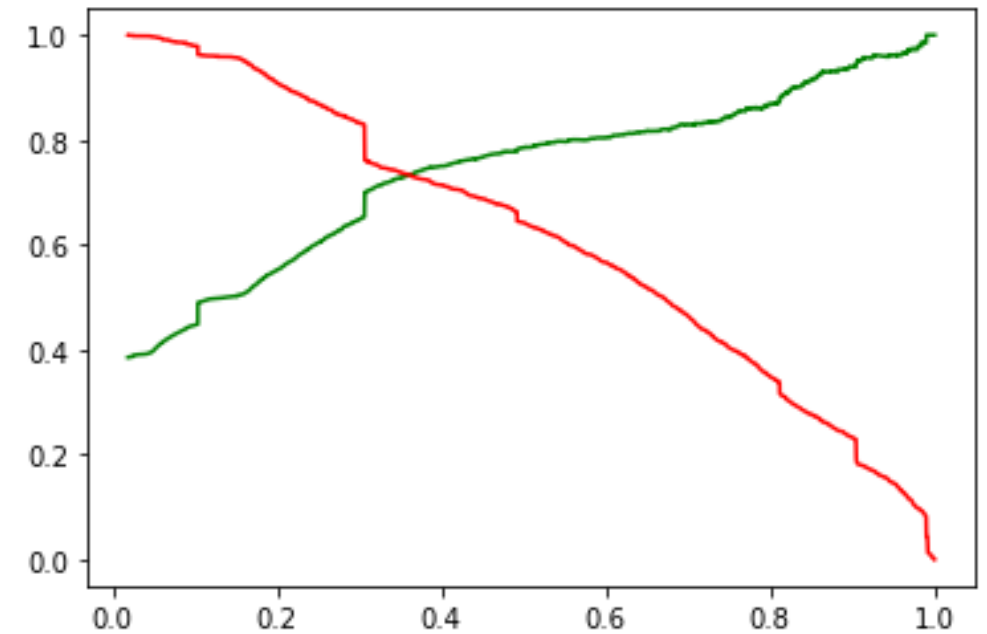


PRECISION VS RECALL

Prediction is also based on tradeoff between precision and recall

- Precision is the accuracy of the prediction.
- Recall, as with sensitivity, is how well the model predicts those customers that converted

Based on the graph, 0.3 is the optimum point to take a cutoff of probability which will give us around 80% conversion of leads



PREDICTION PROBABILITIES

- Based on Accuracy, Sensitivity, Prediction and Precision vs Recall charts, .3 probability cutoff has been made.
- Below are the probabilities based on .3 probability cutoff.

	Training Data Set	Test Data Set
Accuracy	76.44%	76.98%
Sensitivity	83.24%	83.30%
Specificity	72.27%	73.19%
Precision	64.84%	65.08%

Accuracy: percentage of correct prediction the model made

Sensitivity: how well the model predicts those customers that converted

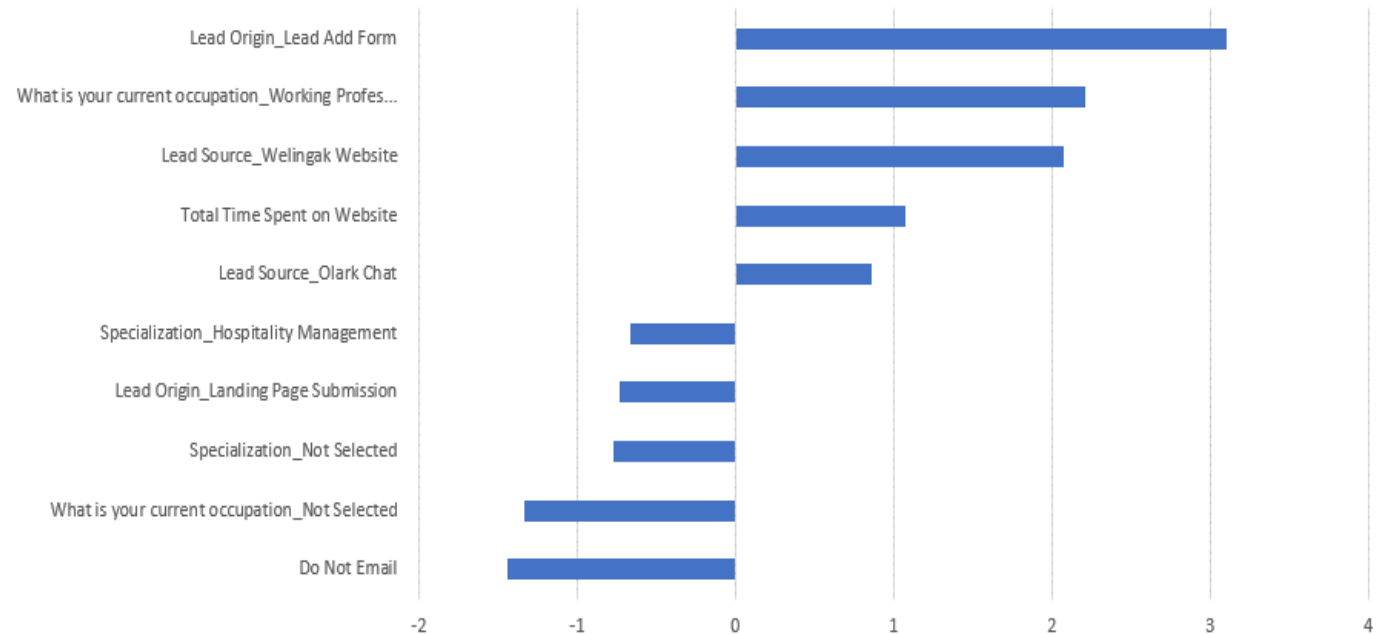
Specificity: how well the model predicts those customers that were not converted

Precision: quality of positive predictions

FEATURES SELECTED

- Positive effect on conversion
 - Lead Origin – Lead Add Form
 - Occupation – Working Professional
 - Lead Source – Welingak Website
 - Total Time Spent on Website
 - Lead Source – Olark Chat
- Negative effect on conversion
 - Do Not Email
 - Occupation – Not Selected
 - Specialization – Not selected
 - Lead Origin – Landing Page Submission

Feature Importance Ranking (4 = High, -2 = Low)



SUMMARY

- The model has an accuracy of about 76%
- The model can predict those customers that convert to about 83%
- Lead score applied to customer data gives probability (on a scale of 0 – 100) that a customer will convert
- Top 3 features that contribute to a lead getting converted are:
 1. Lead Origin: Lead Add Form
 2. Occupation: Working Professional
 3. Lead source: Welingak Website



THANK YOU

David Lee & Satish Singh