Lead Score Case Study

Submission

Group Members:

1. Vigneshwari Sambandan

2. Davinder Khurana

Lead Score Case Study-Abstract

Business Objective:

An education company named X Education sells online courses to industry professionals. The company markets its courses on several platforms. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%. The company wants to increase its conversation rate by only following up with most probable leads.

Goals:

The aim is to identify the 'Hot Leads' to achieve effective conversation rates.

Strategy:

Calculate the lead score based on the probability of the leads and follow up hot leads with better lead scores.

Predictive Algorithm:

Logistic Regression

Assumptions and Imputations

- 'Select' value is considered as a placeholder in many categorical columns.

- All the Columns with more than 30% null values are dropped as they do not give much meaning to the predictive model.

- With columns which have less than 30% null values, calculate the dummy variables and drop the unknown value column.

- Convert Yes/No to 1/0 for categorical values.

- Train/Test Split is at 70/30 range.

Solution

- Important features are selected using RFE with 30 variables as the cut off.

- Features are removed successively by comparing high VIF(>5) and high P-values(>0.05).

- Ideal VIF's with < 5 and Ideal p-values with < 0.05 are achieved whilst maintaining high accuracy(>80%).

- Model is training using logistic regression and probability of the lead to get converted is also predicted with the calculation of the optimum threshold with good accuracy.

- Lead Score is calculated based on this predicted probability.

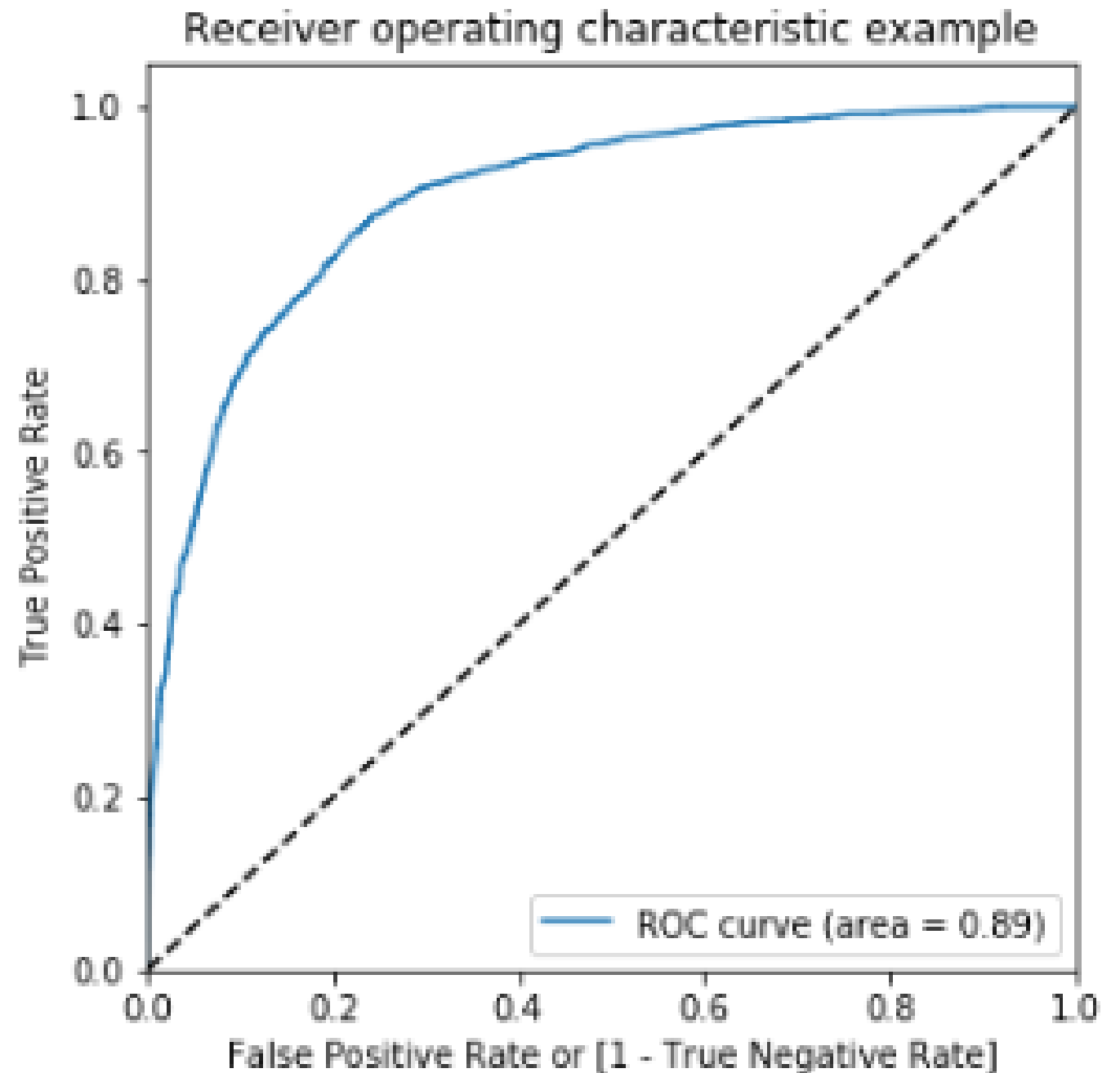- 'Hot Leads' are identified with this predicted score.

Accuracy of the Solution

ROC Curve:

The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.

It is seen that the area is wide (0.89)

Receiver operating characteristic example

ROC curve (area = 0.89)

Conclusion

- Lead Score is calculated based on the predicted probability of the converted leads.

- Conversion rates have been increased from 30% to 88.5% by following up with only these predicted leads.

- Lead Sources - Welingak website and by Reference have the major impact to get a lead converted and Working Professionals are the target audience as these are the top contributors in the model.