

Lead Scoring Case Study

Summary:

An education company named X Education sells online courses to industry professionals. Need to build a logistic regression model which is used by the management to make the process more efficient and understand how exactly the demands vary with different features and accordingly the management can manipulate the business strategy which is used in the conversion of lead to paying customers.

The data is prepared for analysis by using

- Exploratory data Analysis method.
- Feature Engineering (Dummy variables are created for categorical variables).
- Split the data into Training and Test dataset
- Model Building (Using GLM, p-value and VIF's)
- Finally selected the best model that used for converting.

The classes were assigned to all the leads in the test data set. For this, a probability cutoff of 0.5 was used. The model thus made, was very accurate (Accuracy = 90%), Sensitivity (84%) and Specificity (95%). The different cutoff was tried out, i.e. 0.3, which resulted in a model with almost same (Accuracy = 89%), but the sensitivity is more (~89%) and less Specificity (89%). Hence, we learnt that we should use 0.3 as the cutoff for probability for this model because the cutoff that equalizes accuracy, sensitivity and specificity.