Lead Score Case Study

The company ‘X Education’ wanted to improve its lead conversion rate from 27% to 80%.The company wanted to build a model which will predict lead score for every lead. Higher the lead score higher is the probability of the lead to get converted. The data given to us consists of 9240 rows and 37 features. We started with understanding and cleaning the data. We identified the columns containing null values and dropped the columns having more than 35% null values. Few columns still left having null values, which were important from the business point of view were imputed with ‘Not Available’. After cleaning the dataset we were left with 9074 rows and 22 columns to carry our analysis.

After carrying out the data analysis we found that, the probability of lead getting converted is high when ‘Lead Origin’ is ‘Lead add form’, Large number of leads come from Google and direct traffic but referral sites convert most of the leads it gets, means giving referrals can lead to more lead conversion, Leads opting for emailing option have more probability of getting converted. Conversion rate is higher when the information is sent through sms, Unemployed people have more conversion rate as well as more count.

We did model building using logistic regression and RFE, which gave 15 important features for prediction. The optimal cut-off came to be 0.35 and the final model gave the accuracy of 80% with 79% sensitivity and 80% specificity.