

Summary

X Education needs a way to help them filter most promising leads that will eventually buy products and services from them. Presently X Education gets leads from various sources but their conversion rate is just 30%. X Education wants to identify **Hot /Potential** leads in order to push their conversion rate beyond 80%.

The initial dataset consists of 37 variables, the first step was data preparation. Columns with very high null values were removed. In the next step, some columns having missing values (like Specialization, Current occupation, Tags), which appeared to be very important from the business point of view were imputed with appropriate values. Next a univariate analysis was performed, which gave many important insights about the data, and also helped in deciding the most important features for the model building process.

After the exploratory data analysis the dataset contained 14 variables, dummy variables were created and the data was split into test and train sets. After scaling of features the first model was built using the training dataset. Recursive feature elimination was used to reduce the number of features in the model. The final model was built using 11 features. Lead Origin, Lead Source, Last activity, Occupation, Tags etc are the features that contributed most in the prediction of leads that got converted. The final model had an accuracy score of 0.9. The area under the ROC curve is 0.95 showing an overall high accuracy. Accuracy sensitivity and specificity values for various probabilities were plotted and the cutoff probability value obtained was 0.3. A lead score was assigned to each lead based on the conversion probability. Finally predictions were made on the test set with an overall accuracy of 0.9.

- As all the leads have been assigned a lead score the sales team can focus on leads with very high scores, ensuing that the hottest leads are being approached first.
- Lead origin being an important predictor variable sales team can focus API and landing page submissions where the number of leads is high but conversion is very low.
- Google, direct traffic bring in many leads but conversion is low. On the other hand Welingak website is a source that can be used to increase the count of conversion.

- Sales team should work on devising attractive referral policies that would lure existing customers to refer more candidates/leads. We have seen high conversion rate for this category.
- We also observed that amount of time a lead spends on the website they have higher chances of conversion. Sales team should contact product manager to have them look at this aspect so that they can put in more effort on UX testing and to make overall platform experience more engaging.
- Sales team should be sending more promotional mails and messages to working professionals as these are the potential leads who get converted most.
- Majority of the leads come from Mumbai, hence Sales team should try to promote more aggressively in other cities as well to spread awareness.