

Lead Score Case Study Summary Report

X Education gets a lot of leads, its lead conversion rate is very poor. The typical lead conversion rate at X education is around 30%. X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Detailed summary report of Logistic Model:

Data Inspection and Cleaning: Provided data has 9240 entries with 37 Features, we have deleted columns with more than 40% on missing value. For rest of the columns, we have imputed the missing values with required data so that it can be helpful for analysis. Checks for outliers and treat them by capping and flooring technique.

Data Preparation: Then we have created dummy variables for categorical variable and map columns value with Yes\No to binary. Split the dataset into Train and Test Data with 70- 30 ratio. Used feature Scaling for numerical data to normalise the value so that model interpretation will be easy.

Model Building and Evaluation: Used RFE technique for feature selection to select top 20 features out of 140 columns. Then build a model and checked p-Value of all features, if it more than 0.05 simply drop that column one- by-one. Same step has been repeated till we get model where p-value looks significant and then calculate the VIF for reaming variables. If the VIF value > 5 then dropped the column and again build the final Model. Model 5 'm5' is the final model with 16 features. Then checked accuracy score which comes around 91.9%, Build the ROC curve with 0.97 curve area and gives cut-off at 0.33. With this cut-off accuracy, specificity and sensitivity are around 90% and we have also calculated Precision and Recall cut-off that gives value less than 90%. Hence, we stick with optimal cut-off 0.33 and create Lead Score as per the cut-off.

Making Prediction on Test Dataset: Feature scaling and predict the test data based on final model i.e, Model m5. Evaluate all the scores which is similar to train data i.e, 90% Assigned Lead score to test data as well.

Conclusion and Recommendation: We found out top-features that impacs leads conversion into HotLeads:

- Tags_Closed by Horizzon
- Tags_Lost to EINS
- Lead Source_Welingak Website
- What is your current occupation_Working Professional

Company should focus more on Working_Professional, and unemployed people looking to become relevant. There are high chances for them to convert into Hot Leads. Also Leads who spent more time on Website or were generated through the Welingak Website are more likely to convert into Hot Leads.

Company should spend more on advertising as it has also seen some promising lead conversion through Weblinks.