

SUMMARY

LEAD SCORING ASSIGNMENT

With this assignment we were able to build a logistic regression model that could help the X education company to convert the potential leads to their customer

Data understanding (Cleaning/Preparation): The data set provided with leads information was thoroughly checked and cleaned in order to prepare the data for modelling. EDA which goes as part of any modelling activity was also done to understand the behaviour/impacts of few variables of interest.

Model Building: process included various steps like the test-train split (30/70), creating dummy variables of the categorical data, scaling of numeric variables using min-max/standard scalar (we have used min max in our assignment), RFE to select 15 features was applied after which the VIF and P-values were constantly checked to achieve an acceptable value of $VIF \leq 5$ and $P\text{-value} < 0.05$. We were able to boil down to 11 features after the iterative process.

Model Evaluation: Further to this we create confusion matrix to calculate accuracy, sensitivity, specificity for the train data. We also plotted the ROC curve which evidently showed the AUC value as 0.86. The optimal cut off probability value was plotted. Choosing this optimal value however varied from business to business.

Prediction: We then moved to predict on the test data. Where we again checked for confusion matrix along with calculating relevant metrics like sensitivity and specificity. Here we introduced two more metrics namely Recall(how any right predictions happened out of the yes cases present in the data) and Precision(Probability that a predicted “yes” will be actually a yes).

Final list of features (both positive and negative co-eff) that we could come up with were:

- Total visits
- Total time spent on websites
- Lead origin: Lead add forms
- Lead source-Olark chat and -Welingak website
- Last activity phone conversation and SMS sent
- Last notable activity-Unreachable
- Current occupation-Unemployed and students (*Negative*)
- Do not email_YES (*Negative*)

With the above stated features, X-Education company can strategize to reach to their targets (potential leads who would get converted to customers).