

LEAD SCORING CASE STUDY

Logistic Regression Assignment Submission

Student Name: Tarriq Ferrose Khan

Batch : DSC 68

upGrad

upGrad & IITB | Data Science Program - May 2024

Summary Report:

1. Source Data (Leads.csv) was analyzed for the Lead Scoring Case study, which expects to score the Prospective Leads and propose a Lead acquisition strategy to increase the efficiency for a target conversion rate of 80%
2. Data was analyzed for quality and handled for NULL value , value replacements like 'Select'- Ui default values and columns with Yes/No are replaced as 1/0 respectively.
3. Few unwanted columns that are system generated and/or will not contribute for modelling are cleaned up.
4. With the remaining categorical variables dummy variables are created and redundant Categorical columns were dropped.
5. Continuous Variable Columns like Total Visits, Website_Time_spent, Pages Per visit were checked for outliers and removed.
6. Test(30%) and Train data (70%) using Test_Train_split module was done.
7. The continuous variable columns were standardized using Min/Max scaling as they were of different ranges.
8. Model was assessed using statsmodel.api and columns with High VIF (Variable Inflation Factor) and high p-values were removed.
9. Once all the columns are good proceed with Model Evaluation using the ROC, Accuracy , Specificity and Sensitivity metrics , to arrive at a Cut off basis which the Converted Probability is calculated.
10. This model also evaluated using Precision Recall to double check the fitness of the model.
11. Kindly refer to the PPT(PDF) shared for the details on the graphs, Proposed Hot Lead Scoring and the Potential Lead Acquisition Strategy.

-x-End of Document-x-