

Lead Score Case Study Summary

The model building was , started by importing the important Python libraries,after that first step was read the data and inspect is properties like shape, information and numerical data features than clean it in the process we find there are many columns with higher percentage of null values so we dropped the columns having more than 45% of it the impute the rest of the missing values , we also dropped system created columns as they were of no use in the analysis, we also dropped the columns with highly skewed data . After that we started EDA by doing Univariate Analysis of the categorical and numerical columns separately in which we treated outliers using capping and also get useful insights from we also dropped the columns which were not adding any information , after we performed Bi-variate analysis using pairplot in which we found lot of data is skewed and we have noise in the data than we check multicollinearity of the data , after that we started preparing data for modeling by creating dummy variables for categorical variables having more than 2 categorical features , than we split the train set and test set in the ratio of 80 and 20 respectively,than we scale it using StandardScaler. Than we make logistic Regression model on using statsmodel library ,bye using RFE we selected 18 features from which the one with higher p value and VIF are eliminated one by one overall 9 models were created the last model have all the p values less than 0.003 and VIF under 5 after that we assign the predicted probability and converted mark on the cutoff value of 0.5 than accuracy , sensitivity , specificity than we find the optimal cutoff probability using ROC curve and plotting accuracy , specificity and sensitivity which comes out to be 0.35 than make prediction of the test set using the same columns as the last model and find out test data set statistics which comes as follows

Train Data Accuracy :80.95 %

Train Data Sensitivity :80.73 %

Train Data Specificity :81.08 %

Train Data F1 Score :0.72

Test Data Accuracy :81.54 %

Test Data Sensitivity :80.73 %

Test Data Specificity :81.08 %

Test Data F1 Score :0.74

On the basic of which we gave following recommendations to the X education

- Increase user engagement on Welingak website since this helps in higher conversion
- Focus on Working Professional which has high conversion certainty.
- Get Total Time Spent on Website increased by advertising and user

experience which makes the customer engaging in the website. since this helps in higher conversion

- Improve the Olark Chat service since this is affecting the conversion negatively
- Improving Lead add form also improves the lead conversion with high certainty