

Summary

Below are the major steps taken to proceed with our assignments:

1. Data Cleaning: -

- A) First step to clean the dataset we choose to remove the redundant variables/features.
- B) After removing the redundant columns, we found that some columns are having label as 'Select' which means the customer has chosen not to answer this question. The ideal value to replace this label would be null value as the customer has not opted any option. Hence, we changed those labels from 'Select' to null values.
- C) Removed columns having more than 45% null values
- D) For remaining missing values, either we have imputed values with maximum number of occurrences for a column or we have created new columns as pre required.
- E) Also we have dropped missing rows from the dataset.
- F) Handle outliers

2. Data Preparation: -

- A) Changed the multcategory labels into dummy variables and binary variables into '0' and '1'.
- B) Removed all the redundant and repeated columns
- C) Split the dataset into train and test dataset and scaled the dataset
- D) Scaling the dataset.

3. Model Building: -

- A) We created our model having RFE with 15 variables and considering minimum number of VIF and p value and chose our final model with 10 features with more stability and accuracy.
- B) For our final model we checked the optimal probability cutoff by finding points and checking the accuracy, sensitivity and specificity. And also found one convergent points and we chose that point for cutoff and predicted our final outcomes.
- C) We checked the precision and recall with accuracy, sensitivity and specificity for our final model and the tradeoffs. Prediction made now in test set and predicted value was recoded.
- D) We did model evaluation on the test set like checking the accuracy, recall/sensitivity to find how the model is. And found the score of accuracy and sensitivity from our final test model is in acceptable range.
- E) We have given lead score to the test dataset for indication that high lead score are hot leads and low lead score are not hot leads.

4. Conclusion:

X-Education has a better chance of converting a potential lead when:

- ▶ **The total time spent on the Website is high:** Leads who have spent more time on the website have converted.
- ▶ **Current Occupation is specified:** Leads who are working professionals have high chances of getting converted. People who were looking for better prospects like Unemployed, students and Business professionals were also good prospects to focus on.
- ▶ **When the Lead origin was Lead Add form** Leads who have responded/ or engaged through Lead Add Forms have had a higher chances of getting converted
- ▶ **Number of Total Visits were high** Leads who have made a greater number of visits have higher chances of getting converted.
- ▶ **When the last activity was SMS sent or Email opened** Members who have sent an SMS for enquiry or who have opened the email have a higher chance of getting converted.
- ▶ **Approaching the management specialized leads, will produce more conversion.**
- ▶ **Almost all the leads given tag 'Will revert after reading the email' gets converted, they have a high chance of conversion.**