Lead Scoring Case Study Summary

An education company named X education sells online course to industry professionals. The company wants to identify the most promising leads which are most likely to convert into paying customers. Company needs a logistic regression model for the same. Further assigning a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

The following steps are used in building a logistic regression model.

1.Reading and understanding data

The dataset was loaded in python notebook. Data structural check on data is done.

2. Data cleaning

In the data, some variables have a level called 'Select' which is same as null value and is to be replaced by null value. Variables having more than 40% of null values are removed. This step also include filling up less percentage of null values by median in case of numerical column and by mode in case of categorical variable. Unnecessary and zero-variance columns which are not useful in analysis are removed.

3. Visualization of data

In this Step did univariate, bivariate and multivariate analysis of variables.

4. Data Preparation

In this step create dummy variables for categorical variables and use drop first as True.

5. Train -Test Split

Here data is divided into 70% and 30% for train and test data respectively.

6. Feature Scaling

Min-Max scaling is used to scale the numerical variables.

7. Model Building

Firstly the automated approach of Recursive Feature Elimination (RFE) is applied to find the top 15 important features. Further manual approach is applied using the statistics generated by recursively eliminating variables having P-value more than 0.05 and VIF's more than 5.

8. Model Evaluation

In this step finding the predicted value on train set using the final model. Confusion Metric was made. Optimum cut-off value was obtained using ROC curve. Later we find out the accuracy, sensitivity and specificity.

9. Prediction

Finally prediction was done on test data frame with optimal cut-off value. And accuracy, sensitivity and specificity are calculated.

Conclusion of analysis done –

It was found that Variables which contribute most towards the probability of a lead getting converted are -

- Total Time spent on website
- Current Occupation-Working Professional
- Lead Origin-Lead Add Form
- Target lead on the basis of Total time spent on Website, as visitors who are spending more time on website are more interested in the course.
- Sales team should contact lead whose current occupation is working professional. Since this course is specially designed for working professionals and there might be some professionals looking for such course to upgrade their skills or profile.

- Sales team should also try to contact visitors on the basis of Total Visits. As more visits clearly indicates that the person is interested in the course that's why visits website several times for getting complete knowledge about the course.

Keeping in mind the above mentioned point X Education can increase their conversion rate.