Lead Scoring Case Study Summary

Problem Statement:

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Solution:

After going through and understanding the business problem, we came go to know that this is classification problem. Our approach is to use Logistic Regression

Steps involved:

1. **Data Reading and Understanding**

   The data is imported. We check how data looks, like we will see shape, info, data type and data spread.

2. **Data Cleaning**

   We check the missing Null values, here missing null values are imputed. Also drop all the columns having missing values greater than 30%. The check for outliers and removal is done here. The check on data skewness was done and highly skewed columns were dropped.

3. **Data Analysis**

   Univariate analysis is done. Also the check on data skewness was done and highly skewed columns were dropped.

4. **Data Preparation**

   Conversion of some binary variables (Yes/No) to (1/0) is done. Dummy variables was created for Categorical variables, also redundant variables were removed

5. **Test Train Split**

   The data set was divided into Train and Test set (70% - Train set and 30% - Test set)

6. **Feature Scaling**

   Feature scaling is done via Standard Scaler. Check for high correlation variables and remove 2 variables in this step

7. **Model Building**

   Variable selection is done using RFE, in which 15 variables were selected. We check the pvalue to select the most significant that should be present. In this step we dropped 3 variables with pvalue greater that threshold i.e. >0.05(4 repetitive steps). Next we checked VIF of each variable and found to good.

8. **Plotting Roc Curve**

   ROC curve was plotted, where we got ROC curve area coverage around 87%

9. **Finding Optimal cut off Point**

   We plotted graph for Accuracy, Sensitivity and Specificity for different values. The cut off value was found to be 0.35 from the intersection in the graph. We got the values for Accuracy – 80%, Sensitivity-80% and Specificity-80%

10. **Precision and Recall Matrix**

    Precision and Recall value were observed to be 79% and 64% respectively. We got cut off value of 0.43 from precision and recall trade off

11. **Making Predictions On Test Set**

    For the test model we got the values for Accuracy – 79%, Sensitivity-80% and Specificity-79% approx.

Conclusion

1. **The target lead conversion rate is around 80%**
2. **The Accuracy of the model is around 80%**
3. **Accuracy, Sensitivity and Specificity of both Test and Train have closer values**

Important variables which positively contributed for lead conversions are:

1. **What is your current occupation_Working Professional**
2. **Last Activity_Had a Phone Conversation**
3. **Total Time Spent on Website**
4. **Last Activity_SMS Sent**

Recommendation:

1. **Working professionals should be targeted.**
2. **Target the website as more time is spent on website are more likely to get converted. Allocate some budget on the website**
3. **It's good to send SMS along with phone calls**