LEAD SCORE CASE STUDY

SUMMARY REPORT

BY:

DUVVURI SURYATHEJA REDDY

GAURAV RASAL

PROBLEM STATEMENT:

A company named X Education sells online courses to industry professionals. Professionals who are interested in the courses land on their website,once these people land on the website, they might 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. We are required to build a model wherein we need to assign 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 CEO,in particular,has given a ballpark of the target lead conversion rate to be around 80%.

SOLUTION METHODLOGY AND INSIGHTS:

- Imported data,did cleaning and transformation of data.
- EDA is done on columns and insights are:
    - Most of the Page views per visit in converted are in range between 0 and 2.
    - Most of the customers who are converted are unemployed
    - Most of the customers who are converted got 'Will revert after reading the email' tagged.
- Train-test split is done.
- RFE is done on training data and top 15 features are selected after which a model is built using these 15 features using 0.5 as threshold.
- Different models are created by dropping columns based on p values and VIFs and recreating the model with other existing columns. Final model is obtained with statistically significant columns.
- Different evaluation metrics are calculated,and ROC curve is plotted to check stability of our final model. Area of ROC is 0.95 which indicates that model is stable. Also,Precision-recall curve and sensitivity-specificity curve are plotted
- Optimal threshold(0.3) is selected using sensitivity-specificity curve due to business objective being focused on sensitivity.Predictions on test set is done finally.

MODEL SUMMARY:

- Accuracy for Test set: 90.8%
- Accuracy for Training set: 91.0%
- Sensitivity for Test set: 85.4%
- Sensitivity for Training set: 85.8%
- Specificity for Test set: 93.9%
- Specificity for Training set: 94.2%
- Top three variables in model,that contribute to lead conversion are:
    - Last Notable Activity_SMS Sent
    - Tags_Will revert after reading the email
    - Lead Origin_Lead Add Form

RECOMMENDATIONS TO THE COMPANY:

- Company should focus on sending more SMS,since this helps in higher conversion.
- Company should focus on lead add form since customer identification by that produces more conversion
- Also,company should focus on the customers whose current status is 'Will revert after reading the email',so customers who are tagged by this type must be monitored since there is a high potential for these type of customers for lead conversion
- Company should improve its techniques for analyzing quality of lead as it is negatively impacting the conversion
- Company should improve Olark chat service since it is negatively impacting the conversion