LEAD SCORE CASE STUDY SUMMARY

Problem Statement:

X Education is an organization which provides online course to Industry professionals. The company markets its Courses on several websites and search engines like Google.

X Education wants to select the most promising leads that can be converted to paying customers.

Although X education gets a lot of leads, its lead conversion rate is very poor, whereas the company wants a higher lead conversion. Leads come through numerous modes like email, advertisements on websites, google search etc.

The company has had 30% conversion rate through the whole process of turning lead into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating is not efficient in helping conversion.

Solution Summary:

i) **Read and Understanding Data:**

Read and Inspect Data.

ii) **Data Cleaning:**
- First step to clean the dataset we chose was to drop the variables having unique values.
- Then, there were few columns with the value 'Select' which means the leads did not choose any option. We changed those values to Null values.
- We drop the columns having NULL values greater than 3000 missing values.
- Next, we removed the imbalanced and redundant variables. This step also included imputing the missing values as and where required with median values in case of numerical variables and creation of new classification variable in case of categorical variable. Also, column names have been fixed with underscore in between letters converting all column names to small letter case.

iii) **Dummy variable Creation:**
  a. We created dummy variables for the categorical variables.
  b. Removed all the repeated and redundant variables.

iv)   Test Train Split:

- The next step is to divide the data set into test and train sections with a proportion of 70 – 30 Percentage values.

v)   Feature Rescaling:
- We used the Min Max Scaling to scale the original numerical variables.
- Then, we plot Correlation heat map to check correlation among variables. Also, the number of variables are pretty high, heat map of correlation not required.

vi)   Model Building:
- Using the RFE(Recursive Feature Elimination), we went ahead and selected the top 15 Important features.
- Using the statistics generated, we recursively tried looking at the P-values in order to select the most significant values that should be present and dropped the insignificant values.
- Finally, we arrived at the most significant variables. The VIF for these variables were also found to be good.
- For our final model we checked the optimal probability cut off by finding points and checking accuracy , sensitivity and specificity.
- We then plot ROC curve to finding optimal cutoff, the features and the curve came out be pretty decent with an area coverage of 86 % which further solidified the model.
- Then, created different probability cutoffs.
- We checked the precision and recall with accuracy, sensitivity and specificity for the model on train set.
- Next, based on the on the precision and recall trade-off, we got a cut off value of approximately 0.42.
- Then we implemented the learning to the test model and calculated the conversion probability based on the sensitivity and specificity metrics and found out the accuracy value to be 79%; sensitivity 78.40% and specificity 77.71%.
- The precision and Recall used to recheck and cut off of 0.42 was found with precision around 78.30 % and recall around 76.74% on the final data frame.

vii)   Conclusion:
- There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion. First, sort out the best prospects from the leads you have generated. 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the

probability of a lead getting converted. Then, You must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads. A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.