Lead Scoring Case Study Summary

A brief summary report in 500 words explaining how you proceeded with the assignment and the learnings that you gathered.

Answer:

After Importing the data in Jupiter notebook, below are the steps summary how we have proceeded with our assignments:

1. Data Cleaning:

- First step to clean the dataset we choose to remove the redundant variables/features.
- After removing the redundant variables, we found that some variables 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.
- Removed columns having more than 30% null values.
- For remaining missing values, we have imputed values with maximum number of occurrences (mode) for a column.
- We found for one column is having two identical label names in different format (capital letter and small letter). We fixed this issue by changes the labels names into one format.

2. Data Transformation:

- Changed the multi category labels in to dummy variables and binary variables into '0' and '1'.
- Checked the outliers and created bins for them.
- Removed all the redundant and repeated columns.

3. Data Preparation:

- Split the dataset into train and test dataset and scaled the dataset.
- After this, we plot a heatmap to check the correlations among the variables.
- Found some correlations and they were dropped.

4. Model Building:

- We created our model with RFE count 15 and compared the model evaluation score like AUC and choose our final model with RFE 15 variables as it has more stability and accuracy than the other.
- For our final model we checked the optimal probability cut-off by finding points and checking the accuracy, sensitivity and specificity.
- We found one convergence point and we chose that point for cut-off and predicted our final outcomes.
- We checked the precision and recall with accuracy, sensitivity and specificity for our final model and the trade-offs**.**
- Prediction made now in test set and predicted value was recoded.
- We did model evaluation on the test set like checking the accuracy, recall/sensitivity to find how our model is accurate.

- We found the score of accuracy and sensitivity from our final test model is inacceptable range.
- 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.

5. Conclusion:
Below are the learnings which we gathered from final model.

- Test set is having accuracy, recall/sensitivity in an acceptable range.
- In business terms, our model is having stability an accuracy with adaptive environment skills. Means it will adjust with the company's requirement changes made in coming future.
- Top features for good conversion rate:

  1. Lead Origin_Lead Add Form

  2. Tags_Will revert after reading the email
  3. Total Time Spent on Website