Leads Scoring Case Study

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

Answer:

Below are the steps how we have proceeded with our assignments:

1. **Data Cleaning:**
    - In the First step, we cleaned the dataset by choosing to remove the redundant variables/features.
    - Post removing the redundant columns, we found out that some columns are having label with 'Select' as its name which means the customer has not chosen any 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.
    - Next we removed columns that are having more than 30% null values in the data set.
    - For the remaining missing values, we have imputed their values with the maximum number of occurrences for that particular column.
2. **Data Transformation:**
    - First in Data Transformation we have changed the multicategory labels into binaryvariables into '0' and '1'.
    - Next we checked the outliers and created bins for each categories.
    - Then we removed all the redundant and repeated columns in the data set.
3. **Data Preparation:**
    - First in Data Preparation we split the dataset into train and test dataset and scaled the following dataset.
    - After this, we went ahead to plot a heatmap to check the correlations among thevariables in the data set.
    - We had found some attributes with correlations and they were dropped from the data set.

4. **Model Building:**
    - We created our model with rfe count 19 and 15 and compared the model evaluation score like AUC.
    - For our final model created we checked the optimal probability cutoff by finding points and checking the accuracy, sensitivity and specificity.
    - We found out that there is one convergent points and we chose that point for cutoff and predicted our final outcomes.
    - We checked the precision and recall with accuracy, sensitivity andspecificity for our final model and the tradeoffs**.**
    - Prediction made now in the test set and predicted value was recorded.
    - We did a model evaluation on the test set (checking the accuracy,recall/sensitivity to find how the model is).
    - We found out that the score of accuracy and sensitivity from our final testmodel is in an acceptable range.
    - We have given a lead score to the test dataset for an indication that highlead scores are hot leads and low lead scores are not hot leads.

5. **Conclusion:**

The learning gathered are below:

i. Test set is having accuracy, recall/sensitivity in an acceptable range.
ii. In business terms, our model is having stability and accuracy with adaptive environment skills. This means it will adjust with thecompany's requirement changes made in the coming future.
iii. Top features for good conversion rate:
    1. **Last Notable Activity_Had a Phone Conversation**
    2. **Lead Origin_Lead Add Form**
    3. **What is your current occupation_Working Professional**