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:

These are the steps we took to complete our assignments:

1. **Data Cleaning:**
    a. The redundant variables and features were removed as the first stage in cleaning the dataset we selected.
    b. After eliminating the unnecessary columns, we saw that certain columns had the label "Select," indicating that the consumer had decided not to respond to this question. Since the consumer has not chosen any options, the appropriate value to replace this label would be null value. Thus, we modified those labels from 'Select' to null values.
    c. Deleted columns with more than 30% of the values being null.
    d. For remaining missing data, we have imputed values with maximum number of occurrences for a column.
    e. We discovered that two identical label names in a different format were present for one column (capital letter and small letter). By changing the label names to one format, we were able to resolve this problem.
2. **Data Transformation:**
    a. Changed the multicategory labels into dummy variables and binary variables into ‘0’ and ‘1’.
    b. Checked the outliers and created bins for them.
    c. Removed all the redundant and repeated columns.
3. **Data Preparation:**
    a. Scaled the dataset and divided it into train and test datasets.
    b. Next, we create a heatmap to examine the relationships between the variables.
    c. Several correlations were discovered but were later deleted.

4. **Model Building:**
    a. We built our model with rfe counts of 19 and 15, evaluated the model assessment scores like AUC, and chose our final model with rfe counts of 19 as it is more accurate and stable than the other..
    b. For our final model, we determined points and evaluated the accuracy, sensitivity, and specificity to determine the best probability threshold.
    c. We identified a convergent point and used it as the cutoff point while projecting our final results.
    d. For our final model and the tradeoffs, we examined the precision and recall along with accuracy, sensitivity, and specificity.
    e. The predicted value was recoded after the prediction was made in the test set.
    f. In order to evaluate the model, we checked its accuracy, recall, and sensitivity on the test set.
    g. We discovered that our final test model's accuracy and sensitivity scores fall within a reasonable range.
    h. We assigned a lead score to the test dataset to demonstrate that leads with a high lead score are hot leads and those with a low lead score are not.

5. **Conclusion:**

    The findings are below:

        i. The test set has acceptable accuracy, recall, and sensitivity.
        ii. In terms of business, our model is stable and accurate with environment-adaptive abilities. This indicates that it will adapt to any future changes in the company's requirements.
        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**