Summary:

X Education, an online course provider for industry professionals, sought assistance in identifying the most promising leads likely to convert into paying customers. The CEO set a target lead conversion rate of approximately 80%.

The solution involved a multi-step process:

1. **Data Understanding and Cleaning**: The dataset was examined, and steps were taken to clean it, including dropping unique value variables, handling 'Select' values as null, and removing columns with high null percentages. Imbalanced and redundant variables were addressed, outliers identified and removed, and label inconsistencies corrected.
2. **Data Transformation**: Binary variables were converted into '0' and '1'.
3. **Dummy Variables Creation**: Dummy variables were created for categorical variables, removing duplicates and redundancies.
4. **Test-Train Split**: The dataset was divided into 70-30% train-test proportions.
5. **Feature Rescaling**: Min-Max Scaling was used for numerical variables, and highly correlated dummy variables were dropped.
6. **Model Building**: Recursive Feature Elimination was employed to select the top 15 important features, with further refinement through P-value analysis. The final model comprised 11 significant variables, with optimal probability cutoff determined through accuracy, sensitivity, and specificity evaluations. The ROC curve demonstrated an 86% area coverage.
7. **Conclusion**: The final model predicted a lead conversion rate of 83% on the test dataset, exceeding the CEO's target. The model's high sensitivity indicates its effectiveness in identifying promising leads. Key features influencing conversion probability included lead origin via Lead Add Form, current occupation as a Working Professional, and total time spent on the website.

Overall, the developed lead scoring model not only met but exceeded the CEO's expectations, providing a robust framework for identifying high-conversion potential leads and optimizing sales efforts.

Top of Form