SUMMARY

X Education's current lead conversion rate is approximately 30%, but the CEO's target is to achieve an 80% lead conversion rate. To meet this goal, we need to develop a lead scoring model that assigns higher scores to leads with a greater likelihood of conversion. This scoring system will help prioritize leads with higher conversion potential.

In the data, we handled missing values and categorical columns by dropping those with >40% nulls, creating "Others" categories, imputing high-frequency values, and removing low-value columns. For numerical categorical data, imputed with mode and dropped single-response columns. Additional preprocessing steps included outlier treatment, data validation, grouping low-frequency values, and mapping binary categorical values.

Checked data imbalance (38.5% conversion rate) and conducted EDA, highlighting influential variables like 'Lead Origin,' 'Current Occupation,' 'Lead Source,' and the positive impact of website engagement on lead conversion.

Encoded categorical variables, split data into a 70:30 train-test ratio, and standardized features while removing highly correlated columns.

Used RFE to reduce variables to 15, performed manual feature reduction based on p-values, built three models, and selected "logm4" with 12 variables as the final model for train and test predictions.

The confusion matrix led to the selection of a cut-off point at 0.345, balancing accuracy, sensitivity, and specificity with all around 80%. However, when aiming to boost conversion to 80%, metrics dropped with the precision-recall view, so the sensitivity-specificity view was chosen for the optimal cut-off. A lead score of 0.345 was assigned to the training data for final predictions.

Scaled and predicted on the test set with the final model, achieving around 80% evaluation metrics. Assigned lead scores, and identified top 3 influential features: Welingak Website, Reference, and Working Professional as Current Occupation.

Allocate more budget to promote the Welingak Website, incentivize referrals that convert, and target working professionals aggressively due to their higher conversion rates and financial capacity. Focus efforts on converted leads by conducting question-answer sessions to gather valuable insights into their intentions and attitudes toward online course enrollment. Use subsequent inquiries and appointments to gain further insights, ultimately aiding in the conversion of leads into loyal customers.