

Summary

Data:

The data initially had many nan values and redundant/extra columns which had to be removed directly. Some columns were imputed with modes and categorical columns where categories didn't meet a certain numerical threshold were dropped.

Predictions:

The prediction was done with a logistic regression model using RFE to eliminate features at first, then manually removing features until we got a model with low VIF and p-values. We got it at iteration 5.

Statistics:

On test dataset we had an Accuracy of 80%, Sensitivity of 82% and Specificity of 80%, showing good promise in identifying leads efficiently.

Insights:

Professionals, and people who use Welingak websites and sources acquired through reference are good candidates.