Summary Report

The case study aims to build a machine learning model to assign a lead score between 0 and 100 to each lead, which the company can use to target potential leads.

The first step is the EDA. The next step is data preparation. In this step, dummy variables are created. After that, variables are split into test and train sets. After that, feature rescaling is performed. After this, two models are built. First is the logistics regression model and decision tree model.

The decision tree model has a high specificity on the test set, but this is at the expense of accuracy. The decision tree model is also overfitting the training set, as evident by the fact that the accuracy on the training set is 93%, much higher than the 73% accuracy on the test set.

For the logistic regression model, the Accuracy, Sensitivity and Specificity values of the test set are around 77%, 83% and 74%, which are approximately closer to the respective values calculated using the trained set.

Hence, the logistic regression model performs better than the decision tree model in the ‘lead scoring’ dataset.