

Summary

The “Lead Scoring Case Study” assignment proved to be an awesome hands-on experience on python coding and model building and evaluation. Moreover it has also improved our presentation skills. We proceeded by initial data analysis and cleaning which included missing values treatment and outlier treatment followed by exploratory data analysis because of which we were able to find hidden insights in the data.

After that we went ahead with the data preprocessing step as this is a crucial step before building any kind of model. The process involved creation of dummy variables, splitting of data into train and test data, scaling of numerical features (in this case we used min max scaling).

The next important step was to select the independent features, for this we used a technique called Recursive Feature Elimination (RFE) which helped us select initial 15 features to start our model building process. We built a few models keeping an eye on the p-values of the independent variables along with a very important value called the VIF i.e. Variance Inflation Factor which tells us how our independent variables are correlated with each other.

After analyzing the p-values and the VIF we got our set of 14 independent features which seemed to be perfect for our model building process. Using these 14 variables a Logistics Regression model was fit on our train data set and find out conversion probability values, which was then followed by finding the optimal cut-off value to differentiate between converted and not converted. This value came out to be 0.3 in our case.

The next step was to evaluate our model based on certain evaluation metrics such as accuracy, sensitivity, specificity, precision, recall. All these evaluation metrics were evaluated with the help of confusion matrix.

After model evaluation it's time to predict the conversion probabilities on our test data set and calculate the lead scores for each prospect id.

By following the above mentioned approach we were able to determine the top variables which contribute most towards the probability of a lead getting converted.

Working on this assignment was like working on a real world industry problem which boosted the confidence in me as well as my team.

Really looking forward to work on similar projects in the future as well.

Thank You

Jyoti Sharma

Bimla Bisht

Sumit Bhatia