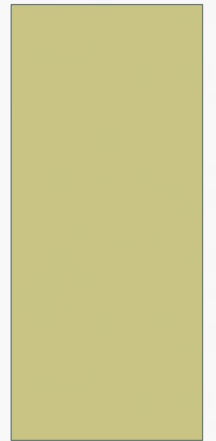


LEAD SCORING CASE STUDY

RISHABH SALEKAR



EDA

- Performed EDA on Leads.csv.
- Started out by gauging out the amount of null values present.
- Dropped those columns which have features which too much and hence are inconsequential to our final model.
- Also dropped Rows as and when required.
- Converted Categorical Data into Dummy Variable as and when required
- Scaled Numerical values as well.

MODEL BUILDING

- After the EDA ended with 74 Features.
- In order to reduce further used RFE to evaluate which feature matter and which don't.
- This gave me 16 Features.
- Use Stats model for Logistic Regression to understand the p-value and VIF.
- After several iterations got a decent VIF and p-value and 11 features.

MODEL EVALUATION

- Applied LR model and got results as set of probabilities. Converted these probabilities to Discrete value by keeping 0.5 as cut of.
- Plotted ROC to get a good cutoff criteria. That came to 0.42.
- Again got a Final Discrete value conversion by keeping 0.42 as cutoff.
- After evaluating the model again got a better accuracy at 0.79.
- Finally, evaluated the trained model on Test Data and getting the accuracy at around 0.78.