

Lead Scoring Case Study Summary

- Tags and Lead Source and What is your current occupation and Last Notable activity are the top 4 variables which contribute most towards the probability of a lead getting converted.
- Country and What matters most to you in choosing a course and Last Notable Activity are the 3 variables which should be focused for the most in order to increase the probability of lead conversion.
- All variables have a good value of VIF. So we need not drop any more variables and we can proceed with making predictions using this model only
- Accuracy at threshold 0.4 is highest.
- A common way to visualize the trade-offs of different thresholds is by using an ROC curve, a plot of the true positive rate (true positives/ total positives) versus the false positive rate (false positives /total negatives) for all possible choices of thresholds.
- A model with good classification accuracy should have significantly more true positives than false positives at all thresholds.
- The optimum position for roc curve is towards the top left corner where the specificity and sensitivity are at optimum levels

Area Under The Curve (AUC)

- The area under the ROC curve quantifies model classification accuracy, the higher the area, the greater the disparity between true and false positives, and the stronger the model in classifying members of the training dataset.
- An area of 0.5 corresponds to a model that performs no better than random classification and a good classifier stays as far away from that as possible. An area of 1 is ideal.
- The closer the AUC to 1 the better.