LEAD SCORING CASE STUDY SUMMARY REPORT:

Learnings and Challenges Faced:

- Process Followed: First I inspected the Data Dictionary and tried to make sense(intuitively) of the variables that could go into the model.
- Visualized or comprehended the EDA process that should go into the model. The main challenge faced was to reduce the levels of categorical variables so that we have less number of dummy variables. This was done by using a data dictionary for respective categorical fields. This was time consuming.
- Also deleted /dropped the least important category while creating dummies for categorical variables. Eg “Others” was the category which was dropped explicitly rather than using “drop_first=True”.
- Another Learning is we just cannot strictly follow the low p_value and High VIF value to be dropped. Along this process we also need to have business sense as to which variable needs to be dropped and what information along with that variable is getting dropped from the model. The best business sense of variables to be included can be got by drawing a heat map of all the features/variables against the Target variable. This gives solid intuition while entering the feature selected process.
- Also the model should be open to scale and change when business requirements change, we should develop the model keeping this in mind.
- Based on business understanding we should either work towards increasing sensitivity or precision. In Our case it was increasing sensitivity while not letting go of Specificity by large values.

Submitted by Achal Kagwad and Surbhi Chaplot.