Summary Report

X Education was facing the problem with their lead conversion rate of 30%. We created a logistic regression model which used 17 features to accurately predict whether a candidate will be converted on not. We were able to achieve an accuracy, specificity and sensitivity of 0.9 in the test set we created. Note that the test set was created at random using 20% of the given data. We where able to find some discrepancies in the given data like, countries other than India was using a city value of ‘Mumbai’. Fixing these kinds of issues can improve the data quality and hence improve the model. Also a lot of missing values was identified and processed using imputations. If we could make a lot of those fields mandatory to fill, we’ll have more data which will again improve the model. There are a lot of categorical values with less than 5-10% occurrences. We had to group them for reducing the model complexity. As the data increases, we can retrain the model with additional categories for these features. The model also explains the influence of different features in the final decision making. This information can also be used to improve the customer engagement while contacting them.