INTRODUCTION

The Case Study has helped me to fully understand the industrial application of Machine Learning Algorithms. The case study presented me with a business problem that need to be resolved using Machine Learning. Below are the steps I took in resolving the problem:

1. Data Engineering

Firstly, I checked for null values from the data set. There were several columns with a lot of null values and imputing these columns would have resulted in data in-balances. I dropped all columns with a lot of null values.

I imputed missing values for the remaining columns with median for numerical variables and mode for categorical variables.

2. EXPLORATORY DATA ANALYSIS

I tried to understand the data through EDA. I plotted several columns on different types of graphs to understand the distribution of values. This helped me to check if there are outliers in numerical columns.

I treated all outliers accordingly by checking the changes in values at each quartile. EDA also help me to identify highly correlated variables. All highly correlated columns were dropped.

3. Building the Model

Firstly, I created categorical dummy variables. Then I used Standard Scaler to standardize the features. I used Logistic Regression to build my model and I used RFE to select top 15 features for my model.

From the 15 features RFE selected for me, I checked P-Values for each feature to see the significance of the variable. I dropped the features with high p-values. I also checked the VIF for all features, all features with VIF greater than 5 were dropped.

I used several metrics to check how good my model is on Train Data. I checked Accuracy, Sensitivity, Specificity, Precision and Recall. I also plotted the model on ROC Curve. These metrics were also plotted on a graph to come up with optimum cut-off point.

Finally, I made prediction on Test Data.

4. Deriving Business Explanation from Model outcome

The Model help me identify variables which contribute most towards the probability of a lead getting converted. There variables were identified by checking the coefficient on model's statistical summary.

The model was also able to predict if a lead is most likely to be converted or not, this can help sales team to improve the efficiency of their lead conversion process because they can only concentrate on communicating to lead that have high probability of converting.

In summary the model will improve conversion rate.

CONCLUSION

I work in telecom industry and this case study has opened my eyes on how to approach business problems. There are a lot of business cases that can be improved or resolved using ML.

One of the most valuable lessons that I learned from the case study is that Logistic Regression can be used to predict the probability of an event occurring. This probability can then be used to make decisions, like in our case; whether to contact a lead or not. In addition to the concepts of logistic regression, I also learned how to implement logistic regression in python.