Abstract:

In this project objective is to improve given portugese bank’s tele marketing campaign efficiency by identifying socio economic attributes of Customers as the driving factor for term deposit product selection. Cross Industry Data Standard Process for data mining(CRISP DM) framework has been used in this project. With this approach Business case understanding was the first step, followed by data exploration, data preparation, modeling, evaluation and recommendation from final model.

In given data set 16 variables related to Customers socio economic conditions have been analyzed for around 41188 Customer records. Three different models have been used in this project- Logistics Regression, Classification tree, RandomForest for classification of binary variable campaign response. Several criteria have been used for evaluation of those three models, some of the key indicators are model accuracy, (AUC), F1 score etc Based on the model comparson data RandomForest has been found as the most efficient model with AUC score of around 92%for the given case scenario. Among predictor variables it is found call “duration” variable is the most important predictor with longer duration calls resulting in more productive discussion and success of the campaign. This was followed by variables euribor3m and nr.employed.

With given data set % of response records are disproportionate compared to the population, around 10% success out of data set. This is real life scenario but creates challenges for the model. To take care of that Area under curve (AUC)metrics has been used for final model selection rather than the accuracy number.