Capstone Project Submission

Team Member's Name, Email and Contribution:

Amit Kundu (amitkundu9593548931@gmail.com)
Contribution:

1) Data cleaning-
Null value treatment, Duplicate values treatment, outlier detection and change the data into correct format

2)EDA and Data Visualization-
Find the relation between dependent and independent variables

3) Splitting data set into train and test set

4)Creating model –
1) K-Nearest Neighbors
2) XGBoost Classifier

5)Cross validation-
Optimized the hyper parameter tunning on random forest

Sananda Biswas Chatterjee(sananda.uem@gmail.com)
Contribution:

1)Data Wrangling –
Analysed the Data set

2)EDA and Data Visualization-
Analysed the univariate
Find the correlation between variables through heatmap

3)Creating model –
1)Logistic Regression
2) Random Forest Classifier
3) Decision Tree Classifier

4)Comparing ROC AUC curves and its performance.

Github Link:- https://github.com/sananda2005/Bank-Marketing-Effectiveness-Prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Problem Statement

One of the industries most impacted by recent advances in machine learning is the finance industry. Whether it be predicting market prices or, in our instance, predicting a customer's tendency to open a term deposit. As a result, we have created a solution for our project that boosts success rates while enhancing efficiency by placing fewer calls. We train a model in this study to predict whether or not the client would sign up for a term deposit. The data set was compiled using call logs from a Portuguese banking institution's direct marketing operations.

Approach

Our aim is to develop a predictive model that could aid in determining whether the client will subscribe for a term deposit or not.
For that purpose, after importing all libraries and dataset we have performed data set cleaning procedure.
The data set is first subjected to exploratory data analysis. We search for missing data values (none were identified) and outliers, then change them as necessary. Additionally, we use correlation analysis to isolate the most crucial and pertinent feature set. Later, we use feature engineering to change a few already-existing columns and eliminate unnecessary ones.
We distinguish between numerical and categorical columns before analyzing each one separately. Label encoding is also done on categorical columns along with hyper parameter tuning.

To train and test the performances, we used five different kinds of models.

1) Logistic Regression

2) Decision Tree Classifier

3) Random Forest Classifier

4) K-Nearest Neighbors (KNN)

5) XGBoost Classifier

This is how we completed our task at the Bank Marketing Effectiveness Prediction Project. We pick up a lot of new skills throughout the project, from understanding the problem description to comprehending the technical aspects of a data analysis.

Conclusion

- The model makes it easier to focus on the correct customers rather than waste time on the wrong ones.
- When compared to all algorithms, the XGBoost algorithm has the highest accuracy rating and ROC-AUC rating. It is therefore declared to be the best model.
- The model can assist in categorizing the clients according to whether they make deposits or not.