Individual Contribution Statement

ID 5059 Knowledge Discovery and Datamining

Group: **G05**

Student ID: **180025784**

In our group, we decided that each member can choose different algorithms to fit the model after the data being cleaned and then compare the optimal model using different algorithms and choose the one performs best with highest Gini score. The models we have tried to fit with the dataset contains AdaBoost, XGBoost, Neural Nets, Random Forest and Generalized Linear Model. My main contribution to the model selection part is to fit GLM model with different *k*-value in the cross validation and lambda search to find the best GLM to predict the test data. The public and private Gini score of the best GLM is 0.23079 and 0.23761; However, XGBoost model shows much better public and private Gini score of 0.28243 and 0.28641. Therefore, we can concluded XGBoost is the best model for the case of Porto Seguro.