Chapter 5 Experiment Result and Analysis

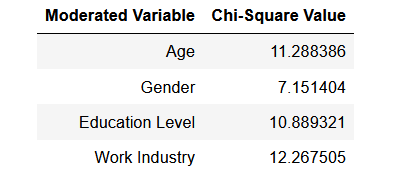
In this part, we test the performance of Correlation Based Filter Feature Selection, Information Gain Ranking Feature Selection, Correlation Attribute Based Ranking Feature Selection and 3 different selected machine learning classifiers to evaluate our model. We applied it in UTAUT Factor Dataset.

The selected machine learning classifier were tree-based algorithm. This is because the benefits of Decision Tree is that it can avoid the multicolinearlity issue where attributes are highly correlated, producing a highly biased and inaccurate result. We also use Random Forest, as it is a bagging method and XGBoost as a boosting method, to determine whether can we improve the performance of our model.

The dataset was preprocessed and analyzed according to the above steps. To ensure the accuracy is valid and all data are used in building the model, we use 10-fold cross validation. The dataset is split 80-20, 80% as the training set and 20% for the testing set.

To assess the performance of our model, we use confusion Matrix.

Relationship between moderated variable and Target Variable

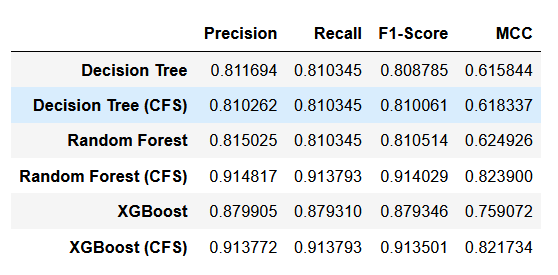


Using Chi-Square Calculation, we experiment “Age”, “Gender”, “Marital Status”, “Education Level”, “Work Industry” and “Work Position” to define whether is there a significant relationship between these variables and the target variables. The above table shows that the intention to use e-wallet has significant relationship between “Age”, “Gender”, “Education Level” and “Work Industry”.

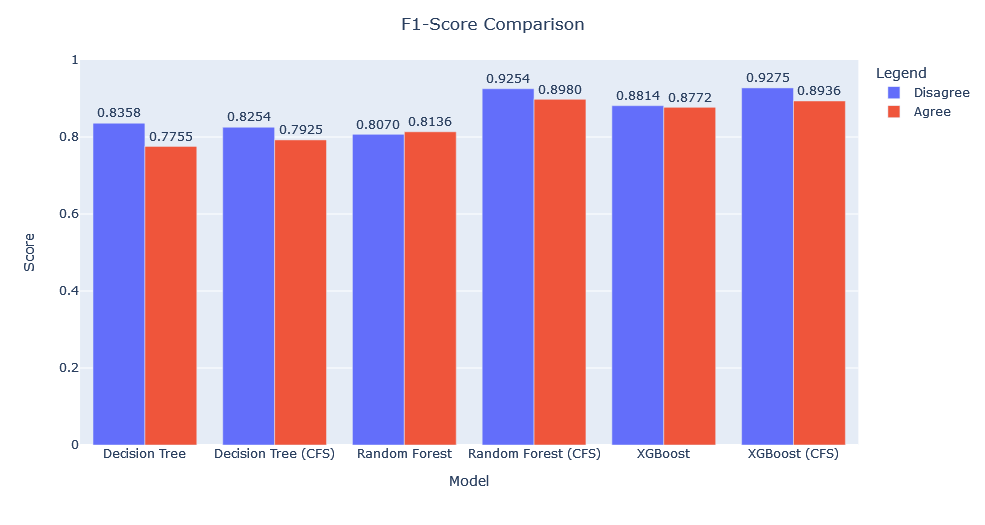
Overall Performance

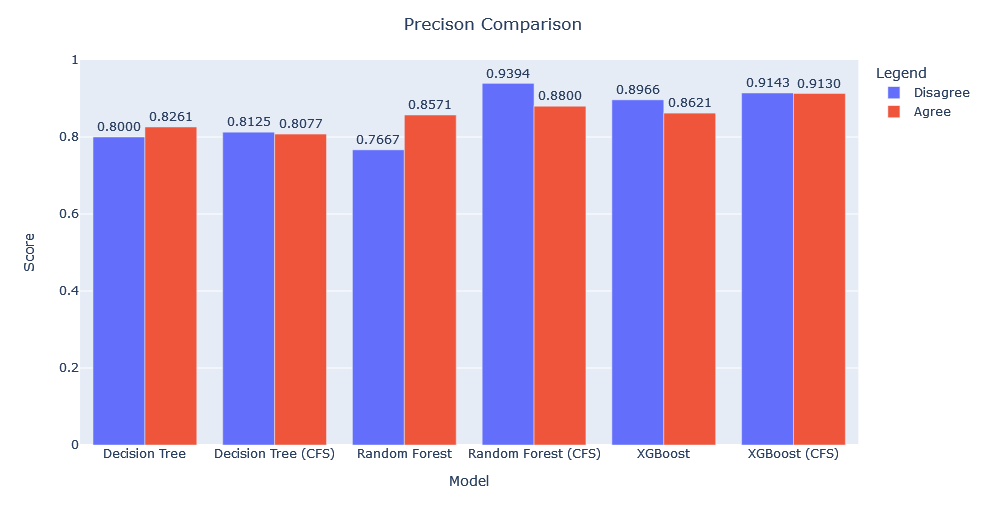
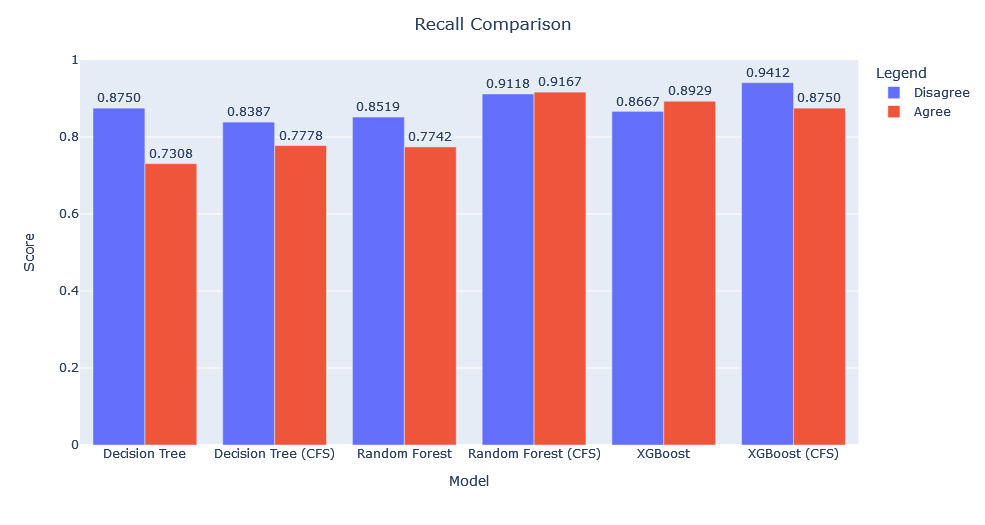
We compare the classification performance metrics against different machine learning classifiers on the number of feature selected. The classification accuracy, precision, F1-Score, recall, Matthew Correlation Coefficient are the average of each classifier to reduce the bias of the result. Time taken to build the model is also recorded as well for analysis.

A good feature selection method is supposed to find out important features accurately. To compare the performance of each feature selection method, we based it on the classification accuracy.



From the above table, we compare our machine learning model based on taking all features and Correlation Based Features. We test it out on 3 different machine learning classifier. The results above indicate that the machine learning model that were trained using correlated features have a much higher accuracy and a much higher MCC Score. This indicates that the feature selection process was a success

The Highest F1-Score was by XGBoost(CFS) in determine disagree values, whereas the highest score for Agree values was by random Forest(CFS). The lowest F1-Score



From the above model, the variable that affects them the most is trust. LMAO

A question that pops up is, does different group of respondents have different results, are they affected by other group of variables?