

Part 5: Summary and Difficulties

The random forest which is the bagged ensemble method of the decision trees is the best type of model for classifying the churn of customers. More detailed insights are subject to be discovered as at the moment there is no much analysis that can be explained due to a quite high misclassification rate for the best model available. Other models other than decision trees are advised to be applied in this study. The high rate of misclassification may be due to the data is generated highly randomly with limited limitations applied. Besides, data from 50 United States and 1 federal district (District of Columbia) is insufficient for 1000 rows, which increases the difficulty of insight generations. With the existence of this limitation, the extension of this study can include the tuning of HP environment model and Gradient Boosting in attempt of generate more valuable insights from this dataset.