CA03 – Ensemble Models

**Q1. Write your observations about the Classifier’s behavior with respect to the number of estimators**

*Random Forest*

The plot returns a polynomial trendline. The accuracy score of the predictions gradually decreases when the n\_estimator ranges between 50 and 250 trees. At 300 trees, the accuracy increases again before decreasing at 350. After, it gradually increases again until it reaches 500 trees. Thus, the model that uses the lowest n\_estimator value produces the optimal results in terms of accuracy.

*ADA Boost*

The accuracy score starts at a low with 50 trees, but then peaks with 100-150 trees. It decreases again once it reaches 200 and stays constant until 500 trees. To obtain the optimal results, a relatively low n\_estimator value should be used. Note that there is an improvement in the overall accuracy scores compared to the Random Forest Model.

*Gradient Boost*

The graph returned a polynomial trendline. With 50 trees, the accuracy score is relatively high but then decreases within the 100-200 tree range. In the range of 250-500 trees, the trendline fluctuates heaviliy but reaches peak accuracy at 350 trees. Thus, a medium-high n\_estimator value should be used to obtain optimal results. Note that the accuracy score using the Gradient Boost model is higher than that using the AdaBoost or Random Forest model.

*XGBoost*

The graph returned a trendline that follows a non-linear downward trend. It returns the highest accuracy score at 100 trees. Despite there are small increases in accuracy, it gradually decreases until it reaches 500 trees. Thus, a relatively low n\_estimator value should be used to obtain optimal results. The range of accuracy scores is comparable to those when using the Gradient Boost Model.

**Q2. Is there an optimal value of the estimator within the given range?**

*Random Forest*

The optimal value of the estimator is 50 trees, which results in the highest accuracy score of 0.8384.

*ADA Boost*

The optimal value of the estimator is 100-150 trees, which results in the highest accuracy scores of 0.844915.

*Gradient Boost*

The optimal value of the estimator is 350 trees, which results in the highest accuracy scores of 0.8472.

*XGBoost*

The optimal value of the estimator is 100 trees, which results in the highest accuracy scores of about 0.84715.