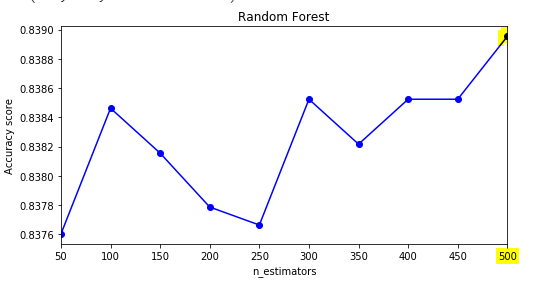
CA-04 Ensemble Models

**Random Forest:**

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

It performs the best when n\_estimator is 500. The random forest model is a bagging algorithm and the performance varies with every n\_estimator and the change is a bit random. As can be seen in the graph below:



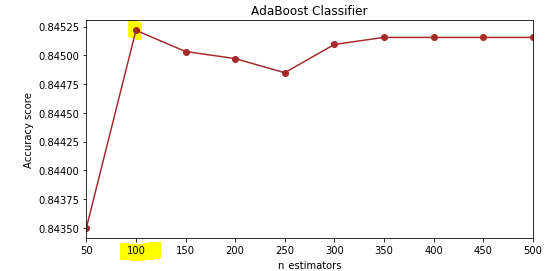
1. **Is there an optimal value of the estimator within the given range?**

The value at n\_estimator =500 is the most optimum value for the random forest performance the accuracy is the highest at this value

**AdaBoost Model:**

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

AdaBoost classifier performance significantly improves after n\_estimator =50 and then kind of performs in the same range with higher accuracy. The best accuracy is at n\_estimator =100. AdaBoost uses a boosting technique where each decision tree learns from the mistakes of previous tree and reduces the error.



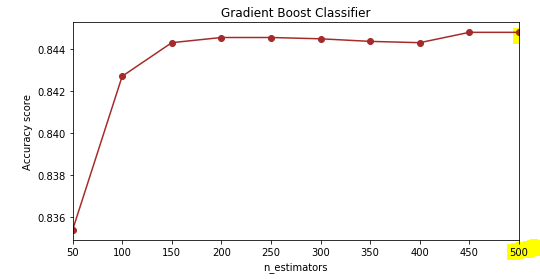
1. **Is there an optimal value of the estimator within the given range?**

The value at n\_estimator =100 is the most optimum value for the random forest performance, the accuracy is the highest at this value

**Gradient Boosting Model:**

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

Gradient Boost classifier performance significantly improves after n\_estimator =50 and then performs in the same range with higher accuracy. The best accuracy is at n\_estimator =500.



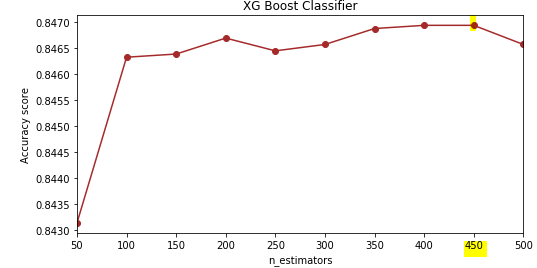
1. **Is there an optimal value of the estimator within the given range?**

The value at n\_estimator =500 is the most optimum value for the random forest performance The accuracy is the highest at this value

**XG Boost Model:**

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

XG Boost classifier performance significantly improves after n\_estimator =50 and then performs in the same range with higher accuracy. The best accuracy is at n\_estimator =450.



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

The value at n\_estimator =450 is the most optimum value for the random forest performance. The accuracy is the highest at this value.

**Compare Performance**

**Keep all common Hyper-parameters same for four models (Random Forest, AdaBoost, Gradient Boost, XGB), run them again and create a performance comparison table within your code and print the same as follows:**

**Upon keeping all the hyper-parameters as the best, we got for Random Forest (n\_estimator =500), we get the following results:**

* **Accuracy and AUC for XGBoost Model is the highest**

