binshu technology data observation and analyses

1. There are 10 features,but “ID” is useless in our analyses
2. To make sure that this is a classification problem, I tried many classifiers such as *LogisticRegression,GaussianNB,MultinomialNB,KNeighborsClassifier,SVC,DecisionTreeClassifier,BaggingClassifier,ExtraTreesClassifier,GradientBoostingClassifier*
3. Then a criterion should be established to figure out which model is best for our problem and at the same time not so much time demanding,after some attempts, GradientBoostingClassifier stands out
4. After determining the model,I need to optimize the parameters to reach the best performance of the model,the parameters that needed to be optimized are: *'n\_estimators', 'max\_depth','min\_samples\_split', 'min\_samples\_leaf', 'max\_features','subsample'*,after these parameters are optimized by a grid search method, we need to decrease the *'learning\_rate'* to save time,meanwhile,we need to increase the *'n\_estimators'* to make up the reduction of *'learning\_rate'.*finally,all the parameters are determined and the roc\_auc score was 0.790813.
5. In the end,I performed a predict\_proba method on the ‘case2\_testing\_data’ to get the final results.
6. The *'max\_features'=5,*which indicates that *'Price'*, *'Review', 'Pic Quality', 'Beds', 'Week day' are the most important factors.*

