1. Training and testing set has been used to train. The result has been summarized as below:

|  |  |  |
| --- | --- | --- |
| n\_estimators | Training Set  Accuracy score | Testing Set  Accuracy Score |
| 1 |  |  |
| 20 |  |  |
| 100 |  |  |
| 9  (highest accuracy score) |  |  |
| 14  (highest accuracy score) |  |  |

Testing data set contains more data, although it contains outliers, it won’t make random forest model overfit as random forest model won’t overfit. Compare the result with train by training set, b

So using the testing set to train more make sense

Random forest advantages-no overfit

Disadvantages- ignore outlier, low accuracy