**Raj Kumar Mannam**

**1002126524**

**Classification Method: Random Forest Classifier**

The reason why the RFC is considered the best model in this context is that it outperformed the other classifiers (such as Linear Support Vector Machine, Decision Tree, KNN and Gaussian Naive Bayes) in terms of accuracy on the test set. It's important to note that the best model selection depends on the specific dataset and the evaluation metric used (in this case, accuracy). Different datasets and evaluation metrics may lead to different best model choices.

**Data Pre-processing:**

Selected only rows with at least 20 minutes of playtime in the data frame.

round attribute values to two decimal places.

**Feature Selection:**

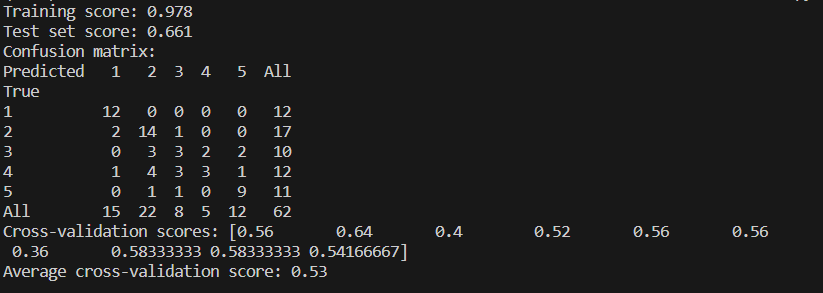
Used these 6 features (AST, TRB, ORB, BLK, DRB, and 3PA) for prediction as these attributes has the best variance among all the attributes

**Model Selection:**

To determine the best number of estimators (trees) for the Random Forest model, the algorithm runs a loop. It starts the RFC with a variety of estimators in a range (from 25 to 124) and assesses the precision of each model on the test set.

The maximum accuracy obtained across all RFCs with various numbers of estimators is tracked by the variable best\_score. The best model is the one that performs the best on the test set; its predictions and accuracy score are kept in the best\_test\_results and best\_score variables, respectively.

**Results:**



The accuracy score for the model is 0.661.

Model's average accuracy score after 10-fold stratified cross-validation was 0.53.