# Introduction

* Goal is to predict if a NBA Team will make it to the playoff or not

# Data preparation

* Manually gathered data from <https://www.basketball-reference.com/leagues/>
* Initial columns chosen for analysis were:
  + Id
  + Year
  + Team
  + 3 Points
  + 2 Points
  + Free Throws
  + Rebounds
  + Assists
  + Steals
  + Block Shots
  + Turnovers
  + Total Points
  + Minutes Played
  + Playoff
* Going over the data we noticed that over the year the number of NBA teams have increased and not all teams have played the same amount of time or games.
* To resolve this issue, we all numeric column by minutes played the final columns that were analysed are:

"Points\_Per\_minute","3Points\_Per\_minute","2Points\_Per\_minute","FThrow\_Per\_minute","Rebound\_Per\_minute","Assists\_Per\_minute","Steals\_Per\_minute","Blocks\_Per\_minute","TurnOvers\_Per\_minute"

# Libraries used

* Gaussian Naïve Bayes
* SVM
* 70 % training 30% Testing
* K-fold

# Observation

Reference : <https://muthu.co/understanding-the-classification-report-in-sklearn/>

**Precision** – What percent of your predictions were correct?

* Precision is the ability of a classifier not to label an instance positive that is actually negative.
* For each class it is defined as the ratio of true positives to the sum of true and false positives.
  + Precision – Accuracy of positive predictions.
  + Precision = TP/ (TP + FP)

**Recall** – What percent of the positive cases did you catch?

* Recall is the ability of a classifier to find all positive instances.
* For each class it is defined as the ratio of true positives to the sum of true positives and false negatives.
  + FN – False Negatives
  + Recall: Fraction of positives that were correctly identified.
  + Recall = TP/(TP+FN)

**F1 score** – What percent of positive predictions were correct?

The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Generally speaking, F1 scores are lower than accuracy measures as they embed precision and recall into their computation. As a rule of thumb, the weighted average of F1 should be used to compare classifier models, not global accuracy.

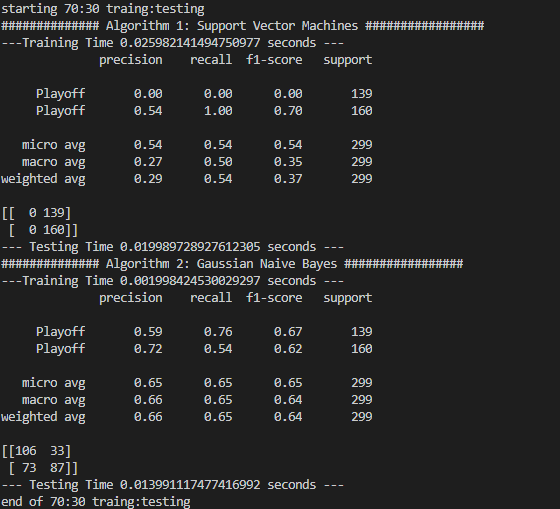
* F1 Score = 2\*(Recall \* Precision) / (Recall + Precision)

Another Definition

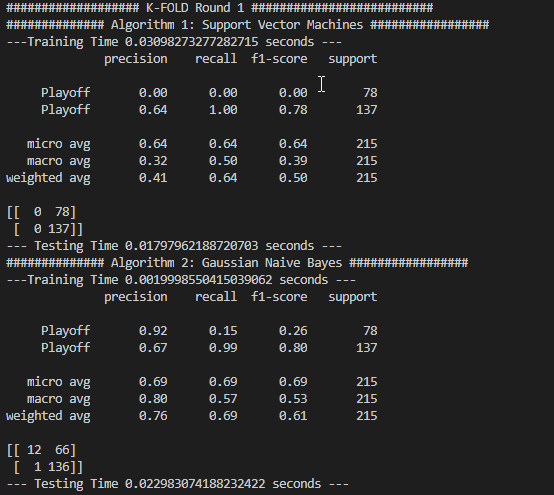
Reference:

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html>

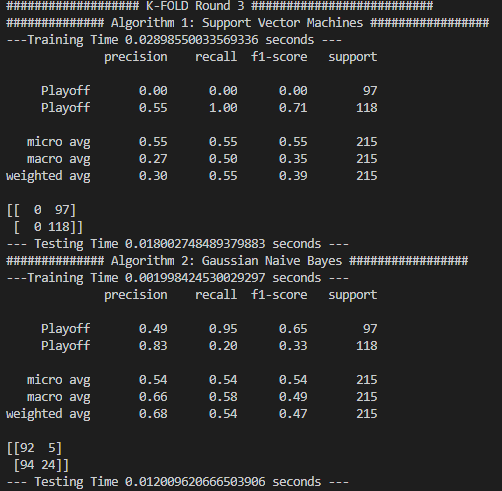
* The precision is the ratio tp / (tp + fp) where tp is the number of true positives and fp the number of false positives. The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
* The recall is the ratio tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples.
* The F-beta score can be interpreted as a weighted harmonic mean of the precision and recall, where an F-beta score reaches its best value at 1 and worst score at 0.
* The F-beta score weights recall more than precision by a factor of beta. beta == 1.0 means recall and precision are equally important.
* The support is the number of occurrences of each class in y\_true.

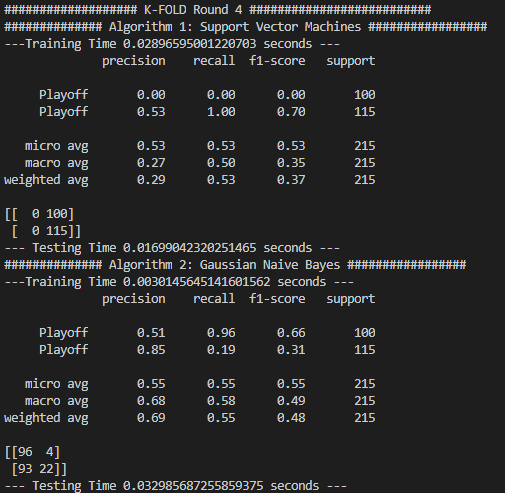


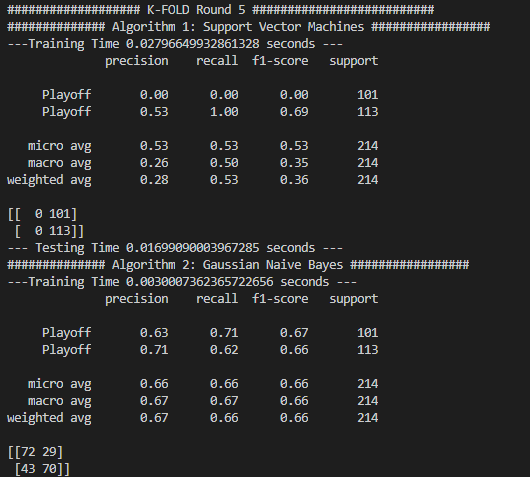
**K-Fold**

****

# 







# Conclusion