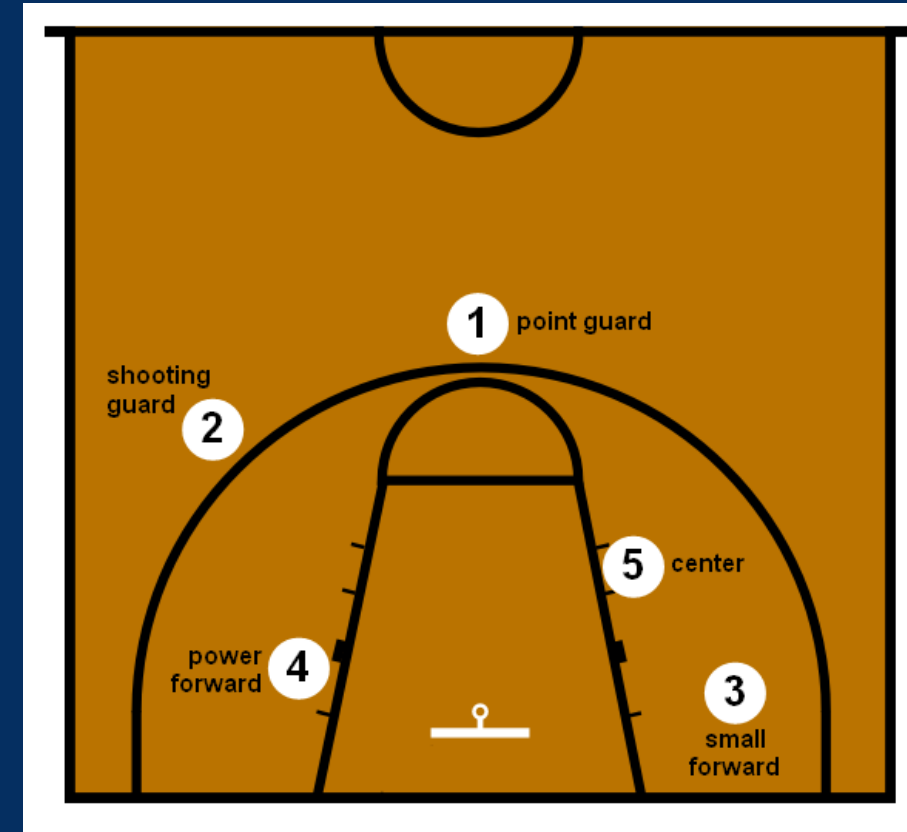


NBA BASKETBALL POSITION CLASSIFICATION ANALYSIS

BY: GABE BLATSTEIN

OVERVIEW

- NBA Classification Modeling Analysis:
 - Background
 - Approach
 - Modeling
 - Key Takeaways



BACKGROUND

- Background:
 - The data set used came from Kaggle.com and ranges from 1950 until 2017
 - The data includes 50 different NBA stats along with each players height and weight
 - Different classification machine learning modeling techniques were used to try and correctly identify an NBA players position based off their physical traits and game stats.
 - Due to the NBA shifting into a position-less basketball style of play, I thought it would be interesting to apply statistical python code to solve this problem.
 - The data set was divided into 4 different time periods due to the historical change in the style of play in the NBA



APPROACH OSEMN FRAMEWORK

Observe/Obtain



Scrub/Clean



Explore



Model

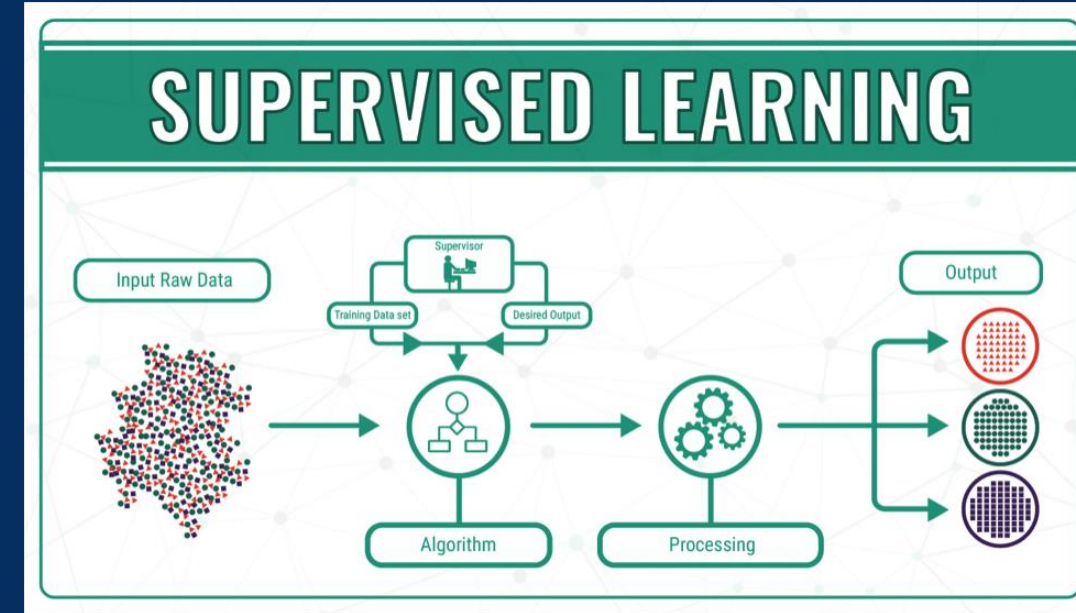


Interpret



MODELING

- Four different modeling techniques were used:
 - K-Nearest Neighbor (KNN)
 - Adaboost
 - Random Forest
 - Support Vector Machine (SVM)



MODELING

KNN

Reference Key (left to right):

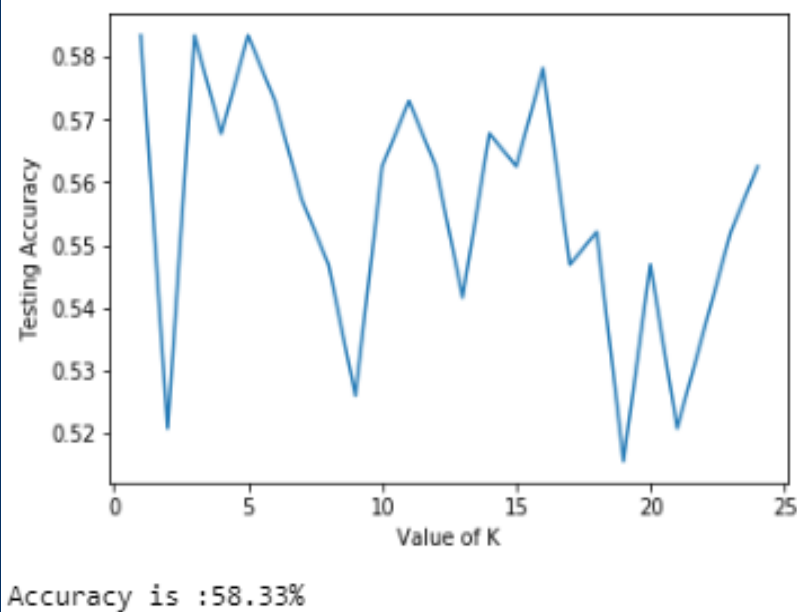
1. 1950-1956

2. 1956-1980

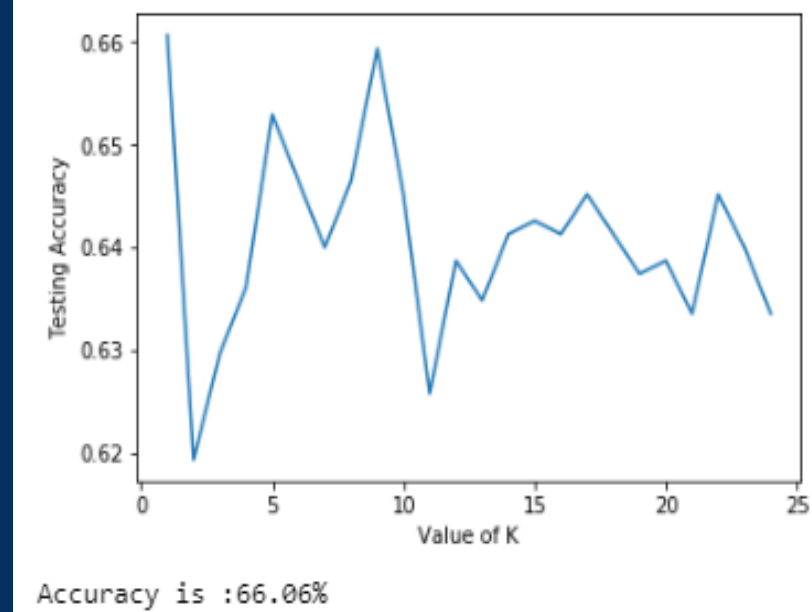
3. 1980-1998

4. 1999-2017

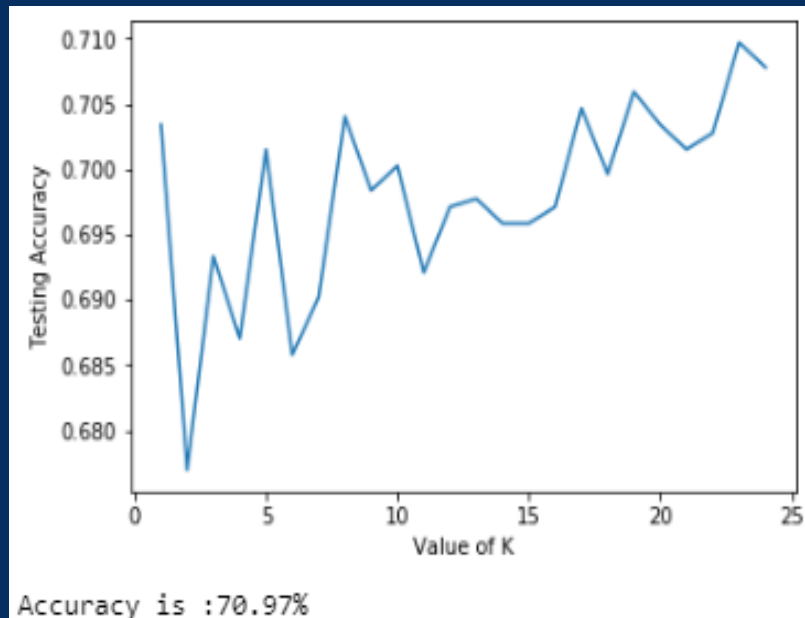
1.



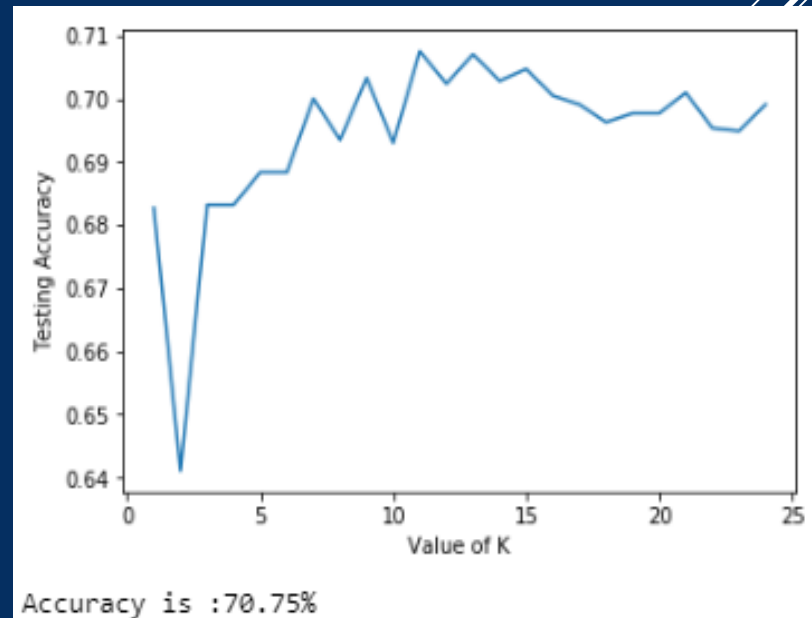
2.



3.



4.



MODELING

ADABOOST

Reference Key (left to right):

1. 1950-1956

2. 1956-1980

3. 1980-1998

4. 1999-2017

1.

Accuracy is :55.21%					
	precision	recall	f1-score	support	
1	0.28	0.53	0.36	15	
2	0.52	0.58	0.55	55	
3	0.60	0.48	0.53	63	
4	0.76	0.41	0.53	32	
5	0.66	0.85	0.74	27	
micro avg	0.55	0.55	0.55	192	
macro avg	0.56	0.57	0.54	192	
weighted avg	0.59	0.55	0.55	192	

2.

Accuracy is :60.0%					
	precision	recall	f1-score	support	
1	0.55	0.66	0.60	117	
2	0.59	0.56	0.57	167	
3	0.74	0.58	0.65	184	
4	0.49	0.41	0.45	154	
5	0.61	0.82	0.70	153	
micro avg	0.60	0.60	0.60	775	
macro avg	0.60	0.60	0.59	775	
weighted avg	0.60	0.60	0.60	775	

3.

Accuracy is :66.37%					
	precision	recall	f1-score	support	
1	0.71	0.96	0.82	319	
2	0.65	0.50	0.56	309	
3	0.70	0.48	0.57	304	
4	0.65	0.47	0.55	350	
5	0.62	0.92	0.74	306	
micro avg	0.66	0.66	0.66	1588	
macro avg	0.67	0.67	0.65	1588	
weighted avg	0.67	0.66	0.65	1588	

4.

Accuracy is :64.02%					
	precision	recall	f1-score	support	
1	0.67	0.96	0.79	467	
2	0.61	0.38	0.47	404	
3	0.69	0.54	0.61	407	
4	0.59	0.29	0.38	419	
5	0.61	0.96	0.75	443	
micro avg	0.64	0.64	0.64	2140	
macro avg	0.63	0.63	0.60	2140	
weighted avg	0.63	0.64	0.61	2140	

NOTE:

- The Adaboost model was the worst performing model out of the four selected modeling techniques.



MODELING

RANDOM FOREST

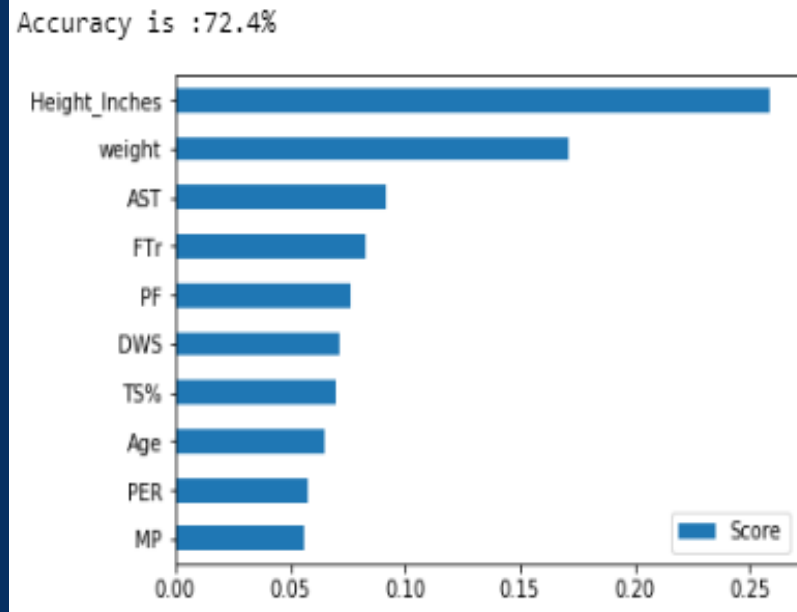
Reference Key (left to right):

1. 1950-1956
2. 1956-1980
3. 1980-1998
4. 1999-2017

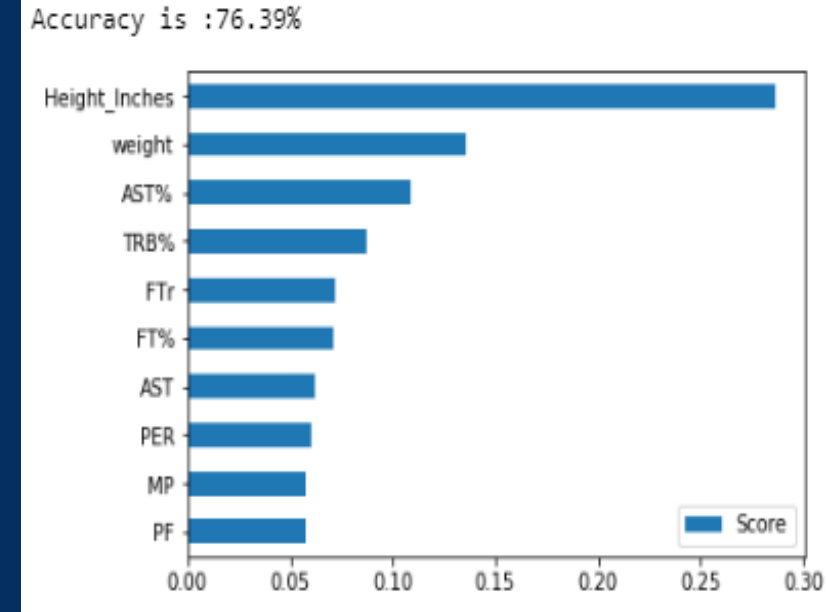
NOTE:

- Random Forest performed the best out of the four modeling techniques used.

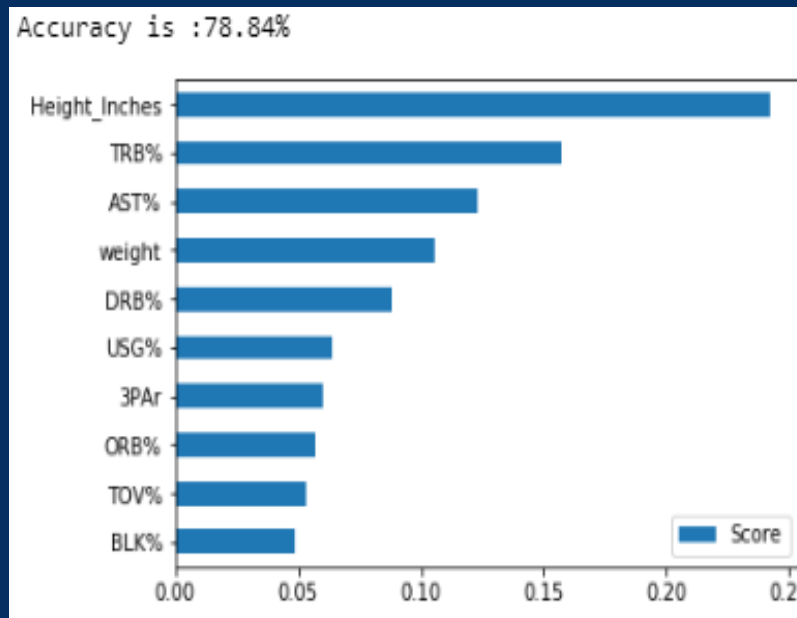
1.



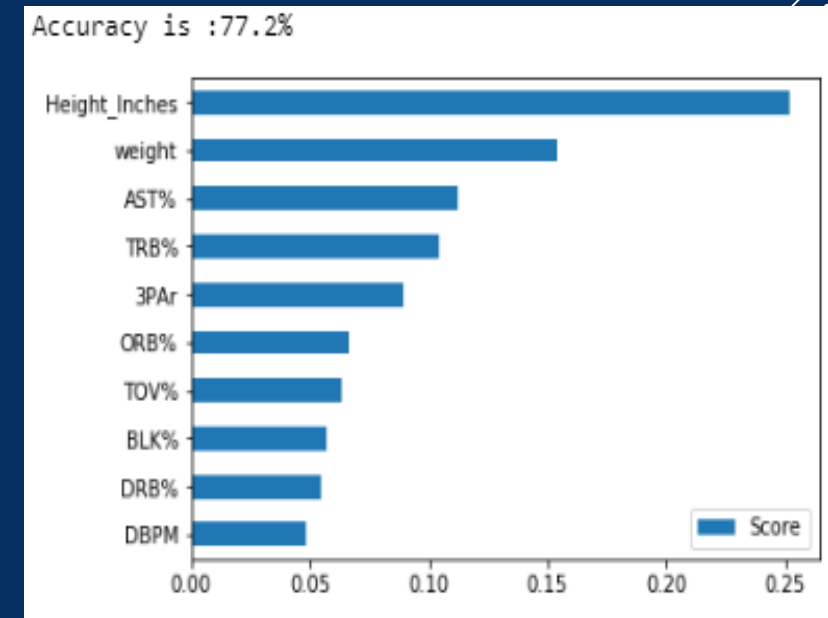
2.



3.



4.



MODELING

SVM

Reference Key (left to right):

1. 1950-1956

2. 1956-1980

3. 1980-1998

4. 1999-2017

1. Accuracy is :67.19%

	precision	recall	f1-score	support
1	0.70	0.35	0.47	20
2	0.53	0.73	0.62	44
3	0.72	0.61	0.66	67
4	0.66	0.78	0.71	32
5	0.89	0.83	0.86	29
micro avg	0.67	0.67	0.67	192
macro avg	0.70	0.66	0.66	192
weighted avg	0.69	0.67	0.67	192

2. Accuracy is :70.45%

	precision	recall	f1-score	support
1	0.71	0.71	0.71	108
2	0.70	0.73	0.71	171
3	0.70	0.66	0.68	191
4	0.59	0.66	0.62	160
5	0.85	0.79	0.82	145
micro avg	0.70	0.70	0.70	775
macro avg	0.71	0.71	0.71	775
weighted avg	0.71	0.70	0.71	775

3. Accuracy is :76.76%

	precision	recall	f1-score	support
1	0.87	0.86	0.87	324
2	0.71	0.71	0.71	316
3	0.68	0.67	0.68	303
4	0.70	0.78	0.74	339
5	0.89	0.81	0.85	306
micro avg	0.77	0.77	0.77	1588
macro avg	0.77	0.77	0.77	1588
weighted avg	0.77	0.77	0.77	1588

4. Accuracy is :75.89%

	precision	recall	f1-score	support
1	0.86	0.91	0.88	455
2	0.73	0.67	0.70	443
3	0.65	0.69	0.67	387
4	0.71	0.68	0.70	446
5	0.82	0.84	0.83	409
micro avg	0.76	0.76	0.76	2140
macro avg	0.76	0.76	0.76	2140
weighted avg	0.76	0.76	0.76	2140

NOTE:

- SVM performed the second best out of the four modeling techniques used.



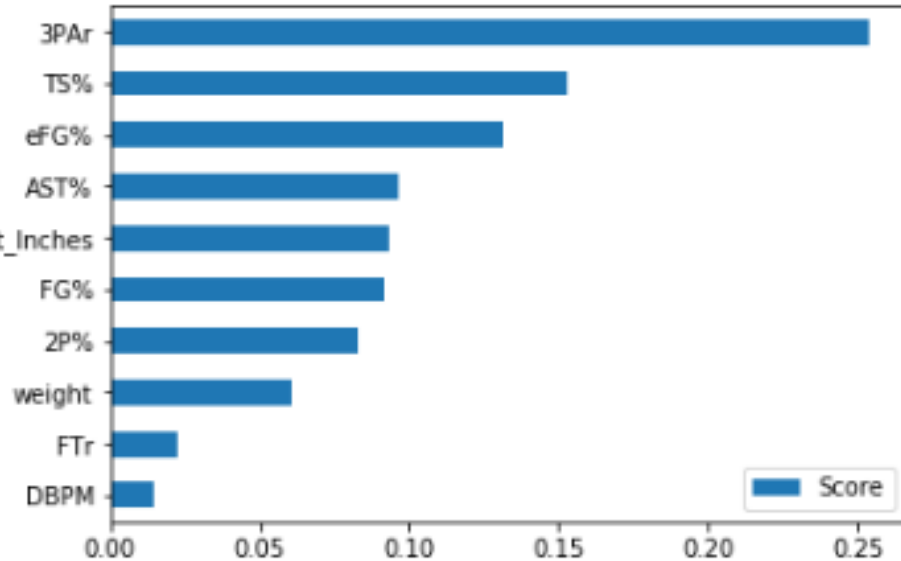
DECISION TO REDUCE CLASSES OF POSITIONS

- To improve the models, the number of positions was reduced to three as follows:
 - PG (1)
 - SG/SF (2)
 - PF/C (3)
- As expected with a reduction in the number of possible classifications, all metrics improved significantly



DECISION TO REDUCE CLASSES OF POSITIONS

Accuracy is :94.67%



	precision	recall	f1-score	support
1	0.86	0.89	0.87	436
2	0.94	0.92	0.93	810
3	1.00	1.00	1.00	894
micro avg	0.95	0.95	0.95	2140
macro avg	0.93	0.94	0.93	2140
weighted avg	0.95	0.95	0.95	2140

- A Random Forest model was used to classify the new data
 - The accuracy score jumped almost 20 percentage points to 94%
- The model predicted the 3 position, Center/Power Forward at 100 percent across all metrics
- The hardest position to classify was the 1, Point Guard.
- Note: The top three features are:
 - 3PAr: 3 Point Attempt Rate
 - TS%: True Shooting Percentage
 - eFG%: Effective Field Goal Percentage



KEY TAKEAWAYS

- In the current day and age, we see a basketball floor spread with Power Forward's playing Center and vice versa.
 - This was apparent in the last model when the classes were reduced from 5 to 3.
- The Center and Point Guard positions were classified accurately the most compared to any other position amongst all models
 - A further deep dive into these positions is needed
- Random Forest and Support Vector Machine models performed the best.
 - Both models could be tuned a bit more to get even more accurate results for the 5-position classification.
- The results improved over the eras of basketball due to the availability of new statistics.

