



NBA

# Predicting Game Winners

Javier Herbas, Aug 2020

# Introduction & Objectives

- Predict NBA game winners based on Home Court Advantage (Home Team)

How?

- Evaluating different ML Classification Algorithms and shortlisting a set of stats that could translate into wins



# Questions raised...

- ~70 years of NBA data available
  - Over 100 stats per team are available
  - Game has evolve through the years, therefore Feature Importance will be key
- 
- How far behind should the models go?
  - Which feature combination could translate into wins?
  - Can we get any insight from this game evolution to help us predict the winners?



# Data Gathering and Conditioning

Data sources:

- <https://stats.nba.com/teams/>
- <https://www.espn.co.uk/nba/stats/ /view/team>
- <https://www.basketball-reference.com>
- <https://watchstadium.com>
- <https://www.si.com>
- <https://www.basketball-reference.com/>



- 1 year data (2018 – 2019)
- 10 year data (2010 – 2020)
- 20 year data (2000 – 2020)
- 30 year data (1990 - 2020)

	Team1	Team1Score		Team2	Team2Score	year	month	Game_Result												
0	Charlotte Hornets		106	Atlanta Hawks		82	2000	10	0											
	TEAM	year	month	PTS	FGM	FGA	FG_P	3PM	3PA	3P_P										
0	Atlanta Hawks	2010	10	107.3	37.3	77.0	48.5	6.3	17.3	36.5										
1	Los Angeles Lakers	2010	10	111.0	41.7	93.0	44.8	9.0	22.3	40.3										
	TEAM	year	month	GP	W	L	MIN	OFFRTG	DEFRTG	NETRTG	AST_P	AST/TO	ASTRATIO	OREB_P	DREB_P	REB_P	TOV_P	EFG_P	TS_P	PACE
0	Miami Heat	2010	10	4	3	1	192	103.6	90.0	13.6	58.9	1.43	15.9	26.0	69.0	49.1	14.7	49.5	54.6	90.00
1	Dallas Mavericks	2010	10	3	2	1	144	102.5	91.2	11.2	68.5	1.49	19.7	22.8	69.4	48.9	18.0	53.3	56.8	94.67

Team1 = Visiting

Team2 = Home

Game Result = 0 >>> Game Lost by Home Team

Game Result = 1 >>> Game Won by Home Team

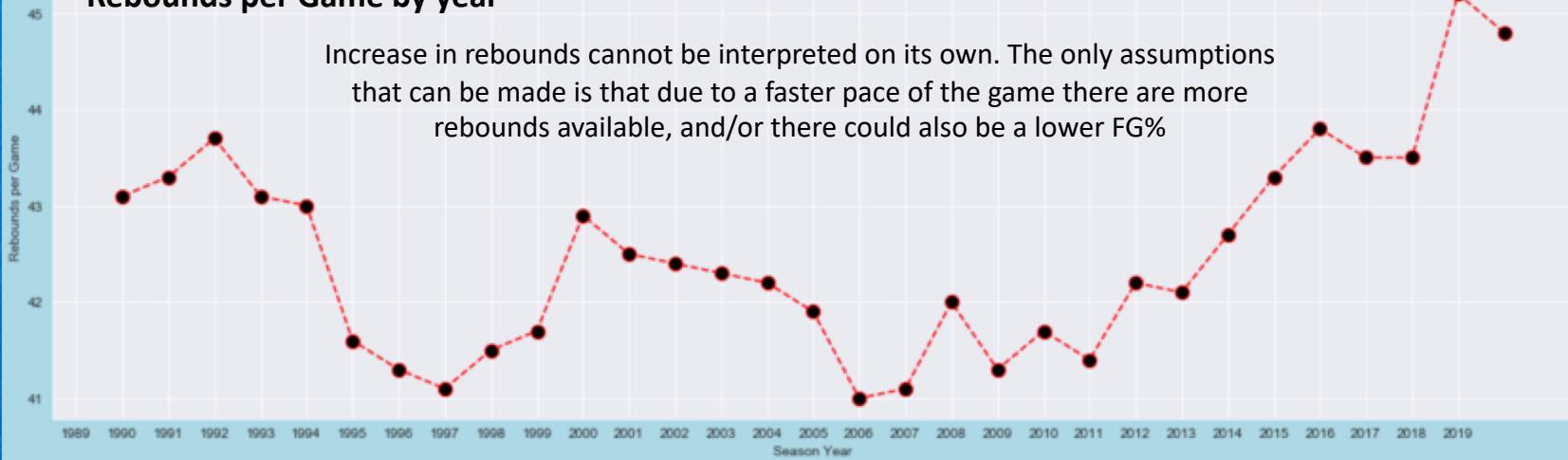
Game_Result	Team1_OFFRTG	Team1_DEFRTG	Team1_AST/TO	Team1_REB_P	Team1_FG_P	Team1_FGA	Team1_PACE	Team1_3PA	Team1_3P_P	
0	1	106.3	107.5	1.83	51.1	43.5	93.1	104.39	35.9	33.8
1	1	102.4	105.0	1.30	50.8	42.8	93.8	106.08	29.7	27.5
2	0	113.2	98.2	1.57	53.4	48.0	91.4	106.00	40.6	38.4
3	1	108.0	111.4	1.46	48.7	45.4	87.5	99.15	35.8	38.1
4	1	104.9	101.5	1.73	47.9	43.5	82.7	98.67	29.0	36.8



# EDA – Game's Evolution

## Rebounds per Game by year

Increase in rebounds cannot be interpreted on its own. The only assumptions that can be made is that due to a faster pace of the game there are more rebounds available, and/or there could also be a lower FG%

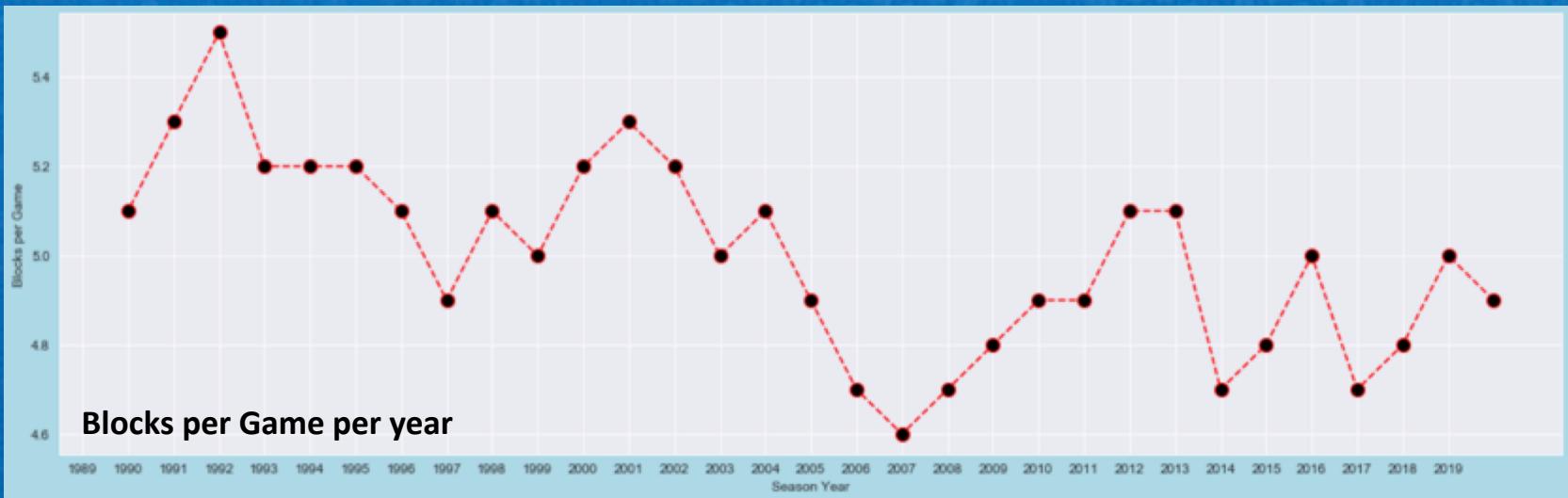
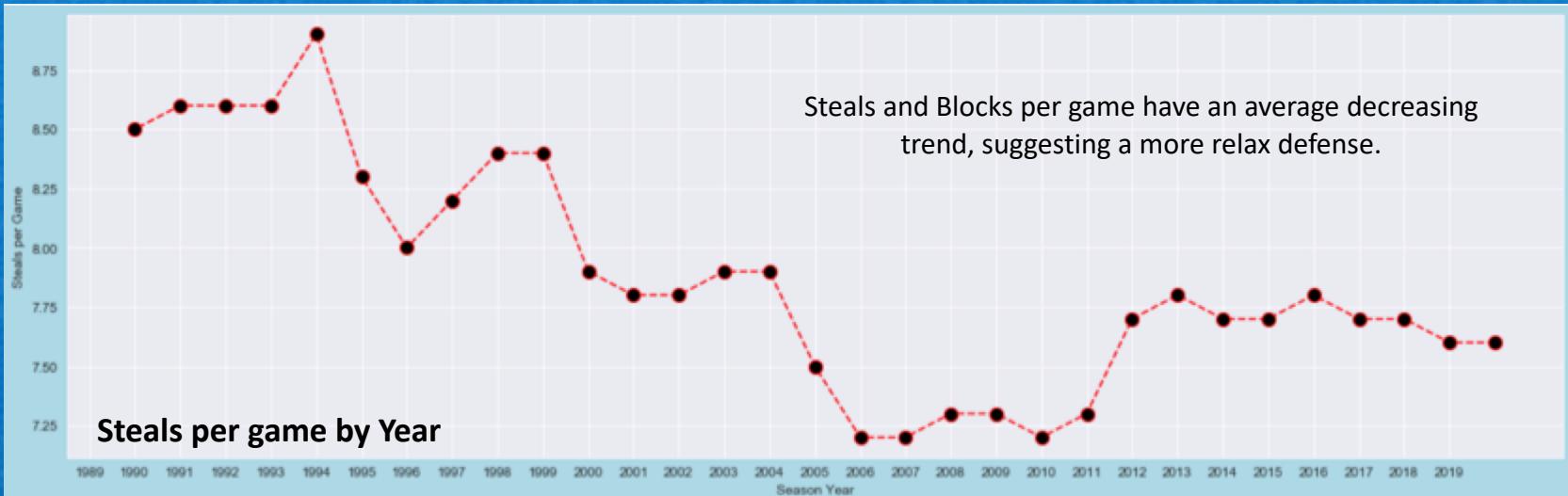


## Points per Game by year

Increase in points per game suggest a faster pace and a weaker defense. In the 90s though, the decrease is due a tougher/better defense

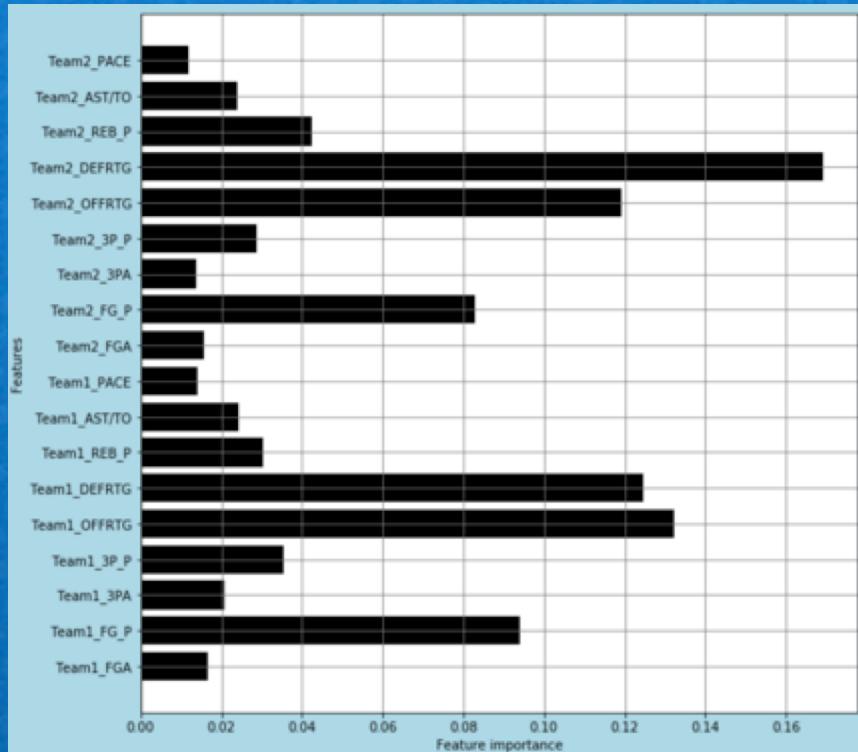


# EDA – Game's Evolution

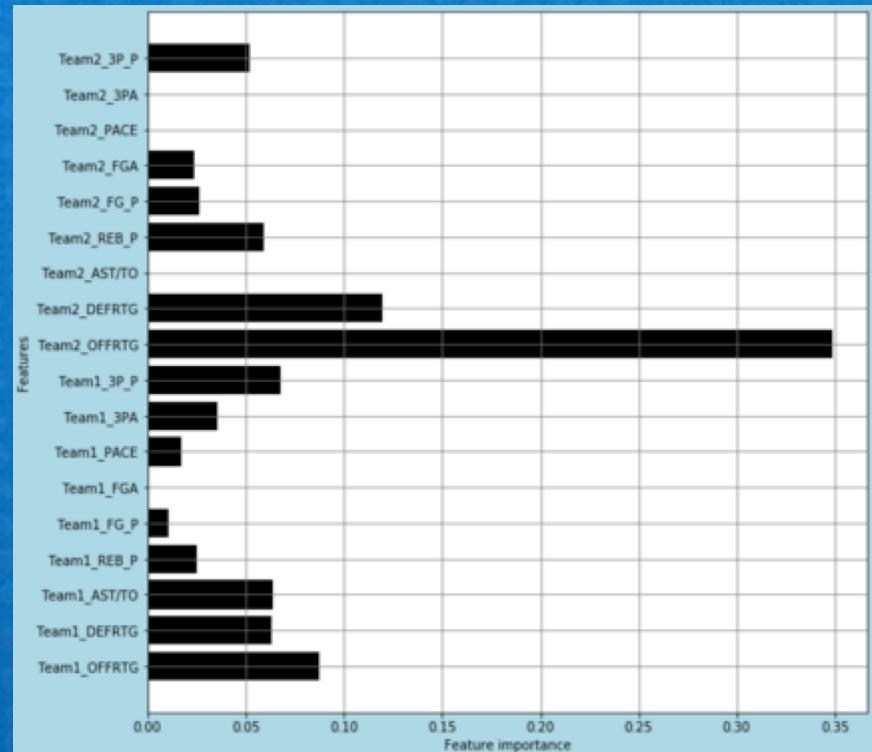


# EDA – Feature Importance

2000-2010



2010-2020



Game Evolution through the past 2 decades...

- Defence used to be the most important element in the game in the early 2000s
- A new era of long range shooters arrived, making it possible to win with a strong offense, however defence still plays a key role



# Exploratory Data Analysis

## Machine Learning Algorithm Tested:

- Logistic Regression Classification
- KNN Classification
- Decision Tree
- Bagged Trees
- Random Forest Classification
- SVM Linear Classification
- SVC Prediction Model
- Bagging SVC Ensemble Classification
- Adaboost Classification
- XGBoost Classification

## Initial Statistical Features

Date	Game day
Team1	Visiting Team
Team1Score	Points scored by Team1
GP	Games Played
W	Wins
L	Losses
WIN_P	Win Percentage
MIN	Minutes Played
Team2	HomeTeam
Team2Score	Points scored by Team2
PTS	Points scored
FGA	Field Goals Attempted
FGM	Field Goals Made
FG_P	Field Goal Percentage
EFG_P	Effective Field Goal Percentage
3PA	3 Points Attempted
3PM	3 Points Made
3P_P	3 Point Percentage
OFFRTG	Offensive Rating
DEFRTG	Defensive Rating
NETRTG	Net Rating
OREB	Offensive Rebounds
DREB	Defensive Rebounds
REB	Rebounds
REB_P	Rebound Percentage
AST_P	Assist Percentage
AST/TO	Assit to Turnover Ratio
AST_P	Assit Percentage
Pace	Possessions per 48 min
TOV	Turnovers
FTA	Free Throws Attempted
FTM	Free Throws Made
FT_P	Free Throw Percentage
BLK	Blocks
BLKA	Blocked Field Goal Attempts
PFD	Personal Fouls Drawn

## Shortlisted Features

FGA
FG_P
3PA
3P_P
OFFRTG
DEFRTG
REB_P
AST/TO
Pace



# Selecting number of years for predictions

Model results – All Teams

Accuracy (%)	All Teams				
	1 year	NBA All-Teams (Average)	10 years	20 years	
	2018 to 2019		2010 to 2020	2000 to 2020	
Logistic Regression Class.	76.85		71.57	71.75	
KNN Classification	67.78		67.18	68.74	
Decision Trees	61.75		61.28	62.42	
Ensembles	Baseline	65.36	67.52	67.20	
	Bagged Trees	66.59	70.92	69.20	
	Random Forest	65.69	70.39	69.40	
SVM Linear Classifier			71.88		
SVC Prediction			71.58		
Bagging SVC Ensemble Classifier			71.62		
AdaBoost Classifier			69.35		
XGBoost Classification			70.52		

75%

- Best predictions were using 10 years of data
- SVM Linear Classifier gave the best accuracy (71.88%)
- Bagged Trees used for real time predictions with good results (17 out of 20 matches accurately predicted)
- No model accounts for upset and matchups between players



# Model Testing Results Per Team



Accuracy (%)		San Antonio Spurs								
		1 year		NBA SAS (Average)	10 years		20 years			
		2018 to 2019			2010 to 2020		2000 to 2020			
		True	Predicted		True	Predicted	True	Predicted		
Logistic Regression Class.		72.72			79.80		77.73			
KNN Classification		100.00			75.96		79.83			
Decision Trees		81.82			65.38		62.13			
Ensembles	Baseline	78.56	67.74		80.50	72.35	78.93	72.86		
	Bagged Trees		77.42			76.52		76.37		
	Random Forest		74.19			78.45		78.76		
Support Vector Machine										



Accuracy (%)		Miami Heat								
		1 year		NBA MIA (Average)	10 years		20 years			
		2018 to 2019			2010 to 2020		2000 to 2020			
		True	Predicted		True	Predicted	True	Predicted		
Logistic Regression Class.		40			64.07		75.00			
KNN Classification		64			63.10		67.79			
Decision Trees		75.00			73.53		65.25			
Ensembles	Baseline	52.5	43.33		68.7	60.46	60.91	63.98		
	Bagged Trees		43.33			66.34		70.62		
	Random Forest		53.33			66.01		70.48		
Support Vector Machine										



# Real Time Results

- The table on the right contains all the games from the first and second round of the playoffs with a prediction accuracy (bottom right side) equal to 74.55%.
- This accuracy of almost 75% was achieved by updating the input data every 2 to 3 days with the results of every team playing a minimum of 1 game. That way the upsets that started happening due to the bubble and changed in player performance could be accounted for. Because of this changes on the input data, and because of the reduction on the number of teams going to the next round of the playoffs, the prediction accuracy is expected to keep on increasing, but we will see that on the next slides

 Correct Prediction  
 Wrong Prediction

Matchday	Matchup		Predicted	Results
	Visiting	Home		
20/08/2020	Heat	Pacers	Heat	Heat
	Thunder	Rockets	Rockets	Rockets
	Magic	Bucks	Bucks	Bucks
	Portland	Lakers	Lakers	Lakers
21/08/2020	Raptors	Nets	Raptors	Raptors
	Nuggets	Jazz	Jazz	Jazz
	Celtics	76ers	76ers	Celtics
	Clippers	Mavericks	Mavericks	Clippers
22/08/2020	Bucks	Magic	Bucks	Bucks
	Heat	Pacers	Heat	Heat
	Rockets	Thunder	Thunder	Thunder
	Lakers	Trailblazers	Lakers	Lakers
23/08/2020	Celtics	76ers	76ers	Celtics
	Clippers	Mavericks	Mavericks	Mavericks
	Raptors	Nets	Raptors	Raptors
	Nuggets	Jazz	Jazz	Jazz
24/08/2020	Bucks	Magic	Bucks	Bucks
	Rockets	Thunder	Thunder	Thunder
	Pacers	Heat	Heat	Heat
	Lakers	Trailblazers	Lakers	Lakers
25/08/2020	Jazz	Nuggets	Nuggets	Nuggets
	Mavericks	Clippers	Clippers	Clippers
30/08/2020 Re-scheduled	Magic	Bucks	Bucks	Bucks
	Thunder	Rockets	Rockets	Rockets
	Trailblazers	Lakers	Lakers	Lakers
30/08/2020 Re-scheduled	Jazz	Nuggets	Nuggets	Nuggets
	Celtics	Raptors	Raptors	Celtics
	Clippers	Mavericks	Mavericks	Clippers
31/08/2020	Heat	Bucks	Bucks	Heat
	Rockets	Thunder	Thunder	Thunder
01/09/2020	Celtics	Raptors	Raptops	Celtics
	Jazz	Nuggets	Nuggets	Nuggets
02/09/2020	Heat	Bucks	Bucks	Bucks
	Thunder	Rockets	Rockets	Rockets
03/09/2020	Raptors	Celtics	Celtics	Raptors
	Nuggets	Clippers	Clippers	Clippers
04/09/2020	Bucks	Heat	Heat	Heat
	Rockets	Lakers	Lakers	Rockets
05/09/2020	Raptors	Celtics	Celtics	Raptors
	Nuggets	Clippers	Nuggets	Nuggets
06/09/2020	Bucks	Heat	Bucks	Bucks
	Rockets	Lakers	Lakers	Lakers
07/09/2020	Celtics	Raptors	Raptors	Celtics
	Clippers	Nuggets	Clippers	Clippers
08/09/2020	Heat	Bucks	Bucks	Heat
	Lakers	Rockets	Lakers	Lakers
09/09/2020	Raptors	Celtics	Raptors	Raptors
	Clippers	Nuggets	Clippers	Clippers
10/09/2020	Lakers	Rockets	Lakers	Lakers
	Nuggets	Clippers	Nuggets	Nuggets
11/09/2020	Celtics	Raptors	Celtics	Celtics
	Clippers	Nuggets	Clippers	Clippers
12/09/2020	Rockets	Lakers	Lakers	Lakers
	Nuggets	Clippers	Nuggets	Nuggets
13/09/2020	Clippers	Nuggets	Nuggets	Nuggets
	Heat	Celtics	Celtics	Heat
14/09/2020	Nuggets	Clippers	Clippers	Nuggets
	Heat	Celtics	Celtics	Heat
15/09/2020	Nuggets	Clippers	Clippers	Nuggets
	Heat	Celtics	Celtics	Heat
				74.55

# Real Time Results

## Conference Finals

Matchday	Matchup		Predicted	Results	
	Visiting	Home			
17/09/2020	Heat	Celtics	Celtics	Heat	Wrong Prediction
	Nuggets	Lakers	Lakers	Lakers	Correct Prediction
19/09/2020	Celtics	Heat	Celtics	Celtics	Correct Prediction
20/09/2020	Nuggets	Lakers	Lakers	Lakers	Correct Prediction
22/09/2020	Lakers	Nuggets	Nuggets	Nuggets	Correct Prediction
23/09/2020	Celtics	Heat	Heat	Heat	Correct Prediction
24/09/2020	Lakers	Nuggets	Lakers	Lakers	Correct Prediction
25/09/2020	Heat	Celtics	Celtics	Celtics	Correct Prediction
26/09/2020	Nuggets	Lakers	Lakers	Lakers	Correct Prediction
27/09/2020	Celtics	Heat	Heat	Heat	Correct Prediction
					<b>90.00</b>

## NBA Finals

Matchday	Matchup		Predicted	Results	
	Visiting	Home			
30/09/2020	Heat	Lakers	Lakers	Lakers	Correct Prediction
02/10/2020	Heat	Lakers	Lakers	Lakers	Correct Prediction
04/10/2020	Lakers	Heat	Heat	Heat	Correct Prediction
06/10/2020	Lakers	Heat	Lakers	Lakers	Correct Prediction
09/10/2020	Heat	Lakers	Lakers	Heat	Wrong Prediction
11/10/2020	Lakers	Heat	Lakers	Lakers	Correct Prediction
					<b>83.33</b>

 Correct Prediction  
 Wrong Prediction

- As expected, by updating the input data with all the recent matches results, the upsets and changes in individual players performances, was accounted for and a result, the accuracy in the Conference Finals increased to 90%
- For the NBA Finals, the accuracy decreased compared to the Conference Finals, however it did increase compare to round 1 and 2 of the playoffs. Still a 83.33% is considerably high considering there are only two teams.
- It can easily be interpreted that the chosen predictive algorithm had predicted a 4-1 series in favor of the Lakers, matching what most analysts had predicted for the finals outcome

# Future Work and Recommendations

- This project already went through a first update adding the results of the playoff games every 2 or 3 days, with a minimum of 1 game played by team. This was done to try to account for the upsets due to playing in a bubble environment and/or due to key players injuries.
- On the first update, the specific match-ups has not been accounted for and I still strongly believe that this can affect the results.
- For the next update of this project I will add player stats and playoff experience
- The most impressive result was the prediction of the championship team despite the Los Angeles Clippers being the favorite to win it all

Predicted team to win:



Result:

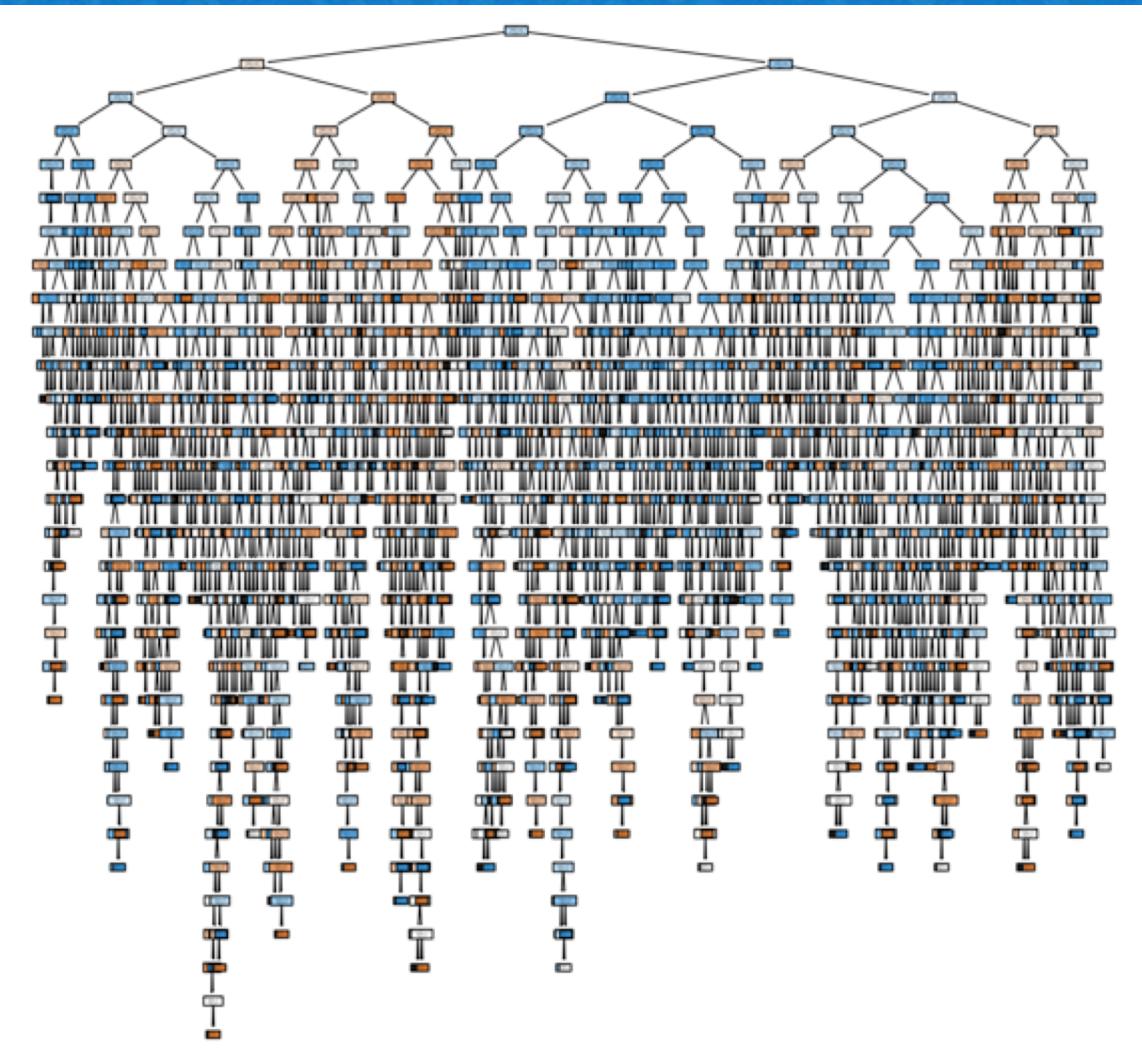


# THANK YOU

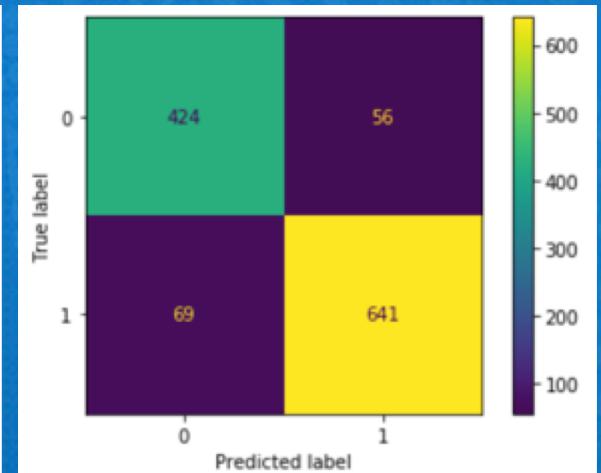


# Appendix

One of the Decision trees:



Confusion Matrix



- Confusion Matrix example comes from the 2018-2019 1 season dataset
- Decision Tree comes from the 2010 – 2020 dataset using all teams as input

