ECE 208: Project Presentation

Kevin Fobare, Ryan Pappania, Rodney Boone, Mukhtar Suleman

Presentation Overview

- Purpose
- Dataset Examples
- Approach
- Model description
- Loss function
- Issues
- Evaluations
- Results
- Summary

PURPOSE

- The original purpose of this project was to create a machine learning model that could be used to predict whether a player would perform over or under their season average performance
- The project was then taken a step further, predicting the performance stats of any player based on their productivity throughout the season and the opponent they would face
- The sport chosen for our model was Basketball

Dataset Player Example

Rk	G	Date	Age	Tm		Opp		GS	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	+/-
1	1	####	35-02	PHO	@	GSW	W (+4	1	####	7	22	0.32	1	2	0.5	3	5	0.6	2	9	11	3	0	1	2	1	18	9.1	0
2	2	####	35-02	PHO	@	LAL	L (-5)	1	####	14	28	0.5	1	5	0.2	10	13	0.77	0	11	11	2	1	1	8	2	39	21.4	18
3	3	####	35-02	PHO		UTA	W (+2	1	####	8	11	0.73	1	4	0.25	9	9	1	0	4	4	7	1	1	2	1	26	26.9	30
4	4	####	35-03	PHO		SAS	L (-1)	1	####	12	19	0.63	1	3	0.33	1	2	0.5	0	2	2	7	2	0	5	0	26	19.6	3
5	5	####	35-03	PHO		SAS	L (-11	1	####	10	15	0.67	3	5	0.6	5	9	0.56	0	6	6	1	0	2	3	3	28	19.6	-17
6	6	####	35-03	PHO	@	PHI	L (-12	1	####	9	20	0.45	1	4	0.25	12	14	0.86	1	7	8	3	1	2	5	1	31	21.7	-13
7	7	####	35-03	PHO	@	DET	W (+1	1	####	14	27	0.52	1	3	0.33	12	12	1	0	4	4	5	1	2	3	1	41	31.4	7
8	8	####	35-04	PHO	@	CHI	W (+1	1	####	7	16	0.44	2	5	0.4	9	9	1	1	6	7	9	2	1	6	2	25	21.3	5
9	9	####	35-04	PHO		LAL	L (-3)	1	####	13	27	0.48	4	6	0.67	8	10	0.8	1	8	9	5	0	0	4	4	38	24.5	8
10	10	####	35-04	PHO		OKC	L (-12	1	####	7	18	0.39	3	5	0.6	11	11	1	0	9	9	4	1	2	3	1	28	22.7	-13
11	11	####	35-04	PHO		MIN	W (+1	1	####	11	15	0.73	2	2	1	7	7	1	0	6	6	6	0	1	2	3	31	28.4	27
12	12	####	35-04	PHO	@	UTA	W (+3	1	####	15	22	0.68	6	8	0.75	2	2	1	0	9	9	9	0	0	1	1	38	36.2	-8
13	13	####	35-05	PHO	@	UTA	W (+3	1	####	14	27	0.52	4	9	0.44	7	7	1	0	8	8	10	2	2	7	2	39	30.7	4
14	14	####	35-05	PHO		POR	W (+1	1	####	13	21	0.62	2	2	1	3	3	1	0	4	4	9	0	1	3	4	31	25.1	20
15	15	####	35-05	PHO		GSW	W (+8	1	####	7	14	0.5	3	4	0.75	15	15	1	0	8	8	2	0	2	6	2	32	23.4	5
16		####	35-05	PHO	@	MEM	W (+2	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inactive
17		####	35-05	PHO	@	NYK	W (+3	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did N	Did Not Play
18	16	####	35-06	PHO	@	TOR	L (-7)	1	####	11	30	0.37	2	8	0.25	6	6	1	0	4	4	6	1	0	1	1	30	18.4	-1
19	17	####	35-06	PHO		DEN	L (-8)	1	####	8	25	0.32	1	3	0.33	13	13	1	0	4	4	11	1	3	1	2	30	25.9	-6
20	18	####	35-06	PHO		MEM	W (+7	1	####	10	14	0.71	2	3	0.67	5	7	0.71	0	2	2	5	1	1	2	1	27	23.8	8
21	19	####	35-06	PHO	@	LAL	L (-3)	1	####	12	17	0.71	4	7	0.57	3	4	0.75	0	7	7	4	0	1	5	5	31	22.1	6
22		####	35-07	PHO		SAC	L (-8)	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inactive
23		####	35-07	PHO		GSW	W (+3	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inacti	Inactive
24	20	####	35-07	PHO		BRK	L (-4)	1	####	9	18	0.5	1	5	0.2	8	10	0.8	1	5	6	4	1	2	2	3	27	21.4	0
25	21	####	35-07	PHO		NYK	L (-17	1	####	10	21	0.48	4	9	0.44	5	7	0.71	0	3	3	6	0	0	0	1	29	22.2	-7

NBA Teams Defensive Stats

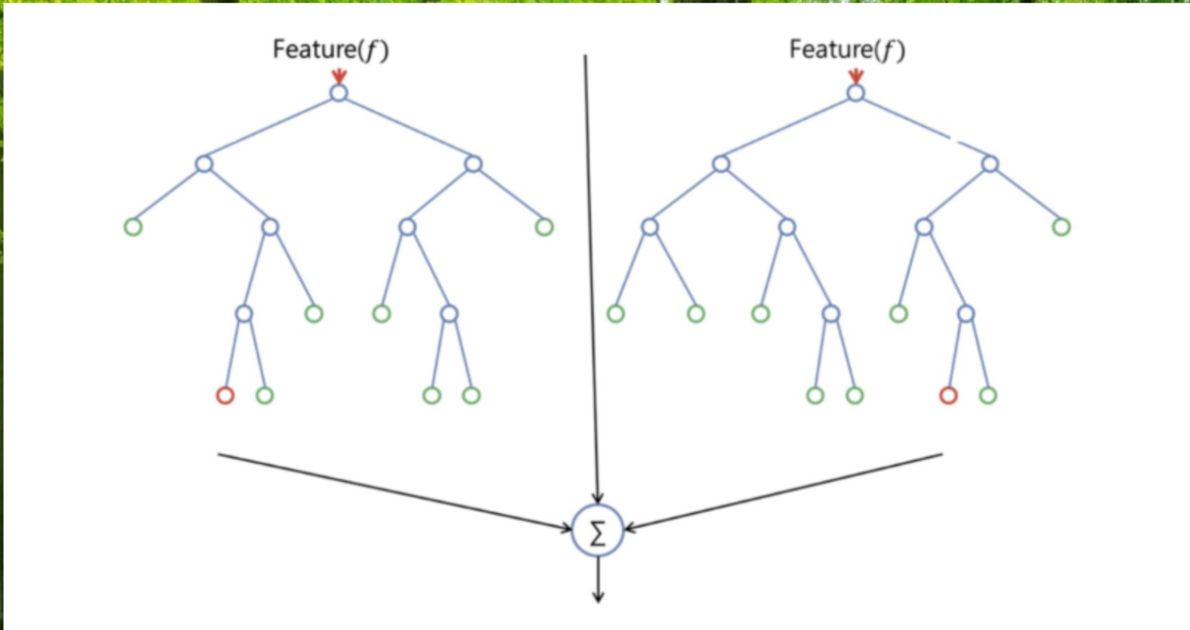
NBA TEAM		PA/G	S/G	B/G	T/G	Defensive Efficency	DRB
ATL		120.5	7.5	4.5	13.5	1.156	32.2
BOS		109.2	6.8	6.5	11.9	1.08	35.6
BKN		113.3	6.8	5.2	13.1	1.124	32.6
CHA		116.8	6.9	4.5	13.8	1.164	31
CHI		113.7	7.8	4.9	12.2	1.127	32.6
CLE		110.2	7.4	4.6	13.6	1.09	33.4
DAL		115.6	6.9	5	12.5	1.118	33.2
DEN		109.6	7.1	5.4	12.6	1.095	33.7
DET		119	6.5	4.7	15.2	1.151	32.8
GSW		115.2	7	4.6	14.3	1.116	34.6
HOU		113.2	7.8	4.6	12.7	1.096	34
IND		120.2	7.5	5.8	12.7	1.143	31.4
LAC		112.3	7.8	5	13.2	1.115	32.9
LAL		117.4	7.3	5.5	13.8	1.118	34.9
MEM		112.8	8.2	6.1	15.1	1.106	31.7
MIA		108.4	7.4	3.4	12.7	1.089	33
MIL		116.4	6.7	5	12.8	1.124	34.8
MIN		106.5	7.9	5.9	14.2	1.056	34.2
NOP		110.7	8.3	4.6	13.3	1.093	33.6
NYK		108.2	7.3	4.3	13.1	1.095	32.5
OKC		112.7	8.4	6.5	12.7	1.079	33.2
ORL		108.4	8.1	5.2	14.7	1.071	31.8
PHI		111.5	8.2	6	12	1.099	31.9
PHX		113.2	7.4	5.9	14.9	1.113	33.9
POR		115.4	7.6	4.3	15.2	1.133	30.1
SAC		114.8	7.7	4.2	13.1	1.114	33.2
SAS		118.6	7.1	6.3	15.1	1.13	33.9
TOR		118.8	7.7	4.7	14	1.151	31.8
UTA		120.5	6.5	5.6	15.7	1.163	33.2
WAS		123	7.6	5.1	14	1.161	31.9

Approach

- We started out using the Linear regression model a baseline for predicting whether a player would exceed or fall short of their season average
- Random Forest Model to predict a player performance stats individually based on their efficiency throughout their season and the opponent team
- We switched to the Random Forest Model because of the randomness it provides and additional stats we can add easily

Random Forest Model

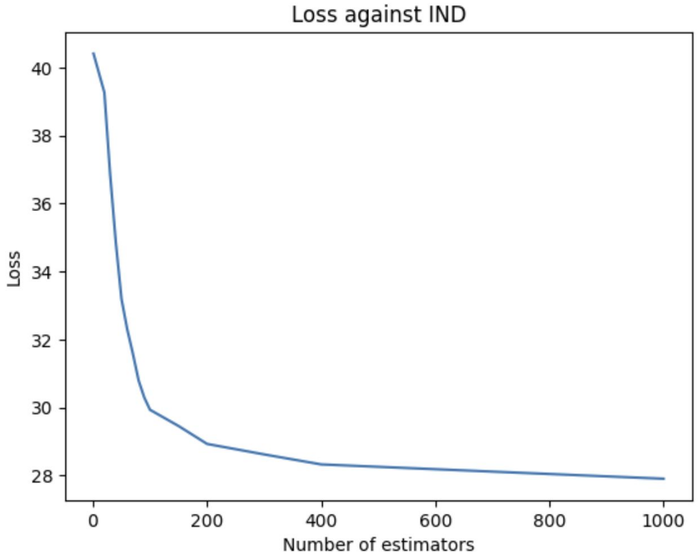
- Builds Multiple decision trees
- Chooses best feature from random subset of features
- Change number of estimators

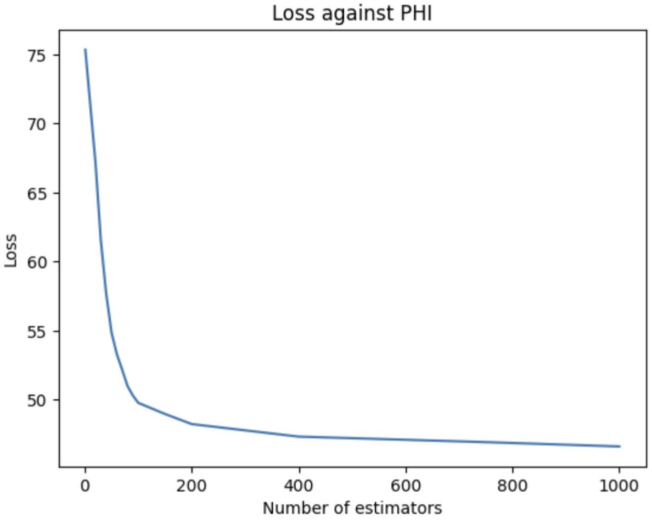


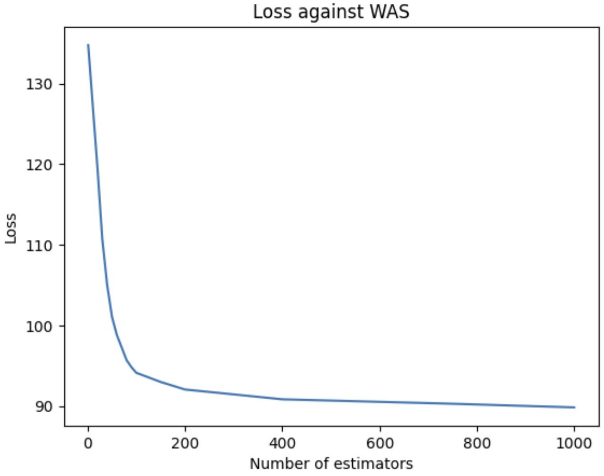
https://builtin.com/data-science/random-forest-algorithm

Loss function

- Used Squared error
- Showed that around 300 estimators was the best balance between computation time and minimizing loss

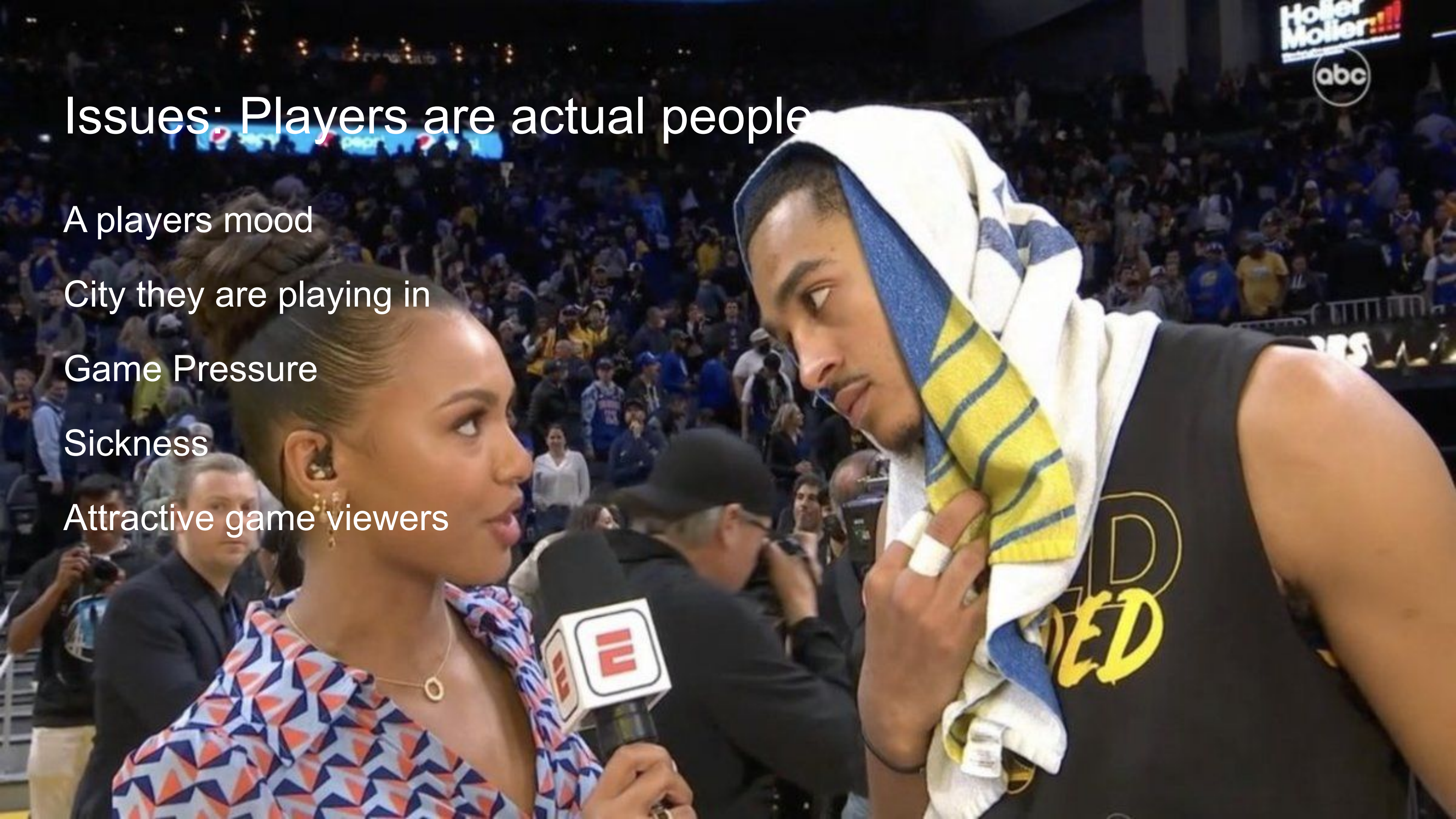


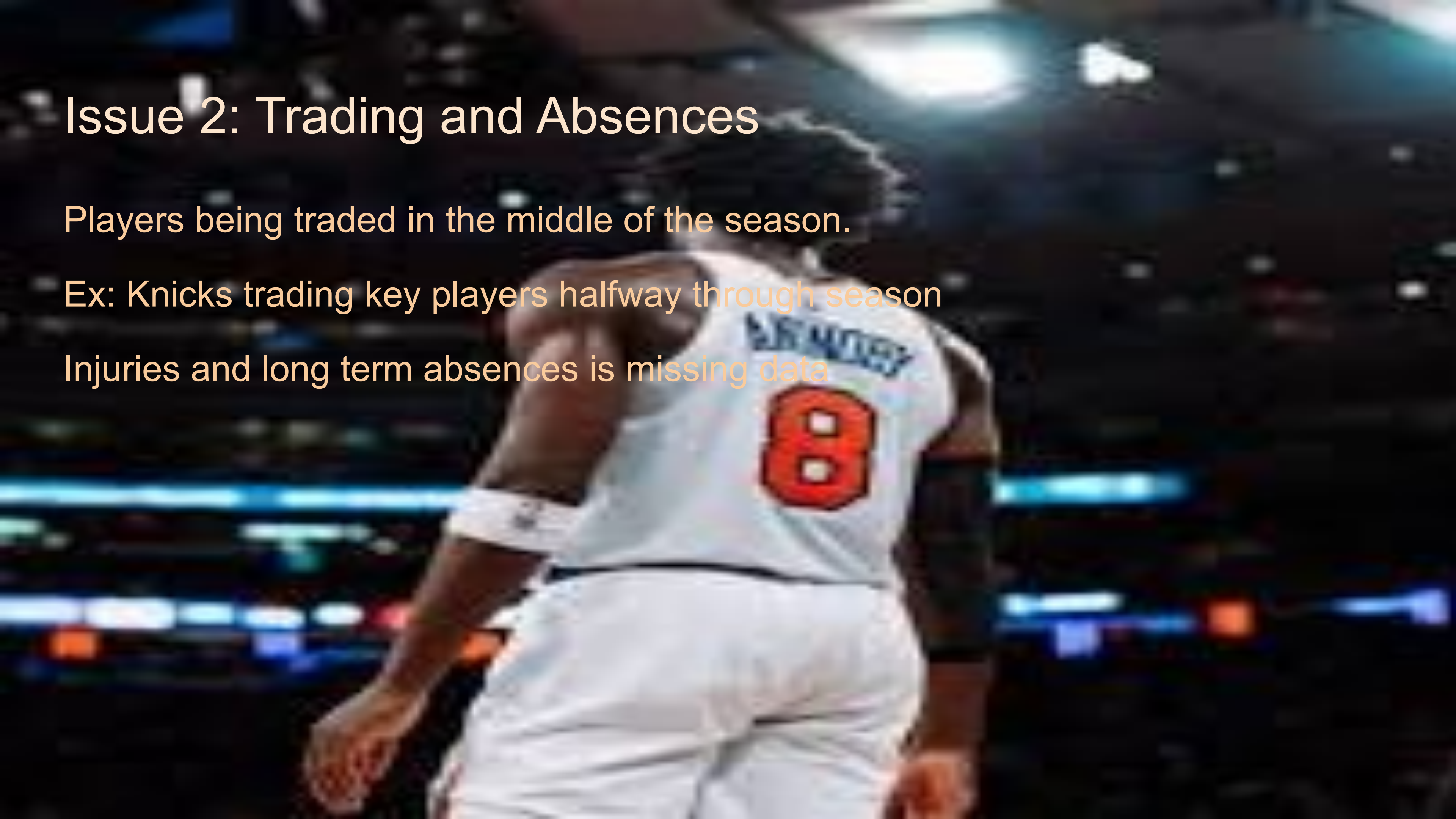




Issues: Players are actual people

- A players mood
- City they are playing in
- Game Pressure
- Sickness
- Attractive game viewers





Issue 2: Trading and Absences

Players being traded in the middle of the season.

Ex: Knicks trading key players halfway through season

Injuries and long term absences is missing data

Issue 3: Bench player

Data does not incorporate a players rest per game

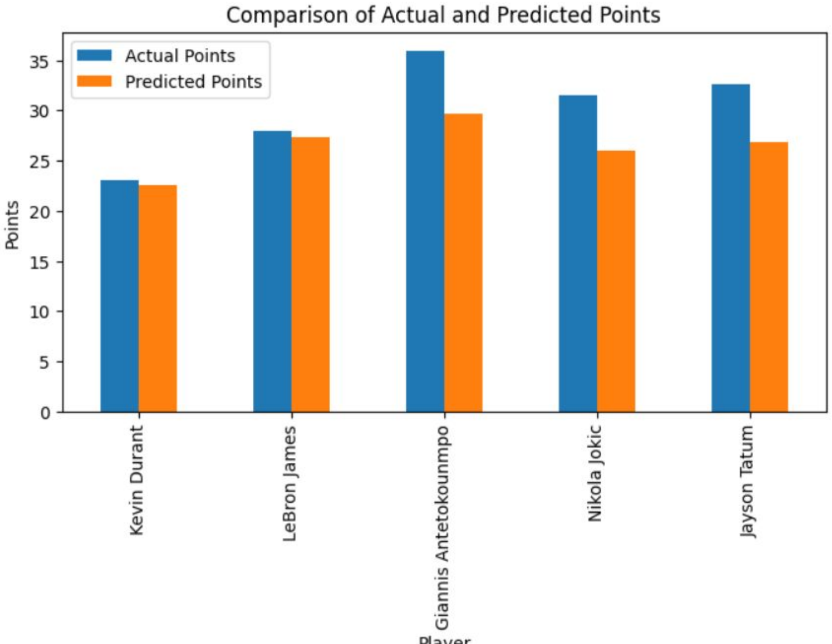
Bench player may sub in during playoff/overtime to save energy of starters

Assumes no bench players have to replace injured player

Evaluation

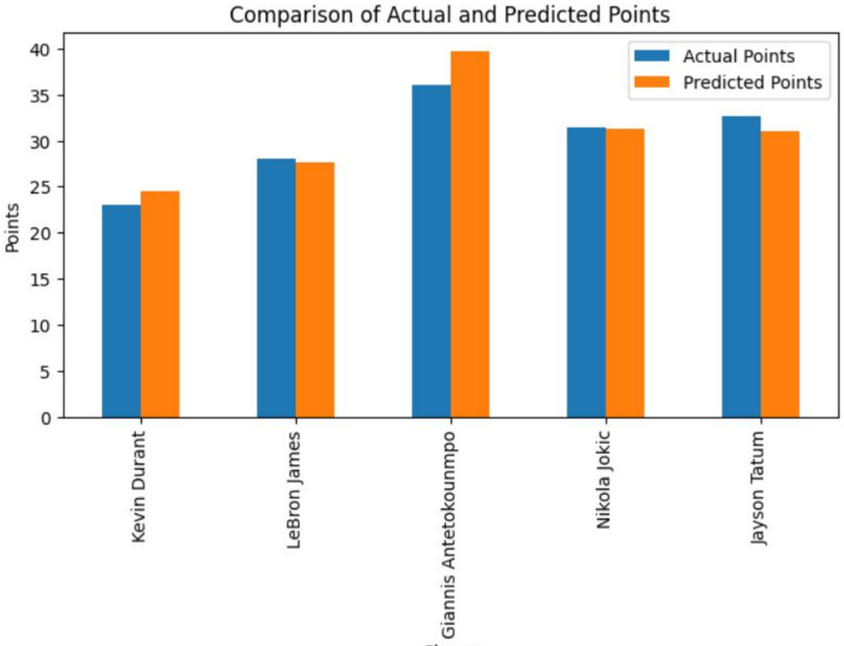
Linear Regression with

no opponent data



Random Forest Model

with opponent defensive data



Results

N-estimators = 100,

Used starting 5,

Results: either sweep or losing team picked up one game

Big gap in points per game

Game Results

<u>Mavericks</u>

doncic.csv against BOS: 35

kyrie.csv against BOS: 24

jonesjr.csv against BOS: 5

washington.csv against BOS: 15

gafford.csv against BOS: 5

Total: 84

<u>Celtics</u>

jrue.csv against DAL: 12

white.csv against DAL: 14

brown.csv against DAL: 30

tatum.csv against DAL: 34

porzingis.csv against DAL: 21

Total: 111

- Celtics have best starting 5 players
- On paper this team should win the playoffs/finals with full sweeps

Results

N - estimator = 100

1 bench player

1 point win games

Finals Prediction Match up

N - estimator = 300

(best estimator)

1 bench player

Most accurate representation

Most tiebreakers

(close)

Our finals championship prediction is…

DALLAS MAVERICKS

Game Results

Mavericks

doncic.csv against CLE: 41

kyrie.csv against CLE: 28

jonesjr.csv against CLE: 7

washington.csv against CLE: 14

gafford.csv against CLE: 10

hardawayjr.csv against CLE: 9

Total: 109

Cavaliers

garland.csv against DAL: 11

mitchell.csv against DAL: 31

strus.csv against DAL: 15

mobley.csv against DAL: 13

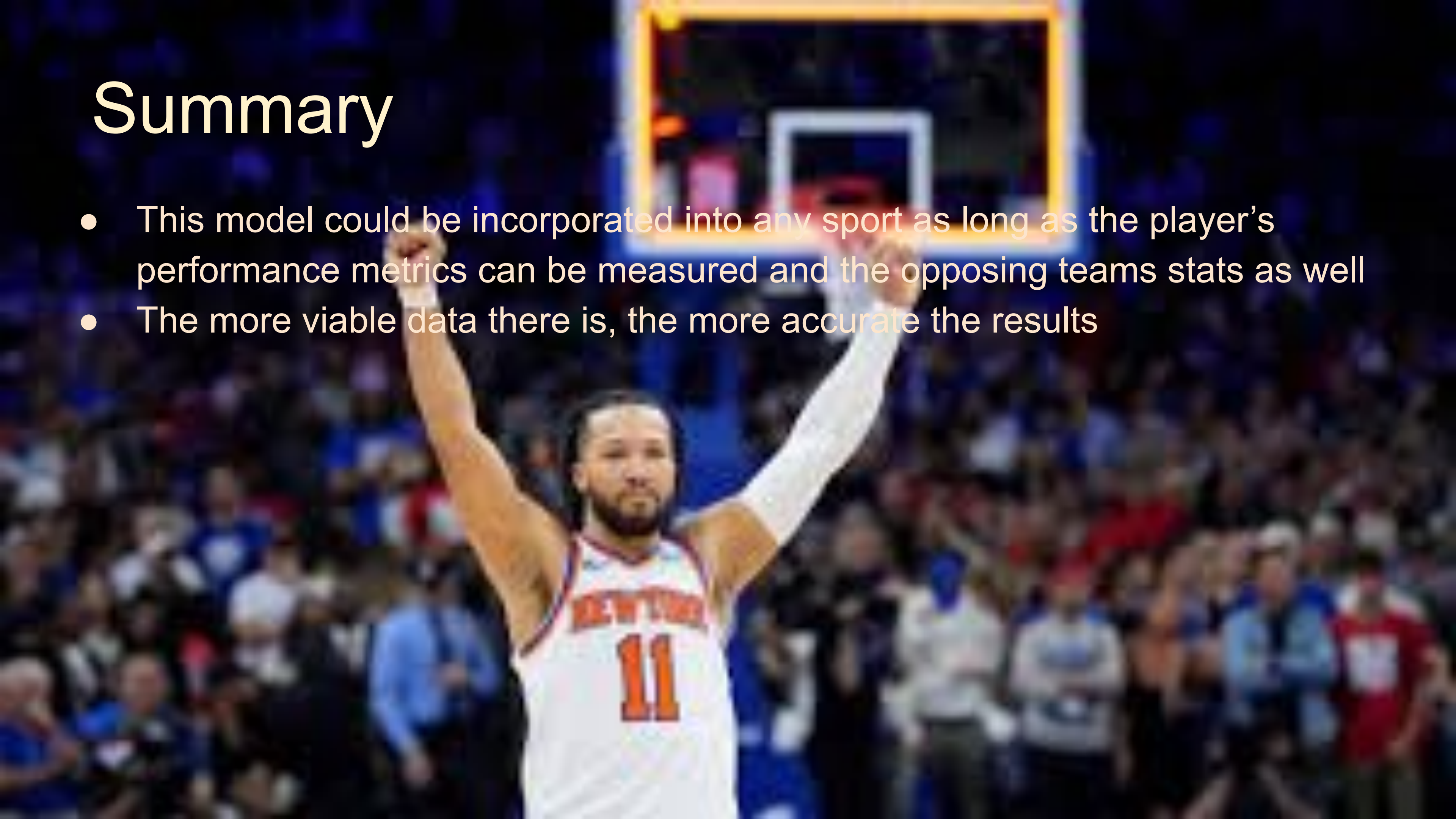
allen.csv against DAL: 21

levert.csv against DAL: 17

Total: 108

Summary

- This model could be incorporated into any sport as long as the player's performance metrics can be measured and the opposing teams stats as well
- The more viable data there is, the more accurate the results



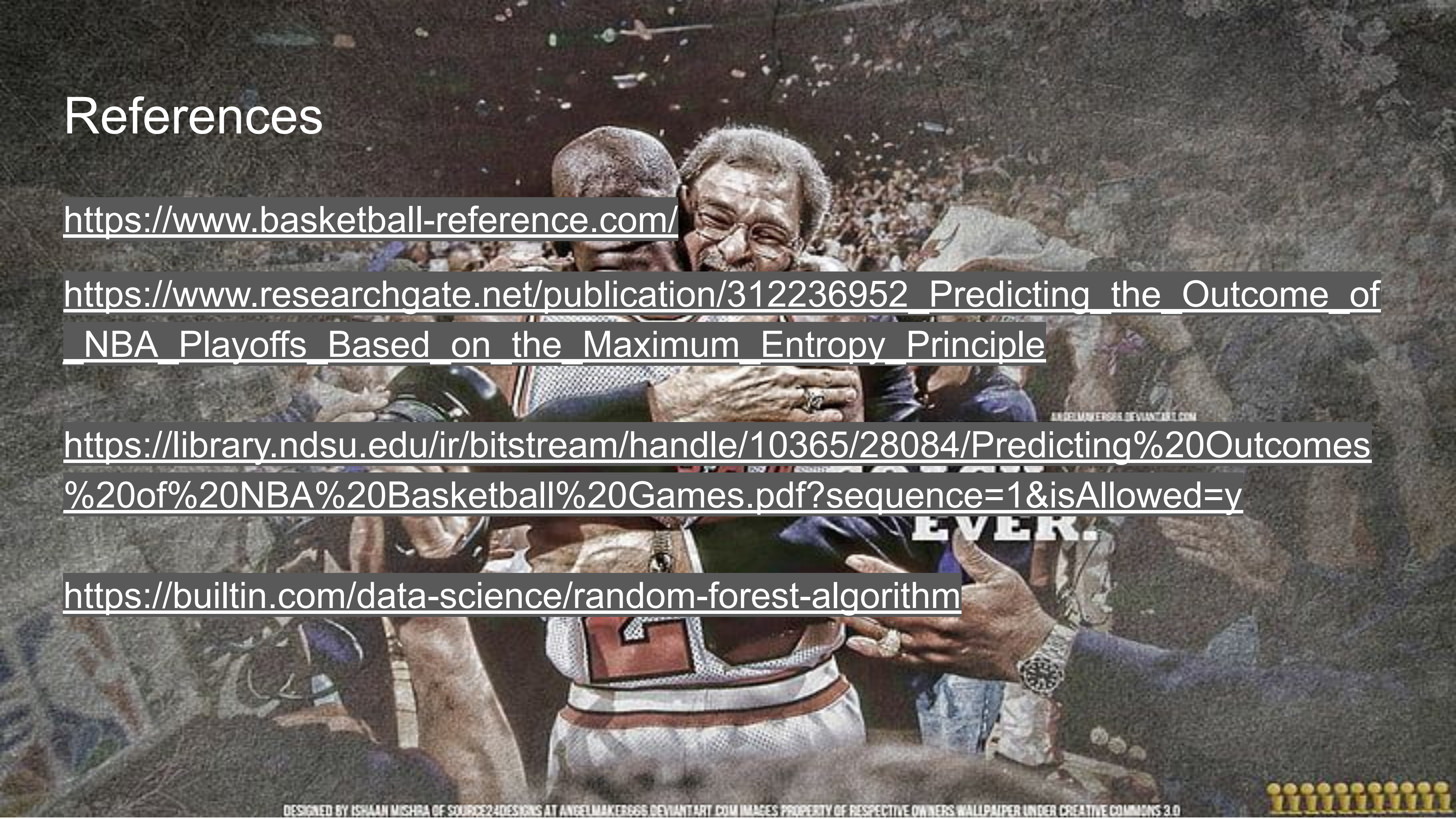
References

https://www.basketball-reference.com/

https://www.researchgate.net/publication/312236952_Predicting_the_Outcome_of_NBA_Playoffs_Based_on_the_Maximum_Entropy_Principle

https://library.ndsu.edu/ir/bitstream/handle/10365/28084/Predicting%20Outcomes%20of%20NBA%20Basketball%20Games.pdf?sequence=1&isAllowed=y

https://builtin.com/data-science/random-forest-algorithm



Questions