Minefield / Sports Betting AI

Eli Souvannarath, Sarah Cullen, Ty Pleasant, Alex Mull-Dreyer
March 5, 2022

Overview

- Goals of Project
- Original Data
- Final Data
- Models
- Results/Conclusion

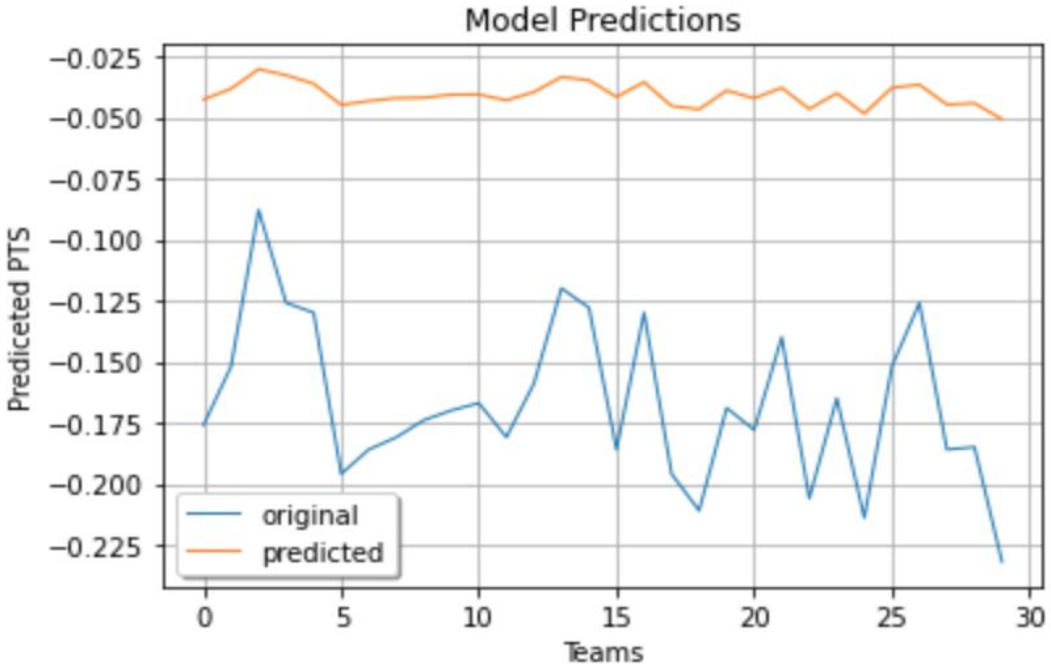
Overview

- **Project Goal**
  - Build a model to aid betting on NBA teams
- **Scraping**
  - Scraped team data from Basketball-reference.com using NBA data from 1980-2022 to predict 2022 season data
- **Technologies**
  - Python (Pandas, BeautifulSoup, Matplotlib), Jupyter, SciKit-Learn, HTML5, CSS3, VS Code
- **Modeling**
  - RandomForestRegressor using SciKit-Learn, focused on Win-Loss %
  - Linear Regression per season per team

Original Data

Season	Lg	Tm	Age	Ht.	Wt.	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV
2021-22	NBA	ATL	2.8%	0.3%	-1.4%	62	-14.5%	-13.2%	-13.4%	+.001	-12.8%	-13.6%	+.003	-13.3%	-13.3%	-.000	-22.9%	-22.2%	-.007	-20.0%	-15.8%	-16.8%	-13.5%	-15.1%	-20.2%	-20.8%
2020-21	NBA	ATL	5.4%	-0.1%	-0.9%	72	6.9%	7.9%	3.5%	+.019	11.2%	-0.6%	+.039	6.5%	6.2%	+.001	14.6%	11.4%	+.022	15.0%	12.9%	13.4%	8.2%	-3.8%	0.3%	-12.2%
2019-20	NBA	ATL	-4.0%	-0.8%	0.5%	67	-18.0%	-19.7%	-19.4%	-.002	-24.6%	-20.4%	-.018	-17.5%	-18.7%	+.008	-14.3%	-18.4%	+.038	-30.8%	-20.8%	-23.3%	-24.2%	-22.5%	-18.6%	-22.3%
2018-19	NBA	ATL	-1.2%	0.3%	1.4%	82	0.8%	8.4%	7.3%	+.005	16.4%	19.3%	-.009	5.1%	0.4%	+.023	11.2%	16.0%	-.032	28.5%	4.9%	10.0%	8.8%	5.8%	20.4%	9.5%
2017-18	NBA	ATL	-9.0%	0.3%	-3.2%	82	-0.9%	0.2%	1.4%	-.005	25.8%	19.0%	+.019	-7.6%	-6.5%	-.006	-12.5%	-18.9%	+.057	-11.8%	-3.6%	-5.5%	0.4%	-5.1%	-12.3%	-1.4%
2016-17	NBA	ATL	-1.1%	0.3%	0.9%	82	0.3%	-1.4%	-0.1%	-.006	-10.6%	-8.1%	-.009	1.7%	4.0%	-.011	15.8%	24.5%	-.055	24.0%	0.8%	5.3%	-7.7%	-10.0%	-18.3%	5.5%
2015-16	NBA	ATL	1.4%	0.0%	-0.5%	82	0.5%	1.5%	3.3%	-.008	-0.4%	8.1%	-.030	2.2%	1.1%	+.005	-5.0%	-5.6%	+.005	-5.0%	6.2%	3.8%	-0.5%	0.4%	27.9%	5.1%
2014-15	NBA	ATL	0.7%	0.3%	-0.9%	82	-0.5%	2.0%	0.2%	+.008	6.5%	1.7%	+.017	0.4%	-0.5%	+.005	-3.1%	-2.6%	-.004	0.3%	1.8%	1.5%	3.4%	9.4%	16.6%	-6.7%
2013-14	NBA	ATL	1.8%	-0.3%	2.3%	82	-0.1%	-0.7%	0.7%	-.006	8.8%	11.3%	-.008	-3.6%	-3.6%	+.000	20.2%	10.1%	+.066	-5.9%	-1.1%	-2.2%	1.7%	2.4%	-11.7%	2.6%
2012-13	NBA	ATL	-2.9%	-0.6%	0.0%	82	22.8%	27.0%	24.2%	+.010	43.5%	42.9%	+.001	22.8%	18.0%	+.019	13.0%	16.9%	-.025	16.3%	25.5%	23.3%	35.5%	23.9%	21.8%	32.4%
2011-12	NBA	ATL	1.8%	-0.4%	-1.8%	66	-18.1%	-18.2%	-16.8%	-.008	-2.0%	-6.8%	+.018	-21.5%	-19.7%	-.012	-23.8%	-19.8%	-.039	-14.4%	-16.0%	-15.6%	-17.8%	7.8%	-11.1%	-17.6%
2010-11	NBA	ATL	3.4%	-0.1%	0.0%	82	-0.6%	-6.6%	-5.5%	-.006	-4.2%	-1.9%	-.008	-7.1%	-6.4%	-.003	-7.3%	-9.7%	+.020	-21.4%	0.4%	-5.8%	0.7%	-16.0%	-17.4%	14.1%
2009-10	NBA	ATL	2.7%	0.6%	0.0%	82	0.8%	7.6%	5.4%	+.010	-12.2%	-10.9%	-.005	12.6%	11.0%	+.007	-5.5%	-8.2%	+.021	11.4%	1.6%	4.2%	8.0%	-1.8%	9.8%	-6.5%
2008-09	NBA	ATL	6.6%	-0.6%	-0.5%	82	-0.8%	-0.6%	-1.5%	+.004	55.5%	51.5%	+.009	-9.0%	-12.0%	+.016	-10.6%	-6.4%	-.035	-13.7%	-1.7%	-5.2%	-8.1%	0.5%	-16.1%	-14.3%
2007-08	NBA	ATL	0.4%	0.4%	1.9%	82	-0.1%	5.1%	2.8%	+.010	12.6%	3.9%	+.028	4.1%	2.6%	+.007	2.6%	1.1%	+.011	3.3%	6.1%	5.3%	14.7%	-1.5%	0.4%	-6.4%
2006-07	NBA	ATL	6.2%	-0.6%	-2.3%	82	0.0%	-3.9%	-1.9%	-.009	-19.6%	-10.1%	-.039	-1.3%	-0.1%	-.005	1.3%	-0.2%	+.011	-8.7%	3.6%	-0.4%	-3.2%	3.7%	13.2%	1.7%
2005-06	NBA	ATL	-12.7%	0.5%	0.5%	82	0.1%	0.1%	-2.6%	+.013	39.5%	18.6%	+.055	-4.4%	-6.3%	+.009	16.9%	10.7%	+.040	-2.8%	-4.4%	-3.9%	0.7%	-6.7%	14.5%	-2.7%
2004-05	NBA	ATL	-0.8%	0.4%	2.8%	82	-0.3%	4.0%	2.2%	+.008	-27.4%	-22.1%	-.023	9.5%	7.9%	+.006	-7.6%	0.9%	-.066	10.4%	-6.9%	-1.9%	-2.1%	0.3%	-15.7%	-2.3%
2003-04	NBA	ATL	-3.7%	0.5%	-0.5%	82	0.0%	-1.0%	1.5%	-.011	4.2%	9.5%	-.017	-1.9%	-0.2%	-.008	-3.8%	-1.7%	-.016	6.3%	-2.0%	0.2%	-1.8%	2.6%	-13.7%	-1.2%

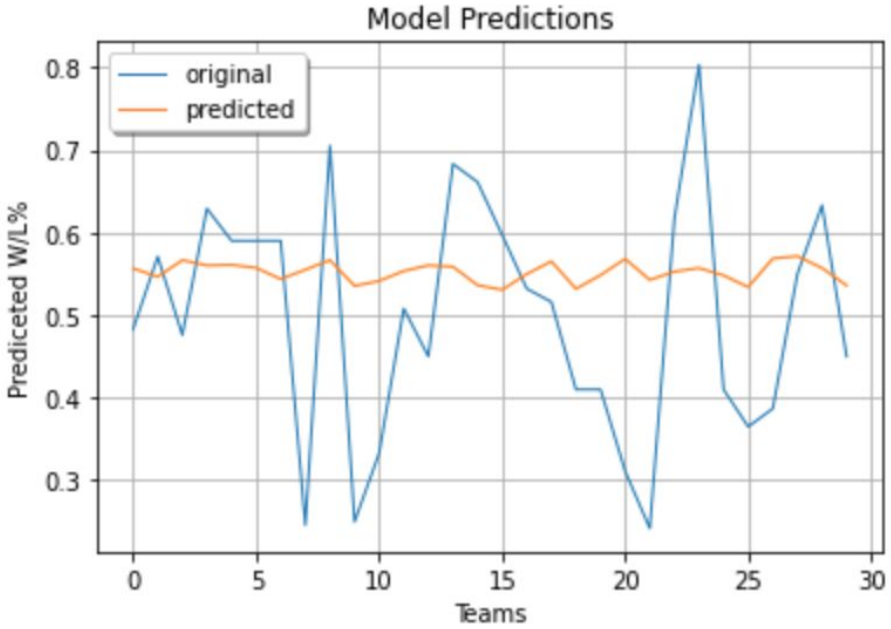
Model 1 - Predicted vs. Actual PTS



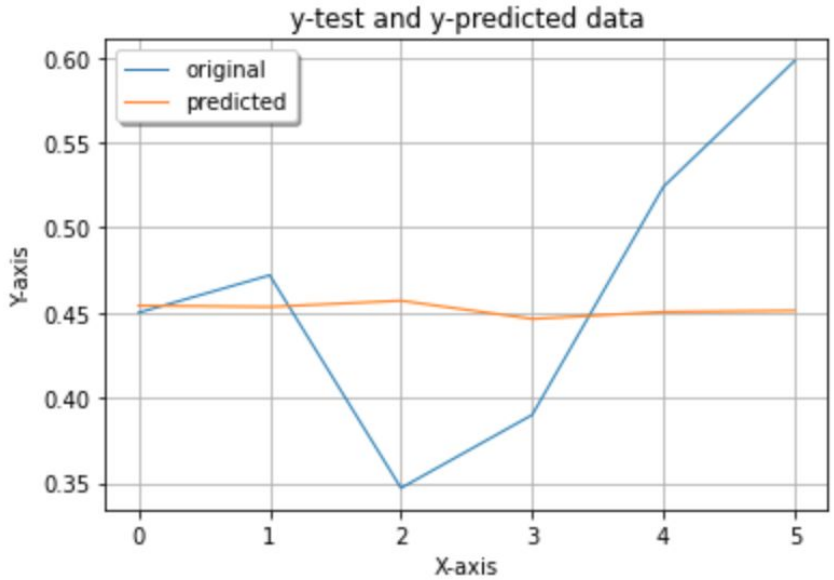
Model 2

- **Random Forest Regression** (Combined & Team by Team)
  - Runs model on all past seasons as training data and the current season as test data
    - Low r-squared
  - Team by Team
    - Each team with separate model
      - R-squared results mixed
    - 12 teams have negative R-squared values

Combined Model - predict Win/Loss% for 2022 season



Team by Team Model



HTML Pages

Results

Conclusions

- Strengthening the model
- Formatting of the data
- Weighting recent data more heavily
- Betting difference than straight odds