NCAA & NBA Playoff Teams Prediction Presentation

—

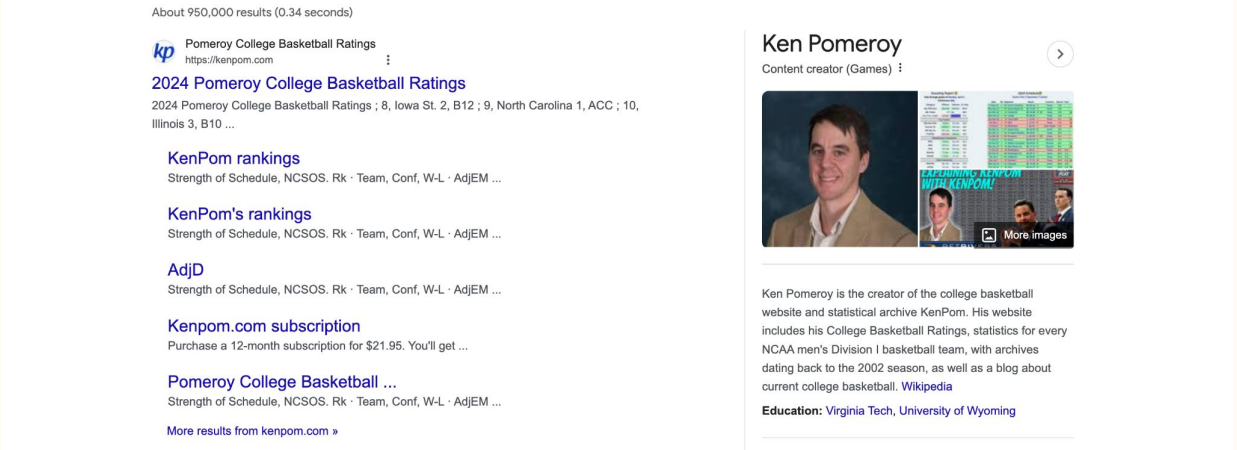
Cole Jennings and Thomas Sniezek

Tractable Data

- NBA Data
  - We added a new variable for making the playoffs (1 for making playoffs, 0 for missing playoffs)
- NCAA Data
  - We added a variable for making the NCAA Tournament (1 for making playoffs, 0 for missing playoffs)
  - We added a binary variable for what conference the team belongs to
  - Scraped Financial data from Sportico
    - OpEx and OpRev

Data Retrieval

- NBA Data was downloaded from Kaggle
  - This was scraped from stats.nba.com
- NCAA Data was downloaded from Kaggle
  - KenPom
  - Scaped Sportico data



Model Specification

- To explore the factors that are involved in both NBA teams making the playoffs and NCAA teams making march madness, we used XGBoost and Logistic Regression Models
- We gathered the feature importances from the XGBoost playoff berth prediction models and compared them to each other
- Compared the accuracies of the two models and tweaked them until they were as similar as possible
- Created a logistic regression model for each set of data
- Using the marginal effects and the mean for each playoff, we found how each feature impacts the model and increases/decreases the likelihood of making the playoffs

Variable Explanations - NCAA

- Adjusted defensive and offensive efficiency (ADJDE and ADJOE)
- ADJOE – Points scored per 100 offensive possessions
- ADJDE – Points allowed per 100 defensive possessions
- Possessions are not recorded officially by statisticians, so estimated using:
- FGA-OR+TO+0.475xFTA
- EFG_O – (FGM + 0.5*3PM) / FGA (on offense)
- EFG_D – (FGM + 0.5*3PM) / FGA (on defense)

Variable Explanation- NBA

ThreePP- Three Point Percentage

OREB- Offensive Rebounds

DREB- Defensive Rebounds

TOV- Turnovers

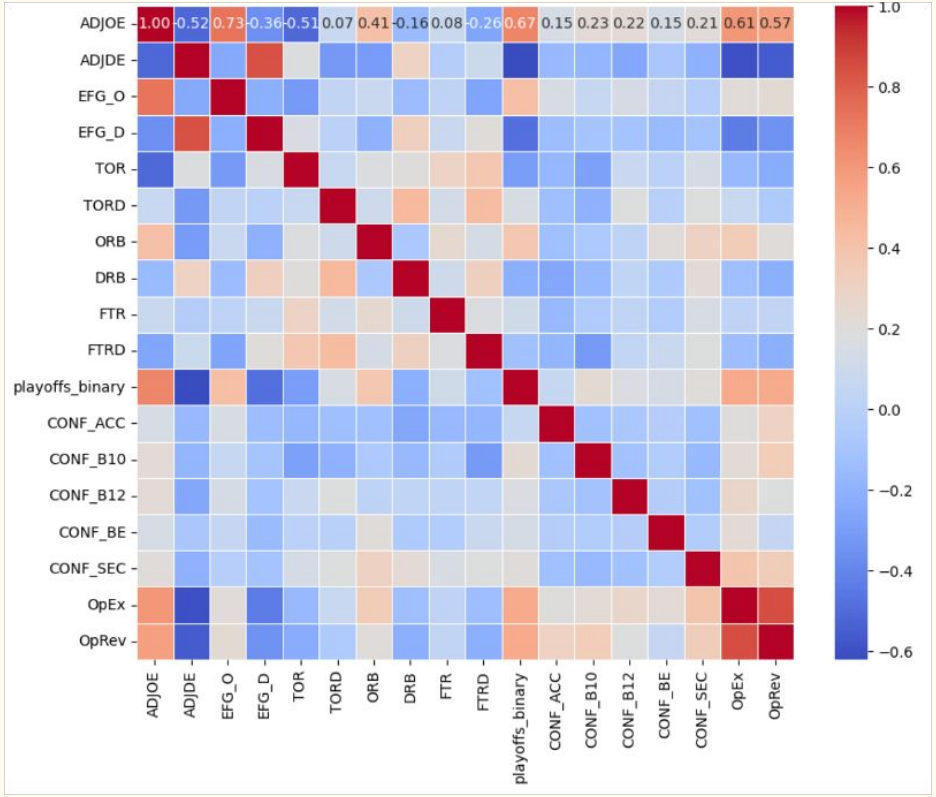
STL- Steals

BLK- Blocks

BLKA- Blocks Against

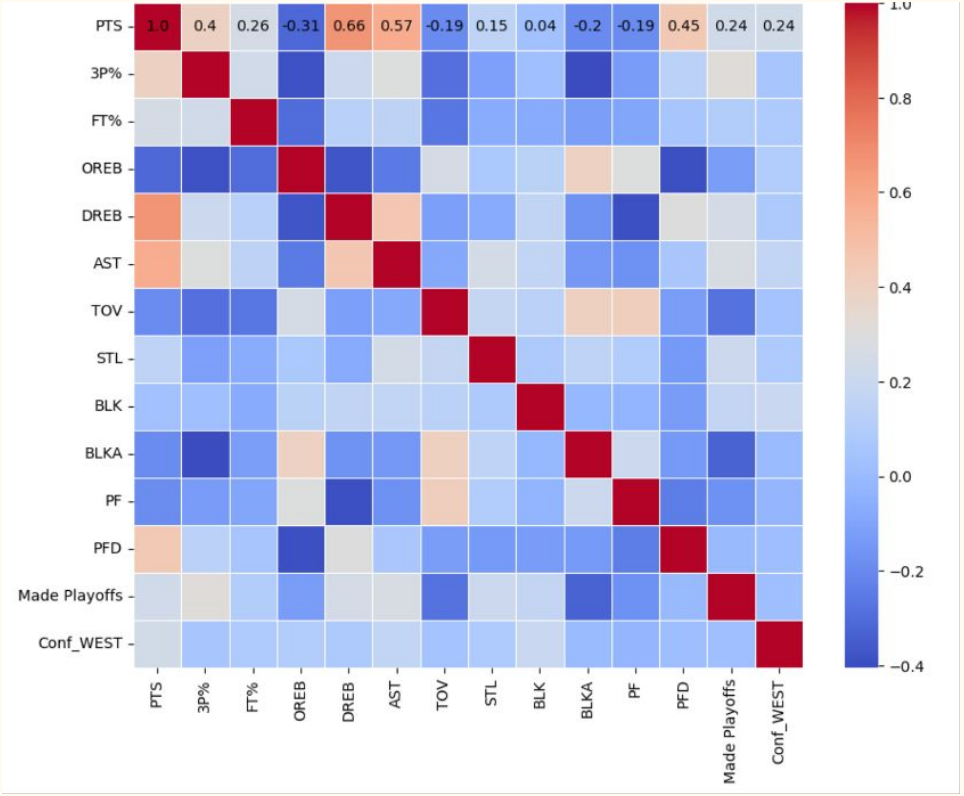
NCAA Variables

- Variables of note
  - ADJOE
  - ADJDE
  - EFG_O
  - EFG_D
  - OpEx
  - OpRev



NBA Variables

- Variables of note
  - BLKA
  - 3P%
  - DREB
  - Steals



Results and Marginal Effects – NCAA

Logit Regression Results

Dep. Variable:	playoffs_binary	No. Observations:	1074
Model:	Logit	Df Residuals:	1060
Method:	MLE	Df Model:	13
Date:	Tue, 30 Apr 2024	Pseudo R-squ.:	0.4744
Time:	13:59:47	Log-Likelihood:	-274.40
converged:	True	LL-Null:	-522.12
Covariance Type:	nonrobust	LLR p-value:	1.369e-97

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-8.3668	5.624	-1.488	0.137	-19.390	2.657
TOR	-0.2182	0.087	-2.514	0.012	-0.388	-0.048
ADJOE	0.2262	0.032	7.130	0.000	0.164	0.288
ADJDE	-0.1792	0.030	-5.960	0.000	-0.238	-0.120
TORD	0.1378	0.068	2.028	0.043	0.005	0.271
ORB	0.0595	0.034	1.732	0.083	-0.008	0.127
DRB	-0.0793	0.047	-1.701	0.089	-0.171	0.012
FTR	0.1014	0.027	3.746	0.000	0.048	0.155
FTRD	-0.0174	0.029	-0.607	0.544	-0.074	0.039
CONF_ACC	-0.7920	0.464	-1.706	0.088	-1.702	0.118
CONF_B10	0.0545	0.455	0.120	0.905	-0.836	0.945
CONF_B12	-0.3101	0.531	-0.584	0.559	-1.350	0.730
CONF_BE	-0.0121	0.474	-0.026	0.980	-0.942	0.918
CONF_SEC	-0.4680	0.492	-0.952	0.341	-1.432	0.496

Logit Marginal Effects

Dep. Variable:	playoffs_binary
Method:	dydx
At:	mean

	dy/dx	std err	z	P>|z|	[0.025	0.975]
TOR	-0.0100	0.004	-2.474	0.013	-0.018	-0.002
ADJOE	0.0104	0.002	5.744	0.000	0.007	0.014
ADJDE	-0.0082	0.002	-5.264	0.000	-0.011	-0.005
TORD	0.0063	0.003	1.996	0.046	0.000	0.013
ORB	0.0027	0.002	1.681	0.093	-0.000	0.006
DRB	-0.0036	0.002	-1.689	0.091	-0.008	0.001
FTR	0.0047	0.001	3.608	0.000	0.002	0.007
FTRD	-0.0008	0.001	-0.606	0.544	-0.003	0.002
CONF_ACC	-0.0364	0.021	-1.701	0.089	-0.078	0.006
CONF_B10	0.0025	0.021	0.120	0.905	-0.039	0.044
CONF_B12	-0.0143	0.024	-0.588	0.557	-0.062	0.033
CONF_BE	-0.0006	0.022	-0.026	0.980	-0.043	0.042
CONF_SEC	-0.0215	0.022	-0.958	0.338	-0.066	0.023

Results and Marginal Effects – NCAA

- (dy/dx)/(mean(playoffs_binary))
- .0104/.1899 = .055; .055*100 = 5.5%
- Every marginal increase of ADJOE (additional point scored/100 possessions) is associated with a 5.5% increase in the likelihood of making March Madness
- Every marginal increase of ADJDE (additional point scored on/100 defensive possessions) is associated with a 4.3% decrease in the likelihood of making March Madness

Results and Marginal Effects – NBA

Logit Regression Results

Dep. Variable:	MadePlayoffs	No. Observations:	566
Model:	Logit	Df Residuals:	552
Method:	MLE	Df Model:	13
Date:	Wed, 24 Apr 2024	Pseudo R-squ.:	0.2912
Time:	18:07:24	Log-Likelihood:	-276.96
converged:	True	LL-Null:	-390.76
Covariance Type:	nonrobust	LLR p-value:	2.802e-41

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-27.4919	5.753	-4.779	0.000	-38.768	-16.216
PTS	-0.1040	0.035	-3.002	0.003	-0.172	-0.036
ThreePP	0.3724	0.072	5.178	0.000	0.231	0.513
FTP	0.0575	0.041	1.407	0.159	-0.023	0.138
OREB	0.3669	0.111	3.307	0.001	0.149	0.584
DREB	0.4637	0.091	5.101	0.000	0.286	0.642
TOV	-0.5504	0.119	-4.616	0.000	-0.784	-0.317
AST	0.1236	0.073	1.703	0.089	-0.019	0.266
STL	1.1986	0.161	7.465	0.000	0.884	1.513
BLK	0.4413	0.145	3.036	0.002	0.156	0.726
BLKA	-1.0473	0.194	-5.409	0.000	-1.427	-0.668
PF	0.0472	0.078	0.608	0.543	-0.105	0.199
PFD	0.0117	0.015	0.780	0.435	-0.018	0.041
Conf_WEST	-0.2409	0.227	-1.063	0.288	-0.685	0.203

Logit Marginal Effects

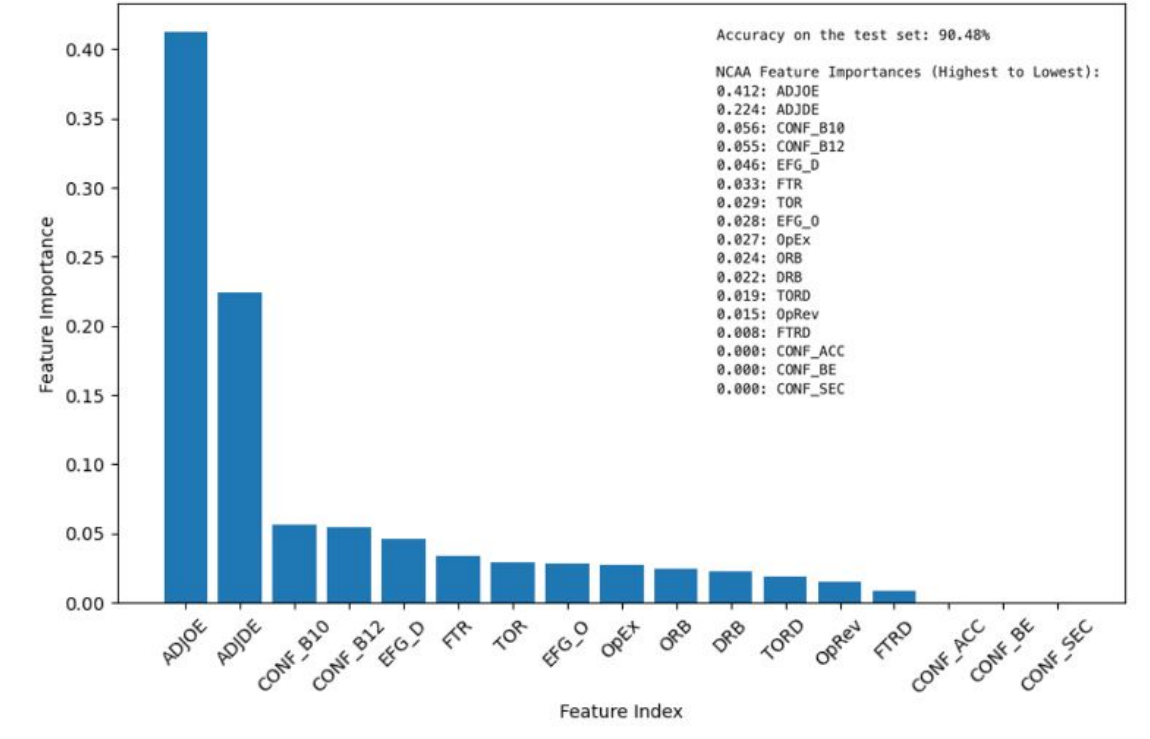
Dep. Variable:	MadePlayoffs
Method:	dydx
At:	mean

	dy/dx	std err	z	P>|z|	[0.025	0.975]
PTS	-0.0257	0.009	-3.002	0.003	-0.042	-0.009
ThreePP	0.0920	0.018	5.180	0.000	0.057	0.127
FTP	0.0142	0.010	1.407	0.159	-0.006	0.034
OREB	0.0906	0.027	3.309	0.001	0.037	0.144
DREB	0.1145	0.022	5.097	0.000	0.070	0.159
TOV	-0.1360	0.030	-4.608	0.000	-0.194	-0.078
AST	0.0305	0.018	1.704	0.088	-0.005	0.066
STL	0.2960	0.040	7.464	0.000	0.218	0.374
BLK	0.1090	0.036	3.038	0.002	0.039	0.179
BLKA	-0.2587	0.048	-5.417	0.000	-0.352	-0.165
PF	0.0116	0.019	0.608	0.543	-0.026	0.049
PFD	0.0029	0.004	0.780	0.435	-0.004	0.010
Conf_WEST	-0.0595	0.056	-1.062	0.288	-0.169	0.050

Results and Marginal Effects-NBA

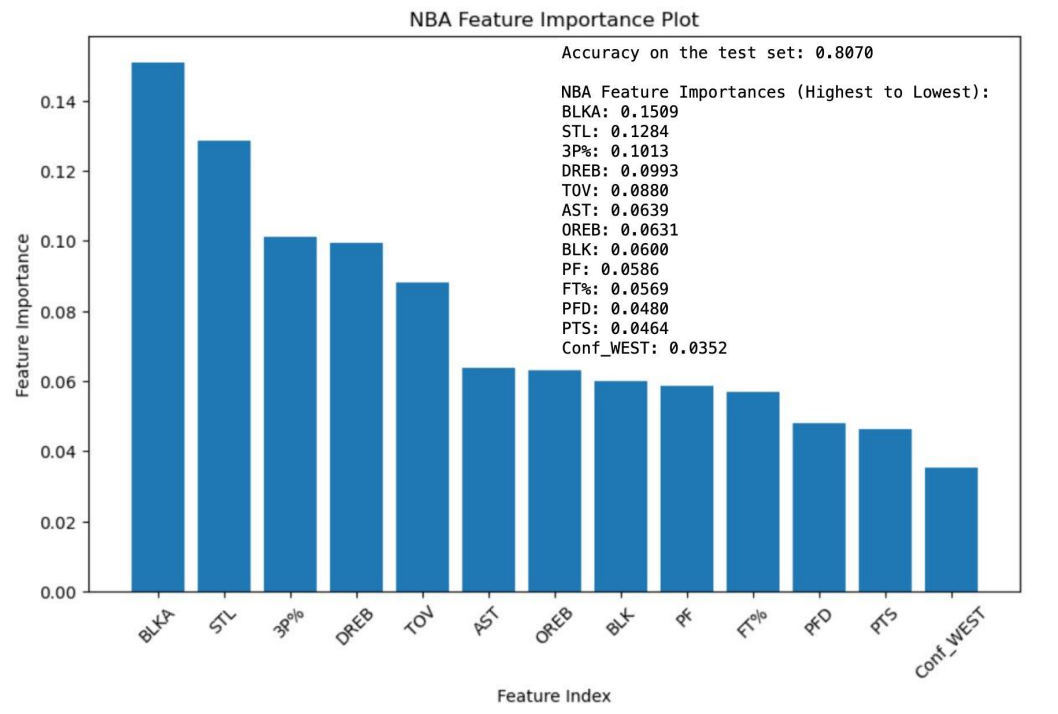
- (dy/dx)/(mean(MadePlayoffs))
- -.136/.537 = -.253 x 100 = 25.3%
- A one unit increase in Turnovers in leads to a 25.3% less of a chance that the NBA team makes the playoffs
  - More turnovers typically means more points against and less points for
- A one unit increase in Defensive Rebounds per game leads to a 21.3% more likely of a chance for the NBA team to make the playoffs
  - More defensive rebounds means less offensive rebounds for the other teams so less points against and more opportunities for the team that grabbed the rebound to score

XGBoost Classifier Model – NCAA



ADJOE and ADJDE were the most important for prediction, followed by Big 10 and Big 12 conferences

XGBoost Classifier Model – NBA



Blocks against and Steals were the most important features the model used to decide if the teams would make the playoffs or not

Stakeholder Implications

The stakeholders are the coaches for both the NBA and NCAA teams, the players for both the NCAA and NBA teams, the GM for the NBA teams, and the fans

College athletics generate large amounts of money for their respective universities. Those working in college athletics will be interested to see which factors are associated with making playoffs.

Making it to March Madness generates significant publicity for universities

Ethical

Bias and Fairness: We ensured the model is built without any bias towards certain teams or players as biased models could perpetuate inequalities or stereotypes

Privacy: We need made sure the data that we have collected does not contain any sensitive information about individuals or teams

Accountability: We are transparent on how our model works and that we are responsible for the outcomes

Legal

Data Privacy: We made sure all of the data we are using are compliant with all data protection laws

Intellectual Property: We made sure we do not infringe on any intellectual property rights so we can avoid any and all legal disputes

Discrimination Laws: The model does not discriminate against any certain groups and we do not violate any anti-discrimination laws

Social

Equity in Sports: We can promote equity and access in sports by figuring out and understanding what factors contribute to success in basketball at different levels

Talent Development: Insights from the model could inform players and coaches on how to develop talent among basketball programs and could lead to more effective training

Economic Impact: Understanding the factors that contribute to success in basketball can lead to investments in different basketball programs and new economic development through sports