

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one in front of the green one.

Bracket Buster

Sean Costello

NCAA MARCH MADNESS

01 Virginia

16 UMBC

08 Creighton

09 Kansas St.

05 Kentucky

12 Davidson

04 Arizona

13 Buffalo

06 Miami (FL)

11 Loyola-Chi

03 Tennessee

14 Wright St.

07 Nevada

10 Texas

02 Cincinnati

15 Georgia St.

South
Atlanta, GA



FINALIST



NCAA CHAMPION
Mon April 2



FINALIST



West
Los Angeles, CA

01 Xavier

16 NC Central / Texas So.

08 Missouri

09 Florida St.

05 Ohio St.

12 South Dakota St.

04 Gonzaga

13 UNC Greensboro

06 Houston

11 San Diego St.

03 Michigan

14 Montana

07 Texas A&M

10 Providence

02 North Carolina

15 Lipscomb

East
Boston, MA

Villanova 01

LIU Brooklyn / Radford 16

Virginia Tech 08

Alabama 09

West Virginia 05

Murray St. 12

Wichita St. 04

Marshall 13

Florida 06

St. Bonaventure / UCLA 11

Texas Tech 03

Stephen F. Austin 14

Arkansas 07

Butler 10

Purdue 02

Cal St. Fullerton 15

Kansas 01

Pennsylvania 16

Seton Hall 08

NC State 09

Clemson 05

New Mexico St. 12

Auburn 04

Charleston 13

TCU 06

Arizona St. / Syracuse 11

Michigan St. 03

Bucknell 14

Rhode Island 07

Oklahoma 10

Duke 02

Iona 15

Midwest
Omaha, NE





Bracket Buster

- Machine Learning model using Gradient Boosted classification and K-Means clustering to predict March Madness winners.

Skip the research and get in the game!

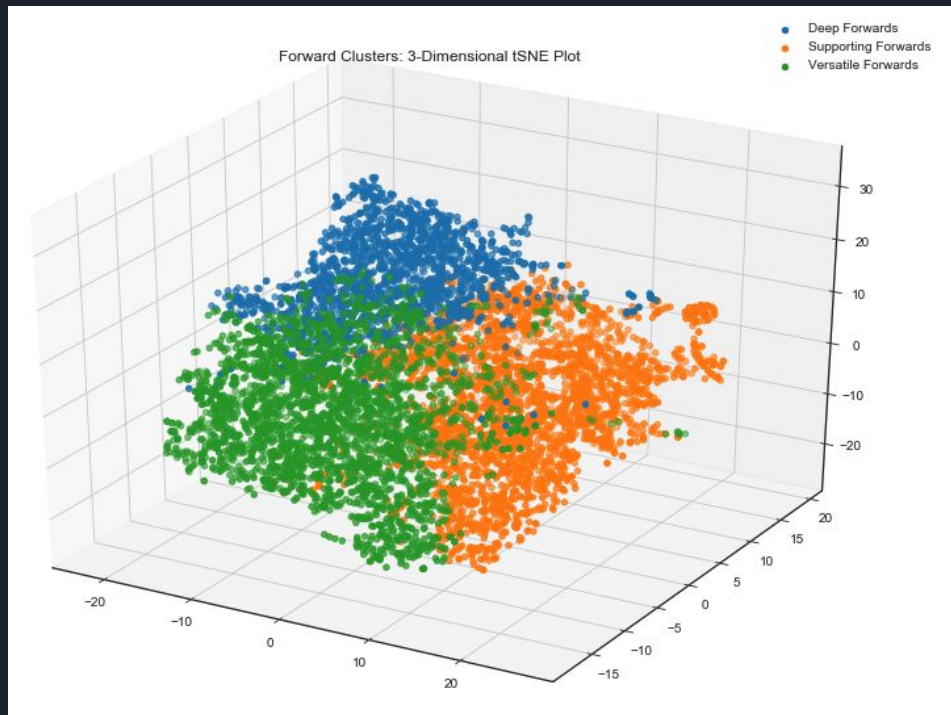
Data and Feature Engineering

Team game logs and player stats
over the past five seasons from:
www.sports-reference.com

Features	Calculated
Win percentage	Cumulative
Strength of Schedule	Season Average
Team Experience	Season Average normalized by minutes played
Team Composition	Season Average normalized by minutes played
Points	5-game rolling average
Points Allowed	5-game rolling average
Field Goal %	5-game rolling average
3-Point %	5-game rolling average
Free Throw %	5-game rolling average
Assists	5-game rolling average
Blocks	5-game rolling average
Steals	5-game rolling average
Rebounds	5-game rolling average
Offensive Rebounds	5-game rolling average
Turnover	5-game rolling average
Personal Fouls	5-game rolling average

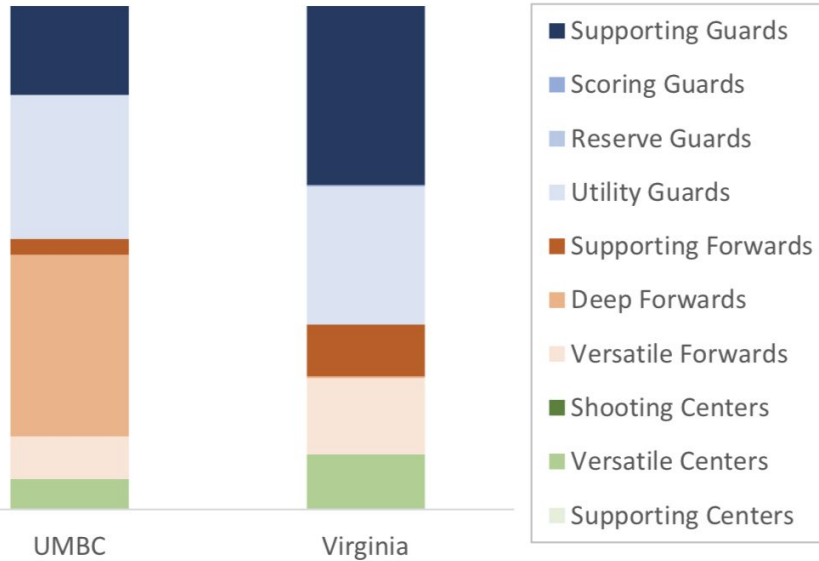
Player Archetypes with K-Means Clustering

Position	Archetype
Centers	Supporting Centers
	Versatile Centers
	Shooting Centers
Forwards	Versatile Forwards
	Deep Forwards
	Supporting Forwards
Guards	Utility Guards
	Reserve Guards
	Scoring Guards
	Supporting Guards



Team Composition Features

UMBC Over Virginia

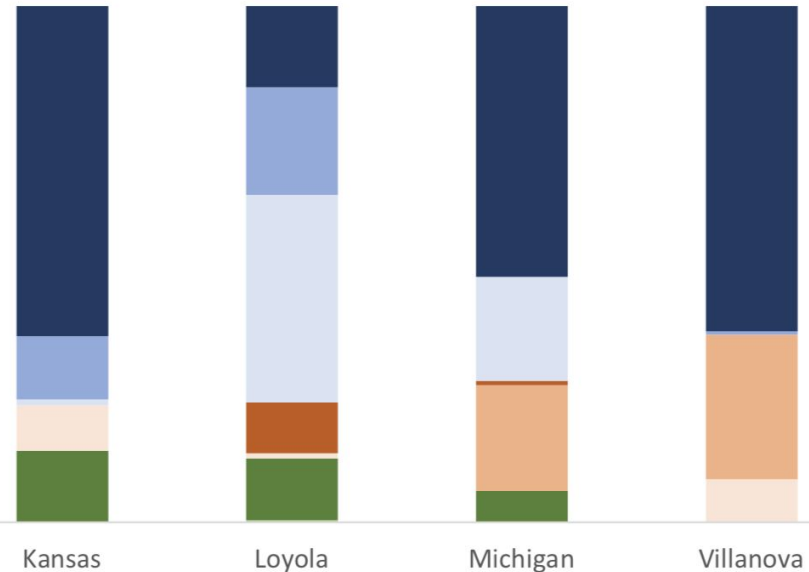


UMBC

Virginia



Final Four



Kansas

Loyola

Michigan

Villanova



Model & Feature Importances

Model: Two Gradient Boosted Classifier Models

Training Data: 2014 - 2017 season games

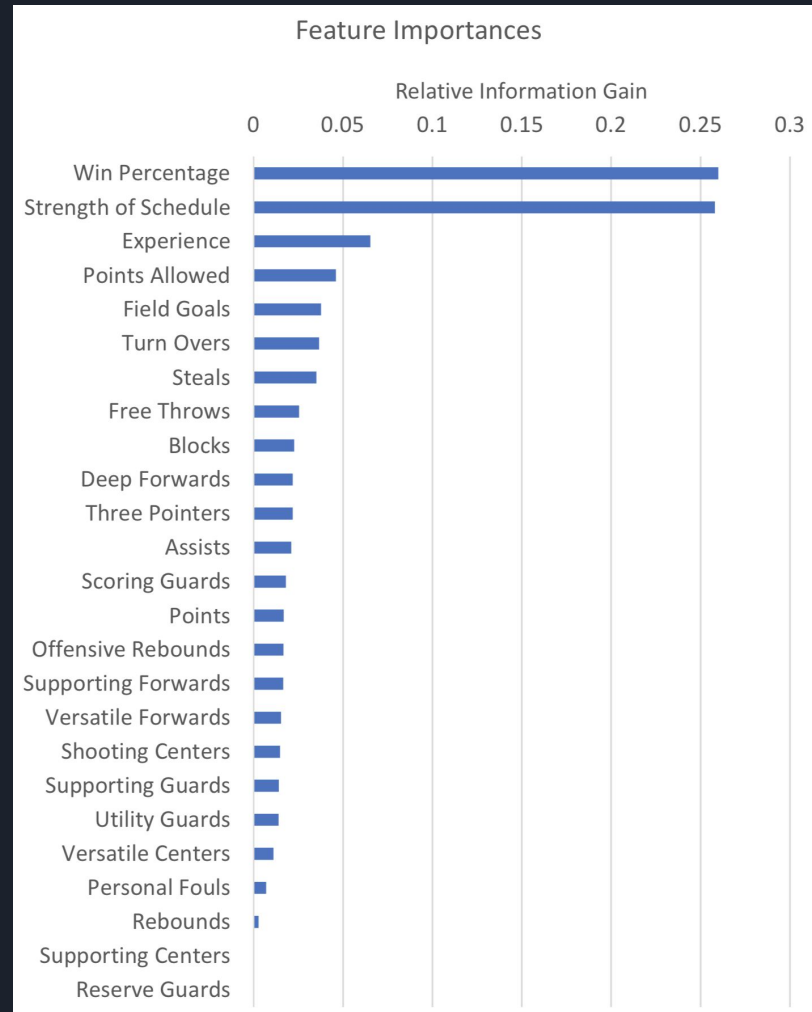
- 4,471 games

Testing Data: 2018 season games

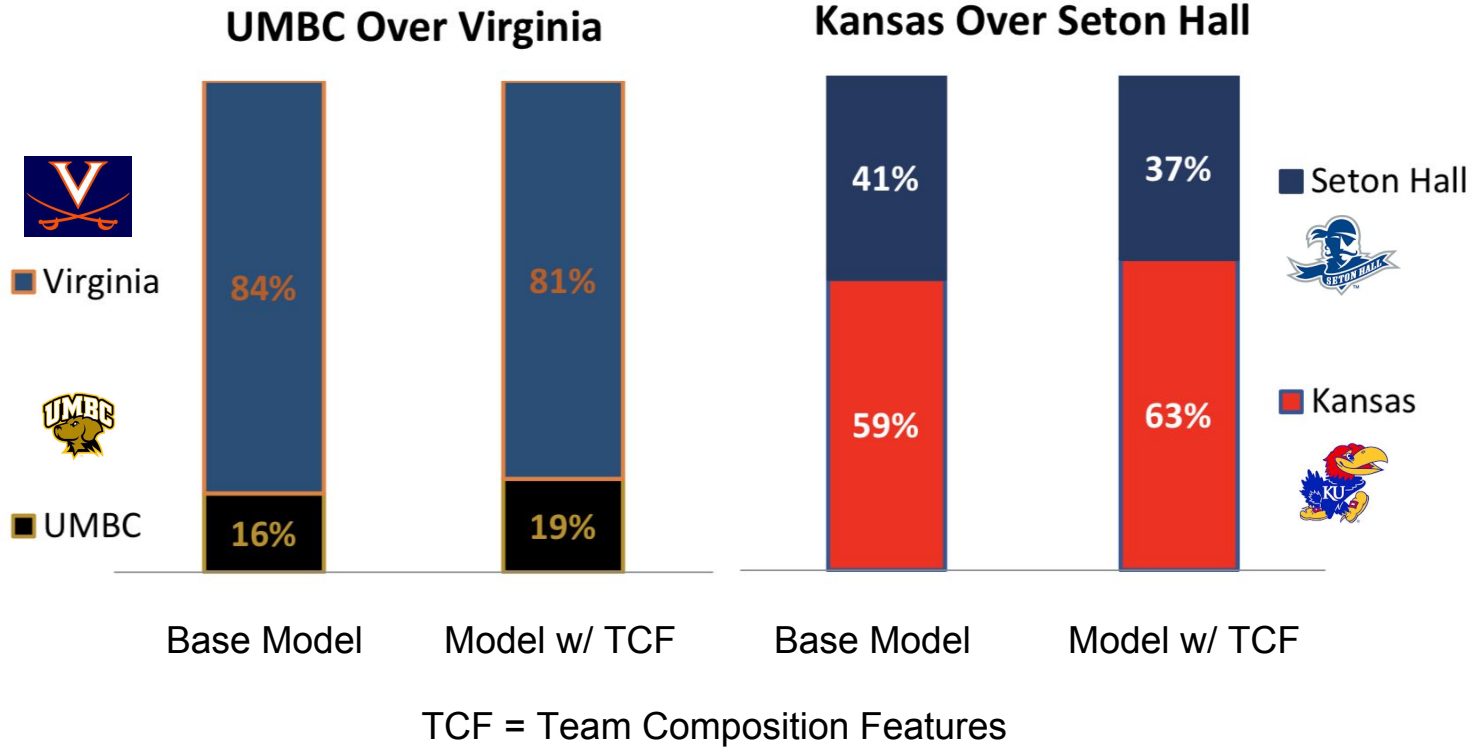
- 1,101 games

Test Accuracy:











- 66% (base model)
- 67% (model with team composition features)



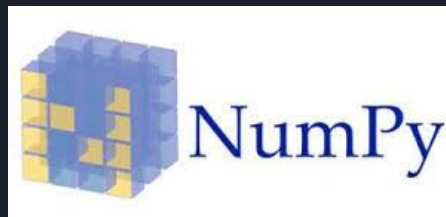
Prediction Examples



2018 March Madness Bracket Results

Rank	Bracket Owner	Total Score	Winner	Logo
1	ESPN Bracket Challenge Winner	167	Villanova	
2	Model w/o Team Comp Features	120	Villanova	
3	Taryn	112	Villanova	
4	Reid Forgrave (CBS)	99	Villanova	
5	Chris Feller	82	Michigan State	
6	Seed/Ranking	81	Virginia	
7	Model w/ Team Comp Features	81	Michigan	
8	ESPN People's Bracket	74	Virginia	
9	Charles Barkley	60	Michigan State	
10	Obama	56	Michigan State	

Tech Stack



Contact

Name:
Sean Costello

Email:
scostello28@gmail.com

Github:
<https://github.com/scostello28>

Repo:
<https://github.com/scostello28/Bracket-Buster>

Linkedin:
www.linkedin.com/in/seanecostello



Appendix

- Model brackets
- Model performance over rounds
- Future Updates
- Hypothetical matchup
- Larger Comparison results



Future Updates

- Test on previous tournaments
- Test with per 100 possession data instead of per game data
- Explore clustering
 - Alternate group numbers (k's)
 - Fuzzy Clustering (GMM)
- Deep Learning
 - MLP
- Auto Bracket generation
 - Produce each team's probability of winning
- Predict Point Spread (change to regression)
- Betting algorithm
- Web app
- Auto Update model with Airflow

Hypothetical Matchup



?



Hypothetical Matchup



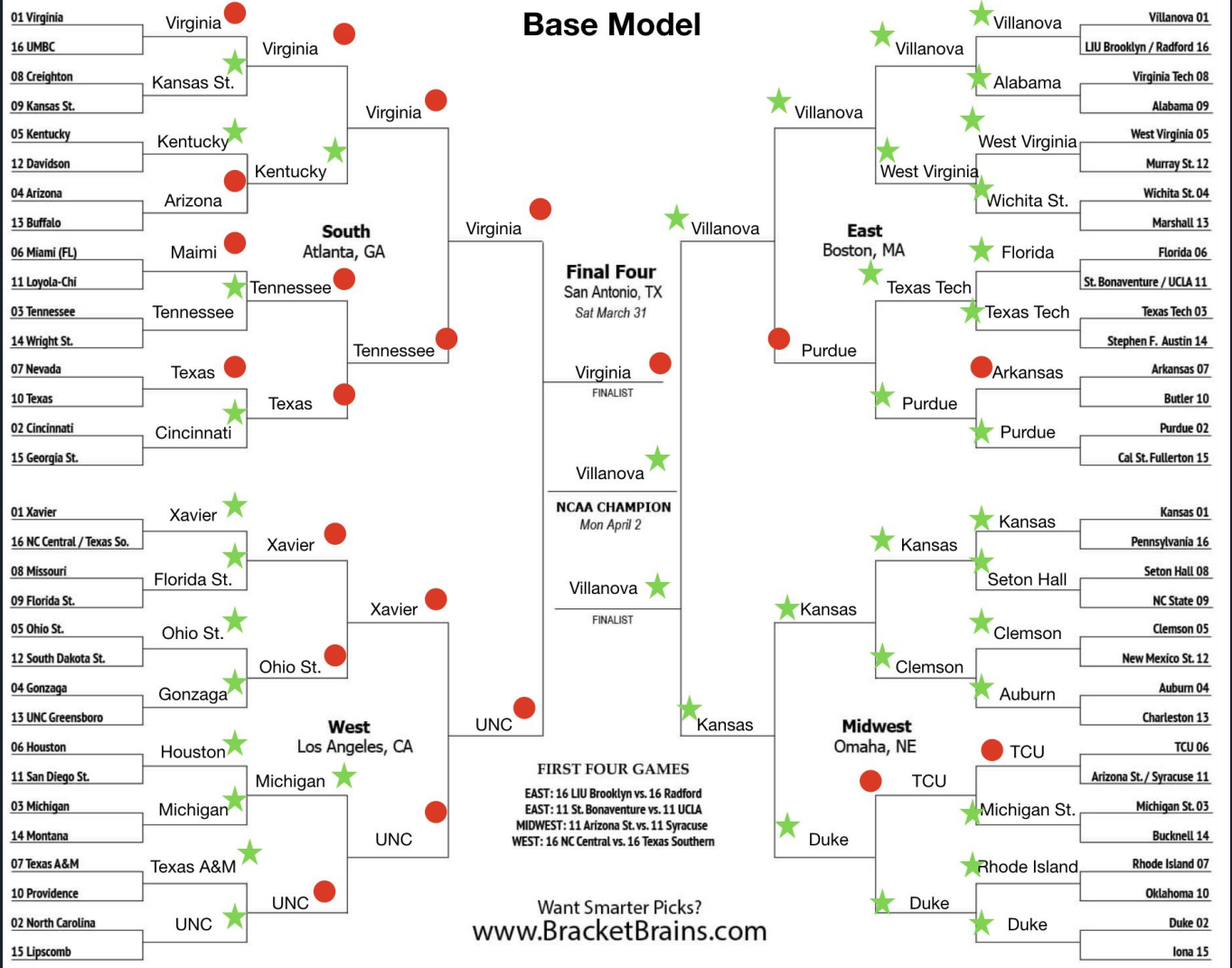
45%



55%



Base Model



Want Smarter Picks?
www.BracketBrains.com

