



D.

MARCH DATA CRUNCH MADNESS



PRESENTED BY THE

NAÏVE BAYESKETBALLERS

***RYAN WAFFLE
SRIM SRJAD***

***KONSTANTIN WEIS
NADEEM CHOUDHURY***



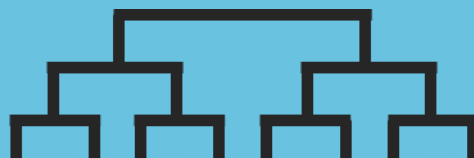


INTRODUCTION

MARCH MADNESS **ANNUAL COLLEGE BASKETBALL TOURNAMENT**



64 TEAMS



**9.2 QUINTILLION
POSSIBLE BRACKETS**



63 GAMES



CHAMPION



PROBLEM STATEMENT

Apply analytical tools to predict the outcome of games

1. Which variables have an impact in models predicting the 2020 tournament?
2. Which teams would have been most likely to upset a higher seed in the 2020 tournament?
3. What would our predictions for 2019 have looked like?



VARIABLE SELECTION

Additional variables added to the dataset

- **Distance** from hometown to each host location
- **Altitude** at hometown and host locations
- **Time zone** for hometown and host locations
- **Upset** for games between different seeded teams

Further steps

- Calculate the **difference (Δ)** between the teams for every variable
 - Difference used to focus on matchup between each team
 - Keep individual seeds and time difference, where the values may have more impact
- Feature selection in SPSS



VARIABLE SELECTION

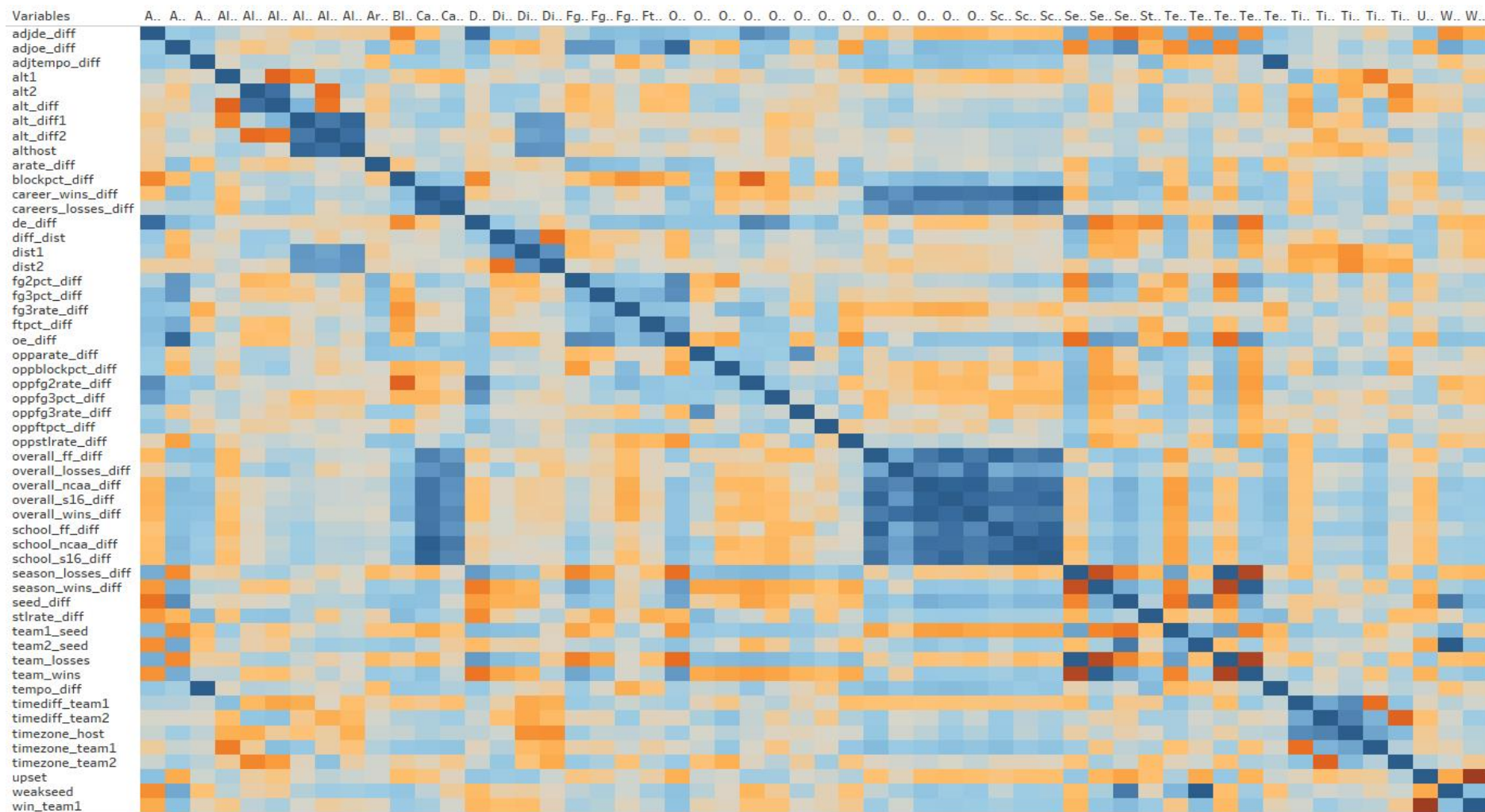


Fig. 1: Correlation Matrix including all the variables



VARIABLE SELECTION

Seed Difference	Coach Overall Wins Difference	School Final Four Appearances Difference
Team 2 Seed	School NCAA Sweet 16 Appearances Difference	3 PT FG % Difference
Adjusted Offensive Efficiency Difference	Opponent 2 PT FG % Difference	Opponent Free Throw % Difference
Adjusted Defensive Efficiency Difference	Time Difference Team 1 to Host	3 PT FG Rate Difference
Team Season Wins Difference	2 PT FG % Difference	Adjusted Tempo Difference
Coach Season Wins Difference	School NCAA Tournament Appearances Difference	Opponent Assist Rate Difference
Coach Season Losses Difference	Steal Rate Difference	Time Difference Team 1 to Host
Team 1 Seed	Free Throw % Difference	Assist Rate Difference
Team Season Losses Difference	Altitude Difference	Opponent Block % Difference
Distance Difference	Opponent Steal Rate Difference	Coach Overall Losses Difference
Block % Difference	Coach Final Four Appearances Difference	Coach Career Losses Difference
Coach NCAA Sweet 16 Appearances Difference	Coach Career Wins Difference	Opponent 3 PT FG Rate
Coach NCAA Tournament Appearances Difference	Opponent 3 PT FG % Difference	

■ used for modeling

Fig. 2: Variables before (38) and after (24) feature selection



QUESTION 1

2020 PREDICTIONS



2020 PREDICTIONS

MODEL PERFORMANCE XG BOOST TREE



Adjusted Offensive Efficiency Difference

Steal Rate Difference

Adjusted Defensive Efficiency Difference

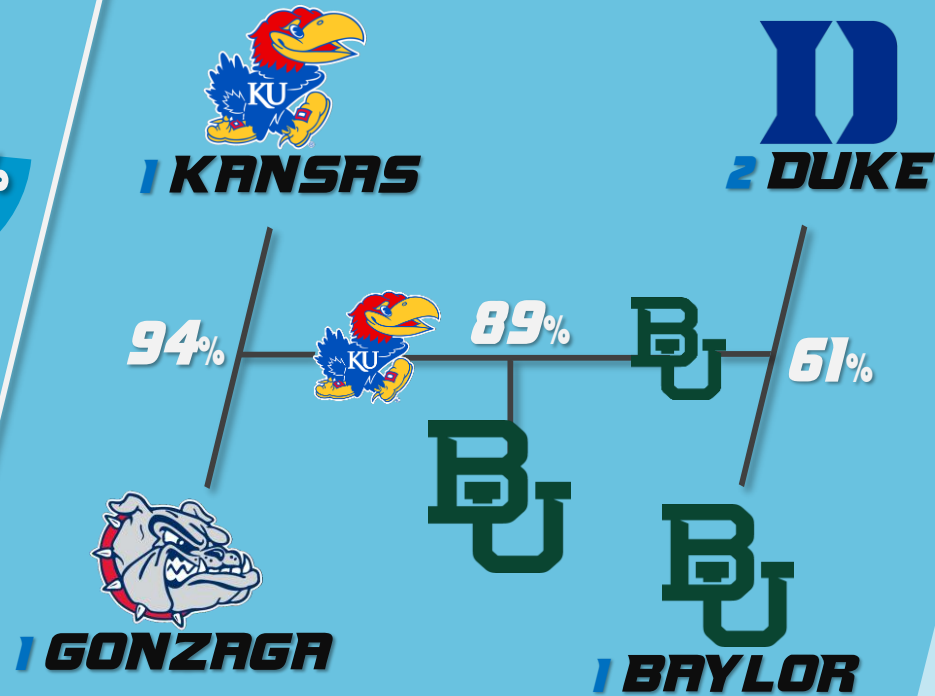
Opponent Steal Rate Difference

Altitude Difference

0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.11 0.12 0.13 0.14 0.15 0.16 0.17 0.18 0.19 0.20

Fig. 3: Top 5 predictors in the model

PREDICTED FINAL FOUR





QUESTION 2

PREDICTING UPSETS IN 2020



PREDICTING UPSETS IN 2020

- Build models based on data from 2002-2019
 - Target Variable: **Upset**
 - Only use games with unequal seed numbers
 - Logistic Regression with 80% Training/20% Testing

**PERFORMANCE ON
HISTORIC DATA**

0.51 **LOGISTIC REGRESSION** **75.8%**

LOG LOSS **ACCURACY**

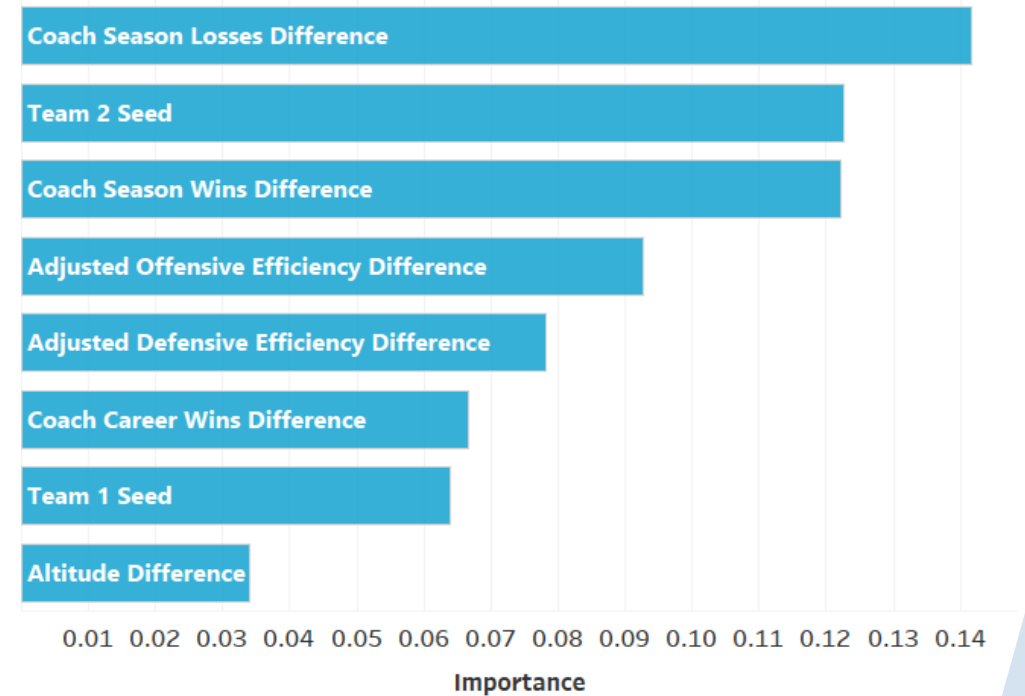


Fig. 4: Top 8 predictors for logistic regression model



PREDICTING UPSETS IN 2020

PREDICTED FIRST ROUND UPSETS

MOST LIKELY UPSET

82.3%
10 HOUSTON
VS
7 RUTGERS

(10) **Houston** vs (7) Rutgers

(11) **BYU** vs (6) Colorado

(11) **Mississippi St** vs (6) Iowa

(10) **Indiana** vs (7) Marquette

(9) **Texas Tech** vs (8) St Marys CA

BIGGEST UPSET

11 BYU
VS
6 COLORADO



0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85

Fig. 5: All five predicted first round upsets sorted by probability

TOP 5 TEAMS TO HAVE A BIG UPSET IN THE TOURNAMENT

AVERAGE PROBABILITY WITH AT LEAST 5 SEED DIFFERENCE

10 HOUSTON
73.2%

11 MISSISSIPPI ST
66.4%

11 WICHITA ST
55.1%

11 PURDUE
50.5%

11 VIRGINIA
46.2%



QUESTION 3

PREDICTIONS FOR 2019

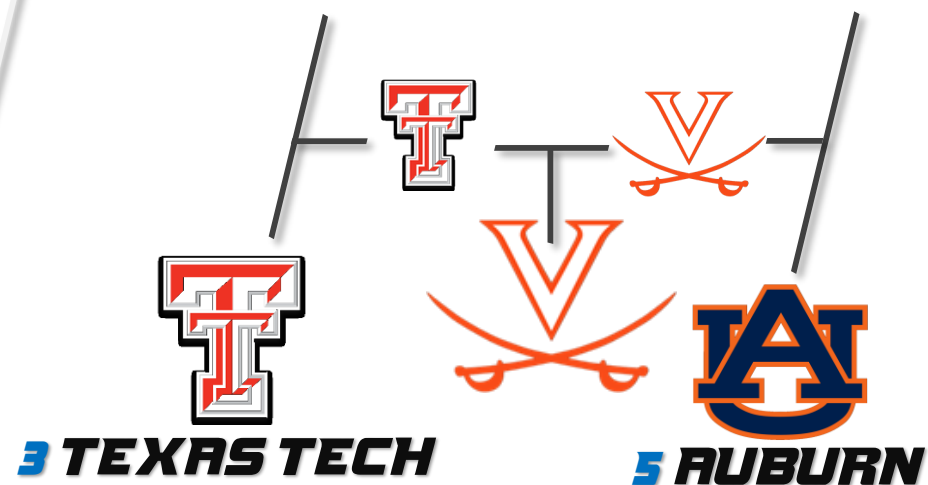


2019 PREDICTIONS

PREDICTED FINAL FOUR



ACTUAL FINAL FOUR



**PREDICTED 1 OUT OF 4 CORRECTLY
PREDICTED CORRECT TOURNAMENT CHAMPION**



RESOURCES

<https://collegebasketball.nbcsports.com/wp-content/uploads/sites/5/2020/02/GettyImages-1139392990-e1581180525220.jpg?w=610&h=343&crop=1>
https://www.gannett-cdn.com/presto/2019/04/07/PDFTF/ab41cf89-eee1-424f-8586-4c03200dce85-UST_040619_Final_Four_M_2.jpg?crop=5075,2855,x0,y0&width=3200&height=1680&fit=bounds
https://upload.wikimedia.org/wikipedia/en/thumb/2/28/March_Madness_logo.svg/1200px-March_Madness_logo.svg.png
<https://cdn.mos.cms.futurecdn.net/XnuD6ptsf8PovnmH5c6WUR.jpg>
<https://cdn.images.express.co.uk/img/dynamic/4/590x/Who-will-be-crowned-this-year-s-national-champion-930854.jpg>
[https://cdn.vox-cdn.com/thumbor/BqTt9GZKTLWEvt9j09wWM04FVO8=/0x0:3000x1687/920x613/filters:focal\(1600x714:2080x1194\):format\(webp\)/cdn.vox-cdn.com/uploads/chorus_image/image/66477024/932020360.jpg.0.jpg](https://cdn.vox-cdn.com/thumbor/BqTt9GZKTLWEvt9j09wWM04FVO8=/0x0:3000x1687/920x613/filters:focal(1600x714:2080x1194):format(webp)/cdn.vox-cdn.com/uploads/chorus_image/image/66477024/932020360.jpg.0.jpg)
<https://i2.wp.com/globalenglish.com/wp-content/uploads/2017/03/deloitte-logo-d.png?resize=600%2C600&ssl=1>
https://upload.wikimedia.org/wikipedia/en/thumb/6/6f/Gabelli_School_of_Business_seal.svg/800px-Gabelli_School_of_Business_seal.svg.png
https://www.google.com/imgres?imgurl=https%3A%2F%2Fa57.foxnews.com%2Fcf-images.us-east-1.prod.boltdns.net%2Fstatic%2F694940094001%2F0273b072-c7a1-47fa-a9d5-a13d8342cef9%2F2c711b7f-7c17-4fb1-8450-89570fe8d8e4%2F1280x720%2Fmatch%2F931%2F524%2Fimage.jpg%3Fve%3D1%26t%3D1&imgrefurl=https%3A%2F%2Fwww.foxnews.com%2Fsports%2Fncaa-basketball-tournaments-cancel-in-response-coronavirus-threat&tbnid=5r_Yiuq26NY9VM&vet=12ahUKEwiGy8zDxLnoAhV7gHIEHQLTAL4QMyhOegUIARDhAQ.i&docid=8dkeGInJRIYhXM&w=931&h=524&q=ncaa%20basketball&ved=2ahUKEwiGy8zDxLnoAhV7gHIEHQLTAL4QMyhOegUIARDhAQ
<https://media.syracuse.com/orangebasketball/photo/9211894-large.jpg>
https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.gannett-cdn.com%2Fpresto%2F2019%2F02%2F05%2Fpkns%2Fb787b97d-e752-4880-9f14-1d00dc1157d6-tennessee_dunk.jpg&imgrefurl=https%3A%2F%2Fwww.knoxnews.com%2Fstory%2Fsports%2Fcollege%2Funiversity-of-tennessee%2Fmens-basketball%2F2019%2F02%2F05%2Ftennessee-vols-basketball-dunk-pregame-dwyane-wade-ja-morant%2F2766196002%2F&tbnid=KeE6HnQyZGqipM&vet=12ahUKEWjT9O3JxbnoAhWfg3IEHUEKBSIQMygBegUIARDRAQ.i&docid=ohpj46_Ph13oaM&w=4554&h=3036&q=ncaa%20basketball%20pregame&ved=2ahUKEWjT9O3JxbnoAhWfg3IEHUEKBSIQMygBegUIARDRAQ
https://upload.wikimedia.org/wikipedia/commons/thumb/5/5a/Defense.gov_photo_essay_120112-D-VO565-015.jpg/1200px-Defense.gov_photo_essay_120112-D-VO565-015.jpg
https://www.gannett-cdn.com/presto/2020/03/02/USAT/8a9f9688-8afe-4eba-a209-5723ab8ad882-USP_NCAA_Basketball__Southern_California_at_UCLA.JPG
<https://arc-anglerfish-arc2-prod-tronc.s3.amazonaws.com/public/KE7AJITRFRFVJP4EHIOOM4CPKQ.jpg>
<https://i.ytimg.com/vi/1aN4ow8i6uw/maxresdefault.jpg>
<https://bloximages.newyork1.vip.townnews.com/greensboro.com/content/tncms/assets/v3/editorial/7/27/727ba388-ac90-541a-b11b-9df32dc90803/5c93d146c368d.image.jpg?resize=1200%2C821>
<https://usatlonzowire.files.wordpress.com/2019/10/GettyImages-627409468.jpg?w=1000&h=600&crop=1>
https://img.bleacherreport.net/img/images/photos/002/697/853/hi-res-156660651-doug-mcdermott-of-the-creighton-bluejays-reacts-after-a_crop_north.jpg?h=1066&w=1600&q=90
https://www.gannett-cdn.com/presto/2019/04/07/PDFTF/ab41cf89-eee1-424f-8586-4c03200dce85-UST_040619_Final_Four_M_2.jpg?crop=5075,2855,x0,y0&width=3200&height=1680&fit=bounds
https://upload.wikimedia.org/wikipedia/en/thumb/d/d1/Virginia_Cavaliers_sabre.svg/1200px-Virginia_Cavaliers_sabre.svg.png
https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT79lvkXt0pZpg_rr_w2j_nQ1Njy8ZimPkiH09k8gT_0MkW7ha6Q&s
https://upload.wikimedia.org/wikipedia/en/thumb/9/90/Wichita_State_Shockers_logo.svg/1200px-Wichita_State_Shockers_logo.svg.png
https://upload.wikimedia.org/wikipedia/commons/thumb/9/95/BYU_Cougars_logo.svg/1200px-BYU_Cougars_logo.svg.png
https://upload.wikimedia.org/wikipedia/commons/e/ec/University_of_Houston_classic_logo.png
https://upload.wikimedia.org/wikipedia/commons/thumb/3/36/Mississippi_State_Bulldogs_logo.svg/1200px-Mississippi_State_Bulldogs_logo.svg.png
https://upload.wikimedia.org/wikipedia/commons/thumb/e/e1/Duke_Athletics_logo.svg/1200px-Duke_Athletics_logo.svg.png
https://upload.wikimedia.org/wikipedia/en/thumb/f/f4/Kansas_Jayhawks_logo.svg/1200px-Kansas_Jayhawks_logo.svg.png
<https://a.espncdn.com/combiner/i?img=/i/teamlogos/ncaa/500/2250.png>
https://upload.wikimedia.org/wikipedia/commons/thumb/c/c4/Baylor_Athletics_logo.svg/1200px-Baylor_Athletics_logo.svg.png
https://upload.wikimedia.org/wikipedia/commons/thumb/b/b6/Kentucky_Wildcats_logo.svg/2000px-Kentucky_Wildcats_logo.svg.png
https://cdn.bleacherreport.net/images/team_logos/328x328/texas_tech_basketball.png
https://upload.wikimedia.org/wikipedia/commons/thumb/1/15/Auburn_Tigers_logo.svg/1200px-Auburn_Tigers_logo.svg.png
https://upload.wikimedia.org/wikipedia/en/thumb/a/a7/Michigan_State_Athletics_logo.svg/1200px-Michigan_State_Athletics_logo.svg.png