Predicting the Outcomes of March Madness Games

By Aviva Mazurek

Objectives and Methodology

Objective
- Predict individual outcomes of march madness games
    - Final SVM model predicts game outcomes with 87% accuracy

**Methodology**
- Data sources
    - Sports Reference - historical brackets, seeding, advanced stats from regular season
    - ESPN - Stats from regular season
    - Kenpom - Ranking System, wins/losses per team in regular season
- Created multiple dictionaries to merge all data from 2002-2019

Feature engineering creates larger disparity between wins and losses - allows model to learn better
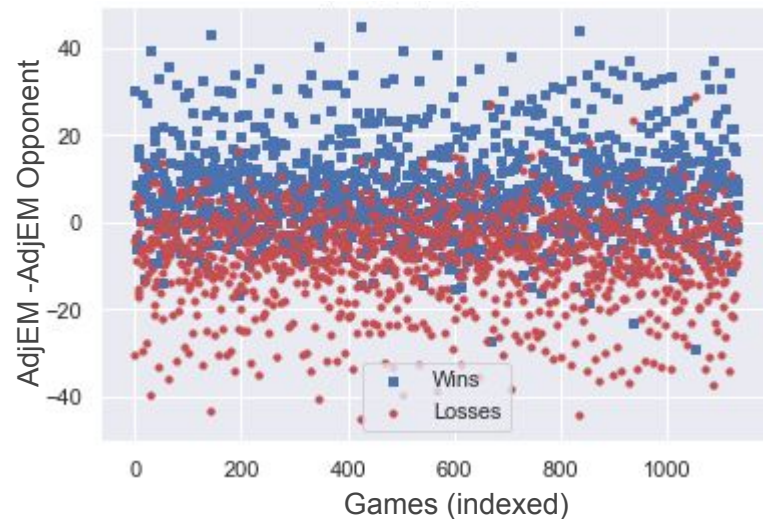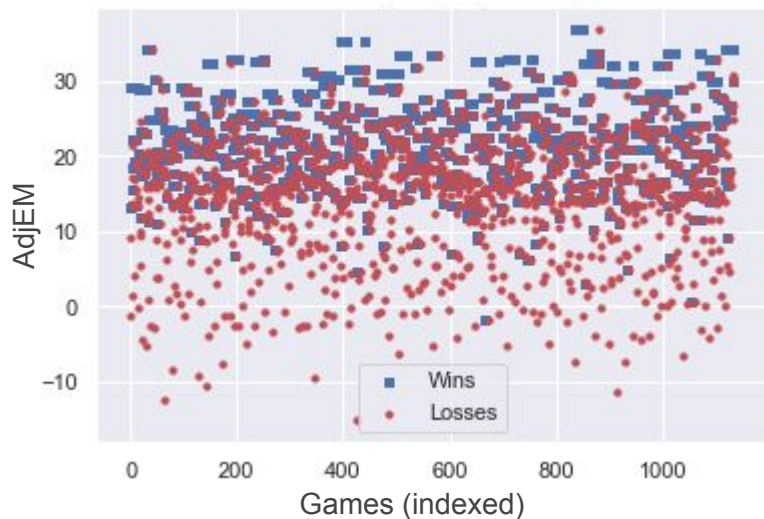
Which features are more significant?
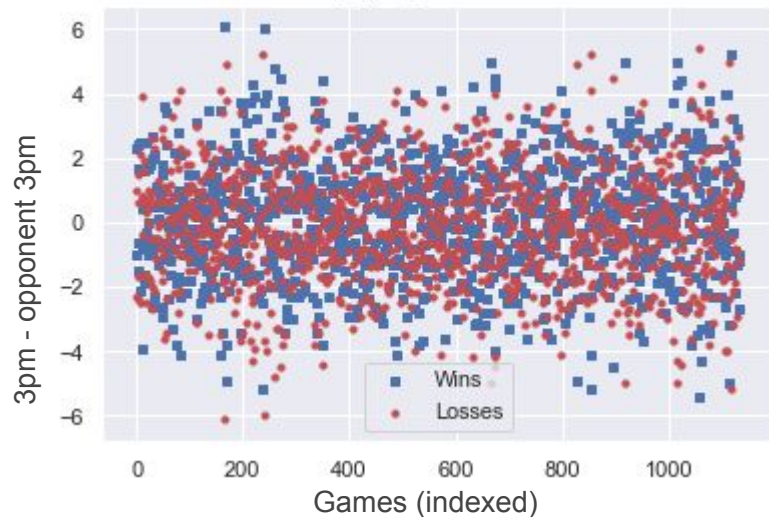
vs.
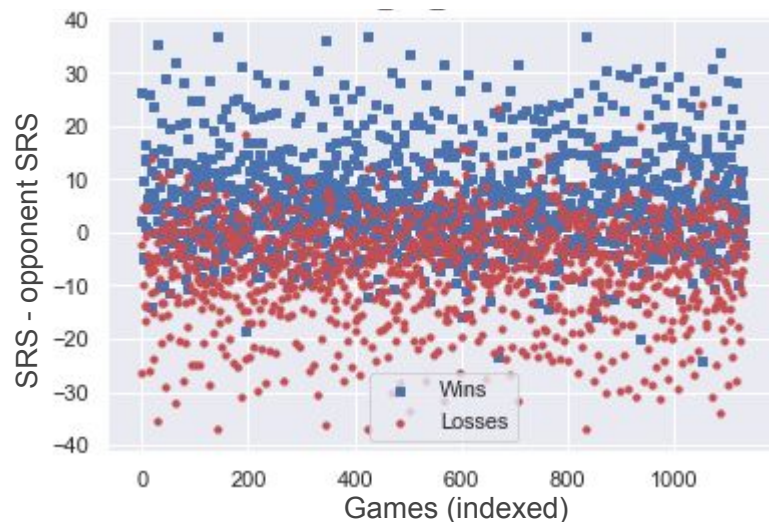
→ Most significant features: Kenpom ratings, wins from regular season, losses from regular season, SOS adv, SRS adv, seeding

Modeling Data - Train and Tested on Data From 2002-2019

Model	Accuracy (%)
Decision Tree	75.62
Gradient Boosted Regression Tree	75.62
XGBoost	85.76
Support Vector Machine (SVM)	86.78

```
           precision   recall   f1-score

SVM:    0     0.85       0.88      0.86
        1     0.89       0.85      0.87
```

Experiment: Used SVM model to predict individual game outcomes for 2019 brackets

- Trained and tested data on all data from 2002-2018 (excluded 2019 data)
- Used the model to predict 2019 march madness results

Results ⟶

Eastern Region

1 Duke — 85					
16 North Dakota State — 62		**1 Duke** — 77			
8 VCU — 58				**1 Duke** — 75	
9 UCF — 73		9 UCF — 76			
					1 Duke — 67
5 Mississippi State — 76					
12 Liberty — 80		12 Liberty — 58		**1 Duke** — ...	
4 Virginia Tech — 66				4 Virginia Tech — 73	
13 Saint Louis — 52		**4 Virginia Tech** — 67			
6 Maryland — 79					**2 Michigan State**
11 Belmont — 77		6 Maryland — 67			
3 LSU — 79				3 LSU — 63	
14 Yale — 74		**3 LSU** — 69			
					2 Michigan State — 68
7 Louisville — 76					
10 Minnesota — 86		10 Minnesota — 50		**2 Michigan State** — 80	
2 Michigan State — 76					
15 Bradley — 65		**2 Michigan State** — 70			

at Columbia, SC
at Columbia, SC
at San Jose, CA
at San Jose, CA
at Jacksonville, FL
at Jacksonville, FL
at Des Moines, IA
at Des Moines, IA
at Columbia, SC
at San Jose, CA
at Jacksonville, FL
at Des Moines, IA
at Washington, DC
at Washington, DC
at Washington, DC

Western Region

1 **Gonzaga** — 87			
at Salt Lake City, UT	1 **Gonzaga** — 83		
16 Fairleigh Dickinson — 49	at Salt Lake City, UT	1 **Gonzaga** — 72	
8 Syracuse — 69	9 Baylor — 71		1 Gonzaga — 69
at Salt Lake City, UT		at Anaheim, CA	
9 **Baylor** — 78		4 Florida State — 58	
5 Marquette — 64	12 Murray State — 62		
at Hartford, CT	at Hartford, CT		
12 **Murray State** — 83	4 **Florida State** — 90		
4 **Florida State** — 76			3 Texas Tech
at Hartford, CT			at Anaheim, CA
13 Vermont — 69			
6 **Buffalo** — 91	6 Buffalo — 58		
at Tulsa, OK	Tulsa, OK	3 **Texas Tech** — 63	
11 Arizona State — 74	3 **Texas Tech** — 78		3 **Texas Tech** — 75
3 **Texas Tech** — 72		at Anaheim, CA	
at Tulsa, OK		2 Michigan — 44	
14 Northern Kentucky — 57			
7 Nevada — 61	10 Florida — 49		
at Des Moines, IA	at Des Moines, IA		
10 **Florida** — 70	2 Michigan — 44		
2 **Michigan** — 74	2 **Michigan** — 64		
at Des Moines, IA			
15 Montana — 55			

Midwest Region

Round 1	Round 2	Round 3	Round 4	Champion
1 UNC 88				
16 Iona 73	**1 UNC** 81			
8 Utah State 61		**1 UNC** 80		
9 Washington 78	9 Washington 59			
5 Auburn 78			**5 Auburn** 77	
12 New Mexico State 77	**5 Auburn** 89			
4 Kansas 87		**5 Auburn** 97		
13 Northeastern 53	4 Kansas 75			
6 Iowa State 59				**5 Auburn**
11 Ohio State 62	11 Ohio State 59			
3 Houston 84		3 Houston 58		
14 Georgia State 55	**3 Houston** 74			
7 Wofford 84			2 Kentucky 71	
10 Seton Hall 68	7 Wofford 56			
2 Kentucky 79		**2 Kentucky** 62		
15 Abilene Christian 44	**2 Kentucky** 62			

Game locations:
- at Columbus, OH
- at Salt Lake City, UT
- at Tulsa, OK
- at Jacksonville, FL
- at Kansas City, MO

Southern Region

1 **Virginia** — **71**
at Columbia, SC
1 **Virginia** — **63**

16 Gardner-Webb — **56**

8 Ole Miss — **72**
at Columbia, SC
9 Oklahoma — **51**

9 **Oklahoma** — **95**

1 **Virginia** — **53**

at Columbia, SC

1 **Virginia** — **80**

5 Wisconsin — **54**
at San Jose, CA
12 **Oregon** — **73**

12 **Oregon** — **72**

12 Oregon — **49**

4 Kansas State — **64**
at San Jose, CA
13 UC-Irvine — **54**

13 **UC-Irvine** — **70**

6 **Villanova** — **61**
at Hartford, CT
6 Villanova — **61**

11 Saint Mary's — **57**

3 **Purdue** — **99**

3 **Purdue** — **61**
at Hartford, CT
3 **Purdue** — **87**

14 Old Dominion — **48**

at Louisville, KY

3 Purdue — **75**

7 Cincinnati — **72**
at Columbus, OH
10 Iowa — **77**

10 **Iowa** — **79**

2 Tennessee — **94**

2 **Tennessee** — **77**
at Columbus, OH
2 **Tennessee** — **83**

15 Colgate — **70**

at Louisville, KY

1 Virginia

at Louisville, KY

1 Virginia

Final 4

1 Virginia		**63**
5 Auburn		62

at Minneapolis, MN ✓

3 Texas Tech		**61**
2 Michigan State		51

at Minneapolis, MN ✗

1 Virginia		**85**
3 Texas Tech		77

at Minneapolis, MN ✓

1 Virginia

Conclusions

- Can predict individual march madness games with 87% accuracy
  - Feature engineering and inclusion of Kenpom stats increased accuracy tremendously

Future Work

- T-Pot pipeline determined logistic regression coupled with SGDC Classifier produces 90% accuracy
    - Model difficult to interpret - requires investigation
- Incorporate point outcomes of games
- Conduct same experiment as for 2019, but for every year and determine where the model fails the most
- Incorporate more features/engineering

Acknowledgements

- Sean Abu
- Eric Rosner
- Matthew Wasserman
- Yish Lim
- FangFang Lee
- Ben Johnson-Laird
- Andrew Triola
- Julia Chong
- Ed Haracz
- Aliza Layman