

Predicting the Outcomes of March Madness Games

By Aviva Mazurek

Objectives and Methodology

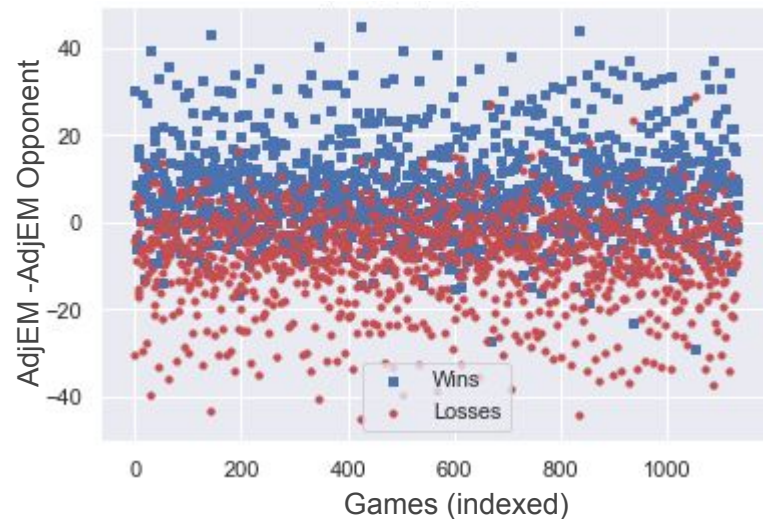
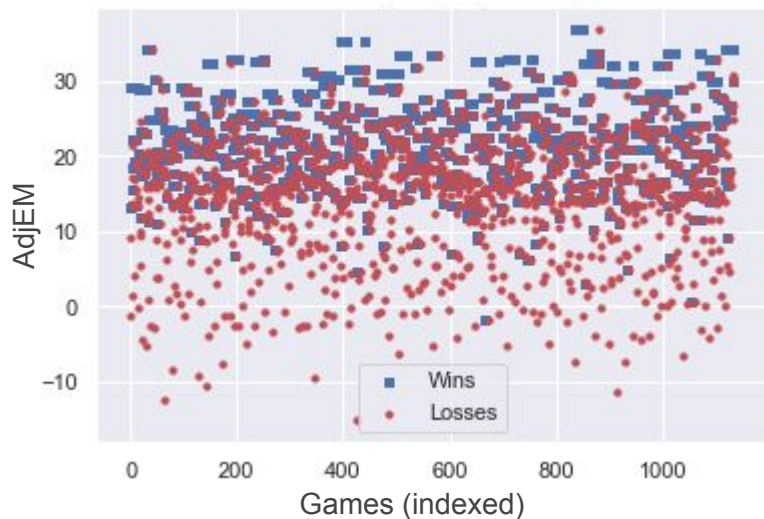
Objective

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

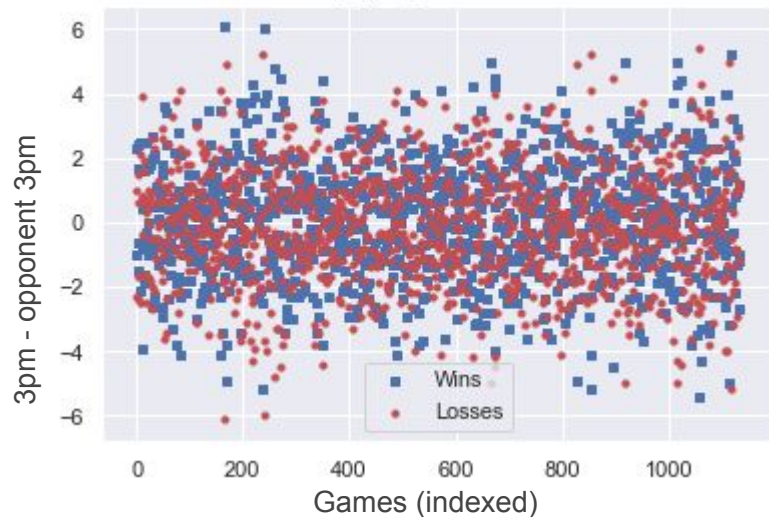
Methodology

- 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

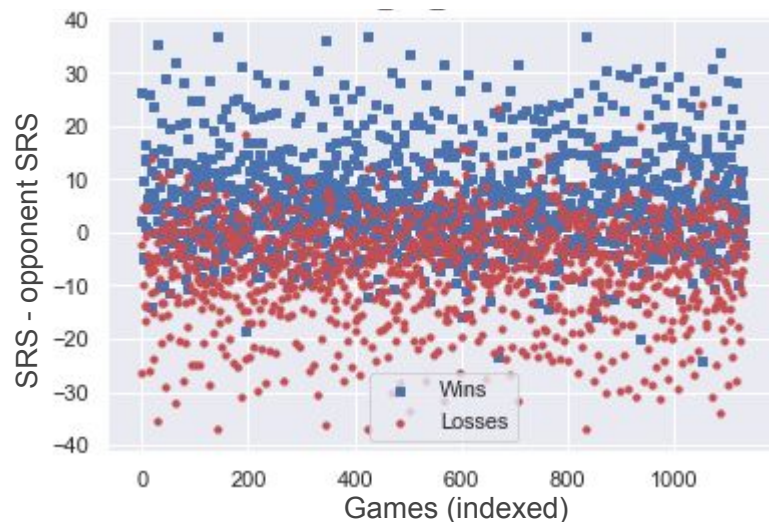
Feature engineering creates larger disparity between wins and losses - allows model to learn better



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 - Trained and Tested on Data From 2002-2019

Model	Accuracy (%)
Decision Tree	75.62
KNN	69.68
XGBoost	85.14
Logistic Regression	84.43
Support Vector Machine (SVM)	86..84

**The accuracy is an average of 10 random train test split accuracies (except for the decision tree)

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

Rank	Team	Score	Game	Score
1	Duke	85		
16	North Dakota State	62	at Columbia, SC	1 Duke 77
8	VCU	58		
9	UCF	73	at Columbia, SC	1 Duke 75
5	Mississippi State	76		
12	Liberty	80	at San Jose, CA	1 Duke 67
4	Virginia Tech	66		
13	Saint Louis	52	at San Jose, CA	4 Virginia Tech 73
6	Maryland	79		
11	Belmont	77	at Jacksonville, FL	1 Duke 68
3	LSU	79		
14	Yale	74	at Jacksonville, FL	3 LSU 63
7	Louisville	76		
10	Minnesota	86	at Des Moines, IA	2 Michigan State 80
2	Michigan State	76		
15	Bradley	65	at Des Moines, IA	2 Michigan State 70

Western Region

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

Midwest Region

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

Southern Region

1 Virginia 71	at Columbia, SC	1 Virginia 63		1 Virginia 53		
16 Gardner-Webb 56			at Columbia, SC			
8 Ole Miss 72		9 Oklahoma 51				
9 Oklahoma 95	at Columbia, SC					
5 Wisconsin 54		12 Oregon 73				1 Virginia 80
12 Oregon 72	at San Jose, CA			at Louisville, KY		
4 Kansas State 64		13 UC-Irvine 54	at San Jose, CA	12 Oregon 49		
13 UC-Irvine 70	at San Jose, CA					
6 Villanova 61		6 Villanova 61				
11 Saint Mary's 57	at Hartford, CT					
3 Purdue 61		3 Purdue 87	at Hartford, CT	3 Purdue 99		
14 Old Dominion 48	at Hartford, CT					
7 Cincinnati 72		10 Iowa 77				
10 Iowa 79	at Columbus, OH			at Louisville, KY	3 Purdue 75	1 Virginia
2 Tennessee 77		2 Tennessee 83	at Columbus, OH	2 Tennessee 94		
15 Colgate 70	at Columbus, OH					

Final 4

1 [Virginia](#) [63](#)

5 [Auburn](#) [62](#)

3 [Texas Tech](#) [61](#)

2 [Michigan State](#) [51](#)



at Minneapolis, MN

at Minneapolis, MN



1 [Virginia](#) [85](#)



at Minneapolis, MN

3 [Texas Tech](#) [77](#)

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