AI Academy
Capstone Presentation

Jessica Pasquesi

MAKING AN
IMPACT THAT
MATTERS
since 1845

Agenda

1 | Overview and Business Problem

2 | Data Understanding & Analysis

3 | Model & Evaluation

4 | Conclusion

Business Problem

Using historical college basketball data, can we better predict the winner of games in the NCAA March Madness tournament? This would be useful given the large unpredictability in creating college brackets.

Data Understanding and Data Preparation

NCAATourneySeeds
- Season*
- Seed
- TeamID

* Season starting at 1985

NCAATourneyCompactResults
- Season*
- DayNum
- WTeamID
- LTeamID
- WScore
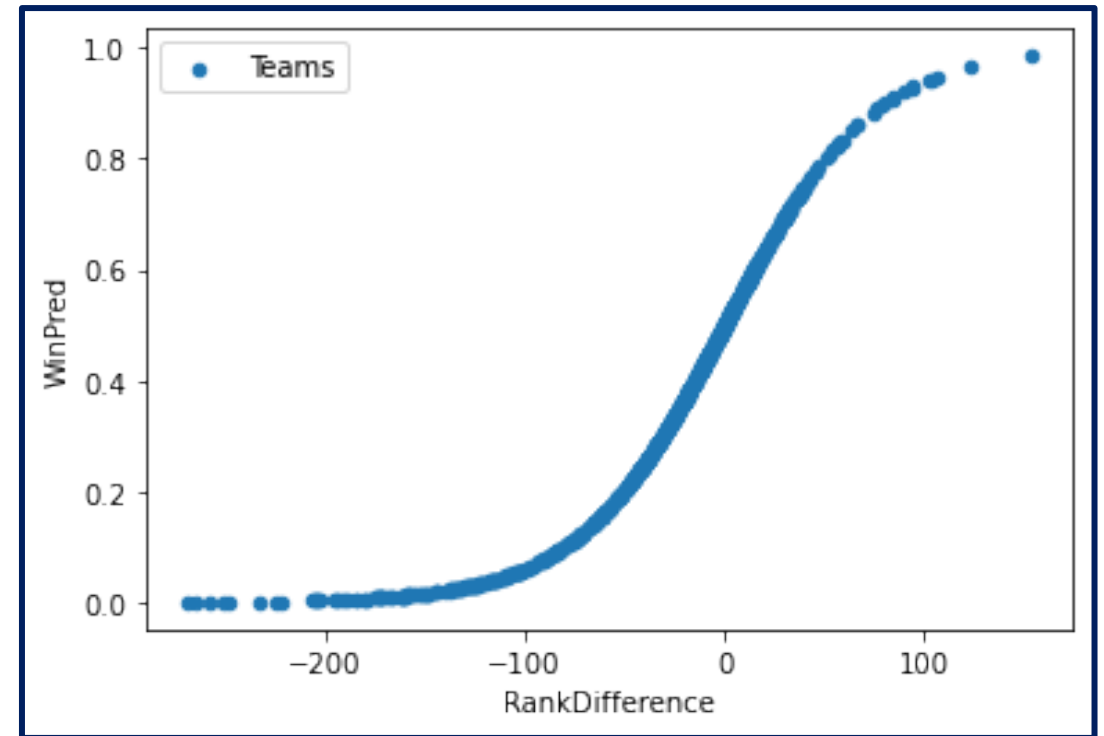- Lscore
- WLoc
- NumOT

* Season starting at 1985

MasseyOrdinals
- Season*
- RankingDayNum
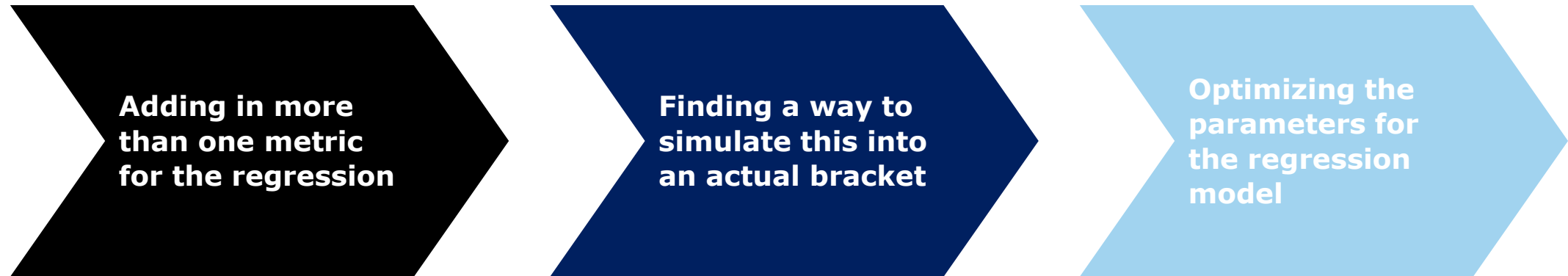- SystemName
- TeamID
- OrdinalRank

* Season starting at 2003

Modeling and Evaluation

	Season	WTeamID	LTeamID	WAvgRank	LAvgRank	RankDifference	WinPred	LossPred
0	2003	1421	1411	240.343750	239.281250	1.062500	0.507270	0.492730
2	2003	1113	1272	36.000000	21.705882	14.294118	0.596588	0.403412
3	2003	1141	1166	45.687500	20.735294	24.952206	0.664410	0.335590
4	2003	1143	1301	36.406250	50.312500	-13.906250	0.405970	0.594030
5	2003	1163	1140	27.500000	28.593750	-1.093750	0.492516	0.507484
...	...	...	...	...	...	...	...	...
975	2017	1376	1124	38.097222	14.346667	23.750556	0.657037	0.342963
977	2017	1332	1242	11.946667	6.040000	5.906667	0.540333	0.459667
978	2017	1314	1246	5.426667	4.813333	0.613333	0.504197	0.495803
981	2017	1314	1332	5.426667	11.946667	-6.520000	0.455501	0.544499
982	2017	1314	1211	5.426667	2.013333	3.413333	0.523341	0.476659

Next Steps

Adding in more than one metric for the regression

Finding a way to simulate this into an actual bracket

Optimizing the parameters for the regression model