

March Machine Learning Mania 2024

50.000\$ en premios



kaggle

Jorge Martinez, 100508957



¿Qué es el March Madness?

- NCAA
- Estados Unidos
- Torneo baloncesto
- División I
- 64 equipos
- Eliminación directa

FIRST ROUND
MARCH 16-17

S	16 Texas A&M-CC (23-10)	75	MW	11 Mississippi St. (21-12)	59	DAYTON MARCH 14-15	61 Texas Southern (14-20)	16	E	98 Arizona St. (22-12)	11	W
	16 SE Missouri St. (19-16)	71		11 Pittsburgh (22-11)	60		84 FDU (19-15)	16		73 Nevada (22-10)	11	





Objetivo

- Predecir correctamente todos los partidos del March Madness 2024



Datos

- Kaggle
- Estadísticas de los partidos de temporada regular
- Posición tras la temporada regular
- Estadísticas de los partidos de March Madness



Preprocesamiento de los datos

- Duplicación de estadísticas de temporada regular para equipos 1 y 2
- Agregación de Estadísticas de Temporada Regular por Equipo y Temporada
- Fusión de Estadísticas de Temporada con Datos del Torneo
- Clasificación al final de la temporada
- Calidad del equipo

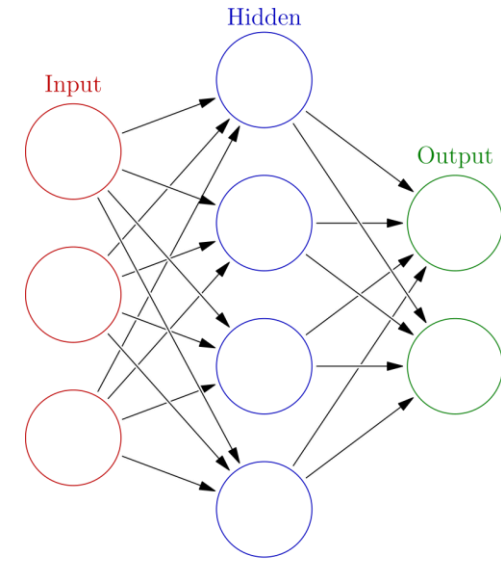
Preprocesamiento de los datos

Season	DayNum	T1_TeamID	T1_Score	T2_TeamID	T2_Score	T1_FGM	...
2003	134	1421	92	1411	84	24.38	...
2003	134	1411	84	1421	92	24.73	...
2003	136	1436	51	1112	80	24.83	...

- 2630 filas
- 63 columnas

T2_opponent_Stl	T2_opponent_Blkl	T2_opponent_PFl	T2_PointDiff	T1_seed	T2_seed	Seed_diff
8.00	2.60	21.63	1.97	16	16	0
8.83	4.24	18.69	-7.24	16	16	0
5.96	2.39	22.07	14.96	16	1	15

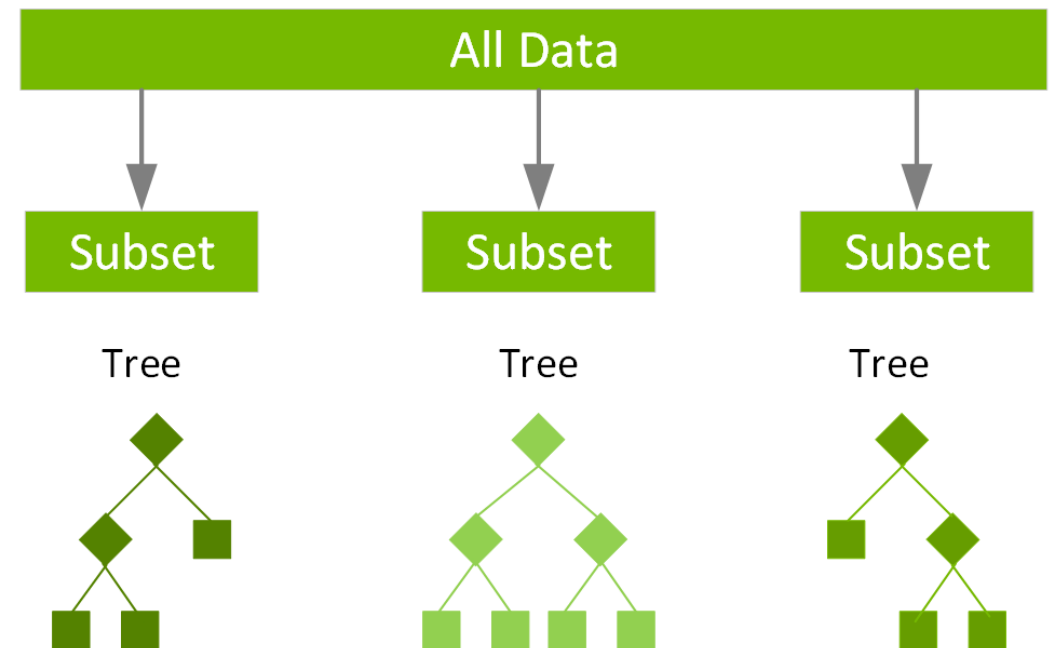
Modelos



dmlc
XGBoost

¿Qué es XGBoost?

- eXtreme Gradient Boosting
- Similar a Random Forest
- Ensamblado de árboles de decisión
- Aprendizaje automático supervisado
- Regresión
- Ganador de 2023



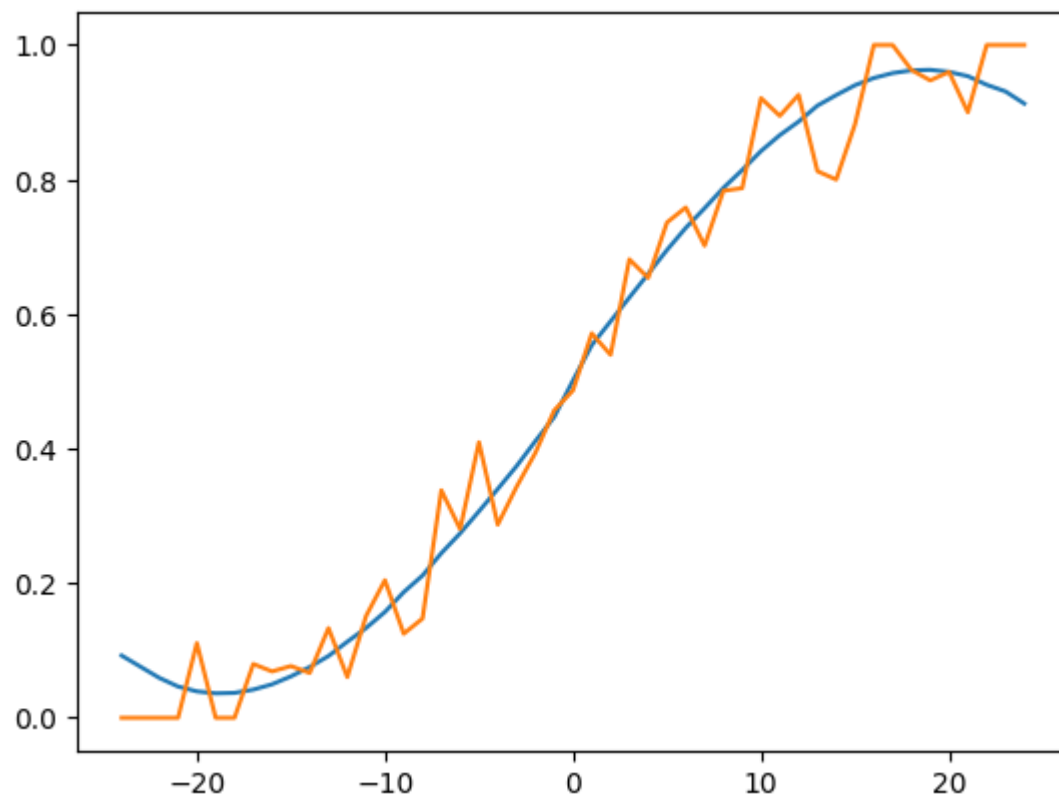
Features y target

	T1_FGM	T1_FGA	T1_FGM3	T1_FGA3	T1_FTM	T1_FTA	T1_OR	...
0	24.379310	56.793103	6.482759	18.000000	15.965517	20.931034	12.275862	...
1	24.733333	55.266667	5.933333	18.500000	17.400000	28.066667	13.166667	...
2	24.827586	55.862069	5.275862	15.482759	12.862069	19.551724	12.965517	...
3	30.321429	65.714286	7.035714	20.071429	17.535714	25.000000	15.178571	...

Index	Target
0	8
1	-8
2	-29
3	29

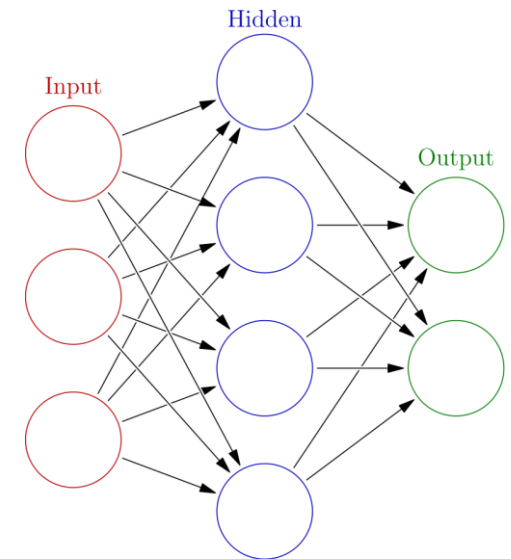
Resultados

Log loss: 0.53



¿Qué son las redes neuronales?

- Modelo computacional basado en el cerebro humano
- Estructurada en capas
- Ajusta pesos en conexiones para minimizar errores
- Aplicada en reconocimiento de imágenes, NLP, etc.
- Aprendizaje supervisado



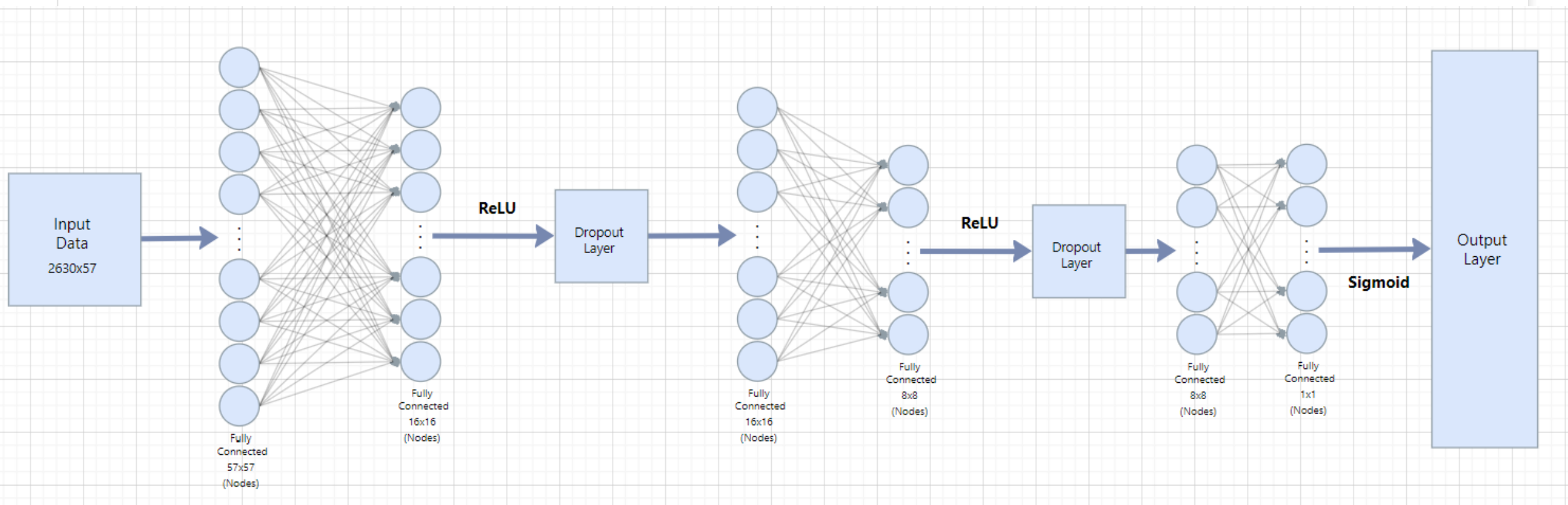
Features y target

	T1_FGM	T1_FGA	T1_FGM3	T1_FGA3	T1_FTM	T1_FTA	T1_OR	...
0	24.379310	56.793103	6.482759	18.000000	15.965517	20.931034	12.275862	...
1	24.733333	55.266667	5.933333	18.500000	17.400000	28.066667	13.166667	...
2	24.827586	55.862069	5.275862	15.482759	12.862069	19.551724	12.965517	...
3	30.321429	65.714286	7.035714	20.071429	17.535714	25.000000	15.178571	...

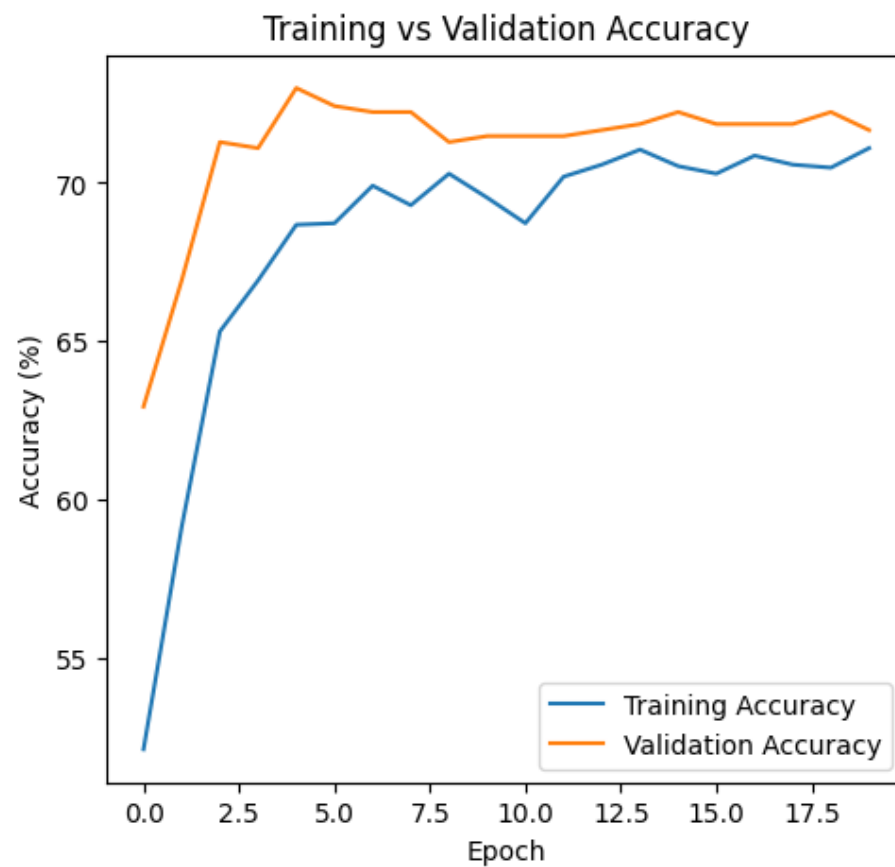
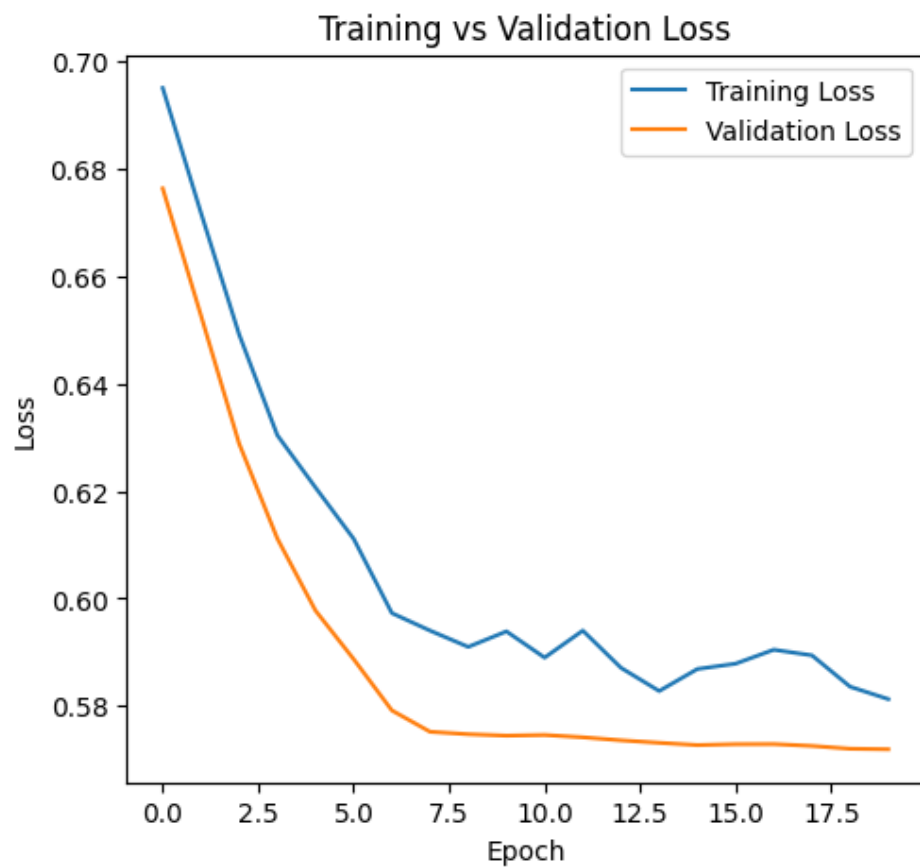
Index	Target
0	1
1	0
2	0
3	1

Standard Scaler

Arquitectura



Resultados



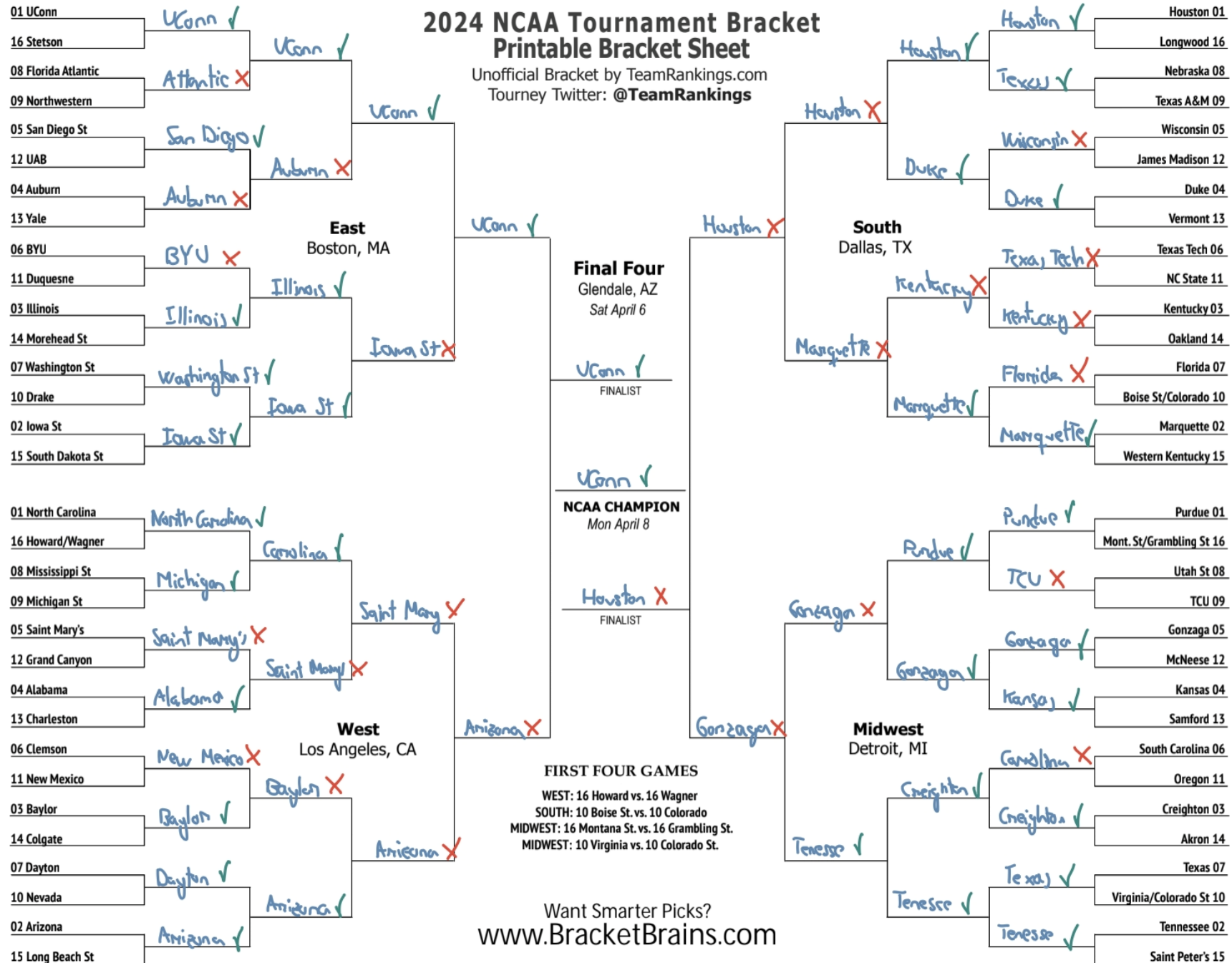


Predicción March Madness 2024

- Red Neuronal
- Mas experiencia de uso
- Mejores resultados al simular el torneo

2024 NCAA Tournament Bracket Printable Bracket Sheet

Unofficial Bracket by TeamRankings.com
Tourney Twitter: @TeamRankings



Comparación con expertos



Analista ESPN

Precisión: 0.67
Puntos: 800
Errores: 42



Maño

Precisión: 0.60
Puntos: 650
Errores: 50



Host de GameDay ESPN

Precisión: 0.57
Puntos: 640
Errores: 54



Conclusiones

- Gran dificultad de predicción
- Los analistas también fallan mucho
- Buen trabajo de la Red Neuronal
- Base para 2025

GRACIAS

Jorge Martinez,
100508957@alumnos.uc3m.es