

A grayscale background image showing a tennis court. In the foreground, several white tennis balls are visible, with one prominently showing the 'HEAD ATP' logo. In the background, a tennis net is visible, slightly out of focus. A solid blue vertical bar runs along the right edge of the image.

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

PROJETO APLICADO EM CIÊNCIA DE DADOS I MODELING

LICENCIATURA EM CIÊNCIA DE DADOS

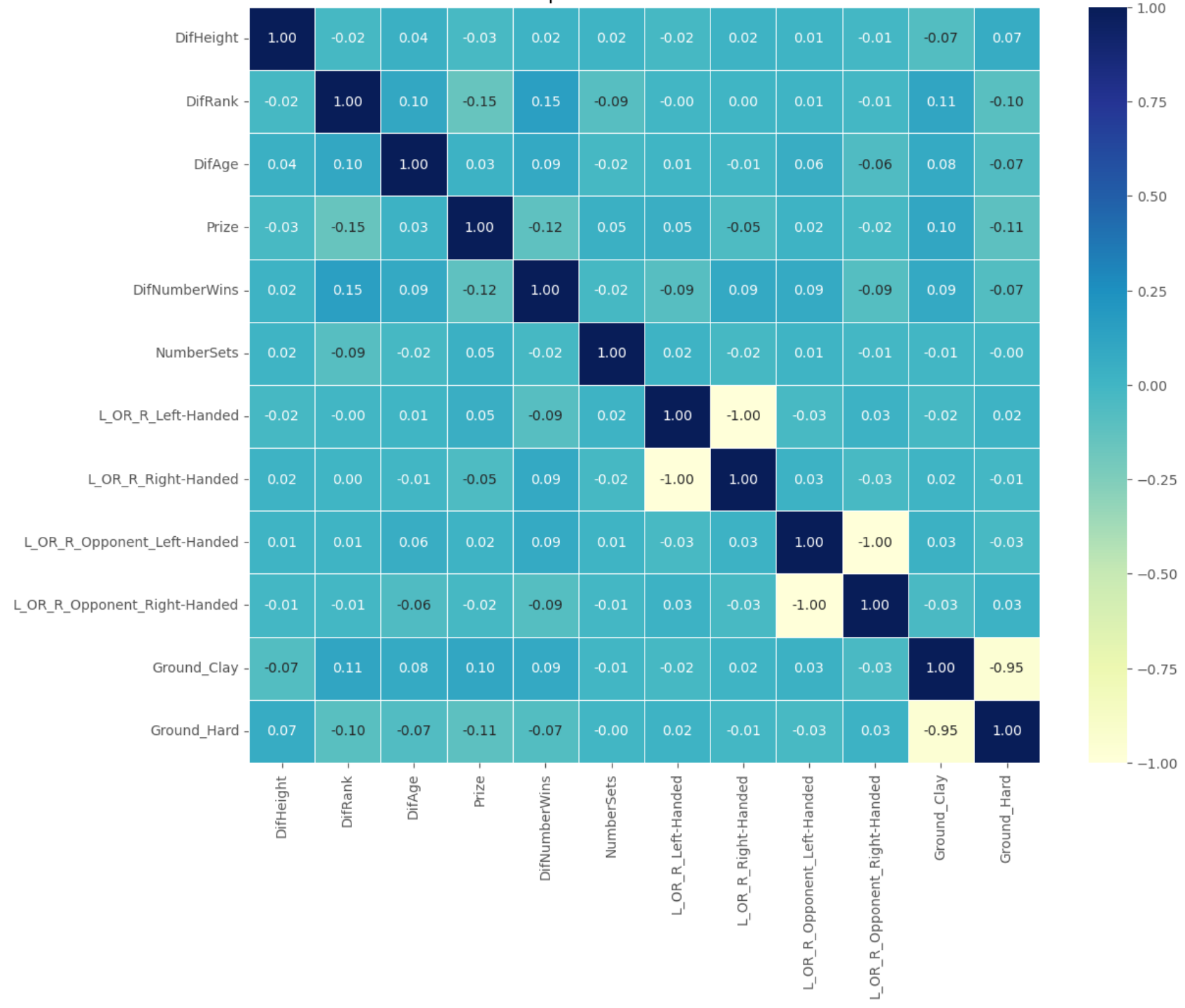
Base de Dados ATP – Brasil

Grupo 5: nº 103303, nº 110451, nº 104716, nº 99239

Docentes: Diana Aldea Mendes e Sérgio Moro

10 de maio de 2023

Correlações entre variáveis



Multicolinearidade

	Variable	VIF
13	DifNumberWins	3.276600
11	NumberWinsPlayer	3.049631
4	RankOpponent	2.150794
8	DifRank	2.092469
0	RankPlayer	1.776492
5	AgeOpponent	1.661207
10	Year	1.491080
12	NumberWinsOpponent	1.432899
9	DifAge	1.424451
3	Prize	1.292864
1	AgePlayer	1.285649
6	HeightOpponent	1.058829
14	DifHeight	1.047947
2	Height	1.014976
7	NumberSets	1.012231

Variáveis consideradas no modelo

Variável	Justificação
DifNumberWins	Mais jogos vencidos -> mais experiência no desporto -> menos sets irão ser necessários
DifRank	Maior rank -> mais experiência e perícia -> mais talento no desporto -> menos sets serão precisos
DifAge	Poderá condicionar o nível de experiência, melhor condição física, recuperação de lesões, mais energia, níveis de resistência
DifHeight	Um jogador mais alto poderá cobrir mais terreno, um jogador mais baixo poderá ter mais vantagem na movimentação -> quanto maior a diferença entre ambos -> mais desequilibrado poderá estar o jogo -> menos sets durará

Total = 4 variáveis para a previsão

Divisão do dataset

Treino

5852 instâncias

41 variáveis

Cerca de 70% do dataset completo

Teste

2509 instâncias

41 variáveis

Cerca de 30% do dataset completo

Modelação – Gradient Boosting

MELHOR DE 3

	precision	recall	f1-score	support
1.0	0.000000	0.000000	0.000000	7.000000
2.0	0.658031	1.000000	0.793750	1651.000000
3.0	0.000000	0.000000	0.000000	851.000000
accuracy	0.658031	0.658031	0.658031	0.658031
macro avg	0.219344	0.333333	0.264583	2509.000000
weighted avg	0.433005	0.658031	0.522312	2509.000000

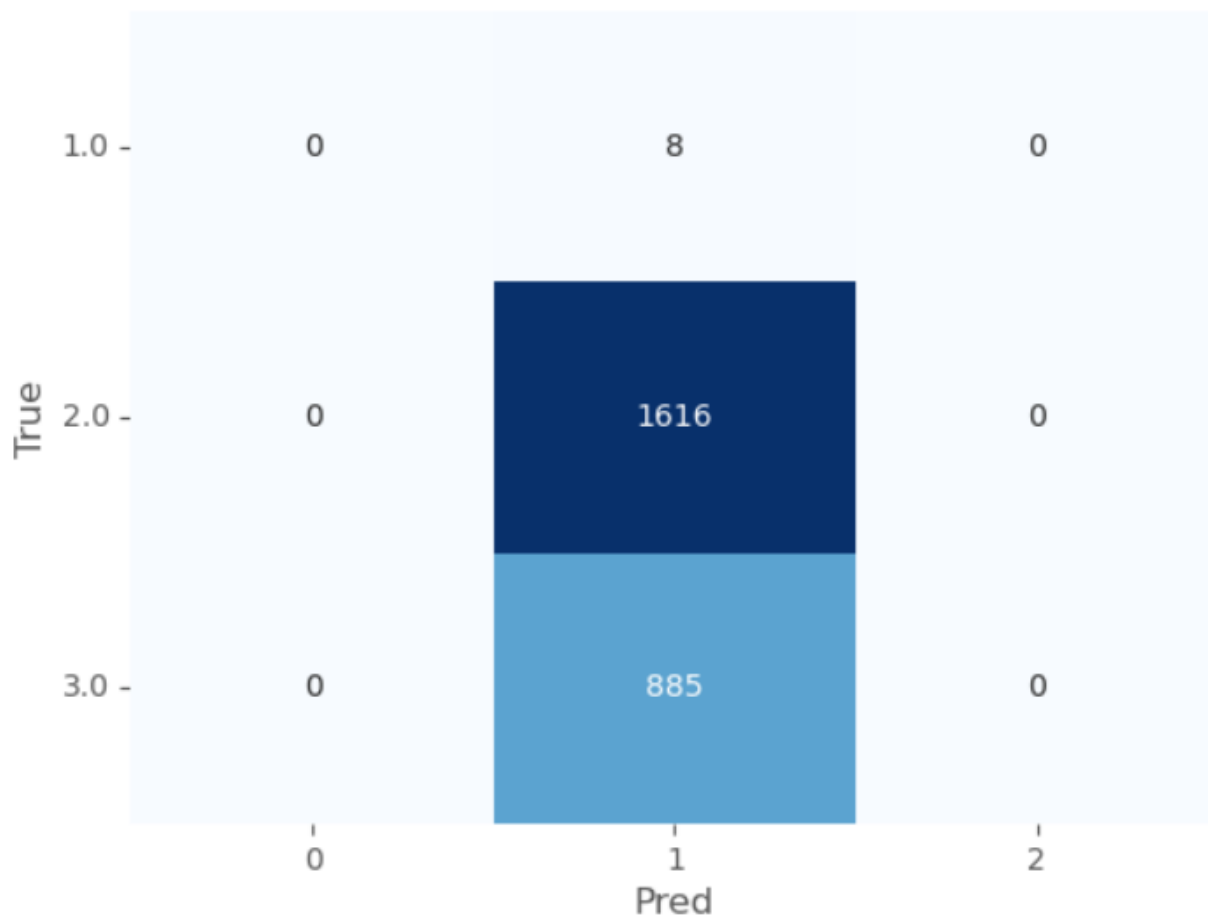
MELHOR DE 5

	precision	recall	f1-score	support
4.0	0.666667	0.200000	0.307692	10.000000
5.0	0.272727	0.750000	0.400000	4.000000
accuracy	0.357143	0.357143	0.357143	0.357143
macro avg	0.469697	0.475000	0.353846	14.000000
weighted avg	0.554113	0.357143	0.334066	14.000000

Avaliação do modelo

MELHOR DE 3

Confusion matrix



MELHOR DE 5

Confusion matrix

