

ATP Players

Apresentação Semanal [5]

UC | Projeto Aplicado a Ciência de Dados I

Docentes | Diana Mendes & Sérgio Moro

Grupo 2

André Silvestre N°104532

Diogo Catarino N°104745

Francisco Gomes N°104944

Rita Matos N°104936

CDB1

Dada a ciclicidade do CRISP-DM recuamos na nossa escolha de eliminar a variável **GameRank** de modo a verificar a sua importância no problema



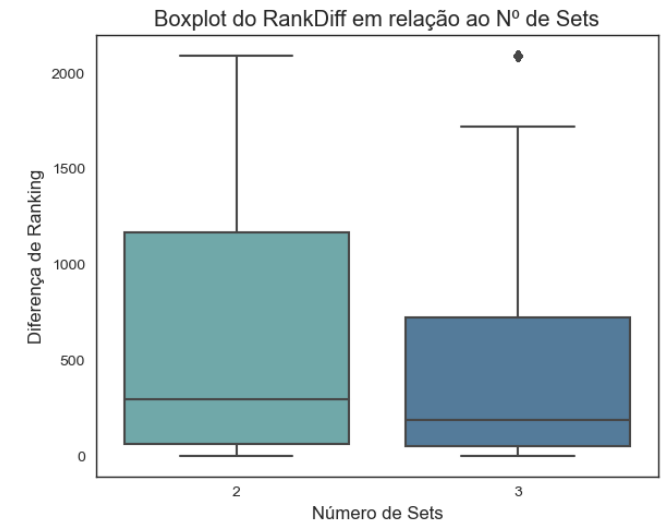
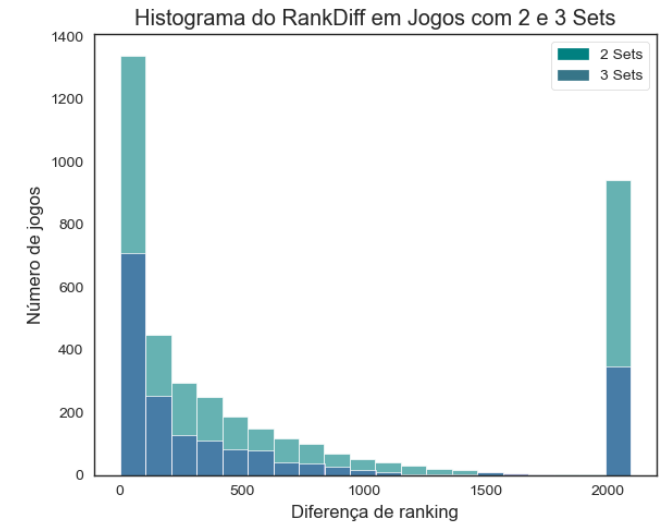
A variável *GameRank* tem
847 NaN (7.36 %)

Associámos a cada jogo
único o *Rank* do **Jogador** e
o *Rank* do **Oponente**

Criámos a variável **RankDiff** como sendo a
diferença absoluta entre os ranks.

Nos casos **NaN** imputámos pelo máximo.

	PlayerRank	OpponentRank	RankDiff
5044	<NA>	1857	2094
800	281	828	547



Imputação dos NAs

Imputámos o máximo de NAs acerca dos jogadores através da pesquisa das suas informações em fontes diversas, como o



```
df_players.to_excel('df_players.xlsx', index=False)
```



Para tal exportámos para o *Excel* o DataFrame com as informações dos jogadores e preenchemos manualmente **≈ 1500** células vazias.

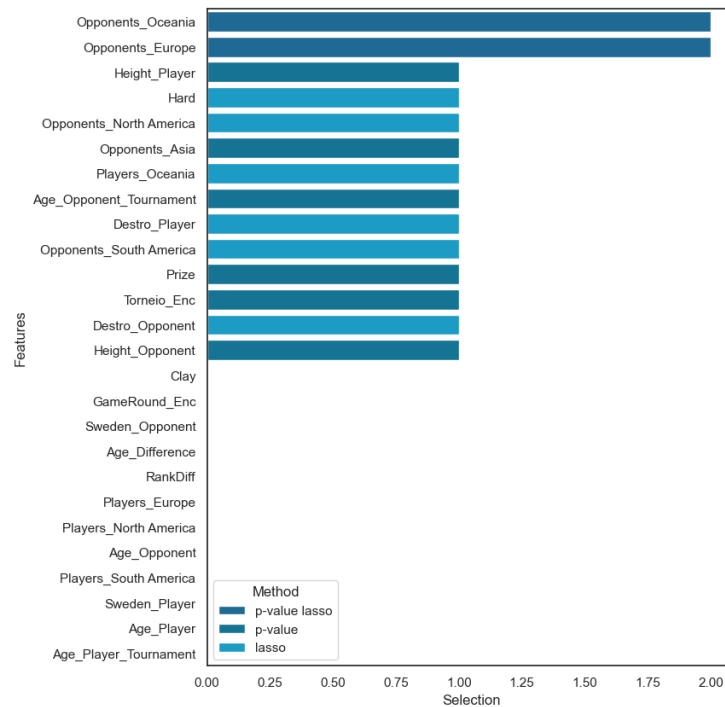


As variáveis **Height** e **Hand** que permaneceram omissas imputámos, respetivamente, pela:

- Média das **Alturas** por país
- Mão Dominante modal

	A	B	C	D	E	G
1	PlayerName	Born_Clean	Height	Hand	Birthdate	Age
2	Aaron Krickstein	United States	183	Right-Handed, Two-Handed Backhand	02/08/1967	55
3	Abdullah Maqdes	Kuwait	180	Right-Handed, Unknown Backhand	31/03/1987	36
4	Adam Chadaj	Poland	178	Left-Handed, Unknown Backhand	01/05/1984	39
5	Adam Heinonen	Sweden	188	Right-Handed, One-Handed Backhand	22/08/2002	20
6	Adam Pavlasek	Slovakia	185	Right-Handed, Two-Handed Backhand	08/10/1994	28
7	Adam Vejmelka	Czech Republic	178	Right-Handed, Unknown Backhand	15/09/1985	37
8	Adil Kalyanpur	India	188	Right-Handed, Two-Handed Backhand	26/01/2000	23
9	Adil Shamasdin	Canada	180	Right-Handed, Two-Handed Backhand	23/05/1982	40
10	Adrian Bodmer	Switzerland	180	Right-Handed, One-Handed Backhand	06/03/1995	28
11	Adrian Forberg Skogeng	Norway	183	Right-Handed, Unknown Backhand	09/08/1990	32

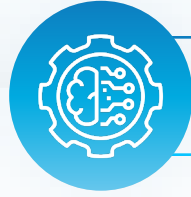
Seleção Avançada das Variáveis



Variáveis Escolhidas

5887 Observações
14 Variáveis





Resumo dos Resultados

Modelos da Semana Anterior 3973 instâncias

	Accuracy	Precision
KNN	0.57	0.52
RL	0.66	0.44
DT	0.56	0.56
RF	0.65	0.58

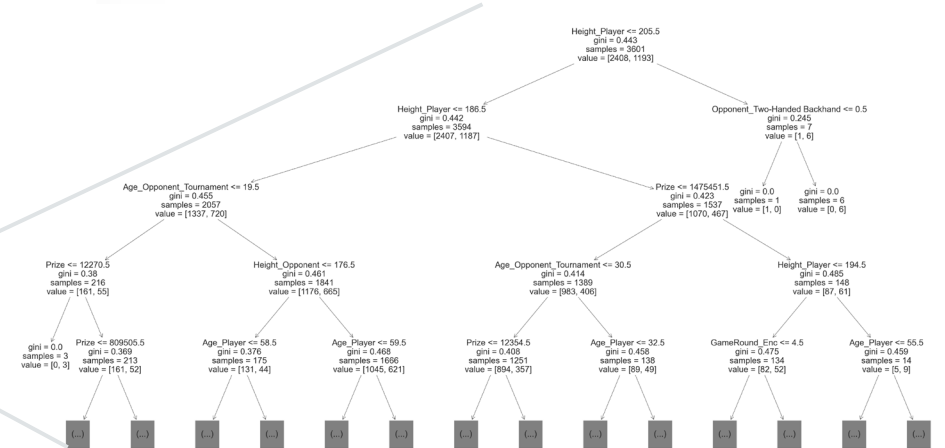
AUC

0.47

0.56

0.51

0.60



Modelos desta Semana 5887 instâncias

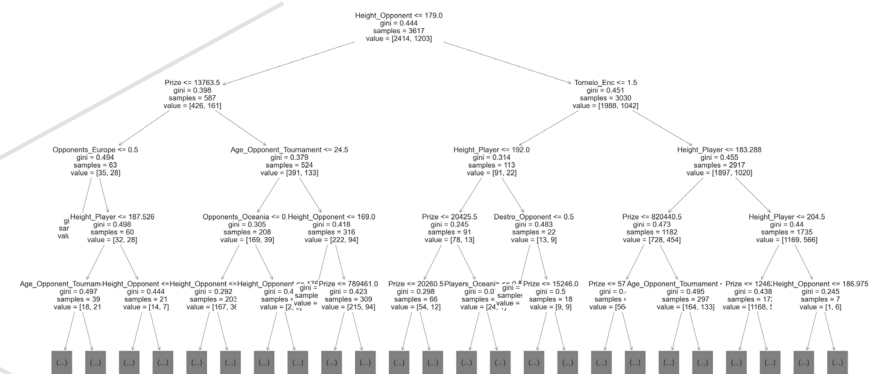
KNN	0.56	0.49
RL	0.67	0.44
DT	0.57	0.58
RF	0.64	0.59

0.44

0.50

0.52

0.52



Tunning do Modelo

Gradient Boosting Classifier

AUC

0.544

