



Evaluation e Deployment Ténis na Bélgica

Grupo 4:

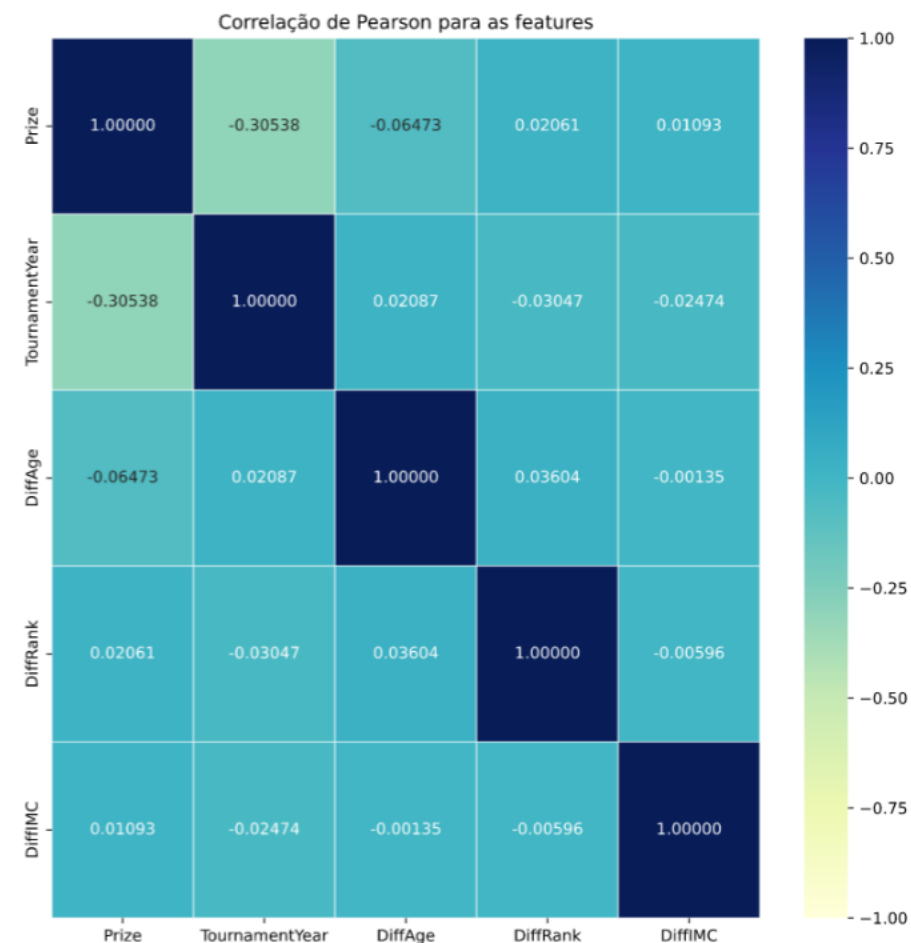
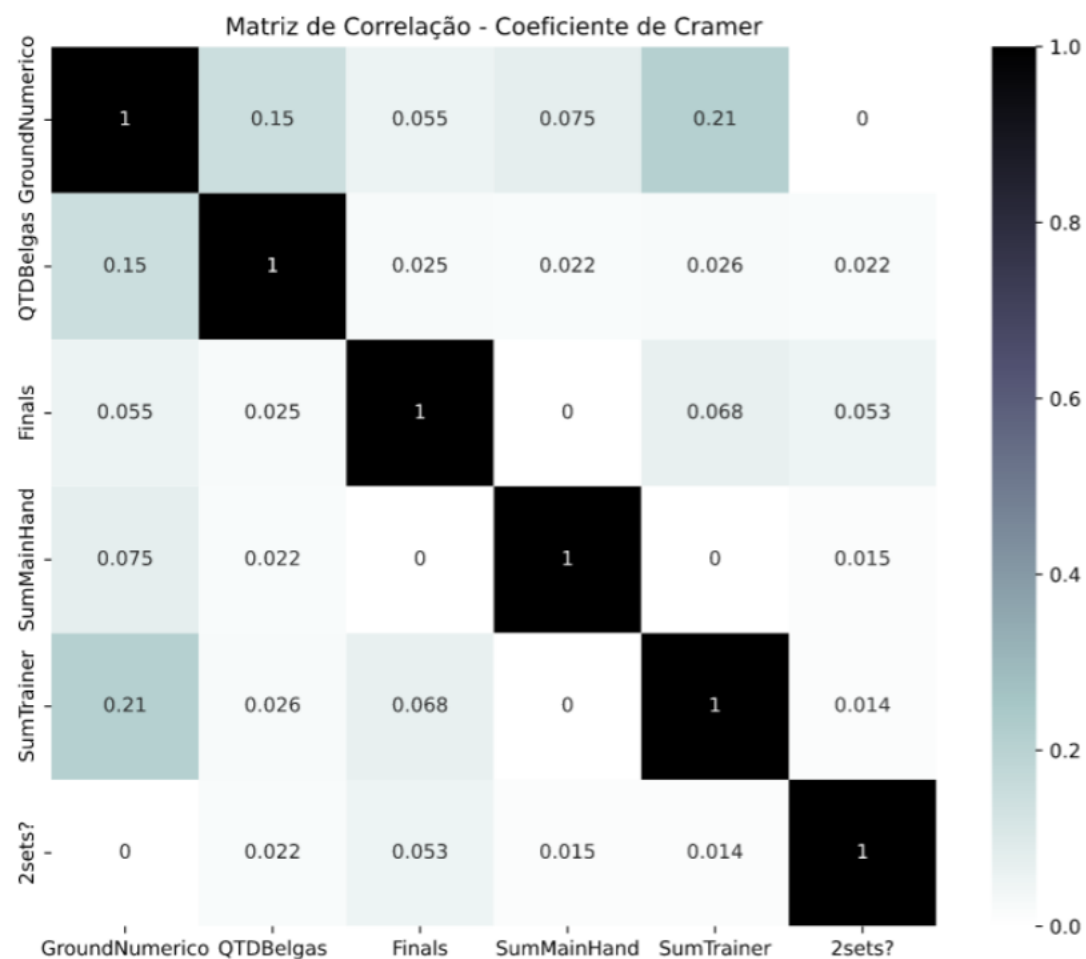
ALLAN KARDEC, N° 103380, TURMA CDB1
DIOGO FREITAS, N°104841, TURMA CDB1
JOÃO BOTAS, N°104782, TURMA CDB1
RICARDO ÂNGELO, N° 104826, TURMA CDB1

Mudanças face às últimas semanas

- Coluna Target atualizada (2sets?)
- Os jogos que possuem RET, foram mantidos, estes deveram ser retirados ou mantidos?
- Não vamos incluir os jogadores que possuem nomes parecidos por incertezas e por muita falta de informação (dados omissos)
- Imputação por Regressão (fancyimpute)
- Mostrar as novas correlações:

NumeroSets	2sets?
2	1
3	0
3	0
3	0
2	1

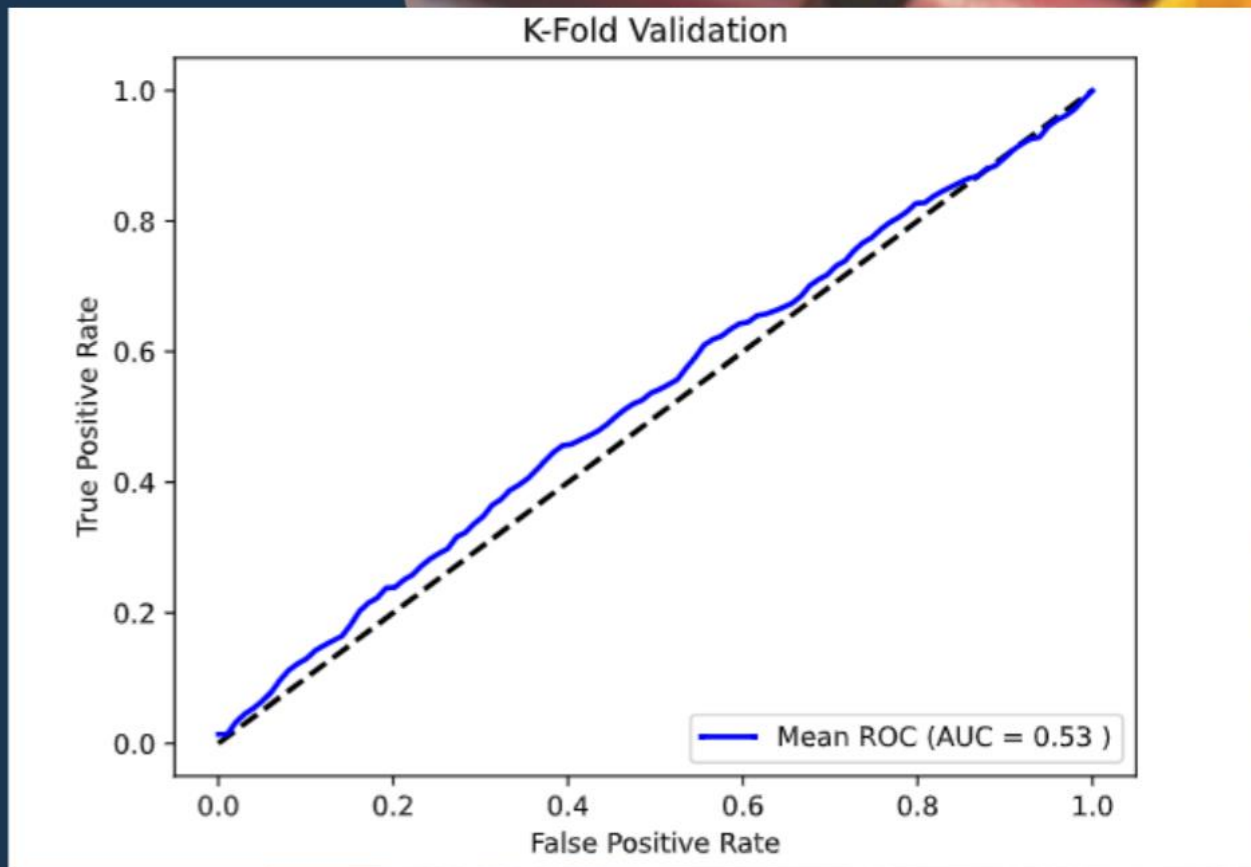
	Score	NumeroSets
33	64 41 (RET)	3
34	63 (RET)	2
35	31 (RET)	2
92	50 (RET)	2
136	41 (RET)	2
...



Coeficiente de Eta entre o (2sets?) e Prize: 0.06606291360597069
 Coeficiente de Eta entre o (2sets?) e TournamentYear: 0.1225714129051667
 Coeficiente de Eta entre o (2sets?) e DiffAge: 0.12140686425617882
 Coeficiente de Eta entre o (2sets?) e DiffRank: 0.6750670776605052
 Coeficiente de Eta entre o (2sets?) e DiffIMC: 0.7637008147133646

Modelo de GB com hiperparâmetros, com K-fold

- Hiperparâmetros iguais aos da semana passada, mas agora com partição k-fold:
 - taxa de aprendizagem: 0.005
 - subsample: 0.7
 - número de árvores: 1250
 - profundidade e features máximas: 6
 - min_size e min_cut (nó folha): 6 e 7
- K-fold com K=10
- Métricas de avaliação:
 - Mean Accuracy: 0.648
 - Mean Precision: 0.505
 - Mean Recall: 0.5
 - Mean F1-score: 0.448



Conclusões e Resultados

- Poucas linhas na base de dados, difícil de tirar conclusões
- Variáveis: Finals não tão importante e IMC e Rank bastante impactantes
- O nº sets é uma característica exigente de classificar com base em informações dos jogadores e do torneio
- Com base nas previsões, conseguimos:
 - ajudar jogadores a verem números de sets de outros jogos
 - marcas a investirem em jogos com 3 sets (+ tempo)
 - talvez completar informação noutras bases de dados

Escrita do relatório:

- Comentar sobre os modelos + importantes
- Incluir gráficos de correlações de Pearson e V de Cramer
- Comentários e decisões gerais desta última semana
- Incluir anexos

Dúvidas:

- seed dos modelos
- estrutura do relatório, no que toca a gráficos e tabelas
- jogos com RET (acerca da decisão tomada)