

Fundação Getulio Vargas
Escola de Matemática Aplicada
Graduação em Ciência de Dados e Inteligência Artificial

Ana Carolina Erthal
Cristiano Larréa
Felipe Lamarca
Guilherme de Melo
Paloma Borges

Field Project CBVela
“Em Busca do Ouro Olímpico”

Field Project desenvolvido com o propósito de conceber modelos de ranking para avaliar e monitorar o desempenho dos atletas da equipe brasileira de vela.

Sumário

Sumário.....	2
Resumo.....	3
Introdução.....	3
Objetivo.....	3
Descrição das Atividades Realizadas.....	4
Entendimento de Negócio.....	4
Entendimento e Mapeamento dos Dados.....	4
Extração dos Dados.....	5
Tratamento e Limpeza de Dados.....	6
Análise Exploratória de Dados.....	6
Modelos de Ranking.....	7
Elo Rating.....	7
Keener's Method.....	7
Avaliação dos Modelos.....	8
K-fold Cross Validation.....	8
Erro de predição.....	8
Resultados e Discussões.....	9
Conclusão.....	9
Considerações Finais.....	9
REFERÊNCIAS.....	10

Resumo

Este projeto se propõe a desenvolver um sistema de ranking para atletas de classes olímpicas de vela, visando proporcionar uma análise mais precisa e justa do desempenho desses competidores. O processo envolveu a compreensão dos conceitos do esporte, pesquisa, mapeamento, extração e tratamento de dados, além da criação e avaliação de modelos estatísticos de *rating*. Este relatório descreve as atividades realizadas e resultados obtidos ao longo dos últimos três meses de trabalho no projeto, apresentando resultados e conclusões obtidas até o momento.

Introdução

Presente nos Jogos Olímpicos desde 1896, a vela tem desempenhado um papel significativo no cenário esportivo brasileiro. Ao longo dos anos, a Confederação Brasileira de Vela (CBVela) assumiu a responsabilidade de administrar e promover a modalidade no país, buscando sempre aprimorar a experiência de atletas, patrocinadores e espectadores.

Fundada em 2013, a CBVela representa oficialmente a vela esportiva brasileira nos cenários nacional e internacional e é filiada à Federação Internacional de Vela (World Sailing) e ao Comitê Olímpico do Brasil. A vela brasileira tem uma história notável, marcada por sete medalhas de ouro olímpicas e um total de 18 medalhas conquistadas em Jogos Olímpicos, tornando-se a modalidade com maior sucesso no país.

No entanto, a CBVela enfrenta um desafio significativo no que diz respeito à avaliação comparativa do desempenho dos atletas: a disparidade nos formatos e padrões de divulgação de resultados, provenientes de diferentes sites e competições, tem impedido uma análise consolidada e padronizada. Esta lacuna na avaliação prejudica a compreensão do progresso dos atletas ao longo do tempo. Além disso, urge a necessidade de um modelo estatístico para classificação dos atletas, já que os modelos atuais de *ranking* no mundo da vela refletem de maneira errônea a realidade dos competidores.

Objetivo

O objetivo central deste projeto é a criação de um sistema de *ranking* abrangente e detalhado, proporcionando uma avaliação justa do desempenho dos atletas da vela. Este sistema visa superar as limitações do modelo de *ranking* atual da World Sailing. Embora este último seja reconhecido internacionalmente, sua ênfase nos últimos 12 meses pode apresentar desafios significativos, especialmente para atletas excepcionais que, por razões diversas, possam ficar afastados das competições por determinados períodos. Isso pode resultar em uma representação inadequada do verdadeiro potencial desses atletas.

A escolha de adotar um novo sistema de ranking centraliza-se na necessidade de oferecer uma avaliação contínua, justa e abrangente do desempenho dos velejadores do cenário internacional atual. Este método permitirá não apenas a classificação dos atletas, mas também irá facilitar a análise histórica e a identificação de tendências de desempenho, oferecendo insights valiosos para o aprimoramento técnico, estratégico e tático, com o objetivo final de assegurar a representação e o sucesso do Brasil nas competições olímpicas de vela.

Descrição das Atividades Realizadas

Entendimento de Negócio

Iniciamos o projeto com uma reunião de *kick-off* com a equipe da CBVela, que permitiu uma compreensão das necessidades, metas e desafios enfrentados pela organização. A partir disso, mergulhamos em uma extensa pesquisa a fim de entender mais sobre a área do negócio e o esporte e criamos um relatório de negócio, que posteriormente nos guiou durante todo o projeto, abordando alguns conceitos e definições essenciais da vela. Neste momento, foram estudados e entendidos conceitos como: funcionamento das regatas; organização das súmulas; classes olímpicas; punições e sistema de pontuação.

Posteriormente, em conjunto com a CBVela, definimos o escopo do projeto, delimitando as classes e eventos relevantes para o ranking. Essa decisão restringiu a elaboração do banco de dados e dos rankings às classes olímpicas de vela. São elas Ilca 7, Ilca 6, 49er, 49erFX, Formula Kite Masculino, Formula Kite Feminino, iQFoil Masculino, iQFoil Feminino, 470 e Nacra17.

Entendimento e Mapeamento dos Dados

Nessa etapa, mapeamos as fontes de dados disponíveis nos diversos websites de competições de vela, selecionando as principais competições a nível internacional utilizando uma janela temporal partindo das competições mais atuais. Esse processo foi fundamental para compreender a estrutura desses dados e decidir a melhor abordagem para trabalhar com eles nas próximas etapas do projeto.

Construiu-se uma proposta inicial de banco de dados orientada pelos requisitos do ranking, definindo as colunas e tipos de dados para armazená-los de forma eficiente. Isso proporcionou uma visão estruturada antes mesmo da extração efetiva dos dados, de forma que foi possível definir quais eram as informações importantes a serem retiradas de cada súmula.

Com os links de todas as competições necessárias em mãos, foi possível constatar que os dados da vela mundial são, acima de tudo, bastante descentralizados e despadronizados. Por isso, realizamos uma clusterização cuidadosa para compreender a estrutura de cada súmula e como ela se comparava às demais. Agrupamos os sites que compartilham padrões semelhantes de divulgação de resultados, visando a eficiência na criação de *scrapers*, *OCR's* e extrações via *API's*. Nessa etapa, foram mapeadas cerca de 150 súmulas, divididas em 13 clusters.

Figura 1 - Planilha de clusterização

Clusters-Surmas
Arquivo Editor Ver Insírtir Formatar Dados Formatações Extensões Ajuda

Menu
100%
132
Pág. 1
10
B
Z
A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022

Figura 1: visão geral da planilha de clusterização utilizada. Ao total, totalizaram-se 13 clusters.

Extração dos Dados

A clusterização das fontes de dados nos permitiu desenvolver diversos métodos para extração dos dados, conforme tratados nos tópicos abaixo.

Web-scrappers

A utilização de *Web-Scrappers* para coleta de dados permitiu uma otimização de tempo considerável. Aproximadamente 33% das súmulas mapeadas puderam ter seus dados coletados dessa forma, através de scrappers. Nesse método, foram utilizados recursos como BeautifulSoup e Selenium (pacotes de Python), além do conhecimento básico de HTML necessário para a montagem dos scripts.

OCR's

Algumas súmulas estavam no formato pdf e em formato de imagem, o que impossibilitou a utilização de scrappers. Portanto, para essas, foram utilizadas ferramentas de OCR online, disponibilizadas na internet, de forma que com um leve tratamento manual, esses documentos pudessem ser convertidos em planilhas.

API's, manipulação de JSON's e expressões regulares

Em determinado momento do projeto, foi observado que algumas súmulas não possuíam seu conteúdo em sua estrutura HTML, mas carregavam esses de uma URL através de uma requisição HTTP. Assim, foram utilizados os pacotes *requests* e *re* da linguagem Python para escrever scripts que carregavam os dados e aplicavam, em seguida, diversas operações de expressões regulares para obter os dados necessários.

Manual

Por fim, alguns clusters não se encaixaram em nenhuma das automações acima, muitas vezes com múltiplas tabelas para uma mesma súmula ou até mesmo com difícil interpretação. Para esses, os dados foram adicionados ao nosso banco de súmulas de maneira manual.

Figura 2 - Distribuição dos métodos de extração de dados

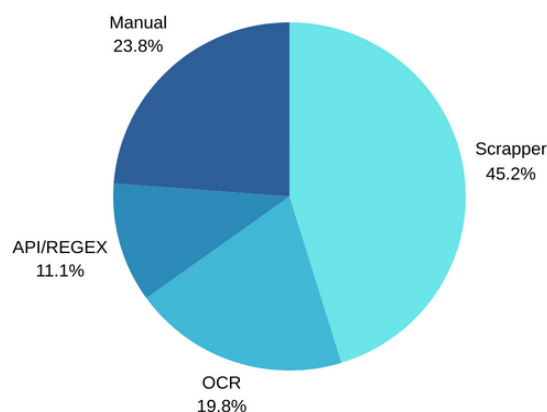


Figura 2: Gráfico de setores com a distribuição dos métodos para as súmulas.

De forma geral, as automações foram cruciais para lidar com a diversidade e o volume de dados presentes nos diferentes websites. Com a descentralização dos dados, conseguimos exercitar diferentes práticas que já tinham sido trabalhadas ao longo do curso.

Além disso, os *scrapers* desenvolvidos são parte dos entregáveis deste projeto à CBVela. Isto é, como utilizamos *scraping* em clusters de súmulas mais recorrentes, espera-se que súmulas futuras sejam frequentemente publicadas seguindo o mesmo padrão. Já que, para manter os rankings e banco de dados atualizados, é necessário que os dados futuros continuem sendo coletados, os *scrapers* desenvolvidos estão prontos para serem utilizados, bastando poucas adaptações.

Essa etapa, apesar de mais demorada, foi concluída de maneira exitosa, de forma que ao final conseguimos chegar a um banco de súmulas com mais de 130 mil dados. Foram coletados dados de 7 competições mundiais, 10 categorias olímpicas e 3170 diferentes competidores, ao longo dos anos 2015 a 2023.

Tratamento e Limpeza de Dados

Com os dados extraídos, iniciamos a etapa crítica de tratamento e limpeza. Começamos conduzindo um *sanity check* para identificar possíveis inconsistências no banco de dados recém-criado. Essa validação inicial foi essencial para uma limpeza eficiente, de forma a garantir a integridade dos dados antes de avançarmos para as etapas subsequentes, proporcionando uma base sólida para o restante do projeto.

Encontramos no *sanity check* divergências variadas, como diferentes nomenclaturas para uma mesma classe e divergência na forma como a pontuação e punição eram registradas. Por fim, o desafio mais complexo foi relacionado à falta de padronização nos nomes dos competidores.

A questão da nomeação dos velejadores foi apontada pela CBVela como problemática desde o início. Na publicação de súmulas de eventos diferentes, é comum que os nomes de competidores sejam publicados de forma diferente (escritas erradas, utilizando sobrenomes diferentes, apelidos, etc), um impedimento muito grande para a construção de um banco organizado e robusto para criação dos rankings.

Por isso, elencamos um processamento automatizado utilizando ferramentas de Processamento de Linguagem Natural, combinando o cálculo de distâncias de Levenshtein, permutações e outras técnicas para buscar diferentes grafias de um mesmo nome, a um processamento manual, contando com uma inspeção minuciosa dos nomes de competidores de todas as classes. Encontramos, nesse processo, competidores que tinham até 9 grafias diferentes em publicações de súmulas, evidenciando a importância dessa etapa.

Análise Exploratória de Dados

Após a conclusão do processo de limpeza, embarcamos em uma análise exploratória de dados detalhada. Essa etapa envolveu a identificação de padrões, tendências e insights relevantes para a construção dos modelos de ranking. Os resultados dessa análise forneceram uma compreensão aprofundada a respeito da dinâmica do esporte, destacando áreas de interesse específicas para a CBVela. Os dados aqui apresentados se referem principalmente às classes Ilca 7, Ilca 6, 49er e 49erFX, que são aquelas que possuem mais informações no banco de dados.

Por exemplo, uma das visualizações produzidas mostra a média de competidores por campeonato. A partir dela, conseguimos identificar que a classe Ilca 7 é a que normalmente possui mais competidores, enquanto o Nacra 17 é o que possui menos¹:

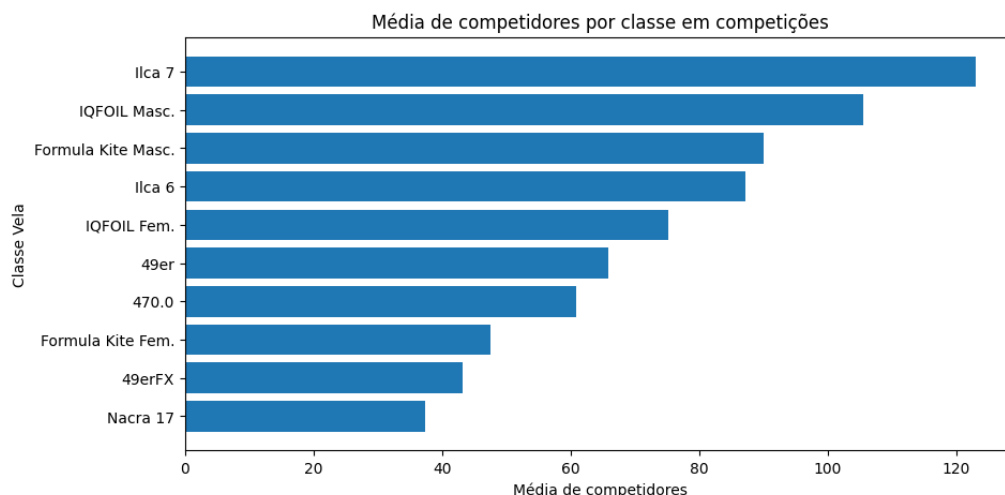


Figura 3: Gráfico da média de competidores por classe em competições

A coleta e análise dos dados também permitiu testar algumas hipóteses prévias da equipe da CBVela. Por exemplo, é de conhecimento comum que há uma atualização de competidores a cada ciclo olímpico (na virada de 2016 para 2017 e na virada de 2020 para 2021), mas os dados coletados nos permitem quantificar essa rotatividade de maneira mais precisa. Observe o gráfico abaixo, que apresenta, para cada edição do Campeonato Mundial, o percentual de competidores que já havia participado de pelo menos uma competição anterior:

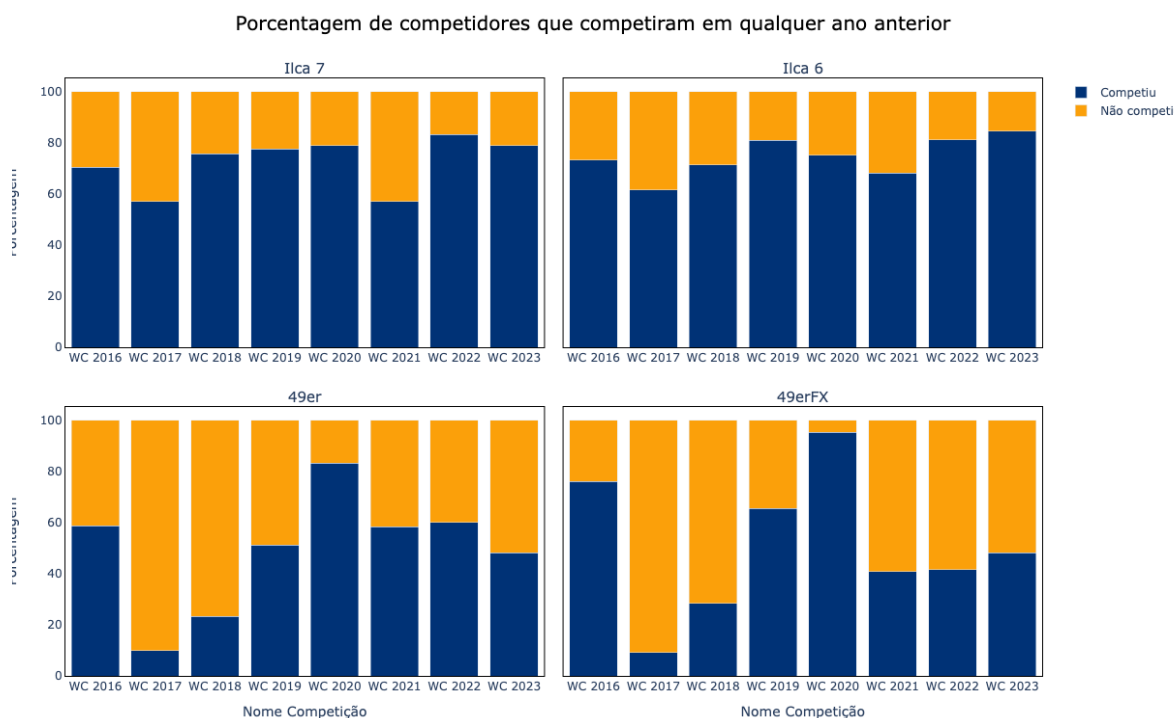


Figura 4: Gráfico da porcentagem de competidores que competiram em ano anterior

¹ Para a análise, considere que uma dupla representa um único competidor.

É possível identificar de imediato que, de fato, há uma grande renovação entre um ciclo olímpico e outro, com altos percentuais de competidores que não haviam participado de competições anteriores do evento.

Outra hipótese que surgiu ao longo da análise dos dados é que é razoável esperar que competidores mais bem-sucedidos no esporte costumam participar de mais competições em média. O gráfico abaixo ajuda a comprovar essa hipótese. Observe, por exemplo, que os competidores do Ilca 7 que ficaram entre as 10% melhores colocações participaram, em média, de 13.3 competições. Esse número de participações é bem mais do que os 10% piores, que participaram em média de 3.36 competições:

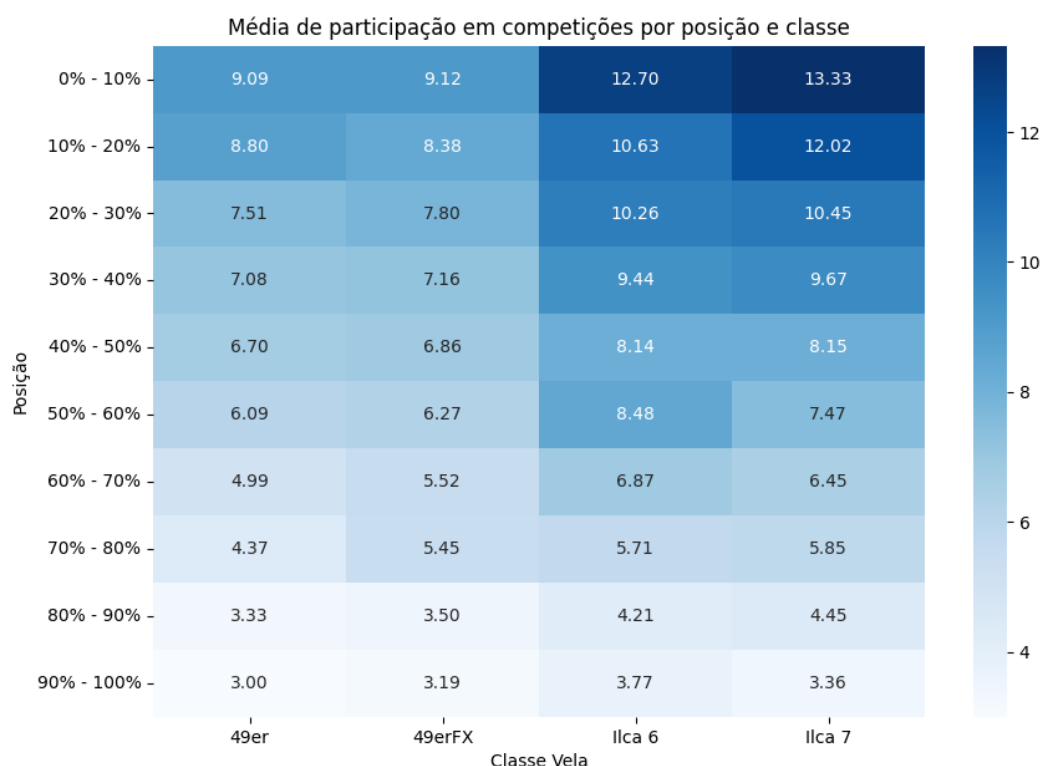


Figura 5: Gráfico da média de participação em competições por posição e classe

Embora não seja possível identificar as razões exatas que justifiquem esse fenômeno, podemos elencar algumas possibilidades: (i) a frustração da derrota, que afasta do esporte jogadores pouco vitoriosos; e (ii) o fato de que se trata de um esporte caro em termos de equipamento - e em termos de participação, já que muitos dos campeonatos importantes são realizados internacionalmente -, tornando-o altamente dependente de patrocinadores e favorecendo a permanência de competidores melhor colocados.

Por fim, o último gráfico mostra a porcentagem de cada punição por cada classe de vela, considerando todos os campeonatos no banco. A punição DNC é a mais comum em todas as classes, casos em que o competidor não compareceu à largada ou não partiu. Nos casos do Ilca 6 e 7, BFD é a segunda punição mais comum, enquanto para o 49er e 49erFX a segunda mais frequente é a UFD. Ambas sinalizam desclassificação, mas indicam infração de [regras diferentes](#).

Porcentagem de punições por classe de vela

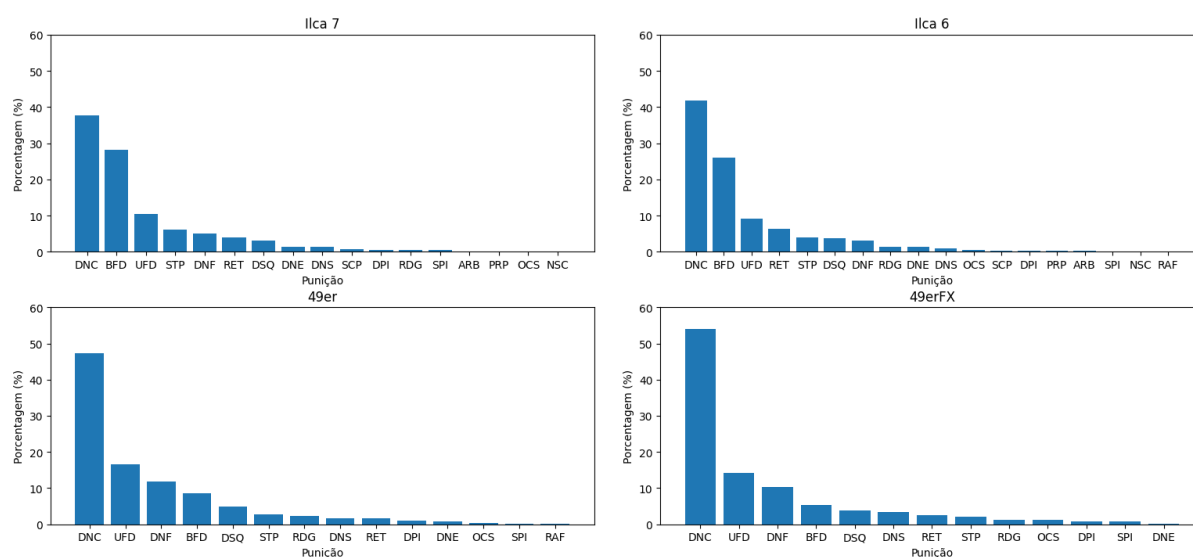


Figura 6: Gráfico da porcentagem de punições por classe da vela

Esta seção não esgota a extensa análise exploratória dos dados realizada para a CBVela, mas contempla alguns dos principais achados ao longo do processo de investigação do banco de dados. A análise completa conta uma série de outras informações relevantes, como a média de competidores por evento, punições mais frequentes por evento, média de punições por competidor, etc. Naturalmente, há outras questões que ainda podem ser exploradas - por exemplo, a rotatividade de competidores em outras competições de vela. Entre os entregáveis da CBVela estão também algumas funções que facilitam esse e outros tipos de análise.

Modelos de Ranking

A elaboração de modelos de ranking é, certamente, uma etapa muito importante do projeto, o que explica sua concomitância com o desenvolvimento de etapas anteriores. Isto é, desde muito cedo foi necessário realizar uma revisão da literatura de ranking e avaliar a relevância de modelos estado da arte no ranqueamento de atletas da vela.

Tratamos de um esporte bastante distinto dos normalmente utilizados em publicações de elaboração e avaliação de rankings. Em geral, têm-se esportes de enfrentamento de duplas, e na vela tratamos de uma única competição em que temos mais de uma centena de competidores.

Inicialmente, avaliamos sete modelos diferentes de ranking para adaptação e aplicação ao banco de súmulas construído. Através de pesquisas, conversas com o Prof. Moacyr (experiente na área de aplicação de sistemas de ranking a esportes) e testes, foi possível destacar dois modelos que melhor se aplicavam ao objetivo: o sistema de Elo Rating e o Keener's Method.

A escolha desses modelos se justifica, principalmente, em uma combinação entre resultados preliminares satisfatórios, limitações matemáticas dos demais modelos aos nossos dados e a explicabilidade dos modelos. Isto é, era desejável que os modelos levassem a rankings coerentes, mas que fossem, a certo ponto, compreensíveis para um atleta que deseja compreender seu ranqueamento. Utilizar, por exemplo, o modelo de Máximo a Posteriori traria uma complexidade indesejada.

Elo Rating

O modelo de elo rating é um modelo clássico desenvolvido para avaliação de jogadores de xadrez. A essência do modelo Elo reside na atribuição de pontuações numéricas, ou ratings, a cada competidor. Inicialmente, os jogadores recebem um rating base e à medida que os competidores enfrentam uns aos outros, os resultados desses confrontos impactam diretamente seus ratings. A diferença entre os ratings define o resultado esperado, e o valor de atualização é baseado na diferença entre o resultado esperado e o observado. Em outras palavras, derrotar um oponente de rating menor acarreta uma atualização menor de rating quando comparada com uma vitória contra um oponente de rating maior.

Os ratings são inicializados arbitrariamente em 1500, comum de acordo com a literatura, e atualizados cronologicamente de acordo com o resultado das competições, o que leva em conta a temporalidade dos dados. Em cada competição calculamos e somamos a mudança para cada par de competidores e dividimos pelo número total de outros competidores.

A atualização unitária se dá a partir do seguinte cálculo:

$$\Delta n = K(N - E)$$

- Δn é a mudança a ser computada
- K é um fator de ponderação (hiperparâmetro)
- S é o resultado da partida (1 para vitória, 0 para derrota)
- E é a expectativa de vitória do jogador, calculada por:

$$E = \frac{1}{1 + 10^{\left(\frac{R_0 - R_c}{400}\right)}}$$

- R_0 é o rating inicial do jogador
- R_c é o rating do oponente

A variação do hiperparâmetro K nos fornece diferentes modelos, que geram rankings diferentes. Por isso, selecionamos valores de K de acordo com a revisão de literatura realizada, mas também nossos experimentos, e avançamos para a realização de testes do modelo utilizando três valores diferentes para esse hiperparâmetro, $K = 40, 100$ e 400 .

Keener's Method

Esse modelo parte da elaboração de uma matriz de vitórias W :

$$W_{ij} = \#\{\text{competidor } i \text{ vence de } j\}$$

A partir dessa matriz, realizamos o cálculo de $D = \text{diag}((W + W^T)1)$, matriz diagonal que descreve o total de competições de i . Utilizamos o autovetor de Perron-Frobenius de $D^{-1}W$, matriz de vitórias normalizada, como ratings, e a ordenação decrescente dos ratings nos fornece o ranking final.

É importante destacar que esse método é dependente de que a matriz W seja não-negativa e irredutível. Nossa matriz é certamente não-negativa, e conforme a literatura, adicionamos uma perturbação a W para forçar a irredutibilidade.

A partir desse modelo, realizamos algumas modificações, gerando mais versões do modelo que poderiam performar melhor que o original. Chamamos a versão original

apresentada acima de modelo α (alpha), e consideramos também uma versão β (beta). A versão β utiliza uma matriz S ao invés de W , dada por:

$$S_{ij} = \sum_{\text{jogos entre } i \text{ e } j} \frac{\# \text{ pontos de } i \text{ contra } j}{\text{total de pontos no jogo}}$$

Aplicamos uma função não linear de *skewing* em S para compensar situações em que um time é muito superior a outro, e obtém uma vitória “de lavada” simplesmente porque podia, recebendo uma quantidade muito alta de pontos, mas contra um time muito inferior. Assim, obtemos:

$$K_{ij} = h\left(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right), \text{ onde } h(x) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(x - \frac{1}{2}\right) \sqrt{|2x - 1|}$$

E então podemos, novamente, calcular $D = \operatorname{diag}((K + K^T)1)$, e obter o autovetor de Perron-Frobenius de $D^{-1}K$ como ratings.

Além disso, consideramos o fato de que esse método não leva em conta questões temporais, sendo comumente utilizado para avaliar uma temporada de um esporte, mas tratamos de avaliações de consecutivos. Sabendo que um resultado de 2015 não deveria ter tanto impacto no ranking atual quando um resultado recente, estabelecemos versões do modelo com *time decay*, atribuindo um peso diferente a vitórias/pontuações de acordo com o ciclo olímpico (cada 4 anos) em que a competição ocorreu.

Obtivemos, assim, a partir do método de Keener's, quatro modelos finais a serem avaliados: Keener's alpha sem *time decay*, Keener's alpha com *time decay*, Keener's beta sem *time decay*, Keener's beta com *time decay*.

Avaliação dos Modelos

Para a avaliação dos modelos de ranking, utilizamos duas métricas de erro para filtrar os modelos mais eficazes na produção de um ranking efetivo. Essa avaliação será somada a uma avaliação qualitativa dos resultados dos rankings desses modelos pela CBVela, que contando com *expertise* no assunto poderá avaliar qual modelo deve ser o produto final deste projeto.

Apresentamos aqui a avaliação de nossos modelos nas classes 49er, 49erFX. Para essas, dentre outras classes, é possível remontar até o ano de 2015 com maior completude dos dados. Outras classes, como o 470, são mais novas, e não possuem tantos dados históricos, essenciais para o funcionamento do modelo. Assim, imaginamos que modelos de ranking serão mais efetivos para essas classes daqui a alguns anos, quando houver maior quantidade de dados para fornecer aos modelos.

K-fold Cross Validation

Uma das métricas utilizadas para avaliação foi o K-fold Cross Validation. A partir do rating de um par de competidores, é possível estimar (de forma diferente para cada modelo) a probabilidade de vitória de cada um. Assim, estimamos o erro considerando a diferença entre a probabilidade de vitória e o resultado da competição (para cada par). Calculamos o erro para uma competição, tirando a média dos erros par a par. Por fim, para que a nossa estimativa seja efetiva, passamos a calcular o rating dos competidores sem levar em conta uma competição por vez, e utilizamos essa competição como resultado real, realizando o K-fold Cross Validation.

Os resultados, como é possível observar na tabela abaixo, mostram que o desempenho utilizando o Elo Rating foi consistentemente melhor que os resultados do Keener's Method. Dentro do modelo Keener's, pode-se observar que utilizar o time decay traz em geral um pouco melhores do que não considerar diferenças temporais, e que a performance do modelo α é muito melhor. No entanto, os erros de validação do Elo se mostram bem menores, mostrando efetividade do modelo.

MODELO	Especificações	49er	49erFX
Elo	k=40	30.18%	30.07%
	k=100	26.21%	26.39%
	k=400	22.48%	23.24%
Keener's alpha (α)	sem time decay	44.73%	54.39%
	com time decay	45.39%	52.11%
Keener's beta (β)	sem time decay	76.32%	81.37%
	com time decay	75.46%	76.41%

Destaca-se ainda, o fato de que utilizar um valor de k mais alto no Elo nos leva a resultados melhores. Discutiremos essa questão mais adiante.

Erro de predição

Utilizamos o erro de predição [4] como segunda métrica de avaliação de nossos modelos. Aqui, a abordagem é realizar uma divisão entre conjunto de treino e teste, utilizar o conjunto de treino para produção do rating e o conjunto de teste na seguinte avaliação:

$$E = \frac{\#\{(competidor\ i\ vence\ de\ j)\ e\ (rating\ de\ i < rating\ de\ j)\}}{\#\{competições\ entre\ i\ e\ j\}}$$

O erro total é calculado como a média entre esse cálculo no conjunto de treino e teste.

MODELO	Especificações	49er	49erFX
Elo	k=40	11.38%	10.2%
	k=100	10.61%	9.90%
	k=400	9.69%	9.41%
Keener's alpha	sem time decay	31.21%	19.38%
	com time decay	32.29%	22.31%
Keener's beta	sem time decay	25.14%	22.07%
	com time decay	25.04%	22.32%

Veja acima que os resultados observados confirmam nossa avaliação anterior, ainda que haja algumas diferenças. Os resultados internos do Keener's tiveram algumas diferenças (nessa métrica o método beta obtém melhor desempenho, por exemplo), refletindo características dos modelos e da própria métrica de erro, que leva em conta previsões de vitórias e derrotas ao invés de considerar o quão distante estamos do que ocorreu.

Novamente, no entanto, o insight mais importante é a prevalência do Elo como modelo de geração de rankings.

Resultados e Discussões

A partir das avaliações apresentadas observa-se que os modelos do Elo geram nossos rankings mais satisfatórios. Resta saber qual K utilizar.

Ao se utilizar de métricas de erro para otimizar o valor de K encontramos problemas. Devido ao pequeno volume de dados, valores baixos de K (no intervalo de 10 a 40) como normalmente utilizados no xadrez são insuficientes para que o modelo descreva corretamente a habilidade dos jogadores. Valores mais altos de K permitem que os melhores atletas se afastem dos demais mais rápido e portanto tem erros mais baixos.

O problema, porém, é mais complexo. O *underfit* do modelo é severo, e aumentar o K indefinidamente, sempre reduz o erro. A resposta para a escolha do ranking, certamente, não é utilizar o maior valor de K possível. Surge então uma decisão a ser tomada, que permita o modelo aprender com os dados disponíveis, mas que não seja ignorante quanto às nuances do problema.

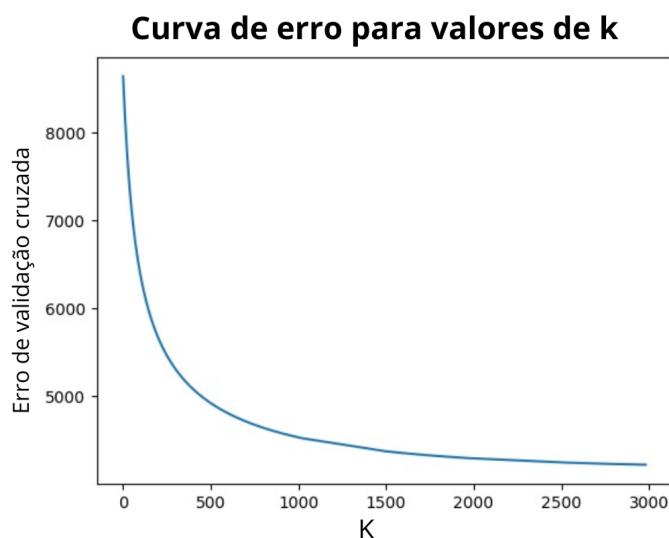


Figura 7: gráfico da curva de erro de validação cruzada conforme o aumento de K

Assim sendo, não é possível determinar, puramente através de métricas de erro, qual valor K fornece o melhor ranking. Por isso, uma qualificação dos rankings gerados pelos modelos finais – Elo rating com $K = 40$, Elo rating com $K = 100$ e Elo rating com $K = 400$ –, se faz necessária para determinar o modelo final do projeto.

Conclusão

O resultado final deste projeto, para nós, foi bastante satisfatório. Conseguimos produzir entregáveis relevantes e que possuem valor para a CBVela, sendo estes principalmente (1) o banco de dados populado com resultados de súmulas dos últimos anos, com mais de 130 mil linhas de dado, (2) *scrapers* para pronta utilização para extração de dados em anos subsequentes e manutenção deste banco, (3) uma análise exploratória desse banco contendo *insights* e informações sobre os campeonatos, atletas e resultados de campeonatos e, por fim (4) os três rankings de melhor performance de acordo com nossas métricas para realização de uma análise qualitativa por parte da CBVela. Após a realização dessa análise, será possível determinar o modelo final e mais efetivo do projeto.

Acreditamos que, independentemente da escolha do ranking final, todos os três modelos escolhidos para análise qualitativa são boas estimativas do cenário atual da vela, e que de fato o ranking é representativo dos dados colhidos – uma satisfação para nós. De toda maneira, a compilação inédita de uma quantidade expressiva de dados de competições de vela em um único banco de dados é um produto que permite a exploração de infinitas possibilidades de agora em diante.

Considerações Finais

Expressamos nossa gratidão a todos os colaboradores, professores, especialistas e demais envolvidos que contribuíram para o sucesso deste Field Project. Em especial, queremos expressar nossos agradecimentos:

Ao Professor Walter Sande, que liderou com maestria a organização deste Field Project. Sua orientação, disposição para esclarecer dúvidas e constante apoio foram fundamentais para a condução bem-sucedida de todas as fases do projeto.

Ao Professor Moacyr, cuja expertise na etapa de modelagem foi importantíssima. Sua contribuição na seleção dos modelos estatísticos de rating, bem como seu esclarecimento de dúvidas, foram cruciais para a construção de bases sólidas para a avaliação do desempenho dos atletas.

À equipe da CBVela, representada por Maria Hackerott, Gustavo Baiano e Ricardo Paranhos. Sua participação ativa nessa jornada foi vital para a compreensão aprofundada do esporte, esclarecimento de dúvidas específicas e conexão com outros profissionais do mundo da vela.

Agradecemos, também, a todos aqueles que, de alguma forma, contribuíram para o sucesso deste projeto, fornecendo insights valiosos, compartilhando conhecimentos e apoiando nossos esforços em cada etapa.

Este Field Project não apenas representa uma conquista acadêmica, mas também fortaleceu nossa compreensão prática sobre a complexidade da gestão esportiva, a importância da padronização de dados e a necessidade de abordagens inovadoras na avaliação do desempenho esportivo.

Mais do que um encerramento, consideramos este um ponto de partida para futuras pesquisas e aprimoramentos neste campo. A construção de um sistema de ranking robusto e justo é apenas o começo de um compromisso contínuo com o avanço do esporte da vela no cenário brasileiro.

REFERÊNCIAS

[1] BARROW, D. et al. Ranking rankings: an empirical comparison of the predictive power of sports ranking methods. **Journal of Quantitative Analysis in Sports**, maio 2013.

Disponível em: <https://doi.org/10.1515/jqas-2013-0013>

[2] KEENER, J. P. THE PERRON-FROBENIUS THEOREM AND THE RANKING OF FOOTBALL TEAMS. **Society For Industrial And Applied Mathematics**, v. 35, p. 80–93, mar. 1993. Disponível em: <https://tinyurl.com/4wxdsfth>

[3] LANGVILLE, A. N.; MEYER, C. D. **Who's #1?: The Science of Rating and Ranking**. [s.l.] Princeton University Press, 2013. Disponível em: <https://www.dcs.bbk.ac.uk/~ale/dsta+dsat/dsta+dsat-3/lm-ch3-keener.pdf>

[4] OCHIENG, P. J.; LONDON, A.; KRÉSZ, M. A Forward-Looking Approach to Compare Ranking Methods for Sports. **Information** **2022**, 13, 232. Disponível em: <https://doi.org/10.3390/info13050232>