

Avaliação de capacidade preditiva de um modelo para identificar preferência de anime por usuário

Bruno de Lima Santos
Instituto Federal do Espírito Santo

Contextualizando o problema

A indústria de anime utiliza de tentativa e erro: cada grande sucesso supera os gastos de uma gama de investimentos sem resultado.

Problema no futuro: com o streaming, muito material “ruim” fica disponível e os bons não se destacam.

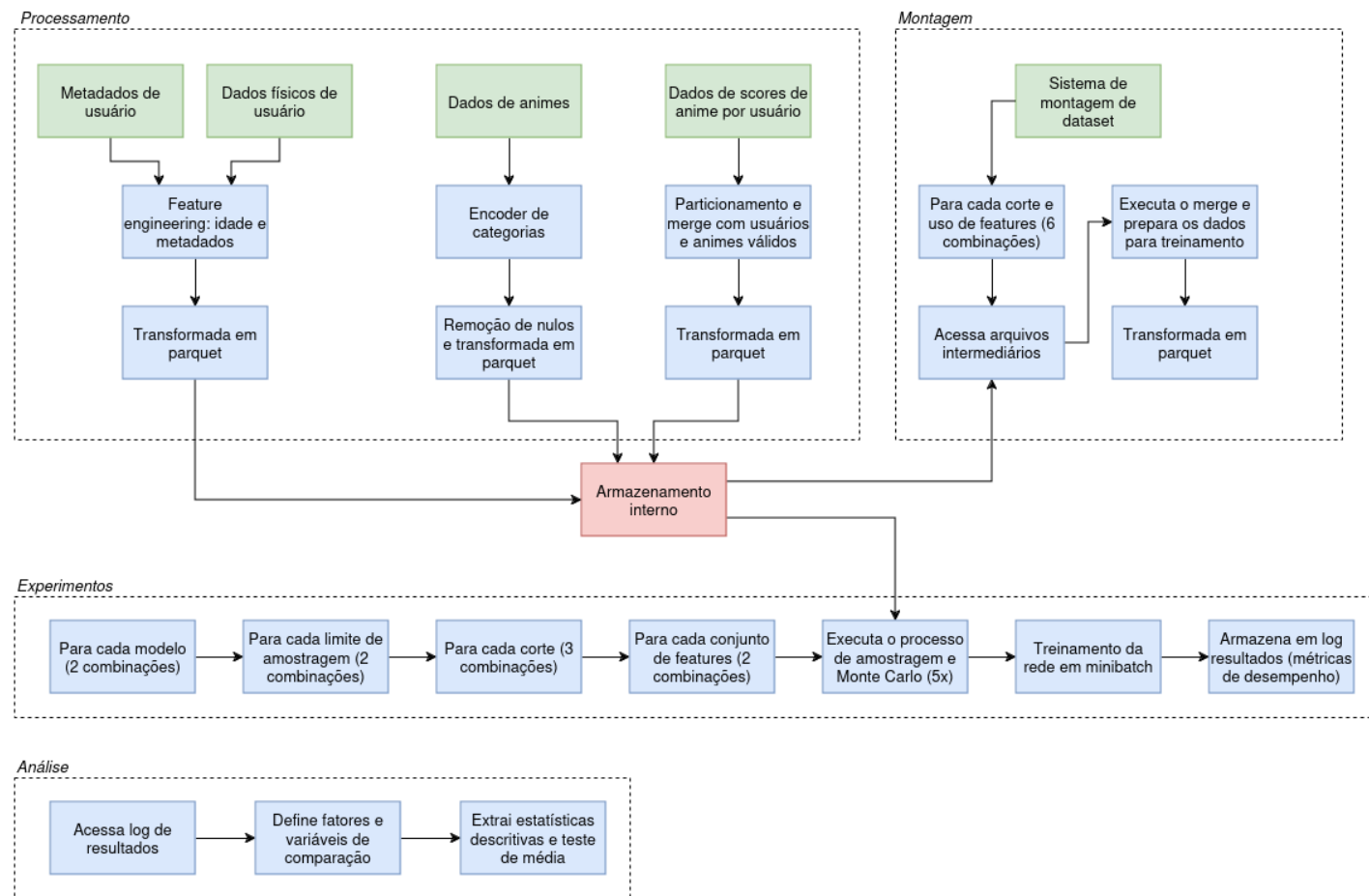
Saber se um anime vai ser bem recebido evita gasto com material ruim e reduz a quantidade de escolhas do usuário.

Desafios de modelagem

O disponível são os scores dados pelos usuários, logo é preciso decidir a regra de negócio mais adequada para converter a escala em binário.

Além dos dados físicos da animação (duração, gênero, material original e outros), os dados do usuário precisam ser considerados. Neste escopo, pode-se utilizar apenas dados físicos ou dados físicos acrescidos de metadados.

Fluxograma de execução



Resultados: metadados

	factor	basic	full	ttest-basic-full
0	arch=1,sample=0.1,cut=cut6	0.703058	0.709564	0.181639
1	arch=1,sample=0.1,cut=cut7	0.706680	0.704263	0.413272
2	arch=1,sample=0.1,cut=cut8	0.703752	0.708675	0.318417
3	arch=1,sample=0.2,cut=cut6	0.707050	0.704225	0.433863
4	arch=1,sample=0.2,cut=cut7	0.706703	0.704651	0.514483
5	arch=1,sample=0.2,cut=cut8	0.707439	0.707936	0.911095
6	arch=2,sample=0.1,cut=cut6	0.707292	0.700698	0.105080
7	arch=2,sample=0.1,cut=cut7	0.700951	0.705855	0.439372
8	arch=2,sample=0.1,cut=cut8	0.707420	0.703235	0.439435
9	arch=2,sample=0.2,cut=cut6	0.702501	0.706176	0.529896
10	arch=2,sample=0.2,cut=cut7	0.705210	0.709543	0.324600
11	arch=2,sample=0.2,cut=cut8	0.707960	0.705540	0.563122

Resultados: local de corte

	factor	cut6	cut7	cut8	ttest-cut6-cut7	ttest-cut6-cut8	ttest-cut7-cut8
0	arch=1,sample=0.1,features=basic	0.703058	0.706680	0.703752	0.424504	0.907202	0.562186
1	arch=1,sample=0.1,features=full	0.709564	0.704263	0.708675	0.116849	0.761126	0.104279
2	arch=1,sample=0.2,features=basic	0.707050	0.706703	0.707439	0.837969	0.910974	0.823197
3	arch=1,sample=0.2,features=full	0.704225	0.704651	0.707936	0.922783	0.419435	0.455492
4	arch=2,sample=0.1,features=basic	0.707292	0.700951	0.707420	0.185875	0.972284	0.214052
5	arch=2,sample=0.1,features=full	0.700698	0.705855	0.703235	0.375080	0.637364	0.688826
6	arch=2,sample=0.2,features=basic	0.702501	0.705210	0.707960	0.654094	0.344183	0.485336
7	arch=2,sample=0.2,features=full	0.706176	0.709543	0.705540	0.402114	0.884352	0.385400

Resultados: arquitetura

	factor	1	2	ttest-1-2
0	sample=0.1,features=basic,cut=cut6	0.703058	0.707292	0.354873
1	sample=0.1,features=basic,cut=cut7	0.706680	0.700951	0.226659
2	sample=0.1,features=basic,cut=cut8	0.703752	0.707420	0.502527
3	sample=0.1,features=full,cut=cut6	0.709564	0.700698	0.046537
4	sample=0.1,features=full,cut=cut7	0.704263	0.705855	0.758477
5	sample=0.1,features=full,cut=cut8	0.708675	0.703235	0.264764
6	sample=0.2,features=basic,cut=cut6	0.707050	0.702501	0.403423
7	sample=0.2,features=basic,cut=cut7	0.706703	0.705210	0.649358
8	sample=0.2,features=basic,cut=cut8	0.707439	0.707960	0.893757
9	sample=0.2,features=full,cut=cut6	0.704225	0.706176	0.643835
10	sample=0.2,features=full,cut=cut7	0.704651	0.709543	0.257530
11	sample=0.2,features=full,cut=cut8	0.707936	0.705540	0.610645

Resultados: amostragem

	factor	0.1	0.2	ttest-0.1-0.2
0	arch=1,features=basic,cut=cut6	0.703058	0.707050	0.348973
1	arch=1,features=basic,cut=cut7	0.706680	0.706703	0.992050
2	arch=1,features=basic,cut=cut8	0.703752	0.707439	0.509266
3	arch=1,features=full,cut=cut6	0.709564	0.704225	0.211191
4	arch=1,features=full,cut=cut7	0.704263	0.704651	0.912950
5	arch=1,features=full,cut=cut8	0.708675	0.707936	0.833433
6	arch=2,features=basic,cut=cut6	0.707292	0.702501	0.400205
7	arch=2,features=basic,cut=cut7	0.700951	0.705210	0.410161
8	arch=2,features=basic,cut=cut8	0.707420	0.707960	0.884968
9	arch=2,features=full,cut=cut6	0.700698	0.706176	0.198687
10	arch=2,features=full,cut=cut7	0.705855	0.709543	0.515265
11	arch=2,features=full,cut=cut8	0.703235	0.705540	0.682302

Repositório

GitHub: <https://github.com/BrunoSantosPK/annproject>