

Mineração de Dados - Trabalho prático 1

Luis Felipe de Deus¹, Nathanael Luchetta¹, Tiago Knorst¹, Yuri Oliveira¹

¹Curso de Engenharia de Computação – Universidade Federal de Santa Maria

{felipe.deus@ecomp.ufsm.br, nathanael.luchetta@outlook.com, tiago.knorst@ecomp.ufsm.br, yuri.alves@ecomp.ufsm.br}

1. Pré-processamento dos dados

Usou-se a ferramenta Excel para a limpeza dos dados, pois possuíam divergência na forma como o nome dos jogadores estava listado, evitando problemas de duplicidade dos dados.

A segunda etapa de pré-processamento é feita após a leitura do arquivo ao início do script em linguagem R. Os dados são transformados em “lista de presença”, onde jogadores presentes no jogo possuem booleano um (1) e ausentes zero (0), seguido de uma coluna com o resultado obtido pela dupla/trio.

2. Resultados

A mineração de dados foi feita a partir do algoritmo de associação Apriori, com objetivo de obter os jogadores, duplas e trios mais e menos vitoriosos, em um primeiro momento foi realizada a busca do jogador que mais teria vitórias, onde Ana destacou-se com 45 vitórias, inferindo que ela estaria na melhor dupla. Foi utilizado o comando subset a fim de descobrir o jogador que teria um alto nível de vitórias com a jogadora Ana, encontrando Ágata, que juntas possuíam 21 vitórias, com a mesma lógica foi realizada a busca do terceiro jogador mais vitorioso com esta dupla, encontrando Fábio, totalizando 4 vitórias, o resultado é descrito na Figura 1, onde são apresentados as consultas no R Studio.

Após encontrar os melhores, deu-se início a busca dos piores, onde Fábio destacou-se com 34 derrotas, e seguindo a mesma lógica, inferindo que Fábio estaria na pior dupla, deu-se início a outra busca onde Fábio tivesse jogado e perdido, obtendo o resultado de que Fábio e Bárbara possuíam 18 derrotas, e então, finalmente para descobrir o pior trio, buscou-se resultados onde Fábio e Bárbara jogaram e o resultado foi de derrota, para então encontrar o terceiro integrante da equipe, Ana, totalizando 4 derrotas com a dupla (Figura 2).

Para verificar a veracidade dos resultados obtidos, foi elaborado um segundo *data frame* contendo a seguinte codificação: 0-Não jogou, 1-Jogou e perdeu e 2-Jogou e ganhou. Os dados obtidos foram executados no software Weka, sendo possível a visualização por meio de gráficos individuais do desempenho dos jogadores, como é descrito na Figura 3.

Console	Terminal	Jobs
~/Desktop/DataMining/T1/ ↗		
[58] {agata=1,Ana=1,Barbara=0,Fabio=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.08000000 0.7058824 1.375095 12
[59] {agata=1,Ana=1,Fabio=0,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.08000000 0.7500000 1.461039 12
[60] {agata=1,Ana=1,Barbara=0,Emanuel=0,Fabio=0,Ricardo=0,Sheldon=0}	=> {Resultado=1}	0.08000000 0.7058824 1.375095 12
[61] {agata=1,Ana=1,Barbara=0,Emanuel=0,Fabio=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.08000000 0.7058824 1.375095 12
[62] {agata=1,Ana=1,Emanuel=0,Fabio=0,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.08000000 0.7500000 1.461039 12
[63] {agata=1,Ana=1,Barbara=0,Fabio=0,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.07333333 0.7333333 1.428571 11
[64] {agata=1,Ana=1,Barbara=0,Emanuel=0,Fabio=0,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.07333333 0.7333333 1.428571 11
[65] {agata=1,Ana=1,Fabio=1}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[66] {agata=1,Ana=1,Barbara=0,Fabio=1}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[67] {agata=1,Ana=1,Fabio=1,Sheldon=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[68] {agata=1,Ana=1,Fabio=1,Ricardo=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[69] {agata=1,Ana=1,Fabio=1,Shelda=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[70] {agata=1,Ana=1,Emanuel=0,Fabio=1}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[71] {agata=1,Ana=1,Barbara=0,Fabio=1,Sheldon=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[72] {agata=1,Ana=1,Barbara=0,Fabio=1,Ricardo=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[73] {agata=1,Ana=1,Barbara=0,Fabio=1,Shelda=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[74] {agata=1,Ana=1,Barbara=0,Emanuel=0,Fabio=1}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[75] {agata=1,Ana=1,Fabio=1,Ricardo=0,Sheldon=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[76] {agata=1,Ana=1,Fabio=1,Shelda=0,Sheldon=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4
[77] {agata=1,Ana=1,Emanuel=0,Fabio=1,Sheldon=0}	=> {Resultado=1}	0.02666667 0.8000000 1.558442 4

Figura 1 - Equipe com melhor desempenho.

Console	Terminal	Jobs
~/Desktop/DataMining/T1/ ↗		
[55] {Ana=0,Barbara=1,Emanuel=0,Fabio=1,Ricardo=0,Sheldon=0}	=> {Resultado=0}	0.08000000 0.6000000 1.2328767 12
[56] {agata=0,Ana=0,Barbara=1,Emanuel=0,Fabio=1,Shelda=0,Sheldon=0}	=> {Resultado=0}	0.08000000 0.6666667 1.3698630 12
[57] {Ana=0,Barbara=1,Fabio=1,Ricardo=0,Shelda=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[58] {agata=0,Ana=0,Barbara=1,Fabio=1,Ricardo=0,Shelda=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[59] {Ana=0,Barbara=1,Fabio=1,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[60] {Ana=0,Barbara=1,Emanuel=0,Fabio=1,Ricardo=0,Shelda=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[61] {agata=0,Ana=0,Barbara=1,Fabio=1,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[62] {agata=0,Ana=0,Barbara=1,Emanuel=0,Fabio=1,Ricardo=0,Shelda=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[63] {Ana=0,Barbara=1,Emanuel=0,Fabio=1,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[64] {agata=0,Ana=0,Barbara=1,Emanuel=0,Fabio=1,Ricardo=0,Shelda=0,Sheldon=0}	=> {Resultado=0}	0.06666667 0.6666667 1.3698630 10
[65] {Ana=1,Barbara=1,Fabio=1}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[66] {agata=0,Ana=1,Barbara=1,Fabio=1}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[67] {Ana=1,Barbara=1,Fabio=1,Sheldon=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[68] {Ana=1,Barbara=1,Fabio=1,Ricardo=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[69] {Ana=1,Barbara=1,Fabio=1,Shelda=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[70] {Ana=1,Barbara=1,Emanuel=0,Fabio=1}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[71] {agata=0,Ana=1,Barbara=1,Fabio=1,Sheldon=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[72] {agata=0,Ana=1,Barbara=1,Fabio=1,Ricardo=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4
[73] {agata=0,Ana=1,Barbara=1,Fabio=1,Shelda=0}	=> {Resultado=0}	0.02666667 0.4000000 0.8219178 4

Figura 2 -Equipe com pior desempenho.

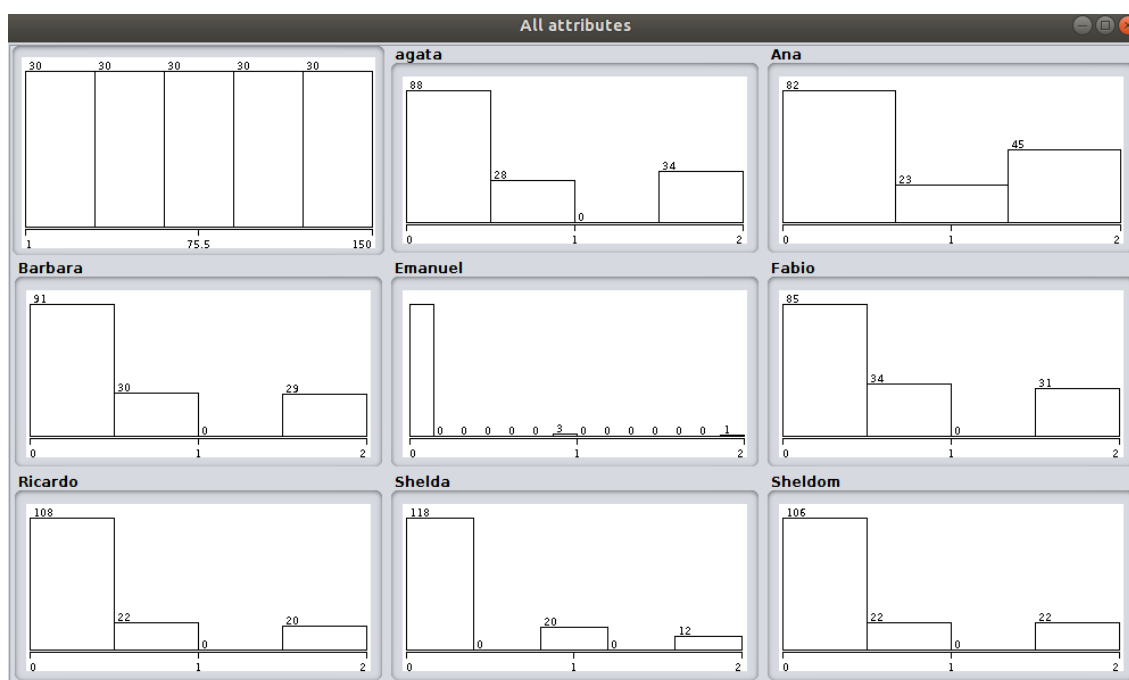


Figura 3 - Desempenho individual dos jogadores (Weka).

Ao final do processo obteve-se:

	Mais vitorioso	Menos vitorioso
Jogador	Ana	Fábio
Dupla	Ana e Ágata	Fábio e Bárbara
Trio	Ana, Ágata e Fábio	Fábio, Bárbara e Ana

Tabela 1 - Resultados obtidos.