

INF-0614 Visualização de Informação

Relatório 03

Dataset “wine-normalizado-versao-esparsa-para-heatmap-Tableau” e Bíblia em formato txt.

Integrantes do grupo:

Daniel Noriaki Kurosawa

Eric Uyemura Suda

Fernando Shigeru Wakabayashi

Objetivo

Análise exploratória e visualização dos dados presentes no dataset de características de vinhos e análise textual da Bíblia em txt.

Análises

A visualização do tableau encontra-se em:

https://public.tableau.com/views/T03_MDC010_INF0614/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

- HeatMaps

- **Eixo X** -> Features dos vinhos; **Eixo Y** -> Classe do vinho, **Cor** -> Média dos valores das features, **Texto** -> Média do valor das features.
- **Informações obtidas:** Para se analisar quais características diferenciam cada um dos vinhos, criamos um heatmap onde através da ordenação da intensidade das cores por classe, podemos identificar quais características mais se destacam entre as classes. Abaixo temos três imagens, cada uma ordenada por uma classe diferente.

A ordenação das características através da classe 3 apresentou o resultado mais interessante (Figura 3) pois podemos distinguir a terceira classe como tendo os menores valores médios de Flavonoides, OD280/OD315, Hue, Proanthocyanins e Total phenols; também possuem os valores médios mais altos para Color Intensity e Malic Acid.

Ordenando pela classe 1 (Figura 1), podemos observar que é a que tem a maior quantidade de Alcohol dentre as três classes e possui características contrárias ao vinho da classe 3, por apresentar os maiores valores de OD280/OD315, Total phenols, Proline e Flavonoids.

Já quando ordenamos a classe 2 (Figura 2), sem contar as Features Proline e Color intensity, todas as outras Features apresentam valores intermediários entre as classes 1 e 3, ordenando a classe 2 na segunda linha facilita a identificação das características intermediárias dessa classe.

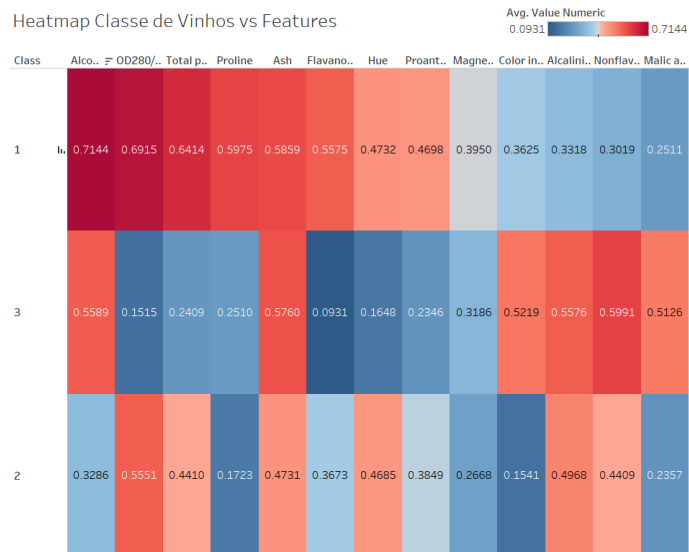


Figura 1 - Ordenação para Class igual a 1

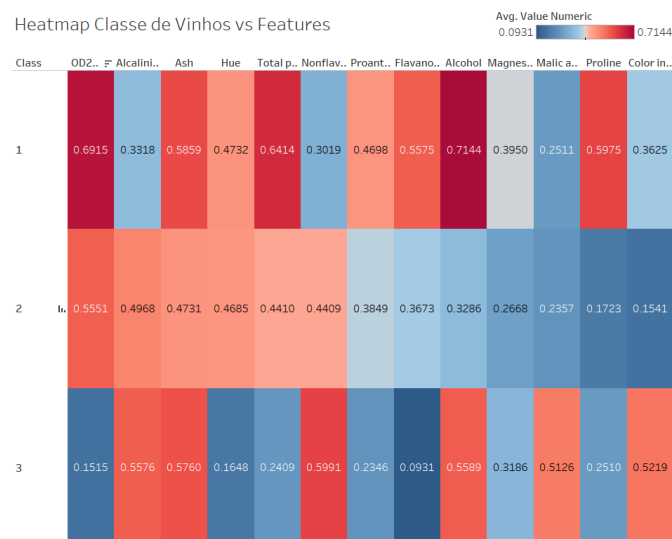


Figura 2 - Ordenação para a Class igual a 2

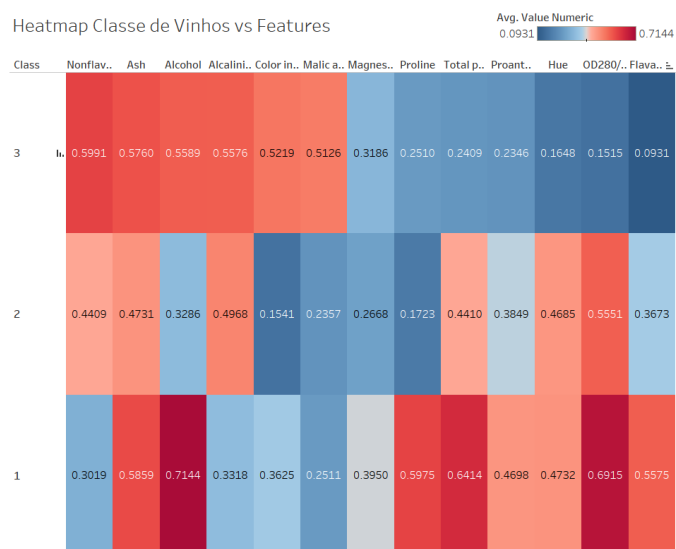


Figura 3 - Ordenação para Class igual a 3

- Projeção dimensional para identificar semelhança entre textos da bíblia:
 - **Cor** -> TF-IDF de uma das palavras utilizadas para busca de distâncias,
 - **Legenda** -> Livro de origem. Projeção espacial de distância não possui representação em escalas de eixo X ou Y.

- **Informações obtidas:**

Filtrando pela palavra Egito, percebemos que esta identifica melhor os livros do antigo testamento, o que era esperado pois estes tratam das origens do povo judeu. A exceção se deu no livro Levíticos, talvez devido a este focar mais nas leis e costumes.

Filtrando pela palavra Senhor, percebemos que esta palavra é demasiadamente comum, não servindo para identificar um livro em específico.

Filtrando pela palavra Jerusalém, percebemos que esta identifica principalmente o livro de Miquéias, devido a suas críticas a mesma e a suas profecias relacionadas a esta, seguido pelos livros do Novo testamento, o que também era esperado, pois a cidade de Jerusalém não estava consolidada nos livros selecionados do antigo testamento.

A Figura 7 apresenta as tabelas utilizadas como input para o programa Orange.

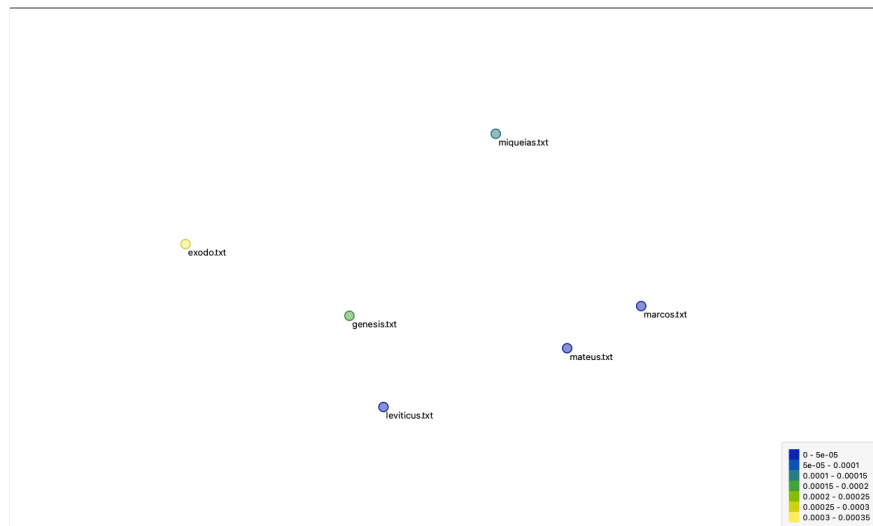


Figura 4 - MDS utilizando o TF-IDF para a palavra Egito.

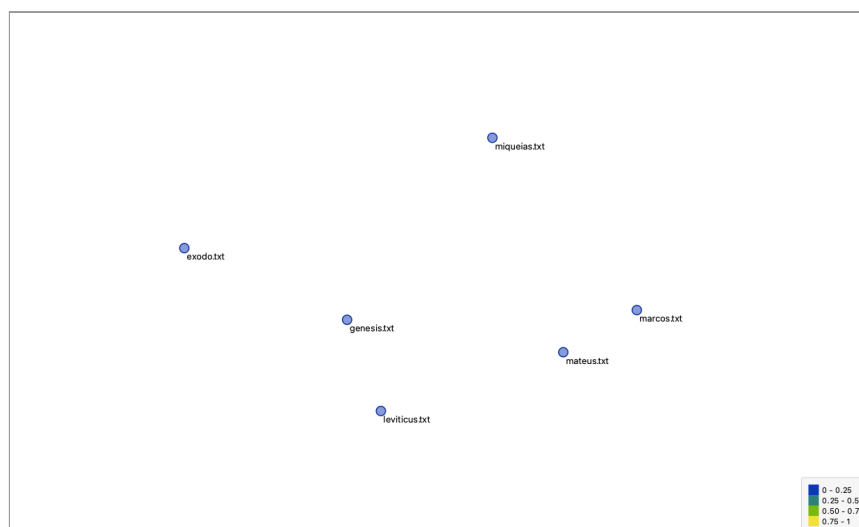


Figura 5 - MDS utilizando o TF-IDF para a palavra Senhor.

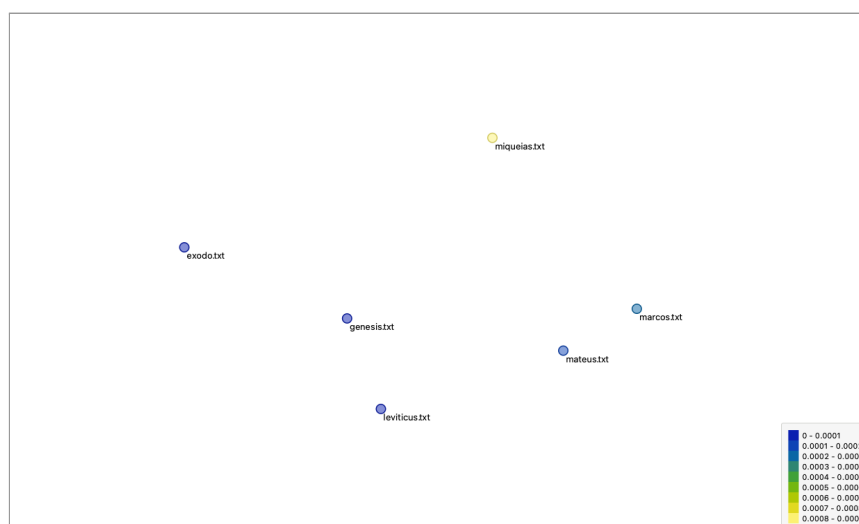


Figura 6 - MDS utilizando o TF-IDF para a palavra Jerusalém.

Frequência dos termos nos textos: Tf(w,d)

	Espírito	Deus	Jeová	Jerusalém	Senhor	Jesus	Egito
genesis.txt	0.0002961	0.0067215	0.0000000	0.0000000	0.0070176	0.0000000	0.0022800
exodo.txt	0.0001744	0.0046744	0.0004186	0.0000000	0.0140580	0.0000000	0.0041511
mateus.txt	0.0007978	0.0025430	0.0000000	0.0006482	0.0039890	0.0091748	0.0001995
marcos.txt	0.0015195	0.0040787	0.0000000	0.0008797	0.0015195	0.0105566	0.0000000
leviticus.txt	0.0000000	0.0024995	0.0000000	0.0000000	0.0150452	0.0000000	0.0004807
miqueias.txt	0.0010799	0.0046796	0.0000000	0.0028798	0.0151188	0.0000000	0.0014399

Ocorrência dos termos em cada documento : Df(w)

	Espírito	Deus	Jeová	Jerusalém	Senhor	Jesus	Egito
genesis.txt	1	1	0	0	1	0	1
exodo.txt	1	1	1	0	1	0	1
mateus.txt	1	1	0	1	1	1	1
marcos.txt	1	1	0	1	1	1	0
leviticus.txt	0	1	0	0	1	0	1
miqueias.txt	1	1	0	1	1	0	1
Df(w)	5	6	1	3	6	2	5

TfIdf(w,d)

	Espírito	Deus	Jeová	Jerusalém	Senhor	Jesus	Egito
genesis.txt	0.0000234	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0001805
exodo.txt	0.0000138	0.0000000	0.0003257	0.0000000	0.0000000	0.0000000	0.0003287
mateus.txt	0.0000632	0.0000000	0.0000000	0.0001951	0.0000000	0.0043775	0.0000158
marcos.txt	0.0001203	0.0000000	0.0000000	0.0002648	0.0000000	0.0050368	0.0000000
leviticus.txt	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000381
miqueias.txt	0.0000855	0.0000000	0.0000000	0.0008669	0.0000000	0.0000000	0.0001140

Figura 7- Tabelas geradas pelo arquivo excel “exemplo tfidf biblia.xlsx”.