

UNISINOS – Tópicos Especiais II

Exercício prático GA

Nomes: Franciele Fagundes, Gabriel Castro, Matheus Hahn, Savanna Denega

Ferramentas e recursos utilizados

- Software Jupyter;
- Biblioteca NLTK;
- Biblioteca TextBlob;
- Biblioteca Pandas;
- Biblioteca Matplotlib;

Referencias (sites, livros, artigos) utilizadas

- <https://textblob.readthedocs.io/en/dev/>
- https://planspace.org/20150607-textblob_sentiment/
- <https://towardsdatascience.com/having-fun-with-textblob-7e9eed783d3f>
- https://textblob.readthedocs.io/en/dev/api_reference.html#textblob.blob.Sentence
- <https://www.nltk.org>

Resultados obtidos com o experimento

- Análise de polaridade das sentenças do conjunto de dados (frases de review dos hotéis da Europa) com a biblioteca TextBlob e NLTK;
- Para a comparação do resultado da avaliação da ferramenta TextBlob, foi realizado uma análise manual de um subset do dataset utilizado;
- Comparação entre as frases, representando em uma tabela comum, com resultados da modelagem e da avaliação manual.

Possibilidades de trabalhos futuros com este experimento

- Realização de uma matriz de confusão para melhor visualização e entendimento sobre quais dados não obtiveram uma boa acurácia dentro da análise realizada;
- Maior uso das opções de limpeza dos dados (data cleaning) para deixar ainda mais modelado o dado;
- Uso de redes neurais para que o modelo se retroalimente e faça a modelagem mais refinada;
- Distribuição dos dados visualmente conforme a localização da review.

Aspectos do aprendizado pessoal durante o experimento

Aprendizado do uso das ferramentas/bibliotecas:

- Jupyter;
- NLTK;
- TextBlob;
- Pandas;
- Matplotlib.

Aprendizado de conceitos:

- Trabalhar com as fases de organização de dados e o conceito de data cleaning;
- Comparação de dados, podendo fazer a real comparação da modelagem com os dados reais da análise;
- Uso das bibliotecas NLTK e TextBlob para reconhecimento de sentimentos em frases, e seu processo para busca de resultados.