



4ª TP – 4ª Tarefa de Programação

Objetivo: Classificação de documentos e análise de sentimentos

Data de entrega: até o dia 10/09/2025 até às 20:50hs.

Membros: Individual ou em dupla.

1) Análise de sentimentos usando a biblioteca TextBlob

a) Leia e processe os tweets presentes no arquivo reforma_previdencia_rotulado.csv (por enquanto, ignore a coluna "Classificação"). Vocês devem pré-processar os tweets usando as técnicas que vocês já sabem: remover acentos, números e pontuação. Remova também URLs e trate os emojis.

b) Traduza cada um dos tweets tratados para o inglês usando a biblioteca Translator.

c) Aplique a biblioteca TextBlob para todos os tweets. Salve os valores fornecidos pela biblioteca (polaridade) em um novo arquivo .csv. Tal arquivo terá duas colunas: tweet tratado e o valor do sentimento calculado pelo TextBlob.

O que deve ser entregue: arquivos .py e o arquivo .csv gerado na letra c)

2) Concordância entre a classificação manual e a classificação automática

a) Considere que valores entre 0,5 e 1 sejam positivos, -1 e -0,5 sejam negativos e os outros sejam neutros. Converta os valores da planilha 1-c) para positivos, negativos e neutros.

b) Abra o arquivo reforma_previdencia_rotulado.csv e compare os valores que foram rotulados por especialistas com os valores calculados pela ferramenta. Conte quantos tweets negativos, positivos e neutros foram classificados de forma incorreta. Calcule a porcentagem de erro por classe (negativo, positivo e neutro).

O que deve ser entregue: um print da porcentagem de erro por classe.

3) Classificação de documentos usando aprendizado de máquina supervisionado

a) Já com os tweets tratados, crie um vocabulário e uma bag of words usando a ponderação TF-IDF (use o arquivo reforma_previdencia_rotulado.csv).



b) Treine um modelo de classificação de documentos usando a técnica de validação cruzada (cross-validation) e o algoritmo Naive Bayes Multinomial.

c) Exiba a acurácia e a matriz de confusão do modelo.

O que deve ser entregue: arquivos .py e um PDF com as medidas de desempenho e uma discussão sobre o desempenho do mesmo.

Referências:

-<https://www.linkedin.com/pulse/criando-um-classificador-de-sentimentos-em-python-usando-augusto/>

-<https://imasters.com.br/desenvolvimento/analise-de-sentimentos-aprenda-de-uma-vez-por-todas-como-funciona-utilizando-dados-do-twitter>

-<https://medium.com/@empowerpython/criando-um-classificador-para-processamento-de-linguagem-natural-8dc27f3642a1>

- <https://medium.com/luisfredgs/classificando-textos-com-machine-learning-e054ca7bf4e0>

-<https://medium.com/@empowerpython/criando-um-analisador-de-sentimentos-para-tweets-a53bae0c5147>