

Disciplina: Inteligência Artificial
Professora: Cristiane Neri Nobre
Data de entrega: 09/12/23
Valor: 4 pontos

Importante: para as duas questões abaixo, é necessário gerar um relatório explicando todas as etapas realizadas, com detalhes. Ou seja, não basta apenas submeter o link do código.

Questão 01 (2 pontos)

Nos endereços

1. <https://www.kaggle.com/code/leandrodoze/sentiment-analysis-in-portuguese/notebook>
2. <https://www.datacamp.com/tutorial/text-classification-python>
3. <https://reintech.io/blog/how-to-create-a-text-classification-model-with-python>

Vocês encontrarão alguns códigos que fazem o processamento e classificação de textos.

O que você precisa fazer:

- 1) Experimentar estes códigos e se basear neles para resolver o problema da base de dados disponível no CANVAS: **ReutersGrain-train.csv** e **ReutersGrain-test.csv**

Esta base de dados é de classificação e classifica se o texto fala sobre grãos ou não.

Desta forma, você deverá usar qualquer algoritmo de aprendizado de máquina visto no semestre para resolver este problema de classificação dos textos.

Questão 02 (2 pontos)

No endereço:

<https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data>

Você encontrará uma base de dados que faz análise de texto em 6 diferentes categorias: toxic, severe_toxic, obscene, threat, insult, identity_hate.

Ou seja, este problema é considerado de classificação multirótulo. Diferentemente de problemas de classificação multiclasse, na classificação multirótulo cada instância pode pertencer a mais de uma classe.

Resolva este problema de classificação e mostre o desempenho do classificador.

Para classificação multirótulo, investigue a biblioteca

Scikit-multilearn - <http://scikit.ml/>