**Plano de avaliação - Python IA**

**Conjunto de dados**

O conjunto de dados a ser utilizado nesta avaliação é uma sub-amostra de um conjunto de dados público disponível no Kaggle. Ele é composto por comentários da Wikipedia que foram rotulados em diferentes tipos de toxidade. O conjunto de dados foi adaptado para ter somente comentários tóxicos e não-tóxicos.

**Avaliação**

O objetivo deste projeto é criar um modelo em que classifica um comentário em tóxicos ou não-tóxicos. Serão avaliados diferentes aspectos neste projeto: tanto a capacidade de apresentar e interpretar informações, como o pré-processamento, treinamento e avaliação dos resultados.

**Etapa de experimentos com diferentes modelos (60 pontos)**

Devem ser realizados experimentos iniciais considerando diferentes métodos de classificação para determinar o modelo mais adequado para o problema em questão.

Questões que serão consideradas para compor a pontuação total:

• Pré-processamento dos dados

• Escolha e implementação do modelo

• Otimização de hiper-parâmetros

• Acurácia

• Escolha de métricas

**Script de treinamento (15 pontos)**

Deve-se implementar um script para a realização do treinamento do modelo que demonstrou o melhor desempenho na etapa de experimentos. O modelo, bem como o método de normalização (se usado), devem ser salvos.

**Script de classificação (15 pontos)**

Deve-se implementar um script para classificar novos textos (os novos textos estarão na mesma estrutura dos dados de treinamento, em um arquivo csv). O script deve carregar o método de normalização (se usado) e o modelo salvo após o treinamento. É necessário salvar um arquivo .csv com os textos e o resultado da classificação. Opcional: salvar csv de resultados.

**Apresentação (10 pontos)**

Deverá ser feita uma apresentação, mostrando o passo a passo da solução.

**Prazo máximo de envio: 24/08/2022**