

Definição das Classes



Coleta de Dados



Criação da Base de Dados Rotulada



Pré-processamento dos Dados



Treinamento dos Algoritmos de Classificação



Avaliação dos Classificadores

Execução de requisições para a API do Mercado Livre com o objetivo de obter os atributos que aparecem no maior número de categorias de produtos; definição dos 39 atributos mais importantes e genéricos para se tornarem as classes.

Importação de perguntas feitas por clientes reais do Mercado Livre através do banco de dados da GoBots; extração manual de mais perguntas de classes pouco presentes na base de dados; criação manual de mais perguntas para atingir um limiar de 15 perguntas/classe.

Criação e execução de um script Python que coleta uma pergunta aleatória das que foram extraídas do banco de dados da GoBots e mostra para o usuário rotular manualmente, de forma iterativa.

Definição das etapas que serão executadas antes do Treinamento dos Algoritmos de Classificação em si: para o classificador BERTimbau, apenas a tokenização com o seu Tokenizador; para o classificador DIETClassifier, outras etapas adicionais.

Treinamento dos Algoritmos de Classificação baseado em diferentes configurações de Pré-processamento dos Dados, diferentes hiperparâmetros de treino e diferentes quantidades de classes; definição sobre o uso ou não de Validação Cruzada.

Criação de uma função responsável pelo cálculo das métricas Acurácia, Precisão, Revocação e F1-Score, e também pela geração das Matrizes de Confusão; aplicação da ponderação sobre as métricas para conter o efeito do desbalanceamento das classes.