1. INTRODUÇÃO

Neste artigo, avaliamos a eficácia de algoritmos de aprendizagem de máquina na classificação de perguntas de acordo com seus assuntos.

A importância disso surge quando um usuário tenta recuperar informações em documentos, pois, dada a potencial grandiosidade da tarefa e o tédio de ter de analisar inúmeros possíveis documentos, delegar essa atividade a um algoritmo de inteligência artificial acaba sendo a melhor alternativa.

No TREC (Text Retrieval Conference), um número de perguntas e respostas tentam responder uma lista de perguntas predefinidas a partir do uso de um conjunto de documentos predeterminado.

1. MATERIAIS

Utilizaremos a base de dados organizada por Xin Li e Dan Roth, que foi utilizada em seu artigo Learning Question Classifiers (2002).

Usaremos três tipos de pré-processamento para cada tipo de algoritmo escolhido. Serão eles o BOW (Bag of Words), o TF (Term Frequency) e o TF-IDF (Term Frequency – Inverse Document Frequency).

1. MÉTODOS

Os algoritmos utilizados serão o KNN (K-Nearest Neighbors) e o Naive Bayes.

1. EXPERIMENTOS E RESULTADOS

4.1. KNN

|  |  |
| --- | --- |
|  | |
|  |  |
|  |  |
|  |  |
|  |  |

4.2. Naive Bayes

1. CONCLUSÃO