

Relatório de Andamento do Projeto

Introdução

Este projeto consiste em uma análise de dados que tem como objetivo prever se um projeto de lei proposto por um deputado estadual na câmara, será aprovado ou não; para isso, utilizar-se-á um meio de treinamento e *machine learning*, através do uso do Python e dos recursos disponíveis na biblioteca *scikit learn*.

Para fazer tal análise, será levado em conta algumas vertentes de dados, retirados dos dados abertos da câmara dos deputados, como por exemplo o partido do autor, sua unidade federativa e o governo vigente na época que o projeto de lei foi proposto, além da própria ementa do projeto (pode-se ver uma descrição completa dos dados no dicionário de dados) Essa análise será feita com dados desde que a constituição atual foi assinada, em 1988, analisando em duas vertentes: a primeira na qual a proposta é arquivada (não vira de fato uma lei) e a segunda na qual a proposta é transformada em norma jurídica.

Métodos

Como o número de propostas arquivadas é cerca de 10 vezes maior que o número de propostas transformadas em norma jurídica, foi feita uma planilha auxiliar para classificação, com cerca de 2600 projetos de lei, distribuídos igualmente entre arquivados e aprovados, como uma forma de não viciar o classificador, que no primeiro cenário ficaria um tanto quanto enviesado a arquivar um projeto de lei.

Primeiramente, realizou-se um classificador com a utilização de variados métodos de *machine learning*, como Naive Bayes, Decision Tree e Random Forest, medindo a acurácia de cada um deles, chegando ao máximo aos 73,26% de precisão, utilizando CountVectorizer e GradientBoostingClassifier, como vetorizador e classificador, respectivamente. Lembrando que foi usada apenas uma variável para estabelecer esse classificador, que foi a ementa propriamente dita.

Falhas observadas

Embora tenha uma precisão que pode ser considerada alta, levando em conta que o classificador foi construído baseado apenas na ementa, algumas falhas foram observadas, como por exemplo, a transformação em norma jurídica de qualquer proposta de lei que apresentasse a expressão “norma XV da constituição”.

Como melhorar?

Como próximos passos, deve-se colocar as demais variáveis para melhorar a precisão do classificador, como por exemplo o partido do candidato e o tamanho da sua bancada. A seguir, criar um classificador treinado manualmente para a separação de cada um dos projetos de lei em categorias diferentes, como políticas, sociais e econômicas, melhorando ainda mais a acurácia do classificador que se pretende construir.