

LÍNGUAS NATURAIS E MÁQUINAS ARTIFICIAIS: APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE TEXTO PARA A CLASSIFICAÇÃO DE SENTENÇAS JUDICIAIS BRASILEIRAS

Lucas Moreira Gomes

Auxiliar de Pesquisa do Subprograma Nacional de Pesquisa para o Desenvolvimento (PNPD) na Assessoria Técnica (Astec) do Ipea. *E-mail:* <lucasmoreira@lamfo.unb.br>.

Jader Martins Camboim de Sá

Auxiliar de Pesquisa do PNPD na Astec/Ipea. *E-mail:* <jader.martins@ipea.gov.br>.

Peng Yaohao

Pesquisador Visitante na Astec/Ipea e Assessor Especial na Secretaria de Política Econômica (SPE) da Secretaria Especial da Fazenda do Ministério da Economia. *E-mail:* <peng.yaohao@fazenda.gov.br>.

DOI: <http://dx.doi.org/10.38116/td2612>

Apresentamos neste artigo a aplicação de técnicas de inteligência artificial e de mineração de texto que possibilitaram, de maneira geral, ensinar ao computador como interpretar grande volume de textos da área do direito e como realizar uma tarefa específica desejada, sem a necessidade de conhecimento prévio da área jurídica pelos autores deste trabalho.

No experimento realizado, desenvolveu-se um modelo computacional para realizar a classificação de decisões judiciais quanto à procedência do pedido do autor da ação em quatro categorias: procedentes, improcedentes, parcialmente procedentes e acordo. Os resultados empíricos mostraram que o sistema desenvolvido foi capaz de classificar sentenças judiciais com boa acurácia, com quase nenhuma informação específica para cada categoria, partindo-se apenas de uma amostra pré-classificada. O conjunto de processos judiciais utilizado restringiu-se àqueles com o assunto de benefício de prestação continuada e aposentadoria rural tramitados na 2ª região da Justiça Federal.

Os dados utilizados foram provenientes do IpeaJus, banco de dados que o Ipea está consolidando, gradativamente, sobre processos judiciais e sobre a estrutura do sistema de justiça no Brasil. O IpeaJus contém, no momento, cerca de 30 milhões de registros relativos a processos judiciais e tem sido insumo para diversos estudos publicados pela instituição.

Detalhamos no artigo a operacionalização do processo de tratamento de dados textuais, a mineração de textos e o emprego da técnica de processamento de linguagem natural (NLP) para classificação não supervisionada. Ao analisar seus resultados, observamos que a

técnica de inteligência artificial empregada apresentou melhorias em relação a outras técnicas que não empregam inteligência artificial, obtendo índices médios superiores a 90% para todas as métricas consideradas.

Desse modo, demonstramos que a inteligência artificial, em geral, e o modelo proposto pelo presente estudo, em especial, permitem explorar ainda mais o valor latente, em termos de conhecimento, presente nos documentos judiciais brasileiros. O sistema proposto neste trabalho pode potencialmente ser refinado metodologicamente e estendido para aplicações mais generalistas, fornecendo índices e indicadores que podem ser utilizados posteriormente para, por exemplo, mensurar a eficiência jurídica no Brasil – o que, por sua vez, pode auxiliar na formulação de políticas públicas direcionadas para o aprimoramento do aparato judicial brasileiro.

SUMÁRIO EXECUTIVO