PROJETO VICTOR

COMO O USO DO APRENDIZADO DE MÁQUINA PODE AUXILIAR A MAIS ALTA CORTE BRASILEIRA A AUMENTAR A EFICIÊNCIA E A VELOCIDADE DE AVALIAÇÃO JUDICIAL DOS PROCESSOS JULGADOS.

por **Pedro Inazawa, Fabiano Hartmann , Teófilo de Campos, Nilton Silva e Fabricio Braz**

O aprendizado de máquina (ou, em inglês, Machine Learning) é uma área da Ciência da Computação que lida com algoritmos que aprendem por experiência e melhoram suas performances com o decorrer do tempo. Essa abordagem é normalmente utilizada para a detecção de padrões em dados, visando à automatização de tarefas complexas ou fazer predições, e vêm se tornando um diferencial em diversas áreas, inclusive no Direito. Nesse contexto, o projeto Victor, parceria entre o Supremo Tribunal Federal (STF) e a Universidade de Brasília, busca a aplicação dos mais novos conceitos e técnicas de Inteligência Artificial e Aprendizado de Máquina para necessidades relevantes em termos de processamento, classificação de peças e classificação de temas na gestão da Repercussão Geral no STF. Os objetivos são o aumento da celeridade de processamento, incremento da precisão e acurácia nas etapas envolvidas, de forma a apoiar os recursos humanos envolvidos nas atividades judiciárias.

Uma etapa importante para o entendimento completo da tecnologia por trás do Victor é a compreensão do campo da inteligência artificial conhecido como Processamento Natural de linguagem (ou, em inglês, Natural Language Processing ou NLP). Este subcampo da inteligência artificial teve seu começo em 1950 como uma intersecção da inteligência artificial e da linguística, inicialmente trabalhando em problemas relacionados à recuperação de informações em texto. Hoje o principal foco é gerar sistemas inteligentes que processem e compreendam a escrita e a fala como os seres humanos o fariam, a partir de metodologias estatísticas. O grande desafio é dar

O projeto Victor, parceria entre o Supremo Tribunal Federal (STF) e a Universidade de Brasília, busca a aplicação dos mais novos conceitos e técnicas de Inteligência Artificial e Aprendizado de Máquina.

a essas grandes, complexas e diversas fontes de informação uma forma estruturada de análise, e que se possam obter contextos, sentimentos, resumos textuais e categorização de conteúdo, dentre outros fatores de interesse.

O funcionamento do Victor no Supremo Tribunal Federal procede da seguinte forma: Inicialmente, o STF disponibiliza sua base de dados de processos jurídicos para que a equipe do Grupo de Aprendizado de Máquina (GPAM) da Universidade de Brasília [1] os processe. Atualmente, o banco de dados do projeto Victor conta com cerca de 952 mil documentos oriundos de cerca de 45 mil processos. Os arquivos são então submetidos a um fluxo de tratamento de documentos que:

1 - Filtra elementos considerados espúrios, como erros de digitalização e imagens;

2 - Divide frases em partes menores e cria símbolos para as partes mais relevantes do texto;

3 - Reduz palavras muito parecidas ou que possuem mesmo radical a símbolos comuns;

4 - Dá uma etiqueta a cada arquivo, classificando-o em uma das peças relevantes ao projeto;

5 - Atribui um rótulo com a repercussão geral do processo.

A partir desse processamento, modelos de NLP são aplicados aos dados visando determinar em qual repercussão geral tal processo se encaixa. Houve a produção também de dois subprodutos ao projeto que são relevantes ao tribunal: transformação de imagens em textos para posteriores buscas e edições e outro classificador capaz de determinar automaticamente se uma peça jurídica é Recurso Extraordinário, Agravo em Recurso Extraordinário, Sentença, Acórdão, Despacho ou outra categoria genérica de documentos. Espera-se que uma vez em produção, o Victor contribua na celeridade e qualidade do fluxo de análises de processos jurídicos, sendo uma solução adequada às necessidades dos servidores e operadores do Direito do Supremo Tribunal Federal.

A equipe do projeto é composta por um time de Direito e o Grupo de Pesquisa em Aprendizado de Máquina [1]. A metodologia de trabalho multidisciplinar dessas equi-

pes também fará parte da entrega final da pesquisa à comunidade acadêmica e ao STF, pois servirá para o desenvolvimento de outras ferramentas ou soluções. Como fruto desse trabalho, já houve algumas publicações em escopo jurídico e de tecnologia e um prêmio de Melhor Artigo em conferência [3]. O projeto também foi veiculado em grandes portais de mídia [2]. Finalmente, cabe ressaltar que esse é o primeiro projeto de inteligência artificial aplicada a tribunais no Brasil e o primeiro do mundo em uma Suprema Corte, desbravando, assim, caminhos para a inovação. ●

Referências

1. Grupo de Aprendizado de Máquina (GPAM). Projeto Victor. http://gpam.unb.br/victor/. [Online; Acessado em 19-Fev-2018].

2. Lydia Medeiros. Supremo do futuro. o Globo, Jun 2018.

3. Nilton Silva, Fabricio Braz, Teofilo Campos, Andre Guedes, Danilo Mendes, Davi Bezerra, Davi Gusmao, Felipe Chaves, Gabriel Ziegler, Lucas Horinouchi, Marcelo Ferreira, Pedro Inazawa, Victor Coelho, Ricardo Fernandes, Fabiano Peixoto, Mamede Maia Filho, Bernardo Sukiennik, Lahis Rosa, Roberta Silva, Taina Junquilho, and Gustavo Carvalho. Document type classification for Brazil's supreme court using a convolutional neural network. In Proceedings of The Tenth International Conference on Forensic Computer Science and Cyber Law. HTCIA, oct 2018.

PEDRO INAZAWA | É engenheiro eletrônico (2016) e mestrando em Engenharia Biomédica pela Universidade de Brasília. É pesquisador colaborador do projeto Victor.

FABIANO HARTMANN | É doutor em Direito pela UnB; professor da Faculdade de Direito e do Programa de Pós-Graduação da Faculdade de Direito - PPGD - UnB. É coordenador Acadêmico do Projeto Victor (PD Machine Learning – Supremo Tribunal Federal/Universidade de Brasília e líder do Grupo de Pesquisa/CNPq: DR.IA (Direito, Racionalidade e Inteligência Artificial).

TEÓFILO DE CAMPOS | É doutor pela Universidade de Oxford (2006), mestre pela USP (2001) e bacharel em Ciência da Computação pela UNESP (1998). De 2005 a 2016, trabalhou como pesquisador nos laboratórios da Sharp, Microsoft, Xerox, Surrey e Sheffield. Atualmente, é professor adjunto no Departamento de Ciência da Computação da UnB e bolsista PQ-CNPq.

NILTON SILVA | É coordenador do GPAM e pesquisador da UnB. Tem conduzido projetos em Inteligência Artificial, especialmente em Aprendizado de Máquina, Aprendizado Profundo e Processamento de Linguagem Natural, voltados à análise de grandes volumes de dados (Big Data): Imagens, Textos e Dados Corporativos.

FABRICIO BRAZ | É professor adjunto do curso de Engenharia de Software da UnB. Também é doutor em Engenharia Elétrica (2009) e tem atuado desde 2012 em pesquisa e desenvolvimento com aprendizado de máquina em imagem, texto e dados estruturados.