



Aplicação da Inteligência Artificial no sistema jurídico brasileiro: VICTOR e ALEI

Application of Artificial Intelligence in the Brazilian legal system: VICTOR and ALEI

NILTON CORREIA DA SILVA ^{1*}, FABRICIO ATAIDES BRAZ^{}, JOSE ROBERTO PIMENTA FERRETTI^{***}, CLENYS
REGES ROSARIO PEREIRA DE CASTRO ^{****}**

Clasificación COLCIENCIAS-PUBLINDEX: Artículo de Revisión

Recibido: 27/09/2019

Revisado: 4/11/2019

Aceptado: 28/11/2019

Resumo

Este artigo relaciona as principais leis brasileiras que estão impulsionando a gestão pública para uma mudança de paradigma e promovendo o uso de novas tecnologias como aliadas de agências governamentais. É mostrado como a Inteligência Artificial pode ser uma grande aliada para automatizar tarefas relacionadas ao fluxo de informações do sistema legal brasileiro. Dois projetos pioneiros de IA em cortes superior e de segunda instância do Brasil são apresentados: o projeto VICTOR, no Supremo Tribunal Federal (STF) e o ALEI, no Tribunal Regional Federal da Primeira Região (TRF1). Estes projetos evidenciam ganhos importantes em termos de economicidade, encurtamento de prazos de respostas, modernização e melhoramento de condições de trabalho na administração pública brasileira.

Palavras-chave: Inteligência Artificial, Sistema Jurídico, Aprendizado de Máquina, Automação.

Abstract

This article lists the main Brazilian laws that are pushing the public management toward a change of its paradigm and promoting the use of new technologies as government agencies allies. It is shown how the Artificial Intelligence can be a great ally to automate tasks related to the information flow of the Brazilian legal system. Two pioneering AI projects in higher courts and second instance courts in Brazil are presented: the VICTOR project at the Federal Supreme Court (STF) and the ALEI at the Federal Regional Court, First Region (TRF1). These projects show important gains in terms of economy, shortening response times, modernizing and improving the of working conditions in the Brazilian public administration.

Keywords: Artificial Intelligence, Legal System, Machine Learning, Automation

¹ *Professor Adjunto, Faculdade do Gama, Universidade de Brasília (UnB), Brasília, Brasil. Doutor em Geofísica pela Universidade de Brasília. Coordenador do Grupo de Pesquisa em Aprendizado de Máquina (GPAM), UnB, Brasília, Brasil. E-mail: niltoncs@unb.br

^{**}Professor Adjunto, Faculdade do Gama, Universidade de Brasília (UnB), Brasília, Brasil. Doutor em Engenharia Elétrica pela Universidade de Brasília. Coordenador do Grupo de Pesquisa em Aprendizado de Máquina (GPAM), UnB, Brasília, Brasil. E-mail: fabraz@unb.br

^{***}Supervisor de Inovação do Tribunal Regional Federal da 1ª Região, Professor de Gestão de Processos no Instituto de Direito Público Brasileiro (IDP), Doutor em Ciências da Informação pela Universidade de Brasília. Pesquisador do Grupo de Pesquisa em Administração de Justiça (Ajus) da Universidade de Brasília. E-mail: ferretti@trf1.jus.br

^{****}Diretora da Secretaria de Análise, Regularidade Processual e Jurisprudência. Bacharel em Direito e Pós-graduada em Direito Público pelo Centro de Ensino Superior de Catalão em Goiás. E-mail: clenys.castro@trf1.jus.br

Cómo citar: Correia da Silva, N., Ataides Braz, F., Pimenta Ferretti, J., y Pereira de Castro, C.(2019). Aplicação da Inteligência Artificial no sistema jurídico brasileiro: VICTOR e ALEI

Revista de Administración Pública del GLAP, 3(4), páginas 45-55.

Introdução

Hoje, no governo federal brasileiro, é muito comum encontrar organizações que tentam melhorar seu desempenho, considerando, na maioria dos casos, soluções digitais e algumas aplicações tecnológicas.

Não é difícil detectar atos normativos relacionados à governança digital. Somente para citar alguns (ordem cronológica), entre atos da Presidência da República e leis em consonância com o Congresso Nacional, foram selecionados:

1. *Eficiência organizacional*, Decreto nº 9.739 / 2019, que visa à administração pública federal ampliar e tratar o arranjo institucional do Sistema de Inovação do Governo Federal (Decreto 9.739, 2019).
2. *Rede Nacional de Governo Digital*, Decreto nº 9.584 / 2018, para promover a colaboração, o intercâmbio e a criação de iniciativas inovadoras relacionadas ao governo digital no setor público (Decreto 9.584, 2018).
3. *Transformação digital*, Decreto nº 9.319 / 2018 que institui o Sistema Nacional para a Transformação Digital e estabelece a estrutura de governança para a implantação da Estratégia Brasileira para a Transformação Digital (Decreto 9.319, 2018).
4. *Responsividade*, Decreto nº 9.203 / 2017 que dispõe sobre a política de governança da administração pública federal (Decreto 9.203, 2017).
5. *Remoção de formalidades e soluções tecnológicas para simplificar o processo cidadão-cliente* (Lei 13.460, 2017).
6. *Incentivo ao desenvolvimento científico, pesquisa, capacitação e inovação científica e tecnológica* (“Lei de Inovação”) (Lei 13.243, 2016).
7. *Política de governança digital*, Decreto nº 8.638 / 2016, que institui a Política de Governança Digital no âmbito da Administração Pública Federal (Decreto 8.638, 2016).

Do lado do Judiciário, o Conselho Nacional de Justiça (CNJ) parece decidir tomar uma direção diferente e não aceitar que seus servidores ainda estejam produzindo decisões judiciais artesanais.

O orçamento em tecnologia cresceu e, em comparação com outros países, conforme mostrado pela Figura 1, o Judiciário brasileiro investiu uma quantidade expressiva de recursos financeiros. No entanto, apesar de o Brasil ter investido R\$ 2,2 bilhões (cerca de US\$ 550 Milhões) em Tecnologia da Informação, o número de ações judiciais em tramitação é de aproximadamente de 80 milhões (CNJ, 2018).

Em 2019 o CNJ instituiu o Laboratório de Inovação (LabInov) e o Centro de Inteligência Artificial (CIA), ambos para o desenvolvimento da Plataforma Judicial Eletrônica (PJe). O CNJ pretende ter essa plataforma como única em todos os tribunais brasileiros (Decreto 25, 2019).

Nas *Premissas e diretrizes para o funcionamento do Centro de Inteligência Artificial aplicado ao PJe*, anexo à norma mencionada, em sua Introdução, o Conselho explica:

A proposta de criar um espaço para pensar, pesquisar e produzir inovação para o processo judicial eletrônico decorre especialmente da necessidade de o Judiciário brasileiro oferecer uma resposta adequada a todos aqueles que buscam serviços de justiça. Assim, como em outros aspectos de suas vidas cotidianas, pessoas e organizações foram atingidas pelo uso da tecnologia da informação e **o ambiente judiciário não pode deixar de buscar novos padrões, aproveitando as melhores soluções possíveis e o seu próprio potencial. Magistrados e funcionários públicos.** (...) A primeira ferramenta disponível para esse ambiente virtual é direcionada ao **uso de IA. Uma solução tecnológica que possibilita pesquisa e produção de serviços inteligentes para auxiliar na construção de módulos para o PJe e seu aprimoramento** (grifo nosso) (Decreto 25, 2019).

Em relação à Justiça Federal, existe outro ator, o Conselho da Justiça Federal (CJF). Cabe a esse Conselho a supervisão administrativa e orçamentária, com poderes correccionais, cujas decisões são de observância obrigatória por todas as unidades da Justiça Federal de primeiro e segundo graus.

Diante da movimentação dos outros entes do Judiciário, o CJF emitiu resolução, cujo Artigo 3º, VII, informa: “aplicação de soluções tecnológicas que visem simplificar os processos e procedimentos de prestação de serviços judiciais, adminis-

trativos e públicos e proporcionar melhores condições para o compartilhamento de informações (grifo nosso)” (Resolução 9.203, 2019).

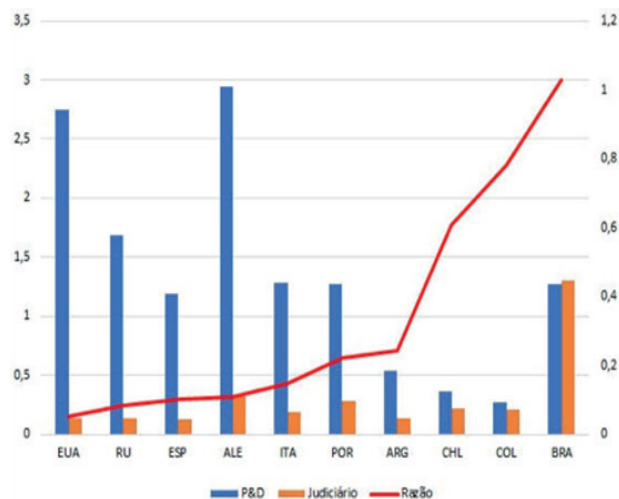


Figura 1. Percentual do PIB investido em P&D, no Judiciário e a razão.

Fonte: Adaptado de (Da Ros, 2015).

A situação de morosidade e o congestionamento que caracterizam a estrutura judicial geram inevitável aumento de seus custos. Os avanços ocorridos em Tecnologia da Informação (TI) não têm sido capazes de superar os gargalos existentes e o constante congestionamento processual.

O mercado, há algum tempo, já vem oferecendo soluções ainda não incorporadas à realidade judicial. Entre outras, citam-se, o *Hackaton* (competições tecnológicas, geralmente com duração de um dia onde equipes concorrem para apresentar a melhor solução a um determinado problema); aplicações de Inteligência Artificial (IA) com o fim de aperfeiçoar os processos de trabalho com emprego de tecnologia sofisticada. Entre estas, destaca-se a aprendizagem de máquina (AM), isto é, “ensinar” o computador a distinguir as decisões certas das erradas. A partir dos seus próprios erros, a máquina “aprende” e, dessa forma, a cada repetição da mesma pro-

blemática, o computador tende a apresentar uma performance melhor.

Entretanto, as soluções oferecidas pelo mercado impõem muito risco para o gestor interessado na contratação, visto que o dado necessário para a tarefa alvo de aprendizado encontra-se em documentos pdf, tendo eles texto embarcado ou, no pior caso, somente imagem². Esse fato, eleva o grau de desconhecimento acerca do sucesso em se solucionar a tarefa, uma vez que há um vale de distância entre o dado disponível e o dado necessário para se iniciar o ciclo da AM, em tempo de contratação.

A realidade do dado, agravada pela inexperiência por parte do tribunal em executar projetos dessa natureza, tornam a associação entre o tribunal e uma instituição federal de ensino e pesquisa, ambos membros do poder público federal, uma saída apropriada, atenuando-se os riscos contratuais, redução de custos e acesso a recursos humanos altamente especializados na matéria em questão.

Este artigo apresenta como se deu o projeto VICTOR, uma parceria entre a Universidade de Brasília e o Supremo Tribunal Federal (STF), que buscou solucionar dois gargalos no fluxo do processo judicial a saber, por meio de AM:

- Classificação de temas de repercussão geral (RG)³.
- Classificação de peças jurídicas.

Além disso, o artigo traz, também, uma visão geral de projeto de pesquisa e desenvolvimento análogo ao do STF, em execução no Tribunal Regional Federal da 1ª Região (TRF1).

O Projeto VICTOR no STF

Do ponto de vista de volume de dados, o STF possui um cenário desafiador. Essa corte recebe aproximadamente 70 mil novos processos por ano para serem analisados e julgados, gerando

² Quando o pdf está no formato imagem, é necessário OCR (Reconhecimento Ótico de Caracteres) para se obter o texto da imagem da página do documento.

³ Instituto processual pelo qual se reserva ao STF o julgamento de temas trazidos em recursos extraordinários que apresentem questões relevantes sob o aspecto econômico, político, social ou jurídico e que ultrapassem os interesses subjetivos da causa. Foi incluído no ordenamento jurídico pela Emenda Constitucional n. 45/2004 e regulamentado pelos arts. 322 a 329 do Regimento Interno do Supremo Tribunal Federal e pelos arts. 1.035 a 1.041 do Código de Processo Civil (Lei n. 13.105/2015). Fonte: <http://www.stf.jus.br/portal/cms/verTexto.asp?servico=estatistica&pagina=entendarg>, Acesso em: 18/11/2019

um estoque de centena de milhares de processos. Esses processos podem ser de diferentes tamanhos e trazem em suas páginas dados textuais e/ou imagens.

A dimensionalidade deste dado também provoca um grande desafio para qualquer tecnologia que vise extrair informações. Há mais de 1 milhão de diferentes termos entre palavras da língua portuguesa, termos jurídicos, códigos jurídicos e outros. Há ainda mais de mil temas de RG para serem analisados com vistas aos processos que adentram o Tribunal.

Realizar o reconhecimento de padrões na base de dados processuais do STF é um problema que agrega uma grande massa de dados não estruturados (são cerca de 350 novos processos por dia) entre textos e imagens de documentos.

Este foi o cenário que motivou o STF apostar no uso de IA no desenvolvimento de uma tecnologia de automação de tarefas jurídicas internas. Assim, o VICTOR, projeto de pesquisa e desenvolvimento (P&D), teve seu início nos primeiros meses de 2018 envolvendo uma equipe multidisciplinar composta por três principais entida-

des em sua execução: colaboradores do próprio tribunal (STF) e duas unidades da UnB representadas pelo GPAM, vinculado a Faculdade do Gama (FGA), e pela Faculdade de Direito (FD).

O principal objetivo do projeto foi o de aplicar métodos de AM para resolver um problema de classificação de textos de processos judiciais que chegam ao STF. A tarefa envolve a classificação (vinculação) de processos em temas de RG.

Os temas versam sobre diferentes assuntos de interesse nacional dos brasileiros, conforme pode ser verificado na Tabela 1, que explica o teor da ementa de alguns deles.

Ao longo do trabalho de investigação sobre classificação de processos em temas de RG, identificou-se que a prática de separação de peças dentro de um processo seria de grande valor para o objetivo final de classificação dos feitos dessa natureza. Esta constatação veio da observação do trabalho dos analistas judiciários. Esses servidores realizavam leituras personalizadas em peças específicas do processo, para em seguida concluir qual tema de RG tal processo estaria relacionado.

Tabela 1. Detalhes de alguns temas de RG classificados pelo VICTOR

Número	Tema	Explicação
5	Compensação da diferença de 11,98%, resultante da conversão em URV dos valores em cruzeiros reais, com o reajuste ocorrido na data-base subsequente.	O enunciado preceitua que há repercussão geral nos Recursos que discutem reposição geral das diferenças salariais de 11,98% decorrentes da implantação do Programa de Estabilização Econômica em 1994, que determinou, à época, a regra de conversão de salários em Unidade Real de Valor – URV, com o reajuste ocorrido na data-base subsequente.
6	Dever do Estado de fornecer medicamento de alto custo a portador de doença grave que não possui condições financeiras para comprá-lo.	O tema 006 refere-se à obrigação do SUS (Sistema Único de Saúde) de dar atendimento médico a todas as pessoas. Diante da falta de recursos, o Poder Judiciário tem de decidir se o Estado pagará o tratamento caro para atender uma única pessoa ou se, com esse dinheiro, paga o tratamento de muita mais gente que depende tratamentos mais baratos.
20	Alcance da expressão “folha de salários”, para fins de instituição de contribuição social sobre o total das remunerações.	A contribuição social a cargo do empregador incide sobre ganhos habituais do empregado, quer anteriores ou posteriores à Emenda Constitucional nº 20/1998.
109	Adoção pelo Poder Judiciário de critérios normativos estaduais como fundamento para extinguir ações de execução fiscal ajuizadas pelo Município.	O Supremo reconheceu a repercussão geral dos casos de extinção de execuções fiscais municipais com base em legislação estadual, pelo princípio da repartição de poderes e pela competência tributária prevista na constituição.
188	Declaração de hipossuficiência para obtenção de gratuidade de justiça.	O STF entende que não há repercussão geral nos casos em que a parte questiona a declaração de hipossuficiência (insuficiência de renda para pagar as custas e despesas com o processo/pobreza) para conseguir a gratuidade na justiça.
460	Prosseguimento de Execução Fiscal lastreada em Certidão de Dívida Ativa supostamente ilíquida.	Não apresenta repercussão geral recurso extraordinário que, tendo por objeto possibilidade de se prosseguir em execução fiscal lastreada em Certidão de Dívida Ativa, supostamente ilíquida, por cobrar, entre outras, taxa declarada inconstitucional pelo Supremo Tribunal Federal, versa sobre tema infraconstitucional.

Fonte: Própria.

Assim, ficou caracterizada uma outra meta do projeto, a identificação de cinco tipos de peças dentro de um processo: Acórdão, Recurso Extraordinário (RE), Agravo de Recurso Extraordinário (ARE), Decisão de Admissibilidade e Sentença.

O VICTOR⁴, denominação dada a solução resultante do projeto, visa à aplicação dos mais novos conceitos e técnicas de IA e AM para necessidades relevantes em termos de processamento, classificação de peças e classificação de temas na gestão da RG no STF (TOLEDO, 2018). Os objetivos são o aumento da celeridade de processamento e o incremento da precisão e acurácia nas etapas envolvidas, de forma a apoiar os recursos humanos envolvidos nas atividades judiciárias (Inazawa et. al., 2019).

O projeto foi executado observando três etapas: extração de texto, modelagem e inferência. Cada uma delas teve orientação metodológica específica, observando a natureza de suas atividades principais.

A primeira etapa, denominada “extração de texto”, assumiu o desafio de transformação de processos em arquivos pdf em texto, o disponibilizando em base de dados com estrutura para reconhecer as referências de cada texto, uma vez que ele é reconhecido em determinada página, de um arquivo, de um processo.

Todas essas relações precisaram ser preservadas e disponibilizadas para as etapas subsequentes. Cumpre ressaltar, que essas atividades são eminentemente relacionadas a desenvolvimento de software. Felizmente, existem várias abordagens que se propõem a orientar ações dessa natureza. A coordenação julgou ser mais adequado às circunstâncias do projeto o uso de método ágil, cujos detalhes são apresentados posteriormente.

A segunda etapa, denominada “modelagem” possuiu o caráter de pesquisa no projeto, uma vez que se esperava por meio dela validar as hipóteses a seguir:

1. É possível construir um modelo baseado em AM que, recebendo como insumo o corpo textual do processo judicial, consiga determi-

nar a RG associada, em desempenho próximo ao do ser humano?

2. É possível construir um modelo baseado em AM que, recebendo como insumo o corpo textual de uma página de um processo judicial, consiga determinar o seu tipo de peça processual, com desempenho próximo ao do ser humano?

A pesquisa para validar ambas, as hipóteses, foi conduzida segundo os preceitos do método científico descrito pelo *Common Task Framework (CTF)*, adaptado a realidade do projeto, apresentado posteriormente.

A última etapa, denominada “inferência”, assumiu a responsabilidade de embarcar os modelos resultantes da etapa anterior em componentes de software. Assim como observado na primeira etapa, a natureza desta se tratou de desenvolvimento de software e, por consequência, fez uso da mesma metodologia ágil mencionada anteriormente, cujos detalhes são apresentados na seção a seguir.

Método Ágil

Os métodos ágeis de desenvolvimento de software estão em franco uso pelo mercado e academia há mais de uma década e, cujos resultados, revelam o seu elevado valor para a Engenharia de Software e para amplificar a capacidade da sociedade em enfrentar os desafios inerentes a construção de produtos de software, quer sejam eles para a indústria, governo ou academia (Meyer, 2018).

Da perspectiva do projeto VICTOR, os seguintes aspectos e orientações do processo ágil foram observados para conduzir as atividades (Williams, 2003):

1. Organização das atividades em blocos a serem realizados em períodos denominados *sprint*. Tais períodos assumiram a duração de uma semana.
2. Planejamento semanal das *sprints* e consequente execução do planejado, com rigoroso controle de mudança após o planejamento e priorização de atividades.

⁴ <http://gpam.unb.br/victor/>

3. Reporte diário em reuniões em que o membro da equipe expusesse o que trabalhou no dia anterior, suas dificuldades e o plano para o dia posterior, sob a luz do planejamento.
4. Estabelecimento do papel do SCRUM⁵ máster, responsável por orquestrar as atividades semanais e conduzir os eventos recorrentes, como reuniões e planejamentos.
5. Controle de configuração e versionamento dos artefatos de software por meio da plataforma git.
6. Integração contínua dos artefatos gerados pelos membros da equipe ao longo do projeto.
7. Uso de quadro Kanban⁶ para acompanhar a evolução das demandas.

A partir dessas orientações o trabalho de desenvolvimento pode ser efetivamente gerenciado.

Common Task Framework (CTF)

O modelo CTF foi descrito por Donoho (2015) como o sendo o segredo do sucesso na modelagem preditiva. Os ingredientes do CTF envolvem:

- a. Uma base de dados que contenha para cada observação, uma lista de medidas de características e rótulos para a classe de cada observação.
- b. Um conjunto de competidores, cuja tarefa comum (*common task*) seja inferir a regra de predição da classe, a partir dos dados de treinamento.
- c. Um avaliador de desempenho centralizado, para o qual os competidores submetem suas regras de predição. O avaliador verifica as regras concebidas pelo competidor, tendo como base um banco de dados de teste, desconhecido pelo competidor, devolvendo apenas o resultado do desempenho do seu modelo (Donoho, 2015).

Obviamente que as regras do CTF, como já foi antecipado, precisaram ser adaptadas para a realidade do projeto. Neste caso, tivemos os ajustes discriminados, segundo os mesmos itens, a saber:

- a. Como resultado da primeira etapa do projeto, além de extrair os textos dos pdfs, o dado armazenado em base dados reteve as classes, segundo rotulação providenciada pelos analistas judiciários do STF. O trabalho foi, então, de promover uma segmentação do dado em treino, teste e validação, observando a proporção 70%, 20%, e 10%. Essa abordagem foi executada tanto nos dados voltados para a classificação de temas de RG, bem como nos dados para a tarefa de classificação de peças jurídicas.
- b. Os competidores foram representados pelos membros do projeto, que foram segregados em equipes menores, com o objetivo de resolver as tarefas de classificação.
- c. A maior adaptação, de fato, aconteceu no processo de aferição de resultado, uma vez que todo o dado estava disponível para a equipe do projeto, apesar de se considerar apenas os resultados dos modelos de predição, segundo seu desempenho frente ao dado de teste.

Assim, o trabalho de modelagem foi conduzido conforme as orientações descritas nos três pontos descritos anteriormente.

Resultados

A etapa inicial, como já foi relatado, referiu-se à extração de texto de cada página do processo judicial. Neste caso, o VICTOR avaliou a qualidade do texto e, em sendo satisfatória, o conteúdo textual da página é armazenado no banco de dados. Caso negativo, a página é transformada em imagem e enviada para a função de OCR (Smith, 2007), conforme ilustrado pela Figura 2.

5 O SCRUM é uma metodologia de desenvolvimento ágil, cujas algumas práticas foram adotadas no projeto.

6 O KANBAN é uma estratégia para otimização de fluxo de valor ao longo do processo por meio de um sistema visual, que evidencia o trabalho em progresso. Fonte: <https://www.scrum.org/resources/kanban-guide-scrum-teams>

Ainda fazendo uso do fluxo apresentado na Figura 2, caso o conteúdo da página seja imagem, o VICTOR a encaminha direto para a função de OCR. Esta função recebe uma imagem como entrada e produz um texto, que é enviado para ser armazenado no banco de dados de processos.

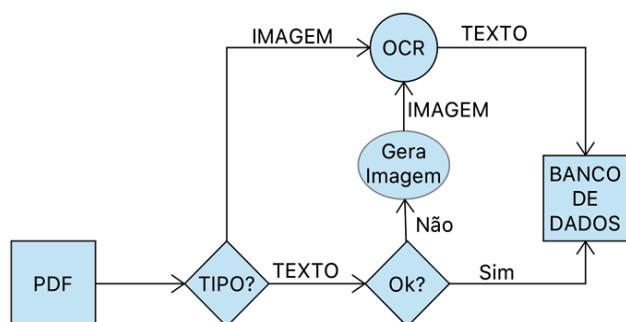


Figura 2. VICTOR: Transformação de dados jurídicos-judiciais.

Fonte: Própria

Para validar as hipóteses descritas na Seção 2, foram realizados vários experimentos orientados segundo a CTF, envolvendo uma variedade ampla de algoritmos áreas de PLN, Sistemas Especialistas e Visão Computacional. Esse conjunto de técnicas e algoritmos pode ser apreciado na Tabela 2.

Dentre os experimentos que apresentaram melhor resultado de modelagem para a tarefa de classificação de temas de RG, houve o XGBOOST, tendo como descritor o TFI-DF, considerando-se os 29 temas apresentados pela Tabela 3, os quais representam mais de 50% do acervo total do STF. Neste caso, o desempenho médio em F1-Score foi acima de 88%. Esse desempenho mostrou-se próximo da capacidade dos analistas judiciários. Confirmando, então a hipótese de pesquisa para classificação de RG.

Tabela 2. Modelos de IA investigados durante desenvolvimento do VICTOR.

Tipo	Nome	Descrição
Pré-processamento	Vocabulário	Indexação de palavras e termos
	Chi-quadrado	Teste estatístico para dados categóricos (Forman, 2003)
	Regex	Expressões Regulares
Descritores	TF-IDF	Medida estatística que tem o intuito de indicar a importância de uma palavra em um documento (Wu et al., 2008)
	Bag of Words	Representação indexada de palavras
	Embedding	Técnica de projetar termos em um outro espaço mais conveniente (Perozzi et al, 2014)
Redes Neurais Artificiais Profundas	CNN	Redes Neurais Convolucionais (Conneau et al., 2016)
	LSTM	Tipo de rede neural recorrente usada para processar dados sequenciais (Hochreiter e Schmidhuber, 1997)
	Bi-LSTM	Redes neurais recorrentes que considera o tanto o estado direto, bem como reverso. (Graves e Schmidhuber, 2005)
Modelos de Aprendizado de Máquina Clássicos	Naive Bayes	Classificador supervisionado probabilístico (Rennie et al, 2003)
	KNN	Classificador supervisionado baseado em similaridade de vizinhança (Muja e Lowe, 2009)
	XGBOOST	Classificador em arvore escalável (Chen e Guestrin, 2016)

Fonte: Própria.

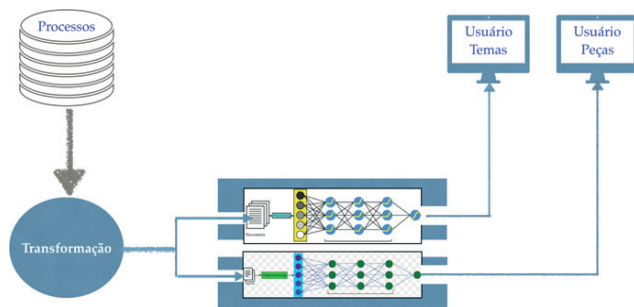
Tabela 3. Números dos temas de RG classificados pelo VICTOR.

5	163	406	597	766
6	232	409	634	773
26	313	424	660	793
33	339	555	695	800
139	350	589	729	810
852	895	951	975	-

Fonte: Própria.

Já com relação a tarefa de classificação de peças jurídicas, o experimento que trouxe o melhor resultado foi o BI-LSTM, usando também como descritor o TF-IDF. Neste caso, o desempenho médio em TF-IDF superou os 90%, considerando F1-Score como métrica (Silva, et. al., 2018, Braz, et. al., 2018). Fato que confirma a hipótese de pesquisa de classificação de peças jurídicas.

Como última etapa, considerando o sucesso dos modelos obtidos pela pesquisa, restou a equipe do projeto encapsular os modelos em componentes que pudessem ser usados para inferência dentro do STF.

**Figura 3.** Arquitetura geral do VICTOR (Silva, 2018).

Como ilustrado pela Figura 4, temos dois componentes de AI na solução, sendo eles o motor de inferência de peças, com capacidade de discernir dentre as 6 peças supracitadas; e o motor de inferência de RG, com capacidade de discernir dentre as 30 possibilidades de RG, incluindo o rótulo sem RG.

Para finalizar, cumpre destacar as estimativas dos gestores da Secretaria de Tecnologia da Informação do STF, o VICTOR deve prover um

aumento de performance geral em todo o fluxo do processo judicial bem como gerar uma economia em torno de R\$ 2 milhões (cerca de USD 500 mil) por ano. Notoriamente o projeto VICTOR se consolidou como um marco importante da aplicação de IA no cenário jurídico-judicial brasileiro.

O Projeto ALEI no TRF1

Em meados de maio deste ano (2019), quando o Projeto de Digitalização de Coleta do TRF1, 2ª instância teve início, observou-se que o Tribunal poderia fazer um trabalho com maior efetividade empregando ações digitais.

Apesar das inúmeras facilidades e economias que a migração de processos físicos para o Processo Judicial Eletrônico⁷ propicia, a existência de um banco de dados (metadados) que permita apenas a busca de informações pelas partes interessadas sobre o andamento processual e o conteúdo das decisões judiciais, por meio da internet, não é capaz de incrementar de maneira significativa os índices de desempenho em termos de produtividade.

O movimento em direção à IA deu-se por meio de uma abordagem feita com o GPAM para confirmar uma possível aplicação de IA sobre processos judiciais em formato digital. A primeira tarefa executada pelo GPAM foi a confirmação do elevado nível de qualidade do processamento de OCR.

Em um segundo momento, foi verificada a possibilidade de desenvolvimento de um projeto para o TRF1 com características semelhantes ao VICTOR do STF.

Tratado por ALEI (Análise Legal Inteligente) o projeto volta-se especificamente para as necessidades dos gabinetes de 2º Grau.

Dando sequência à atual fase de digitalização dos processos judiciais do TRF1, o ALEI, projeto de pesquisa e desenvolvimento (P&D) a ser desenvolvido entre o TRF1 e GPAM deverá utilizar a base de textos jurídicos digitalizados como insumo para o desenvolvimento de tecnologias calcadas em IA para automatizar fases

processuais de segunda instância do TRF1⁸. Especificamente o ALEI deverá automatizar as seguintes etapas de forma integrada ao Processo Judicial Eletrônico (PJe), sistema de gestão de processos jurídicos utilizado pelo TRF1:

1. Análise de precedentes⁹;
2. Formação da Base de Jurisprudência Interna
3. Agrupamento de processos similares.

A *Análise de precedentes* refere-se à tarefa de verificar se um dado processo judicial possui relação com alguma Repercussão Geral no STF, Repetidos no STJ¹⁰ e/ou IRDRs (Incidente de Resolução de Demanda Repetitiva)¹¹ e decisões internas do TRF1. Para tanto, pretende-se conceber um módulo de IA para cada uma destas tarefas, uma vez que as bases de dados de cada uma delas são distintas.

O *Levantamento de jurisprudências* deverá verificar se um dado processo judicial possui decisões monocráticas ou de acórdãos (decisão colegiada) dentro do próprio TRF1. Neste caso, espera-se que um modelo de IA treinado com o banco de dados de jurisprudências do TRF1 possa responder à esta demanda.

A *Formação da Base de Jurisprudência Interna* contará com a aplicação de IA, de comparação das atuais e futuras decisões monocráticas e dos Colégios de desembargadores, Turmas (três desembargadores) e Sessões (seis desembargadores)¹².

Por fim, a base de processos do TRF1 será utilizada para treinar um modelo de IA que tenha a capacidade de agrupar processos jurídicos semelhantes a um dado processo judicial

de interesse, viabilizando uma análise conjunta de processos jurídicos.

Conclusões

As recentes legislações que tratam da modernização da administração pública brasileira fazem forte menção ao uso de tecnologias como aliadas para se otimizar os recursos que estão sendo aplicados na gestão dos órgãos governamentais do Brasil. Soma-se a este cenário nacional, a mesma tendência política e econômica internacional relativa aos vários casos de sucesso onde novas tecnologias estão demonstrando cabalmente que podem gerar economia no processo de gestão em seus diferentes nichos.

Este artigo apresentou como se deu a pesquisa e desenvolvimento do projeto VICTOR realizado em parceria entre o STF e UnB. Discorreu-se acerca dos desafios mais alinhados com desenvolvimento de software, especialmente voltados para a obtenção dos textos dos processos judiciais, passando pelo encapsulamento dos motores de inferência. Além disso, houve especial destaque aos desafios de pesquisa, que redundaram na confirmação das hipóteses acerca da modelagem baseada em AM de classificadores de peças e temas de RG.

Pouco antes das considerações finais, foi apresentada uma visão geral de projeto análogo ao VICTOR, de interesse do TRF1, tratando dos seus desafios e oportunidades para as partes envolvidas.

O modelo de desenvolvimento de tecnologias envolvendo IA em parceria com grupos de pesquisas de universidades públicas brasileiras

8 O ALEI está em fase de formalização de termo de cooperação entre TRF1 e UnB para seu início. Estima-se que seu efetivo desenvolvimento pelo GPAM inicie entre o final de 2019 e início de 2020.

9 Precedentes são julgados anteriores que servem de parâmetro para novos julgamentos. Exemplos, Repercussão Geral no STF, Repetidos no STJ e IRDRs (Incidente de Resolução de Demanda Repetitiva) no TRF1.

10 O art. 1.036 do Código de Processo Civil-CPC/2015 dispõe que, quando houver multiplicidade de recursos especiais com fundamento em idêntica controvérsia, a análise do mérito recursal pode ocorrer por amostragem, mediante a seleção de recursos que representem de maneira adequada, a controvérsia. Recurso repetitivo, portanto, é aquele que representa um grupo de recursos especiais que tenham teses idênticas, ou seja, que possuam fundamento em idêntica questão de direito.

Fonte: <http://www.stj.jus.br/sites/porta1p/Processos/Repetitivos-e-IACs/Saiba-mais/Sobre-Recursos-Repetitivos> Acesso em: 19/11/2019

11 O Incidente de Resolução de Demandas Repetitivas (IRDR) foi uma das inovações processuais trazidas pelo novo Código de Processo Civil (Lei n. 13.105, de 16 de março de 2015). Encontra-se regulamentado nos artigos 976 a 987 do CPC e tem como principal objetivo identificar processos que contenham a mesma questão de direito, para decisão conjunta. Fonte: <https://www.tjdft.jus.br/consultas/gerenciamento-de-precedentes/irdr> Acesso em: 18/11/2019

12 O número de desembargadores depende da natureza do julgamento e está previsto no Regimento Interno.

(como é o caso do GPAM) para atender demandas específicas de órgãos públicos tem sido muito exitoso. A iniciativa mostrou uma diminuição do custo final de desenvolvimento, minorou trâmites burocráticos, contou com recursos humanos com formações multidisciplinares e com alto nível de conhecimento agregado (pesquisadores e acadêmicos em nível de doutorado, mestrado e graduação), contribuiu na formação de mão de obra para o mercado e com o aparelhamento de laboratórios de pesquisas que via de regra são utilizados em outros projetos de mesma natureza.

Num contexto focado em economicidade e produtividade, a IA parece estar colaborando para melhorar o sistema jurídico e judicial brasileiro, encurtando tempos de respostas e economizando o erário público.

Os projetos VICTOR e ALEI detalhados aqui mostram que a IA pode ser implantada para a automação de processos judiciais de forma a gerar economia para o estado¹³, modernizar processos e melhorar a qualidade de trabalho das pessoas que estão envolvidas em tarefas repetitivas e não tão nobres, dada a capacidade intelectual da força de trabalho.

Referências

- Braz, F. A., da Silva, N. C., de Campos, T. E., Chaves, F. B. S., Ferreira, M. H., Inazawa, P. H., ... & Bezerra, D. A. (2018). Document classification using a Bi-LSTM to unclog Brazil's supreme court. arXiv:1811.11569.
- Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794). ACM.
- Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2016). *Very deep convolutional networks for text classification*. arXiv:1606.01781.
- Conselho Nacional de Justiça (CNJ). (2018). Justiça em números. Acesso 28/07/2019 em <https://tinyurl.com/unas48f>
- Da Ros, L. (2015). O custo da Justiça no Brasil: uma análise comparativa exploratória. Newsletter. Observatório de elites políticas e sociais do Brasil. NUSP/UFPR, 2(9), 1-15. ISSN 2359-2826.
- Da Silva, N. C. (2018). Notas Iniciais Sobre a Evolução dos Algoritmos do Victor: O Primeiro Projeto de Inteligência Artificial em Supremas Cortes do Mundo. In: R. Fernandes and A. Gamba, ed., II Congresso Internacional de Direito, Governo e Tecnologia, 1st ed. Belo Horizonte: Fórum. ISBN 978-85-450-0584-1.
- Da Silva, N. C., Braz, F. A., de Campos, T., Gusmao, D., Chaves, F., Mendes, D., ... & Carvalho, G. H. T. A. (2018). Document type classification for Brazil's supreme court using a convolutional neural network. In *The tenth international conference on forensic computer science and cyber law-ICoFCS* (pp. 7-11).
- Decreto n. 9.739. (2019, 28 de março). Estabelece medidas de eficiência organizacional para o aprimoramento da administração pública federal direta, autárquica e fundacional, estabelece normas sobre concursos públicos e dispõe sobre o Sistema de Organização e Inovação Institucional do Governo Federal - SIOrg. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9739.htm
- Decreto n. 25. (2019, 19 de fevereiro). Institui o Laboratório de Inovação para o Processo Judicial em meio Eletrônico – Inova PJe e o Centro de Inteligência Artificial aplicada ao PJe e dá outras providências. Brasília, DF: Conselho Nacional de Justiça. Acesso 18/11/2019 em <https://atos.cnj.jus.br/atos/detalhar/atos-normativos?documento=2829>
- Decreto n. 9.584. (2018, 26 de novembro). Altera o Decreto nº 8.638, de 15 de janeiro de 2016, para instituir a Rede Nacional de Governo Digital. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Decreto/D9584.htm
- Decreto n. 9.319. (2018, 21 de março). Institui o Sistema Nacional para a Transformação Digital e estabelece a estrutura de governança para a implantação da Estratégia Brasileira para a Transformação Digital. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Decreto/D9319.htm
- Decreto n. 9.203. (2017, 22 de novembro). Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Decreto/D9203.htm
- Decreto n. 8.638. (2016, 15 de janeiro). Institui a Política de Governança Digital no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fun-

13 Haja visto o projeto VICTOR, o qual gera uma economia ao STF superior ao seus custos de desenvolvimento nos primeiros seis meses de sua plena implantação.

- dacional. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8638.htm
- Donoho, D. (2015). 50 years of Data Science. In Tukey Centennial Workshop, Princeton, NJ. Disponível em: http://www.ccs.neu.edu/course/cs7280sp16/CS7280-Spring16_files/50YearsOfDataScience.pdf
- Forman, G. (2003). *An extensive empirical study of feature selection metrics for text classification*. *Journal of machine learning research*, 3(Mar), 1289-1305.
- Graves, A., & Schmidhuber, J. (2005). *Framewise phoneme classification with bidirectional LSTM and other neural network architectures*. *Neural networks*, 18(5-6), 602-610.
- Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory*. *Neural computation*, 9(8), 1735-1780.
- Inazawa, P., Peixoto, F., De Campos, T., Da Silva, N., Braz, F. (2019). 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. *Revista da Sociedade Brasileira de Computação - Machine Learning: Desafios para um Brasil Competitivo*, Porto Alegre - RS, v. 1, p. 19 - 24.
- Lei n. 13.243. (2016, 11 de janeiro). Dispõe sobre estímulos ao desenvolvimento científico, à pesquisa, à capacitação científica e tecnológica e à inovação e altera a Lei nº 10.973, de 2 de dezembro de 2004, a Lei nº 6.815, de 19 de agosto de 1980, a Lei nº 8.666, de 21 de junho de 1993, a Lei nº 12.462, de 4 de agosto de 2011, a Lei nº 8.745, de 9 de dezembro de 1993, a Lei nº 8.958, de 20 de dezembro de 1994, a Lei nº 8.010, de 29 de março de 1990, a Lei nº 8.032, de 12 de abril de 1990, e a Lei nº 12.772, de 28 de dezembro de 2012, nos termos da Emenda Constitucional nº 85, de 26 de fevereiro de 2015. Brasília, DF: Presidência da República. Acesso 18/11/2019 em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13243.htm
- Meyer, B. (2018). Making Sense of Agile Methods. *IEEE Software*, 35(2), 91-94.
- Muja, M., & Lowe, D. G. (2009). *Fast approximate nearest neighbors with automatic algorithm configuration*. *VISAPP* (1), 2(331-340), 2.
- Perozzi, B., Al-Rfou, R., Kulkarni, V., & Skiena, S. (2014). *Inducing language networks from continuous space word representations*. In *Complex Networks V* (pp. 261-273). Springer, Cham.
- Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). *Tackling the poor assumptions of naive bayes text classifiers*. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).
- Resolução n. 9.203. (2019, 20 de março). Dispõe sobre a instituição do Programa de Desburocratização e Simplificação da Justiça Federal – FLUI JF, da Rede Colaborativa de Desburocratização e Simplificação da Justiça Federal – Rede FLUI, e do Selo de Desburocratização e Simplificação da Justiça Federal. Brasília, DF: Conselho da Justiça Federal. Acesso 18/11/2019 em <https://www2.cjf.jus.br/jspui/bitstream/handle/1234/50245/Res%20529-2019.pdf>
- Smith, R. (2007, September). *An overview of the Tesseract OCR engine*. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 2, pp. 629-633). IEEE.
- Toledo, E. (2018). Projetos de inovação tecnológica na Administração Pública. In: R. Fernandes and A. Gamba, ed., II Congresso Internacional de Direito, Governo e Tecnologia, 1st ed. Belo Horizonte: Fórum. ISBN 978-85-450-0584-1.
- Williams, L., & Cockburn, A. (2003). Agile software development: it's about feedback and change. *IEEE Computer*, 36(6), 39-43.
- Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). *Interpreting tf-idf term weights as making relevance decisions*. *ACM Transactions on Information Systems (TOIS)*, 26(3), 13.