



PROTEÇÃO DOS DIREITOS HUMANOS E DO MEIO AMBIENTE;  
GARANTIA DA SEGURANÇA JURÍDICA PARA OTIMIZAÇÃO DO  
AMBIENTE DE NEGÓCIOS NO BRASIL; COMBATE À CORRUPÇÃO,  
AO CRIME ORGANIZADO E À LAVAGEM DE DINHEIRO; INCENTIVO  
AO ACESSO À JUSTIÇA DIGITAL; E FORTALECIMENTO DA  
VOCAÇÃO CONSTITUCIONAL DO STF.

## Toth - solução inteligente preditora de classe e assuntos para processos autuados no PJe

Jairo Simão Santana Melo

Verônica Ferreira Nascente

Luiz Eduardo dos Santos

**Resumo:** Na busca por uma classificação processual adequada, o TJDFT desenvolveu o Toth para recomendar classe e assuntos do processo. No momento da autuação, o PJe aciona o Toth que efetua mineração do texto da petição inicial e classifica o processo. A recomendação ocorre quando o analista processual avalia o processo. O piloto realizado produziu resultados promissores em ambiente de produção, o que levou à decisão de efetivamente realizar recomendações ao usuário do PJe. Suas classificações têm atingido relevância satisfatória. Este artigo, primeiro, apresenta a solução Toth. Em seguida, conceitua Inteligência Artificial e descreve como o Toth utiliza processamento de linguagem natural e aprendizado supervisionado. Por fim, os resultados são apresentados.

**Palavras-chave:** Toth. Processamento de Linguagem Natural. Aprendizado de máquina. Mineração de texto. Classificação.

**Abstract:** In the search for an adequate process classification, the TJDFT developed Toth to recommend class and subjects. At the time of the process registration, the PJe triggers Toth, which mines the text of the initial petition and classifies the process. The recommendation occurs when the process analyst evaluates the process. The pilot produced promising results in a production environment, which led to the decision to effectively carry out recommendations to the PJe user. Its classifications have achieved satisfactory relevance. This article first introduces the Toth solution. It then conceptualizes Artificial Intelligence and describes how Toth uses natural language processing and supervised learning. Finally, the results are presented.

**Keywords:** Toth. Natural Language Processing. Machine Learning. Text mining. Classification.

### 1 Introdução

O processo judicial possui, na sua concepção, uma série de importantes informações que servirão como balizadores durante o seu ciclo de vida. Algumas dessas são classe e assuntos, informados pelo advogado no momento da autuação processual, que servirão para indicar, entre outras coisas, a competência do processo. Além de classe, assuntos e competência, existem documentos essenciais e obrigatórios que devem ser informados no momento do cadastro do processo pelo advogado no sistema PJe (Processo Judicial eletrônico), como, por exemplo, o documento de petição inicial. Esse documento autuado pelo advogado relata os fatos que geraram o conflito que não pode ser resolvido na seara informal, motivando assim a necessidade da presença de uma figura imparcial, o magistrado, capaz de analisar os documentos, ouvir as partes e emitir uma decisão final imparcial que deverá ser respeitada.

Na busca pela melhoria da prestação jurisdicional, o Conselho Nacional de Justiça (CNJ) estabeleceu procedimentos para que cada processo judicial receba nomenclatura padrão. A Resolução CNJ nº 46, de 18 de dezembro de 2007, implementou as Tabelas Processuais Unificadas (TPU), que visam, entre outras coisas, à uniformização taxonômica e terminológica de classe e assunto no âmbito da Justiça Estadual, Federal, do

Trabalho e do Superior Tribunal de Justiça. Nesse sentido, a classe processual refere-se ao rito processual, aquilo que diz respeito a como o processo tramitará. Assim, o tipo de procedimento assumido pelos autos durante sua tramitação é definido pela classe atribuída ao processo. O assunto, por sua vez, está relacionado ao pedido concreto, ao objeto do processo. Refere-se, então, ao que se deseja alcançar com o processo (CNJ, 2021).

O Processo Judicial eletrônico (PJe) é um sistema desenvolvido pelo CNJ por meio de parceria com diversos tribunais. O Conselho Nacional do Ministério Público (CNMP), Ordem dos Advogados do Brasil (OAB), Advocacia Pública e Defensorias Públicas participaram da construção de forma consultiva. O PJe busca o ganho de produtividade e eficiência ao criar e aprimorar a padronização na prática de atos jurídicos e no acompanhamento do trâmite processual. O sistema permite escala na oferta de serviço de justiça e automação. As Justiças Federal, Estadual, Eleitoral, do Trabalho, Militar, Tribunais Superiores e Conselhos têm feito uso do PJe (PJE, 2021).

Dessa maneira, novos litígios são inseridos nos tribunais pelo uso do sistema PJe. Após o cadastro inicial realizado pelo advogado no PJe (Figura 1), o processo passará pelo referendo de um analista processual que ficará responsável por analisar a petição inicial e os outros documentos anexos, ava-

liar os dados cadastrais e realizar eventuais correções. Todo esse procedimento, realizado individualmente, é de suma importância pois promoverá, num nível macro, o perfil de um tribunal. Será assim que teremos a oportunidade de verificar, mapear e rastrear as principais fontes de conflito que geram

litígio em cada órgão do Poder Judiciário. A errada classificação processual quanto a classes/assuntos gera inúmeros problemas para a Base Nacional de Dados do Poder Judiciário (DataJud) e, conseqüentemente, para sua correta avaliação.

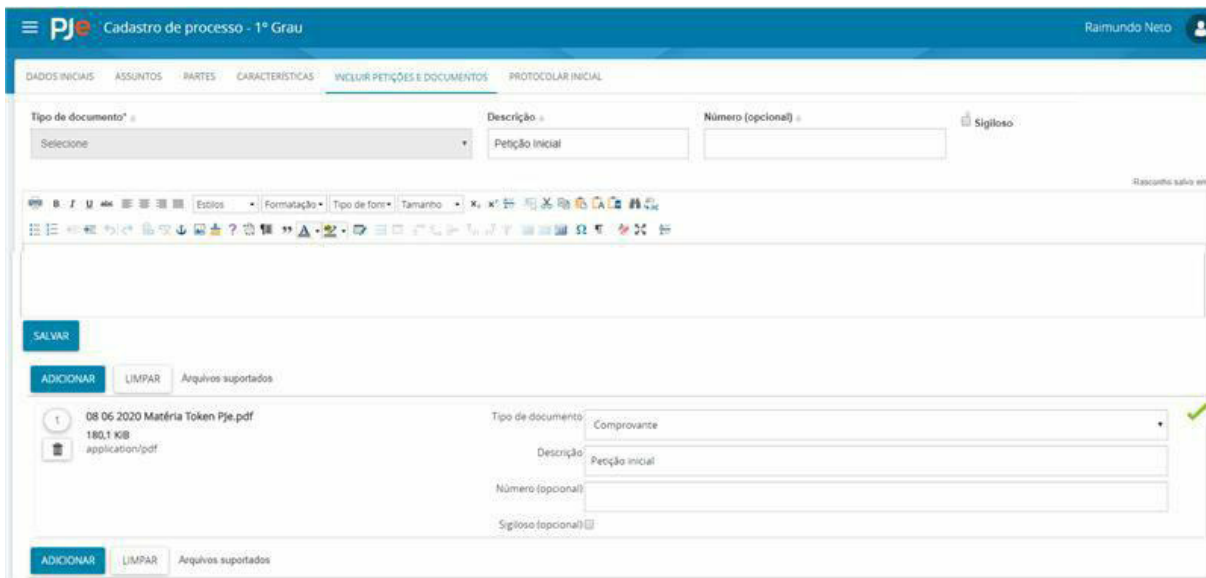


Figura 1: PJe - Tela Cadastro de Processo - 1o Grau

Segundo CNJ (2021), o DataJud é a base de dados responsável por armazenar os dados e metadados processuais de forma centralizada. Foi instituída como fonte primária de dados do Sistema de Estatística do Poder Judiciário (SIESPJ). O DataJud traz transparência ao Poder Judiciário, assim como promove integração de dados e contribui para a construção, acompanhamento e otimização de rotinas de trabalho e políticas públicas. As tecnologias utilizadas pelo CNJ possibilitam o recebimento, consumo e disponibilização de dados e metadados de forma constante e segura.

A correção de qualquer informação errônea informada pelo advogado é menos custosa se for fornecida no início do ciclo de vida processual. Pensando em uma maneira de dirimir ou, pelo menos, diminuir ao máximo erros na classificação processual, o Tribunal de Justiça do Distrito Federal e dos Territórios (TJDFT) desenvolveu o Toth, solução inteligente baseada em NLP (*Natural Language Processing*), que visa fornecer recomendações de possíveis classe e/ou assuntos do processo para os usuários do PJe no momento da análise da petição inicial. O mnemônico Toth tem relação com o Deus Egípcio de todos os ramos da sabedoria.

O Toth foi concebido para alcançar os seguintes objetivos:

- 1) Promover agilidade ao usuário interno do sistema PJe pelo uso da recomendação da classificação que será emitida no momento da análise da petição inicial. Como consequência, o TJDFT passa a garantir maior celeridade na análise dos seus processos.
- 2) Permitir que soluções em Inteligência Artificial (IA) estejam ao alcance de todos os usuários do sistema PJe, possibilitando facilidades, agregando equidade, retirando eventuais resistências, trazendo o público interno a fazer uso da Inteligência Artificial com resultados quantificáveis de maneira que essas soluções tenham atuação cada vez mais importante na busca por inovação e eficiência.
- 3) Atender às recomendações do CNJ no desenvolvimento de micro serviços e conversação entre sistemas via mensageria. Desse modo, o Toth e o PJe mantêm-se independentes entre si, sua integração é possível na forma "plug and play" (pela adição de variável dentro da tarefa de fluxo do PJe) e as funcionalidades são exibidas aos usuários de maneira transparente.

- 4) Não raro, reclassificações exigem uma redistribuição processual, o que leva inevitavelmente a um aumento do seu ciclo de vida, bem como a um retrabalho nas varas comuns ou especializadas. Evitar essas consequências, saneando os dados processuais no seu nascedouro, permite agilizar a justiça.
- 5) Aperfeiçoar o metadado processual do tribunal que será enviado ao DataJud. A classificação processual incorreta gera inúmeros problemas para a estatística do tribunal. Assim, o Toth vem a ser uma importante ferramenta para auxiliar na correta classificação processual, garantindo, dessa maneira, que os dados enviados pelo TJDFT ao CNJ sejam cada vez mais precisos e completos.
- 6) Promover o acesso à Justiça, por meio de ações e projetos desenvolvidos para o uso colaborativo de produtos que empregam novas tecnologias e Inteligência Artificial - Programa Justiça 4.0. Alinhado a esse pilar, o Toth é uma solução que atende à ação Projeto da Plataforma Digital do Poder Judiciário (PDPJ), com possibilidade de ampliar o grau de automação do processo judicial eletrônico e o uso de Inteligência Artificial.

Em sua fase de estudo inicial e na busca pela minimização de riscos e de eventuais custos associados à correção de informação processual, considerou-se as estatísticas iniciais de quantidade de erros gerados nas unidades julgadoras para justificar a intervenção e consequente desenvolvimento do Toth. A Figura 2 apresenta os dez órgãos julgadores com maior quantidade de erros.

Consulta à base analítica do PJe	
Órgão Julgador	Total de Erros
1ª Vara de Família e de Órfãos e Sucessões de Águas Claras	1989
Vara de Precatórias do DF	1849
Vara da Infância e da Juventude do DF	1602
1ª Vara de Família e de Órfãos e Sucessões de Sobradinho	1417
2ª Vara de Família e de Órfãos e Sucessões de Sobradinho	1333
Vara de Família e de Órfãos e Sucessões do Guará	1152
Vara Cível, de Família e de Órfãos e Sucessões do Recanto das Emas	1028
1ª Vara de Família e de Órfãos e Sucessões de Samambaia	1002
2ª Vara de Família e de Órfãos e Sucessões de Samambaia	991
1ª Vara de Família e de Órfãos e Sucessões de Planaltina	984

Figura 2: Dez Órgãos Julgadores com maior

quantidade de erros

O Toth teve sua autorização concedida no Plano Diretor de Tecnologia da Informação e Comunicação 2020 (PDTIC) e foi desenvolvido mediante o consórcio Assessoria de Ciência de Dados (ACID), Coordenadoria de Sistemas e Estatísticas da Primeira Instância (COSIST) e Coordenadoria de Desenvolvimento do Processo Judicial Eletrônico (CODPJE), com o auxílio do Gabinete da Primeira Vice-Presidência (GPVP), da Secretaria de Tecnologia da Informação (SETI) e da Secretaria-Geral (SEG).

## 2 Inteligência Artificial

Considerada um ramo da Ciência da Computação, a Inteligência Artificial, em linhas gerais, tem como meta fazer uso da máquina para tentar raciocinar, aprender e tomar decisões objetivando resolver problemas, do mesmo modo que nós, seres humanos, fazemos (SHARMAL; KUMAR, 2017).

Concebida na década de 1950, sua inspiração inicial foi de natureza biológica. Cientistas indagaram sobre a possibilidade de uma máquina ser capaz de simular o pensamento humano, por meio da criação de modelos matemáticos. Nesse contexto, surgiram os primeiros estudos propondo modelos de Redes Neurais Artificiais (RNAs). Haykin (1999) define RNAs como sendo sistemas baseados no funcionamento do cérebro e nas conexões de seus neurônios. Essa rede adquire conhecimento por meio de aprendizado, que é armazenado em forma de pesos sinápticos representando a força das conexões entre os neurônios.

Os maquinários de *hardware* da época eram muito aquém do necessário, em termos de processamento e armazenamento, então o ramo da Inteligência Artificial acabou relegado ao campo teórico acadêmico. Soni *et al.* (2020) descrevem que somente após a evolução dos processadores de *hardware*, principalmente com a criação da *Graphics Processing Unit* (GPU), bem como a possibilidade da execução em *Cloud*, gerou-se a possibilidade de tornar soluções complexas e robustas de Inteligência Artificial exequíveis.

Hoje, essa tão propalada Inteligência Artificial, que surgiu há quase 70 anos, finalmente tornou-se uma realidade em nossa rotina diária. Está presente nas ferramentas de buscas da Internet, nos *feeds* de notícias das redes sociais, nas recomendações de compras *online*, no atendimento ao cliente

via *chatbots* e no tradutor automático de linguagem, simulação aeroespacial e médica, sistemas de defesa e robôs industriais (SONI *et al.*, 2020). Sua utilização tornou-se tão vasta que é frequente utilizarmos um serviço sem termos ciência de que estamos na verdade consumindo uma solução desenvolvida em Inteligência Artificial. Por exemplo, em equipamentos e aplicações diversas, tais como celulares, assistentes virtuais e *internet banking* (BARBOSA; BEZERRA, 2020).

Nos últimos anos, vários campos de estudos nesta área surgiram promovendo uma grande diversidade nas soluções entregues. Ludermir (2021) cita, dentre as principais, a Visão Computacional, com seus processos para reconhecimento facial em tempo real; e o Processamento de Linguagem Natural, com a criação de tradutores de texto disponíveis em várias linguagens.

## 2.1 Tipos de Aprendizado

Para ser capaz de resolver o problema proposto, um modelo em Inteligência Artificial precisa passar pela etapa de aprendizado, que pode ser: supervisionado ou não supervisionado (HOSAKI; RIBEIRO, 2021).

Conforme descrito em Ludermir (2021), o aprendizado supervisionado ocorre quando o modelo recebe, de antemão, as entradas e as saídas em seu conjunto de dados de treinamento. Dessa maneira, o modelo consegue aprender, usando o arcabouço que lhe foi passado, qual é a saída correta para uma dada entrada. Em termos comparativos, é bem parecido com a figura de um professor em uma sala de aula. A função desse é a de ensinar seu aluno, que absorverá suas instruções e será capaz de inferir outras conclusões similares, baseando-se nos ensinamentos que lhe foram passados.

Já o aprendizado não supervisionado ocorre de maneira oposta. Nesse caso, o modelo receberá o seu conjunto de dados para treinamento sem uma chancela do especialista, ou seja, os dados não terão um rótulo ou uma classificação prévia. Assim, seu aprendizado ocorrerá por analogia, com a identificação de semelhança entre os dados (LUDERMIR, 2021).

## 3 O Toth

Conforme relatado anteriormente, o Toth é uma solução inteligente que objetiva classificar classe/assuntos, após analisar

a petição inicial protocolada pelo advogado, e exibi-la ao usuário interno do sistema PJe na forma de uma recomendação, expondo na mensagem qual é a classificação atual e qual é sua predição para o processo.

Seu fluxo de trabalho inicia-se assim que o advogado autua um processo no sistema PJe do TJDF. Nesse instante, o Toth receberá uma mensagem informando que um novo processo está disponível para classificação. O Toth reconhecerá o processo como sendo válido (através de uma busca no Modelo Nacional de Interoperabilidade (MNI) do PJe) e fará uma avaliação quanto ao seu grau de prioridade. Processos do tipo tutela ou liminar, por exemplo, por sua natureza inerente de urgência/emergência, possuem maior prioridade no processamento.

Passado pela priorização, o Toth realiza a busca pelo documento de petição inicial. Atualmente, o PJe permite que o advogado tenha liberdade ao inserir a documentação processual. Não há qualquer exigência de ordenação específica para se anexar os documentos necessários. Dessa forma, tornou-se necessária a realização de um estudo para encontrar a petição inicial do processo. A solução encontrada baseou-se no tipo de documento principal, anexado pelo advogado, a partir da escolha feita por ele da classe processual autuada. Por meio do uso de serviços *Representational State Transfer* (REST), consulta ao MNI e análise dos dados, foi possível mapear um comportamento padrão realizado pela maioria dos advogados. Pelo fato de ter sido possível realizar este rastreamento de maneira satisfatória, não foi necessário a implementação de Inteligência Artificial nesta busca.

Encontrado o documento de petição inicial, a solução segue para a mineração do texto e respectivas classificações de classe e de assuntos processuais, as quais ficarão armazenadas na base de dados do Toth. Quando o analista processual acessar a tela que trata a petição inicial de um determinado processo, o PJe enviará uma requisição ao Toth. Caso exista recomendação de classe, assuntos ou ambos em sua base de dados, o Toth responderá ao PJe informando quais são as suas predições para o processo em questão. A Figura 3 retrata o fluxo básico da interação entre Toth e PJe.

No momento da distribuição processual:



No momento da análise da petição inicial:

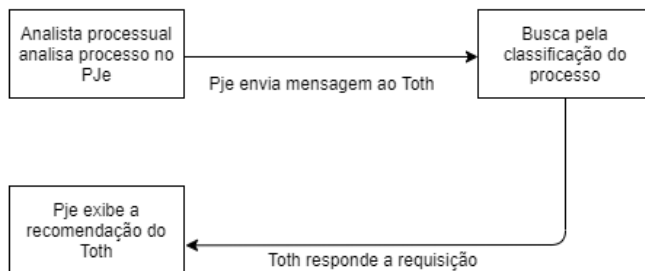


Figura 3: Fluxo básico da interação entre o Toth e o PJe

### 3.1 Processamento de linguagem natural no Toth

O processamento de linguagem natural *Natural Language Processing* (NLP) é a subárea da Inteligência Artificial empenhada na solução de problemas de origem linguística (LUDERMIR, 2021). Seu principal objetivo é compreender a interação humana feita através da comunicação, seja ela via oral ou escrita, para assim fornecer informações relevantes aos seus usuários.

O Toth baseia-se nos princípios do NLP para prover sua solução. Seu objetivo é compreender a comunicação humana através da petição inicial protocolada pelo advogado no sistema PJe e buscar por palavras-chaves que ajudem a obter a essência do litígio em questão, de maneira a colaborar com o usuário do PJe na recomendação de possíveis classe e assuntos para o processo. O Toth leva em consideração o que já sabe sobre os temas em questão (o que aconteceu durante o seu aprendizado e treinamento) e emite, após analisar a petição inicial, uma classificação e seu respectivo grau de relevância para o caso concreto.

Uma etapa fundamental dentro do NLP é a de converter a comunicação humana de uma maneira que o computador consiga compreender este diálogo (LUDERMIR, 2021). Assim, faz-se necessário realizar uma conversão dos textos protocolados pelo advogado no sistema PJe em caracteres acessíveis ao processamento computacional. Para realizar essa tarefa, o Toth possui, no seu arcabouço, dois diferentes conversores.

Um conversor é responsável por transformar arquivos escritos, no formato PDF, em formato acessível. O outro conversor realiza a tarefa de analisar textos escritos em HTML, pois o sistema PJe permite que a petição seja enviada usando seu editor de texto próprio, que salva as informações no formato HTML. Assim, esse conversor realiza *web scraping* (coleta de dados) de páginas no formato HTML (nesse caso, a petição) para obter o texto no formato ideal para realização de análises.

Depois de realizada a conversão dos arquivos em um formato acessível ao processamento computacional, o Toth realiza o processo de mineração do texto propriamente dito (com retirada de dados não essenciais), deixando a petição apenas com as informações relevantes antes de proceder para a classificação. Neste trabalho, foi realizado um estudo sistemático baseado em artigos escritos tanto na comunidade técnica quanto no campo acadêmico e seguindo as orientações de pré-processamento descritas em Gusmão *et al.* (2021) para retirada de *stopwords* (palavras vazias), informações comuns em todas as petições ou irrelevantes para a análise em questão.

### 3.2 Aprendizado no Toth

No Toth foi escolhido o processo de aprendizado supervisionado. Após análise do Núcleo Permanente de Sistemas da Primeira Instância (NUSIS) junto com a Corregedoria do TJDF, ficou decidido que o órgão piloto do Toth seria a 1ª Vara Cível, de

Família e de Órfãos e Sucessões de Santa Maria (IVCFAMOSSMA). A partir da escolha da vara piloto, a equipe da ACID preparou um relatório com as classes e os assuntos mais peticionados naquela vara e emitiu um outro, mais específico, que exibia números processuais dessas classes/assuntos. Com a listagem em mãos, iniciou-se, com o auxílio do GPVP, a coleta das petições iniciais, bem como sua correta classificação.

Durante o treinamento, foram desenvolvidos dois diferentes e independentes modelos: um para classificar classes e outro para classificar assuntos. Em ambos os modelos, foi feito uso do *fastText*, biblioteca gratuita e de código aberto desenvolvida pelo laboratório *AI Research do Facebook*. Como descrito em Joulin et al. (2016), *fastText* tem precisão semelhante a classificadores de aprendizado profundo e muitas ordens de magnitude mais rápidas para treinamento e avaliação. O modelo de classe caracteriza-se em um modelo de classificação *single-label*, ou seja, sua classificação emitirá uma única resposta. O modelo de assuntos já é considerado um modelo de classificação *multi-label*, pois sua resposta poderá ser composta por uma ou mais classificações. Read et al. (2011) explica que *multi-label* tem recebido cada vez mais atenção e é aplicável a uma grande variedade de domínios, incluindo classificação de texto.

Na primeira versão usada em ambiente de produção, o modelo de classe possuía 22 diferentes códigos extraídos após a análise de mais de 16 mil petições, e o modelo de assunto compunha-se de 50 códigos após analisar mais de 12 mil petições. Os dois modelos foram ajustados para emitir classificações quando o grau de relevância da análise fosse igual ou superior a 50%.

Massa Treinamento Rotulado	
Classe CNJ	Quantidade Petições
436	2282
7	1851
1116	1555
12154	1249
69	1059
39	886
81	838
12372	723
12246	673
120	569
40	522
1294	495

10943	445
156	434
283	402
58	391
172	336
74	306
1268	302
12541	301
12078	164
37	162

Figura 4: Massa de treinamento rotulado

### 3.3 Avaliação das previsões do Toth e publicação em ambiente de produção

Finalizado o período de coleta e treinamento inicial do modelo, o Toth passou por uma avaliação de suas classificações por usuários especialistas do GPVP, que ficaram responsáveis por analisar as petições, observar as classificações feitas pelo Toth e julgar o quão satisfatória estava aquela recomendação com relação a sua classificação jurídica pessoal. De acordo com esses especialistas, as recomendações de classe do Toth foram consideradas ótima, boa ou regular em 90% dos casos e as recomendações de assuntos tiveram a mesma avaliação em 84% dos casos analisados.

O relatório de satisfação do usuário especialista foi repassado à equipe comercial do NUSIS, que concluiu pela aprovação do Toth no ambiente de produção para a vara piloto, porém sem exibir recomendação ao usuário. Nesse período, o Toth processou cerca de 360 petições, obtendo insumos para avaliação de seu desempenho pela Corregedoria do TJDFR. Os resultados podem ser visualizados na Figura 5.

Resumo Geral - Processamento do Toth		360
<b>Total de petições processadas</b>		<b>360</b>
Recomendação de classe		<b>61,11%</b>
	Igual ao advogado	72,73%
	Diferente do advogado	27,27%
Recomendação de assunto		<b>55,56%</b>
	Pelo menos 1 indicação confere	67,50%
	Não confere com único assunto *	27,00%
	Não confere com nenhum dos assuntos **	5,50%

\* : Advogado cadastrou 1 assunto no PJe  
 \*\* : Advogado cadastrou 2 ou mais assuntos no PJe

Figura 5: Quadro resumo sobre o processamento do Toth

Após avaliação dos resultados exibidos acima, a Corregedoria do TJDFR decidiu por aprovar a liberação da recomendação no ambiente PJe para a vara piloto.

A receptividade do Toth, segundo usuários, foi bastante animadora, como conta Guilherme Cabral, Diretor de Secretaria

da TVCFAMOSSMA: “A experiência tem sido muitíssimo interessante, já que são feitas análises das petições iniciais, com recomendações para o processo, no tocante à classe e aos assuntos específicos para cada caso em concreto”.

Até o momento, o Toth foi capaz de realizar, para a vara piloto, predições de classe e assuntos em 67% e 64% dos casos analisados, respectivamente. Suas recomendações foram compatíveis às classificações realizadas pelo analista processual em 81% dos casos de classes e 77% dos casos de assuntos.

### 3.4 Expansão

Após a publicação completa do piloto, iniciou-se uma nova fase, ainda mais desafiadora, que foi a expansão do Toth para outras varas com competências distintas da TVCFAMOSSMA. Nessa expansão, foram escolhidas 80 varas de diversas competências. Assim como feito na vara piloto, foi tomada a decisão de somente realizar o processamento das petições, sem exibir recomendações ao usuário. Dessa maneira, ficou decidido que a liberação da recomendação seria feita por vara, após análise dos resultados das predições.

Foi iniciado também um grupo de estudo formado pelas equipes do NUSIS e da ACID para analisar o ciclo de vida processual dentro do PJe e obter informações sobre quais eram os fluxos do sistema responsáveis por escolher classe/assuntos definitivos de um dado processo nas diversas varas, após a análise da petição inicial pelo usuário interno.

Esse estudo foi uma das etapas mais importantes, pois novos códigos surgem a todo momento, bem como outros caem em desuso. Esse aspecto exige que o Toth acompanhe o dinamismo existente no Poder Judiciário, tornando-se assim uma ferramenta inteligente, eficiente e robusta que será capaz de realizar autoaprendizagem de forma contínua.

O Toth executa automaticamente um *job* diário de treinamento que resulta em um novo modelo. Esse mecanismo de autoaprendizagem já exibe resultados expressivos. As Figuras 6 e 7 apresentam, respectivamente, as quantidades atuais de códigos disponíveis para classes e assuntos. Embora o Toth ainda se encontre em processo de expansão para novas Serventias, já foi possível registrar o aumento da quantidade de cóni-

gos disponíveis de 22 para 31, para classes, e de 50 para 104 códigos disponíveis, para assuntos.

Classes		
Classe Raiz	Códigos disponíveis	Relevância nas recomendações
Processo cível e do trabalho	23	95%
Processo criminal	8	98%

Figura 6: Quantitativo de classes e relevância em suas recomendações

Assuntos		
Assunto Raiz	Códigos disponíveis	Relevância nas recomendações
Direito penal	28	88%
Direito processual penal	2	85%
Direito civil	38	89%
Direito tributário	3	82%
Direito do consumidor	9	78%
Registros públicos	1	78%
Direito previdenciário	2	89%
Direito processual civil e do trabalho	15	84%
Direito da criança e do adolescente	2	90%
Direito administrativo e outras matérias de direito público	3	90%
Direito da saúde	1	91%

Figura 7: Quantitativo de assuntos e relevância em suas recomendações

### 3.5 Ferramentas e Serviços

O Toth utilizou um variado conjunto de ferramentas e tecnologias durante seu processo de desenvolvimento. Seguindo as diretrizes da administração pública, 90% da tecnologia utilizada é de uso gratuito. Procurou-se também atender às recomendações do CNJ quanto ao uso de micro serviços, mensageria e outras instruções que tenham como objetivo realizar uma conversação não intrusiva com o PJe.

As principais ferramentas e tecnologias usadas durante o projeto foram: *Python, Java, PostgreSQL, Tesseract OCR, Telegram, GitLab, Rclone, Docker, Minio, RabbitMQ, Keycloak, Spring Boot, Hibernate, Json, Maven, Okd, Flyway, fastText, Oracle e OpenShift*.

Por meio do uso dessas, foi possível realizar a construção de uma gama de serviços diversos, o que permitiu ao Toth tornar-se estável, robusto, inteligente e livre de interação humana permanente. Sua constelação é composta pelos seguintes serviços:



- 1) Classificação: responsável por consumir as mensagens enviadas pelo sistema PJe via Rabbitmq, realizando a mineração do texto via Optical Character Recognition (OCR), limpeza, tratamento dos dados e posterior classificação processual de classe/assuntos.
- 2) Recomendação: responsável por encaminhar ao sistema PJe a sugestão de classe/assuntos do processo, bem como o respectivo grau de relevância de cada uma em termos percentuais.
- 3) Modelos: responsável por obter a classificação final escolhida pelo usuário interno do PJe e utilizar esta informação para treinar os modelos de classes e de assuntos. Desta forma, o Toth possui a capacidade de realizar autoaprendizado, incluindo novas classes e novos assuntos no seu arcabouço sem a necessidade da existência de uma interação humana direta.
- 4) Telemetria: responsável por medir e comunicar à equipe ACID sobre a evolução e uso da solução Toth, bem como informar caso algum erro crítico ocorra. Disponibilizado por diferentes canais, o Toth possui: um *Bot* para fornecimento de informações e geração de relatórios técnicos e gerenciais; um *Newsletters* que informa em tempo real dados sobre o treinamento de seus modelos; um serviço de *e-mail* disparado sempre que novos códigos (de classe ou de assunto) sejam disponibilizados na sua plataforma. De modo adicional, quando algum erro crítico ou fatal acontece, os desenvolvedores do Toth são informados em tempo real tanto por *e-mail*, quanto por mensagem via *Telegram*, permitindo assim rapidez na análise e enfrentamento do problema.

#### 4 Conclusão

Assim como toda solução imaginada dentro do arcabouço de Inteligência Artificial, o Toth tem como primazia facilitar e melhorar o dia a dia do usuário interno do sistema PJe, auxiliando na recomendação de classes e assuntos para os processos que são distribuídos no TJDFT.

Concebido dentro dos paradigmas de NLP, pelo uso do modelo de aprendizado supervisionado, o Toth realiza mineração do texto protocolado pelo advogado, explora a essência do litígio em voga e fornece uma

classificação única de classe e múltipla de assuntos, caso o grau de relevância de sua análise seja igual ou superior a 50%.

Não obstante, sua importância não reside somente em prover facilidade e conforto ao uso do PJe por meio da sua recomendação. Outro ponto relevante é a melhoria na qualidade dos dados que são enviados ao CNJ via DataJud. Instituído pela Resolução CNJ n. 331/2020. Segundo (CNJ, 2020):

DataJud é responsável pelo armazenamento centralizado dos dados e metadados processuais relativos a todos os processos físicos ou eletrônicos, públicos ou sigilosos dos tribunais indicados nos incisos II a VII do art. 92 da Constituição Federal. Os dados do DataJud são usados para estudos e diagnósticos do Poder Judiciário a fim de contribuir com a construção e acompanhamento de políticas públicas, otimizar as rotinas de trabalho com a unificação de sistemas, promover integração de dados entre entes públicos, além de conferir maior transparência ao Poder Judiciário.

A classificação errada dos metadados processuais (dentre eles, classes e assuntos) gera inúmeros problemas para o correto mapeamento do perfil de um tribunal. Assim, o Toth vem a ser uma importante ferramenta para auxiliar na correta classificação processual, garantindo desta maneira que os dados enviados ao CNJ sejam cada vez mais corretos e completos, com a elevada acurácia que sempre é uma exigência constante do TJDFT.

O Toth está atualmente em fase de expansão, analisando dados vindos de 80 varas de diversas competências distintas. O prognóstico, até o momento, é bastante animador com a classificação de classe do Toth sendo igual ao protocolado pelo advogado em 81% dos casos e igual ao escolhido pelo analista processual do PJe em 78% dos casos. No caso da classificação de assuntos, apesar da complexidade inerente ao modelo com múltiplas respostas, o Toth foi capaz de prever pelo menos 1 dos códigos escolhidos pelo usuário em 48% dos casos, para advogado, e 56% dos casos, para analista processual.

Assim sendo, acredita-se que o Toth será capaz de promover, ao mesmo tempo, comodidade ao usuário PJe, assertividade na classificação de metadados de classe/assuntos e melhoria na acurácia dos dados enviados ao DataJud. Todos esses fatores corroboram os valores institucionais do TJ-

DFT, que preza sempre em inovar, buscando cada vez mais ser um modelo de excelência na prestação jurisdicional.

### Referências

BARBOSA, Xênia de Castro; BEZERRA, Ruth Ferreira. Breve introdução à história da inteligência artificial. **Jamax**, v. 4, n. 1, 2020.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução n. 331, 2020**. Institui a Base Nacional de Dados do Poder Judiciário. Brasília: CNJ, 2020. Disponível em: [atos.cnj.jus.br/atos/detalhar/3428](https://atos.cnj.jus.br/atos/detalhar/3428). Acesso em: 2 ago. 2021.

CONSELHO NACIONAL DE JUSTIÇA. **Tabelas processuais unificadas**. Disponível em: [cnj.jus.br](https://cnj.jus.br). Acesso em: 02 ago. de 2021.

GUSMÃO, Camila; FIGUEIREDO, Karla; BRITO, Walkir A. T. Técnicas de processamento de linguagem natural em denúncias criminais: automatização e classificação de texto em português coloquial. *In*: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE (SEMISH), 48., 2021, Evento Online. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 172-182. ISSN 2595-6205. DOI: <https://doi.org/10.5753/semish.2021.15820>.

HAYKIN, Simon. **Neural Networks: a comprehensive foundation**. Upper Saddle River : Macmillan, 1994, p. 363-394.

HOSAKI, Gabriel; RIBEIRO, Douglas. DEEP LEARNING Ensinando a aprender. **Revista de Gestão e Estratégia**, Assis-SP, v. 3, n.1, 2021.

JOULIN, Armand; GRAVE, Edouard; BOJANOWSKI, Piotr; MIKOLOV, Tomas. **Bag of Tricks for Efficient Text Classification**. *In*: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, Spain : Association for Computational Linguistics, p. 427-431, v. 2.

LUDERMIR, Teresa Bernarda. Inteligência artificial e aprendizado de máquina: estado atual e tendências. **Estudos Avançados**, São Paulo, 35(101), p. 85-94. 2021.

PJE. Processo Judicial Eletrônico. **Página Principal**. Disponível em: [<pje.jus.br>](https://pje.jus.br). Acesso em: 2 ago. 2021.

READ, Jesse; PFAHRINGER, Bernhard; HOLMES, Geoff; FRANK, Eibe. **Classifier chains for multi-label classification**. Hamilton, New Zealand : The University of Waikato. Disponível em : <https://www.cs.waikato.ac.nz/~eibe/pubs/chains.pdf> . Acesso em: 30 Jun 2011.

SHARMAL, Diksha; KUMAR, Neeraj. A Review on Machine Learning Algorithms, Tasks and Applications. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, v. 6, Issue 10, October 2017.

SONI, Neha; SHARMA, Enakshi Khular; SINGH, Narotam; KAPOOR, Amita. Artificial Intelligence in Business: From Research and Innovation to Market Deployment. **Procedia Computer Science**, v. 167, p. 2200-2210, 2020.

#### Jairo Simão Santana Melo

Doutor em Engenharia Elétrica pela Universidade de Brasília – PGEA/UnB (2012). Lidera em nível tático diversos projetos de inovação e transformação digital do TJDF, com destaque para projetos de Inteligência Artificial, Data Warehouse, Analytics e BI. Também atua como docente do ensino superior e pós-graduação, desde 2018.

#### Verônica Ferreira Nascente

Pós-Graduada em Deep Learning - IGTI. Pós-Graduada em Gestão e Desenvolvimento de Sistemas de Informação - Fortium. Graduada em Engenharia de Computação - UFC. É servidora do TJDF desde 2013 e está lotada na Assessoria de Ciência de Dados atuando no desenvolvimento de soluções em Inteligência Artificial.

#### Luiz Eduardo dos Santos

Mestre em Gestão de Sistemas da Informação pela Coventry University, UK. Especialista em Gestão de TI pela UNIEURO-DF; Graduado em Processamento de Dados pela UnB-DF. Atua como professor associado no Ibmec-DF e na UPIS-DF. É servidor do TJDF desde 2008 e está lotado na Assessoria de Ciência de Dados.