

QuaTI: Sistema de Classificação de Peças Jurídicas para Identificação de Processos Prescritos

Fernando Peixoto^{1†}, Jorge Cecatto^{1†}, Marcos Salami^{1†}

^{1*}Faculdade de Computação, UFMS, Av. Costa e Silva, Campo Grande, 79070900, MS, Brasil.

Contributing authors: fernando.peixoto@ufms.br;
jorge.cecatto@ufms.br; marcos.salami@ufms.br;

[†]These authors contributed equally to this work.

Resumo

A Procuradoria Geral do Estado do Mato Grosso Do Sul atualmente vive um impasse. O número de peças jurídicas a serem analisadas cresce linearmente a cada ano ao passo que o número de assessores e procuradores não cresce na mesma velocidade. Este trabalho tem por objetivo potencializar o trabalho dos procuradores pelo estudo, desenvolvimento, implementação e implantação de técnicas para automação de rotinas jurídicas a partir do processamento de linguagem natural. Assim, testes com as mais modernas arquiteturas em inteligência artificial foram elaboradas a fim de atingir o estado da arte nesse domínio. Este trabalho apresenta uma base de dados inédita com movimentações jurídicas rotuladas, bem como um sistema construído a partir de modelos de processamento de linguagem natural com taxas de acerto em 95%, apresentando, então, uma mudança de paradigmas para esta tarefa.

Keywords: Natural Language Processing, Expert System, Systems Development, Framework

1 Introdução

A Procuradoria-Geral do Estado (PGE) é a instituição responsável pela representação do Estado e a defesa de seus direitos e interesses nas áreas judicial, extrajudicial e administrativa, em caráter exclusivo. Entre as competências atribuídas à PGE

enquadram-se a cobrança da dívida ativa estadual e das decisões do Tribunal de Contas em favor da Fazenda Pública Estadual. Anualmente, observa-se o crescimento na quantidade de processos e peças elaboradas pela PGE. Em 2019, houve um crescimento de 7,3% no número de peças elaboradas em comparação com o ano anterior. Em 2020, o crescimento atinge 14,5% em relação a 2019. Por fim, em 2021, o número de peças é 15,4% maior que em 2020 como mostra a Figura 1. Este crescimento evidencia que o número de tarefas vinculadas a análise processual cresce linearmente, de modo que o capital humano atual se torna incompatível com o número de tarefas acumuladas. Nesse sentido, buscam-se alternativas para automação de rotinas repetitivas, a fim de acelerar o custo temporal de análises processuais. Uma das possíveis alternativas é a utilização de modelos de inteligência artificial para identificação de marcos, tendências e prescrição de processos.

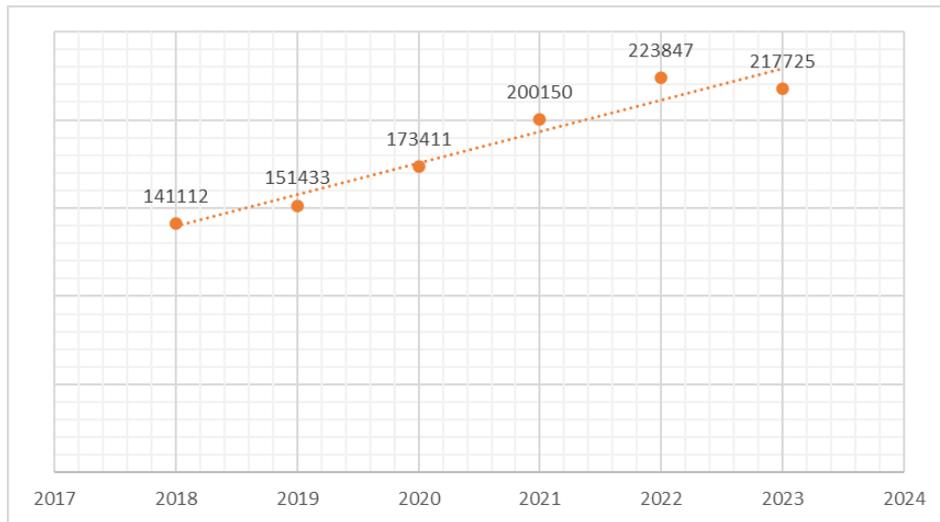


Figura 1: Número de peças finalizadas. Fonte: Unidade de Inteligência e Estatística (UIE/CIGE/PGE), a partir de dados do SAJ.

Ao longo dos anos diversos trabalhos surgiram na tentativa de automatizar rotinas jurídicas, à exemplo do artigo de Barros *et al.*[1], que utiliza modelos Bayseanos para análise de tendências em processos nas cortes brasileiras. De semelhante modo, o trabalho proposto por Silva *et al.*[2] utiliza processamento de linguagem natural (PLN) a fim de prever a decisão final de processos, nos tribunais de Minas Gerais, Brasil. Além disso, é possível constatar abordagens como a de Sleimi *et al.*[3], que utilizam PLN para extração de regras. Outro, notável exemplo é o trabalho de Aletras *et al.*[4], que utilizou modelos de inteligência artificial na predição de decisões judiciais nas Cortes de Direitos Humanos Europeias.

Nota-se, porém a ausência de pesquisas relacionadas a automação de rotinas jurídicas com inteligência artificial vinculadas a prescrição intercorrente. O prazo prescricional é um objeto de interesse dentro dos órgãos públicos, como a PGE, pois uma vez que findado o prazo de prescricional, sendo este não reconhecido pelo Estado, é cabível ao devedor solicitar a prescrição. Com o processo prescrito existe, a imposição da obrigação ao Estado de suportar os valores referentes aos honorários advocatícios. Tal cenário, acarreta prejuízos à fazenda pública, ocasionando a perda de recursos. Dessa forma, propõe-se a criação de uma ferramenta que se utiliza do processamento de linguagem natural para identificar marcos processuais dentro do fluxo da execução fiscal, onde a prescrição intercorrente acontece. A partir disso, montar uma linha do tempo processual, de modo que seja possível o cálculo referente a data de prescrição do processo.

Dentre as principais contribuições desse artigo destacam-se a:

- Criação de uma base de dados jurídica rotulada com auxílio de especialistas
- Treinamento de um modelo para classificação de movimentações jurídicas
- Criação de mecanismo para cálculo de prescrição
- Desenvolvimento do QuaTI, um sistema que engloba os procedimentos de coleta, rotulação e classificação de processos e cálculo do tempo prescricional.

2 Fluxo da Prescrição Intercorrente

Para a compreensão do fluxo da prescrição intercorrente é necessário compreender alguns termos que fazem parte do fluxo. Nesta seção são descritos os principais termos, bem como o fluxo da prescrição.

2.1 Citação

A citação, conforme estabelecido pelo Artigo 238 da Lei nº 13.105, é o ato pelo qual são convocados o réu, o executado ou o interessado para integrar a relação processual, podendo ocorrer pelo correio, considerando-se efetuada na data de entrega da carta no endereço do executado. Caso o aviso de recepção não retorne em quinze dias, a citação será realizada por oficial de justiça ou por edital (Art. 7º Lei No 6.830), observa-se a citação sendo positiva quando o réu é convocado, ou seja, a citação sendo executada.

2.2 Penhora

A penhora, definida pelo Art. 789 do Código de Processo Civil, é o ato em que o devedor responde com todos os seus bens presentes e futuros para cumprir suas obrigações, salvo as restrições legais. A penhora é efetivada mediante a apreensão, sequestro ou arresto e depósito dos bens.

2.3 Artigo 40

O Artigo 40 da Lei 6.830/80 é regido por parágrafos que tratam da suspensão do curso da execução, abertura de vista ao representante judicial da Fazenda Pública, arquivamento dos autos após um ano sem localização do devedor ou bens penhoráveis, e desarquivamento em caso de posterior localização.

2.4 Alvará

O alvará judicial é um documento gerado da solicitação de um requerente, com o objetivo de que determinado ato seja autorizado pelo juiz.

2.5 Sentença

A sentença é o pronunciamento em que o juiz encerra a fase de conhecimento do procedimento comum, ou seja, encerra o processo, analisando ou não a questão principal da ação.

2.6 Dívida Ativa

Conforme disposto na Lei nº 6.830, de 22 de setembro de 1980, sendo considerada Dívida Ativa da Fazenda Pública aquela classificada como tributária ou não tributária. Dessa forma, qualquer dívida relacionada a inadimplência tributária, fiscal, contratual envolvendo os bens públicos será considerado Dívida Ativa da Fazenda Pública. Destaca-se que a Dívida Ativa da União é apurada e registrada na Procuradoria da Fazenda Nacional.

2.7 Prescrição Intercorrente

Prescrição Intercorrente é a decisão que ordena o arquivamento, após decorrer o prazo prescricional, o juiz, após ouvir a Fazenda Pública, pode reconhecer de ofício e decretá-la imediatamente.

2.8 Execução Fiscal

A execução fiscal é um processo que pode ser instaurada contra o devedor conforme previsto na lei, para dívidas, tributárias ou não, de pessoas físicas ou jurídicas de direito privado, incluindo seus sucessores a qualquer título (Art. 4º Lei No 6.830). A petição inicial deverá indicar apenas o juiz a quem é dirigida, além do requerimento para a citação. Esta petição será instruída com a Certidão da Dívida Ativa, integrando-a como se estivesse transcrita. O despacho do juiz que defere a petição inicial autoriza a citação, seguindo as modalidades sucessivas de penhora, qual é visto que existe além da penhora, o arresto e sequestro, visto que todos são medidas diferentes de constrição judicial, que tem e causam os mesmos efeitos no processo.

2.9 Fluxo da Execução Fiscal

O fluxo da prescrição intercorrente, demonstrado na Figura 2 possui marcos processuais definidos, sendo estes as movimentações referentes a: citação, suspensão nos termos do Art. 40, penhora, alvará e sentença.

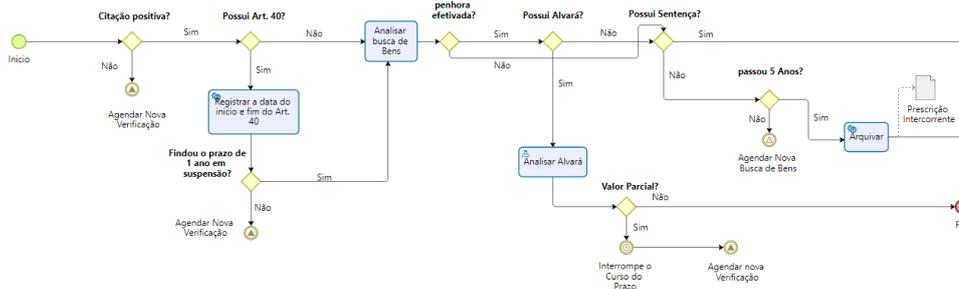


Figura 2: Fluxo da Prescrição Intercorrente em ordem prática.

Assim, o cálculo do prazo prescricional encontra-se diretamente ligado a ordem e ocorrência destes marcos, de modo que sua identificação permite a reconstrução de uma linha do tempo processual, em que é possível analisar o efeito de cada um. As movimentações possuem diferente efeitos em relação ao prazo prescricional, sendo estes: início, suspensão e reinício. Logo, entende-se que a citação positiva trata-se de uma movimentação com efeito de início, uma vez que a partir de sua existência inicia-se o prazo de cinco anos para a prescrição de um processo. Por sua vez, a existência de uma suspensão nos termos do Art.40 tem efeito suspensivo, como o próprio nome indica e aumenta em um ano o prazo de prescrição. Além disso a movimentação de penhora também possui caráter suspensivo, porém por prazo indeterminado, saindo de suspensão com a existência da movimentação de alvará. A movimentação de alvará, além de tirar o processo do efeito de suspensão confere o reinício para o cálculo processual, dessa forma, a partir de sua existência o tempo de contagem de cinco anos recomeça. A movimentação de sentença indica que o processo foi prescrito.

3 Proposta de Trabalho

A fim de estimar o prazo da prescrição interscorrente criou-se, então, um sistema capaz de cobrir este fluxo, com capacidade de prever a data de prescrição, bem como elencar os eventos que levaram a essa previsão. Este sistema utiliza modelos de aprendizado profundo treinados em problemas de processamento de linguagem natural, a fim de oferecer maior confiabilidade e respaldo às previsões. Além disso, o serviço pode funcionar como plataforma de coleta de feedbacks pelo rótulos dos usuários e auto-treinamento. Com essa funcionalidade, o modelo utilizado pode ser melhorado, a partir da rotulação de dados pelo usuário e assim refinado com o seu uso.

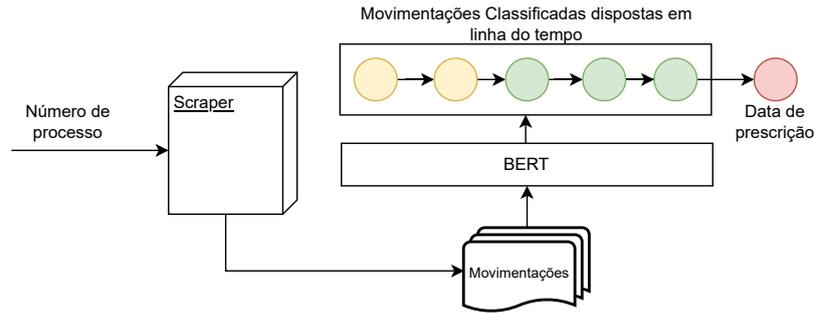


Figura 3: Fluxo de funcionamento do sistema

Conforme indicado na Figura 3, seu funcionamento se dá pela interação do cliente com uma interface gráfica (Seção 3.1), em que o usuário insere o número do processo que deseja analisar. A partir disso, o sistema executa um serviço Scrapper (Seção 3.2) que busca todas as movimentações deste processo. Posteriormente os textos provenientes destas movimentações são utilizados como entrada para um classificador textual, criado a partir de um modelo BERT pré-treinado [5], que tem seu treinamento e instanciação descritos na Seção 3.4. O resultado desse processo são textos classificados e organizados em uma linha do tempo. A partir disso, utiliza-se um algoritmo codificado com os conhecimentos levantados na seção 2 para realizar o cálculo do prazo prescricional. O resultado final é a data de prescrição estimada, em conjunto as movimentações do processo destacadas, que são elementos que definem o prazo prescricional. Na interface o usuário pode indicar um erro vinculado a classificação de uma movimentação, assim o prazo é recalculado com base nessa nova informação e este dado é coletado a fim de refinar o modelo.

Em termos de aprendizado de máquina supervisionado, as entradas do modelo são os textos das movimentações e as classes são: Citação, Alvará, Penhora Sentença/Arresto, Artigo 40, que foram explicados anteriormente e que definem os principais losangos ilustrados no fluxograma da Figura 2.

3.1 Interface Gráfica

Para que o usuário acesse as funções dos serviços e as funcionalidades do sistema em geral, criou-se uma interface gráfica. As telas foram divididas em: tela inicial e tela de visualização do processo.



Figura 4: Campo de inserção do processo na tela inicial

Como apresentado na Figura 4, a tela inicial possui um campo de *input* textual em que o usuário insere o número do processo.

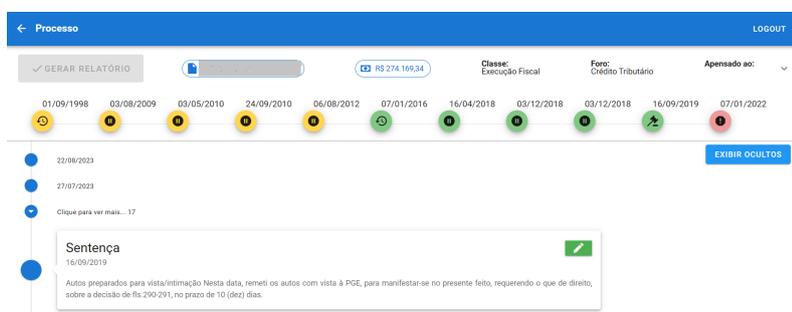


Figura 5: Tela de visualização do processo

A tela na Figura 5, demonstra a visão do processo dentro do sistema do Quati, sendo possível ver os dados do processo, como juiz ou vara que o processo está correndo. Também é apresentado o fluxo temporal das movimentações do processo, por exemplo, as movimentações onde que ocorreram penhoras ou sentenças.

Para facilitar o uso e entendimento do usuário, a tela de visualização é dividida em duas partes principais: *Timeline* horizontal(7) e *Timeline* vertical(6).



Figura 6: Timeline vertical



Figura 7: Timeline horizontal

A *timeline* vertical, Figura 6, apresenta para o usuário a lista com todas as movimentações retiradas do processo no site esaj.tjms.jus.br via Scraper (Seção 3.2). A *timeline* tem como função principal destacar as movimentações classificadas pelo Quati. Essas classificações podem ser alteradas com o mecanismo de *feedback*, representado pelo botão verde mostrado na Figura 6. Ao clicar no botão, o usuário é direcionado a tela ilustrada na Figura 8, onde tem a opção de corrigir a classificação da movimentação selecionada.

Na Figura 8 é exibida a tela de *feedback* do sistema, composta por campos de texto, em que um destes contem a movimentação selecionada, enquanto o outro é destinado a seleção da classificação adequada para a movimentação.

Já a *timeline* horizontal, Figura 7, apresenta o fluxo do processo de forma mais simples e usual, de forma a evidenciar as datas dos marcos e as decisões que ocorreram dentro do processo. O sistema coloca em evidência todas as movimentações definidas como importantes, mostradas na *timeline* vertical, e as coloca em ordem crescente pela data da ocorrência. Ao final de todas as movimentações o sistema calcula o prazo prescricional do processo em questão e o exhibe ao final da *timeline* horizontal.

Dessa forma, o usuário analisa a classificação do processo feita pelo Quati e, caso seja necessária uma correção, o mesmo pode retornar um *feedback* para o sistema.

Qual tipo de movimentação é a correta ?

Certidão Cartorária PJMS - Certidão de Registro de Sentença

Selecione um tipo de movimentação

Sentença

CANCELAR
ENVIAR

Figura 8: Tela de feedback

O usuário através do *feedback* corrige a classificação feita e a *timeline* horizontal é atualizada em tempo real, recalculando a possível data da prescrição. O *feedback* coletado é utilizado posteriormente para melhoras e novas rotinas de treinamento do modelo.

3.2 Raspagem de Dados

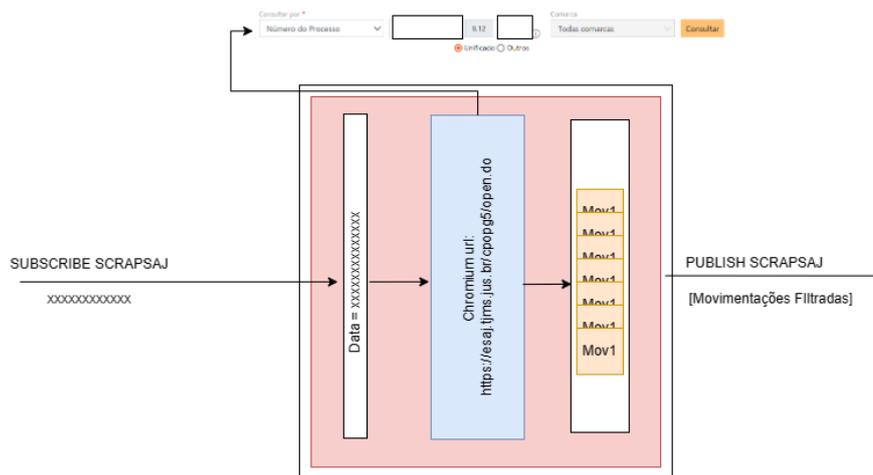


Figura 9: Fluxo Scrapper

As informações necessárias para alimentar o sistema são de domínio público, pertencentes ao Tribunal de Justiça do Estado do Mato Grosso do Sul e estão disponíveis em <https://esaj.tjms.jus.br/>. Porém, tais informações não estão agrupados em tabelas ou bancos de dados que possuam APIs ou *web services* públicos. Dessa forma, para construção do sistema, tanto em relação aos dados necessários para o treinamento do

modelo de inteligência artificial, quanto para a rotina de análise de novos processos fez-se necessário a criação de um módulo de raspagem de dados. Esta técnica consiste na extração automática de informações em páginas da internet, que anteriormente só seriam possíveis via acesso manual em um navegador[6].

Além disso, existe uma grande desvantagem relacionada, a possível detecção como mecanismo de automação, ficando suscetível a restrição de acesso pelo site em questão. Isso ocorre como forma de proteção dos responsáveis pelo sistema, a fim de evitar ataques maliciosos. Nesse sentido, torna-se de suma importância que o serviço escolhido não seja tão facilmente detectável como automação, evitando possíveis bloqueios. Verifica-se, então, a possibilidade de utilizar serviços de raspagem de dados sem a detecção dos sites alvos[7].

Assim, seu funcionamento necessita de um número de processo como entrada. A partir disso, uma instância do navegador Chromium é carregada em memória e acessa o site <https://esaj.tjms.jus.br/>, em seguida, uma busca é realizada no site com o valor enviado anteriormente como entrada. Após esses passos, a página carregada no navegador contém as informações necessárias ao sistema, denominadas movimentações, estas são extraídas por meio de buscas nos seletores CSS(Cascading Style Sheets) e HTML (HyperText Markup Language). Por fim, a informação coletada é enviada ao serviço de classificação descrito na Seção 3.5 contendo o modelo BERT.

3.3 Conjunto de Dados

Os dados deste domínio estão organizados de forma hierárquica, existindo processos e movimentações. Um processo é constituído por informações singulares como seu número de registro, valor de dívida, juiz responsável, bem como um grupo de movimentações. As movimentações são o que dão sentido a um processo e tratam-se de textos contendo decisões, pedidos, definições e informações. Dessa forma um processo e suas respectivas movimentações são organizados conforme a Figura 10.

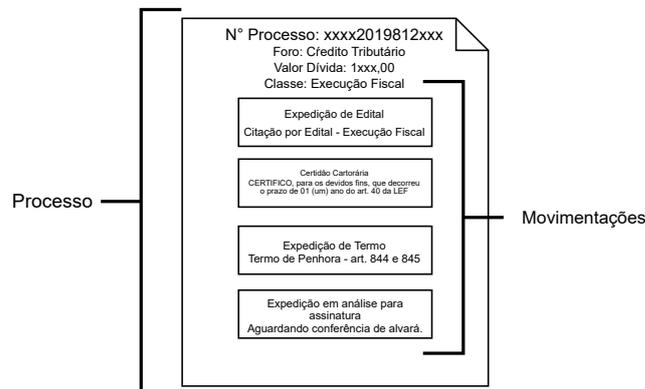


Figura 10: Organização Processo e Movimentações

A construção do conjunto de dados se deu pelo levantamento de números de processos relacionados a prescrição intercorrente, um número de processo é um identificador que permite a consulta de todas as movimentações relacionadas ao processo no site de um tribunal. A abordagem se deu pela busca de processos prescritos, pois estes possivelmente contém a existência de todas as movimentações relacionadas ao fluxo de prescrição. Para isso, solicitou-se exemplos de processo a Juíza responsável pela Vara de Execução Fiscal, a fim de construir a base de dados do sistema. Assim, 750 números de processos prescritos no Estado de Mato Grosso do Sul foram enviados. A partir disso, o sistema de Scraper (Seção 3.2) foi alimentado com esses números e suas respostas foram enviadas a um filtro. No filtro os textos recebidos eram submetidos a uma comparação buscando o radical das palavras citação, artigo 40, sentença, penhora, arresto, sequestro e alvará. Para a palavra citação por exemplo buscava-se a existência do radical "cit", de semelhante modo esse processo foi realizado nas demais palavras, caso positivo o texto era adicionado ao conjunto de dados juntamente ao número de processo do qual pertencia.

Posteriormente, os textos provenientes do procedimento anterior foram adicionados a uma planilha, assim, deu-se início ao processo de rotulação. Este processo consistiu na inserção de cinco novas colunas, denominadas: Alvará, Citação, Penhora/Arresto/-Sequestro, Artigo. 40 e Sentença, preenchidas com o valor numérico 0, caso o texto indicado possuísse o radical da palavra que nomeia a coluna, mas não fosse uma movimentação representativa para aquela classe, -1 para textos que nem possuíam o radical da palavra indicada na coluna, nem fossem uma movimentação representativa daquela

classe e 1 para movimentações representativas para a classe e que possuíam o radical da palavra que dá nome à coluna.

O banco de dados inicial apresenta 2299 movimentações divididas em seis classes: "Alvará", "Citação", "Penhora/Arresto/Sequestro", "Artigo. 40", "Sentença" e "Não relevantes". Os textos foram rotulados por 7 pesquisadores, com o auxílio de 2 especialistas da área durante o período de 17/02/2023 a 06/03/2023. Cada processo foi rotulado por uma única pessoa, sendo pesquisador ou especialista, e posteriormente o rótulo adicionado foi analisado por um dos especialistas. Nota-se, porém um certo desequilíbrio na distribuição de classes, enquanto a classe "Alvará" apresenta 31 exemplos a classe "Não Relevantes" apresenta 1362, isso ocorre graças à natureza do problema em si, uma vez que a ocorrência de determinadas movimentações é naturalmente maior que outras. Assim, a distribuição do conjunto de dados original seguiu como na Tabela 1.

Tabela 1: Distribuição da base de dados original

Classe	Rótulo	Quantidade
Alvará	0	31
Sentença	1	144
Penhora/Arresto	2	154
Citação	3	166
Artigo 40	4	442
Não Relevantes	5	1362
Total		2299

As classes aparecem de forma variada nas movimentações de um processo, observa-se como exemplo de cada uma:

- Alvará: "Expedição de Alvará de Soltura"
- Sentença: "Registro de Sentença"
- Penhora/Arresto: "Expedição de Termo; Termo de Penhora de bem móvel"
- Citação: "Expedição de Edital; Citação por Edital - Execução Fiscal"
- Artigo 40: "Certidão Cartorária CERTIFICO, para os devidos fins, que em 13/01/2020 decorreu o prazo de 01 (um) ano do art. 40 da LEF. Certifico, ainda, que estes autos serão remetidos ao arquivo provisório pelo prazo da prescrição quinquenal, nos termos do art. 40, §2º da LEF, conforme r. Despacho de f. 538."
- Não Relevantes: "Despacho Interlocutório Vistos, etc. Tendo em vista que não transcorreu o prazo de 05 (cinco) anos desde a suspensão requerida pela Fazenda Pública Estadual (f. 233), retornem-se os autos ao arquivo provisório. Cumprindo o prazo prescricional, nos termos do § 4º do Artigo 40 da Lei de Execução Fiscal, manifeste-se a Fazenda Pública em 05 (cinco) dias, após voltem conclusos para sentença. Intime-se."

3.4 Treinamento e Exportação de Modelos

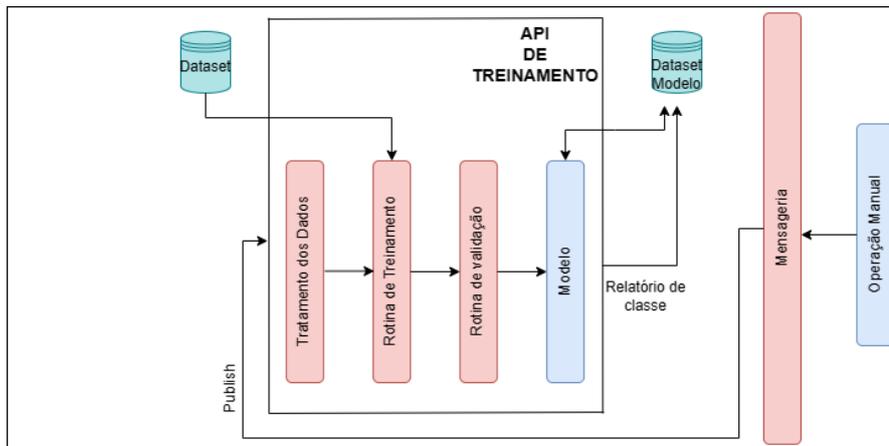


Figura 11: Fluxo de funcionamento da API de Treinamento

3.4.1 Tratamento dos dados

Os dados utilizados para treinamento são provenientes de duas fontes, a primeira sendo o conjunto de dados inicial (Seção 3.3), enquanto a segunda refere-se à coleção "feedbacks" existente no banco de dados, conforme Seção 3.7. Um *feedback* é um dado proveniente diretamente dos usuários, sendo constituído por uma movimentação relacionada a uma classe. Durante o processo de *feedback* um usuário pode escolher entre atribuir a uma movimentação uma das classes existentes e previamente mapeadas (Seção 2), ou criar uma nova classe. Para que uma nova classe seja adicionada ao conjunto de dados é necessário que existam, no mínimo, 50 exemplos desta.

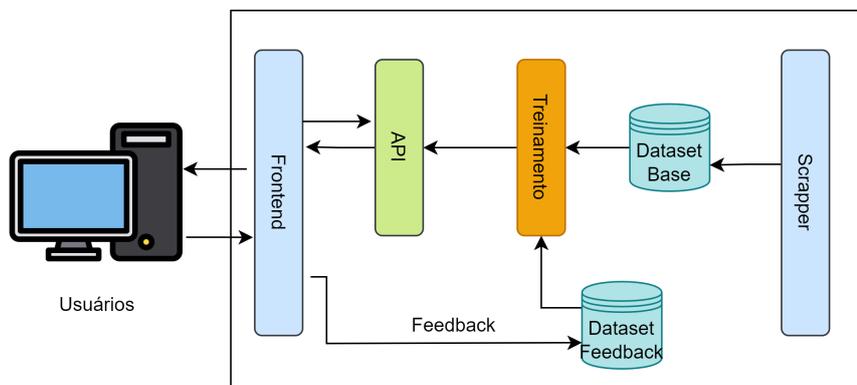


Figura 12: Fluxo do tratamento de dados

Os dados estão distribuídos no formato movimentação, que armazena o texto a ser classificado e rótulo, que indica a classe da qual se trata aquela movimentação. Para exemplos provenientes da coleção "feedbacks" do banco de dados o rótulo é adicionado diretamente com a classe informada pelo usuário. Como saída do procedimento de tratamento de dados, obtêm-se um novo conjunto de dados, mesclando o original com *feedbacks* de usuário, além disso exporta-se também um arquivo de texto informando quais as classes serão utilizadas na rotina de treinamento (Seção 3.4.4). Além disso, aqui ocorre a divisão do conjunto de dados em três conjuntos, sendo estes treino, teste e validação, em que 70% da base original vai para a parte de treino 9% para validação e 21% para teste. A partição obedece o processo de estratificação, ou seja, cada uma das partes contém uma distribuição proporcional a original. Além disso, a seleção é feita de forma aleatória, mas com um número fixo de geração, garantindo que o procedimento seja reproduzível.

3.4.2 Modelo

A partir da publicação de Vaswan *et al.*[8], o campo de NLP (Natural Language Processing) passou por uma grande mudança de paradigmas. A nova arquitetura proposta, baseada unicamente em mecanismos de atenção ganhou rápida popularidade e expandiu as perspectivas para tarefas relacionadas a NLP. Dentre os principais avanços, observa-se a criação de redes pré-treinadas capazes de atingir o estado da arte apenas com ajuste fino[5], bem como a simplicidade para exportar e gerenciar esses modelos.

3.4.3 Escolha do Modelo

A fim de selecionar o melhor modelo base para a tarefa de treinar um classificador textual de dados jurídicos em português levantou-se os possíveis candidatos, sendo estes: *bert-base-multilingual-cased*, um modelo com 179 milhões de parâmetros treinado com textos do wikipedia em 104 idiomas [5] e o *bertimbau-base*, um modelo treinado exclusivamente em português brasileiro com 110 milhões de parâmetros [9]. Utilizando o conjunto de dados descrito em 3.3, com o otimizador AdamW[10] realizou-se um comparativo, como ilustrado nas Tabelas 2 e 3

Tabela 2: Resultados modelo bert-base-multilingual-cased

	Precision	Recall	F1-score	Support
Alvará	0.60	0.86	0.71	7
Sentença	0.97	0.93	0.95	30
Penhora/Arresto	1.00	0.91	0.95	32
Citação	0.94	0.89	0.91	35
Art.40	0.93	0.98	0.95	93
Não Relevante	0.96	0.95	0.96	286
accuracy			0.95	483
macro avg	0.90	0.92	0.90	483
weighted avg	0.95	0.95	0.95	483

Tabela 3: Resultados modelo Bertimbau

	Precision	Recall	F1-score	Support
Alvará	0.50	0.57	0.53	7
Sentença	0.97	0.97	0.97	30
Penhora/Arresto	1.00	0.88	0.93	32
Citação	0.91	0.89	0.90	35
Art.40	0.94	0.95	0.94	93
Não Relevante	0.95	0.96	0.95	286
accuracy			0.94	483
macro avg	0.88	0.87	0.87	483
weighted avg	0.94	0.94	0.94	483

Ao observar os resultados de *F1-score* nota-se alta semelhança, com exceção a classe alvará que apresenta uma diferença evidente. De maneira geral o modelo *bert-base-multilingual-cased* apresentou melhores resultados e a classe alvará foi melhor representada.

3.4.4 Rotina de Treinamento

Trata-se de uma função responsável por treinar e exportar um modelo de inteligência artificial baseado nas métricas descritas na Seção 3.4.3. Primeiramente, instancia-se o modelo selecionado como classificador, utilizando o número de classes proveniente da função de tratamento de dados, conforme mostrado na Figura 11. Em seguida os conjuntos de treino e validação são utilizados para alimentar o modelo. Na rotina os pesos do modelo e do otimizador são atualizados e medidos no conjunto de validação, caso haja melhora a rotina segue, do contrário em um número limitado de tentativas mal sucedidas a rotina é encerrada e o modelo com melhor avaliação é retornado.

3.4.5 Rotina de Testes e Exportação

A fim de validar o modelo treinado na Seção 3.4.4, desenvolveu-se uma função para avaliar o modelo no conjunto de teste, esse conjunto apresenta dados que ainda não foram vistos pelo modelo nas etapas de treino. Assim, seu funcionamento consiste avaliar as respostas com as métricas precisão, revocação e média harmônica neste conjunto. Por fim, o modelo avaliado é armazenado juntamente ao de dados utilizado para seu treinamento e um arquivo no formato JSON (*JavaScript Object Notation*) com o nome das classes utilizadas durante esta etapa.

3.5 Classificação

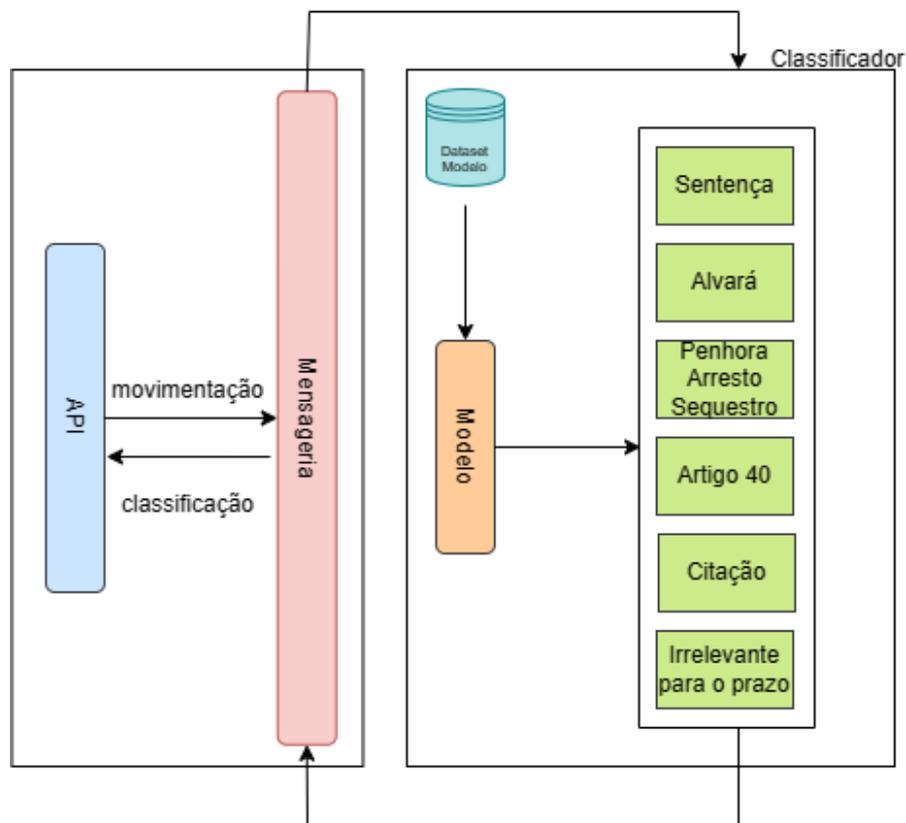


Figura 13: Diagrama do Classificador

O serviço de classificação exerce papel central no funcionamento do sistema. Seu funcionamento se dá pela instanciação de modelos previamente treinados (Seção 3.4) que passam a receber movimentações provenientes do serviço de raspagem de dados (Seção 3.2) como entrada. Em seguida, o modelo realiza a classificação de cada uma das movimentações, que posteriormente são organizadas na interface gráfica. O resultado dessa rotina serve de entrada para um algoritmo que permite o cálculo do prazo prescricional.

É necessário ressaltar que esta etapa pode se tornar um grande gargalo em nível arquitetural e de custos para um projeto, pois, é comum que outras abordagens necessitem de mais de um modelo carregado em memória, aumentando o custo computacional. Além disso, o tempo de predição pode variar a depender da arquitetura escolhida. Torna-se prudente buscar formas de otimização para esta rotina. No sistema QuaTI, adotou-se a medida de enviar movimentações filtradas aos classificadores, ou seja, uma

movimentação só é enviada para tarefa de classificação caso contenha em seu corpo o radical de alguma palavra que nomeia classes alvo, descritas na Seção 3, uma vez que é improvável a ocorrência de um marco sem que estas palavras estejam presentes.

3.6 Cálculo da Prescrição

O cálculo do prazo prescricional segue um conjunto de regras muito bem estabelecidas. A partir da identificação dos marcos processuais, descritos na Seção 2.9, torna-se possível a criação de um serviço a fim de realizar este cálculo. Logo, se faz necessário a transformação dos conceitos jurídicos outrora explicitados em decisões lógicas. Dessa forma a aplicação QuaTI, não se utiliza do processamento de linguagem apenas como objetivo fim, mas também como um passo intermediário, que permite em conjunto ao conhecimento humano na tarefa de prescrição intercorrente a criação de um algoritmo para predição de datas estimadas. Vale ressaltar, que um usuário interage diretamente com o algoritmo de cálculo prescricional ao realizar o processo de *feedback*, descrito na Seção 3.1, pois o cálculo é imediatamente atualizado conforme a entrada do usuário.

3.7 QuaTI: Arquitetura Completa

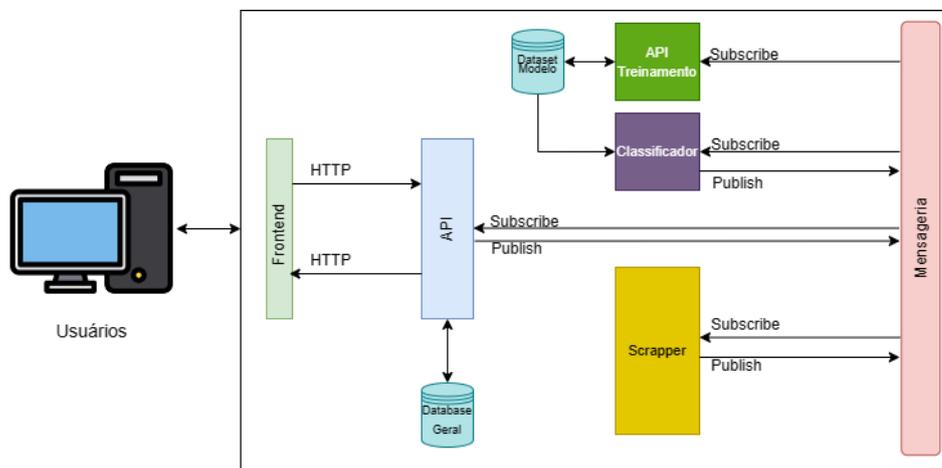


Figura 14: Diagrama do fluxo geral do sistema

O sistema Quati é uma ferramenta de análise processual construída a partir do mapeamento de uma rotina específica, a prescrição intercorrente. Porém, sua construção foi planejada de modo a permitir comutação de módulos, a fim de que fosse possível escalar o sistema, bem como alterar o enfoque das análises. Além disso, trata-se de uma ferramenta voltada ao usuário final, sendo assim desejável a existência de uma interface gráfica agradável e um serviço veloz, capaz de oferecer boa usabilidade ao usuário.

Nesse sentido, adotaram-se decisões arquiteturais que visassem a interoperabilidade do sistema, bem como seu desempenho. Assim, optou-se por uma arquitetura orientada a eventos, utilizando como fundamento central um serviço de mensagens. Esta abordagem permite ao sistema escalabilidade e o desacoplamento em nível temporal, espacial e de sincronização [11].

Além disso, outra decisão do projeto é a utilização de micro serviços, que são módulos desenvolvidos e implantados individualmente, possuindo baixo acoplamento entre si. Espera-se com isso melhoras em relação ao processo de desenvolvimento, bem como melhor eficiência no uso de recursos do servidor [12].

A alteração de um módulo do sistema necessita apenas atender os critérios de comunicação especificados com os demais serviços. Logo, além da eficiência desejada a arquitetura deste sistema oferece a possibilidade de expansão do escopo atual.

4 Resultados

O produto criado é um serviço capaz de retornar um processo com as movimentações relacionadas a prescrição intercorrente devidamente destacadas, conforme as regras descritas na Seção 2. Além disso, possui também, a capacidade de calcular a possível data de prescrição de um processo. Ao utilizar o sistema em 104 processos, constatou-se que as 21500 movimentações neles existentes foram classificadas pelo modelo em 1042 movimentações com classes que definem o prazo prescricional. Assim um usuário não necessita analisar as 21500 movimentações, mas apenas as 1042 identificadas pelo sistema. Desse modo, o número de movimentações a serem analisadas é reduzido em mais de vinte vezes, sendo o tempo médio de análise para cada um desses processos de oito segundos. Assim, observa-se um impacto direto na metodologia de trabalho, lançando novas perspectivas sobre a capacidade produtiva destacadas na Seção 1.

Além disso, a função de *feedbacks* e os demais serviços acoplados tornam o sistema facilmente adaptável a resolução de novos problemas textuais relacionados ao contexto jurídico, de modo que o QuaTI possui a capacidade de funcionar como um *framework* para tarefas de classificação e rotulação do mesmo domínio.

É necessário destacar, também, como a arquitetura em micro serviços, combinada a orientação a eventos tornou-se uma grande vantagem ao longo do projeto, considerando que ao trabalhar com um domínio pouco explorado e modelos de inteligência artificial é comum o abandono, ou substituição de alguns serviços, assim, a combinação da comunicação de um serviço de mensagens ligado externamente aos demais permitiu a escalabilidade e os desacoplamento desejados.

4.1 Teste entre classificador multi classe e classificador padrão

Na primeira tentativa de indução de um modelo, foi treinado um classificador binário para cada uma das classes, pois seria possível alterar o sistema para uma abordagem multi classe caso fosse necessário. Assim, em conjunto a arquitetura de micro serviços criava-se a necessidade de cinco módulos classificadores, como os descritos na Seção 3.5. A abordagem em micro serviços permitiria que recursos fossem alocados de acordo com a utilização de cada um destes, porém esse ainda era um procedimento custoso do ponto de vista computacional. Dessa forma, buscando otimizar o uso de memória, bem

como a facilitação de comunicação dentro do sistema foram realizados experimentos com a abordagem de um classificador multi classe, ou seja, não existiriam mais seis modelos retornando positivo ou negativo, mas sim um único modelo que retornaria a qual classe pertence a movimentação de entrada. Nesse sentido, realizaram-se testes comparativos, em rotinas de treinamento mais simples utilizando o número de épocas fixo em dez e comparando os resultados, assim, a base de dados, também, foi dividida apenas em treino e teste. A partir disso verificou-se a possibilidade de construir serviços com uma eficiência similar, porém com uso de recursos muito menor.

As métricas levantadas na comparação estão dispostas nas Tabelas 4, 5 e 6, onde as Tabelas 4 e 5 representam o desempenho em cada uma das classes e a métrica geral no classificador multi classe. Por sua vez a Tabela 6 refere-se as métricas dos classificadores binários. Nota-se alta semelhança em relação aos valores de *F1-score* nas Tabelas 4 e 5, a principal diferença refere-se a classe Alvará em que o classificador multi classe tem desempenho relativamente inferior. Porém, este resultado indicou a possibilidade de substituição dos classificadores binários para um classificador multi classe, uma vez que é possível utilizar técnicas para melhorar o resultado da classe minoritária.

Tabela 4: Métricas de Desempenho classificador multi classe

Classe	Precisão	Revocação	F1-Score	Suporte
Alvará	1.00	0.33	0.50	9
Sentença	0.98	0.94	0.96	48
Arresto/Penhora	0.98	0.93	0.95	57
Citação	0.90	0.93	0.91	46
Art.40	0.97	0.91	0.94	128
Não Influencia no cálculo	0.95	0.99	0.96	402

Tabela 5: Métricas Globais

Métrica	Precisão	Revocação	F1-Score	Acurácia
Macro Avg	0.96	0.84	0.87	-
Weighted Avg	0.95	0.95	0.95	0.95

Tabela 6: Métricas Globais Classificador Padrão

Classe	Métrica	Precisão	Revocação	F1-Score
Citação	Macro Avg	0.93	0.95	0.94
Sentença	Macro Avg	0.90	0.90	0.90
Alvará	Macro Avg	0.92	0.69	0.73
Artigo.40	Macro Avg	0.93	0.95	0.94
Arresto	Macro Avg	0.89	0.78	0.82
Penhora	Macro Avg	0.94	0.96	0.95

5 Discussão

Atualmente, constata-se que a área de inteligência artificial aplicada ao contexto jurídico ainda está em desenvolvimento, tendo um crescimento em potencial recentemente. Assim, como explicitado na Seção 1 existem diversas abordagens para esta temática, porém, estas em sua totalidade trazem o processamento com inteligência artificial como objetivo fim de seus sistemas. Além disso, os trabalhos de Barros *et al.*[1], Silva *et al.*[2], Sleimi *et al.*[3], e Aletras *et al.*[4] atualmente utilizam técnicas como *Word2Vec*. Este trabalho aplica modelos baseados em mecanismos de atenção, que possuem resultado mais próximo ao estado da arte.

Destaca-se, também, a apresentação de um conjunto de dados inédito com grande número de exemplos, um entrave em diversas publicações existentes. É necessário salientar a assertividade dos modelos utilizados, uma vez que a acurácia geral é de 95% nas tarefas propostas.

6 Conclusão

Conclui-se, portanto, que as decisões arquiteturais e testes com usuários ao longo da construção do sistema permitiram a criação de um produto capaz de atender todas as demandas levantadas inicialmente, ou seja, o QuaTI possui a capacidade de emular o comportamento humano na rotina de análise processual vinculada a processos inscritos na Dívida Ativa, conforme descrito na Seção 2.9.

Além disso, é necessário também destacar o sucesso vinculado a modelos "*Transformers*" para tarefas deste domínio, é possível notar desde os primeiros testes que a assertividade nas tarefas propostas é alta, de modo que tornou-se difícil qualquer melhora em relação aos modelos padrões.

É possível observar, também, o pioneirismo relacionado a criação deste sistema, sendo possível que a aplicação QuaTI sirva como forma de expansão deste domínio.

Referências

- [1] Barros, R., Peres, A., Lorenzi, F., Krug Wives, L., Silva Jaccottet, E.: Case law analysis with machine learning in brazilian court. In: Recent Trends and Future Technology in Applied Intelligence: 31st International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2018, Montreal, QC, Canada, June 25-28, 2018, Proceedings 31, pp. 857–868 (2018). Springer
- [2] Silva, A.C., Maia, L.C.G.: The use of machine learning in the classification of electronic lawsuits: An application in the court of justice of minas gerais. In: Brazilian Conference on Intelligent Systems, pp. 606–620 (2020). Springer
- [3] Sleimi, A., Sannier, N., Sabetzadeh, M., Briand, L., Dann, J.: Automated extraction of semantic legal metadata using natural language processing. In: 2018 IEEE 26th International Requirements Engineering Conference (RE), pp. 124–135 (2018). IEEE

- [4] Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ computer science* **2**, 93 (2016)
- [5] Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bi-directional transformers for language understanding. *CoRR* **abs/1810.04805** (2018)
- [6] Singrodia, V., Mitra, A., Paul, S.: A review on web scrapping and its applications. In: *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6 (2019). IEEE
- [7] Färholt, F.: *Less Detectable Web Scraping Techniques* (2021)
- [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
- [9] Souza, F., Nogueira, R., Lotufo, R.: Bertimbau: pretrained bert models for brazilian portuguese. In: *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I* 9, pp. 403–417 (2020). Springer
- [10] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017)
- [11] Eugster, P.T., Felber, P.A., Guerraoui, R., Kermarrec, A.-M.: The many faces of publish/subscribe. *ACM computing surveys (CSUR)* **35**(2), 114–131 (2003)
- [12] Surantha, N., Utomo, O.K., Lionel, E.M., Gozali, I.D., Isa, S.M.: Intelligent sleep monitoring system based on microservices and event-driven architecture. *IEEE Access* **10**, 42069–42080 (2022) <https://doi.org/10.1109/ACCESS.2022.3167637>

A Apêndice

A.1 Experimento de OverSample com dados textuais

Observa-se que um dos grandes desafios relacionados ao projeto consiste no desbalançamento do conjunto de dados, uma vez que em uma tarefa de classificação classes minoritárias tendem a ser diretamente afetadas. Nesse sentido, uma das possíveis abordagens para solução deste entrave é a técnica de OverSample, que consiste em aumentar artificialmente o número de exemplos das classes minoritárias a fim de possíveis melhoras no desempenho do modelo. Nota-se porém, a escassez relacionada a existência de técnicas dessa natureza para dados textuais, em contrapartida a técnicas já consolidadas em relação a dados tabulares. Além disso, existe outra problemática relacionada a criação de novos exemplo, pois, tratando-se de movimentações jurídicas faz-se necessário que a linguagem utilizada seja altamente técnica, porém o sentido

original das sentenças precisa ser mantido. Assim, três abordagens foram elencadas, sendo estas:

- Troca randômica por sinônimos
- Utilização de rede Bert, para gerar sentenças com fillMask
- Criação de novos textos com modelo generativo Mistral 7B

É importante ressaltar que para validar a eficiência dos experimentos utilizou-se o modelo selecionado na seção 3.4.3 e que o conjunto de dados descrito na seção 3.3 foi expandido exclusivamente no conjunto de treino, mantendo os conjuntos de validação e teste inalterados. É necessário, também, salientar que a classe majoritária não foi alterada e as demais tiveram novos exemplos criados a fim de atingir uma distribuição homogênea.

Para a primeira abordagem, denominada como troca randômica por sinônimos primeiro fez-se necessário a busca por bibliotecas e base de dados que permitissem a implementação da rotina. Levantou-se então a existência da biblioteca SpaCy (biblioteca *python open-source* para tarefas NLP), em conjunto a uma base textual disponível na biblioteca NLTK (Natural Language Toolkit), amplamente difundida em problemas NLP. Com isso, este processo consistiu em selecionar aleatoriamente palavras úteis, como artigos, verbos e adjetivos dentro de uma movimentação e troca-los por seus sinônimos indicados na biblioteca em questão. Esperava-se uma possível melhora, em relação a classe "Alvará", que possuía a menor quantidade de exemplos. O resultado, demonstrado na Tabela 7, indica que esta técnica, pouco alterou o resultado anterior, piorando as métricas em relação a classe minoritária.

De semelhante modo, o segundo processo consistia em criar *tokens* [MASK] randômicos substituindo palavras dentro de uma frase e passá-la como entrada para um modelo BERT pré-treinado[5], no modo *Fill Mask*. Nesse modo o modelo é capaz de prever 5 possíveis palavras para completar o *token*. Porém, como os textos de entrada do sistema são relativamente diferentes daqueles em que o modelo foi pré-treinado realizou-se um processo de *fine-tuning*[5] com textos relativos ao domínio. Para isso, foi utilizado o serviço de *Scraper*, descrito na Seção 3.2, porém diferente do processo descrito na seção 3.3, aqui, aplicou-se o mesmo filtro de forma contrária, então, criou-se uma base de dados com todas as movimentações descartadas anteriormente. Essa base foi utilizada para o processo de *fine-tuning* e o índice "perplexidade" que avalia o desconhecimento do modelo em relação ao novo vocabulário foi medido. Por fim, este processo indicou melhora em relação ao aprendizado do modelo em relação ao contexto jurídico. Logo, este foi colocado no modo "*fill-mask*", a fim de que novas frases fossem criadas a partir da predição do modelo. Com isso, esperava-se que as novas sentenças criadas fossem gramaticalmente diferente das originais, mas que o sentido fosse mantido. Os resultados expressos na Tabela 8, indicam que o processo não apresentou a melhora desejada.

Por fim, para a última abordagem, instanciou-se um modelo de texto generativo Mistral quantizado, com 7 bilhões de parâmetros em sua versão "instrução", capaz de interpretar comandos, a fim de que fosse possível gerar novos textos a partir dos existentes. Porém, não foi possível seguir essa abordagem uma vez que o modelo não apresentava respostas consistentes, por vezes alterando até o idioma da sentença.

Tabela 7: Resultados modelo pós Oversample com sinonimos

	Precision	Recall	F1-score	Support
0	0.57	0.57	0.57	7
1	0.97	0.93	0.95	30
2	0.97	0.88	0.92	32
3	0.97	0.89	0.93	35
4	0.94	0.97	0.95	93
5	0.95	0.97	0.96	286
accuracy			0.95	483
macro avg	0.89	0.87	0.90	483
weighted avg	0.95	0.95	0.95	483

Tabela 8: Resultados modelo pós Oversample com fillMask

	Precision	Recall	F1-score	Support
0	1.00	0.43	0.60	7
1	1.00	0.90	0.95	30
2	1.00	0.84	0.92	32
3	0.97	0.89	0.93	35
4	1.00	0.84	0.91	93
5	0.91	1.00	0.95	286
accuracy			0.94	483
macro avg	0.98	0.82	0.88	483
weighted avg	0.94	0.94	0.93	483