

TECNOLOGIA DE MINERAÇÃO DE TEXTO NO DIREITO: O CASO DA IDENTIFICAÇÃO DE ENDEREÇOS EM PROCESSOS DE USUCAPIÃO

FLÁVIO JOSÉ FERREIRA JUNIOR¹

¹ TJPE / Cesar School – flavio.junior@tjpe.jus.br

RESUMO – O artigo explora o uso de mineração de texto no setor jurídico, com foco na identificação de endereços em processos de usucapião. A pesquisa utiliza o Processamento de Linguagem Natural (PLN) para estruturar informações extraídas de petições iniciais, otimizando a análise documental. Com 10 petições iniciais extraídas da base de dados do TJPE, analisadas no modelo GPT-4.o-mini, foi possível extrair dados completos em 80% dos casos, evidenciando a eficiência e apresentando as limitações da tecnologia.

No contexto jurídico, destaca-se a relevância da usucapião para a regularização fundiária e a necessidade de automação no Processo Judicial Eletrônico (PJE). A aplicação de IA pode reduzir o tempo de tramitação de processos e contribuir para a eficiência jurisdicional, com isso alinhando-se às diretrizes do Conselho Nacional de Justiça (CNJ) para a modernização da Justiça.

ABSTRACT – This paper explores the use of text mining in the legal sector, focusing on the identification of addresses in adverse possession cases. The research employs Natural Language Processing (NLP) to structure information extracted from initial petitions, optimizing document analysis. Using 10 initial petitions sourced from the TJPE database and analyzed with the GPT-4.o-mini model, complete data extraction was achieved in 80% of cases, demonstrating the technology's efficiency while also highlighting its limitations.

In the legal context, the study underscores the significance of adverse possession for land regularization and the necessity of automation in the Electronic Judicial Process (PJe). The application of AI can reduce case processing time and enhance judicial efficiency, thereby aligning with the National Justice Council's (CNJ) guidelines for modernizing the judiciary.

1 INTRODUÇÃO

O avanço acelerado das tecnologias de inteligência artificial (IA) tem transformado profundamente os processos de trabalho em diversas áreas do conhecimento humano, promovendo maior eficiência, automação e precisão. No setor jurídico, marcado pela intensa produção de documentos textuais e informações não estruturadas, a aplicação de IA apresenta um potencial significativo para otimizar a análise e o processamento de dados, reduzindo atrasos e aumentando a eficácia das decisões.

Entre as tecnologias de IA, destaca-se o processamento de linguagem natural (PLN), definido por Guimarães (2019) como "um campo interdisciplinar que se concentra em permitir que computadores compreendam, interpretem e gerem linguagem humana". Essa tecnologia permite que máquinas analisem e interpretem textos de maneira semelhante ao raciocínio humano, alinhando-se ao conceito de interação homem-máquina delineado por Alan Turing (1950). A mineração de texto, uma aplicação prática do PLN, tem se mostrado uma ferramenta poderosa para extrair padrões, estruturar informações e gerar insights a partir de grandes volumes de dados textuais.

A área jurídica enfrenta desafios significativos no processamento de documentos como petições, contratos e decisões judiciais. Nesse contexto, a mineração de texto aplicada ao direito facilita a identificação de padrões, a extração de informações e a classificação de processos. Um exemplo prático é o sistema Elis, desenvolvido pelo Tribunal de Justiça de Pernambuco (TJPE), que automatiza a análise de certidões de dívida ativa (CDAs) e petições iniciais. Segundo o TJPE (2025), o Elis reduz prazos de análise de 18 meses para apenas 15 dias, evidenciando sua eficiência e impacto positivo na prestação jurisdicional.

Além disso, o TJPE (2015) destaca diversas aplicações da IA no setor jurídico, como triagem de processos judiciais, previsão de reincidência criminal, análise de contratos e identificação de advocacia predatória. Essas ferramentas não

apenas melhoram a produtividade e a celeridade, mas também liberam os operadores do direito para se concentrarem em atividades estratégicas.

O processo judicial é o instrumento pelo qual um cidadão (ou grupo de pessoas) requer formalmente ao Poder Judiciário que julgue determinada causa.

Atualmente o processo judicial tramita de forma eletrônica pelo sistema Processo Judicial Eletrônico (PJE), ferramenta desenvolvida pelo Conselho Nacional de Justiça (CNJ), utilizado em 13 tribunais estaduais. No entanto, esse sistema ainda apresenta limitações importantes no que diz respeito à estruturação dos dados de identificação do endereço dos imóveis, impossibilitando a automação de atividade e decisões, especialmente em processos de usucapião.

As informações relacionadas ao endereço são inseridas no texto da petição inicial do processo de usucapião. Essa falta de organização dificulta o acesso rápido e preciso a dados, condição essencial para a análise de dados e automatização de processos operacionais.

Este artigo explora o potencial da mineração de texto para analisar petições iniciais em processos judiciais de usucapião que tramitam pelo PJE, com foco na identificação e estruturação de informações sobre o endereço do imóvel.

O objetivo é demonstrar como essa tecnologia pode contribuir para a análise de dados, contribuindo para a tomada de decisões estratégicas, ações que visem a celeridade dos processos, como por exemplo mutirões de regularização fundiária, contribuindo para a melhora da qualidade do serviço jurisdicional, promovendo a eficiência operacional e subsidiando os auxiliando magistrados com informações mais assertivas.

Vale salientar que não é a pretensão desse artigo propor a utilização de tecnologias de IA em substituição a decisão do magistrado, mas avaliar o uso de tecnologias de inteligência artificial como apoio ao processo jurisdicional.

Destacamos também a iniciativa em desenvolvimento pelo Supremo Tribunal Federal (STF), que utiliza tecnologias de mineração de texto para otimizar o processamento de informações judiciais. Segundo o Conselho Nacional de Justiça (CNJ, 2024), o projeto envolve uma inteligência artificial capaz de receber dados de processos judiciais e gerar resumos contendo informações das instâncias percorridas pelo processo.

A modernização do sistema judicial é uma prioridade do Conselho Nacional de Justiça (CNJ), que destaca a eficiência como um fator essencial para atender à crescente demanda por Justiça no Brasil.

2 METODOLOGIA

2.1 Fundamentação Teórica

A composição de uma peça jurídica consiste, tradicionalmente, na demanda inicial e nos dispositivos jurídicos anexos ao processo. Durante décadas, esses documentos foram elaborados e organizados fisicamente, em papel. Com a informatização dos processos judiciais, essa dinâmica foi transposta para o meio digital, mantendo, no entanto, o mesmo *modus operandi*. A anexação de documentos passou a ser realizada virtualmente, transformando o processo eletrônico em um registro de dados predominantemente semiestruturados.

Apesar de facilitar a gestão logística, esse modelo apresenta limitações significativas, especialmente no que se refere à análise de dados processuais. A ausência de estruturação impede a extração eficiente de informações relevantes que poderiam melhorar a celeridade processual e a eficácia das decisões judiciais.

Nesse contexto, a mineração de dados se mostra uma ferramenta promissora. Ao estruturar informações de processos judiciais, possibilita a identificação de padrões e tendências, auxiliando na formulação de políticas públicas e na tomada de decisões com maior grau de assertividade. Sua aplicação no Direito, especialmente no campo da regularização fundiária, é estratégica para otimizar operações e atender às demandas crescentes por justiça eficiente e acessível.

Em destaque são apresentadas as tecnologias empregadas neste trabalho, bem como o conceito de usucapião para melhor entendimento da discussão envolta no artigo.

2.2 Mineração de Texto e Processamento de Linguagem Natural

A mineração de texto pode ser definida como o processo ou prática de examinar grandes coleções de recursos escritos para gerar novas informações. Segundo a definição apresentada pelo *Oxford Dictionary*, trata-se de “o processo ou prática de examinar grandes coleções de recursos escritos para gerar novas informações”. Essa técnica visa identificar informações relevantes contidas em textos, analisando padrões linguísticos e transformando-os em dados estruturados.

De acordo com Fernandes (2019), o Processamento de Linguagem Natural (PLN; em inglês, *Natural Language Processing*) é uma área de estudo da Ciência da Computação que investiga as capacidades e limitações das máquinas em compreender a linguagem humana. O PLN combina conhecimentos das áreas de Computação e Linguística, com o

propósito de, por meio do estudo do funcionamento do cérebro humano, aproximar os computadores da linguagem natural dos seres humanos.

O Processamento de Linguagem Natural (PLN) utiliza, segundo Caelen e Blete (2024), uma classe de algoritmos conhecida como aprendizado profundo (*deep learning*, em inglês). Esses algoritmos, inspirados na estrutura do cérebro humano, são chamados de redes neurais artificiais.

O PLN é um componente essencial na mineração de texto, realizando análises linguísticas que permitem às máquinas "lerem" textos de forma eficiente. Essa área utiliza metodologias diversas para lidar com as ambiguidades inerentes à linguagem humana, como sumarização automática, marcação de classes gramaticais, desambiguação, extração de entidades e de relações, além da compreensão e reconhecimento de linguagem natural (expert.ai, 2024).

O desafio da programação de linguagem natural é a ambiguidade dos termos linguísticos e a diversificação termos utilizados para expressar o mesmo significado. Não obstante a utilização de expressões em latim, tão comum no jargão jurídico.

2.2.1 GPT-4

Entre as ferramentas modernas que aplicam conceitos avançados de PLN está o GPT-4, desenvolvido pela OpenAI. Segundo Stefanovskyi (2024), o GPT é um *Large Language Model* (LLM) projetado para entender e gerar textos semelhantes à linguagem humana com base em entradas fornecidas. Demonstra grande utilidade no processo de mineração de texto.

Segundo Caelen (2024):

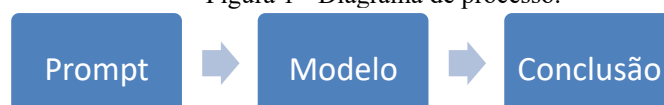
O GPT-4 foi baseado em um tipo específico de algoritmo de aprendizado profundo chamado transformadores. Os transformadores são como máquinas de leitura. Eles prestam atenção em diferentes partes da frase ou bloco de texto para entender seu contexto e produzir uma resposta coerente. Também conseguem entender a ordem das palavras em uma frase e seu contexto. (CAELEN, 2024)

Apoiar a extração de entidades e relações: Sua capacidade de compreensão contextual permite identificar entidades específicas e suas conexões em textos extensos e não estruturados.

Fornecer respostas contextualizadas: O modelo consegue responder a consultas específicas, adaptando as respostas ao contexto fornecido pelo usuário, tornando-o uma ferramenta versátil na análise de dados textuais.

O processo de utilização do modelo LLM é apresentado pela Figura 1.

Figura 1 - Diagrama de processo.



Fonte: O autor (2025).

De acordo com Caelen e Blete (2024), os modelos da família GPT recebem um *prompt* como entrada e, como resposta, geram um texto, em um processo conhecido como conclusão de texto.

Dessa forma, GPT-4 apresenta-se como uma poderosa ferramenta de instrumentalização na mineração de texto, alinhando-se às práticas de PLN para otimizar a análise e interpretação de informações em grande escala.

2.2.2 Direito à Moradia e o Processo de Usucapião

O direito à moradia digna ganhou relevância mundial após a Segunda Guerra Mundial, sendo enfatizado na Declaração Universal dos Direitos Humanos, no artigo 25, parágrafo 1º, que reconhece o direito de todos a um padrão de vida adequado. No Brasil, essa garantia foi incorporada à Constituição Federal de 1988, como um direito fundamental no artigo 6º e reforçado nos termos do artigo 5º, que assegura a dignidade humana como um dos pilares constitutivos.

Nesse contexto, a regularização fundiária surge como um importante instrumento para concretizar o direito à moradia, especialmente diante da expressiva informalidade habitacional no Brasil, que compromete a dignidade das pessoas. Como enfatizam Freiria e Dosso (2021), “morar irregularmente é o mesmo que navegar em permanente insegurança”, citando o jurista José Renato Nalini.

Destaca Bazilio (2022):

A usucapião é uma forma de aquisição da propriedade imóvel, sendo que pode ser requerida em caso em que o possuidor seja residente do imóvel, em regra, fazendo uso constante por mais de 10 anos (o prazo pode variar conforme o tipo de usucapião), sem qualquer reclamação do proprietário original. Portanto, o possuidor que requerer os direitos de posse da propriedade não pode tê-los obtido de forma violenta, clandestina ou ilegal. (BAZILIO, 2022)

A usucapião, por sua vez, desempenha um papel central na regularização fundiária. Segundo Costa (1999), tratasse de um modo autônomo de aquisição de propriedade, aplicável tanto a bens móveis quanto imóveis, desde que se cumpram os requisitos legais relacionados à posse qualificada pelo período estipulado em lei. A origem etimológica do termo deriva do latim *usus* (posse) e *capio, capere* (tomar, adquirir), ou seja, adquirir pela posse prolongada.

Conforme Freiria e Dosso (2021), a usucapião apresenta-se como uma ferramenta jurídica para transformar situações de ocupação informal em regularidade, permitindo que indivíduos e famílias conquistem a titularidade do imóvel que ocupam, promovendo, assim a dignidade e a segurança jurídica.

O crescimento desordenado das cidades tem provocado o fenômeno da ocupação de áreas sem a devida formalização legal, criando um cenário em que o poder público não consegue atender às necessidades mínimas exigidas pela sociedade moderna. O acesso ao saneamento básico e à energia elétrica são condições fundamentais para garantir uma vida digna.

A formalização da posse da moradia é o primeiro passo para dar visibilidade ao contingente populacional que vive na obscuridade da informalidade.

Segundo Lobo (2024):

Também há de se destacar as políticas habitacionais inadequadas e a especulação imobiliária. Quando áreas nas quais comunidades se assentam de forma irregular tornam-se mais valorizadas, passam a figurar como objeto de desejo dos especuladores imobiliários, o que faz com que a pressão pela expulsão dos assentados aumente drasticamente e se efetive, na maior parte das vezes. (LOBO, 2024)

2.3 Estudo de Caso: Aplicação da Mineração de Texto em Processos de Usucapião

Segundo Ueno e Soares (2024), na jornada da inteligência artificial (IA), a análise de provas de conceito (*Proofs of Concept*, POCs), os testes de ferramentas e a adoção de projetos piloto podem ser estratégias eficazes antes de escalar um projeto de grande magnitude.

Para demonstrar a aplicação da mineração de texto no contexto de identificação de endereços em processos de usucapião, desenvolvemos uma aplicação que analisou 10 processos de usucapião selecionados de forma aleatória arquivados na base de dados do TJPE.

Com destaque para os desafios da mineração de texto, Alura (2024) enfatiza que:

... um ponto importante é a ambiguidade como característica comum na linguagem natural. Por isso, temos uma etapa de desambiguação, na qual o objetivo é tentar resolver as ambiguidades e determinar o significado mais apropriado com base no contexto. Assim, podemos realizar a extração de informações específicas, como nomes de pessoas, datas, localizações, eventos importantes etc. (ALURA, 2024)

A interação com o modelo foi realizada por meio da linguagem de programação Python, utilizando o ambiente de desenvolvimento web *Google Colab* para execução do código, a biblioteca *OpenAI* e o modelo GPT-4o-mini. O objetivo foi extrair, de maneira estruturada, os dados relacionados aos imóveis, incluindo logradouro, número, cidade, estado e CEP.

Segundo Caelen (2024), a *OpenAI* disponibiliza o GPT-4 e o *ChatGPT* como serviços. Isso implica que os usuários não têm acesso direto ao código-fonte dos modelos, nem podem executá-los em seus próprios servidores. Desta forma é necessário a utilização de (API) para a interação com o modelo GPT, exigindo do usuário uma chave de acesso.

O processo de extração dos dados do endereço do imóvel foi conduzido por meio do seguinte prompt: “Atue como analista judiciário e analise o texto a seguir e identifique a localização do imóvel a ser usucapido de forma estruturada separando os dados em: logradouro ou loteamento, número, bairro, cidade, CEP e estado.”

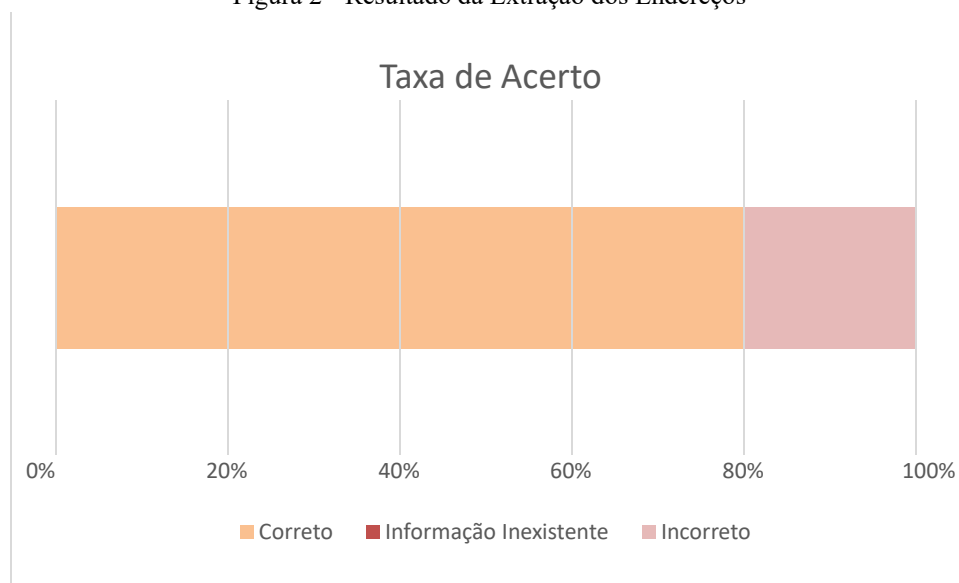
Mediante a definição do prompt foi anexado o texto da petição inicial do processo de usucapião e submetido a análise do modelo GTP.

3 RESULTADOS

Após a execução da mineração de texto sobre as 10 petições iniciais, os resultados foram categorizados em três grupos: (1) Identificação Correta; (2) Informação Inexistente e (3) Dados incorretos. Sendo assim a classificação correta se refere as informações constantes na petição inicial com os dados do imóvel usucapiendo e a mesma selecionada pela IA. Para a classificação de informação inexistente se aplica a definição que a IA não conseguiu definir o endereço do imóvel. Contudo, a classificação de dados incorretos compreende total divergência entre os dados de endereço constantes na petição e os dados definidos pela IA para o imóvel.

Esse detalhamento permitiu avaliar com precisão a eficiência do modelo e identificar limitações específicas, conforme Figura 2.

Figura 2 - Resultado da Extração dos Endereços



Fonte: O autor (2025).

Endereços identificados de forma correta: Em 8 dos 10 processos, os dados de logradouro, número, cidade, estado e CEP foram extraídos de forma como definido na petição inicial.

Endereços incorretos: Em 2 casos, os endereços estavam descritos nas petições, porém o modelo não conseguiu interpretá-los de forma estruturada, identificando outro endereço constante na petição como sendo o do imóvel usucapiendo.

Informação Inexistente: Não houve ocorrência de casos em que a aplicação não encontrou dados referentes a endereço de imóvel.

4 CONCLUSÃO

Os resultados demonstram a eficiência do modelo na extração de informações estruturadas a partir de textos jurídicos não estruturados. Ao analisar os dois casos em que a aplicação não conseguiu definir de forma correta o endereço

do imóvel objeto da ação verificou-se que se tratava de imóveis localizados na zona rural em que a definição da localização do imóvel estava baseada em coordenadas geográficas e pontos de referência.

Este estudo serviu como prova de conceito para a utilização da tecnologia de mineração de texto, comprovando que o emprego da tecnologia impacta significativamente o processo de tomada de decisão auxiliando os operadores do Direito na busca por eficiência.

Apesar do aparente sucesso em 80% dos casos as limitações identificadas em 20% dos exemplos destacam a importância de fornecer textos bem formatados e claros, além de possíveis ajustes no prompt para melhorar o desempenho do modelo.

Esse estudo de caso evidencia o potencial da mineração de texto como ferramenta para agilizar e otimizar a análise documental especialmente em processos de usucapião onde a definição clara do objeto de pesquisa, no caso os dados o endereço do imóvel usucapiendo, facilita a análise computacional, oferecendo soluções práticas para a organização e padronização de dados que servirão de base para iniciativas de ações visando promover justiça social.

As limitações identificadas no estudo abrem caminhos para pesquisas futuras incluindo o desenvolvimento de algoritmos mais robustos e a criação de padrões textuais nos sistemas de tramitação eletrônica. A integração dessas soluções com sistemas o Processo Judicial Eletrônico (PJE) tem o potencial de revolucionar a administração da Justiça no Brasil.

Desta forma, evidenciamos que a adoção de tecnologia de inteligência artificial não deve ser encarada como uma inovação futurística, mas como uma ferramenta real de apoio ao poder judiciário que possibilita transformar a base de conhecimento em um ativo estratégico capaz de apoiar tomadas de decisões e promover um judiciário operacionalmente mais eficiente.

4.1 Proposta de Trabalho Futuro

Este trabalho propõe o desenvolvimento de um algoritmo para a extração automática de endereços contidos em processos de usucapião. Como extensão dessa iniciativa, será criada uma base de dados estruturada que integrará os metadados dos processos, os dados de endereço e as respectivas coordenadas geográficas. O objetivo principal é possibilitar o mapeamento preciso dos imóveis localizados em áreas de interesse social ou em situação de risco, contribuindo para a formulação de políticas públicas voltadas à regularização fundiária e ao planejamento urbano.

Como benefício colateral, este trabalho também colabora para a desjudicialização dos processos de regularização fundiária em imóveis situados em zonas de interesse social (ZEIS), especialmente quando a regularização segue uma tramitação extrajudicial, evitando com isso a judicialização desnecessária de um processo de usucapião. Além disso, contribui para a agilização dos processos de usucapião já judicializados, tornando sua tramitação mais célere e auxiliando de forma significativa no alcance das metas de qualidade propostas pelo Conselho Nacional de Justiça (CNJ).

5 REFERÊNCIAS

ALURA. **O que é PLN?** Disponível em: <https://www.alura.com.br/artigos/o-que-e-pln>. Acesso em: 13 jan. 2025.

BAZILIO, JAQUELINE FERNANDES. **Como funciona o processo de usucapião judicial.** Disponível em: <https://www.jusbrasil.com.br/artigos/como-funciona-o-processo-de-usucapiao-judicial/1369531396>. Acesso em: 13 jan. 2024.

BRASIL. CONSTITUIÇÃO (1988). **Constituição da República Federativa do Brasil.** Brasília, DF: Senado Federal, 1988.

CAELEN, O.; BLETE, M.-A. **Desenvolvendo aplicativos com GPT-4 e ChatGPT.** São Paulo: Novatec, 2024.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). **Programa Justiça 4.0 divulga resultados de pesquisa sobre IA no Judiciário brasileiro.** Disponível em: <https://www.cnj.jus.br/programa-justica-4-0-divulga-resultados-de-pesquisasobre-ia-no-judiciario-brasileiro/>. Acesso em: 13 jan. 2025.

COSTA, DILVANIR JOSÉ DA. **Usucapião: conceitos e fundamentos jurídicos.** Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/524/r143-25.PDF>. Acesso em: 13 jan. 2024.

EXPERT.AI. **Natural Language Processing and Text Mining.** Disponível em: <https://www-expertai.translate.goog/blog/natural-language-processing-and-text-mining/>. Acesso em: 13 jan. 2024.

FREIRIA, R. C.; DOSSO, T. C. **Direito Agrário.** 3. ed. Salvador: Juspodivm, 2021.

FERNANDES, Vitor. **Processamento de Linguagem Natural**. Disponível em: <https://www.each.usp.br/petsi/jornal/?p=2577>. Publicado em: 12 ago. 2019. Acesso em: 13 jan. 2024.

NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. [S.l.]: Nações Unidas, 1948. Disponível em: <https://www.un.org/>. Acesso em: 13 jan. 2024.

STEFANOVSKYI, OLEKSANDR. **Como o ChatGPT da OpenAI funciona**. Intelliarts. Disponível em: <https://intelliarts-com.translate.goog/authors/oleksandr-stefanovskyi/>. Acesso em: 13 jan. 2024.

UENO, GISELE; SOARES, JÚLIO CÉSAR. **A Jornada de IA como Estratégia no Escritório Tributário**. In: COELHO, Alexandre Zavaglia; BARBOSA, Maria Juliana do P. (Org.). *Inteligência Artificial Aplicada aos Serviços Jurídicos*. São Paulo: Thomson Reuters, 2024.