

HÁ DIFERENÇA DE SENTIMENTO NA SENTENÇA DE JUÍZES E JUÍZAS NO BRASIL?

Uma análise empírica de decisões do Tribunal de Justiça do Ceará

Igo Costa

Rafael Carvalho dos Santos

Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa - IDP

Curso de Mestrado Acadêmico em Ciência de Dados e Inteligência Artificial

Disciplina: Deep Learning

Brasília – DF

2025

Sumário

1. Resumo.....	3
2. Introdução	3
3. Definição do Problema	4
4. Coleta e Preparação dos Dados	5
4.1 Fonte de dados e critérios de seleção.....	5
4.2 Recorte institucional e temático	6
4.3 Web scraping e obtenção das sentenças	6
5. Construção e Limpeza do Corpus	7
5.1 Padronização inicial.....	7
5.2 Identificação do magistrado.....	7
6. Inferência do sexo do magistrado	7
7. Metodologia de Análise de Sentimento.....	8
7.1 Unidade textual: o dispositivo da sentença	8
7.2 Modelo de análise de sentimento e estratégia de inferência	8
8. Resultados	8
8.1 Estatísticas descritivas.....	8
8.2 Análise gráfica	9
8.3 Testes estatísticos e regressão.....	10
9. Discussão	10
10. Limitações	11
11. Implicações para políticas públicas	11
12. Considerações finais.....	11

1. Resumo

O presente trabalho acadêmico realiza uma análise empírica com o objetivo de investigar a existência de diferenças no sentimento textual das sentenças judiciais proferidas no Brasil em função do gênero do(a) magistrado(a). Parte-se da pergunta de pesquisa central: **há diferença de sentimento na sentença de juízes e juízas no Brasil?**

A pesquisa baseia-se em um corpus de decisões do Tribunal de Justiça do Estado do Ceará (TJCE), selecionadas a partir de processos relacionados a crimes ambientais. Utilizando técnicas de web scraping, extração automatizada de texto a partir de arquivos PDF, inferência de sexo do magistrado baseada em nomes próprios e análise de sentimento por meio de modelos de linguagem pré-treinados, construiu-se um pipeline computacional reprodutível para coleta, tratamento e análise dos dados. O corpus final é composto por 516 sentenças válidas, extraídas do dispositivo da sentença. Os resultados empíricos, obtidos por meio de análise gráfica, testes estatísticos não paramétricos e regressão linear simples, indicam ausência de diferenças estatisticamente significativas no sentimento das decisões entre magistrados do sexo masculino e feminino. Os achados sugerem elevado grau de padronização linguística institucional na parte decisória das sentenças analisadas.

2. Introdução

A crescente **digitalização do Poder Judiciário brasileiro** tem ampliado significativamente a disponibilidade de dados judiciais em formato eletrônico, abrindo novas possibilidades para a realização de **análises empíricas em larga escala**. A informatização dos processos, aliada à adoção de sistemas eletrônicos de tramitação processual e à disponibilização de bases públicas de dados, tem permitido que pesquisadores explorem decisões judiciais sob perspectivas antes inviáveis, especialmente por meio de métodos computacionais.

Nesse contexto, técnicas de Ciência de Dados, Aprendizado de Máquina e Processamento de Linguagem Natural (PLN) vêm sendo progressivamente incorporadas ao estudo do Direito, dando origem a uma agenda interdisciplinar frequentemente denominada ***Legal Analytics*** ou ***Computational Legal Studies***. Essas abordagens permitem investigar padrões linguísticos, decisórios e institucionais em grandes volumes de textos jurídicos, contribuindo para debates sobre transparência, previsibilidade, coerência decisória e eventuais vieses no sistema de justiça.

Um dos temas recorrentes na literatura internacional refere-se à influência de **características individuais dos julgadores** — como gênero, orientação ideológica, experiência profissional ou formação acadêmica — sobre o conteúdo das decisões judiciais. Estudos empíricos em diferentes países têm investigado se magistrados e magistradas decidem de forma distinta em determinados tipos de casos ou se utilizam estilos redacionais diferentes. No Brasil, entretanto, ainda são relativamente escassas as pesquisas que abordam essa questão de forma sistemática, especialmente a partir da **análise automatizada** do texto das decisões judiciais.

Este trabalho insere-se nesse debate ao investigar se há diferenças no **sentimento expresso na redação das sentenças judiciais** em função do sexo do magistrado. Diferentemente de estudos que analisam apenas o resultado decisório (por exemplo, condenação ou absolvição), o foco aqui recai sobre o **tom linguístico** da decisão, mensurado por meio de técnicas de **análise de sentimento** aplicadas ao texto judicial.

O recorte empírico concentra-se no **Tribunal de Justiça do Estado do Ceará (TJCE)**, com análise de sentenças relacionadas a **crimes contra a fauna**. A escolha desse recorte busca reduzir a heterogeneidade temática do corpus e controlar variações linguísticas associadas a diferentes áreas do direito, favorecendo uma análise mais consistente e comparável.

3. Definição do Problema

Em um primeiro momento, a pesquisa foi estruturada com base em dados provenientes do **Tribunal de Justiça do Estado do Rio de Janeiro (TJRJ)**, obtidos por meio da **API pública DataJud**, mantida pelo Conselho Nacional de Justiça (CNJ). A coleta inicial foi delimitada pelo código de assunto 3619 (“Crimes contra a Fauna”), resultando em um conjunto preliminar de aproximadamente 700 processos. No entanto, durante a fase de implementação do web scraping, foram identificadas **limitações técnicas na extração automatizada e consistente do texto integral das sentenças** a partir do portal do tribunal, o que comprometeria a padronização do corpus e a reprodutibilidade da análise em larga escala.

Diante dessa limitação operacional, optou-se por **redirecionar o escopo empírico do estudo para o Tribunal de Justiça do Estado do Ceará (TJCE)**, cujo portal eletrônico apresentou maior estabilidade e previsibilidade para a automação da coleta das

movimentações processuais e das sentenças. Essa mudança permitiu a construção de um pipeline de extração mais robusto, com controle de erros, registro de logs e validação individual dos processos, assegurando a qualidade e a consistência do corpus textual utilizado nas etapas subsequentes da pesquisa.

O objetivo prático central do projeto consiste em extrair, por meio de métodos de PLN, o nome do(a) juiz(a) responsável pela sentença (utilizando técnicas como reconhecimento de entidades nomeadas — NER — e regras linguísticas híbridas) e, a partir do prenome identificado, inferir o gênero do(a) magistrado(a). Essa inferência será realizada com base em bases públicas de prenomes e heurísticas linguísticas, sendo os casos ambíguos classificados como “indeterminados”, em consonância com boas práticas de transparência metodológica. Essa etapa estabelece a **variável independente fundamental** para a análise de sentimento subsequente, que constitui o núcleo da investigação sobre **diferenças de sentimento nas sentenças judiciais**.

4. Coleta e Preparação dos Dados

A etapa de coleta e preparação dos dados constitui o **alicerce empírico** deste estudo, sendo conduzida de forma sistemática, documentada e reprodutível. O objetivo foi construir um corpus de sentenças judiciais suficientemente amplo e consistente para permitir a aplicação de técnicas de Processamento de Linguagem Natural (PLN) e análise estatística comparativa.

4.1 Fonte de dados e critérios de seleção

Os dados utilizados na pesquisa foram obtidos por meio da **API pública DataJud**, mantida pelo Conselho Nacional de Justiça (CNJ), que disponibiliza informações estruturadas sobre processos judiciais em tramitação e arquivados no Judiciário brasileiro, incluindo metadados como número do processo, tribunal, classe processual, assuntos cadastrados e movimentações processuais. A escolha dessa fonte justifica-se por oferecer vantagens metodológicas importantes, como abrangência nacional, padronização dos dados e aderência a princípios de transparência e acesso a dados públicos. Além disso, a API permite a aplicação de filtros temáticos e temporais de forma reproduzível.

4.2 Recorte institucional e temático

Como recorte institucional, foi selecionado o **Tribunal de Justiça do Estado do Ceará (TJCE)**. A escolha do TJCE também se mostrou adequada do ponto de vista técnico, uma vez que seu portal eletrônico apresentou maior estabilidade para a extração automatizada de informações processuais.

O recorte temático concentrou-se em processos relacionados a **crimes contra a fauna**, com códigos de sentença, conforme classificação disponível nos metadados do DataJud. Esse recorte buscou reduzir a heterogeneidade material do corpus, uma vez que diferentes áreas do direito tendem a apresentar estilos redacionais distintos, bem como encontrar processos que já tivessem decisão.

A consulta à **API pública DataJud**, mantida pelo Conselho Nacional de Justiça (CNJ), resultou em uma amostra inicial de aproximadamente **1.000 processos**, cujos identificadores e metadados básicos (número do processo, órgão julgador, classe processual e assunto) foram armazenados em formato tabular. A partir dessa lista, procedeu-se à extração automatizada das movimentações processuais e das sentenças diretamente no portal do TJCE por meio de técnicas de web scraping. Essa etapa encontra-se documentada no notebook **Deep Learning Final.ipynb**, com os dados intermediários e finais organizados em arquivos CSV, os quais compõem a base empírica do estudo.

4.3 Web scraping e obtenção das sentenças

Após a obtenção dos metadados via DataJud, foi desenvolvido um robô de web scraping para acesso ao portal e-SAJ do TJCE. Esse robô consultou individualmente cada processo a partir do número único (NUP), expandiu todas as movimentações processuais disponíveis e realizou o **download automático dos arquivos PDF das sentenças**.

Foram adotadas boas práticas técnicas, incluindo pausas controladas entre requisições, uso de esperas explícitas, controle de janelas e iframes, e registro estruturado de logs. Essas medidas permitiram mitigar falhas de execução e assegurar a continuidade da coleta em larga escala.

5. Construção e Limpeza do Corpus

Após a etapa de coleta das sentenças judiciais, procedeu-se à **preparação e limpeza do corpus textual**, etapa fundamental para garantir a qualidade analítica das técnicas de Processamento de Linguagem Natural (PLN) aplicadas posteriormente.

5.1 Padronização inicial

A base inicial continha **1.000 processos judiciais** selecionados via API DataJud. Essa base passou por diversas etapas de limpeza, incluindo remoção de duplicatas, padronização do número único de processo (NUP) e exclusão de registros sem documentos disponíveis para análise.

5.2 Identificação do magistrado

A identificação do magistrado responsável pela decisão foi realizada por meio de uma **abordagem híbrida**. Quando disponível, utilizou-se a **extração direta** do nome a partir do HTML do portal e-SAJ. Nos casos em que essa informação não estava estruturada, aplicaram-se expressões regulares ao **texto extraído dos PDFs das sentenças**, especialmente nas páginas finais, onde usualmente consta a assinatura do julgador.

Essa abordagem permitiu ampliar significativamente a cobertura de identificação do magistrado, resultando em **mais de 700 decisões com julgador identificado** em alguma das fontes.

6. Inferência do sexo do magistrado

A inferência do sexo do magistrado foi realizada a partir do nome identificado, utilizando uma abordagem hierárquica e transparente. Inicialmente, aplicaram-se regras jurídicas explícitas baseadas em **marcadores textuais** (juiz, juíza etc.) presentes na sentença. Em seguida, recorreu-se à **base pública de prenomes do IBGE**, utilizando um limiar de confiança para classificar o prenome como masculino ou feminino. Por fim, nos **casos remanescentes**, aplicaram-se heurísticas linguísticas baseadas em **sufixos nominais**.

Foram admitidas três categorias: **masculino, feminino e indeterminado**. Casos ambíguos foram explicitamente classificados como indeterminados, evitando inferências forçadas. Para a análise final, apenas os processos classificados como masculino ou feminino foram mantidos.

7. Metodologia de Análise de Sentimento

A metodologia adotada neste trabalho combina técnicas de **Processamento de Linguagem Natural (PLN)**, **aprendizado de máquina pré-treinado** e **análise estatística não paramétrica**, estruturadas em um pipeline reproduzível, modular e documentado em notebooks versionados.

7.1 Unidade textual: o dispositivo da sentença

A unidade textual adotada para a análise de sentimento foi o **dispositivo da sentença**, entendido como a parte decisória formal do texto judicial. Essa escolha fundamenta-se em três razões principais: sua centralidade normativa, maior padronização linguística e menor variabilidade discursiva em comparação com outras partes da sentença.

A extração do dispositivo foi realizada por meio de expressões regulares aplicadas ao texto dos PDFs, com validação automática e exclusão dos casos em que a extração não foi possível.

7.2 Modelo de análise de sentimento e estratégia de inferência

Para a inferência de sentimento, utilizou-se o modelo **cardiffnlp/twitter-xlm-roberta-base-sentiment**, baseado na arquitetura **XLM-RoBERTa**. O modelo classifica textos em três categorias (negativo, neutro e positivo) e fornece probabilidades associadas a cada classe.

Foi utilizado um **escore contínuo de sentimento**, definido como a diferença entre as probabilidades de polaridade positiva e negativa. Para lidar com textos longos, adotou-se uma estratégia de segmentação em janelas com sobreposição, seguida pela agregação das probabilidades médias por sentença.

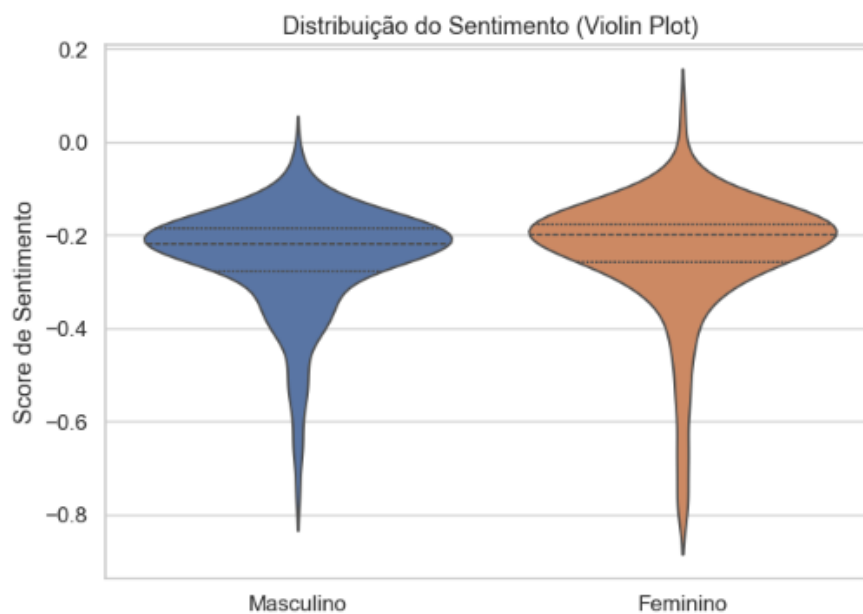
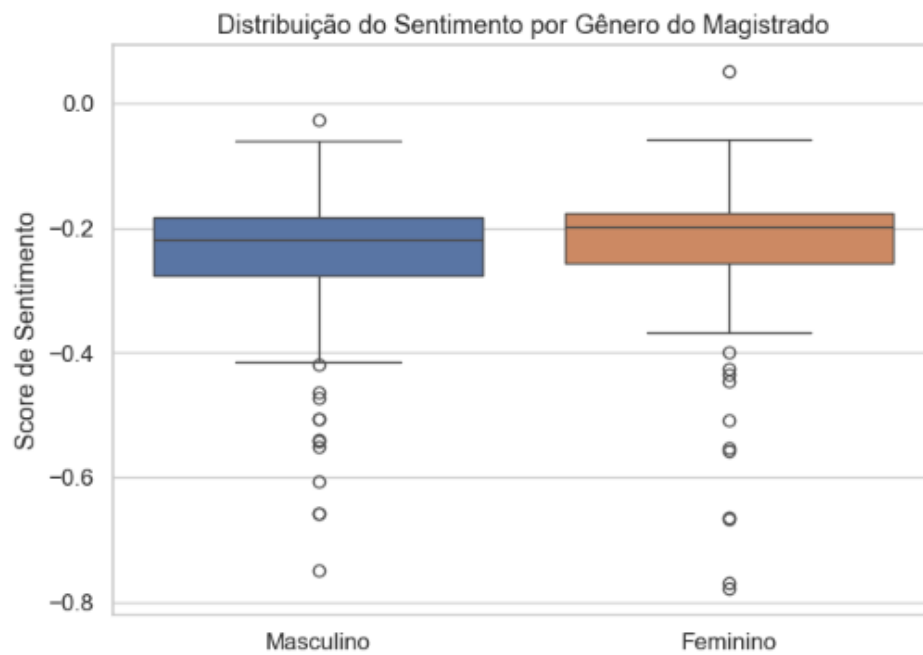
8. Resultados

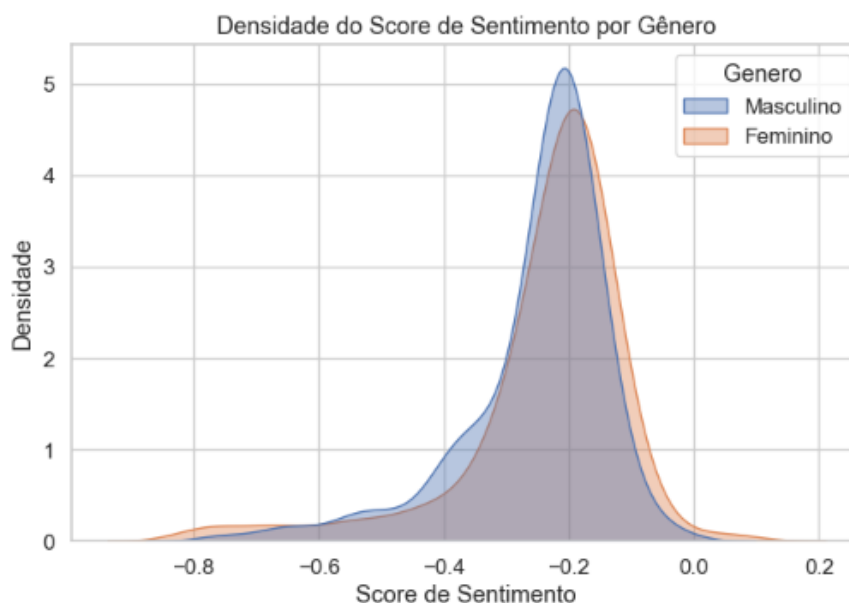
8.1 Estatísticas descritivas

O corpus final de análise é composto por **516 sentenças válidas**, sendo 336 proferidas por magistrados do sexo masculino e 180 por magistradas do sexo feminino. Em ambos os grupos, o sentimento médio apresentou **valores levemente negativos**, compatíveis com a linguagem formal e técnica do contexto penal.

8.2 Análise gráfica

As análises gráficas, incluindo **boxplots**, **violin plots** e **curvas de densidade**, revelaram **forte sobreposição** entre as distribuições de sentimento dos dois grupos. As medianas e os intervalos interquartis são praticamente idênticos, não havendo evidência visual de diferenças sistemáticas.





8.3 Testes estatísticos e regressão

Testes de normalidade indicaram violação da hipótese de normalidade, motivando o uso do **teste não paramétrico de Mann–Whitney U**. O teste não identificou diferença estatisticamente significativa entre os grupos. O tamanho de efeito, medido por Cliff's Delta, foi praticamente nulo. A **regressão linear simples** reforçou esses achados, com coeficiente associado ao sexo feminino próximo de zero e baixo poder explicativo.

9. Discussão

Os resultados indicam **elevado grau de padronização linguística** institucional no dispositivo das sentenças analisadas. A ausência de diferenças substantivas sugere que **normas processuais, modelos redacionais e rotinas organizacionais** exercem papel central na conformação do texto decisório, reduzindo a influência de características individuais do julgador.

Sendo assim, o estudo evidencia que **análises automatizadas de sentimento** em textos jurídicos devem ser **interpretadas com cautela**, especialmente quando baseadas em **modelos genéricos treinados fora do domínio jurídico**.

10. Limitações

Apesar de cumprir os requisitos metodológicos e operacionais propostos no projeto, este estudo apresenta limitações relevantes que devem ser consideradas na interpretação dos resultados.

Entre as limitações destacam-se o **foco em um único tribunal**, o uso de um modelo de sentimento não treinado especificamente em **linguagem jurídica brasileira**, que embora seja amplamente utilizado em contextos acadêmicos, tem a aplicação ao domínio jurídico podendo não capturar nuances semânticas próprias da linguagem judicial, como formalismo, tecnicidade e neutralidade institucional, assim como o processo de **inferência automática de gênero a partir do nome do julgador**, que embora documentado e transparente, está sujeito a ambiguidades, erros de reconhecimento de entidades e limitações inerentes ao uso de heurísticas linguísticas e modelos de PLN, especialmente em casos de nomes abreviados, títulos genéricos ou referências indiretas.

11. Implicações para políticas públicas

O estudo demonstra o potencial de métodos computacionais para o **monitoramento empírico do Judiciário**. Ferramentas desse tipo podem subsidiar políticas de governança judicial, transparência institucional e avaliação de padrões decisórios, permitindo análises contínuas sobre linguagem decisória e padrões institucionais no Judiciário. Mesmo quando o resultado empírico aponta ausência de diferença estatística, essa informação é valiosa para subsidiar decisões baseadas em evidências.

Além disso, o estudo reforça a importância da **transparência e da abertura de dados judiciais em formatos estruturados**, incluindo o acesso mais facilitado ao texto integral das decisões, o que ampliaria significativamente o potencial analítico de pesquisas futuras e permitiria avaliações mais precisas sobre vieses linguísticos, estilos decisórios e impactos institucionais.

12. Considerações finais

Este trabalho investigou empiricamente se há diferenças no sentimento das sentenças judiciais em função do sexo do magistrado, utilizando dados reais do TJCE e técnicas modernas de PLN. Os resultados indicam ausência de diferenças estatisticamente

significativas, contribuindo para o debate sobre **padronização institucional e neutralidade linguística no Judiciário** brasileiro.

Além de seus resultados substantivos, o presente trabalho oferece uma contribuição metodológica relevante ao demonstrar a viabilidade de um **pipeline computacional reproduzível** para a análise empírica de decisões judiciais no contexto brasileiro. O estudo integrou, de forma sistemática, **técnicas de web scraping, extração automatizada de texto a partir de arquivos PDF e métodos de Processamento de Linguagem Natural (PLN)**, permitindo a construção de um corpus textual padronizado e analiticamente consistente a partir de dados judiciais públicos.

Destaca-se, em particular, a **extração automática do dispositivo da sentença** como unidade textual de análise, etapa frequentemente negligenciada em estudos empíricos, mas fundamental para reduzir ruído linguístico e alinhar a análise ao núcleo decisório do documento judicial. Adicionalmente, a aplicação de uma **estratégia explícita de segmentação textual (chunking)**, com posterior agregação das probabilidades de sentimento, permitiu contornar limitações técnicas associadas ao tamanho máximo de entrada dos modelos **Transformer**, preservando a coerência semântica dos textos analisados.

O conjunto dessas escolhas metodológicas — desde a coleta automatizada até a modelagem estatística final — reforça o caráter **replicável, transparente e extensível** do pipeline desenvolvido, o qual pode ser adaptado para outros tribunais, ramos do direito ou tarefas analíticas, contribuindo para o avanço de pesquisas empíricas baseadas em dados no âmbito do Poder Judiciário brasileiro.