

A dualidade da Inteligência Artificial: a cadência do futuro?

Sistema preditivo de criticidade processual com LightGBM: validação temporal e priorização semanal sob capacidade limitada

Juliano Souza de Assis

juliano.souza.assis@gmail.com

Cesar School / TJPE

<https://orcid.org/0009-0002-1154-6503>

Anderson Tenório Sergio

ats@cesar.school

Cesar School

<https://orcid.org/0000-0002-7131-7991>

Resumo – Este artigo apresenta a construção e a avaliação de um artefato de software baseado em inteligência artificial para apoiar a priorização semanal de processos de execução fiscal. A pesquisa aplicada foi orientada por *Design Science Research* (DSR) e utilizou 4.130.466 registros do Tribunal de Justiça de Pernambuco, distribuídos em 12 cargas semanais. A preparação incluiu estatística descritiva, quartis por grupo órgão-tarefa e agrupamento das quatro faixas em três classes ordinais. O LightGBM foi avaliado por validação temporal *walk-forward*, com janela expansiva e embargo de uma semana, produzindo 817.505 previsões fora da amostra. Nas cinco janelas, alcançou F1 ponderado médio de 0,8387 e acurácia média de 0,8392. Na ablação, a retirada da posição percentual intragrupo (PCTL_OT) reduziu o F1 ponderado médio para 0,5688. A comparação das listas TOP-N registrou ganho interno de triagem de até 39,52% em relação à ordenação pelos dias de permanência na tarefa. As saídas apoiam a triagem sem automatizar decisões, com minimização e anonimização dos dados, processamento local, rastreabilidade e registro de versões.

Palavras-chave: inteligência artificial; validação temporal; sistemas de apoio à decisão; priorização processual; LightGBM.

Predictive system for procedural criticality with LightGBM: temporal validation and weekly prioritization under limited capacity

Abstract – This article presents the development and evaluation of an artificial intelligence-based software artifact to support the weekly prioritization of tax enforcement cases. The applied research followed Design Science Research (DSR) and used 4,130,466 records from the Court of Justice of Pernambuco, distributed across 12 weekly snapshots. Data preparation included descriptive statistics, quartiles within judicial unit-task groups, and coarsening of four bands into three ordinal classes. LightGBM was evaluated using walk-forward temporal validation with an expanding window and a one-week embargo, producing 817,505 out-of-sample predictions. Across five windows, it achieved a mean weighted F1 of 0.8387 and mean accuracy of 0.8392. In the ablation analysis, removing the within-group percentile position (PCTL_OT) reduced mean weighted F1 to 0.5688. The TOP-N comparison recorded an internal triage gain of up to 39.52% over ordering by days spent in the current task. The outputs support triage without automating decisions, with data minimization and anonymization, local processing, traceability, and version logging.

Keywords: artificial intelligence; temporal validation; decision support systems; case prioritization; LightGBM.

Agradecimentos

Soli Deo Gloria.

Deo gratias ago, qui per Jesum Christum me sustentavit et illuminavit.

A dualidade da Inteligência Artificial: a cadência do futuro?

Sistema preditivo de criticidade processual com LightGBM: validação temporal e priorização semanal sob capacidade limitada

1 INTRODUÇÃO

Órgãos do Judiciário lidam diariamente com grandes volumes de dados e precisam decidir quais processos receberão atenção primeiro. O problema se torna mais evidente quando a equipe consegue analisar apenas parte do acervo a cada semana. Nessa situação, prever risco não é suficiente, o resultado do modelo precisa chegar ao usuário como uma lista compreensível, passível de conferência e compatível com a capacidade real de trabalho. A decisão final continua sob a supervisão humana.

O Relatório Justiça em Números 2026 atualiza a dimensão desse problema. Ao final de 2025, o Poder Judiciário contabilizava aproximadamente 75,5 milhões de processos pendentes, após uma redução de 3,4 milhões em relação ao ano anterior. No mesmo período, o acervo de execuções fiscais diminuiu em 4,4 milhões de processos, uma queda de 21,3%. Apesar desse avanço, as execuções fiscais ainda representavam cerca de 22% dos casos pendentes e 45% das execuções pendentes do Poder Judiciário, com taxa de congestionamento de 72,4%. O tempo médio de tramitação das execuções fiscais baixadas foi de 8 anos e 2 meses. Esses dados mostram que a redução do estoque não elimina a necessidade de critérios para priorizar o acervo que ainda demanda intervenção (Conselho Nacional de Justiça, 2026).

Na rotina estudada na Diretoria Estadual das Varas de Execução Fiscal, Fazenda Pública e Acidentes de Trabalho (DEFFA), do Tribunal de Justiça de Pernambuco (TJPE), os processos já eram acompanhados pelo tempo de permanência na tarefa atual, registrado como DIAS_ABERTOS. Ordenar a fila por esse valor é uma regra conhecida pelos gestores. Ela, porém, trata do mesmo modo processos de órgãos e tarefas com ritmos diferentes e não informa como o caso tende a se posicionar na semana seguinte. A pesquisa examinou se um sistema preditivo poderia complementar essa referência.

A mineração de processos e o monitoramento preditivo permitem usar o histórico dos registros para acompanhar o estado atual de um caso e estimar sua evolução. Mesmo assim, uma boa métrica geral não resolve sozinha o problema da fila semanal. A avaliação precisa respeitar o tempo, a lista deve caber na capacidade K da equipe, o resultado deve ser comparado com a regra já utilizada e os erros mais importantes precisam ser conhecidos por quem fará a triagem (Van der Aalst; Carmona, 2022; Ceravolo *et al.*, 2024).

A pergunta que orientou o estudo foi a seguinte: como construir e avaliar um sistema que use o estado atual do processo e sua tendência de curto prazo para apoiar a priorização semanal quando a capacidade de análise é limitada? Para responder, foram observados quatro pontos: desempenho ao longo do tempo, contribuição da posição relativa no grupo órgão-tarefa, utilidade das listas TOP-N e condições para um uso rastreável e sujeito à revisão humana.

O sistema usa os dias de permanência na tarefa e a posição relativa do processo entre os casos do mesmo órgão e da mesma tarefa para prever a classe da semana seguinte. Os dois primeiros quartis são reunidos, e as duas faixas superiores permanecem separadas. As probabilidades estimadas formam um escore contínuo de priorização. O desempenho foi medido em cinco janelas de validação temporal, as matrizes por recorte foram utilizadas para diagnosticar os erros.

A contribuição do estudo está no conjunto da solução, e não na criação de um novo algoritmo. O trabalho documenta um fluxo que parte da descrição dos dados, passa pela

A dualidade da Inteligência Artificial: a cadência do futuro?

formação das classes e pelo treinamento temporal e termina em uma lista que pode ser auditada. Também verifica, por ablação, a contribuição da posição percentual intragrupo e mede quanto a nova ordenação pode reduzir o volume de casos a percorrer em comparação com a ordenação pelos dias de permanência na tarefa.

O restante do artigo está organizado da seguinte forma: a Seção 2 apresenta os conceitos, a seleção da literatura e os trabalhos relacionados; a Seção 3 descreve o método de pesquisa; as Seções 4 e 5 detalham o sistema e o protocolo experimental; a Seção 6 reúne a análise e a discussão dos resultados; a Seção 7 apresenta as limitações; e a Seção 8 expõe as recomendações e as conclusões.

2 FUNDAMENTAÇÃO E TRABALHOS RELACIONADOS

A seleção da literatura foi orientada ao problema de priorização processual. As buscas combinaram termos em português e inglês, como “jurimetria”, “mineração de processos judiciário”, “process mining judiciary”, “predictive process monitoring”, “time series forecasting evaluation” e “judicial AI governance”. Foram consultados Google Scholar, portais científicos e bibliotecas digitais, como Springer Link, PLOS ONE, Frontiers, MDPI e ACM Digital Library, além de fontes institucionais do Conselho Nacional de Justiça (CNJ), do TJPE e da Organisation for Economic Co-operation and Development (OECD). Priorizou-se a produção de 2020 a 2025, com referências anteriores de fundamentação e atualização por trabalhos pertinentes de 2026.

Foram incluídos estudos com dados processuais reais, trabalhos relacionados a duração, classificação e priorização, referências metodológicas necessárias às decisões do sistema e documentos sobre gestão e governança de IA. Foram excluídos trabalhos sem relação com esses objetivos, sem detalhamento metodológico suficiente ou redundantes em relação a fontes mais pertinentes. Os estudos comparados na Tabela 1 foram escolhidos pela proximidade com a tarefa preditiva, o contexto judicial ou a formação de listas sob capacidade limitada.

2.1 Sistemas de apoio à decisão e monitoramento preditivo de processos

A mineração de processos parte dos registros de eventos deixados pelos sistemas para reconstruir fluxos, verificar conformidade e localizar tempos de espera e gargalos. No Judiciário, esses registros ajudam a entender a tramitação e a apoiar a gestão do acervo. A análise, contudo, depende do significado correto de tarefas e movimentações e precisa considerar que o fluxo pode mudar ao longo do tempo (Lima *et al.*, 2023; Cruz *et al.*, 2023; Van der Aalst; Carmona, 2022).

No monitoramento preditivo, o estado atual de um processo é usado para estimar duração, resultado, desvio ou algum risco futuro. A previsão só tem utilidade quando seu horizonte coincide com a decisão que será tomada. Por isso, modelos destinados a uma rotina semanal devem ser avaliados na mesma ordem em que os dados surgem e revistos quando o comportamento do fluxo muda (Ceravolo *et al.*, 2024; Bousdekis *et al.*, 2023).

Um sistema de apoio à decisão reúne dados, modelos e formas de apresentar o resultado para ampliar a análise do usuário, sem assumir a responsabilidade pelo ato final (Sharda; Delen; Turban, 2019). Essa distinção é central neste estudo. A saída do sistema é uma lista ordenada. Classe prevista, probabilidades, score e DIAS_ABERTOS continuam visíveis, para que o gestor possa comparar, revisar ou desconsiderar a sugestão.

A dualidade da Inteligência Artificial: a cadência do futuro?

2.2 Estatística descritiva, discretização e *coarsening*

Quartis e percentis descrevem a posição de um valor em uma distribuição. Os quartis dividem os dados nos pontos de 25%, 50% e 75%, o percentil indica a posição relativa de cada observação. Essas medidas são úteis para tempos processuais porque tais distribuições costumam ser assimétricas e conter valores extremos, situação em que a média, sozinha, oferece uma leitura limitada (Wheelan, 2016).

A escolha dos pontos de corte define as categorias produzidas pela discretização. Rim e Liu (2020) utilizam cortes baseados na média e no desvio padrão para discretizar atributos de entrada do algoritmo C4.5. Neste trabalho, os dias de permanência na tarefa (DIAS_ABERTOS) são divididos em quartis por órgão julgador, tarefa e carga semanal, formando o QOT (quartil órgão–tarefa). O campo QOT_FIXO registra o quartil atual, de QOT1 a QOT4. Quando o mesmo processo é localizado na carga seguinte, seu quartil observado recebe o nome QOT_FUTURO. O alvo de três classes é formado pelo agrupamento dos dois primeiros quartis futuros. A posição percentual órgão tarefa intragrupo (PCTL_OT), por sua vez, conserva uma medida contínua da posição do processo no grupo corrente.

Após a divisão intragrupo em quartis, QOT1 e QOT2 são reunidos, enquanto QOT3 e QOT4 permanecem separados. Esse agrupamento, denominado *coarsening*, foi definido como uma decisão de projeto, que reúne a metade inferior da distribuição e preserva a distinção entre as duas faixas superiores, de maior interesse para a priorização relativa. O resultado são três classes ordinais: QOT1+QOT2, QOT3 e QOT4.

Na DEFFA, a classificação operacional adotada no período do estudo considera saneados os processos com até 49 dias de permanência na tarefa, pré-críticos aqueles entre 50 e 99 dias e críticos aqueles com 100 dias ou mais. Essa é a referência utilizada no local de trabalho.

Os quartis adicionam contexto com outra função, pois situam o processo entre os casos do mesmo órgão julgador e da mesma tarefa em uma carga semanal. Seus limites variam entre grupos e semanas. Assim, QOT1+QOT2 não equivale necessariamente a saneados, QOT3 a pré-críticos ou QOT4 a críticos. A proporção de observações em cada classe QOT não informa a proporção de processos nas faixas institucionais de dias.

2.3 Validação temporal, governança e reprodutibilidade

Quando os dados têm ordem cronológica, exemplos futuros não podem participar do treinamento de um modelo que será avaliado no passado. Uma divisão aleatória pode misturar esses momentos e produzir um resultado mais otimista do que o uso real permitiria. A janela expansiva evita esse problema: o modelo aprende com as cargas anteriores e é testado em uma carga posterior. A repetição do procedimento também mostra se o desempenho se mantém ao longo das semanas (Cerqueira; Torgo; Mozetič, 2020; Hyndman; Athanasopoulos, 2021; Hewamalage; Ackermann; Bergmeir, 2023).

A separação temporal é apenas uma parte da avaliação. Estudos com aprendizado de máquina precisam informar quais dados foram usados, como foram transformados, quais configurações foram adotadas e onde estão as limitações. TRIPOD+AI e REFORMS reúnem recomendações nessa direção, e trabalhos sobre reprodutibilidade mostram os problemas causados por descrições incompletas do fluxo computacional (Collins *et al.*, 2024; Kapoor *et al.*, 2024; Semmelrock *et al.*, 2025). Neste estudo, F1 ponderado e acurácia são apresentados em conjunto porque oferecem leituras diferentes do problema multiclasse (Rainio; Teuho; Klén, 2024).

A dualidade da Inteligência Artificial: a cadência do futuro?

No serviço público, desempenho técnico não autoriza o uso automático do resultado. É necessário saber quem responde pela decisão, manter registros do ciclo de vida do sistema e permitir revisão. A Resolução CNJ nº 615/2025 estabelece diretrizes para o desenvolvimento e o uso de inteligência artificial no Poder Judiciário. Por isso, o sistema foi tratado como apoio à triagem: versões, parâmetros e execuções são registrados, e nenhuma previsão produz ato ou movimentação processual por conta própria (Conselho Nacional de Justiça, 2025; Cheong, 2024; OECD, 2025; Zuiderwijk; Chen; Salem, 2021).

2.4 Lacuna identificada na literatura

A Tabela 1 compara tarefa, método, validação e tratamento da capacidade nos estudos relacionados. Como alvos, horizontes e métricas diferem, a comparação situa as escolhas metodológicas, sem ordenar os trabalhos por desempenho.

Tabela 1 – Trabalhos relacionados: contexto, método e distinção em relação ao estudo.

Estudo	Contexto e tarefa	Método e validação	Priorização
Cruz <i>et al.</i> (2023)	Justiça eleitoral: duração processual.	LightGBM e mineração de processos; divisão aleatória.	Sem lista sob capacidade.
Lage-Freitas <i>et al.</i> (2022)	Tribunais brasileiros: decisões.	Classificação binária; modelos clássicos e aprendizado profundo; divisão aleatória.	Sem ordenação sob capacidade.
Oliveira <i>et al.</i> (2022)	Justiça trabalhista: duração processual.	Regressão e <i>ensembles</i> ; divisão aleatória.	Sem lista semanal.
Resck <i>et al.</i> (2025)	Supremo Tribunal Federal: temas.	Regressão logística e outros modelos; divisão estratificada e validação cruzada.	Triagem temática, sem horizonte semanal.
Sobrino <i>et al.</i> (2026)	Promotoria de Zacatecas: finalização em seis meses.	<i>Random Forest</i> e outros modelos; validação temporal.	Lista semanal de 300 casos.
Este estudo	Execuções fiscais: classe relativa na semana seguinte.	LightGBM; cinco janelas temporais com embargo.	Escore e capacidade K, comparados aos dias de permanência.

Fonte: elaboração própria a partir dos trabalhos citados.

Cruz *et al.* (2023) e Oliveira *et al.* (2022) abordam a duração processual e sustentam o uso de atributos de movimentações e de bases judiciais de grande volume. Lage-Freitas *et al.* (2022) e Resck *et al.* (2025) tratam da classificação de decisões e temas, com alvos distintos da posição temporal futura. Sobrino *et al.* (2026) aproxima-se da rotina operacional ao produzir listas semanais sob capacidade fixa, embora utilize outro alvo e horizonte. Neste estudo, o sistema integra contexto órgão-tarefa, previsão semanal, validação temporal com embargo e ordenação sob capacidade K. A comparação situa essas escolhas, não permite ordenar os estudos por desempenho, pois tarefas, horizontes e métricas diferem.

3 MÉTODO DE PESQUISA

3.1 Tipo de pesquisa

Trata-se de pesquisa aplicada, quantitativa e observacional, orientada por *Design Science Research* (DSR, ou pesquisa em ciência do design), com avaliação empírica em um contexto institucional único, a DEFFA/TJPE. Foram utilizados dados processuais secundários, estruturados e previamente existentes. O objeto empírico compreende processos de execução fiscal (EF) organizados em cargas semanais, o objeto técnico é o sistema construído para apoiar sua priorização.

A dualidade da Inteligência Artificial: a cadência do futuro?

O desenvolvimento partiu do problema de ordenação semanal sob capacidade limitada e da definição de requisitos de uso, rastreabilidade e supervisão humana. Em seguida, foram implementados a preparação das cargas, os atributos, o alvo futuro e a política de ordenação. A demonstração utilizou cargas reais em análise retrospectiva. A avaliação seguiu os quatro critérios apresentados na Seção 3.2. Decisões, versões e resultados foram documentados no artigo e nos registros de execução (Hevner *et al.*, 2004; Vom Brocke; Gau; Mädche, 2021).

As cargas foram extraídas do sistema corporativo TJPE Reports, disponibilizado pela SEPLAN (Secretaria de Planejamento e Gestão Estratégica) do Tribunal de Justiça de Pernambuco, com autorização institucional formalizada em processo do Sistema Eletrônico de Informações (SEI). Foram utilizados registros estruturados previamente existentes, sem intervenção nos fluxos processuais. Identificadores pessoais e conteúdo de peças processuais ficaram fora da base analítica. O processamento ocorreu localmente, e nenhuma classificação gerou ato, movimentação ou consequência processual automática.

3.2 Procedimentos e critérios de avaliação

Os símbolos t e $t+1$ representam, respectivamente, a semana corrente e a seguinte. As 12 cargas históricas são identificadas por s_0 a s_{11} em ordem cronológica, o número após s corresponde ao índice da carga no campo `SNAPSHOT_IDX`. O mesmo processo é ligado entre duas cargas consecutivas pelo Número Processual Único (NPU).

A avaliação foi organizada em quatro critérios. C1 verifica se F1 ponderado e acurácia permanecem estáveis nas cinco janelas temporais. C2 mede a contribuição de `PCTL_OT` por importância interna e por ablação. C3 compara o esforço de triagem da ordenação pelo escore contínuo de priorização (`CRIT_SCORE_%`) com o da ordenação por dias em aberto (`DIAS_ABERTOS`), em listas TOP-N limitadas pela capacidade semanal de análise, representada por K . C4 examina os erros por recorte e as condições de rastreabilidade, revisão e responsabilidade humana.

Duas perguntas foram mantidas separadas durante a análise. A primeira é preditiva: a classe relativa prevista pelo modelo (`QOT_PRED`), calculada em t , corresponde à classe relativa observada na semana seguinte (`QOT_FUTURO`), em $t+1$? A segunda é operacional: para uma capacidade K , a lista ordenada por `CRIT_SCORE_%` encontra as classes previstas de interesse antes da lista ordenada por `DIAS_ABERTOS`? Essa distinção impede que uma medida interna da lista seja apresentada como impacto real sobre a organização.

4 SISTEMA PREDITIVO DE CRITICIDADE PROCESSUAL

4.1 Descrição e componentes técnicos

O sistema recebe cargas semanais e produz uma lista ordenada para triagem. `DIAS_ABERTOS` continua presente, mas passa a ser acompanhado pela classe prevista, pelas probabilidades e pelo `CRIT_SCORE_%`. Entre a entrada e a lista final, o programa verifica os dados, calcula quartis e percentis por grupo, forma as três classes relativas, executa o LightGBM e guarda as informações necessárias para conferir o resultado.

A escolha do LightGBM apoiou-se em quatro dimensões: desempenho empírico no Judiciário brasileiro, viabilidade computacional, aderência operacional e conformidade normativa. Cruz *et al.* (2023) aplicaram o algoritmo a dados processuais dos Tribunais Regionais Eleitorais do Nordeste, enquanto Resck *et al.* (2025) o avaliaram, junto a outros modelos, na classificação de recursos do Supremo Tribunal Federal. Neste estudo, a eficiência de memória e processamento em bases tabulares favoreceu a execução local sobre 4.130.466

A dualidade da Inteligência Artificial: a cadência do futuro?

registros, enquanto as probabilidades por classe permitiram compor o CRIT_SCORE_% e produzir saídas passíveis de conferência (Ke *et al.*, 2017). A escolha não pressupõe superioridade universal, mas adequação ao problema, ao volume de dados e à forma de uso avaliada.

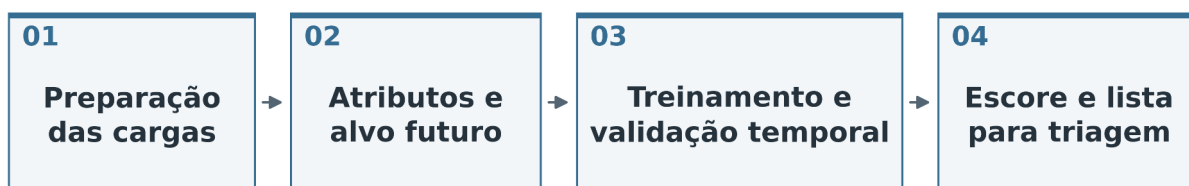
A implementação em Python utiliza pandas e NumPy no tratamento dos dados, *scikit-learn* no pré-processamento e nas métricas e LightGBM na classificação multiclasse (Ke *et al.*, 2017). O pré-processamento utiliza *OneHotEncoder* para órgão e tarefa e *StandardScaler* para os atributos numéricos.

O experimento foi executado localmente em notebook com processador AMD Ryzen 3 5300U, frequência de 2,60 GHz, 8 GB de memória RAM e aproximadamente 238 GB de armazenamento. O processamento ocorreu localmente, sem compartilhamento da base em rede. Os tempos reportados correspondem a esse equipamento e às condições da execução documentada.

4.2 Como o sistema foi organizado

A Figura 1 resume o sistema em quatro blocos, da preparação das cargas à geração da lista semanal. O fluxo distingue as informações disponíveis na semana t da classe observada em $t+1$, utilizada na construção do alvo futuro e na avaliação.

Figura 1 – Arquitetura metodológica com validação temporal.



As previsões utilizam informações disponíveis em t ;
o alvo observado é obtido em $t + 1$.

Fonte: elaboração própria.

No primeiro bloco, as cargas semanais são lidas, conferidas e padronizadas. No segundo, calculam-se os quartis e a posição percentual dos dias de permanência na tarefa em cada grupo definido por órgão julgador, tarefa e carga semanal, antes dos filtros de elegibilidade para o treino. Em seguida, realiza-se o pareamento do mesmo NPU entre t e $t+1$. Esse pareamento permite formar o alvo futuro, com QOT1 e QOT2 agrupados em uma única classe. O terceiro bloco reúne o LightGBM e a validação em cinco janelas temporais, com treino expansivo e embargo de uma carga. No quarto, as probabilidades estimadas são convertidas no escore de priorização, os K primeiros processos compõem a lista TOP- N e os resultados são exportados para conferência. Na aplicação semanal, a carga em t gera a lista, a carga em $t+1$ fornece a classe observada para a análise diagnóstica, sem retornar ao treinamento daquela rodada.

4.3 Preparação, discretização do QOT e representação contextual

Cada carga é uma fotografia semanal do acervo. Antes do uso, o programa confere colunas e tipos, padroniza nomes e trata valores ausentes. SNAPSHOT_IDX preserva a ordem das semanas, e o NPU liga o mesmo processo em cargas consecutivas. Processos suspensos, arquivados ou sobrestados permanecem registrados para fins descritivos e de auditoria, mas não entram no treino enquanto estão nessas situações.

A dualidade da Inteligência Artificial: a cadência do futuro?

DIAS_ABERTOS mede o número de dias que o processo permanece na tarefa atual. Em cada semana, os registros são agrupados por órgão julgador e tarefa. Os quartis são calculados sobre postos ordenados pelos dias de permanência, com desempate pela ordem de ocorrência dos registros. Assim, valores iguais de dias podem pertencer a quartis distintos. Esse cálculo é realizado antes dos filtros de elegibilidade para o treino. QOT_FIXO guarda o quartil atual, de 1 a 4. Quando o processo é encontrado na carga seguinte, o quartil de $t+1$ recebe o nome QOT_FUTURO e passa a ser o alvo do modelo. Essa variável não faz parte das entradas usadas para prever.

Neste estudo, a previsão refere-se à posição relativa do processo na semana seguinte, entre os casos do mesmo órgão e da mesma tarefa. As classes QOT não determinam se o processo atingiu o limite institucional de 100 dias, que continua sendo verificado por DIAS_ABERTOS. A previsão não expressa urgência jurídica, mérito, probabilidade de êxito nem autorização para qualquer ato automático.

PCTL_OT complementa DIAS_ABERTOS ao informar a posição percentual do processo entre os demais casos do mesmo órgão e da mesma tarefa na carga corrente:

$$PCTL_OT(p,s) = rank_pct(DIAS_ABERTOS(p,s) \mid \text{órgão}(p,s), \text{tarefa}(p,s), \text{carga } s)$$

Nessa expressão, p identifica o registro processual e s , a carga semanal. A função $rank_pct$ calcula $(r - 1)/(n - 1)$, em que r é o posto do registro, com média dos postos em caso de empate, e n é o número de observações do grupo órgão–tarefa–carga. Para grupos com uma observação, a posição percentual é definida como zero.

QOT e PCTL_OT representam aspectos diferentes do tempo. QOT divide a distribuição em faixas ordinais, PCTL_OT mantém a posição contínua dentro do grupo. Um processo pode, portanto, compartilhar o mesmo QOT com vários casos e ainda ter uma posição percentual diferente. Os dois valores são calculados somente com a carga corrente, sem consultar QOT_FUTURO.

4.4 Classificação, escore e lista semanal

Para cada processo, o LightGBM estima QOT_PRED e as probabilidades P_QOT12 , P_QOT3 e P_QOT4 . QOT_FUTURO é o alvo preditivo observado em $t+1$, CRIT_SCORE_% é o critério contínuo que transforma as probabilidades estimadas em uma ordem de triagem. O escore, portanto, não cria uma quarta classe nem deve ser interpretado como probabilidade calibrada de um evento. Ele é calculado por:

$$CRIT_SCORE_ \% = (P_QOT4 + 0,6 \times P_QOT3) \times 100$$

No escore, P_QOT4 recebe peso integral por representar a classe relativa superior prevista, enquanto P_QOT3 recebe peso 0,6 como classe intermediária de prioridade. P_QOT12 não recebe peso. Essa escolha privilegia as duas faixas superiores sem convertê-las nas faixas institucionais fixas de dias. O valor 0,6 é uma decisão de projeto, não um parâmetro aprendido pelo LightGBM. O CRIT_SCORE_% não representa a probabilidade de atingir 100 dias e pode ter seus pesos revistos em outra aplicação mediante análise de sensibilidade e validação temporal.

A posição percentual, calculada por órgão, tarefa e carga semanal, integra o modelo e contextualiza os dias de permanência. As probabilidades previstas são convertidas pela mesma regra no CRIT_SCORE_%, que ordena conjuntamente os processos de diferentes órgãos e tarefas, do maior para o menor escore. Os K primeiros formam a lista semanal, com DIAS_ABERTOS disponível para comparação. A lista indica uma ordem inicial de leitura, sujeita à revisão do responsável pela triagem, sem definir as providências a adotar.

A dualidade da Inteligência Artificial: a cadência do futuro?

5 PROTOCOLO EXPERIMENTAL

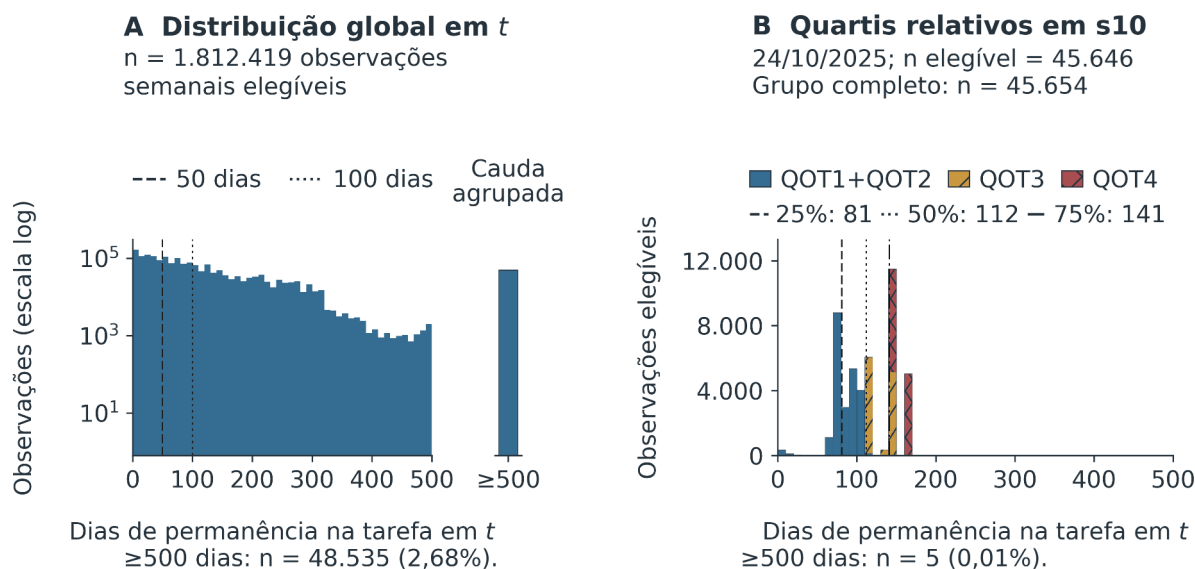
5.1 Cargas e estatística descritiva

Foram usadas 12 cargas históricas semanais, de 15 de agosto a 31 de outubro de 2025. Juntas, elas contêm 4.130.466 registros. Uma carga de 7 de novembro foi usada na aplicação do modelo, e a carga de 14 de novembro forneceu as classes observadas para os recortes das matrizes. Após os filtros de escopo e elegibilidade, 1.812.419 observações permaneceram na base efetiva de treino e validação. Os arquivos preservam a data e o horário da extração no prefixo AAAAMMDDhhmmss. Por exemplo, 20250815080214 identifica a carga extraída em 15/08/2025 às 08h02min14s.

A seleção para o treino considerou a classe Execução Fiscal, restrita a quatro unidades especializadas, sem limitar a base ao local de trabalho Diretoria. Após o pareamento semanal, o filtro por classe manteve 2.739.546 observações, e o recorte pelas quatro unidades manteve 2.084.293. Dessas, 271.874 estavam em tarefas de suspensão, arquivamento ou sobrestamento e foram segregadas, restando 1.812.419 observações elegíveis.

O pareamento entre semanas foi feito pelo NPU. A base efetiva contém observações semanais e um processo pode aparecer em mais de uma carga. No alvo futuro de três classes, QOT1+QOT2 reuniu 906.818 observações (50,03%), QOT3 reuniu 454.368 (25,07%) e QOT4, 451.233 (24,90%). Esses percentuais refletem o agrupamento de dois quartis na primeira classe e não a proporção de saneados, pré-críticos e críticos na classificação institucional. As cinco janelas de teste produziram 817.505 predições fora da amostra. A Tabela 2 apresenta os volumes de cada etapa e a Figura 2 apresenta a distribuição dos dias de permanência na tarefa e o agrupamento dos quartis em três classes.

Figura 2 – Distribuição dos dias de permanência na tarefa e agrupamento dos quartis.



Intervalos de 10 dias; valores ≥500 em categoria aberta.

Cortes projetados do ranking; empates podem atravessar quartis.

Faixas institucionais e classes relativas distintas; órgão e tarefa omitidos.

Fonte: elaboração própria a partir da base analítica do experimento.

A dualidade da Inteligência Artificial: a cadência do futuro?

O painel A reúne as 1.812.419 observações semanais elegíveis e destaca os limites institucionais de 50 e 100 dias. O painel B apresenta um grupo órgão–tarefa da carga s10, com os dois primeiros quartis reunidos em uma classe e o terceiro e o quarto mantidos separadamente. Essa classificação relativa expressa a posição das observações no grupo e não equivale às faixas institucionais. Nos dois painéis, as observações com 500 dias ou mais foram reunidas em categoria aberta, sem exclusão de registros. A identificação nominal do órgão e da tarefa foi omitida para preservar informações institucionais.

Tabela 2 – Volumes utilizados na construção e avaliação do sistema.

Conjunto ou etapa	Volume	Função
12 cargas históricas	4.130.466	Ingestão bruta e preparação temporal.
Registros preparados para treino	3.768.574	Base anterior aos filtros do domínio.
Observações semanais elegíveis	1.812.419	Base elegível para treino e validação.
Estados segregados	271.874	Suspensos, arquivados ou sobrestados preservados para auditoria.
Classes futuras QOT1+QOT2 / QOT3 / QOT4	906.818 / 454.368 / 451.233	Distribuição das classes na base elegível.
Predições fora da amostra	817.505	Agregação das cinco janelas temporais.
Recorte EF Geral	164.165	Conjunto geral representado nas matrizes de confusão por recorte.

Fonte: elaboração própria a partir dos registros do experimento.

Os Recortes A e B correspondem, respectivamente, às perspectivas de execução fiscal estadual (VEFE) e municipal (VEFM). Nas matrizes, são apresentados recortes gerais e por local de trabalho. Na comparação TOP-N, A corresponde a VEFE na Diretoria e B a VEFM no Gabinete. Todos são subconjuntos examinados pelo mesmo sistema, não classificadores treinados separadamente.

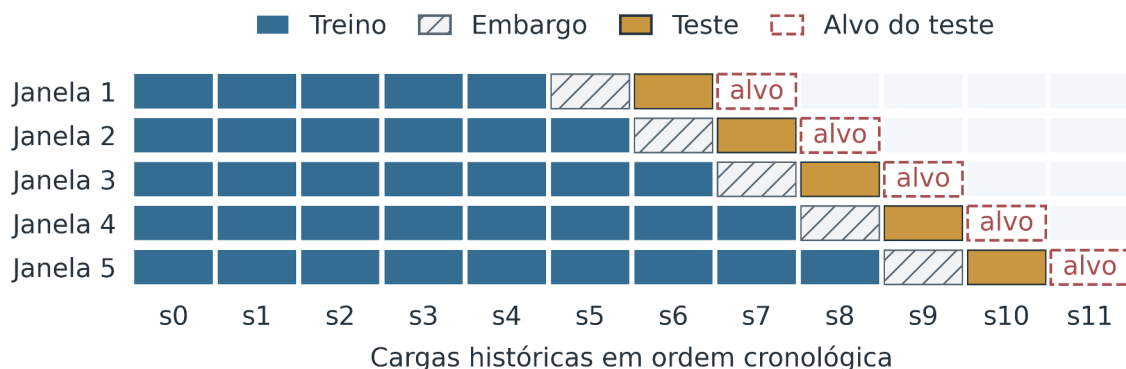
5.2 Validação temporal com embargo

A Figura 3 representa as cinco janelas expansivas, e a Tabela 3 detalha as cargas e os volumes de cada partição. A primeira janela usa s0 a s4 para treino, s5 como embargo e s6 para teste. Na quinta, s0 a s8 compõem o treino, s9 é o embargo e s10 fornece as entradas do teste. A carga s11 fornece os rótulos observados na semana seguinte para os processos de s10, por isso, não constitui uma sexta janela de teste. O pareamento t com t+1 também é realizado para as demais cargas. Assim, as entradas supervisionadas estão em s0 a s10, embora tenham sido necessárias 12 cargas históricas.

Após a validação temporal, o modelo foi aplicado à carga de 07/11/2025, identificada nos arquivos como CARGA_0 e adotada como referência temporal t dessa aplicação. A carga de 14/11/2025, denominada CARGA_t+1, forneceu os rótulos observados uma semana depois, permitindo confrontá-los com as predições produzidas em t. Essas duas cargas foram utilizadas na etapa posterior de aplicação e análise diagnóstica, sem integrar as cinco janelas da validação temporal.

A dualidade da Inteligência Artificial: a cadência do futuro?

Figura 3 – Validação temporal com janela expansiva e embargo de uma carga.



Fonte: elaboração própria.

Nota (figura 3): O contorno tracejado identifica a carga que fornece as classes observadas na semana seguinte ao teste. A carga de embargo é excluída das entradas de treino e teste, mas fornece os rótulos futuros da última carga de treino. Na primeira janela, s5 fornece os rótulos de s4, enquanto s7 fornece os de s6.

Tabela 3 – Estrutura das cinco janelas temporais com embargo.

Janela	Treino	Embargo	Teste	n treino	n teste
1	s0–s4	s5	s6	830.754	163.861
2	s0–s5	s6	s7	994.914	163.705
3	s0–s6	s7	s8	1.158.775	163.424
4	s0–s7	s8	s9	1.322.480	163.077
5	s0–s8	s9	s10	1.485.904	163.438

Fonte: elaboração própria a partir dos registros do experimento.

O intervalo de uma carga entre treino e teste reduz a influência imediata do mesmo ciclo semanal sobre a avaliação. Ele não torna as observações independentes, porque um processo pode continuar no acervo por várias semanas. Sua função é mais específica: impedir que informações posteriores ao ponto de corte sejam usadas para ajustar o modelo que será testado.

5.3 Configurações testadas e métricas

A configuração adotada no modelo final foi `class_weight=None`, que mantém a distribuição observada das classes. Como teste de sensibilidade, o treinamento também foi executado com `class_weight="balanced"`. Na ablação, `PCTL_OT` foi retirada e as demais entradas: `ORGAO_JULGADOR`, `TAREFA` e `DIAS_ABERTOS` foram mantidas. Esses testes avaliam escolhas do sistema, não pretendem comparar todas as famílias de modelos possíveis.

O experimento foi executado em ambiente Windows, com Python 3.13.2 e as bibliotecas `pandas 2.2.3`, `NumPy 2.2.6`, `scikit-learn 1.6.1` e `LightGBM 4.6.0`. O `LightGBM` foi configurado para classificação multiclasse, com três classes, 500 iterações de treinamento, taxa de aprendizagem de 0,1, limite de 64 folhas por árvore, profundidade máxima de 9 e mínimo de 100 observações por folha.

A fração de atributos foi fixada em 0,85, e a amostragem de observações, em 0,9, realizada a cada duas iterações. Os coeficientes de regularização L1 e L2 foram definidos como 0,2 e 0,4, respectivamente, com ganho mínimo para divisão igual a zero e semente pseudoaleatória igual a 42. Essa configuração foi utilizada nas cinco janelas de validação

A dualidade da Inteligência Artificial: a cadência do futuro?

temporal e no ajuste final, sem ponderação de classes e com embargo de uma carga entre treino e teste.

Os volumes e as proporções das classes foram descritos por carga. Para o classificador, a métrica principal é o F1 ponderado, que combina precisão e revocação de cada classe e considera seu suporte. A acurácia foi mantida como leitura complementar. Média e desvio-padrão mostram a variação entre as cinco semanas de teste, e as matrizes permitem identificar em que direção os erros ocorreram (Rainio; Teuho; Klén, 2024).

A análise TOP-N compara duas listas formadas com os mesmos processos: uma ordenada por CRIT_SCORE_% e outra por DIAS_ABERTOS. Para cada capacidade K, conta-se quantas posições a segunda lista precisa percorrer para reunir a mesma quantidade de casos que a primeira colocou na faixa prevista de interesse. O ganho é calculado por:

$$Ganho_TOP-N = 1 - K / N$$

N é o número de posições percorridas na lista ordenada por dias de permanência na tarefa para alcançar a mesma quantidade de casos da classe prevista de interesse obtida no lote de tamanho K. A expressão retorna uma proporção, multiplicada por 100 para apresentação em porcentagem.

A contagem usa QOT_PRED_LABEL, isto é, a classe prevista pelo próprio sistema. Portanto, TOP-N mede a eficiência interna da ordenação e não o acerto futuro, a redução do acervo ou outro efeito institucional. O desempenho preditivo é medido separadamente nas cinco janelas com QOT_FUTURO. As matrizes de confusão por recorte servem apenas para detalhar os erros observados nas perspectivas filtradas.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

6.1 Desempenho fora da amostra e execução computacional

A Tabela 4 apresenta o desempenho por janela temporal, e sua variação ao longo dos cinco testes temporais. Nas cinco janelas, o F1 ponderado ficou entre 0,8233 e 0,8488. A média foi 0,8387, com desvio-padrão amostral de 0,0094. A acurácia variou de 0,8216 a 0,8490, com média de 0,8392 e desvio-padrão de 0,0103. Ao todo, foram avaliadas 817.505 predições fora da amostra. A proximidade dos resultados entre as semanas indica que o desempenho não dependeu de uma única janela e atende ao critério C1.

Tabela 4 – Desempenho fora da amostra por janela temporal.

Janela	F1 ponderado	Acurácia	n teste
1	0,8378	0,8411	163.861
2	0,8233	0,8216	163.705
3	0,8488	0,8490	163.424
4	0,8417	0,8413	163.077
5	0,8418	0,8428	163.438
Média	0,8387	0,8392	817.505 (total)

Fonte: elaboração própria a partir dos registros do experimento.

A leitura sequencial das 12 cargas levou cerca de 47 minutos. O treinamento e o teste das cinco janelas consumiram aproximadamente 14 minutos, e o intervalo da primeira leitura ao salvamento do modelo final foi de 65min44s. Esses valores mostram que a execução foi

A dualidade da Inteligência Artificial: a cadência do futuro?

possível no equipamento utilizado, mas podem mudar com o hardware, as versões dos programas e outras tarefas em execução.

6.2 Sensibilidade, importância e ablação

Com balanceamento automático, o F1 ponderado chegou a 0,8435 e a acurácia, a 0,8424. Em relação à configuração final, os aumentos foram de 0,48 e 0,32 ponto percentual, acompanhados por maior variação entre as janelas. Como o ganho médio foi pequeno, `class_weight=None` foi mantido por apresentar comportamento mais estável e uma configuração mais simples.

Na importância por ganho do LightGBM, PCTL_OT respondeu por 78,80%, DIAS_ABERTOS por 9,56%, TAREFA por 9,53% e ORGAO_JULGADOR por 2,11%. Esses valores mostram quais variáveis foram mais usadas para melhorar as divisões das árvores, mas não devem ser lidos isoladamente. PCTL_OT é calculada a partir das outras três informações; por isso, sua contribuição também foi testada com a retirada controlada da variável (Molnar *et al.*, 2024; Ibebuchi, 2025).

A contribuição de PCTL_OT foi examinada por uma ablação controlada. O LightGBM foi retreinado com os mesmos dados, filtros, classes, janelas, embargo, hiperparâmetros e semente, retirando-se apenas essa variável. Nessa configuração, o F1 ponderado médio passou de 0,8387 para 0,5688 e a acurácia média, de 0,8392 para 0,5730, diferenças de -26,99 e -26,62 pontos percentuais, respectivamente, observadas de modo consistente nas cinco janelas. O resultado indica que ORGAO_JULGADOR, TAREFA e DIAS_ABERTOS não recuperaram desempenho equivalente sem a posição percentual intragrupo. Assim, a evidência de C2 decorre da comparação temporal sob protocolo idêntico, e não da leitura isolada da importância interna.

A Tabela 5 separa a sensibilidade ao balanceamento do efeito da retirada de PCTL_OT, mantendo como referência a configuração final.

Tabela 5 – Sensibilidade ao balanceamento e ablação de PCTL_OT.

Configuração	F1 ponderado	Acurácia	Leitura
Final, sem ponderação	0,8387	0,8392	Maior estabilidade entre janelas.
Ponderação automática	0,8435	0,8424	Pequeno ganho médio e maior oscilação.
Final com PCTL_OT	0,8387	0,8392	Representação contextual preservada.
Ablação sem PCTL_OT	0,5688	0,5730	Diferenças: -26,99 e -26,62 p.p.

Fonte: elaboração própria a partir dos registros do experimento.

Esse resultado esclarece a função de PCTL_OT na base tabular: a variável não substitui DIAS_ABERTOS, mas acrescenta a posição contínua do processo na distribuição de seu próprio grupo. Assim, tempos absolutos semelhantes podem ter significados diferentes quando pertencem a órgãos e tarefas com ritmos distintos. Essa construção dialoga com abordagens de aprendizado de máquina que utilizam quantis como forma de representação dos dados (Dempster; Schmidt; Webb, 2024). Como a redução de desempenho ocorreu em uma ablação fora da amostra, mantendo-se dados, janelas, embargo, hiperparâmetros e semente, a evidência sustenta a contribuição preditiva de PCTL_OT no contexto estudado, sem atribuir efeito causal à variável.

A dualidade da Inteligência Artificial: a cadência do futuro?

6.3 Matrizes de confusão e padrões de erro

A Figura 4 reúne seis matrizes de confusão para EF Geral e seus recortes institucionais. O painel geral contém 164.165 registros pareados entre a carga de aplicação de 7 de novembro de 2025 e a observação de 14 de novembro. As matrizes comparam a classe prevista em t com a observada em $t+1$ para diagnosticar os padrões de erro. O volume pareado pode diferir do acervo da carga corrente, pois a presença e a elegibilidade dos processos podem mudar entre semanas. A estimativa de desempenho fora da amostra permanece a das cinco janelas temporais. As linhas indicam a classe observada, as colunas a prevista e os percentuais são calculados por linha.

Figura 4 – Matrizes de confusão entre QOT_PRED(t) e QOT_FUTURO($t+1$), por recorte operacional.

Linhas: classe observada em $t + 1$.

Colunas: classe prevista em t .



Fonte: elaboração própria a partir das saídas do sistema.

A dualidade da Inteligência Artificial: a cadência do futuro?

Em todos os recortes, a maioria das previsões coincidiu com a classe observada na semana seguinte, como mostram as células da diagonal. Entre as observações pertencentes a QOT4 nessa semana, o modelo havia previsto corretamente essa classe em 82,4% a 98,2% dos casos, conforme o recorte. Esses percentuais correspondem à revocação de QOT4. Os erros ocorreram principalmente entre classes vizinhas, envolvendo tanto QOT1+QOT2 e QOT3 quanto QOT3 e QOT4. A previsão de QOT1+QOT2 para observações que foram identificadas como QOT4 na semana seguinte foi menos frequente, representando entre 0,8% e 4,1% dos QOT4 observados.

6.4 Eficiência das listas TOP-N

A Tabela 6 compara as listas de CRIT_SCORE_% e DIAS_ABERTOS para capacidades de 4.000 e 8.000 processos. Em todos os cenários, a lista pelo escore exigiu a leitura de um volume igual ou menor para alcançar a mesma quantidade de classes previstas de interesse. O ganho variou de 0,89% a 39,52%. Os maiores valores apareceram para QOT4 nos Recortes A e B, quando QOT3 e QOT4 foram reunidos, os ganhos foram menores. Portanto, o resultado depende do recorte, da classe procurada e do tamanho do lote.

Tabela 6 – Ganho de triagem do CRIT_SCORE_% em relação a DIAS_ABERTOS.

K	Recorte	Alvo	Captura no escore	N por dias	Ganho
4.000	Geral	QOT4	4.000	4.345	7,94%
4.000	A	QOT4	4.000	6.614	39,52%
4.000	B	QOT4	4.000	4.427	9,65%
4.000	Geral	QOT3+QOT4	4.000	4.037	0,92%
4.000	A	QOT3+QOT4	4.000	4.895	18,28%
4.000	B	QOT3+QOT4	4.000	4.051	1,26%
8.000	Geral	QOT4	8.000	9.952	19,61%
8.000	A	QOT4	5.227	11.443	30,09%
8.000	B	QOT4	8.000	11.435	30,04%
8.000	Geral	QOT3+QOT4	8.000	8.072	0,89%
8.000	A	QOT3+QOT4	8.000	10.101	20,80%
8.000	B	QOT3+QOT4	8.000	8.417	4,95%

Fonte: elaboração própria. Nota: a captura usa a classe prevista; não representa impacto prospectivo confirmado.

No Recorte A, com capacidade de 8.000 e alvo QOT4, havia somente 5.296 registros previstos nessa classe, dos quais 5.227 entraram na lista. Para manter a comparação possível, a ordenação por DIAS_ABERTOS foi medida contra esses 5.227 casos, e não contra uma quantidade que não existia na carga.

6.5 Contribuições do estudo

A primeira contribuição prática é tornar visível a regra usada para formar a lista. Previsão, preferência de ordenação e capacidade de trabalho ficam separadas. Assim, o peso ou o tamanho do lote pode mudar sem que a lógica fique escondida do usuário.

A segunda contribuição é o protocolo temporal. Esse procedimento pode ser adaptado a outros sistemas alimentados por cargas periódicas.

A terceira contribuição é integrar contexto local e ordenação conjunta. Quartis e posição percentual são calculados sobre os dias de permanência, por órgão e tarefa, a cada semana. A

A dualidade da Inteligência Artificial: a cadência do futuro?

posição percentual integra o modelo, cujas probabilidades geram o escore. A ablação evidenciou a contribuição preditiva dessa variável, e a comparação das listas mostrou vantagem interna sobre a ordenação pelos dias, nos cenários avaliados.

A quarta contribuição é aproximar a avaliação do limite real de trabalho. Em vez de olhar apenas para F1 e acurácia, o estudo pergunta quantos casos precisam ser percorridos para montar uma lista de tamanho K. TOP-N responde a essa pergunta dentro do *ranking*. As matrizes, por sua vez, mostram quais erros merecem atenção durante a triagem.

6.6 Uso prático e cuidados

Em um uso futuro, CRIT_SCORE_% pode definir a ordem inicial da lista, enquanto DIAS_ABERTOS continua visível como referência. O lote deve passar por revisão humana antes de qualquer providência. Também precisam ser guardados a versão do modelo, a data da carga, o valor de K, o peso dado a QOT3 e as justificativas para alterações feitas pelo usuário.

K não deve ser fixado como se todas as equipes tivessem a mesma capacidade. Quando o lote aumenta, os casos mais evidentes já foram incluídos e a proporção de QOT4 tende a diminuir. Cada unidade precisa escolher K de acordo com os recursos disponíveis e acompanhar o benefício de ampliar a lista, o tempo de revisão e o risco de deixar casos relevantes de fora.

O acompanhamento deve observar a distribuição das classes, as métricas por período e recorte, os erros $QOT4 \rightarrow QOT1 + QOT2$, a qualidade das cargas e mudanças em órgãos ou tarefas. Uma alteração relevante exige investigação e pode levar à suspensão do uso, a novo treinamento ou a ajustes da política. Em qualquer cenário, a saída continua sendo uma recomendação de triagem, e não uma ordem sobre o processo.

7 LIMITAÇÕES DO ESTUDO

O estudo foi realizado com execuções fiscais de uma única organização. Os resultados não podem ser transferidos automaticamente para outras classes, unidades ou tribunais. Em outro contexto, os grupos órgão-tarefa precisam ser refeitos e o modelo deve passar por nova validação temporal. Os Recortes A e B valem somente para os subconjuntos observados.

O experimento concentrou-se no LightGBM, com análise de sensibilidade ao balanceamento e ablação de PCTL_OT. A comparação com modelos lineares, outras famílias de árvores e técnicas de calibração permanece como possibilidade de extensão, pois não foi realizada neste experimento.

As cargas semanais abrangem agosto a novembro de 2025 e preservam a ordem em que o acervo foi observado. A permanência do mesmo processo em semanas sucessivas faz parte da dinâmica estudada, o embargo impede o uso de informação posterior ao corte, embora não torne as observações independentes. Os resultados sustentam o comportamento do sistema no período e nos recortes analisados. Mudanças de tarefas, regras ou qualidade dos registros devem ser acompanhadas e, quando relevantes, exigir novo treinamento e nova validação temporal.

O código-fonte e a base institucional não são disponibilizados publicamente neste artigo, o que limita a reprodução independente do estudo. O ambiente computacional e os hiperparâmetros estão descritos na Seção 5.3, mas a reprodução depende do acesso às rotinas utilizadas e aos dados, cuja obtenção exige autorização institucional.

A dualidade da Inteligência Artificial: a cadência do futuro?

8 CONCLUSÕES

Como recomendações para a continuidade da pesquisa, propõe-se aplicar o protocolo temporal à comparação de outras famílias de modelos, avaliar a calibração probabilística, testar pesos alternativos para a classe relativa intermediária (QOT3) e repetir o estudo em outras unidades. As explicações locais devem considerar a dependência entre os atributos. A avaliação prospectiva deverá registrar a capacidade semanal de análise, as sugestões aceitas ou alteradas, o tempo de revisão e os resultados, permitindo comparar períodos ou grupos com e sem o sistema. Testes de qualidade das cargas, monitoramento de mudanças na distribuição dos dados e controle de versões também devem integrar essa etapa.

Como continuidade, propõe-se comparar as ordenações pelos dias de permanência, pela posição percentual corrente e pelo escore preditivo. As classes observadas na semana seguinte permitiriam avaliar a contribuição adicional da previsão. Também cabe investigar a aplicação da arquitetura em outros contextos, como demandas distribuídas entre regiões ou países com dinâmicas temporais distintas. A adaptação exige definir grupos de referência, medida temporal, horizonte de previsão e objetivo de priorização, com nova avaliação empírica.

O estudo demonstrou a viabilidade de executar localmente o sistema com dados processuais organizados em cargas semanais e de avaliá-lo respeitando sua ordem temporal. Nas cinco janelas, o LightGBM obteve F1 ponderado médio de 0,8387 e acurácia média de 0,8392. A ablação mostrou a contribuição da posição percentual intragrupo, enquanto as matrizes evidenciaram predominância de acertos, erros principalmente entre classes vizinhas e baixa frequência de previsões de QOT1+QOT2 para observações pertencentes a QOT4 na semana seguinte. A comparação das ordenações apresentou ganhos internos de triagem entre 0,89% e 39,52%, medidos a partir das classes previstas pelo modelo.

No contexto analisado, a pergunta de pesquisa foi respondida pela combinação da preparação das cargas, da representação do estado atual do processo, da previsão de sua classe relativa na semana seguinte e da ordenação sob capacidade limitada. O sistema produziu uma lista semanal para apoiar a priorização, preservando a possibilidade de conferir e reavaliar suas saídas. Essa leitura relativa complementa o acompanhamento dos dias de permanência na tarefa e não altera a classificação institucional de criticidade. A escolha dos processos a analisar e das providências cabíveis permanece sob responsabilidade humana.

9 REFERÊNCIAS

BOUSDEKIS, Alexandros et al. Modelling and predictive monitoring of business processes under uncertainty with reinforcement learning. **Sensors**, v. 23, n. 15, art. 6931, 2023. DOI: 10.3390/s23156931. Disponível em: <https://www.mdpi.com/1424-8220/23/15/6931>. Acesso em: 10 nov. 2025.

CERAVOLO, Paolo et al. Predictive process monitoring: concepts, challenges, and future research directions. **Process Science**, v. 1, art. 2, 2024. DOI: 10.1007/s44311-024-00002-4. Disponível em: <https://link.springer.com/article/10.1007/s44311-024-00002-4>. Acesso em: 9 nov. 2025.

CERQUEIRA, Vitor; TORGO, Luis; MOZETIČ, Igor. Evaluating time series forecasting models: an empirical study on performance estimation methods. **Machine Learning**, v. 109, n. 11, p. 1997–2028, 2020. DOI: 10.1007/s10994-020-05910-7. Disponível em: <https://link.springer.com/article/10.1007/s10994-020-05910-7>. Acesso em: 10 nov. 2025.

A dualidade da Inteligência Artificial: a cadência do futuro?

CHEONG, Ben Chester. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. **Frontiers in Human Dynamics**, v. 6, art. 1421273, 2024. DOI: 10.3389/fhumd.2024.1421273. Disponível em: <https://www.frontiersin.org/articles/10.3389/fhumd.2024.1421273>. Acesso em: 16 jan. 2026.

COLLINS, Gary S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. **BMJ**, v. 385, e078378, 2024. DOI: 10.1136/bmj-2023-078378. Disponível em: <https://www.bmj.com/content/385/bmj-2023-078378>. Acesso em: 10 nov. 2025.

CONSELHO NACIONAL DE JUSTIÇA. **Justiça em Números 2026**. Brasília: CNJ, 2026. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2026/06/justica-em-numeros-2026-1.pdf>. Acesso em: 6 jul. 2026.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 615, de 11 de março de 2025**. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília, DF: CNJ, 2025. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/6001>. Acesso em: 6 jul. 2026.

CRUZ, Jaqueline K. L. da et al. O uso da mineração de processos na análise do tempo das movimentações processuais. **Revista de Engenharia e Pesquisa Aplicada**, v. 9, n. 1, p. 97–104, 2023. DOI: 10.25286/rep.v9i1.2785. Disponível em: <https://doi.org/10.25286/rep.v9i1.2785>. Acesso em: 23 out. 2025.

DEMPSTER, A.; SCHMIDT, D. F.; WEBB, G. I. Quant: a minimalist interval method for time series classification. **Data Mining and Knowledge Discovery**, v. 38, p. 2377–2402, 2024. DOI: 10.1007/s10618-024-01036-9. Disponível em: <https://doi.org/10.1007/s10618-024-01036-9>. Acesso em: 16 jun. 2026.

HEVNER, Alan R.; MARCH, Salvatore T.; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75–105, 2004. DOI: 10.2307/25148625. Disponível em: <https://aisel.aisnet.org/misq/vol28/iss1/6/>. Acesso em: 10 nov. 2025.

HEWAMALAGE, Hansika; ACKERMANN, Klaus; BERGMEIR, Christoph. Forecast evaluation for data scientists: common pitfalls and best practices. **Data Mining and Knowledge Discovery**, v. 37, p. 788–832, 2023. DOI: 10.1007/s10618-022-00894-5. Disponível em: <https://link.springer.com/article/10.1007/s10618-022-00894-5>. Acesso em: 11 nov. 2025.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: principles and practice**. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 10 nov. 2025.

IBEBUCHI, Chibuike Chiedozie. Uncertainty in machine learning feature importance for climate science: a comparative analysis of SHAP, PDP, and gain-based methods. **Theoretical and Applied Climatology**, v. 156, art. 476, 2025. DOI: 10.1007/s00704-025-05703-9. Disponível em: <https://doi.org/10.1007/s00704-025-05703-9>. Acesso em: 16 jun. 2026.

KAPOOR, Sayash et al. REFORMS: consensus-based recommendations for machine-learning-based science. **Science Advances**, v. 10, n. 18, eadk3452, 2024. DOI: 10.1126/sciadv.adk3452. Disponível em: <https://www.science.org/doi/10.1126/sciadv.adk3452>. Acesso em: 10 nov. 2025.

A dualidade da Inteligência Artificial: a cadência do futuro?

KE, Guolin et al. LightGBM: a highly efficient gradient boosting decision tree. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS**, v. 30, p. 3146–3154, 2017. Disponível em: <https://proceedings.neurips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf>. Acesso em: 9 nov. 2025.

LAGE-FREITAS, André et al. Predicting Brazilian court decisions. **PeerJ Computer Science**, v. 8, e904, 2022. DOI: 10.7717/peerj-cs.904. Disponível em: <https://peerj.com/articles/cs-904/>. Acesso em: 9 nov. 2025.

LIMA, Ricardo Massa F. et al. Mineração de processos no Judiciário brasileiro. **Computação Brasil**, Sociedade Brasileira de Computação, p. 39–43, abr. 2023. Disponível em: <https://journals-sol.sbc.org.br/index.php/comp-br/article/view/4061>. Acesso em: 16 out. 2025.

MOLNAR, Christoph; KÖNIG, Gunnar; BISCHL, Bernd; CASALICCHIO, Giuseppe. Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach. **Data Mining and Knowledge Discovery**, v. 38, p. 2903–2941, 2024. DOI: 10.1007/s10618-022-00901-9. Disponível em: <https://doi.org/10.1007/s10618-022-00901-9>. Acesso em: 16 jun. 2026.

OECD. Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. Paris: **OECD Publishing**, 2025. DOI: 10.1787/795de142-en. Disponível em: https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html. Acesso em: 9 nov. 2025.

OLIVEIRA, Raphael Souza de et al. Predicting the number of days in court cases using artificial intelligence. **PLOS ONE**, v. 17, n. 5, e0269008, 2022. DOI: 10.1371/journal.pone.0269008. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269008>. Acesso em: 9 nov. 2025.

RAINIO, Oona; TEUHO, Jarmo; KLÉN, Riku. Evaluation metrics and statistical tests for machine learning. **Scientific Reports**, v. 14, art. 6086, 2024. DOI: 10.1038/s41598-024-56706-x. Disponível em: <https://doi.org/10.1038/s41598-024-56706-x>. Acesso em: 10 nov. 2025.

RESCK, Lucas et al. LegalAnalytics: bridging visual explanations and workload streamline in Brazilian Supreme Court appeals. **Artificial Intelligence and Law**, 2025. DOI: 10.1007/s10506-025-09446-w. Disponível em: <https://link.springer.com/article/10.1007/s10506-025-09446-w>. Acesso em: 10 nov. 2025.

RIM, Patrick; LIU, Erin. Optimizing the C4.5 decision tree algorithm using MSD-Splitting. **International Journal of Advanced Computer Science and Applications (IJACSA)**, [S. l.], v. 11, n. 10, 2020. DOI: 10.14569/IJACSA.2020.0111006. Disponível em: <https://thesai.org/Publications/ViewPaper?Code=IJACSA&Issue=10&SerialNo=6&Volume=11>. Acesso em: 9 nov. 2025.

SEMMELOCK, Harald et al. Reproducibility in machine-learning-based research: overview, barriers, and drivers. **AI Magazine**, v. 46, n. 2, e70002, 2025. DOI: 10.1002/aaai.70002. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/aaai.70002>. Acesso em: 26 fev. 2026.

SOBRINO, Fernanda et al. **Designing AI for prosecutorial governance: case prioritization and statutory oversight in Mexico**. arXiv:2601.00396, 2026. DOI: 10.48550/arXiv.2601.00396. Disponível em: <https://arxiv.org/abs/2601.00396>. Acesso em: 25 maio 2026.

A dualidade da Inteligência Artificial: a cadência do futuro?

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business intelligence e análise de dados para gestão do negócio**. 4. ed. Porto Alegre: Bookman, 2019.

VAN DER AALST, Wil M. P.; CARMONA, Josep (org.). **Process Mining Handbook**. Cham: Springer Nature, 2022. DOI: 10.1007/978-3-031-08848-3. Disponível em: <https://link.springer.com/book/10.1007/978-3-031-08848-3>. Acesso em: 25 out. 2025.

VOM BROCKE, Jan; GAU, Michael; MÄDCHE, Alexander. Journaling the design science research process: transparency about the making of design knowledge. In: CHANDRA KRUSE, L.; SEIDEL, S.; HAUSVIK, G. I. (org.). **The next wave of sociotechnical design**. Cham: Springer, 2021. p. 131–136. DOI: 10.1007/978-3-030-82405-1_15. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-82405-1_15. Acesso em: 10 nov. 2025.

WHEELAN, Charles. **Estatística: o que é, para que serve, como funciona**. 1. ed. Rio de Janeiro: Zahar, 2016. ISBN 9788537815120.

ZUIDERWIJK, Anneke; CHEN, Yu-Che; SALEM, Fadi. Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda.

Government Information Quarterly, v. 38, n. 3, 101577, 2021. DOI:

10.1016/j.giq.2021.101577. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0740624X21000137>. Acesso em: 9 nov. 2025.