Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
MBA em *Advanced Analytics & Business Optimization*

David Robert de Oliveira

# MODELAGEM E GESTÃO DE ACERVO PROCESSUAL: PREDIÇÃO E PRIORIZAÇÃO NA SEGUNDA INSTÂNCIA DO TJRJ

**Curitiba**
**2025**

David Robert de Oliveira

# MODELAGEM E GESTÃO DE ACERVO PROCESSUAL: PREDIÇÃO E PRIORIZAÇÃO NA SEGUNDA INSTÂNCIA DO TJRJ

Monografia apresentada ao MBA em *Advanced Analytics & Business Optimization* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Cassius Tadeu Scarpin

Curitiba
2025

# Modelagem e Gestão de Acervo Processual: Predição e Priorização na Segunda Instância do TJRJ

**Modeling and Managing Case Backlog: Prediction and Prioritization in the TJRJ Appellate Court**

David Robert de Oliveira[1], Cassius Tadeu Scarpin[2]

[1]Departamento de Gestão da Informação, Universidade Federal do Paraná, Av. Prefeito Lothário Meissner, 632, Jardim Botânico - Curitiba/PR - 80210-070*
[2]Professor do Departamento de Estatística - DEST/UFPR, cassiusts@ufpr.br

A gestão de acervo processual na segunda instância do TJRJ enfrenta o desafio de priorizar milhares de processos considerando metas institucionais, prioridades legais e capacidade operacional dos órgãos julgadores. Neste artigo, buscou-se construir um sistema de apoio à decisão capaz de estimar o tempo de tramitação de cada processo e gerar rankings de priorização. Para isso, foram utilizadas técnicas de aprendizado de máquina aplicadas a dados públicos do DataJud, selecionando automaticamente o algoritmo de melhor desempenho para cada grupo de processos. A partir disso, foi possível implementar um protótipo funcional que produz relatórios estruturados por órgão julgador, ordenando o acervo pendente conforme critérios de maturidade processual e elegibilidade para metas do CNJ. A abordagem pode ser aprimorada com a incorporação de novas fontes de dados e técnicas de modelagem mais avançadas.
**Palavras-chave:** Priorização Processual, Aprendizado de Máquina, Gestão Judicial, TJRJ

Managing the caseload at the appellate level of TJRJ involves prioritizing thousands of cases while considering institutional targets, legal priorities, and the operational capacity of adjudicative bodies. This paper describes the development of a decision-support system capable of estimating processing time for each case and generating prioritization rankings. Machine learning techniques were applied to public DataJud data, automatically selecting the best-performing algorithm for each group of cases. A functional prototype was implemented that produces structured reports by adjudicative body, ranking pending cases according to procedural maturity criteria and eligibility for CNJ targets. The approach can be enhanced by incorporating additional data sources and more advanced modeling techniques.
**Keywords:** Case Prioritization, Machine Learning, Judicial Management, TJRJ

## 1. Introdução

O Tribunal de Justiça do Estado do Rio de Janeiro (TJRJ) destaca-se nacionalmente por sua produtividade. Segundo o relatório Justiça em Números 2024 [1], cada magistrado do TJRJ, considerando o primeiro e o segundo graus de jurisdição, baixou em média 4.128 processos em 2023, desempenho 60% superior à média nacional de 2.574 processos. O tribunal alcançou pela 15ª vez consecutiva o maior índice de produtividade entre os tribunais estaduais, com 69% de eficiência no segundo grau conforme o Índice de Produtividade Comparada da Justiça (IPC-Jus), métrica que relaciona casos solucionados e pendentes ponderados pelo estoque inicial, permitindo comparação entre tribunais de diferentes portes.

A segunda instância do TJRJ conta com 210 desembargadores (ampliados de 180 em setembro de 2024)[1] distribuídos em 40 câmaras especializadas: 22 de Direito Privado, 10 de Direito Público (incluindo as 9ª e

[1]AMAERJ: Estado do Rio de Janeiro terá 20 novos desembargadores. Disponível em: https://amaerj.org.br/noticias/estado-do-rio-de-janeiro-tera-20-novos-desembargadores/. Acesso em: 27 nov. 2025.

*contato@david.dev.br

10ª inauguradas em janeiro de 2025)[2] e 8 Criminais[3]. Esta estrutura processa anualmente centenas de milhares de recursos através de dois sistemas eletrônicos distintos: o ejud, que gerencia o acervo legado, e o eproc, plataforma modernizada para novos processos.

O volume processual, no entanto, permanece desafiador. Dados do CNJ indicam que o TJRJ mantinha 7,3 milhões de processos pendentes em abril de 2023, tendo julgado 382.642 casos no primeiro quadrimestre daquele ano[2]. A taxa de congestionamento no segundo grau alcança 32%, relativamente baixa comparada a outros tribunais, mas ainda representando significativo acervo residual[4].

As metas nacionais do CNJ direcionam esses esforços institucionais. A Meta 1 exige julgar mais processos que os distribuídos no ano corrente. A Meta 2 estabelece prazos para julgamento de processos antigos, variando conforme matéria e instância. Estes objetivos competem por recursos limitados, exigindo estratégias de priorização que equilibrem produtividade quantitativa com qualidade decisória.

### 1.1. Caracterização do Problema

Como priorizar julgamento e baixa de processos na segunda instância, considerando a carga de trabalho estimada para cada processo, prioridades legais e metas institucionais simultâneas?

O desafio central da gestão de acervo judicial está em identificar, com segurança, quais processos estão efetivamente próximos de serem julgados ou baixados. Sem essa informação, a priorização torna-se arbitrária ou baseada apenas em critérios formais (como antiguidade), ignorando a realidade operacional de cada caso. A proposta deste trabalho é estimar, a partir das características de cada processo, o tempo médio que ele levará para ser concluído — permitindo assim priorizar aqueles com maturidade processual adequada para decisão.

Para abordar este problema, propõe-se um sistema de apoio à decisão que integra modelagem preditiva do tempo para julgar ou baixar com priorização multicritério. O sistema estrutura o ciclo de vida processual em três eixos: (1) **Distribuição**, que compreende a atribuição de processos aos órgãos julgadores — fenômeno observado, não decisão a otimizar neste trabalho; (2) **Julgamento**, que abrange a prolação de decisões em sessão colegiada — alvo de priorização para cumprimento das Metas 1 e 2 do CNJ; e (3) **Baixa**, que corresponde ao encerramento definitivo dos processos — também alvo de priorização para reduzir a taxa de congestionamento.

### 1.2. Objetivos

O objetivo é desenvolver um sistema de apoio à decisão para gestão do acervo processual na segunda instância do TJRJ. Especificamente: (a) treinar modelos preditivos para estimar tempo até julgamento e baixa; (b) formular função de scoring combinando maturidade, metas e prioridades legais; e (c) implementar pipeline que gera rankings por órgão julgador.

## 2. Experiências no Judiciário

A aplicação de métodos quantitativos ao Direito — conhecida como Jurimetria, definida por Marcelo Nunes como "a disciplina do conhecimento que utiliza a metodologia estatística para investigar o funcionamento de uma ordem jurídica"[3] cap. 5 — e técnicas de Pesquisa Operacional[4] têm fundamentado sistemas de apoio à decisão em organizações judiciárias. Diversos tribunais brasileiros implementaram soluções específicas para enfrentar desafios relacionados à gestão de acervo processual, produtividade e cumprimento de metas institucionais.

### 2.1. Iniciativas dos Tribunais Brasileiros

O Tribunal de Justiça do Rio Grande do Norte desenvolveu o GPS-Jus, ferramenta que fornece informações estatísticas de todas as unidades judiciárias e monitora cumprimento das metas CNJ[5]. O sistema permite acesso à quantidade de processos por meta e envia automaticamente dados de produtividade para o CNJ.

[2]Consultor Jurídico: TJ do Rio inaugura duas novas câmaras de Direito Público. Disponível em: https://www.conjur.com.br/2025-jan-24/tj-rj-inaugura-camaras-de-direito-publico-e-observatorio-penal-em-prisao/. Acesso em: 27 nov. 2025.

[3]TJRJ: 2ª Instância - Órgãos Julgadores. Disponível em: https://www.tjrj.jus.br/servicos/processo-eletronico/rel-orgaos-julg-eletr/2-instancia. Acesso em: 27 nov. 2025.

[4]TJRJ: Painel de Produtividade da 2ª Instância. Disponível em: https://www.tjrj.jus.br/estatisticas/segunda-instancia/painel-de-produtividade-da-2-instancia. Acesso em: 27 nov. 2025.

[5]Tribuna da Justiça: Sistema GPS-Jus permite acompanhamento de metas do CNJ por unidades judiciárias do TJRN. Disponível em: https://tribunadajustica.com.br/sistema-gps-jus-permite-acompanhamento-de-metas-do-cnj-por-unidades-judiciarias-do-tjrn/. Acesso em: 27 nov. 2025.

A solução atraiu atenção de outros tribunais, incluindo visita técnica da Corregedoria do TJMA.

O Tribunal de Justiça da Bahia implementou Painel de Priorização que atribui peso a cada processo, estabelecendo ranking baseado em múltiplos critérios alinhados com o Prêmio Qualidade CNJ[6]. Resultados incluem redução de mais de 100 processos na 1ª Vara Cível de Salvador, substituindo critérios subjetivos por ranking quantitativo.

### 2.2. Coordenação pelo CNJ

O Conselho Nacional de Justiça coordena o Programa Justiça 4.0, desenvolvido com o PNUD desde 2020, que busca integrar sistemas judiciais em plataforma unificada[7]. O número de projetos de IA cresceu de 112 em 2022 para 178 em 2024. A ferramenta Apoia, primeira IA generativa integrada à Plataforma Digital do Poder Judiciário, auxilia na criação de relatórios, revisão de textos e detecção de litigância predatória.

A Justiça do Trabalho implementou diversos sistemas nos Tribunais Regionais, incluindo análise preditiva (TRT 1ª Região), clustering de processos (TRT 4ª Região) e sistema Gemini para agrupamento de recursos (TRTs 5ª, 7ª, 15ª e 20ª)[8]. O TST criou o sistema Bem-te-vi para gestão processual em gabinetes, implementado desde 2018.

## 3. Modelagem Conceitual

O sistema de apoio à decisão proposto fundamenta-se em quatro pilares: o ciclo de vida processual, as metas institucionais do CNJ, a arquitetura de predição e priorização, e a lógica de scoring que combina esses elementos.

[6]TJBA: 1ª Vara Cível de Salvador reduz mais de 100 processos com o uso do Painel de Priorização de Processos. Disponível em: https://www.tjba.jus.br/portal/1a-vara-civel-de-salvador-reduz-mais-de-100-processos-com-o-uso-do-painel-de-priorizacao-de-processos/. Acesso em: 27 nov. 2025.

[7]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-pesquisa-sobre-ia-no-judiciario-brasileiro/. Acesso em: 27 nov. 2025.

[8]CSJT: Justiça do Trabalho está na vanguarda da inovação e inteligência artificial. Disponível em: https://www.csjt.jus.br/web/csjt/-/justica-do-trabalho-esta-na-vanguarda-da-inovacao-e-inteligencia-artificial. Acesso em: 27 nov. 2025.

### 3.1. Ciclo de Vida Processual

O sistema estrutura o ciclo de vida processual em três eixos sequenciais, conforme ilustrado na Figura 1. O **Eixo 1 (Distribuição)** compreende a atribuição de processos a relatores, sendo fenômeno observado e não decisão a otimizar neste trabalho. O **Eixo 2 (Julgamento)** abrange a inclusão em pauta e sessão colegiada, constituindo alvo de priorização. O **Eixo 3 (Baixa)** corresponde ao trânsito em julgado e arquivamento definitivo, também sendo alvo de priorização.




**Figura 1:** Ciclo de vida processual: da distribuição à baixa

### 3.2. Metas Institucionais

As Metas Nacionais do CNJ constituem o principal instrumento de gestão estratégica dos tribunais[5]. Para a segunda instância, as principais metas são: **Meta 1**, que exige julgar mais processos que os distribuídos no ano corrente; **Meta 2**, que determina julgar 90% dos processos distribuídos até 31/12/2022; e **Meta 2 Ant**, que estabelece julgar 100% dos processos com mais de 15 anos.

### 3.3. Estratificação por Período de Metas

As Metas Nacionais do CNJ estabelecem prazos diferenciados para julgamento de processos conforme sua antiguidade. Essa diferenciação institucional cria três horizontes temporais com significados e tratamentos distintos, apresentados na Tabela 1.

**Tabela 1:** Definição dos grupos temporais de metas

| Grupo | Critério | Meta |
|---|---|---|
| Meta 2 Ant | Até 31/12/2010 | 100% |
| Meta 2 | 01/01/2011 a 31/12/2022 | 90% |
| Pós-Meta 2 | A partir de 01/01/2023 | CN |

A estratificação por período de metas reflete a própria lógica institucional do CNJ: cada grupo sofre cobranças distintas, processos antigos têm características sistematicamente diferentes (classes predominantes, complexidade), e um único modelo para todo o acervo pode ter desempenho inferior a modelos específicos por grupo. A Figura 2 visualiza essa estratificação temporal.

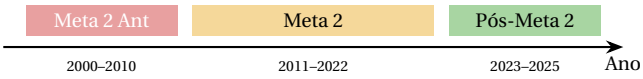



**Figura 2:** Estratificação por período de distribuição

### 3.4. Arquitetura do Sistema

O sistema de apoio à decisão integra duas etapas metodológicas, ilustradas nas Figuras 3 e 4. Na **Etapa 1 (Predição)**, três algoritmos são comparados para estimar o tempo de tramitação: Random Forest, XGBoost e Elastic Net. O modelo com menor erro (MAE[9]) é selecionado para cada grupo temporal. Na **Etapa 2 (Priorização)**, o tempo estimado é combinado com critérios institucionais para gerar um ranking de prioridade.




**Figura 3:** Etapa 1: Pipeline de predição

Uma vez selecionado o melhor modelo para cada grupo, a Etapa 2 aplica as previsões ao acervo pendente e calcula o score de priorização.

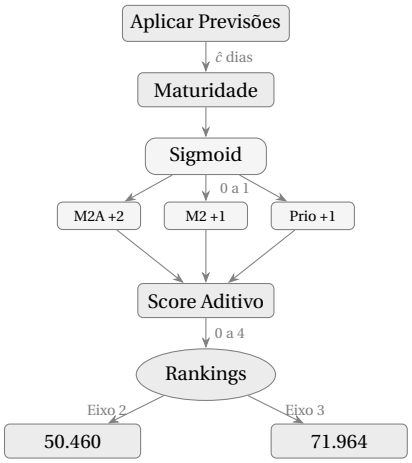



**Figura 4:** Etapa 2: Pipeline de priorização com score aditivo

[9]MAE (*Mean Absolute Error*): média do valor absoluto dos erros de previsão, em dias. Quanto menor, melhor o modelo.

### 3.5. Fonte de Dados: DataJud

O DataJud é a Base Nacional de Dados do Poder Judiciário, instituída pela Resolução CNJ nº 331/2020[6]. Contém movimentações processuais de todos os tribunais brasileiros em formato padronizado. Para este trabalho, foram extraídos **1.675.350 processos** da segunda instância do TJRJ, abrangendo o período de 2019 a 2025. As principais variáveis utilizadas incluem identificadores de órgão e classe processual, datas de distribuição e julgamento, além de flags indicando prioridades legais como violência doméstica, feminicídio e infância.

### 3.6. Acervo Ativo e Inativo

A base de dados contém tanto processos concluídos quanto pendentes. O **Acervo Inativo** compreende processos baixados, identificados pelo preenchimento da data de primeira baixa. O **Acervo Ativo** compreende processos pendentes, subdivididos em: **Pendente de Julgamento** (sem data de resolução, Eixo 2) e **Pendente de Baixa** (com resolução mas sem baixa, Eixo 3).

Essa distinção é fundamental para a modelagem: o treino utiliza o Acervo Inativo para aprender padrões históricos, enquanto a previsão é aplicada no Acervo Ativo para estimar tempo restante. Esta separação evita vazamento de informação (*data leakage*).

### 3.7. Modelagem Preditiva

O objetivo da modelagem é estimar o tempo típico de tramitação ($\hat{c}_i$) para cada processo pendente. Três algoritmos são comparados: **Random Forest** (ensemble de árvores, resistente a outliers), **XGBoost** (gradient boosting, alta performance) e **Elastic Net** (regressão regularizada, interpretável). Os modelos são treinados separadamente para cada grupo temporal, evitando que padrões de processos antigos distorçam previsões de processos recentes.

### 3.8. Maturidade Processual

O elemento central do sistema é a **maturidade processual**: a relação entre tempo decorrido e tempo esperado. Para cada processo $i$, define-se:

$$\text{tempo_faltando}_i = \hat{c}_i - \text{tempo_hoje}_i$$

onde $\hat{c}_i$ é a previsão do modelo e $\text{tempo_hoje}_i$ é o tempo desde a distribuição. Valores positivos altos indicam processos recém-chegados; valores próximos de zero indicam processos no ponto esperado; valores

negativos indicam processos que passaram do ponto e devem ser priorizados.

### 3.9. Lógica de Priorização

O sistema combina a maturidade com critérios institucionais através de um **score aditivo**:

$$S_i = s_i^{tempo} + w_i^{grupo} + w_i^{prioridade}$$

O score varia de 0 a 4, onde cada componente contribui de forma independente. A **maturidade temporal** ($s^{tempo}$) é normalizada via função sigmoid: $s_i^{tempo} = 1/(1 + e^{\text{tempo_faltando}_i/k})$, com $k = 30$ dias. A Figura 5 ilustra o comportamento desta função: processos atrasados (tempo_faltando negativo) recebem scores próximos de 1, enquanto processos recentes recebem scores próximos de 0.

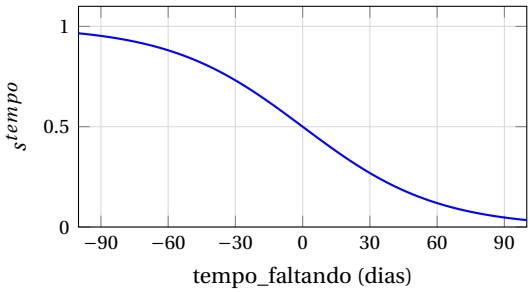



**Figura 5:** Função sigmoid aplicada à maturidade processual ($k = 30$)

O **peso do grupo** ($w^{grupo}$) atribui 2 pontos para Meta 2 Ant, 1 ponto para Meta 2 e 0 para Pós-Meta 2. A **prioridade legal** ($w^{prioridade}$) adiciona 1 ponto para processos com marcadores de violência doméstica, feminicídio ou infância.

Esta abordagem aditiva permite balanceamento flexível: um processo antigo sem prioridade legal pode ter score equivalente a um processo novo com prioridade.

## 4. Dados e Análise

A análise exploratória segue a estratificação por grupos temporais definida na seção anterior. Gráficos e tabelas são apresentados separadamente para cada grupo — Meta 2 Ant, Meta 2 e Pós-Meta 2 — evidenciando padrões relevantes à modelagem preditiva.

### 4.1. Base de Dados

Os dados foram obtidos do DataJud (Resolução CNJ nº 331/2020) através do portal Justiça em Números[10]. A Tabela 2 resume as características da base.

**Tabela 2:** Características da base de dados

| Parâmetro | Valor |
|---|---|
| Tribunal | TJRJ (Segunda Instância) |
| Tabela | `metas_g2` |
| Registros | 1.675.350 processos |
| Ano | 2025[11] |

### 4.2. Composição do Acervo

A distribuição por período de metas (Figura 6) revela três perfis: Meta 2 Ant com 4,0% (66.887), Meta 2 com 52,9% (886.377) e Pós-Meta 2 com 43,1% (722.086).

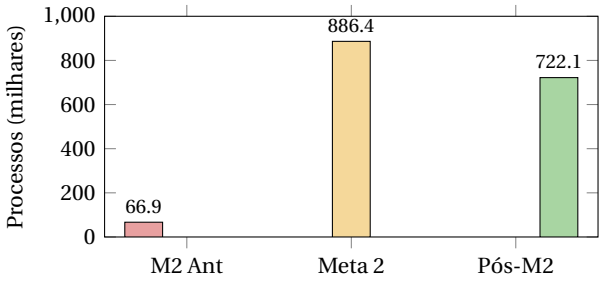



**Figura 6:** Composição do acervo por grupo de metas (N = 1.675.350)

### 4.3. Situação Processual

Cada processo pode estar em uma de três situações: **Baixado** (concluído), **Pendente de Baixa** (julgado, aguardando trânsito) ou **Pendente de Julgamento** (aguardando decisão). A Tabela 3 apresenta a distribuição.

O Meta 2 Ant tem apenas 0,7% de pendência, mas sua meta de 100% torna críticos os 471 processos restantes. O Meta 2 tem 3,1% e o Pós-Meta 2 tem 14,3% (ciclo natural). A Figura 7 visualiza essa distribuição.

[10] Disponível em: `https://justica-em-numeros.cnj.jus.br/painel-estatisticas/`. Acessar aba "Metas" → "Downloads" → selecionar TJRJ → filtrar por grau G2 (segunda instância). Acesso em: 24-26 nov. 2025.

[11] A base inclui registros históricos de acervo inativo desde 1907. Ano de referência: 2025.

**Tabela 3:** Situação processual por grupo (%)

| Grupo | Baixado | Pend. Baixa | Pend. Julg. |
|---|---|---|---|
| Meta 2 Ant | 99,3 | 0,4 | 0,3 |
| Meta 2 | 96,9 | 2,0 | 1,1 |
| Pós-Meta 2 | 85,7 | 7,0 | 7,3 |

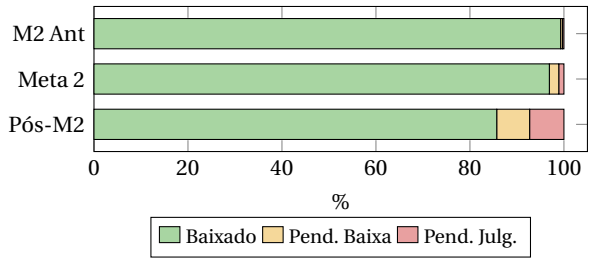



**Figura 7:** Situação processual por grupo (%)

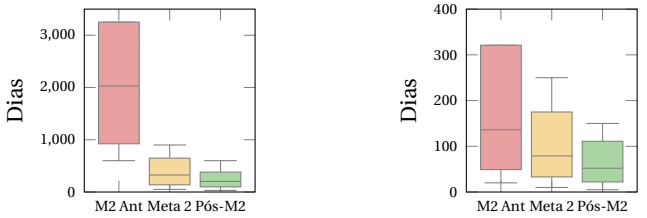
### 4.4. Tempo de Tramitação

A análise do tempo de tramitação considera dois eixos: **Eixo 2** (tempo até julgamento) e **Eixo 3** (tempo do julgamento até a baixa). A Tabela 4 apresenta as estatísticas descritivas para processos já baixados (Acervo Inativo), que servem de base para a modelagem preditiva.

**Tabela 4:** Tempo de tramitação por eixo (dias)

| Eixo | Grupo | Med. | Q25 | Q75 |
|---|---|---|---|---|
| 2 | M2 Ant | 2.029 | 923 | 3.251 |
| | Meta 2 | 325 | 138 | 648 |
| | Pós-M2 | 202 | 99 | 381 |
| 3 | M2 Ant | 136 | 49 | 321 |
| | Meta 2 | 79 | 33 | 175 |
| | Pós-M2 | 52 | 22 | 111 |

O Eixo 2 domina o ciclo de tramitação: processos do grupo Meta 2 Ant levam em mediana 2.029 dias até julgamento, enquanto o Eixo 3 acrescenta apenas 136 dias. Esta assimetria justifica o foco prioritário na predição do tempo até julgamento. A variação entre grupos confirma a necessidade de modelos estratificados, pois um modelo único não capturaria as diferenças de perfil temporal entre processos antigos e recentes. A Figura 8 ilustra essa distribuição.




**(a)** Eixo 2: Tempo até Julgamento

**(b)** Eixo 3: Tempo até Baixa

**Figura 8:** Distribuição do tempo de tramitação por grupo (Acervo Inativo)

### 4.5. Variáveis para Modelagem

Das 173 variáveis da base, apenas 14 são candidatas à modelagem preditiva: 5 características do processo e 9 flags de prioridade legal. A maioria (131 variáveis) são derivadas retrospectivamente para cálculo de metas e não podem ser usadas como preditores. A Tabela 5 lista as variáveis selecionadas.

**Tabela 5:** Variáveis selecionadas

| Variável | Tipo | Papel |
|---|---|---|
| `grupo` | Categ. (3) | Estratif. |
| `id_ultimo_oj` | Categ. (~100) | Feature |
| `id_ultima_classe` | Categ. (~50) | Feature |
| `ano` | Numérica | Feature |
| `recurso` | Categ. (2) | Feature |
| `flg_*` | Binária | Scoring |

A variável `id_assunto` foi excluída apesar de sua relevância conceitual. São aproximadamente 2.500 valores distintos, com 71% aparecendo menos de 100 vezes na base. Esta alta cardinalidade e distribuição desbalanceada introduzem ruído estatístico que prejudica a generalização do modelo. Em testes empíricos, substituir assunto por `ano` e `recurso` melhorou o desempenho preditivo, pois estas variáveis capturam parte da informação com menor variância.

A variável `grupo` não é feature, mas estratificador: o modelo é treinado separadamente por grupo, evitando vazamento de informação e permitindo capturar padrões específicos de cada período.

### 4.6. Qualidade dos Dados

A variável `dt_recebimento` é essencial para o cálculo do tempo de tramitação. Entretanto, 1,06% dos processos (17.763) não possuem esta data. No acervo pendente de julgamento, a taxa sobe para 7,45% (4.064 de

54.524 processos). As classes mais afetadas são Apelação Cível (62%) e Apelação/Remessa Necessária (15%). Não há datas alternativas disponíveis para imputação. O tratamento adotado foi excluir estes processos do pipeline de modelagem, documentando a limitação.

### 4.7. Síntese

A análise exploratória revela padrões consistentes que orientam a modelagem. Os grupos temporais apresentam perfis distintos de tempo e pendência, justificando modelos separados. O Eixo 2 domina a tramitação total, direcionando o foco preditivo. As variáveis classe, órgão julgador, ano e recurso são determinantes do tempo, enquanto prioridades legais aceleram o julgamento e devem integrar diretamente o scoring. A alta cardinalidade do assunto levou à sua exclusão em favor de variáveis mais estáveis. Estes achados fundamentam a implementação do pipeline descrito na próxima seção.

## 5. Predição e Priorização

O **OPTJUS** é o sistema desenvolvido como produto desta pesquisa. Implementado em R com os frameworks tidyverse e tidymodels, integra modelagem preditiva e priorização de processos judiciais com foco no cumprimento das metas do CNJ. O código-fonte está disponível em repositório público[12].

### 5.1. Limitações

O sistema apresenta limitações em duas frentes. Quanto aos **dados**, o DataJud apresenta inconsistências cadastrais reconhecidas pelo próprio CNJ, que disponibiliza ferramentas para saneamento contínuo da base. Os dados foram tratados conforme descrito na Seção 4, aplicando filtros de qualidade. Quanto aos **modelos**, o tempo de tramitação depende de fatores não capturáveis em bases estruturadas — complexidade fática, comportamento das partes, carga momentânea do relator, eventos imprevisíveis. Os modelos explicam apenas parte da variância observada (MAE entre 32 e 199 dias conforme o grupo). O sistema foi concebido como **demonstração de conceito**, estabelecendo arquitetura passível de aperfeiçoamento.

[12]Repositório OPTJUS: `https://github.com/david-dev-br/OPTJUS`. O projeto está aberto para aperfeiçoamentos, testes e contribuições da comunidade.

### 5.2. Demonstração do Pipeline

O OPTJUS opera em duas etapas complementares. Na primeira, **modelos preditivos** são treinados com dados históricos para estimar o tempo esperado de tramitação. Três algoritmos são comparados — Random Forest, XGBoost e Elastic Net — e o melhor para cada combinação de grupo e eixo é selecionado pelo menor MAE.

Na segunda etapa, as previsões alimentam um **sistema de priorização** baseado em score aditivo. Cada processo recebe pontuação de 0 a 4 combinando: maturidade temporal via função sigmoid, peso de grupo conforme elegibilidade para Meta 2, e flag de prioridade legal. O ranking ordena processos do maior para o menor score.

O pipeline principal foi executado com os dados do TJRJ. Após aplicação dos filtros de qualidade (cf. Seção 4.6), o carregamento processou **1.657.587 registros** da tabela `metas_g2`. Após aplicação dos filtros para identificar processos pendentes, foram encontrados 50.460 processos pendentes de julgamento (Eixo 2) e 71.964 processos julgados pendentes de baixa (Eixo 3).

### 5.3. Modelos Selecionados

Para cada combinação de grupo temporal e eixo, o pipeline comparou três algoritmos e selecionou o melhor pelo menor MAE. A Tabela 6 apresenta a performance completa de todos os modelos testados.

**Tabela 6:** Performance dos modelos: MAE (dias) e $R^2$

| Grupo | Eixo | RF | XGB | EN |
|---|---|---|---|---|
| M2 Ant | 2 | **199 / 8%** | 215 / 6% | 248 / 3% |
| | 3 | **177 / 8%** | 182 / 4% | 188 / 6% |
| Meta 2 | 2 | 96 / 25% | **96 / 25%** | 108 / 15% |
| | 3 | 76 / 20% | **76 / 18%** | 81 / 7% |
| Pós-M2 | 2 | 42 / 32% | **40 / 36%** | 47 / 19% |
| | 3 | 34 / 22% | **32 / 24%** | 36 / 15% |

O XGBoost apresentou melhor desempenho nos grupos Meta 2 e Pós-Meta 2, enquanto o Random Forest foi superior para processos do grupo Meta 2 Ant. O baixo $R^2$[13] dos modelos (8% a 36%) reflete a alta variabilidade inerente aos tempos de tramitação judicial, influenciados por fatores não capturáveis em dados

[13]$R^2$ (coeficiente de determinação): proporção da variância explicada pelo modelo. Valores próximos de 1 indicam ajuste perfeito; valores baixos indicam que fatores não modelados influenciam o resultado.

estruturados. O MAE, métrica principal de seleção, variou de 32 dias (Pós-Meta 2, Eixo 3) a 199 dias (Meta 2 Ant, Eixo 2).

## 6. Relatórios por Órgão Julgador

Os rankings globais do OPTJUS podem ser transformados em listas de trabalho personalizadas para cada câmara ou gabinete. O template de relatório gera documentos HTML individuais com a priorização de processos pendentes por órgão julgador.

### 6.1. Geração de Relatórios

O template aceita parâmetros para personalização: código do órgão julgador (ex: 85612) e data do ranking (YYYYMMDD). A renderização pode ser feita individualmente para um órgão específico ou em lote para os órgãos com maior volume.

Cada relatório inclui: **Panorama do Acervo** (volume total, processos por grupo temporal, prioridades legais), **Composição por Classe e Assunto** (distribuição das principais classes processuais), **Análise de Maturidade** (gráfico relacionando tempo decorrido vs. tempo estimado), e **Rankings de Julgamento e Baixa** (listas ordenadas por score com exportação para Excel).

### 6.2. Estrutura do Relatório

A Figura 9 ilustra a estrutura do relatório gerado. Os indicadores consolidados incluem total de processos, volume por grupo de metas, prioridades legais e processos com score crítico (≥ 3,0).




**Figura 9:** Exemplo de relatório por órgão julgador

Os rankings são apresentados em tabelas separadas para julgamento (Eixo 2) e baixa (Eixo 3), contendo posição no ranking, identificador do processo, nível de urgência (CRÍTICO / ALTO / MÉDIO / NORMAL), grupo temporal, idade em dias, previsão do modelo e score de priorização. O arquivo Excel gerado permite análises adicionais e integração com sistemas internos.

### 6.3. Uso Operacional

Os relatórios podem ser gerados sob demanda para análise pontual, em lote para todos os órgãos julgadores, ou periodicamente integrados a rotinas de atualização do ranking.

## 7. Considerações Finais

Este trabalho apresentou o OPTJUS, um sistema para priorização de processos judiciais desenvolvido com dados públicos do DataJud[14]. O pipeline processou 1,66 milhão de registros e gerou rankings para mais de 122 mil processos pendentes na segunda instância do TJRJ.

A escolha entre os algoritmos testados — Random Forest, XGBoost e Elastic Net — foi feita automaticamente com base no menor erro absoluto médio, que variou de 32 a 199 dias dependendo da antiguidade dos processos. Os resultados refletem tanto as características dos dados disponíveis quanto as limitações conhecidas da base do DataJud (cf. Seção 5.1).

O sistema foi desenvolvido como protótipo funcional, com código e documentação disponíveis online[15]. A ferramenta não substitui a avaliação humana, mas oferece forma estruturada de visualizar o acervo pendente e pode ser aperfeiçoada com novas fontes de dados e técnicas de modelagem.

## Agradecimentos

À minha esposa Carolina, pelo apoio constante. Ao Prof. Dr. Cassius Tadeu Scarpin pela orientação dedicada. Ao corpo docente do MBA em *Advanced Analytics & Business Optimization* da UFPR.

[14] Base Nacional de Dados do Poder Judiciário: `https://www.cnj.jus.br/sistemas/datajud/`

[15] Monografia OPTJUS: `https://david.dev.br/OPTJUS/MBA-AABO-UFPR`

## Referências

[1] Conselho Nacional de Justiça. Justiça em números 2024. Technical report, CNJ, Brasília, 2024. URL https://www.cnj.jus.br/wp-content/uploads/2024/05/justica-em-numeros-2024-v-28-05-2024.pdf.

[2] Conselho Nacional de Justiça. Justiça em números 2023. Technical report, CNJ, Brasília, 2023. URL https://www.cnj.jus.br/wp-content/uploads/2023/08/justica-em-numeros-2023.pdf.

[3] Marcelo Guedes Nunes. *Jurimetria: Estatística, Tecnologia e Direito*. Thomson Reuters Brasil, São Paulo, 3 edition, 2024.

[4] Patrícia Belfiore and Luiz Paulo Fávero. *Pesquisa Operacional: Para Cursos de Engenharia*. Elsevier, Rio de Janeiro, 2013.

[5] Conselho Nacional de Justiça. Glossário e esclarecimentos: Metas nacionais do poder judiciário 2025 — justiça estadual. Technical report, Secretaria de Estratégia e Projetos, Departamento de Gestão Estratégica, Brasília, jun 2025. URL https://www.cnj.jus.br/wp-content/uploads/2025/08/glossario-metas-nacionais-do-poder-judiciario-2025-justica-estadual-versao-4.pdf. Versão 4.

[6] Conselho Nacional de Justiça. Resolução cnj nº 331/2020 - institui a base nacional de dados do poder judiciário - datajud, ago 2020. URL https://atos.cnj.jus.br/atos/detalhar/3428.