

# Previsão de ajuizamento de Inscrição da Dívida Ativa da União de Devedores PJ: um estudo usando Random Forest

1<sup>st</sup> Thiago Alves Flor  
PPCA - UNB  
Brasília, Brasil  
thiagoflor83@gmail.com

2<sup>nd</sup> Nicolas Rodrigues de Oliveira  
PPCA - UNB  
Brasília, Brasil  
nk.oliveira@gmail.com

**Abstract**—A Dívida Ativa da União (DAU) representa os débitos não quitados por pessoas físicas ou jurídicas junto ao governo federal, cuja cobrança é realizada pela Procuradoria-Geral da Fazenda Nacional (PGFN). O ajuizamento dessas dívidas, embora legalmente necessário em muitos casos, revela-se uma estratégia onerosa. Um tempo de tramitação de um processo de execução fiscal pode ultrapassar dez anos, com baixíssimo índice de recuperação. Diante desse cenário, este trabalho propõe a aplicação de técnicas de aprendizado de máquina — especificamente, o algoritmo Random Forest — para prever a probabilidade de ajuizamento de dívidas inscritas na DAU por pessoas jurídicas. Utilizando dados das características da dívida e do devedor (como valor, tempo em cobrança, situação cadastral, porte e atividade econômica), referentes aos anos de 2022 e 2023, o modelo foi treinado e posteriormente aplicado sobre os dados de 2024. O objetivo é subsidiar a PGFN com uma ferramenta preditiva que permita priorizar estratégias alternativas de cobrança mais eficazes e econômicas, contribuindo para a racionalização dos recursos públicos e a elevação da eficiência arrecadatória.

**Index Terms**—Aprendizado de Máquina, Dívida Ativa da União, Modelagem Preditiva, Procuradoria-Geral da Fazenda Nacional (PGFN), Ajuizamento, Cobrança Judicial, Cobrança Administrativa, Redução de Litigiosidade, Random Forest.

## I. INTRODUÇÃO

A dívida ativa administrada pelo Governo Federal, em 2024, tinha um valor um pouco acima de 3 trilhões de reais, segundo dados divulgados pela Procuradoria Geral da Fazenda Nacional (PGFN), divididos entre pessoas físicas e jurídicas, sendo que no público PJ concentra-se mais 90% desse valor. Na composição desses valores constam três grandes blocos de origem da dívida, sendo: dívidas oriundas do FGTS (Fundo de Garantia por Tempo de Serviço), da Previdência Social (INSS) e demais tributos federais. Atualmente, é possível recuperar essas informações desde o primeiro trimestre de 2020 a até o final de 2024. Especificamente para esse trabalho, utilizaremos dados do primeiro e quarto trimestres dos anos 2022, 2023 e 2024, tendo em vista o objetivo do trabalho descrito mais adiante.

Sempre que um contribuinte deixa de cumprir uma obrigação legal, formaliza-se a geração de uma inscrição na dívida ativa da União. Resumidamente, esse é o ato no qual

o contribuinte passa a ser um devedor e essa inscrição (débito com a União), que pode ser traduzida como a constituição formal da exigibilidade do débito, entra na esteira de cobrança da PGFN. O valor de 3 trilhões citado acima está dividido em aproximadamente 40 milhões de inscrições. É válido destacar que a inscrição é a informação com menor granularidade na base de dados de cobrança da PGFN e será o objeto desse estudo no que se refere ao ato de ser ou não ajuizada. Depois de constituído o débito, são iniciadas as etapas de cobrança, divididas em dois grandes momentos, sendo: cobrança administrativa e cobrança judicial. A cobrança administrativa é a principal forma de cobrança priorizada pelas estratégias da PGFN e tenta-se, ao máximo, receber o débito dentro das alternativas definidas na esfera administrativa, aplicando esforços para que a cobrança não seja por meio do ajuizamento, por conta do custo e da morosidade da cobrança, quando seguido esse caminho. Um estudo do IPEA (Instituto de Pesquisa Econômica Aplicada), divulgado no comunicado IPEA - 2011 - Dezembro - nº 127, com o título: Custo e tempo do processo de execução fiscal promovido pela Procuradoria Geral da Fazenda Nacional (PGFN) [1], mostrou que o custo unitário médio de uma ação fiscal promovida pela PGFN é de R\$5.606,67 e um tempo médio total de tramitação de 9 anos, 9 meses e 16 dias, com uma probabilidade de recuperação integral de 25,8%.

No ano de 2024, de um total de 34 milhões de inscrições que iniciaram o ano - dados extraídos da divulgação do primeiro de 2024 - como 'não ajuizadas', ou seja, na esteira de cobrança administrativa, 805 mil inscrições terminaram o ano como 'ajuizadas', (segundo os dados divulgados no último trimestre de 2024). Levando em consideração que o estudo do IPEA estimou o custo de ajuizamento em R\$5,6 mil (dado de 2011, esse valor deve ser ainda maior, hoje), podemos estimar que 805 mil inscrições tiveram um custo total de R\$4,5 bilhões para a instituição no ano de 2024. Levando em conta que a arrecadação da PGFN foi R\$61 bilhões em 2024, acreditamos que seja bastante relevante a construção e aplicação de um modelo preditivo para prever quais inscrições têm maior probabilidade de entrarem na esteira da cobrança judicial.

Assim, o objetivo desse estudo consiste na construção de um modelo preditivo, onde, primeiramente, será configurada a leitura e gravação paralelizada de vários arquivos contendo informações de mais 100 milhões de inscrições dos anos de 2022, 2023 e 2024. Posteriormente, filtrar as inscrições que fazem parte do objetivo desse estudo e, usando um algoritmo de Machine Learning, mais especificamente Random Forest, construir um modelo que possa prever quais inscrições têm maior probabilidade de entrar em ajuizamento do público PJ, de acordo com as características da dívida e do devedor, sendo: valor da dívida, situação atual da dívida, tempo em cobrança, situação cadastral do devedor, porte do devedor PJ, atividade econômica, dados geográficos, entre outras informações.

## II. REVISÃO DA LITERATURA

### A. Dívida Ativa da União (DAU)

A Dívida Ativa da União (DAU) compreende o conjunto de débitos de pessoas físicas ou jurídicas com o governo federal que não foram quitados dentro do prazo legal. Quando um tributo, contribuição, multa ou financiamento federal não é pago, a dívida é inscrita na DAU e passa a ser gerida e cobrada pela Procuradoria-Geral da Fazenda Nacional (PGFN). Essa cobrança visa recuperar valores devidos ao Estado e contribuir para o equilíbrio fiscal do país.

Entre os exemplos mais comuns de débitos inscritos na DAU estão o Imposto de Renda não recolhido, contribuições previdenciárias em atraso, multas aplicadas por órgãos federais (como Ibama, ANTT e ANVISA), débitos com o FGTS (para empresas) e financiamentos estudantis ou rurais não quitados, como o FIES e o Pronaf.

A gestão e recuperação desses créditos públicos é tema recorrente na literatura e em estudos institucionais [2], dado seu impacto na arrecadação e na eficiência do Estado.

Segundo o relatório de prestação de contas da Presidência da República referente ao exercício de 2022, publicado em 2023 [3], houve uma redução de aproximadamente 18

TABLE I  
VALORES TOTAIS POR SITUAÇÃO DE AJUIZAMENTO E EXERCÍCIO

Exercício	Não Ajuizado	Ajuizado
2021	478.080.196,45	2.190.226.480,21
2022	423.141.909,40	1.742.020.705,62
<b>Total</b>	<b>901.222.105,85</b>	<b>3.932.247.185,83</b>

<sup>a</sup> Valores em R\$ (em reais) - milhares

A Tabela I, com dados do relatório supracitado, apresenta a comparação dos valores totais de créditos ajuizados e não ajuizados por exercício. Observa-se uma redução de aproximadamente 20,46

### B. Ajuizamento de Dívida Ativa

Após a inscrição do débito na DAU, a PGFN pode optar por ajuizar ou não a dívida. O ajuizamento corresponde à abertura de uma ação judicial de execução fiscal com o objetivo de recuperar o crédito público. Contudo, esse processo é oneroso e pode ser bastante demorado, sobretudo diante dos custos operacionais e da morosidade do Judiciário brasileiro [?].

1) *Custos Operacionais*: Dado o elevado custo envolvido, o ajuizamento de dívida ativa é uma decisão estratégica, baseada na análise de fatores como o valor da dívida, sua recuperabilidade e o custo-benefício da judicialização.

Os custos operacionais englobam diversas etapas, como gastos com pessoal (servidores, juízes e procuradores), manutenção da infraestrutura do Judiciário, sistemas de informação, e recursos administrativos. Esses custos impactam negativamente a eficiência do sistema de cobrança da dívida ativa.

A situação é agravada pelo tempo médio de tramitação das ações, que pode ultrapassar dez anos, retardando a recuperação efetiva dos valores devidos e tornando o processo ainda mais oneroso [?].

Como resposta a esse cenário, a PGFN tem adotado critérios objetivos para definir quais casos devem ser judicializados, priorizando dívidas de maior valor e com maior probabilidade de recuperação. Paralelamente, tem-se incentivado o uso de mecanismos extrajudiciais, como o protesto de certidões de dívida ativa, por serem mais ágeis e menos custosos [?].

De acordo com [?], os principais critérios adotados atualmente pela PGFN para o ajuizamento de dívidas ativas incluem:

- Valor da dívida superior a R\$ 10.000,00 (sujeito a atualizações);
- Expectativa razoável de recuperação do crédito;
- Existência de garantias reais ou de bens com liquidez.

Nos demais casos, a cobrança ocorre por meio administrativo, utilizando ferramentas como a plataforma REGULARIZE, protesto em cartório, negativação em cadastros de inadimplência (como o SERASA), bloqueio de ativos via BacenJud/Sisbajud e, quando aplicável, a oferta de transações tributárias para facilitar o pagamento.

2) *Trabalhos Relacionados*: Diversos estudos têm abordado o tema da dívida ativa, combinando metodologias quantitativas e qualitativas. Pesquisas costumam utilizar análise de decisões judiciais, dados de sistemas como o DataJud, informações obtidas via Lei de Acesso à Informação (LAI) junto a órgãos da administração tributária, e entrevistas com especialistas e gestores públicos. A perspectiva da análise econômica do direito tem sido frequentemente empregada para compreender os incentivos e os custos relacionados ao litígio tributário [?].

### *C. Análise preliminar dos Dados*

TODO: colocar informações sobre a análise de dados e como foi utilizado para entender o problema e orientar a construção do modelo. (citar brevemente o CRIPS-DM) (Avaliar o melhor lugar: aqui ou na metodologia)

### *D. Volume dos dados*

TODO: descrever sobre o volume dos dados a importância do volume de dados, impactos, como isso afeta o modelo e como resolver (gancho pro paralelismo)

### *E. Paralelismo*

TODO: explicar a importância do paralelismo, etapas com possibilidade de aplicação do paralelismo; avaliar citação de ferramentas

### *F. Técnicas de Machine Learning para Classificação*

TODO: explicar brevemente sobre técnicas utilizadas para classificação e porque a escolha do Random Forest.

1) *Classificação com Random Forest*: TODO: explicar o funcionamento do Random Forest e como ele é utilizado para classificação. Explicar a importância da escolha de hiper-parâmetros e como isso afeta o desempenho do modelo.

### *G. Validação do modelo - Cross Validation*

### III. METODOLOGIA

#### A. Coleta de dados

Os dados utilizados nesse trabalho vêm da divulgação trimestral da PGFN atendendo às melhores práticas de transparência ativa previstas na Lei de Acesso à Informação (Lei nº 12.527/2011), na Política de Dados Abertos (Decreto nº 8.777/2016) e no Acórdão nº 2497/2018 - TCU - Plenário (páginas 47-49). Destaca-se que os débitos inscritos em dívida ativa não estão cobertos por sigilo, conforme disposto no art. 198, §3º, II, do Código Tributário Nacional. Destaca-se que foram/serão utilizados os dados divulgados no primeiro e quarto trimestres de 2022, 2023 e 2024, com o objetivo de identificar quais inscrições não são ajuizadas no início do período e se tornam ajuizadas no final do período. Os dados de 2022 e 2023 serão utilizados para fazer a modelagem, onde a nossa target será o campo indicador\_ajuizado (que identifica se a inscrição está ou não em ajuizamento), tal que "SIM" indica ajuizamento e "NÃO" indica não ajuizamento. Após a obtenção de um modelo que apresente uma boa performance, utilizaremos o ano de 2024 para fazer uma simulação dos resultados e responder quanto poderia ter sido economizado com a aplicação desse modelo. O arquivo possui os seguintes campos, conforme descrito na tabela 1 - Dicionário de Campos - abaixo:

#### B. Preparação e pré-processamento de dados

Para os seis períodos citados acima no desenvolvimento do trabalho, serão processados cerca de 162 arquivos csv, totalizando cerca de 27 GB de dados, contendo mais de 100 milhões inscrições da dívida ativa da União. Dados o grande volume de dados e grande quantidade de arquivos, se faz necessária a utilização do paralelismo.

Poderá ser utilizada a biblioteca dask e/ou pyspark. Ainda em análise qual a melhor alternativa para os nossos objetivos.

#### C. Análise Exploratória de Dados

A análise descritiva, dentro do processo de mineração de dados e modelagem preditiva, é fundamental, pois ajuda a identificar padrões, valores de tendência central, medidas de variabilidade, compreender a distribuição dos dados e, também, identificar valores discrepantes, além da existência de valores ausentes e o que fazer para tratá-los.

Acrescenta-se aqui que neste trabalho é fundamental uma análise bivariada, sendo possível assim identificar o impacto das variáveis preditoras na variável target.

#### D. Transformação e seleção de variáveis

Aqui nessa etapa serão construídas as combinações de dados dando origem a novas informações e insights que possam enriquecer a análise e maximizar o alcance do objetivo proposto.

Importante destacar, também, que nessa etapa do trabalho será avaliada a necessidade da normalização ou padronização dos dados com o objetivo de evitar o peso excessivo de uma informação preditora em detrimento das demais.

TABLE II  
DICIONÁRIO DE DADOS

Variável	Descrição
Nome do arquivo	após o termo "arquivo_lai", há indicação do sistema de extração do dado, seguida da UF da unidade da PGFN responsável pela cobrança do devedor e o ano/mês em que o dado foi extraído
CPF_CNPJ	Número identificador do contribuinte no cadastro de pessoas físicas ou no cadastro nacional de pessoas jurídicas
DATA_INSCRICAO	data em que o crédito foi inscrito em dívida ativa
ENTIDADE_RESPONSAVEL	indica se o débito de FGTS está sendo cobrado pela PGFN ou pela Caixa Econômica Federal
INDICADOR_AJUIZADO	indica se o crédito está sendo cobrado judicialmente
NOME_DEVEDOR	nome do devedor
NUMERO_INSCRICAO	número da inscrição em dívida ativa
RECEITA_PRINCIPAL	receita do crédito que está sendo cobrado
SITUACAO_INSCRICAO	situação da inscrição no sistema de controle de créditos
TIPO_DEVEDOR	indica se o devedor é principal (titular original da dívida) ou corresponsável (foi vinculado posteriormente à dívida)
TIPO_PESSOA	indica se é uma pessoa física ou jurídica
TIPO_SITUACAO_INSCRICAO	indica se a inscrição está em cobrança (situação irregular), em benefício fiscal (em parcelamento ou moratória), em negociação, suspenso por decisão judicial, garantia (integralmente garantida)
UF_UNIDADE_RESPONSAVEL	unidade federativa da unidade da PGFN responsável pela cobrança do devedor
UNIDADE_INSCRICAO	indica a unidade da PGFN que realizou a inscrição em dívida ativa
UNIDADE_RESPONSAVEL	unidade da PGFN responsável pelo acompanhamento do devedor
VALOR_CONSOLIDADO	valor do débito na data de extração, com acréscimos legais

#### E. Modelagem

Dentre os algoritmos de Machine Learning disponíveis, inicialmente, elegemos o Random Forest para construir o nosso modelo. A escolha se deve pelo sucesso dessa técnica quando se trata de modelos de classificação, especialmente para target binária, como o nosso trabalho. Também está no escopo da etapa de modelagem comparar duas abordagens diferentes de modelagem e observar qual delas apresenta melhor performance para os nossos dados, sendo:

1) criar um modelo único, unindo todas as informações da dívida;

2) criar três modelos diferentes, sendo um para cada tipo de dívida (FGTS, Previdenciária e Demais tributos);

Pode ser extremamente interessante, também, além das duas abordagens descritas acima, a tentativa de desenvolver modelos usando outros algoritmos de Machine Learning, a depender da variável tempo para finalização deste trabalho.

Não sendo possível, deixaremos como sugestões para trabalhos futuros.

#### REFERENCES

- [1] A. d. S. Cunha, I. d. V. Klin, O. A. Gomes, and I. de Pesquisa Econômica Aplicada (Ipea), “Custo e tempo do processo de execução fiscal promovido pela procuradoria geral da fazenda nacional (pgfn),” Instituto de Pesquisa Econômica Aplicada (Ipea), Comunicado do Ipea 127, 2012, [Accessed: 03/05/2025]. [Online]. Available: <http://repositorio.ipea.gov.br/handle/11058/887>
- [2] IPEA, “Custo Unitário do Processo de Execução Fiscal na Justiça Federal,” Instituto de Pesquisa Econômica Aplicada (Ipea), Tech. Rep., 2011, [Accessed: 03/05/2025]. [Online]. Available: <https://shorturl.at/j55NC>
- [3] Presidência da República, “Mensagem (CN) nº 1, de 2023: Encaminha prestação de contas do Governo Federal,” 2023, [Accessed: 03/04/2025]. [Online]. Available: <https://www.congressonacional.leg.br/materias/pesquisa/-/materia/156683>