



GOVERNO DO DISTRITO FEDERAL

SECRETARIA DE ESTADO DE ECONOMIA DO DISTRITO FEDERAL

SECRETARIA-EXECUTIVA DE FAZENDA

SUBSECRETARIA DA RECEITA

COORDENAÇÃO DE COBRANÇA TRIBUTÁRIA

METODOLOGIA DO RATING

Brasília/DF, setembro de 2025

Sumário

1	Definição do Rating	3
2	<i>Machine Learning</i> – Algoritmo <i>Gradient Boosting</i>	4
2.1	Justificativa para o uso do <i>Gradient Boosting</i>	4
3	Variáveis do Rating	5
3.1	Variáveis do ICMS	6
3.2	Variáveis do ISS	7
3.3	Variáveis do IPVA	9
3.4	Variáveis do IPTU	10
3.5	Variáveis ITBI, ITCD e Taxas	11
4	Interpretação dos Resultados	12
4.1	ICMS	13
4.2	ISS	14
4.3	IPVA	15
4.4	IPTU	16
4.5	ITBI, ITCD e Taxas	17
5	Revisão individual do Rating	17
	Referências.....	18

1 Definição do Rating

O Rating da Dívida Ativa (DA) é uma iniciativa da Secretaria de Economia do Distrito Federal voltada à gestão estratégica da dívida pública. Instituído pela Lei Complementar nº 1.026/2023 e regulamentado pelo Decreto nº 47.090/2025, o sistema classifica os créditos inscritos em dívida ativa conforme sua expectativa de recuperação, promovendo maior eficiência na cobrança e no uso de recursos públicos.

O sistema organiza os créditos em quatro classes, com base na probabilidade calculada de regularização nos próximos cinco anos, entendida como o encerramento do débito por meio de pagamento integral ou parcelamento mantido ativo. Conforme o Decreto nº 47.090/2025, as classes são:

- Classe A (alta, probabilidade $\geq 35\%$);
- Classe B (média, 10% a 35%);
- Classe C (baixa, 1% a 10%);
- Classe D (baixíssima ou irrecuperável, $< 1\%$).

Ainda conforme o Decreto nº 47.090/2025, a classificação decorre de uma análise tridimensional do débito, avaliando:

- Dados da dívida;
- Dados do devedor;
- Dados do objeto do crédito.

Essa análise é operacionalizada por um algoritmo de *machine learning* (aprendizado de máquina), treinado com dados históricos da Secretaria, que calcula a probabilidade de regularização de cada dívida. Com isso, a Secretaria busca apoiar decisões mais eficazes, baseadas em evidências e alinhadas à gestão responsável da dívida pública. A **Figura 1** ilustra o processo de cálculo do Rating.

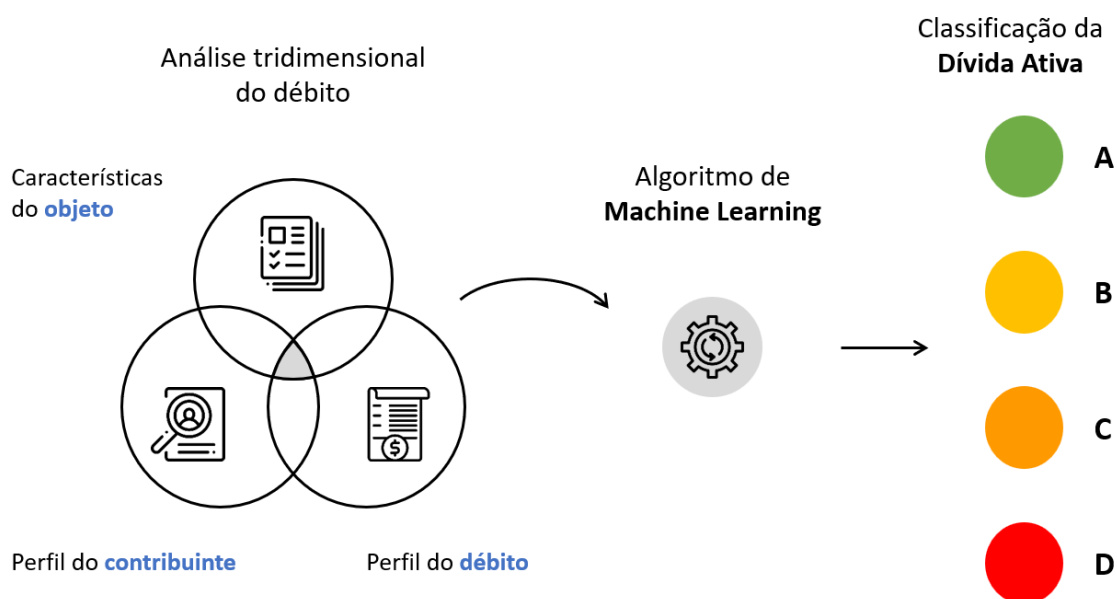


Figura 1: processo de cálculo do Rating.

2 *Machine Learning* – Algoritmo *Gradient Boosting*

O modelo escolhido para o problema em questão é denominado *Gradient Boosting Decision Trees* (GBDT), um modelo de aprendizagem supervisionada baseado em árvores de decisão e comumente utilizado para tarefas de classificação e regressão.

O modelo é baseado em um processo de construção sequencial de preditores parciais (árvores de decisão) que são combinados para formar um modelo robusto e resistente ao problema do *overfitting*, que ocorre quando o modelo é incapaz de generalizar. Cada árvore da sequência é construída a partir de variáveis preditoras buscando reduzir o residual antecedente. O resultado final é uma transformação logarítmica da soma ponderada das árvores de decisão, onde cada árvore contribui com uma pequena parte das previsões, gerando a probabilidade calculada.

A implementação GBDT selecionada foi a biblioteca Python XGBoost [1], uma das mais utilizadas em algoritmos GBDT, bem como algoritmos de classificação e regressão no geral, e que inclui técnicas adicionais como regularização, *sampling*, *pruning*, entre outras.

2.1 Justificativa para o uso do *Gradient Boosting*

A escolha do algoritmo *Gradient Boosting* para o cálculo da probabilidade de recuperação das dívidas em DA do DF foi feita com base em 3 critérios: a) embasamento técnico e teórico; b) poder preditivo apresentado pelo algoritmo; e c) interpretabilidade.

Para problemas complexos e não-lineares, como é o caso de muitos desafios que envolvem o comportamento humano, são necessários algoritmos capazes de modelar esses padrões. A combinação de múltiplos preditores parciais no GBDT, somada às técnicas de regularização e *sampling* da implementação XGBoost, torna a solução adotada particularmente eficaz para o problema em questão. Soma-se a isso o fato que os dados disponíveis são em grande parte multicategóricos, dados do tipo que algoritmos baseados em árvore de decisão lidam especialmente bem.

No campo teórico-prático, acadêmicos vêm registrando o poder preditivo do GBDT aplicado a problemas de escore de crédito, muito semelhante ao Rating da DA. Tian et al. [2] concluiu que o GBDT tem “a melhor habilidade de classificação e generalização” para o problema estudado, de risco de crédito. De forma semelhante, para Zou et al. [3], o XGBoost foi o 1º colocado dentre as implementações já estabelecidas em todas 4 bases de dados estudadas.

Na prática, um dos critérios mais importantes é a capacidade preditiva comprovada. Por esse motivo, múltiplas técnicas de *machine learning* foram testadas e comparadas. A métrica escolhida para a comparação é métrica AUC, ou área sob a curva ROC (*Receiver Operating Characteristic*), uma medida de desempenho utilizada em modelos de classificação binária – no caso, se haverá ou não regularização. Ela avalia a capacidade do modelo de discriminar entre duas classes, considerando a taxa de verdadeiros positivos e a taxa de falsos positivos em diferentes limiares de decisão. Essa métrica é especialmente relevante porque valida a capacidade discriminativa do modelo, cujo resultado probabilístico é utilizado para definir as classificações A, B, C ou D. Nos múltiplos testes realizados, o GBDT se provou o com maior performance.

Por fim, a interpretabilidade do algoritmo também foi levada em consideração. A interpretabilidade refere-se à capacidade de compreendermos, enquanto auditores, gestores e contribuintes, os resultados gerados por um modelo de aprendizado de máquina. Modelos mais

complexos, embora capazes de capturar padrões sofisticados nos dados e, assim, potencialmente produzir previsões mais precisas, tendem a ser menos transparentes em sua lógica de funcionamento. O modelo escolhido, por ser um baseado em árvores de decisão, oferece um bom equilíbrio entre complexidade e interpretabilidade, permitindo tanto resultados robustos quanto análises compreensíveis dos fatores que influenciam o cálculo. A **Figura 2** ilustra essa relação.

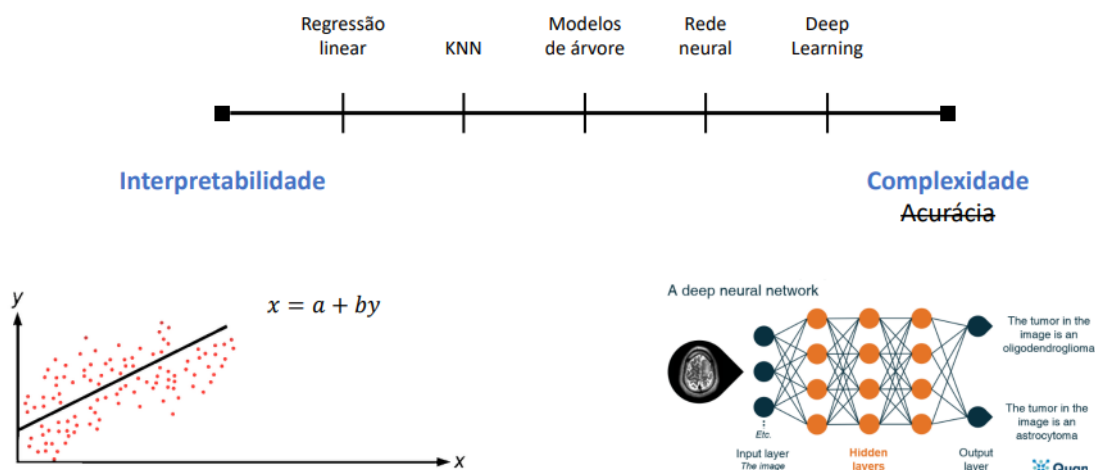


Figura 2: interpretabilidade e complexidade em modelos de *machine learning*.

Cabe destacar, como sugere a figura, que complexidade não necessariamente é sinônima de acurácia. A eficácia de um algoritmo depende de quão bem sua lógica interna se ajusta à estrutura dos dados e dos padrões que se busca modelar. Em problemas de relação perfeitamente linear, por exemplo, uma regressão linear terá acurácia máxima. O GBDT, como já mencionado, se encaixa bem no tipo dos dados disponíveis, e apresentou o melhor desempenho, seguindo a conclusão da literatura.~

O algoritmo adotado é passível de alteração futura com base nos mesmos três critérios considerados.

Com o intuito de ampliar a interpretabilidade dos resultados do modelo por parte do contribuinte, a seção a seguir detalha as variáveis incluídas no modelo e fornece algumas ferramentas para a interpretação.

3 Variáveis do Rating

A seleção de variáveis para o Rating consistiu na identificação daqueles fatores que, com base nos dados históricos disponíveis à Secretaria, mais contribuíram para distinguir entre os contribuintes e os débitos que tendem à regularização e aqueles que não o fazem. Essa etapa é fundamental para garantir que o modelo trabalhe com informações relevantes e diretamente ligadas ao comportamento buscado – a regularização da dívida.

Durante esse processo, a Secretaria avaliou todo o histórico de regularização da dívida ativa do Distrito Federal, desde o início da série temporal em formato digital, o que representa uma análise de mais de 30 anos de ocorrências ou mudanças de situação dos débitos. Foram avaliadas diversas variáveis relacionadas às dimensões Dívida, Devedor, e Objeto, sendo a

dimensão Objeto reservada a impostos patrimoniais ou outras dívidas com um bem relacionado. A avaliação combinou as informações disponíveis à Secretaria, ferramentas avançadas de análise de dados e o conhecimento técnico acumulado por sua equipe de auditores fiscais.

Da dimensão Dívida, foram avaliados dados a respeito do valor da dívida, idade da dívida, histórico de ocorrências (situações), origem, entre outros. Da dimensão Devedor, foram avaliados dados a respeito do histórico de regularizações e pagamentos do devedor, situação cadastral, grupo econômico, total de dívidas, entre outros. Da dimensão Objeto, foram avaliados dados a respeito da idade do objeto, valor venal, entre outros. Em conjunto, esses dados compõem um quadro geral da dívida e do devedor que, analisados sob a perspectiva do comportamento histórico, permitem à Secretaria fazer uma avaliação da probabilidade de regularização, por meio do modelo de *machine learning*.

O processo de seleção envolve, ainda, a remoção de variáveis redundantes, pouco informativas ou que pioram a capacidade preditiva do modelo. Um fator importante na avaliação de cada variável é sua dependência temporal, isto é, se a variável mantém suas correlações com o comportamento esperado ao longo do tempo.

Ao fim, o objetivo é o desenvolvimento de um modelo robusto, com capacidade de previsão adequada e devidamente testado com métodos robustos de teste em séries temporais. Além da capacidade preditiva, buscou-se garantir que as variáveis selecionadas fossem compreensíveis e úteis para orientar ações de cobrança e política fiscal.

Dada a especificidade de cada tributo, identificado por um conjunto de códigos de receita, o processo de seleção é conduzido de forma individualizada. Isso significa que, para cada tributo, são aplicados os critérios estabelecidos de avaliação, levando em conta apenas as variáveis pertinentes ao seu contexto operacional e normativo. Essa abordagem considera as particularidades de cada tributo, assegurando adequação contextual e acurácia ao modelo.

3.1 Variáveis do ICMS

Para o Imposto sobre Circulação de Mercadorias e Serviços (ICMS), um total de 24 variáveis foram definidas para o cálculo, descritas na **Tabela 1**.

Tabela 1: variáveis do modelo ICMS.

Variável	Dimensão	Unidade	Descrição
Código de receita	Dívida	código	Código de receita da dívida
Idade da dívida em meses	Dívida	meses	Idade da dívida desde a inscrição em dívida ativa
Ano de declaração	Dívida	ano	Ano de referência da declaração do débito inscrito
Valor original da dívida corrigido	Dívida	R\$	Valor original no momento de inscrição em DA, corrigido para desconsiderar valores em moedas anteriores ao Real
Valor relativo da multa	Dívida	razão	Razão entre o valor original da multa e o valor original total da dívida no momento de inscrição em DA
Situação da dívida	Dívida	código	Código de situação atual da dívida
Quantidade de situações	Dívida	quantidade	Quantidade de mudanças de situação que a dívida passou

Variável	Dimensão	Unidade	Descrição
Quantidade de parcelamentos cancelados	Dívida	quantidade	Quantidade de parcelamentos cancelados da dívida
Tempo inativo	Dívida	anos	Tempo desde a última mudança de situação da dívida
Se ativo	Devedor	sim/não	Se a empresa está ativa ou não (incluindo inscrição condicional e em fase de suspensão como ativo)
Quantidade de ativos do núcleo	Devedor	quantidade	Quantidade de empresas ativas do núcleo CNPJ
Regime tributário ICMS	Devedor	código	Código do regime tributário do ICMS
Regime tributário ISS	Devedor	código	Código do regime tributário do ISS
Idade da empresa no momento de inscrição	Devedor	categoria	Idade da empresa desde a 1ª situação cadastral ou regime tributário registrado, no momento de inscrição da DA (0 p/ menos de 1 anos, 1 p/ menos de 2 anos, 2 p/ menos de 4 anos)
Quantidade de imóveis	Devedor	quantidade	Quantidade de imóveis em posse do devedor (considerando apenas imóveis do DF), limitado a um valor máximo de 3
Valor de Dívidas Abertas do CNPJ	Devedor	R\$	Valor total das dívidas em aberto do CNPJ
Quantidade de DAs em aberto	Devedor	quantidade	Quantidade de dívidas inscritas em DA em aberto do núcleo CNPJ
Histórico de regularização do CNPJ (5a)	Devedor	razão	Histórico de regularização de dívidas do CNPJ, últimos 5 anos
Histórico de regularização do núcleo (1a)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, último 1 ano
Histórico de regularização do núcleo (5a)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, últimos 5 anos
Histórico de pagamento do núcleo	Devedor	razão	Histórico de pagamento em cheio de dívidas do núcleo CNPJ
Histórico de regularização do núcleo	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, desconsiderando dívidas canceladas e suspensas
Código de receita codificado	Dívida	razão	Código de receita codificado*
Código de origem codificado	Dívida	razão	Código de origem codificado*

* Variáveis codificadas são variáveis categóricas (como o código de receita) transformadas em valores numéricos em que o valor representa a taxa de regularização histórica da categoria (ex.: se a categoria X possui um histórico de pagamento de 20% o valor 0.20 será utilizado como codificação).

3.2 Variáveis do ISS

Para o Imposto sobre Serviços de Qualquer Natureza (ISS), um total de 21 variáveis foram definidas para o cálculo, descritas na **Tabela 2**.

Tabela 2: variáveis do modelo ISS.

Variável	Dimensão	Unidade	Descrição
Código de receita	Dívida	código	Código de receita da dívida
Idade da dívida	Dívida	meses	Idade da dívida desde a inscrição em DA
Valor original da dívida corrigido	Dívida	R\$	Valor original no momento de inscrição em DA, corrigido para desconsiderar valores em moedas anteriores ao Real
Valor relativo da multa	Dívida	razão	Razão entre o valor original da multa e o valor original total da dívida no momento de inscrição em DA
Quantidade de situações	Dívida	quantidade	Quantidade de mudanças de situação que a dívida passou
Quantidade de parcelamentos cancelados	Dívida	quantidade	Quantidade de vezes que a dívida teve um parcelamento cancelado
Tempo de inatividade	Dívida	anos	Tempo desde a última mudança de situação da dívida
Se ativo	Devedor	sim / não	Se a empresa está ativa ou não (incluindo inscrição condicional e em fase de suspensão como ativo)
Quantidade de ativos do núcleo	Devedor	quantidade	Quantidade de empresas ativas do núcleo CNPJ
Regime tributário ISS	Devedor	código	Código do regime tributário do ISS
Idade da empresa	Devedor	semestres	Idade da empresa desde a 1a situação cadastral ou regime tributário registrado
Quantidade de imóveis	Devedor	quantidade	Quantidade de imóveis em posse do devedor (considerando apenas imóveis do DF)
Valor de dívida em aberto do CNPJ	Devedor	R\$	Valor total das dívidas em aberto do CNPJ, pelo valor original
Histórico de regularização do CNPJ (5a)	Devedor	razão	Histórico de regularização de dívidas do CNPJ, últimos 5 anos
Histórico de regularização do núcleo (1a)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, último 1 ano
Histórico de regularização do núcleo (5a)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, últimos 5 anos
Histórico de pagamento do núcleo	Devedor	razão	Histórico de pagamento em cheio de dívidas do núcleo CNPJ
Histórico de regularização do núcleo	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ, desconsiderando dívidas canceladas e suspensas
Regime tributário ISS codificado	Devedor	razão	Regime tributário ISS codificado*
Código de origem codificado	Dívida	razão	Código de origem codificado*

Variável	Dimensão	Unidade	Descrição
Situação inicial codificado	Dívida	razão	Situação inicial codificado*

* Variáveis codificadas são variáveis categóricas (como o regime tributário) transformadas em valores numéricos em que o valor representa a taxa de regularização histórica da categoria (ex.: se a categoria X possui um histórico de pagamento de 20% o valor 0.20 será utilizado como codificação).

3.3 Variáveis do IPVA

Para o Imposto sobre a Propriedade de Veículos Automotores (IPVA), foram definidas 25 variáveis para o cálculo, descritas na **Tabela 3**.

Tabela 3: variáveis do modelo IPVA.

Variável	Dimensão	Unidade	Descrição
Idade da dívida	Dívida	meses	Idade da dívida desde a inscrição em dívida ativa
Valor original da dívida corrigido	Dívida	R\$	Valor original no momento de inscrição em DA, corrigido para desconsiderar valores em moedas anteriores ao Real
Valor relativo da multa	Dívida	razão	Razão entre o valor original da multa e o valor original total da dívida no momento de inscrição em DA
Situação da dívida	Dívida	código	Código de situação atual da dívida
Quantidade de situações	Dívida	quantidade	Quantidade de mudanças de situação que a dívida passou
Quantidade de parcelamentos cancelados	Dívida	quantidade	Quantidade de vezes que a dívida teve um parcelamento cancelado
Tempo inativo	Dívida	anos	Tempo desde a última mudança de situação da dívida
Se PJ e se ativo	Devedor	categoria	Se PF (=0), se PJ e está ativa (incluindo inscrição condicional e em fase de suspensão como ativo, =1) ou não (=2)
Quantidade de veículos	Devedor	quantidade	Quantidade de veículos (do DF) registrados em nome do devedor
Quantidade de imóveis	Devedor	quantidade	Quantidade de imóveis em posse do devedor (considerando apenas imóveis do DF), limitado a um valor máximo de 5
Idade da pessoa	Devedor	anos	Idade do devedor pessoa física, vazio caso pessoa jurídica
Anos desde o óbito	Devedor	anos	Tempo desde o óbito do devedor (0 se não aplicável, anos + 1 se houve óbito)
Indicador de estrangeiro	Devedor	sim/não	Indica se o devedor é estrangeiro
Quantidade de DAs em aberto	Devedor	quantidade	Quantidade de dívidas inscritas em DA em aberto do núcleo CNPJ
Quantidade de DAs do veículo	Objeto	quantidade	Quantidade de dívidas inscritas em DA em aberto referentes ao veículo
Histórico de pagamento do núcleo	Devedor	razão	Histórico de pagamento integral de dívidas do núcleo CNPJ (ou CPF)

Variável	Dimensão	Unidade	Descrição
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	razão	Histórico de regularização de dívidas do CPF/CNPJ nos últimos 5 anos
Valor de aquisição do veículo	Objeto	R\$	Valor de aquisição do veículo, corrigido para desconsiderar valores em moedas anteriores ao Real
Código do tipo de veículo	Objeto	código	Código do tipo do veículo
Indicador de subcategoria do veículo	Objeto	sim/não	Indica se existe código da subcategoria do veículo
Código do combustível	Objeto	código	Código do tipo de combustível do veículo
Idade da última atualização	Objeto	dias	Tempo desde a última atualização de dados do veículo
Idade de fabricação do veículo	Objeto	anos	Idade do veículo desde a fabricação
Indicador de revendedor de veículos	Objeto	sim/não	Indica se o proprietário é CNPJ revendedor de veículos
Quantidade de transferências de veículo	Objeto	quantidade	Quantidade de transferências de propriedade do veículo

3.4 Variáveis do IPTU

Para o Imposto sobre a Propriedade Predial e Territorial Urbana (IPTU), um total de 22 variáveis foram definidas para o cálculo, descritas na **Tabela 3**.

Tabela 4: variáveis do modelo IPTU.

Variável	Dimensão	Unidade	Descrição
Idade da dívida	Dívida	meses	Idade da dívida desde a inscrição em dívida ativa
Valor original da dívida corrigido	Dívida	R\$	Valor original no momento de inscrição em DA, corrigido para desconsiderar valores em moedas anteriores ao Real
Valor relativo da multa	Dívida	razão	Razão entre o valor original da multa e o valor original total da dívida no momento de inscrição em DA
Situação da dívida	Dívida	código	Código de situação atual da dívida
Código de origem	Dívida	código	Código de origem da dívida
Quantidade de situações	Dívida	quantidade	Quantidade de mudanças de situação que a dívida passou
Tempo inativo	Dívida	anos	Tempo desde a última mudança de situação da dívida
Indicador de pessoa jurídica (CNPJ)	Devedor	sim/não	Indica se o devedor é pessoa jurídica ou não

Variável	Dimensão	Unidade	Descrição
Se ativo	Devedor	sim/não	Se a empresa está ativa ou não (incluindo inscrição condicional e em fase de suspensão como ativo)
Quantidade de veículos	Devedor	quantidade	Quantidade de veículos (do DF) registrados em nome do devedor
Quantidade de imóveis	Devedor	quantidade	Quantidade de imóveis em posse do devedor (considerando apenas imóveis do DF)
Anos desde o óbito	Devedor	anos	Tempo desde o óbito do devedor (0 se não aplicável, anos + 1 se houve óbito)
Quantidade de DAs em aberto	Devedor	quantidade	Quantidade de dívidas inscritas em DA em aberto do núcleo CNPJ
Histórico de pagamento do núcleo	Devedor	razão	Histórico de pagamento integral de dívidas do núcleo CNPJ (ou CPF)
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	razão	Histórico de regularização de dívidas do CPF/CNPJ nos últimos 5 anos
Bairro do imóvel (grupo)	Objeto	categoria	Bairro do endereço do imóvel mais atualizado, categorizado em grupos
Idade do último registro	Objeto	anos	Tempo desde a última atualização do registro do imóvel
Idade do imóvel	Objeto	anos	Idade do imóvel desde a construção
Percentual de alíquota	Objeto	%	Percentual de alíquota aplicada para o cálculo do último IPTU
Valor da construção	Objeto	R\$	Valor venal da construção
Valor do terreno	Objeto	R\$	Valor venal do terreno
Valor do m2	Objeto	R\$/m ²	Valor do metro quadrado construído

3.5 Variáveis ITBI, ITCD e Taxas

Para as dívidas: a) Imposto de Transmissão de Bens Imóveis (ITBI), b) Imposto de Transmissão Causa Mortis e Doação (ITCD), e c) Taxas, Multas e outras dívidas não-tributárias, um total de 18 variáveis foram definidas para o cálculo, descritas na **Tabela 3**.

Tabela 5: variáveis do modelo ITBI, ITCD e Taxas.

Variável	Dimensão	Unidade	Descrição
Idade da dívida (meses)	Dívida	meses	Idade da dívida desde a inscrição em dívida ativa
Código da receita	Dívida	categoria	Código de receita da dívida, categorizado em grupos
Situação da dívida	Dívida	código	Código de situação atual da dívida

Variável	Dimensão	Unidade	Descrição
Se houve ajuizamento	Dívida	sim/não	Indica se, em algum momento, houve ajuizamento da dívida
Quantidade de situações	Dívida	quantidade	Quantidade de mudanças de situação que a dívida passou
Se ativo	Devedor	sim/não	Se a empresa está ativa ou não (incluindo inscrição condicional e em fase de suspensão como ativo)
Quantidade de ativos do núcleo	Devedor	quantidade	Quantidade de empresas ativas do núcleo CNPJ
Regime tributário ICMS	Devedor	código	Código do regime tributário do ICMS
Regime tributário ISS	Devedor	código	Código do regime tributário do ISS
Idade da empresa	Devedor	meses	Idade da empresa desde a 1ª situação cadastral ou regime tributário registrado
Idade da pessoa	Devedor	anos	Idade do devedor pessoa física, vazio caso pessoa jurídica
Quantidade de imóveis	Devedor	quantidade	Quantidade de imóveis em posse do devedor (considerando apenas imóveis do DF)
Quantidade de veículos	Devedor	quantidade	Quantidade de veículos (do DF) registrados em nome do devedor
Quantidade de DAs em aberto	Devedor	quantidade	Quantidade de dívidas inscritas em DA em aberto do núcleo CNPJ
Histórico de pagamento do núcleo	Devedor	razão	Histórico de pagamento integral de dívidas do núcleo CNPJ (ou CPF)
Histórico de regularização do núcleo (últimos 5 anos)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ nos últimos 5 anos
Histórico de regularização do núcleo (último ano)	Devedor	razão	Histórico de regularização de dívidas do núcleo CNPJ no último 1 ano
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	razão	Histórico de regularização de dívidas do CPF/CNPJ nos últimos 5 anos

4 Interpretação dos Resultados

Para auxiliar a interpretação dos resultados do algoritmo, utilizamos duas medidas que visam avaliar a contribuição de cada variável para o modelo: a correlação e a importância.

A correlação é uma medida estatística cuja força varia entre 0 e 1 e indica o grau de associação entre uma variável e o resultado do modelo – neste caso, a probabilidade de regularização da dívida. Correlações positivas, ou de polaridade positiva, indicam que o aumento do valor da variável está associado a um aumento da probabilidade prevista; correlações de polaridade negativa, por outro lado, indicam que valores mais altos da variável estão relacionados a uma menor probabilidade de regularização. A título de exemplo, o tempo de inatividade da dívida, que nos diz quanto tempo a dívida está sem haja alguma ocorrência,

apresenta correlação negativa: quanto maior o tempo, menor tende a ser a probabilidade de regularização.

É importante destacar que as relações entre o cálculo do modelo e as variáveis não são lineares – em outras palavras, um aumento na variável não necessariamente acarreta um aumento ou diminuição correspondente no cálculo do modelo – mas não-lineares, refletindo o comportamento complexo do processo de regularização. É o caso da variável "idade da empresa" que, olhada isoladamente, tem correlação negativa, refletindo o efeito negativo do passar do tempo. Se combinada com uma empresa em situação ativa e uma dívida recente, entretanto, tem efeito positivo, refletindo a robustez da empresa. As correlações são, portanto, medidas estatísticas, que ajudam a compreender a direção geral de cada variável.

A importância de uma variável, por sua vez, refere-se à contribuição que ela oferece ao processo de decisão do modelo, considerando as interações com outras variáveis. Mesmo variáveis com correlação isolada baixa podem apresentar elevada importância se, em combinação com outras informações, ajudam o modelo a diferenciar melhor os casos. Pode ser entendida como o peso médio da variável no cálculo do modelo.

Ao analisar os resultados do modelo, devemos avaliar, em especial, as variáveis de alta correlação e as de alta importância.

4.1 ICMS

Tabela 6: polaridade, correlação e importância das variáveis do modelo ICMS.

Variável	Dimensão	Polaridade	Correlação	Importância calculada
Histórico de regularização do núcleo	Devedor	Positiva	0,402	0,040
Histórico de regularização do núcleo (5a)	Devedor	Positiva	0,384	0,205
Histórico de regularização do CNPJ (5a)	Devedor	Positiva	0,371	0,020
Tempo inativo	Dívida	Negativa	0,335	0,143
Histórico de pagamento do núcleo	Devedor	Positiva	0,304	0,019
Se ativo	Devedor	Positiva	0,289	0,053
Histórico de regularização do núcleo (1a)	Devedor	Positiva	0,238	0,021
Código de origem codificado	Dívida	Positiva	0,179	0,016
Valor de Dívidas Abertas do CNPJ	Devedor	Negativa	0,177	0,053
Idade da dívida em meses	Dívida	Negativa	0,175	0,032
Quantidade de ativos do núcleo	Devedor	Positiva	0,169	0,024
Ano de declaração	Dívida	Positiva	0,164	0,028
Quantidade de parcelamentos cancelados	Dívida	Positiva	0,135	0,022
Quantidade de DAs em aberto	Devedor	Negativa	0,125	0,028

Código de receita codificado	Dívida	Positiva	0,066	0,016
Quantidade de situações iniciais	Dívida	Positiva	0,061	0,023
Valor original da dívida corrigido	Dívida	Negativa	0,035	0,029
Valor relativo da multa	Dívida	Negativa	0,027	0,023
Idade da empresa no momento de inscrição	Devedor	Positiva	0,016	0,023
Quantidade de imóveis	Devedor	Negativa	0,003	0,031
Regime tributário ICMS	Devedor			0,087
Situação da dívida	Dívida			0,027
Regime tributário ISS	Devedor			0,019
Código de receita	Dívida			0,019

4.2 ISS

Tabela 7: polaridade, correlação e importância das variáveis do modelo ISS.

Variável	Dimensão	Polaridade	Correlação	Importância calculada
Histórico de regularização do núcleo	Devedor	Positiva	0,533	0,123
Situação inicial codificado	Dívida	Positiva	0,498	0,025
Histórico de pagamento do núcleo	Devedor	Positiva	0,465	0,03
Tempo de inatividade	Dívida	Negativa	0,413	0,092
Histórico de regularização do núcleo (5a)	Devedor	Positiva	0,399	0,125
Histórico de regularização do CNPJ (5a)	Devedor	Positiva	0,394	0,021
Código de origem codificado	Dívida	Positiva	0,291	0,019
Se ativo	Devedor	Positiva	0,256	0,04
Histórico de regularização do núcleo (1a)	Devedor	Positiva	0,256	0,034
Quantidade de imóveis	Devedor	Positiva	0,241	0,06
Idade da dívida	Dívida	Negativa	0,204	0,033
Quantidade de parcelamentos cancelados	Dívida	Positiva	0,202	0,035
Quantidade de situações	Dívida	Positiva	0,15	0,119
Regime tributário ISS codificado	Devedor	Positiva	0,115	0,019

Valor relativo da multa	Dívida	Negativa	0,052	0,039
Idade da empresa	Devedor	Negativa	0,034	0,02
Valor de dívida em aberto do CNPJ	Devedor	Negativa	0,026	0,05
Valor original da dívida corrigido	Dívida	Negativa	0,006	0,039
Quantidade de ativos do núcleo	Devedor			0,031
Regime tributário ISS	Devedor			0,025
Código de receita	Dívida			0,021

Obs.: não é possível calcular a correlação para variáveis não numéricas (categóricas).

4.3 IPVA

Tabela 8: polaridade, correlação e importância das variáveis do modelo IPVA.

Variável	Dimensão	Polaridade	Correlação	Importância calculada
Histórico de pagamento do núcleo	Devedor	Positiva	0,459	0,041
Idade da última atualização	Objeto	Negativa	0,467	0,174
Tempo inativo	Dívida	Negativa	0,384	0,041
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	Positiva	0,370	0,041
Idade de fabricação do veículo	Objeto	Negativa	0,366	0,028
Valor da dívida corrigido	Dívida	Positiva	0,317	0,018
Idade da dívida	Dívida	Negativa	0,294	0,026
Quantidade de imóveis	Devedor	Positiva	0,257	0,095
Quantidade de veículos	Devedor	Positiva	0,187	0,032
Quantidade de transferências de veículo	Objeto	Positiva	0,173	0,014
Quantidade de parcelamentos cancelados	Dívida	Positiva	0,146	0,032
Indicador de revendedor de veículos	Objeto	Positiva	0,117	0,027
Anos desde o óbito	Devedor	Negativa	0,109	0,015
Quantidade de situações	Dívida	Positiva	0,084	0,028
Valor da multa	Dívida	Negativa	0,071	0,020
Quantidade de DAs do veículo	Objeto	Positiva	0,043	0,133
Quantidade de DAs em aberto	Devedor	Positiva	0,029	0,045

Se ativo	Devedor	Negativa	0,021	0,041
Idade da pessoa	Devedor	Negativa	0,018	0,011
Valor de aquisição do veículo	Objeto	Negativa	0,010	0,015
Indicador de subcategoria do veículo	Objeto	Positiva	0,002	0,001
Código do combustível	Objeto	—	—	0,044
Código do tipo de veículo	Objeto	—	—	0,026
Indicador de estrangeiro	Devedor	—	—	0,004

4.4 IPTU

Tabela 9: polaridade, correlação e importância das variáveis do modelo IPTU.

Variável	Dimensão	Polaridade	Correlação	Importância calculada
Histórico de pagamento do núcleo	Devedor	Positiva	0,526	0,074
Tempo inativo	Dívida	Negativa	0,429	0,061
Valor do m2 construído	Objeto	Positiva	0,407	0,039
Percentual de alíquota	Dívida	Negativa	0,381	0,033
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	Positiva	0,359	0,016
Quantidade de DAS em aberto	Devedor	Negativa	0,347	0,163
Indicador de pessoa jurídica (CNPJ)	Devedor	Negativa	0,323	0,047
Idade da dívida	Dívida	Negativa	0,264	0,032
Idade do último registro	Objeto	Positiva	0,195	0,059
Quantidade de veículos	Devedor	Negativa	0,186	0,108
Quantidade de imóveis	Devedor	Negativa	0,175	0,030
Valor do terreno	Objeto	Positiva	0,168	0,121
Quantidade de situações	Dívida	Positiva	0,086	0,021
Se ativo	Devedor	Negativa	0,082	0,008
Valor da multa	Dívida	Negativa	0,074	0,054
Valor da construção	Objeto	Positiva	0,022	0,041
Valor da dívida corrigido	Dívida	Negativa	0,010	0,010
Anos desde o óbito	Devedor	Positiva	0,004	0,009
Situação da dívida	Dívida	—		0,030
Código de origem	Dívida	—		0,014
Bairro do imóvel (grupo)	Objeto	—		0,018

4.5 ITBI, ITCD e Taxas

Tabela 10: polaridade, correlação e importância das variáveis do modelo ITBI, ITCD e Taxas.

Variável	Dimensão	Polaridade	Correlação	Importância calculada
Histórico de regularização do núcleo (últimos 5 anos)	Devedor	Positiva	0,344	0,065
Histórico de regularização do CPF/CNPJ (últimos 5 anos)	Devedor	Positiva	0,337	0,086
Idade da empresa	Devedor	Negativa	0,310	0,017
Idade da dívida	Dívida	Negativa	0,265	0,045
Se houve ajuizamento	Dívida	Negativa	0,264	0,016
Histórico de pagamento do núcleo	Devedor	Positiva	0,261	0,024
Quantidade de veículos	Devedor	Negativa	0,214	0,045
Quantidade de DAS em aberto	Devedor	Negativa	0,191	0,038
Histórico de regularização do núcleo (último ano)	Devedor	Positiva	0,128	0,017
Quantidade de situações	Dívida	Negativa	0,080	0,024
Idade da pessoa	Devedor	Negativa	0,079	0,015
Quantidade de imóveis	Devedor	Negativa	0,001	0,384
Se ativo	Devedor	—		0,062
Código de receita	Dívida	—		0,047
Situação da dívida	Dívida	—		0,038
Quantidade de ativos do núcleo	Devedor	—		0,032
Regime tributário ICMS	Devedor	—		0,020
Regime tributário ISS	Devedor	—		0,017

5 Revisão individual do Rating

O contribuinte poderá solicitar através do atendimento virtual os valores dos parâmetros considerados para classificação das CDAs.

Caso o contribuinte discorde dos valores considerado para classificação, deverá solicitar a solicitação no atendimento virtual, apresentando toda a documentação comprobatória que justifique a alteração do valor dessas variáveis.

Após a validação e atualização dos dados cadastrais pela SEEC, o rating dos débitos será recalculado considerando os novos valores das variáveis, com a possibilidade de alterar a classe do rating.

Referências

1. Get Started with XGBoost — Xgboost 3.0.3 Documentation Available online: https://xgboost.readthedocs.io/en/stable/get_started.html (accessed on 1 August 2025).
2. Tian, Z.; Xiao, J.; Feng, H.; Wei, Y. Credit Risk Assessment Based on Gradient Boosting Decision Tree. *Procedia Computer Science* **2020**, *174*, 150–160, doi:10.1016/j.procs.2020.06.070.
3. Zou, Y.; Xia, M.; Lan, X. Interpretable Credit Scoring Based on an Additive Extreme Gradient Boosting. *Chaos, Solitons & Fractals* **2025**, *194*, 116216, doi:10.1016/j.chaos.2025.116216.