

**UNIVERSIDADE ESTADUAL DO MARANHÃO**  
**CENTRO DE CIÊNCIAS TECNOLÓGICAS**  
**MESTRADO EM ENGENHARIA DE COMPUTAÇÃO E SISTEMAS**

**DANIELLE COSTA BRANDÃO**

**CIÊNCIA DE DADOS APLICADA À GESTÃO FISCAL: UM ESTUDO SOBRE A  
INADIMPLÊNCIA DO IMPOSTO PREDIAL E TERRITORIAL URBANO EM SÃO  
LUÍS-MA**

**SÃO LUÍS-MA**

**2026**

**DANIELLE COSTA BRANDÃO**

**CIÊNCIA DE DADOS APLICADA À GESTÃO FISCAL: UM ESTUDO SOBRE A  
INADIMPLÊNCIA DO IMPOSTO PREDIAL E TERRITORIAL URBANO EM SÃO  
LUÍS-MA**

Dissertação apresentada como requisito  
parcial para a obtenção do grau de Mestre em  
Engenharia de Computação e Sistemas.

**Orientador:** Dr. Luis Carlos Costa Fonseca  
**Coorientador:** Dr. Sofiane Labidi

**SÃO LUÍS-MA**

**2026**

Brandão, Danielle Costa.

Ciência de dados aplicada à gestão fiscal: um estudo sobre a inadimplência do Imposto Predial e Territorial Urbano em São Luís-MA./ Danielle Costa Brandão. - São Luís - MA, 2026.

104f.

Dissertação (Mestrado em Engenharia de Computação e Sistemas) - Universidade Estadual do Maranhão, 2026.

Orientador: Prof. Dr. Luis Carlos Costa Fonseca.

Coorientador: Prof. Dr. Sofiane Labidi.

1. IPTU. 2. Inadimplência. 3. Ciência de dados. 4. Aprendizado de máquinas. 5. Arrecadação municipal. I. Título.

CDU: 351.71(812.1)

**Elaborado por Luciana de Araújo - CRB 13/445**

**DANIELLE COSTA BRANDÃO**

**CIÊNCIA DE DADOS APLICADA À GESTÃO FISCAL: UM ESTUDO SOBRE A INADIMPLÊNCIA DO IMPOSTO PREDIAL E TERRITORIAL URBANO EM SÃO LUÍS-MA**

Dissertação apresentada junto ao curso de Mestrado em Engenharia de Computação e Sistemas da Universidade Estadual do Maranhão (UEMA) para obtenção de grau de Mestre em Engenharia de Computação e Sistemas.

Aprovada em 27/02/2026.

**BANCA EXAMINADORA**

Documento assinado digitalmente  
**LUIS CARLOS COSTA FONSECA**  
Data: 09/03/2026 14:04:19-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Luís Carlos Costa Fonseca (UEMA)**  
Presidente/Orientador

Documento assinado digitalmente  
**SOFIANE BEN EL HEDI LABIDI**  
Data: 11/03/2026 14:29:40-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Sofiane Labidi (UFMA)**  
Membro Externo

Documento assinado digitalmente  
**FABIO DE JESUS LIMA GOMES**  
Data: 09/03/2026 18:02:21-0300  
Verifique em <https://validar.iti.gov.br>

---

**Prof. Dr. Fábio de Jesus Lima Gomes (IFPI)**  
Membro Externo

**SÃO LUÍS-MA**

**2026**

## **AGRADECIMENTOS**

Agradeço, primeiramente, à minha família, pelo apoio constante, incentivo e compreensão ao longo de toda a minha trajetória acadêmica.

Aos professores que contribuíram para minha formação, em especial ao meu orientador Luís Carlos e ao meu coorientador Sofiane Labidi, cujo apoio, orientação e confiança foram fundamentais para a realização deste trabalho e sem os quais não teria sido possível a concretização do intercâmbio acadêmico realizado.

À Universidade Estadual do Maranhão (UEMA), à Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) e à EURECOM, pela oportunidade do intercâmbio, que contribuiu de forma significativa para minha formação acadêmica e científica.

À Secretaria Municipal da Fazenda de São Luís (SEMFAZ-SL) e aos colegas envolvidos, pela disponibilização dos dados e pelo estímulo à ampliação de conhecimentos.

Por fim, agradeço aos colegas de turma e aos amigos em geral, pelo convívio, acompanhamento e apoio ao longo desse percurso.

## RESUMO

O Imposto Predial e Territorial Urbano (IPTU) é uma das principais fontes de receita própria dos municípios brasileiros, porém sua arrecadação é frequentemente afetada por elevados níveis de inadimplência e fragilidades cadastrais. No município de São Luís, observa-se uma defasagem significativa entre o potencial arrecadatário e os valores efetivamente arrecadados. Diante desse contexto, este estudo investiga os fatores associados à inadimplência do IPTU por meio da aplicação de técnicas de ciência de dados e algoritmos de aprendizado de máquina. Foram utilizados dados cadastrais e financeiros do IPTU referentes ao período de 2019 a 2023, combinando análises descritivas, modelagem supervisionada e métodos não supervisionados. O diagnóstico inicial evidenciou inadimplência estrutural, com cerca de 70% dos imóveis sem benefício fiscal permanecendo inadimplentes após cinco anos do lançamento e aproximadamente 45% do valor lançado não arrecadado ao final do prazo prescricional. Na etapa de modelagem supervisionada, aplicou-se um classificador multiclasse baseado no algoritmo *XGBoost*, capaz de distinguir imóveis adimplentes, adimplentes parciais e inadimplentes, mesmo em contexto de forte desbalanceamento entre classes. Os resultados indicaram que a inadimplência está fortemente associada à baixa qualidade cadastral, especialmente à ausência de CPF/CNPJ válido, além de variáveis relacionadas à infraestrutura urbana, às características construtivas dos imóveis e à localização espacial. A análise de clusterização com o algoritmo *K-prototypes* revelou elevada concentração do potencial arrecadatário em um número reduzido de grupos de imóveis inadimplentes, com maior valor médio de IPTU e maior viabilidade operacional de cobrança. Adicionalmente, a identificação de padrões atípicos e inconsistências cadastrais demonstrou o potencial dessas abordagens para apoiar a qualificação do cadastro imobiliário e a priorização de ações administrativas. Conclui-se que a inadimplência do IPTU em São Luís decorre predominantemente de fatores estruturais, cadastrais e territoriais. Os resultados evidenciam que o uso integrado de técnicas de aprendizado de máquina pode contribuir de forma significativa para o aprimoramento da gestão tributária municipal, promovendo maior eficiência arrecadatária e justiça fiscal sem a necessidade de elevação de alíquotas.

**Palavras-chave:** IPTU; inadimplência; ciência de dados; aprendizado de máquina; arrecadação municipal.

## ABSTRACT

The Urban Property Tax (IPTU) is one of the main sources of own-source revenue for Brazilian municipalities; however, its collection is often affected by high levels of tax delinquency and weaknesses in cadastral records. In the municipality of São Luís, there is a significant gap between the revenue potential and the amounts effectively collected. Within this context, this study investigates the factors associated with IPTU delinquency through the application of data science techniques and machine learning algorithms. Cadastral and financial IPTU data for the period from 2019 to 2023 were used, combining descriptive analyses, supervised modeling, and unsupervised methods. The initial diagnosis revealed structural delinquency, with approximately 70% of properties without tax exemptions remaining delinquent five years after the tax assessment, and about 45% of the assessed value not collected by the end of the statutory limitation period. In the supervised modeling stage, a multiclass classifier based on the XGBoost algorithm was applied, enabling the distinction between compliant, partially compliant, and delinquent properties, even in a context of strong class imbalance. The results indicate that tax delinquency is strongly associated with poor cadastral quality, particularly the absence of a valid CPF/CNPJ, as well as variables related to urban infrastructure, property construction characteristics, and spatial location. The clustering analysis using the K-prototypes algorithm revealed a high concentration of revenue potential within a small number of groups of delinquent properties, characterized by higher average IPTU values and greater operational feasibility for collection efforts. In addition, the identification of atypical patterns and cadastral inconsistencies demonstrated the potential of these approaches to support the improvement of property records and the prioritization of administrative actions. It is concluded that IPTU delinquency in São Luís is predominantly driven by structural, cadastral, and territorial factors. The findings show that the integrated use of machine learning techniques can significantly contribute to the improvement of municipal tax administration, promoting greater revenue efficiency and tax fairness without the need to increase tax rates.

**Keywords:** Urban Property Tax; delinquency; data science; machine learning; municipal revenue collection.

## LISTA DE FIGURAS

<b>Figura 1:</b> Mapa dos setores fiscais de São Luís.....	77
<b>Figura 2:</b> Modelo <i>XGBoost</i> - Top 20 <i>Feature Importances</i> (Global).....	81
<b>Figura 3:</b> <i>SHAP</i> do Modelo <i>XGBoost</i> - Classe adimplente.....	83
<b>Figura 4:</b> <i>SHAP</i> do Modelo <i>XGBoost</i> - Classe adimplente parcial .....	84
<b>Figura 5:</b> <i>SHAP</i> do Modelo <i>XGBoost</i> - Classe inadimplente .....	85
<b>Figura 6:</b> Método do Cotovelo ( <i>K-Prototypes</i> ).....	87

## LISTA DE GRÁFICOS

<b>Gráfico 1:</b> Arrecadação de IPTU em capitais do Nordeste (2019 a 2024).....	12
<b>Gráfico 2:</b> Percentual médio da quantidade de imóveis com IPTU pago por ano ....	59
<b>Gráfico 3:</b> Percentual médio do valor arrecadado de IPTU por ano.....	61

## LISTA DE TABELAS

<b>Tabela 1:</b> Detalhamento da quantidade de lançamentos de IPTU .....	46
<b>Tabela 2:</b> Pesos e Parâmetros para Infraestrutura e Serviços (PGV-SL) .....	53
<b>Tabela 3:</b> Detalhamento dos lançamentos cujo valor de IPTU foi maior que zero ....	57
<b>Tabela 4:</b> Detalhamento dos valores lançados de IPTU .....	58
<b>Tabela 5:</b> Detalhamento do percentual da quantidade de imóveis com IPTU pago por ano .....	60
<b>Tabela 6:</b> Detalhamento do percentual dos valores arrecadados de IPTU por ano ..	62
<b>Tabela 7:</b> Resumo dos dados de IPTU em relação à situação de adimplência .....	63
<b>Tabela 8:</b> Validade do CPF/CNPJ e situação de adimplência.....	64
<b>Tabela 9:</b> Uso do imóvel e situação de adimplência .....	64
<b>Tabela 10:</b> Condomínio e situação de adimplência .....	65
<b>Tabela 11:</b> Destinação do imóvel e situação de adimplência .....	65
<b>Tabela 12:</b> Conservação do imóvel e situação de adimplência .....	66
<b>Tabela 13:</b> Estrutura do imóvel e situação de adimplência.....	66
<b>Tabela 14:</b> Padrão construtivo e situação de adimplência .....	67
<b>Tabela 15:</b> Tipologia do imóvel e situação de adimplência .....	67
<b>Tabela 16:</b> Conservação da via e situação de adimplência .....	68
<b>Tabela 17:</b> Limpeza pública e situação de adimplência .....	68
<b>Tabela 18:</b> Acesso a transporte e situação de adimplência .....	69
<b>Tabela 19:</b> Coleta de lixo e situação de adimplência .....	69
<b>Tabela 20:</b> Arborização e situação de adimplência.....	70
<b>Tabela 21:</b> Existência de meio-fio e situação de adimplência.....	70
<b>Tabela 22:</b> Pavimentação e situação de adimplência .....	70
<b>Tabela 23:</b> Iluminação pública e situação de adimplência .....	71
<b>Tabela 24:</b> Drenagem e situação de adimplência .....	71
<b>Tabela 25:</b> Existência de calçada e situação de adimplência .....	71
<b>Tabela 26:</b> Rede telefônica e situação de adimplência .....	72
<b>Tabela 27:</b> Rede elétrica e situação de adimplência.....	72
<b>Tabela 28:</b> Rede de esgoto e situação de adimplência .....	73
<b>Tabela 29:</b> Rede de água e situação de adimplência .....	73
<b>Tabela 30:</b> Média do valor de IPTU por bairro (Top 20) .....	74

<b>Tabela 31:</b> Valor lançado de IPTU por bairro e situação de adimplência .....	75
<b>Tabela 32:</b> Quantidade de lançamentos de IPTU por bairro e situação de adimplência .....	76
<b>Tabela 33:</b> Média do valor de IPTU por setor fiscal .....	76
<b>Tabela 34:</b> Caracterização dos <i>clusters</i> obtidos a partir do algoritmo K- prototypes.	89
<b>Tabela 35:</b> <i>Clusters</i> reorganizados a partir de suas características.....	90

## SUMÁRIO

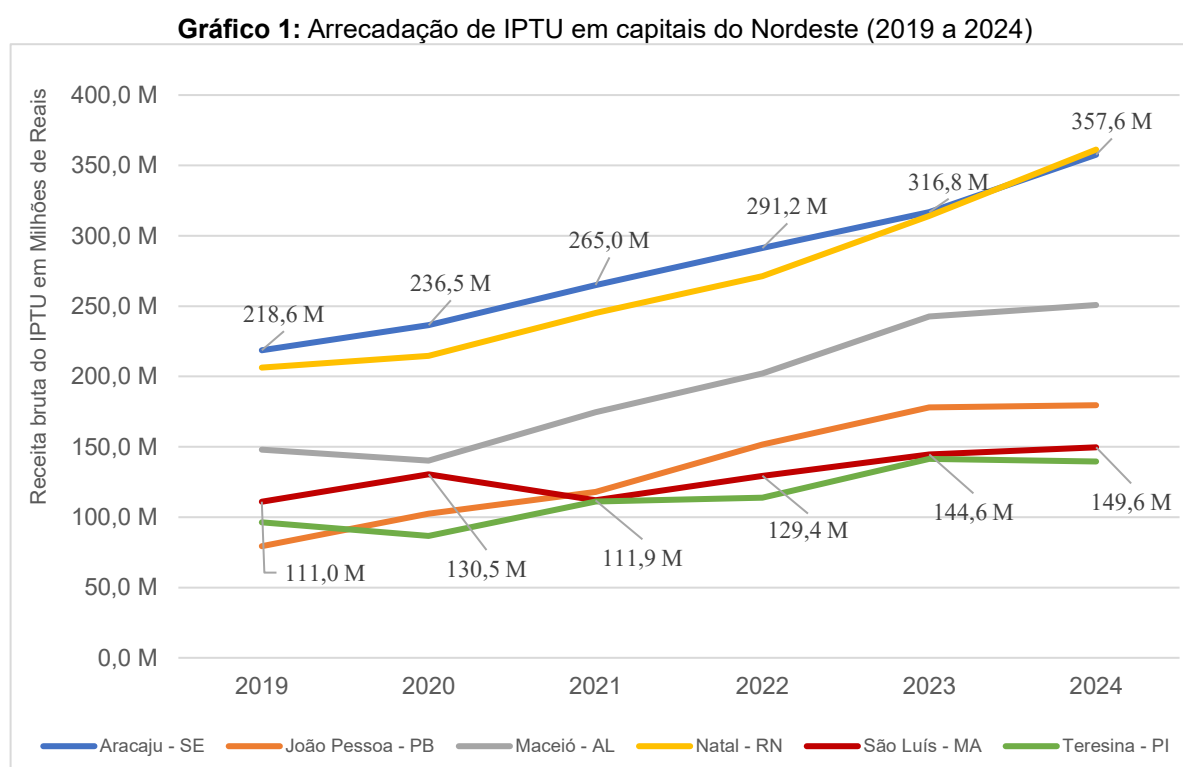
<b>1</b>	<b>INTRODUÇÃO: CONTEXTO E JUSTIFICATIVA</b>	12
<b>2</b>	<b>OBJETIVOS</b>	14
2.1	Objetivo geral	14
2.2	Objetivos específicos	14
<b>3</b>	<b>REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS</b>	15
<b>3.1</b>	<b>Tributação Municipal: Potencialidades e Dificuldades</b>	16
3.1.1	Estrutura da tributação no Brasil	17
3.1.2	O IPTU no contexto da arrecadação municipal	18
3.1.3	Inadimplência tributária	19
<b>3.2</b>	<b>Governança, Inovação e Transformação Digital na Administração Pública</b>	20
3.2.1	Gestão fiscal moderna	21
3.2.2	Governança e transformação digital no setor público	21
<b>3.3</b>	<b>Ciência de Dados e Aprendizado de Máquina</b>	23
3.3.1	Conceitos fundamentais	23
3.3.2	Técnicas de ciências de dados	26
3.3.3	Modelo de aprendizagem supervisionada: <i>XGBoost</i>	29
3.3.4	Modelo de aprendizagem não supervisionada: <i>K-prototypes</i>	31
3.3.5	Aplicações de Ciência de Dados no setor fiscal público brasileiro	32
<b>3.4</b>	<b>Análise preditiva e justiça tributária</b>	37
<b>3.5</b>	<b>Administração Tributária Municipal, tecnologias e justiça fiscal</b>	40
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	42
4.1	Pré-processamento de dados para Diagnóstico do IPTU	45
4.2	Pré-processamento de dados para Modelo de Classificação Supervisionada	46
4.3	Pré-processamento de dados para Clusterização	51
<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b>	55
<b>5.1</b>	<b>Diagnóstico do IPTU em São Luís</b>	56
5.1.1	Análise da base financeira: lançamentos com valor de IPTU maior do que zero	56
5.1.2	Análise da base financeira: lançamentos com valor de IPTU maior do que zero e sem benefício fiscal	59

<b>5.1.3</b>	<b>Análise da base cadastral e financeira sob a perspectiva da série histórica .....</b>	<b>62</b>
5.1.3.1	Aspectos fiscais e cadastrais dos imóveis.....	63
5.1.3.2	Características de infraestrutura e serviços urbanos dos imóveis ..	68
5.1.3.3	Aspectos fiscais sob a ótica espacial: setores e bairros.....	73
<b>5.2</b>	<b>Modelo Supervisionado: Classificador Multiclasse .....</b>	<b>78</b>
<b>5.3</b>	<b>Clusterização.....</b>	<b>86</b>
<b>5.4</b>	<b>Identificação de padrões atípicos e inconsistências cadastrais .....</b>	<b>92</b>
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>94</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>97</b>

## 1 INTRODUÇÃO: CONTEXTO E JUSTIFICATIVA

O Imposto Predial e Territorial Urbano (IPTU) constitui uma das principais fontes de receita própria dos municípios brasileiros, sendo essencial para o financiamento de políticas públicas, investimentos em infraestrutura e a oferta de serviços urbanos. Com a recente aprovação da Reforma Tributária no Brasil, sua relevância torna-se ainda maior, uma vez que, mesmo após as mudanças, o IPTU permanece integralmente sob competência municipal.

No entanto, como se pode observar no gráfico 1, verifica-se que a arrecadação de algumas capitais do Nordeste brasileiro apresenta desempenho substancialmente superior a outras, o que evidencia um potencial arrecadatório frequentemente não explorado.



**Fonte:** Elaborado pelo autor. Dados Primários: SICONFI - Tesouro Nacional.

Adicionalmente, de acordo com o Censo Demográfico de 2022, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), dentre as seis capitais analisadas, São Luís é a que apresenta o maior número de domicílios recenseados,

possuindo um total de 406.974 unidades, seguida por Maceió (403.096), João Pessoa (377.756), Teresina (355.470), Natal (337.029) e Aracaju (267.071).

Ao confrontar as informações dos parágrafos anteriores, observa-se que Aracaju e Natal, embora possuam o menor número de domicílios, registram as maiores arrecadações de IPTU. São Luís, por sua vez, mesmo tendo 52% mais domicílios que Aracaju, arrecada menos da metade do que esta, revelando uma expressiva defasagem em relação ao seu potencial tributário.

Outrossim, de acordo com Carvalho Júnior (2018), em 2012, municípios com população entre 70.000 e 600.000 habitantes registraram inadimplência média de aproximadamente 60% nas regiões Norte e Nordeste, enquanto nas regiões Sul, Sudeste e Centro-Oeste esse percentual foi de cerca de 33%. Nos municípios com mais de 600.000 habitantes, a inadimplência média alcançou 50% no Norte e Nordeste e 25% no Sul, Sudeste e Centro-Oeste. Ainda, segundo o autor, essa disparidade também se manifesta entre as capitais estaduais: em 2015, os níveis de inadimplência variaram de apenas 10% a 15% em Campo Grande, São Paulo, Florianópolis e Curitiba, enquanto chegaram a aproximadamente 70% em São Luís, Porto Velho, Macapá e Rio Branco.

Para mais, diversos estudos têm contribuído para a compreensão dos fatores que influenciam a inadimplência do IPTU. Gering, Pinto e Vieira (2021), ao analisarem o município de Santa Maria/RS, destacaram que variáveis como alíquota, valor venal, irregularidades cadastrais e localização do imóvel estão diretamente relacionadas à probabilidade de inadimplência.

Bartasson Ferreira Rosa, Pereira e Penedo (2023), por sua vez, demonstraram que, em Uberlândia/MG, a inadimplência está concentrada em regiões com menor renda domiciliar média, sendo influenciada pela localização do imóvel.

Bernardes (2020), em estudo desenvolvido em Fortaleza/CE, identificou que a inadimplência é mais elevada em áreas de menor valor venal, reforçando o papel das desigualdades socioespaciais nesse contexto.

Já Costa, Pecini e Fukumi Tsunoda (2021), utilizando técnicas de mineração de dados aplicadas aos registros do IPTU de Curitiba/PR, identificaram padrões de inadimplência associados ao tipo de construção, material dos imóveis e perfil de pagamento.

Essas evidências sugerem que a inadimplência do IPTU é um fenômeno multifatorial, influenciado por variáveis econômicas, urbanísticas, cadastrais e

comportamentais, e que pode ser mais bem compreendida por meio da aplicação de algoritmos e métodos de análise de dados. Ao identificar padrões ocultos nos dados, inconsistências nos cadastros e perfis de imóveis com maior propensão à inadimplência, é possível traçar estratégias mais eficazes para a recuperação de receita, promovendo justiça fiscal e eficiência na gestão tributária.

No caso de São Luís, a escassez de estudos voltados à inadimplência do IPTU e à análise sistemática de seu potencial arrecadatário indica uma lacuna relevante na literatura e na prática da administração pública local. Assim, este estudo justifica-se por razões de ordem prática, como a necessidade de fortalecer a receita pública sem aumento de alíquotas, e por razões de ordem intelectual, ao contribuir para o avanço do conhecimento sobre a relação entre estrutura urbana, perfil dos imóveis e comportamento do contribuinte.

Diante desse contexto, esta pesquisa propõe-se a investigar os fatores que influenciam a inadimplência do IPTU no município de São Luís, com base na aplicação de técnicas de análise de dados e algoritmos voltados à identificação de padrões e inconsistências. Ao integrar abordagens empíricas e tecnológicas, espera-se contribuir com subsídios concretos para o aprimoramento da gestão fiscal municipal, promovendo maior equidade tributária e eficiência na arrecadação.

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Investigar, por meio de técnicas de análise de dados e aplicação de algoritmos, os fatores associados à inadimplência do IPTU no município de São Luís, visando à identificação de padrões recorrentes, inconsistências cadastrais e potenciais oportunidades de incremento na arrecadação tributária municipal.

### **2.2 Objetivos específicos**

- a) Avaliar fatores que afetam a inadimplência do IPTU do município de São Luís.
- b) Identificar padrões de imóveis que possuem inadimplência.
- c) Mapear imóveis inadimplentes que, por características pontuais, como informações relativas ao responsável tributário, localização ou outros atributos, possuam maior potencial de arrecadação.

- d) Detectar inconsistências ou ausência de informações cadastrais.
- e) Avaliar se a inadimplência está relacionada à infraestrutura urbana e serviços públicos.

### **3 REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS**

O presente capítulo tem como finalidade estruturar o referencial teórico e apresentar os trabalhos relacionados que orientam a análise da inadimplência do Imposto Predial e Territorial Urbano (IPTU) no município de São Luís.

A abordagem adotada organiza-se em eixos temáticos que integram contribuições de diferentes áreas do conhecimento, reconhecendo que o problema em estudo exige uma leitura jurídica, econômica, administrativa e tecnológica.

O objetivo não é apenas contextualizar o IPTU no âmbito da tributação municipal, mas também evidenciar como a literatura tem tratado o fenômeno da inadimplência, quais estratégias de modernização e transformação digital vêm sendo aplicadas na administração tributária, e de que forma métodos de ciência de dados e aprendizado de máquina têm sido incorporados a esse campo de investigação.

Inicialmente, apresenta-se uma breve discussão sobre a tributação municipal e o IPTU, recuperando a estrutura constitucional e legal que define as competências tributárias dos municípios e a relevância do IPTU como fonte de receita própria. Essa parte detalha tanto os fundamentos normativos do tributo, previstos na Constituição Federal de 1988 e no Código Tributário Nacional, quanto os desafios específicos enfrentados pelos municípios brasileiros, especialmente no que diz respeito à subutilização do imposto, à defasagem dos cadastros e à elevada inadimplência.

Na sequência, discute-se a gestão fiscal moderna, a governança e a transformação digital na administração pública. Nessa parte, o foco recai sobre como a literatura tem tratado a relação entre eficiência arrecadatória e justiça fiscal, destacando propostas para modernizar os cadastros, integrar sistemas e utilizar tecnologias digitais.

Adiante, introduz-se os conceitos fundamentais da ciência de dados e do aprendizado de máquina, bem como as técnicas aplicáveis à gestão fiscal. A literatura revisada mostra como a ciência de dados emerge da integração entre estatística, computação e matemática, oferecendo um conjunto de métodos para organizar e

interpretar grandes bases de dados. São detalhadas as principais categorias de aprendizado de máquina e discutidas suas aplicações em diversas áreas, especialmente relacionada à Administração Pública.

A seção 3.4 dedica-se à análise preditiva e sua relação com a justiça tributária, destacando como técnicas estatísticas e computacionais têm sido utilizadas para antecipar riscos e otimizar auditorias.

Por fim, a seção 3.5 integra os debates anteriores ao discutir a administração tributária municipal, as tecnologias e a justiça fiscal. Nessa parte, o foco está na articulação entre gestão tributária, uso de técnicas analíticas e promoção de equidade, reconhecendo que o enfrentamento da inadimplência do IPTU exige tanto avanços institucionais quanto soluções tecnológicas.

Ao longo das seções 3.3, 3.4 e 3.5 são apresentados os trabalhos relacionados ao uso de ciência de dados no setor público brasileiro, que constituem referência direta para esta dissertação.

Assim, o referencial teórico e os trabalhos relacionados aqui apresentados foram estruturados de modo a oferecer uma base sólida para a análise da inadimplência do IPTU em São Luís. A construção do capítulo segue uma progressão lógica: parte da estrutura do sistema tributário e da função do IPTU, avança para os desafios da inadimplência, introduz as propostas de modernização administrativa e de transformação digital, explora as técnicas da ciência de dados e do aprendizado de máquina aplicáveis ao problema, revisa experiências nacionais e internacionais e, por fim, integra esses elementos em um quadro que relaciona administração tributária, inovação tecnológica e justiça fiscal. Essa organização permite não apenas situar o estudo em um contexto multidisciplinar, mas também evidenciar de forma clara como a literatura fundamenta a metodologia adotada nesta pesquisa.

### **3.1 Tributação Municipal: Potencialidades e Dificuldades**

Esta seção está organizada em três partes. Primeiro, apresenta-se a estrutura da tributação no Brasil e a posição do IPTU dentro do federalismo fiscal. Depois, discute-se a relevância do imposto para a arrecadação municipal e suas funções fiscal, extrafiscal e social. Por fim, analisa-se a inadimplência tributária, destacando suas causas, consequências e os desafios que impõe à sustentabilidade financeira dos municípios, com ênfase no caso de São Luís.

### 3.1.1 Estrutura da tributação no Brasil

O sistema tributário brasileiro organiza-se em um modelo de federalismo fiscal, no qual União, Estados, Distrito Federal e Municípios detêm competências próprias para instituir e arrecadar tributos, conforme previsto na Constituição Federal de 1988 (CF/88). O artigo 145 estabelece os impostos como principal espécie tributária, enquanto os artigos 153 a 156 detalham a competência de cada ente federativo. Especificamente, o artigo 156, inciso I e III, atribui aos municípios a competência para instituir o Imposto Predial e Territorial Urbano (IPTU) e o Imposto sobre Serviços (ISS), tributos que constituem parte relevante de sua receita própria.

O federalismo fiscal brasileiro é marcado por forte centralização da arrecadação na União, contrastando com a crescente descentralização das responsabilidades administrativas e de prestação de serviços devidas pelos municípios à população.

Esse desenho da repartição de competências revela que, apesar de o Brasil adotar sistema fiscal federativo, a União detém, em termos quantitativos e qualitativos, poderes mais robustos de arrecadação, comparado aos demais entes. Para mitigar essa distinção foram criados mecanismos de repartição de receitas, dentre as quais destacam-se o Fundo de Participação dos Estados (FPE) e o Fundo de Participação dos Municípios (FPM), conforme aponta Rezende (2018) *apud* Müller; Da Silveira; Cademartori (2025).

Essa assimetria gera desafios relacionados à autonomia financeira municipal. Adiciona-se que, embora os municípios tenham competência constitucional para instituir tributos sobre a propriedade urbana, a subutilização dessa competência limita sua capacidade de financiar políticas públicas locais. Castro e Afonso (2017) apontam que a maioria dos municípios brasileiros não explora plenamente o potencial arrecadatório do IPTU, o que é especialmente crítico em localidades menores, mais dependentes de transferências intergovernamentais.

Assim, a estrutura da tributação no Brasil revela-se tensionada entre dois polos: de um lado tem-se a competência constitucional dos municípios, que garante instrumentos próprios de arrecadação e reforça sua autonomia financeira, mas que muitas vezes é subaproveitada; de outro, há a dependência prática de transferências, que limita a plena efetividade do federalismo fiscal. Esse cenário evidencia a

necessidade de fortalecer os instrumentos de gestão tributária municipal como forma de reduzir a vulnerabilidade fiscal.

### 3.1.2 O IPTU no contexto da arrecadação municipal

Como já mencionado, a Constituição Federal do Brasil de 1988, em seu artigo 156, inciso I, atribui aos Municípios a competência para instituir o Imposto Predial e Territorial Urbano (IPTU), consolidando-o como um dos principais instrumentos de arrecadação tributária local.

Essa competência é detalhada pelo Código Tributário Nacional, que disciplina os elementos essenciais do tributo: o artigo 32 estabelece como fato gerador a propriedade, o domínio útil ou a posse de bem imóvel localizado em zona urbana; o artigo 33 define que a base de cálculo do imposto é o valor venal do imóvel; e o artigo 34 determina que o contribuinte é o proprietário, o titular do domínio útil ou o possuidor a qualquer título.

Em âmbito local, a Lei nº 6.289, de 28/12/2017 que institui o Novo Código Tributário do Município de São Luís, regulamenta o IPTU no Município de São Luís, disciplinando hipóteses de incidência e alíquotas (inclusive diferenciações por destinação/faixa de valor venal), em consonância com os princípios da capacidade contributiva e da função social da propriedade.

À luz desses princípios, é importante ressaltar que o IPTU possui função fiscal, como fonte de receita própria municipal; função extrafiscal, ao induzir o adequado aproveitamento do solo urbano, como por exemplo ao utilizar progressividade e instrumentos urbanísticos compatíveis com o artigo 182, §4º, da CF/88; e função social, vinculada à efetivação da função social da propriedade e à correção de distorções na ocupação do espaço urbano.

No contexto de São Luís, o IPTU integra o núcleo das receitas próprias e tem papel estratégico para financiar serviços locais e reduzir a dependência de transferências. A Lei nº 6.289/2017 organiza os parâmetros de cálculo que utiliza a Planta Genérica de Valores Imobiliários do Município (PGV), define alíquotas e permite isenções e calibragens compatíveis com os objetivos fiscais (sustentação orçamentária) e extrafiscais/sociais (ordenamento urbano e justiça tributária).

Um ponto de destaque acerca deste imposto, evidenciado pela literatura, refere-se às falhas na definição do valor venal e na condução de procedimentos administrativos, que comprometem a equidade do IPTU. Borges e Souza (2004)

demonstram que práticas inadequadas de apuração da base de cálculo podem resultar em uma tributação desproporcional, onerando relativamente mais os imóveis de menor valor e favorecendo os de maior valor de mercado. Esse desequilíbrio gera impactos negativos sobre a justiça fiscal e contribui para a reprodução de desigualdades na distribuição de renda.

E para reforçar, a literatura supracitada já apontou que melhorias em avaliação, cadastro e cobrança, tendem a elevar o uso do potencial do IPTU, com mitigação de iniquidades e aumento de arrecadação quando bem implementadas.

### **3.1.3 Inadimplência tributária**

A inadimplência tributária, no âmbito do IPTU, corresponde ao não pagamento do imposto lançado pelo município, configurando-se como um dos principais entraves à eficiência arrecadatória. De acordo com Bahl (1979) *apud* Carvalho Júnior (2018), a arrecadação do imposto imobiliário depende de fatores como valores de mercado dos imóveis, o nível de cobertura cadastral, o nível de avaliação dos imóveis com relação ao mercado, da alíquota média, considerando-se as isenções e os descontos concedidos, e o nível de inadimplência.

Os estudos sobre o Brasil mostram que a inadimplência no IPTU atinge patamares elevados e apresenta forte heterogeneidade entre regiões e portes municipais. Carvalho Júnior (2018) aponta que, em 1998, os índices médios de inadimplência do IPTU em municípios de até 600 mil habitantes alcançavam cerca de 70% nas regiões Norte e Nordeste e 50% nas regiões Sul, Sudeste e Centro-Oeste. Em 2012, os percentuais caíram, mas ainda permaneceram altos: aproximadamente 60% no Norte/Nordeste e 33% no Sul/Sudeste/Centro-Oeste. Mesmo entre as capitais, a discrepância era significativa: enquanto São Paulo e Curitiba registraram taxas entre 10% e 15%, São Luís, Porto Velho e Macapá apresentaram índices próximos de 70%.

Ademais, Carvalho Júnior (2018) identifica três problemas estruturais na administração do IPTU no Brasil. O primeiro refere-se à necessidade de cadastros imobiliários e de contribuintes atualizados e integrados, o que exige o uso de tecnologias como georreferenciamento e informatização dos registros. O segundo diz respeito à base de cálculo do imposto, o valor venal dos imóveis, cuja defasagem compromete a equidade tributária. Esse problema é agravado pela Súmula 160 do Superior Tribunal de Justiça, que condiciona sua atualização à aprovação legislativa,

atribuindo caráter político a uma atividade eminentemente técnica. E o terceiro problema relaciona-se à cobrança da dívida ativa, considerada complexa, onerosa e pouco efetiva, o que, segundo estimativas, apenas 25% da dívida ativa federal é recuperada, a um custo médio de R\$ 6.000 por processo (CUNHA; KLIN; PESSOA, 2011, *apud* CARVALHO JÚNIOR, 2018). Em municípios de pequeno porte, a insuficiência de estrutura fazendária, a dificuldade de localização de contribuintes e os custos políticos associados à execução fiscal intensificam esse cenário, restringindo a capacidade arrecadatória local.

Os impactos da inadimplência são profundos e atingem diretamente a efetividade do IPTU. Segundo Kelly (2013) *apud* Carvalho Júnior (2018), nenhum objetivo extrafiscal do IPTU se efetiva se o imposto não for pago, pois a progressividade ou a indução ao uso adequado do solo urbano tornam-se ineficazes diante do não cumprimento da obrigação principal. Assim, a inadimplência não apenas compromete a função fiscal do tributo, reduzindo a arrecadação e a capacidade de investimento municipal, mas também inviabiliza a realização de políticas urbanas de caráter redistributivo e social.

Carvalho Júnior (2018) sugere diversas medidas para mitigar esse problema, entre as quais: programas de educação fiscal para estimular a adimplência voluntária; facilitação do pagamento com parcelamento e débito em conta; recadastramentos periódicos e integração com cadastros de outros órgãos; protesto da dívida ativa em cartório, nos termos da Lei nº 12.767/2012 (que alterou a Lei nº 9.492/1997); além de consórcios intermunicipais de administração tributária e aperfeiçoamento do Programa de Modernização da Administração Tributária (PMAT).

Em síntese, a inadimplência tributária é uma das principais causas da baixa eficiência arrecadatória do IPTU no Brasil. Sua persistência compromete tanto a sustentabilidade fiscal dos municípios quanto a efetividade do imposto como instrumento de justiça social e de ordenamento urbano.

### **3.2 Governança, Inovação e Transformação Digital na Administração Pública**

Esta seção discorre acerca do conceito de gestão fiscal moderna, discutindo como eficiência e justiça podem atuar de forma complementar na administração tributária e posteriormente aborda a governança e a transformação digital no setor público, destacando a relevância de tecnologias aplicadas à gestão

fiscal, bem como experiências nacionais e internacionais que apontam para práticas mais eficientes, transparentes e inclusivas.

### **3.2.1 Gestão fiscal moderna**

A gestão fiscal moderna busca conciliar eficiência arrecadatória e justiça fiscal como princípios fundamentais da administração tributária. Segundo Vettorato, Afonso e Fuck (2020, p. 97), “a eficiência da Administração Tributária pode ser entendida como a busca de recursos públicos pela arrecadação de tributos com fins de garantir direitos fundamentais com menor geração de custos, tanto para o poder público quanto para o contribuinte”. Esse conceito representa um ponto de equilíbrio entre justiça tributária e neutralidade tributária, uma vez que implica que os custos do sistema não devem comprometer de forma desproporcional nem a atuação estatal, nem a capacidade contributiva do cidadão.

Desse modo, eficiência e justiça não se apresentam como princípios antagônicos, mas complementares. A eficiência arrecadatória permite que o município maximize a captação de receitas próprias com menor dispêndio de recursos administrativos, ao passo que a justiça fiscal assegura que essa arrecadação se realize em conformidade com a capacidade contributiva e a isonomia tributária, princípios consagrados nos artigos 145, §1º, e 150, II, da CF/88.

Nesse contexto, a modernização da gestão fiscal municipal envolve o aperfeiçoamento de cadastros, a utilização de tecnologias de informação e a integração entre órgãos fazendários, de forma a reduzir desigualdades de tratamento, aumentar a transparência e garantir maior previsibilidade orçamentária. Ao incorporar práticas que alinham eficiência e equidade, o poder público fortalece não apenas a arrecadação, mas também a legitimidade do sistema fiscal perante a sociedade.

### **3.2.2 Governança e transformação digital no setor público**

A transformação digital no setor público, frequentemente associada ao conceito de governo eletrônico (e-gov), constitui uma das principais estratégias para enfrentar os desafios da gestão fiscal. O e-gov compreende a utilização de tecnologias da informação e comunicação (TICs) para aprimorar a eficiência administrativa, a transparência e a participação cidadã na gestão pública.

No Brasil, iniciativas de governo eletrônico ainda apresentam caráter incipiente e fragmentado, sobretudo no nível municipal. Guimarães, Araújo e Costa (2020) destacam que apesar de os nove municípios do Nordeste que estão entre as

100 cidades que figuraram no *Ranking Connected Smart Cities* de 2018 possuem ações *smart*, como alta qualidade e cobertura de internet, elas ainda permanecem pontuais e desintegradas, não sendo suficientes para solucionar os problemas urbanos. Os autores ressaltam que projetos inteligentes devem estar orientados para o desenvolvimento das funções sociais da cidade, mediadas pela tecnologia, com vistas ao bem-estar da população.

Nesse sentido, a governança digital deve ser compreendida não apenas como informatização de serviços, mas como uma mudança estrutural na administração pública, que exige integração de sistemas, interoperabilidade entre órgãos e participação social efetiva. Nessa mesma linha, a noção de cidades inteligentes ultrapassa a mera aplicação de tecnologias, pressupondo a articulação entre sistemas físicos, digitais e humanos. Partindo da concepção de que a cidade inteligente é aquela voltada à construção de um ambiente sustentável, próspero e inclusivo (NETTO, 2017, *apud* RODRIGUES, 2022, p. 7), observa-se que a efetividade da *smart city* depende tanto da dimensão tecnológica quanto do engajamento dos atores sociais no processo de governança.

Por conseguinte, a incorporação de tecnologias de gestão fiscal às estratégias de cidades inteligentes pode contribuir para ampliar a arrecadação de tributos como o IPTU e, ao mesmo tempo, orientar políticas urbanas mais justas e sustentáveis. O uso de dados georreferenciados, inteligência artificial e plataformas de participação social reforça a integração entre eficiência arrecadatória e justiça fiscal, tornando a gestão pública mais responsiva às demandas da sociedade.

Ademais, um dos avanços mais relevantes da transformação digital é a utilização de dados e sistemas inteligentes para subsidiar a tomada de decisão na administração tributária. Experiências internacionais reforçam o potencial desses instrumentos. Khan et al. (2025) analisam a aplicação de Sistemas de Informação Geográfica (SIG ou GIS) no processo de avaliação de impostos imobiliários na cidade de Shillong, na Índia. Os autores mostram que a adoção dessa tecnologia aumentou a precisão na identificação de imóveis, reduziu falhas de valoração, ampliou a transparência e contribuiu para maior justiça distributiva. Apesar desses avanços, alertam que a manutenção de dados atualizados e a mitigação de disparidades socioeconômicas ainda são desafios centrais. Esse tipo de experiência revela que o uso de tecnologias baseadas em dados pode beneficiar diretamente a arrecadação municipal, ao tornar os processos mais eficientes e justos. A incorporação de sistemas

inteligentes também favorece a governança, uma vez que cria condições para decisões mais baseadas em evidências e mais transparentes para a sociedade.

Portanto, as evidências discutidas nos trabalhos de Vettorato, Afonso e Fuck (2020), Guimarães, Araújo e Costa (2020), Khan et al. (2025) e Rodrigues (2022) convergem para o entendimento de que a modernização da gestão fiscal só se realiza plenamente quando associada à inovação tecnológica e ao fortalecimento da governança democrática. Em outras palavras, a eficiência arrecadatória só alcança sua finalidade maior que é o financiamento de políticas públicas, quando acompanhada da justiça fiscal e de uma governança inclusiva.

### **3.3 Ciência de Dados e Aprendizado de Máquina**

Esta seção apresenta os fundamentos da Ciência de Dados e do Aprendizado de Máquina, explorando suas origens, conceitos e classificações, com ênfase em como essas abordagens se consolidaram como instrumentos centrais no tratamento e interpretação de grandes volumes de informação. Em seguida, são discutidas as principais técnicas aplicáveis à gestão fiscal, destacando seu potencial para análise de inadimplência, modernização cadastral e segmentação de contribuintes. Por fim, são analisadas aplicações da literatura nacional voltadas ao setor fiscal público brasileiro, evidenciando experiências recentes que demonstram ganhos de eficiência, justiça fiscal e sustentabilidade na arrecadação a partir do uso dessas metodologias.

#### **3.3.1 Conceitos fundamentais**

Segundo Borges e Almeida (2024), a Ciência de Dados (CD) apresenta o seu prólogo nas interseções entre a Estatística, a Matemática e a Ciência da Computação. Os autores ressaltam que a CD é reconhecida como uma área interdisciplinar, ainda em expansão, mas capaz de agregar valor em diferentes aplicações.

De acordo com Matos, Condurú e Benchimol (2022) *apud* Borges e Almeida (2024), a CD surgiu no início da década de 2000, em um contexto de crescente robustez dos dados digitais que afetam ambientes acadêmicos e organizacionais. Para esses autores, a CD tem como premissa “[...] a criação de técnicas, tecnologias e teorias para a utilização de dados provenientes de fontes heterogêneas e complexas” (Matos; Condurú; Benchimol, 2022, p. 2 *apud* Borges; Almeida, 2024, p.

48), constituindo-se como suporte para a tomada de decisões e a geração de valor a partir da informação.

Segundo Urs e Minhaj (2022) *apud* Borges e Almeida (2024), a constituição da CD como disciplina está relacionada à visão de Cleveland, que sugeriu a criação de uma nova área oriunda da Estatística, mas incorporando práticas da Computação e da Mineração de Dados. Ainda conforme Urs e Minhaj (2022) *apud* Borges e Almeida (2024), a trajetória da CD evoluiu do simples processamento de dados para a descoberta de conhecimento em bases de dados, o uso de ontologias e a extração de informações da web semântica, integrando mineração de dados, aprendizado de máquina e análise descritiva.

Pacheco e Disconzi (2019) destacam que, embora a expressão Data Science remonte aos anos 1960, trata-se de uma ciência recente. Para Amaral (2016) *apud* Pacheco e Disconzi (2019), a CD pode ser definida como o conjunto de processos, modelos e tecnologias voltados ao estudo do ciclo de vida do dado. Os autores complementam que se trata de um campo em constante evolução, no qual novos saberes vêm sendo incorporados ao longo do tempo.

Na visão de Soares (2020), a CD constitui um estudo sistemático sobre a organização, propriedades e análise de dados estruturados e não estruturados, incluindo inferências. Acrescenta-se que, segundo Cetax (2019) *apud* Soares (2020), a CD combina estatística, matemática e soluções computacionais para capturar dados, detectar padrões e realizar atividades de limpeza, preparação e organização. Essa definição é considerada abrangente por incluir estatística, matemática, mineração de dados, big data, visualização e aprendizado de máquina.

O aprendizado de máquina, segundo Nascimento, Santos et al. (2018) *apud* Soares (2020), representa um segmento da Inteligência Artificial que constrói modelos matemáticos baseados em dados existentes, sem a necessidade de programação explícita. Essa técnica é amplamente utilizada para seleção de atributos, detecção de fraudes e otimização de parâmetros de modelos. Complementarmente, Domingos (2012) *apud* Soares (2020) destaca que o aprendizado de máquina é um processo contínuo, no qual grande parte do tempo é dedicada à coleta, integração e limpeza de dados, além de sucessivas etapas de tentativa e erro.

A mineração de dados, conforme o *SIGKDD Curriculum Committee* (2006) *apud* Soares (2020), consiste na extração de padrões a partir de grandes volumes de

dados, utilizando métodos estatísticos, de inteligência artificial e de sistemas de banco de dados.

Blei e Smyth (2017) *apud* Morettin e Singer (2020) analisam a CD sob três perspectivas: estatística, computacional e humana. Sob a ótica estatística, a CD se beneficia da coleta e análise de dados complexos; sob a computacional, da criação de algoritmos capazes de processar grandes volumes em alta velocidade; e sob a humana, do uso de modelos e métodos para resolver problemas específicos, selecionar dados relevantes, processá-los e comunicar os resultados de forma compreensível.

Cao (2016) *apud* Matos, Condurú e Benchimol (2022) também reforça que a CD busca desenvolver técnicas, tecnologias e teorias que permitam a utilização de dados heterogêneos e complexos para apoiar a tomada de decisão.

Adicionalmente, o Aprendizado de Máquina (AM) é uma área da Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais que permitam a construção de sistemas capazes de adquirir conhecimento de forma automática. Segundo Monard e Baranauskas (2003) *apud* Pacheco e Disconzi (2019), um sistema de aprendizado pode ser definido como um programa de computador que toma decisões com base em experiências acumuladas por meio da resolução de problemas anteriores. Nesse sentido, Pacheco e Disconzi (2019) destacam que o processo consiste essencialmente em adquirir conhecimento a partir de experiências bem-sucedidas.

A Inteligência Artificial, de modo geral, pode ser compreendida como o estudo de algoritmos computacionais que simulam comportamentos inteligentes, interpretando dados externos e utilizando essa aprendizagem para atingir objetivos específicos (Sayad, 2019, *apud* Soares, 2020). Dentro dessa perspectiva, o AM constitui um subcampo da IA em que os algoritmos ajustam automaticamente seus modelos à medida que processam os dados, segundo Soares (2020).

As técnicas de AM podem ser divididas em três grandes categorias: supervisionadas, não supervisionadas e híbridas. De acordo com Leskovec, Rajaraman e Ullman (2014) *apud* Soares (2020), os métodos supervisionados trabalham com dados rotulados, isto é, utilizam atributos previamente classificados para extrair padrões que permitam realizar previsões ou classificações. Por outro lado, os métodos não supervisionados operam sobre dados não rotulados, em que não há

categorias ou classes pré-definidas, buscando agrupar ou estruturar os dados com base em similaridades.

Morettin e Singer (2020) complementam que, no aprendizado supervisionado, o objetivo é utilizar variáveis de entrada (*inputs*) para prever respostas de saída (*outputs*), que podem ser quantitativas (previsão) ou qualitativas (classificação). Problemas de classificação incluem, por exemplo, identificar clientes de um banco como bons ou maus pagadores a partir de variáveis como salário, idade e classe social, utilizando métodos como regressão logística, análise discriminante, vizinho mais próximo, árvores de decisão e máquinas de vetor de suporte (*Support Vector Machines*). Já os problemas de previsão consistem em estimar valores futuros, como o saldo médio de clientes, por meio de técnicas de regressão, árvores de decisão ou algoritmos de suporte vetorial.

Além desses, o AM conta com algoritmos não supervisionados, tradicionalmente empregados em técnicas de agrupamento e redução de dimensionalidade. Métodos como Análise de Componentes Principais, Análise Fatorial, Análise de Componentes Independentes e técnicas de *clustering* são utilizados para identificar padrões em dados sem rótulos, conforme apontam Morettin e Singer (2020).

Nos últimos anos, métodos híbridos ou avançados ganharam destaque, englobados no campo da Ciência de Dados. Entre eles, estão algoritmos como *Random Forests*, *Bagging* e *Boosting*, que combinam características tanto de abordagens supervisionadas quanto de não supervisionadas, buscando melhorar a precisão das previsões e a robustez dos modelos, afirmam Morettin e Singer (2020).

Assim, o Aprendizado de Máquina apresenta-se como um campo central da Inteligência Artificial, no qual coexistem diferentes paradigmas de algoritmos. Enquanto os supervisionados buscam prever resultados a partir de dados rotulados, os não supervisionados identificam padrões em dados brutos, e os híbridos integram ambos os enfoques, oferecendo soluções mais completas e eficazes.

### **3.3.2** Técnicas de ciências de dados

O cenário contemporâneo de gestão fiscal enfrenta desafios cada vez mais complexos diante da expansão contínua das bases de dados tributárias. Os registros cadastrais e financeiros, produzidos e armazenados por administrações públicas, alcançam volume e variedade de difícil tratamento por métodos convencionais.

Informações heterogêneas, compostas por atributos numéricos, textos descritivos, imagens de documentos digitalizados e até dados geoespaciais, tornam o processo de análise mais oneroso e suscetível a erros. Nesse contexto, emergem as técnicas de mineração de dados e de aprendizado de máquina, que possibilitam novas estratégias de exploração e interpretação de grandes volumes de informação, gerando conhecimento útil para apoiar a tomada de decisão governamental.

No âmbito tributário, a inadimplência representa uma das maiores dificuldades enfrentadas pelas administrações municipais, pois compromete não apenas o fluxo de arrecadação, mas também a capacidade de planejamento e execução de políticas públicas. O emprego de métodos avançados de análise de dados permite identificar padrões de comportamento entre contribuintes, prever riscos de inadimplência e detectar casos atípicos ou fraudulentos. Dessa forma, a integração entre mineração de dados, modelos preditivos e técnicas de clusterização e detecção de anomalias pode oferecer um caminho promissor para modernizar a administração tributária e aumentar a eficiência da arrecadação municipal.

A mineração de dados pode ser compreendida como o processo de extração de conhecimento a partir de grandes bases, envolvendo técnicas estatísticas, algoritmos de aprendizado de máquina e tecnologias de banco de dados, segundo Shu e Ye (2022). No caso das bases tributárias, os dados apresentam alta complexidade, visto que englobam tanto variáveis estruturadas (valores venais, áreas de imóveis, histórico de quitação) quanto registros não estruturados (anotações relacionadas a processos judiciais, justificativas de revisão de lançamentos, imagens de plantas, vistorias em campo ou escrituras). Esse conjunto exige não apenas a consolidação e integração das informações, mas também procedimentos rigorosos de limpeza, transformação e padronização.

De acordo com Shu e Ye (2022), três áreas acadêmicas compõem a base da mineração de dados: a estatística, que fornece técnicas para identificar relações entre variáveis e representar padrões; a inteligência artificial, que treina computadores a reconhecer estruturas complexas; e os sistemas de banco de dados, que garantem o armazenamento e a recuperação eficiente de grandes volumes de informação. Esse tripé é especialmente relevante no domínio tributário, em que a heterogeneidade dos dados aumenta os riscos de inconsistência e exige soluções híbridas.

A diversidade dos dados (numéricos, textuais, geográficos) torna imprescindível o uso de técnicas de pré-processamento, capazes de converter

entradas desestruturadas em formatos compatíveis com algoritmos de mineração, consoante Shu e Ye (2022). Isso é particularmente visível em cadastros municipais, em que descrições de uso do imóvel, registros de fiscalizações e observações de atendimentos precisam ser transformados em categorias normalizadas para análise quantitativa.

Segundo Khalid e Prieto-Alhambra (2019), o processo de seleção de atributos (*feature selection*) é uma das etapas mais importantes, pois reduz a dimensionalidade, elimina variáveis redundantes e aumenta a precisão dos modelos. No caso da inadimplência, atributos como histórico de quitação, tipo de uso do imóvel, zona de localização e valor venal podem ser priorizados para identificar padrões significativos.

Assim, a mineração de dados em bases tributárias não se limita à organização de informações, mas constitui um processo estratégico para transformar dados brutos em conhecimento acionável, capaz de orientar políticas fiscais e subsidiar decisões de cobrança.

Adicionalmente, o uso de modelos preditivos em análise de risco ganhou força com o avanço das técnicas de aprendizado de máquina. Esses modelos permitem estimar a probabilidade de inadimplência a partir de atributos históricos e contextuais, possibilitando não apenas antecipar perdas de receita, mas também adotar medidas preventivas. Nogueira, Gama e Ferreira (2020) desenvolveram um arcabouço baseado em variáveis causais probabilísticas, capazes de capturar relações estruturais entre atributos e melhorar o desempenho de algoritmos como *Random Forest*. Para dados fiscais, tal abordagem é útil para diferenciar relações de correlação simples (por exemplo, sazonalidade nos pagamentos) de relações causais mais profundas (como choques econômicos que afetam categorias específicas de contribuintes).

Enquanto os modelos preditivos se concentram em estimar riscos individuais, a clusterização busca compreender a estrutura coletiva dos dados, agrupando contribuintes em perfis homogêneos. A clusterização é especialmente útil para segmentar contribuintes inadimplentes em grupos que compartilham características semelhantes, como localização geográfica, faixa de valor venal ou padrão histórico de atraso. Essa segmentação permite que a administração pública adote estratégias diferenciadas, seja oferecendo renegociação para determinados perfis, seja intensificando a fiscalização em outros.

A detecção de anomalias, por sua vez, constitui um recurso essencial para identificar contribuintes cujo comportamento diverge do padrão esperado. Chenniappanadar et al. (2022) destacam que, em segurança cibernética, a detecção baseada em anomalias ganhou espaço justamente por sua eficácia em reconhecer eventos não previstos por assinaturas conhecidas. A analogia com o campo tributário é direta: contribuintes que apresentam saltos abruptos de dívida, declarações inconsistentes ou padrões de pagamento incomuns podem ser sinalizados como potenciais inadimplentes ou até fraudadores. Assim, a combinação de clusterização e detecção de anomalias permite tanto compreender a estrutura coletiva do problema da inadimplência quanto identificar casos individuais críticos, oferecendo aos gestores fiscais instrumentos mais precisos para a formulação de estratégias.

Por conseguinte, a integração entre mineração de dados, modelos preditivos e técnicas de clusterização e anomalia representa uma evolução significativa no tratamento da inadimplência tributária. Enquanto a mineração de dados garante a preparação e a organização das bases, os modelos preditivos oferecem estimativas probabilísticas robustas do risco de inadimplência, e as técnicas de clusterização e anomalia ampliam a compreensão coletiva e individual dos padrões de comportamento dos contribuintes.

Essas abordagens não apenas aumentam a eficiência administrativa e a capacidade arrecadatória, mas também promovem maior justiça fiscal, ao permitir que medidas diferenciadas sejam aplicadas conforme os perfis de risco. Para os municípios, isso significa aliar inovação tecnológica a sustentabilidade financeira, transformando dados em conhecimento estratégico para o aprimoramento da gestão pública.

### **3.3.3** Modelo de aprendizagem supervisionada: *XGBoost*

Consoante Chen e Guestrin (2016), o *XGBoost* (*eXtreme Gradient Boosting*) é um modelo de aprendizado de máquina de *tree boosting* escalável e altamente eficiente, amplamente utilizado por cientistas de dados para obter resultados de estado da arte em diversos desafios competitivos. Ele é uma implementação otimizada do algoritmo de *Gradient Boosting Machine* (GBM), projetado para ser rápido e preciso.

Ainda de acordo com os autores, o funcionamento do *XGBoost* baseia-se em um modelo de conjunto de árvores, que utiliza múltiplas árvores de decisão cujas

previsões são somadas para gerar o resultado. O processo de treinamento é aditivo, o que significa que o modelo adiciona novas árvores sequencialmente para minimizar uma função de perda. Eles mencionam que o *XGBoost* possui os seguintes mecanismos técnicos:

a) Otimização de segunda ordem, o que permite o algoritmo ser mais rápido e genérico;

b) Regularização que penaliza a complexidade do modelo, como o número de folhas, o que ajuda a suavizar os pesos aprendidos e evitar o *overfitting*;

c) Algoritmo *sparsity-aware*, para lidar com dados esparsos (valores ausentes, zeros ou codificações *one-hot*), aprendendo automaticamente uma "direção padrão" para esses valores em cada nó da árvore, o que pode acelerar o processamento em até 50 vezes;

d) Inovações de Sistema: Ele utiliza estruturas de blocos em colunas para aprendizado paralelo, acesso consciente ao cache e computação *"out-of-core"* para processar conjuntos de dados que não cabem na memória principal.

O algoritmo é considerado por muitos a melhor escolha para dados estruturados ou tabulares de pequeno a médio porte. Suas aplicações abrangem diversos tipos de dados: classificação de eventos físicos; previsão de vendas e comportamento de clientes; análise de satisfação em serviços públicos e detecção de fraudes, além de funcionarem bem em dados com alta dimensionalidade e presença de ruído ou valores faltantes.

Chen e Guestrin (2016) também explicam que a popularidade do *XGBoost* se deve a três fatores principais:

a) Escalabilidade: ele consegue lidar com bilhões de exemplos usando muito menos recursos computacionais do que outros sistemas.

b) Velocidade: em uma única máquina, pode rodar mais de dez vezes mais rápido que soluções populares anteriores.

c) Eficácia Comprovada: Em 2015, das 29 soluções vencedoras de competições no *Kaggle*, 17 utilizaram o *XGBoost*. Ele venceu todas as dez principais equipes no *KDDCup* 2015.

Adicionalmente, consoante Meng *et al* (2020), o *SHAP* (*SHapley Additive exPlanations*) é uma técnica de atribuição de características baseada na teoria dos jogos usada para interpretar as previsões do modelo. Como modelos de *boosting* são frequentemente vistos como "caixas-pretas", o *SHAP* resolve esse problema ao:

a) Quantificar a contribuição: atribui um valor para cada variável, indicando quanto ela contribuiu para uma previsão específica, tanto em nível global quanto individual;

b) Identificar Interações: o método *SHAP* permite capturar e quantificar como duas variáveis interagem conjuntamente para influenciar o resultado.

c) Garantir Consistência: é o único método de atribuição de características individualizado que é consistente e localmente preciso para modelos baseados em árvores.

Em resumo, enquanto o *XGBoost* fornece o poder preditivo, o *SHAP* oferece a explicabilidade necessária para entender quais fatores, como idade, renda ou localização, estão de fato movendo as previsões do modelo.

### 3.3.4 Modelo de aprendizagem não supervisionada: *K-prototypes*

A análise de agrupamento (*clustering*) é uma técnica de aprendizado não supervisionado que organiza dados em segmentos ou grupos (*clusters*) com base em similaridades, permitindo a simplificação de informações complexas.

Os algoritmos de agrupamento variam dependendo da natureza dos dados e do objetivo da análise. Segundo Valderrama e Flores (2021), na clusterização por centróide (particional) cada grupo é representado por um centro ou centróide. O algoritmo busca particionar os dados em um número específico de grupos (*k*) minimizando a distância dos pontos em relação aos seus respectivos centros. Exemplos clássicos incluem o *K-means* e o *K-prototypes*.

Conforme explica Ferreira (2015), diferente dos baseados em centróides (que tendem a criar grupos esféricos), algoritmos por densidade, como o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) identificam *clusters* com base na concentração de pontos no espaço. Um grupo é formado onde a densidade de pontos é alta, permitindo identificar grupos de formatos irregulares e tratar dados ruidosos.

Ademais, De Matos (2007) explicita que a clusterização hierárquica organiza os dados em uma estrutura de árvore (dendograma), podendo ser aglomerativa, começa com cada ponto como um grupo e os une gradualmente, ou divisiva, começa com todos em um único grupo e os separa.

Ferreira (2015) adiciona que o *K-means* é um dos algoritmos particionais mais utilizados e foca em dados numéricos. Ele divide o conjunto de dados em *k*

grupos pré-especificados, atribuindo cada observação ao centróide mais próximo, com base na distância euclidiana, de forma a minimizar a variância intragrupo. Contudo, o algoritmo apresenta limitações no tratamento de dados categóricos, uma vez que não é possível definir médias ou distâncias euclidianas para categorias sem transformações adicionais, o que pode aumentar a complexidade do modelo.

Ainda, segundo o autor, o *K-prototypes* é uma extensão poderosa que integra os algoritmos *K-means* e *K-modes*, sendo projetado especificamente para agrupar dados de tipo misto, numéricos e categóricos simultaneamente.

Conforme Valderrama e Flores (2021), no *K-prototypes* a dissimilaridade entre elementos com atributos mistos é calculada pela combinação da distância euclidiana ao quadrado para variáveis numéricas e da dissimilaridade de concordância simples para variáveis categóricas, na qual a distância é 0 quando os valores são iguais e 1 quando são diferentes. Essa soma é frequentemente ponderada por um fator ( $\gamma$ ) para equilibrar a influência entre os dois tipos de dados.

Ferreira (2015), explica que, diferentemente do *K-means*, o *K-prototypes* atualiza os centros dos *clusters* utilizando a média para variáveis numéricas e a moda para variáveis categóricas.

Valderrama e Flores (2021) detalham que o *K-prototypes* inicia com  $k$  protótipos, atribui os elementos aos *clusters* mais próximos, atualiza esses protótipos e repete o processo até a convergência das alocações.

Tanto no *K-means* quanto no *K-prototypes*, o número de *clusters* ( $k$ ) deve ser definido previamente, sendo o Método do Cotovelo uma técnica visual e quantitativa utilizada para estimar esse valor ideal. Brevemente, Crocco et al (2023), explicam que o Método do Cotovelo consiste em executar o algoritmo para diferentes valores de  $k$ , plotar a função de custo em relação ao número de *clusters* e identificar o ponto em que a redução do custo passa a ser pouco significativa, formando o chamado “cotovelo” no gráfico.

### 3.3.5 Aplicações de Ciência de Dados no setor fiscal público brasileiro

A literatura nacional recente tem consolidado um corpo de evidências que respalda o uso de ciência de dados e aprendizado de máquina na gestão tributária e, mais amplamente, na sustentabilidade fiscal do setor público. Em linhas gerais, esses estudos convergem para três frentes complementares:

a) previsão e planejamento de receitas em contextos sujeitos a choques macroeconômicos e heterogeneidade territorial;

b) qualificação de contribuintes e priorização de esforços de fiscalização e cobrança, por meio de modelos classificatórios e/ou segmentação baseada em comportamento de pagamento; e

c) modernização de cadastros e avaliação em massa para fins fiscais, com ênfase no IPTU, além de aplicações correlatas na detecção de anomalias do lado da despesa.

Tais frentes articulam-se diretamente com o objetivo desta dissertação, que é analisar a inadimplência do IPTU em São Luís-MA a partir de métodos de aprendizado de máquina, produzindo subsídios técnicos para decisões de arrecadação, comunicação e melhoria da justiça fiscal.

A literatura tem abordado modelos para apoiar a fiscalização e a recuperação de créditos. Alvarenga (2025), ao desenvolver classificadores para prever o pagamento de autos de infração na SEFAZ-DF, estrutura o problema com a metodologia CRISP-DM e testa regressão logística, árvores de decisão e *Random Forest*, com avaliação por f1-score, precisão e recall. Os resultados apontam a *Random Forest* como a melhor alternativa e destacam o papel central de variáveis financeiras (valor da autuação e indicadores de arrecadação) na discriminação dos casos pagos e não pagos. Tal evidência converge com a proposta desta dissertação ao mostrar que, além de modelar “se” haverá pagamento, é possível derivar instrumentos operacionais para segmentar a régua de cobrança, alocando recursos (por exemplo, o custo elevado de cartas) aos contribuintes com maior probabilidade de regularização e calibrando intervenções para perfis de maior resistência.

Ainda no campo da priorização de cobrança, Guimarães (2023) utiliza *XGBoost* para classificar a recuperabilidade de créditos de ICMS em dívida ativa, organizando os resultados como um *rating* de créditos. Entre as variáveis de maior contribuição destacam-se “quantidade de dias em débito”, “valor total da dívida” e “dias de empresa”, além de características cadastrais e regionais. O trabalho evidencia que, ao substituir julgamentos exclusivamente profissionais por modelos quantificados, é possível obter medidas mais objetivas de perda e de probabilidade de recuperação, com implicações para provisões contábeis e para a ordenação de filas de cobrança. A transposição para o IPTU municipal é direta: ratings por imóvel/débito podem orientar tanto o sequenciamento de comunicações quanto a

definição de metas realistas de recuperação por ciclo, mantendo rastreabilidade técnica das escolhas.

Para além do enfoque na propensão a pagar, a modernização cadastral e a avaliação em massa constituem uma frente indispensável para enfrentar inadimplência estrutural associada a percepções de iniquidade na base de cálculo. Dos Anjos e Pinheiro (2023), ao discutir inovações de inteligência artificial na fiscalização do IPTU, enfatizam que a adoção de ferramentas de aprendizado de máquina e mineração de dados pode contribuir desde o lançamento até a cobrança, com ganhos de eficiência e de justiça fiscal. O artigo resgata experiências de atualização cadastral e de uso de modelos automatizados de avaliação (AVM) que resultaram em aumentos expressivos de arrecadação logo após a modernização do cadastro e maior transparência na determinação de valores cadastrais.

Na mesma direção, Oliveira (2020) aplica regressão linear múltipla, *Random Forest* e *XGBoost* à avaliação em massa de terrenos urbanos em Fortaleza e encontra desempenho consistentemente superior do *XGBoost* em todas as métricas analisadas, culminando na proposição de uma Planta Genérica de Valores (PGV) georreferenciada. O autor destaca a dificuldade prática de cumprimento dos pressupostos do modelo linear clássico em contextos massivos e heterogêneos, reforçando a adequação de métodos de aprendizado de máquina para capturar não linearidades e interações complexas entre atributos físicos, locais e de vizinhança.

No domínio específico do IPTU e do comportamento de pagamento, Alves et al. (2023) propõem um modelo não supervisionado para qualificar contribuintes/imóveis a partir do histórico de pagamentos e estados do débito (aberto, extinto, imunidade, isenção, suspenso), incluindo marcadores operacionais (cota única, desconto, parcelamento, inscrição em dívida ativa) e contagens de comunicações (e-mail, carta, WhatsApp). Em experimentos com dados de Fortaleza, a segmentação em *clusters* interpretáveis, “bom pagador”, “mau pagador”, “possível bom pagador”, “possível mau pagador”, sustentou uma política de envio da primeira carta apenas a perfis com maior propensão a regularizar, com F1-Score de 92,30% e precisão efetiva compatível, além de estimativa de ganhos relevantes de arrecadação e redução de dispêndios com correspondências desnecessárias.

Embora o estudo adote uma abordagem conservadora (selecionando bons pagadores para a primeira notificação), a contribuição central é metodológica:

historizar o comportamento, qualificar perfis de risco e conectar tais perfis a decisões operacionais mensuráveis. Esta dissertação se beneficia diretamente dessa lógica ao propor um *pipeline* que combine *clustering* de comportamento com classificação de risco de inadimplência por exercício e, em seguida, experimente réguas de cobrança diferenciadas por perfil, integrando avaliação de custo por real recuperado.

Do ponto de vista do ISS, duas contribuições antecipam o uso de mineração de dados e classificação de irregulares com benefícios potenciais para a arrecadação. Piccirilli (2013) constrói modelos com árvore de decisão para classificar contribuintes com irregularidades no município de Goiânia, alcançando acurácias superiores a 90% quando combina atributos socioeconômicos e informações de auditoria cuidadosamente selecionadas para evitar vazamento de informação. O estudo é exemplar por mostrar que a curadoria das variáveis, mais do que o algoritmo em si, condiciona o desempenho e a utilidade prática do modelo para direcionar auditorias em contexto de recursos limitados.

Em perspectiva complementar, Franco et al. (2023) apresentam um comitê de algoritmos para geração de indícios de evasão no ISS em Fortaleza, com um traço distintivo: a capacidade de inferir custos e faturamento mesmo na ausência de dados completos, por meio de heurística de similaridade entre empresas de porte e localização comparáveis. A abordagem reconhece e qualifica *outliers* de comportamento, tendo identificado 1.839 empresas com fortes indícios de evasão num universo de 22.071, com validação por auditores e estimativa de perdas relevantes de receita. A analogia para o IPTU é imediata: quando dados cadastrais ou histórico de pagamento de determinados imóveis estão incompletos, técnicas de inferência por similaridade espacial e tipológica podem mitigar lacunas e manter a cobertura do modelo, sem renunciar à prudência na interpretação.

Embora o foco desta dissertação seja a receita tributária, convém observar que a mesma racionalidade de ciência de dados tem sido aplicada ao lado da despesa pública com bons resultados. Marcelo (2021) propõe parametrização de despesas municipais e detecção de anomalias com modelos preditivos estimados a partir de um painel amplo de municípios paulistas, reportando  $R^2$  em torno de 92%. Para além do indicador de ajuste, o autor adverte para um risco metodológico pertinente também ao lado da receita: a retroalimentação de vieses quando decisões futuras se baseiam, sem crítica, em classificações passadas e a necessidade de utilizar o modelo como instrumento de triagem e apoio, não como substituto da atividade de auditoria.

A transposição dessa cautela para a presente pesquisa é central: modelos que priorizam contribuintes para comunicação ou *enforcement* devem ser continuamente avaliados quanto a efeitos distributivos indesejados e quanto ao risco de reforçar padrões históricos de cobertura desigual entre bairros, usos ou faixas de valor.

Em síntese, os trabalhos revisados são consistentes em pelo menos quatro aspectos relevantes para o desenho desta dissertação. Primeiro, modelos não lineares e *ensembles* (*SVM*, *Random Forest*, *XGBoost*) tendem a superar abordagens lineares na presença de múltiplos determinantes e relações complexas, desde que acompanhados de validação rigorosa e controle de sobreajuste.

Segundo a engenharia de atributos, combinando histórico de pagamento, estados do débito, marcadores de comunicação e sinais comportamentais/digitais, é determinante para discriminar risco de inadimplência e propensão à regularização, abrindo espaço para réguas de cobrança baseadas em dados.

Terceiro, modernização cadastral e avaliação em massa com aprendizado de máquina contribuem para justiça fiscal e para a aceitabilidade social do IPTU, reduzindo a inadimplência “estrutural” associada a bases defasadas; o cruzamento entre risco de inadimplência e suspeita de defasagem cadastral pode, portanto, orientar estratégias distintas (comunicação versus revisão).

Quarto, modelos devem ser concebidos como instrumentos de apoio à decisão, com monitoramento contínuo de desempenho e de possíveis vieses, além de diálogo permanente com as áreas de negócio (arrecadação, cadastro, fiscalização) para que resultados preditivos se traduzam em ganhos operacionais observáveis.

Dessa base, a contribuição pretendida desta dissertação situa-se em integrar os aprendizados dos estudos resenhados num arcabouço aplicado a São Luís-MA, com dados volumosos e de alta dimensionalidade: (i) segmentar contribuintes/imóveis por padrões de pagamento, em linha com abordagens não supervisionadas; (ii) classificar risco de inadimplência por exercício com ensembles de árvores, priorizando explicabilidade e estabilidade temporal; (iii) estimar probabilidade de recuperação condicionada a diferentes políticas de comunicação, permitindo simulações de custo-efetividade; e (iv) dialogar com o cadastro e a avaliação em massa para separar casos de inadimplência comportamental daqueles cuja origem provável é a percepção de injustiça tributária.

Ao fazê-lo, pretende-se produzir um capítulo aplicado da literatura de ciência de dados em finanças públicas que, ao mesmo tempo, respeite especificidades locais, entregue métricas operacionais úteis e reforce princípios de equidade e transparência na gestão do IPTU.

### 3.4 Análise preditiva e justiça tributária

A inadimplência tributária representa um dos principais desafios para a administração pública municipal. Nesse contexto, a análise preditiva, alicerçada em técnicas de ciência de dados e aprendizado de máquina, tem se consolidado como um instrumento promissor para apoiar a identificação de riscos, a otimização de auditorias fiscais e a formulação de estratégias de arrecadação mais eficazes.

A análise preditiva é definida como o conjunto de técnicas estatísticas e computacionais que utilizam dados históricos para identificar padrões e antecipar resultados futuros (Mitchell, 1997 *apud* Ippolito e Lozano 2020). No campo tributário, essa abordagem possibilita prever quais contribuintes apresentam maior probabilidade de inadimplir, de fraudar ou de cometer irregularidades fiscais. Assim, a administração pode direcionar suas ações de fiscalização de forma mais assertiva, aumentando o impacto das auditorias e otimizando recursos.

Estudos nacionais e internacionais reforçam essa tendência. Ippolito e Lozano (2020) demonstraram que o emprego de algoritmos de aprendizado de máquina em auditorias fiscais permitiu prever crimes contra o sistema tributário de São Paulo, alcançando resultados robustos ao aplicar *Random Forests*. Os autores destacam que a sinergia do conjunto de árvores decisórias não apenas garantiu maior precisão, mas também proporcionou capacidade de adaptação a novos dados, assegurando validade dos modelos em diferentes períodos fiscais.

No contexto coreano, Lee e Kim (2024) aplicaram modelos avançados *de deep learning*, como CNN, para previsão de contribuintes individuais em atraso no pagamento de tributos locais. Os resultados indicaram acurácia superior a 87%, com destaque para a capacidade do modelo de não classificar equivocadamente contribuintes adimplentes como inadimplentes. Tal achado é relevante, pois evita prejuízos administrativos e fortalece a credibilidade do fisco junto à sociedade.

O setor financeiro acumula décadas de experiência no desenvolvimento de modelos de risco de crédito, os quais classificam clientes em perfis de maior ou menor propensão ao inadimplemento (Thomas e Edelman, 2017, *apud* Oliveira et al., 2021).

Esses modelos, tradicionalmente baseados em regressão logística, foram progressivamente aprimorados com a incorporação de algoritmos de aprendizado de máquina, como redes neurais, máquinas de vetores de suporte (SVM) e *Random Forests*, conforme apontam Maher, 2016 e Wu, 2015 *apud* Oliveira et al., 2021. A literatura recente aponta que tais modelos podem ser adaptados ao campo tributário, dada a similaridade conceitual entre prever o risco de inadimplência financeira e prever o risco de inadimplência fiscal. Oliveira et al. (2021) validaram essa hipótese ao propor a “*Full Discretization Methodology*” para preparar dados de empresas registradas no Distrito Federal. Com base em informações socioeconômicas e fiscais, os autores alcançaram resultados expressivos, com destaque para o desempenho de redes neurais artificiais (AUC = 0,9568).

A adaptação, no entanto, exige a incorporação de variáveis específicas da esfera tributária, como histórico de recolhimento, regime de tributação, natureza da atividade econômica e valores de débitos em aberto. Nesse sentido, Ippolito e Lozano (2020) destacam que crimes fiscais apresentam características próprias, como a recusa em fornecer documentos ou a omissão de informações em declarações, o que diferencia os modelos de risco de crédito tradicionais. Ao mesmo tempo, a convergência entre as áreas é inevitável e promissora. Assim como os bancos utilizam *scores* de crédito para definir limites e taxas de juros, a administração tributária pode se beneficiar de *scores* de risco de inadimplência para priorizar fiscalizações, renegociações ou políticas de estímulo à conformidade voluntária.

Adicionalmente, a adoção de políticas públicas sustentadas por evidências empíricas e análises preditivas vem se consolidando em diferentes países. No caso brasileiro, experiências como a de São Paulo demonstram que auditorias baseadas em risco, sustentadas por algoritmos de aprendizado de máquina, elevaram em 15% a arrecadação do ISS apenas pelo aumento da percepção de risco entre contribuintes, consoante pesquisa realizada por Ippolito e Lozano (2020).

No Chile e na Espanha, governos têm utilizado redes neurais para previsão de fraudes fiscais, alcançando resultados significativos na detecção de anomalias (González e Velásquez, 2012; López et al., 2019 *apud* Ippolito e Lozano, 2020). Embora os contextos legais sejam distintos, essas experiências confirmam que modelos preditivos podem ser aplicados a diferentes sistemas tributários, desde que devidamente adaptados às especificidades locais.

No âmbito global, Pamisetty (2024) defende o desenvolvimento de sistemas integrados de apoio à decisão, baseados em *big data* e computação em nuvem, para fortalecer a gestão de tributos e de propriedades não reclamadas. A proposta integra técnicas de detecção de anomalias, visualização de riscos em mapas e ferramentas de análise acessíveis até mesmo a usuários não técnicos, ampliando a transparência e a efetividade da gestão fiscal.

Nos Estados Unidos, Kokogho et al. (2021) descrevem uma transformação conceitual impulsionada por inteligência artificial, envolvendo *machine learning*, NLP, RPA e *blockchain*. Essas inovações não apenas fortalecem a fiscalização e a detecção de fraudes, mas também aprimoram a interação com os contribuintes, tornando os processos mais transparentes e menos custosos.

A principal expectativa relacionada à adoção de análise preditiva em administrações tributárias municipais é o incremento da eficiência arrecadatória. Ao direcionar esforços de fiscalização para contribuintes com maior risco de inadimplência, reduz-se o custo de auditorias aleatórias e aumenta-se o retorno das ações fiscais (Oliveira et al., 2021). Experiências empíricas demonstram que a simples comunicação preventiva aos contribuintes, alertando sobre a possibilidade de fiscalização, já é capaz de elevar os índices de conformidade voluntária (Alm, 2019 *apud* Ippolito e Lozano, 2020).

Outro benefício relevante refere-se à formulação de políticas fiscais mais justas. Modelos preditivos permitem identificar padrões de comportamento que podem estar associados a dificuldades econômicas específicas de determinados setores ou regiões. Com base nesses insights, governos podem propor programas de parcelamento, incentivos fiscais ou ajustes normativos que busquem equilibrar a capacidade contributiva com a necessidade de arrecadação (Lee e Kim, 2024).

Além disso, a disponibilidade de análises baseadas em dados reforça a transparência das decisões fiscais, reduzindo o espaço para arbitrariedades e aumentando a legitimidade da ação estatal.

Por fim, a incorporação da análise preditiva à gestão fiscal contribui para a sustentabilidade financeira dos municípios. Como destacado por Ippolito e Lozano (2020), em São Paulo, os tributos municipais representam mais da metade da receita da cidade, sendo o ISS e o IPTU suas principais fontes de arrecadação. A redução da inadimplência, portanto, impacta diretamente na capacidade de investimento em áreas essenciais como saúde, educação e infraestrutura.

Além do aspecto imediato de aumento de receita, a utilização de modelos preditivos fortalece a governança fiscal no longo prazo. Pamisetty (2024) argumenta que sistemas inteligentes de apoio à decisão ampliam a resiliência financeira, ao permitirem uma gestão mais dinâmica frente a contextos de crise econômica, mudanças legislativas ou transformações sociais.

Portanto, o combate à inadimplência tributária exige novas abordagens, capazes de superar os limites da fiscalização tradicional. A análise preditiva, ao integrar técnicas de aprendizado de máquina, big data e sistemas inteligentes de apoio à decisão, apresenta-se como um caminho promissor para modernizar a gestão fiscal municipal.

A adaptação de modelos de risco de crédito ao contexto tributário, como demonstrado em Brasília por Oliveira et al. (2021), bem como as experiências de São Paulo (Ippolito e Lozano, 2020), da Coreia do Sul (Lee e Kim, 2024) e de outros países, reforçam a validade e a aplicabilidade desse paradigma.

Os benefícios esperados incluem não apenas o aumento da eficiência arrecadatória, mas também a promoção de maior justiça fiscal e a consolidação da sustentabilidade financeira dos municípios. Nesse sentido, a adoção da análise preditiva deve ser compreendida como parte de uma agenda mais ampla de transformação digital e de fortalecimento da governança pública, alinhada às demandas contemporâneas por eficiência, transparência e equidade.

### **3.5 Administração Tributária Municipal, tecnologias e justiça fiscal**

A análise da inadimplência do Imposto Predial e Territorial Urbano (IPTU) exige uma abordagem multidimensional, que articule fundamentos das áreas de finanças públicas, ciência de dados, economia urbana e justiça fiscal. Esta pesquisa se ancora em três eixos teóricos principais: a gestão tributária municipal, o uso de tecnologias analíticas e algoritmos na administração pública e o princípio da equidade na tributação urbana.

No campo da gestão tributária municipal, o IPTU é regulamentado pelo Código Tributário Nacional e instituído constitucionalmente como um tributo de competência dos municípios, com base de cálculo no valor venal do imóvel. Segundo Varsano (1977) *apud* Costa, Pecini e Fukumi Tsunoda (2021), os impostos sobre propriedades imobiliárias são utilizados há séculos para financiar as atividades públicas, sendo indicados por especialistas de finanças públicas como uma das

melhores fontes de financiamento dos municípios, pois apresentam três vantagens principais: (i) não geram conflitos entre entes federativos; (ii) o contribuinte só pode evitar o pagamento ao transferir a propriedade; e (iii) o valor dos imóveis é relativamente estável, o que confere ao imposto arrecadação menos suscetível às oscilações econômicas. Esta última característica é também ressaltada por De Cesare *et al.* (2020) *apud* Costa, Pecini e Fukumi Tsunoda (2021), ao defender que, mesmo em contextos de crise como a pandemia de Covid-19, o IPTU se mantém como fonte relevante e constante de financiamento dos serviços públicos essenciais.

No entanto, a eficácia do IPTU tem sido comprometida por fatores como subavaliação de imóveis, cadastros desatualizados, dificuldades na fiscalização e inadimplência elevada (Afonso, Araújo, Nóbrega, 2009; Carvalho Júnior, 2016, *apud* Costa, Pecini e Fukumi Tsunoda, 2021). Estudos como o de Bernardes (2020), ao analisar a inadimplência do IPTU em Fortaleza, revelam que imóveis localizados em regiões de menor valor venal são os que mais concentram inadimplência, sugerindo um padrão socioeconômico associado ao comportamento fiscal.

O segundo eixo refere-se à aplicação de técnicas de ciência de dados e inteligência artificial no setor público, especialmente em contextos urbanos e fiscais. A evolução da tecnologia e o crescimento exponencial de bases de dados disponíveis nas administrações municipais possibilitaram a adoção de métodos avançados de análise preditiva e detecção de padrões, como aprendizado de máquina e mineração de dados.

Conforme exposto por Wang, Hu e Zhu (2009) *apud* Costa, Pecini e Fukumi Tsunoda (2021), a mineração de dados no setor público começou como ferramenta para detectar fraudes e desperdícios, mas atualmente tem se expandido para a avaliação e melhoria do desempenho de programas e políticas públicas. Bach (2003) *apud* Costa, Pecini e Fukumi Tsunoda (2021) identifica as áreas do setor público que mais aplicam técnicas de mineração de dados, como finanças e economia, saúde, justiça criminal e segurança, trabalho e bem-estar social, educação e transporte. Ainda segundo Bach (2003), nessas áreas, especialmente nas de finanças, os modelos preditivos têm sido aplicados para estimar arrecadação, identificar padrões de inadimplência e avaliar a eficiência de programas públicos.

Além disso, constata-se que a análise espacial e georreferenciada é muito útil na compreensão da inadimplência, como mostram os trabalhos de Bernardes (2020) e Rosa *et al.* (2023). A geolocalização dos imóveis inadimplentes e sua

associação com infraestrutura urbana, renda média e presença de serviços públicos pode revelar desigualdades territoriais que ajudam a compreender a relação entre estrutura urbana e arrecadação tributária.

O terceiro eixo teórico está ancorado no princípio da justiça fiscal urbana, que exige uma distribuição equitativa da carga tributária entre os contribuintes e a garantia de que os tributos sejam cobrados com base em critérios objetivos e transparentes.

A tributação ineficiente, baseada em cadastros defasados ou valores venais incoerentes, tende a aprofundar desigualdades socioespaciais. A justiça fiscal, nesse contexto, não se limita ao aspecto normativo, mas demanda instrumentos técnicos capazes de identificar distorções e propor ajustes no sistema de arrecadação.

Esses fundamentos sustentam diretamente os objetivos e a metodologia da presente pesquisa. A articulação entre conceitos clássicos da administração tributária e métodos contemporâneos de análise de dados permite abordar o problema da inadimplência de forma inovadora e eficaz.

A proposta não apenas busca compreender os fatores associados à inadimplência do IPTU no município de São Luís, mas também propor soluções técnicas, sustentáveis e replicáveis, que contribuam para a justiça fiscal, a eficiência arrecadatória e o fortalecimento da gestão pública municipal.

#### **4 MATERIAIS E MÉTODOS**

A presente pesquisa tem como propósito identificar e analisar os fatores determinantes para a inadimplência do Imposto Predial e Territorial Urbano (IPTU) no município de São Luís-MA. Com isso, pretende-se fornecer subsídios técnicos à gestão fiscal municipal para o aprimoramento das estratégias de arrecadação. Para alcançar tal objetivo, serão utilizadas técnicas de ciência de dados e algoritmos de aprendizado de máquina aplicados a uma base de dados históricos da Secretaria Municipal da Fazenda de São Luís (SEMFAZ-SL).

A abordagem metodológica adotada foi quantitativa, uma vez que a análise se dará com base em dados estruturados extraídos de registros administrativos. O estudo tem natureza aplicada, pois busca resolver um problema concreto da administração pública, e tem caráter exploratório e descritivo. A pesquisa é classificada, quanto aos procedimentos técnicos, como documental, tendo em vista

que utilizou dados oficiais da administração tributária municipal. Foram também incorporadas características descritivas, ao buscar compreender e caracterizar os diferentes perfis de inadimplência observados nos dados.

A Secretaria Municipal da Fazenda (SEMFAZ) disponibilizou arquivos estruturados relativo a:

a) informações cadastrais dos imóveis, situação cadastral referente ao mês de abril de 2025;

b) informações financeiras dos imóveis, relativas ao lançamento de IPTU, por ano, no período de 2019 a 2023.

A base de dados referente às informações cadastrais do imóvel apresenta as seguintes colunas, que englobam, entre outros, dados construtivos, zoneamento e infraestrutura urbana:

1. INSCRICAOFICT: número fictício da inscrição imobiliária, anonimizado para respeitar o sigilo fiscal e a Lei Geral de Proteção de Dados (LGPD).

2. USOIMOVEL: finalidade de uso do imóvel (predial ou territorial).

3. BAIRRO: bairro de localização do imóvel.

4. CONDOMINIO: informação binária (sim/não) sobre se o imóvel pertence a condomínio.

5. AREA\_TOT\_EDIF: área total edificada do imóvel, em metros quadrados.

6. AREA\_TERRENO: área do terreno, em metros quadrados.

7. SETOR: setor fiscal vinculado ao imóvel.

8. QTD\_CPFCNPJFICT: quantidade de CPFs/CNPJs associados ao imóvel.

9. CPFCNPJ\_VALIDO: se há CPF ou CNPJ válido, considerando o método de cálculo dos dígitos verificadores

10. CONSERVACAO\_VIA: informação binária sobre conservação da via (tem/não tem).

11. LIMPEZA\_PUB: existência ou não de serviço de limpeza pública na via.

12. TRANSPORTE: presença ou não de transporte público (direto, próximo ou não tem).

13. COLETA\_LIXO: tipo de coleta de lixo na região (alternada, diária ou não tem).

14. ARBORIZACAO: presença e tipo de arborização urbana (não tem, uma, até três árvores ou mais de três).

15. MEIO\_FIO: presença de meio-fio na via (tem ou não tem).

16. PAVIMENTACAO: tipo de pavimentação da via, por exemplo, asfalto, lajota, paralelepípedo etc.

17. ILUMINACAO\_PUB: existência de iluminação pública (tem ou não).

18. DRENAGEM: presença de rede de drenagem urbana (tem ou não).

19. CALCADA: presença de calçada no logradouro (tem ou não).

20. REDE\_TEL: tipo de rede de telefonia disponível (aérea, subterrânea ou não tem).

21. REDE\_ELE: tipo de rede elétrica disponível, por exemplo, baixa, alta, irregular etc.

22. REDE\_ESG: presença de rede de esgoto (tem ou não).

23. REDE\_AGUA: presença de rede de abastecimento de água (tem ou não).

24. DESTINACAO: destinação de uso do imóvel, por exemplo, residencial, comercial, institucional, religioso etc.

25. CONSERVACAO: estado de conservação do imóvel (boa, ótima, precária e regular).

26. TIPOLOGIA: tipo de construção, por exemplo, casa, depósito, apartamento, sobrado etc.

27. ESTRUTURA: tipo de estrutura construtiva predominante, por exemplo, alvenaria, concreto, madeira, metálica etc.

28. PADRAO: padrão de acabamento do imóvel (baixo, médio, alto).

Já as bases de dados relativas às informações financeiras contêm:

1. INSCRICAOFICT: número fictício da inscrição imobiliária, anonimizado para respeitar o sigilo fiscal e a Lei Geral de Proteção de Dados (LGPD).

2. ANO: ano de lançamento do IPTU.

3. VALOR\_IPTU: valor do imposto para o imóvel.

4. DATA\_PAG: data de pagamento do imposto, quando realizado.

5. VALOR\_PAGO: valor efetivamente pago do IPTU.

6. TEM\_BNF: informação binária (sim/não) indicando se o imóvel possui benefício fiscal.

Na etapa inicial, foi realizado o pré-processamento dos dados detalhado nos itens 4.1 a 4.3. Em seguida, foi conduzida uma análise exploratória inicial dos dados, buscando depreender informações preliminares acerca das características dos imóveis, da localização e da infraestrutura e serviços urbanos. Foram calculadas estatísticas descritivas de variáveis financeiras, cadastrais e territoriais, explanadas no item 5.1. Estas etapas foram realizadas com auxílio de planilhas eletrônicas.

Posteriormente, de forma a promover uma análise mais profunda, foram aplicados algoritmos de aprendizagem supervisionada e não supervisionada. Estas etapas estão detalhadas no item 5.2 a 5.4.

Os algoritmos foram implementados com o auxílio de ferramentas computacionais, especialmente a linguagem *Python*. Os experimentos foram desenvolvidos no ambiente *Google Colab*. Para o algoritmo supervisionado de classificação, foi utilizada a biblioteca *XGBoost*, em função de sua eficiência e bom desempenho em dados tabulares. Já a clusterização foi realizada por meio do algoritmo *K-prototypes*, implementado com a biblioteca *kmodes*, apropriada para bases de dados que combinam variáveis numéricas e categóricas. Adicionalmente, foram empregadas bibliotecas consolidadas do ecossistema *Python*, como *pandas* e *NumPy* para manipulação e tratamento dos dados, e *scikit-learn* para pré-processamento, validação cruzada e avaliação dos modelos, assegurando consistência metodológica e confiabilidade nos resultados obtidos.

#### **4.1 Pré-processamento de dados para Diagnóstico do IPTU**

De acordo com informações da SEMFAZ, em abril de 2025, havia 384.564 inscrições imobiliárias cadastradas. A base de dados utilizada neste estudo, entretanto, inicialmente, compreende 383.507 inscrições imobiliárias, devido à exclusão de registros com inconsistências de estrutura e alinhamento entre variáveis na etapa de extração, inviabilizando seu aproveitamento analítico.

Considerando como universo de análise as 383.507 inscrições imobiliárias da base cadastral, foram identificados, nas bases financeiras disponibilizadas, os lançamentos de IPTU correspondentes aos anos de 2019 a 2023. Nesse recorte, registraram-se 378.915 lançamentos em 2023; 377.202 em 2022; 373.293 em 2021; 369.762 em 2020; e 368.154 em 2019.

Considerando a importância da coluna “VALOR\_IPTU” na base financeira, verificou-se que algumas não possuíam qualquer valor ou valor zero. A tabela 1 a seguir apresenta um resumo dessas quantidades:

**Tabela 1:** Detalhamento da quantidade de lançamentos de IPTU

Ano	Quantidade de lançamentos de IPTU (A) = (B) + (C)	Quantidade de lançamentos de IPTU sem benefício fiscal (B)	Quantidade de lançamentos de IPTU com benefício fiscal (C)	Total de registros com “VALOR_IPTU” zero ou ausente (D)	Quantidade de lançamentos de IPTU com “VALOR_IPTU” > 0 (E) = (A) – (D)
2019	368.154	315.738	52.416	4.222	363.932
2020	369.762	317.348	52.414	3.372	366.390
2021	373.293	321.029	52.264	4.795	368.498
2022	377.202	328.300	48.902	7.578	369.624
2023	378.915	330.242	48.673	7.492	371.423

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Portanto, as análises que serão realizadas no item 5.1.1 terão como base os quantitativos apresentados na última coluna (E) da tabela 1, ou seja, somente os imóveis que tiveram IPTU lançado maior do que zero em cada ano.

É importante destacar que este estudo se concentra nos imóveis com valor de IPTU lançado maior do que zero, pois são os contribuintes com maior potencial de arrecadação, bem como os que estão diretamente associados ao problema da inadimplência. Não obstante, ressalta-se que os registros ausentes ou zerados representam potenciais anomalias cadastrais, como será discutido no item 5.4.

A análise do item 5.1.2, em sequência, removerá dos dados citados no parágrafo anterior, os imóveis que possuem benefício fiscal. Portanto, para o item 5.1.2, a análise se concentrará apenas nos imóveis sem benefício fiscal.

#### **4.2 Pré-processamento de dados para Modelo de Classificação Supervisionada**

Para a construção do modelo classificador multiclasse, adotaram-se como base os dados financeiros referentes ao ano de 2023, por se tratar do exercício mais recente disponível e, portanto, mais próximo da realidade base cadastral dos imóveis.

Conforme apresentado no item anterior, no ano de 2023 havia 371.423 inscrições imobiliárias com lançamento de IPTU superior a zero. Agora, tomando esse

conjunto como universo de análise, foi elaborada uma nova base de dados a partir da reorganização das variáveis referentes a 2023, com a padronização das colunas VALOR\_IPTU2023, DATA\_PAG2023, VALOR\_PAG2023 e TEM\_BNF2023, mantendo-se a variável identificadora INSCRICAOFICT. Adicionalmente, foram incorporadas à base as informações históricas dos exercícios anteriores, compreendendo:

- a) VALOR\_IPTU2022 a VALOR\_IPTU2019;
- b) DATA\_PAG2022 a DATA\_PAG2019;
- c) VALOR\_PAG2022 a VALOR\_PAG2019;
- d) TEM\_BNF2022 a TEM\_BNF2019.

As inscrições imobiliárias referentes aos exercícios de 2019 a 2022 que não apresentaram correspondência com as inscrições presentes na base financeira de 2023 foram mantidas como valores ausentes.

A partir dessa base consolidada, foram construídas as seguintes variáveis derivadas:

a) qtd iptu: número de ocorrências (células) em que os valores do conjunto VALOR\_IPTU2023 a VALOR\_IPTU2019 são superiores a zero, assumindo valor mínimo igual a 1, uma vez que 2023 foi adotado como ano-base, e valor máximo igual a 5;

b) qtd pag: número de ocorrências (células) em que os valores do conjunto VALOR\_PAG2023 a VALOR\_PAG2019 são superiores a zero, com limite mínimo de 0 e máximo de 5;

c) qtd\_bnf\_SIM: número de ocorrências (células) em que o conjunto TEM\_BNF2023 a TEM\_BNF2019 apresenta valor igual a "SIM", com limite mínimo de 0 e máximo de 5;

d) adimplência: variável categórica definida a partir da relação entre qtd iptu e qtd pag, classificada em:

- (i) Adimplente: quando  $qtd\_iptu = qtd\_pag$ ;
- (ii) Adimplente parcial: quando  $qtd\_pag > 0$  e  $qtd\_pag \neq qtd\_iptu$ ;
- (iii) Inadimplente: quando  $qtd\_pag = 0$

e) sem\_bnf\_all: variável binária indicativa da existência de benefícios fiscais no período analisado, classificada em:

- (i) Com BNF: quando  $qtd\_bnf\_SIM > 0$ ;
- (ii) Sem BNF: quando  $qtd\_bnf\_SIM = 0$ .

É importante destacar que, em razão da baixa proporção da quantidade de pagamentos em relação ao total de lançamentos de IPTU, bem como do fato de que, na coluna VALOR\_PAG, algumas inscrições apresentarem valores pagos inferiores ao valor originalmente lançado, em geral decorrentes de descontos concedidos para pagamento, adotou-se, do ponto de vista metodológico, que a existência de qualquer pagamento caracteriza o contribuinte como adimplente. Ressalta-se que, nesses casos, a simples observação de um valor pago inferior ao valor lançado não permite inferir inadimplência, uma vez que a diferença pode decorrer de descontos legalmente previstos, tais como abatimentos por pagamento em cota única. Assim, embora o valor registrado como pago seja inferior ao valor lançado, assume-se que, em geral, o contribuinte não possui débito de IPTU, tendo realizado o pagamento de forma regular, conforme a legislação vigente.

Dessa forma, considerando especialmente que a classe de adimplentes é substancialmente menor do que a de inadimplentes, e tendo em vista que a criação de novas categorias de classificação não poderia ser realizada com segurança nem contribuiria de forma relevante para o objetivo central do estudo, que é a análise do comportamento dos contribuintes inadimplentes, optou-se por considerar como adimplente, no ano de referência, todo contribuinte que tenha efetuado qualquer pagamento, independentemente da diferença entre os valores lançados e pagos. Como será demonstrado nas análises subsequentes, mesmo sob essa definição favorável à classificação dos adimplentes, a classe de contribuintes inadimplentes permanece consideravelmente majoritária, o que reforçou a decisão metodológica tomada. Portanto, privilegia-se neste estudo a análise da inadimplência total, entendida como a ausência completa de pagamento no exercício de lançamento do IPTU, por refletir com maior precisão o núcleo do problema enfrentado pelo município.

Para estudos futuros, sugere-se a incorporação de informações adicionais à base de dados, como a variável valor do débito, a fim de permitir a verificação precisa se o contribuinte realizou integralmente o pagamento do IPTU no período considerado, distinguindo de forma inequívoca situações de pagamento com desconto legal de casos de pagamento parcial ou efetiva inadimplência.

Continuando as análises do pré-processamento, foram excluídas 60.821 inscrições imobiliárias classificadas como “Com BNF” na variável *sem\_bnf\_all*, mantendo-se apenas aquelas que não apresentaram benefícios fiscais no período de 2019 a 2022. Após essa etapa, a base passou a conter 310.602 inscrições imobiliárias.

Em seguida, foi construída uma nova base de dados por meio da integração, via chave INSCRICAOFICT, das informações cadastrais e financeiras referentes a essas 310.602 inscrições imobiliárias. Do conjunto financeiro, foram incorporadas exclusivamente as variáveis ADIMPLENCIA e VALOR\_IPTU2023.

Nessa base integrada, quando a variável USOIMOVEL assumia o valor “TERRITORIAL”, foi atribuída a classe “TERRITORIAL” às variáveis DESTINACAO, CONSERVACAO, TIPOLOGIA, ESTRUTURA e PADRAO, originalmente ausentes nesses casos, uma vez que tais características se aplicam apenas a imóveis prediais, isto é, imóveis edificadas.

Após as etapas anteriores, foram excluídas 1.300 observações, em razão dos seguintes critérios:

a) Ausência de informações de infraestrutura e serviços urbanos, incluindo as variáveis CONSERVACAO\_VIA, LIMPEZA\_PUB, TRANSPORTE, COLETA\_LIXO, ARBORIZACAO, MEIO\_FIO, PAVIMENTACAO, ILUMINACAO\_PUB, DRENAGEM, CALCADA, REDE\_TEL, REDE\_ELE, REDE\_ESG e REDE\_AGUA, correspondendo a 1.122 inscrições;

b) Ausência de informação de DESTINACAO em registros classificados como imóveis prediais (USOIMOVEL ≠ TERRITORIAL), correspondendo a 4 registros;

c) AREA\_TERRENO igual a zero, em 34 registros;

d) AREA\_TERRENO ausente, em 140 registros.

Optou-se pela exclusão das observações com valores ausentes ou inconsistentes, em detrimento da imputação de dados, uma vez que tais casos representam uma fração reduzida do conjunto total e sua remoção simplifica o processo de modelagem, sem comprometer a representatividade da amostra. Ademais, esses registros podem ser interpretados como anomalias cadastrais, aspecto que será explorado no item 5.4.

Após a exclusão das 1.300 observações, a base final utilizada na etapa de modelagem passou a conter 309.302 registros e 30 variáveis, a saber: INSCRICAOFICT, USOIMOVEL, CONDOMINIO, AREA\_TOT\_EDIF, AREA\_TERRENO, BAIRRO, SETOR, CPFCNPJ\_VALIDO, QTD\_CPF\_CNPJ, CONSERVACAO\_VIA, LIMPEZA\_PUB, TRANSPORTE, COLETA\_LIXO, ARBORIZACAO, MEIO\_FIO, PAVIMENTACAO, ILUMINACAO\_PUB, DRENAGEM, CALCADA, REDE\_TEL, REDE\_ELE, REDE\_ESG, REDE\_AGUA, DESTINACAO,

CONSERVACAO, TIPOLOGIA, ESTRUTURA, PADRAO, VALOR\_IPTU2023 e ADIMPLENCIA.

Na etapa de modelagem, o atributo BAIRRO foi excluído da base de dados por apresentar elevada cardinalidade, com mais de 200 categorias distintas. Embora a codificação *one-hot* seja aplicada às variáveis categóricas no *pipeline* de modelagem adotado, a inclusão dessa variável resultaria na criação de um número excessivo de colunas adicionais, tornando o processo computacionalmente oneroso e aumentando significativamente o tempo de treinamento do modelo *XGBoost*. Ademais, a expansão excessiva da dimensionalidade poderia comprometer a eficiência do algoritmo, sem necessariamente gerar ganhos proporcionais em desempenho preditivo, sobretudo diante da presença de outras variáveis estruturais e territoriais já consideradas na análise. Optou-se por deixar a variável setor fiscal, como representativa da localização geográfica do imóvel.

Adicionalmente, com o objetivo de otimizar a eficiência computacional e garantir maior estabilidade na execução do algoritmo, foram realizadas as seguintes transformações adicionais nas variáveis categóricas:

a) A variável USOIMOVEL foi recodificada para valores binários, assumindo 1 para imóveis prediais e 0 para imóveis territoriais;

b) As variáveis originalmente representadas por categorias do tipo “tem/não tem” ou “sim/não” foram convertidas para valores binários 1/0, incluindo CPFCNPJ\_VALIDO, CONDOMINIO, CONSERVACAO\_VIA, LIMPEZA\_PUB, MEIO\_FIO, ILUMINACAO\_PUB, DRENAGEM, CALCADA, REDE\_ESG e REDE\_AGUA.

c) Na variável DESTINACAO, as classes de baixa frequência, PREST SERVICOS, RELIGIOSO, SERV PUB COMUNITARIO, ENSINO, INDUSTRIAL, INSTITUCIONAL, FUNDACAO e OUTROS, foram agregadas em uma nova categoria denominada “OUTRA”. Em conjunto, essas classes totalizavam 10.033 registros, correspondendo a aproximadamente 3,2% do total de observações. Dessa forma, a variável DESTINACAO passou a apresentar apenas quatro categorias: COMERCIAL, RESIDENCIAL, TERRITORIAL e OUTRA.

Essas transformações visaram reduzir a dimensionalidade da base, facilitar o processamento dos dados e assegurar a compatibilidade com o pipeline de modelagem adotado.

Por conseguinte, essa base de dados será utilizada para a construção de um modelo supervisionado de classificação multiclasse (adimplente, adimplente parcial e inadimplente), visando tanto prever a classe de adimplência dos imóveis quanto a analisar os fatores que contribuem para a atribuição de cada classe.

### 4.3 Pré-processamento de dados para Clusterização

Para a modelagem do algoritmo de clusterização, partiu-se da base de dados utilizada na etapa de modelagem supervisionada, sobre a qual foram realizadas as seguintes modificações e exclusões:

a) Foram excluídas 94.437 observações correspondentes às classes adimplente e adimplente parcial da variável ADIMPLÊNCIA, mantendo-se apenas as 214.865 inscrições imobiliárias classificadas como inadimplentes. Em razão da homogeneidade da variável-alvo após esse filtro, a coluna ADIMPLENCIA foi excluída da base;

b) Foram excluídas as variáveis AREA\_TOT\_EDIF, AREA\_TERRENO, SETOR, QTD\_CPF\_CNPJ, CONSERVACAO, TIPOLOGIA, ESTRUTURA e PADRAO, uma vez que o objetivo da clusterização é identificar perfis homogêneos de inadimplência a partir de características passíveis de subsidiar políticas fiscais e urbanas. Essas variáveis, embora relevantes para a modelagem preditiva, não contribuem diretamente para a segmentação estratégica pretendida ou introduzem ruído redundância no processo de agrupamento;

c) As variáveis relacionadas à infraestrutura urbana e aos serviços públicos, CONSERVACAO\_VIA, LIMPEZA\_PUB, TRANSPORTE, COLETA\_LIXO, ARBORIZACAO, MEIO\_FIO, PAVIMENTACAO, ILUMINACAO\_PUB, DRENAGEM, CALCADA, REDE\_TEL, REDE\_ELE, REDE\_ESG e REDE\_AGUA, foram consolidadas em uma única variável sintética, denominada PT\_INFRA\_SERV, detalhada mais à frente;

d) Ao final dessas etapas, a base de dados destinada à clusterização passou a conter 214.865 observações e 7 variáveis, a saber: INSCRICAOFICT, USOIMOVEL, CONDOMINIO, CPFCNPJ\_VALIDO, DESTINACAO, VALOR\_IPTU2023 e PT\_INFRA\_SERV.

A seleção dessas seis variáveis explicativas, além da variável identificadora INSCRICAOFICT, fundamenta-se em critérios fiscais, administrativos, urbanos e sociais, considerando sua potencial contribuição para a interpretação dos padrões de

inadimplência e para a segmentação de estratégias de gestão tributária. A seguir, apresentam-se as justificativas associadas a cada variável:

a) USOIMOVEL: sob a perspectiva da administração tributária, imóveis prediais tendem, em geral, a apresentar maior facilidade de identificação e localização de seus responsáveis tributários, uma vez que se encontram edificados e ocupados. Em contraste, imóveis territoriais, por não possuírem edificação, podem demandar maior esforço administrativo para fins de fiscalização e cobrança. Ademais, os imóveis territoriais assumem relevância estratégica no contexto das políticas urbanas, por se relacionarem diretamente aos instrumentos previstos no § 4º do art. 182 da Constituição Federal de 1988 e no art. 359 do Código Tributário do Município de São Luís, que tratam do cumprimento da função social da propriedade urbana, incluindo o IPTU Progressivo no Tempo e a exigência de adequado aproveitamento do solo urbano não edificado, subutilizado ou não utilizado.

b) CONDOMINIO: do ponto de vista operacional, imóveis situados em condomínios podem apresentar maior viabilidade administrativa para ações de fiscalização, em razão da concentração espacial das unidades e da possibilidade de interlocução institucional com representantes legais, como síndicos. Além disso, tais imóveis, em termos gerais, tendem a dispor de infraestrutura urbana consolidada e a arcar regularmente com despesas condominiais, o que pode sinalizar maior capacidade contributiva, ainda que essa relação deva ser interpretada com cautela.

c) CPFENPJ\_VALIDO: a presença de informações cadastrais válidas relativas ao CPF ou CNPJ do responsável tributário é um elemento relevante para a efetividade das estratégias de cobrança, pois possibilita a adoção de instrumentos administrativos e extrajudiciais, como notificações formais, envio de comunicações e protesto em cartório, especialmente em casos de inadimplência recorrente.

d) DESTINACAO: a destinação do imóvel constitui uma variável relevante para fins de segmentação, uma vez que imóveis com uso comercial ou não residencial podem, em determinados contextos, apresentar dinâmicas econômicas mais favoráveis das unidades exclusivamente residenciais ou territoriais, o que justifica sua consideração na definição de estratégias diferenciadas de acompanhamento e cobrança.

e) VALOR\_IPTU2023: o valor lançado de IPTU representa um indicador direto da relevância fiscal do imóvel. Em termos de gestão tributária, valores mais elevados concentrados em um número reduzido de inscrições podem indicar maior

potencial arrecadatório, tornando essas unidades prioritárias do ponto de vista do custo-benefício das ações de fiscalização e cobrança, sem prejuízo da observância dos princípios da equidade e da justiça fiscal.

f) PT\_INFRA\_SERV: essa variável sintetiza informações relativas à infraestrutura urbana e aos serviços públicos disponíveis no entorno do imóvel, incorporando uma dimensão urbanística e social à análise. Em termos analíticos, imóveis com maior pontuação nessa variável indicam maior provisão de bens e serviços públicos pelo município, o que, sob a ótica da justiça urbana e fiscal, pode estar associado a uma maior expectativa de cumprimento das obrigações tributárias, ainda que tal relação não seja presumida de forma determinística.

Destaca-se que foram realizados testes exploratórios de clusterização incluindo as variáveis SETOR e AREA\_TERRENO; contudo, esses experimentos não produziram agrupamentos interpretáveis ou úteis para a formulação de políticas fiscais e urbanas, razão pela qual tais variáveis não foram mantidas na configuração final do modelo.

Por fim, descreve-se o processo de construção da variável PT\_INFRA\_SERV. A Planta Genérica de Valores (PGV) do Município de São Luís estabelece uma tabela de Pesos e Parâmetros para Infraestrutura e Serviços, utilizada como fator de correção a partir dos itens constantes no Cadastro de Logradouro, detalhada na tabela 2 a seguir.

**Tabela 2:** Pesos e Parâmetros para Infraestrutura e Serviços (PGV-SL)

ORD.	DESCRIÇÃO	PONTUAÇÃO
01	Conservação de Vias	03
02	Limpeza Pública	06
03	Estacionamento	02
04	Transporte Coletivo	
04.1	Direto	05
04.2	Indireto	05
05	Coleta de Lixo	
05.1	Diária	04
05.2	Alternada	03
06	Arborização	03
07	Sarjeta / Meio-fio	05
08	Pavimentação	

ORD.	DESCRIÇÃO	PONTUAÇÃO
08.1	Asfalto	10
08.2	Paralelo	08
08.3	Concreto	08
08.4	Lajota	09
08.5	Pedra Irregular	08
09	Iluminação Pública	05
10	Drenagem	06
11	Calçada	04
12	Rede Telefônica	
12.1	Aérea	03
12.2	Subterrânea	04
13	Rede Elétrica	
13.1	Alta Tensão	
13.2	Baixa Tensão	12
13.3	Baixa e Alta Tensão	12
13.4	Ligação Irregular	
14	Rede de Esgoto	08
15	Rede de Água	17
16	Tipo de Logradouro	
16.1	Via Primária	06
16.2	Via Secundária	05
16.3	Via Coletora Principal	04
16.4	Via Coletora Secundária	03
MÁXIMO DE PONTOS		100

**Fonte:** Planta Genérica de Valores de São Luís.

A partir dessa tabela de pesos e parâmetros, os valores correspondentes a cada item de infraestrutura e serviços urbanos foram associados às variáveis CONSERVACAO\_VIA, LIMPEZA\_PUB, TRANSPORTE, COLETA\_LIXO, ARBORIZACAO, MEIO\_FIO, PAVIMENTACAO, ILUMINACAO\_PUB, DRENAGEM, CALCADA, REDE\_TEL, REDE\_ELE, REDE\_ESG e REDE\_AGUA da base de dados analisada. Para cada inscrição imobiliária, as classificações observadas nessas variáveis foram convertidas em pontuações individuais, conforme os pesos definidos na Planta Genérica de Valores, e posteriormente agregadas, resultando em uma pontuação sintética de infraestrutura e serviços urbanos, representada pela variável PT\_INFRA\_SERV.

Destaca-se que, em razão de limitações da base de dados disponibilizada, os itens 03 (Estacionamento) e 16 (Tipo de Logradouro) não foram incorporados no cálculo da pontuação. Dessa forma, com o item Estacionamento possui peso de 2 pontos e o item Tipo de Logradouro apresenta peso máximo de 6 pontos, a pontuação máxima originalmente prevista de 100 pontos foi ajustada para 92 pontos ( $100 - 2 - 6$ ). Por conseguinte, a variável PT\_INFRA\_SERV passou a assumir valores no intervalo de 0 a 92.

Adicionalmente, no item 06 (Arborização), a base de dados fornecida pela SEMFAZ apresentava quatro categorias: uma árvore, até três árvores, mais de três árvores e não tem. Como a classificação da PGV não distingue quantitativamente o número de árvores, as categorias uma árvore, até três árvores e mais de três árvores foram agrupadas e atribuídas à pontuação 3, conforme previsto na tabela de pesos, enquanto a categoria não tem recebeu pontuação zero.

Por fim, no que se refere à variável Pavimentação, identificou-se a presença da categoria “ANTEPO” em 4.921 inscrições imobiliárias. Considerando que essa classificação não possui correspondência direta na tabela da PGV e que se refere a um tipo de solo laterítico, atribuiu-se a essa categoria a pontuação zero.

Esse conjunto de ajustes permitiu a construção de uma variável sintética de infraestrutura e serviços urbanos compatível com as informações disponíveis, preservando, na medida do possível, a lógica e os pesos definidos na Planta Genérica de Valores do município.

## **5 RESULTADOS E DISCUSSÕES**

Este capítulo apresenta os resultados obtidos a partir da aplicação das etapas metodológicas descritas anteriormente, articulando análises descritivas da base cadastral e financeira do IPTU de São Luís com técnicas de aprendizado de máquina supervisionadas e não supervisionadas. Inicialmente, é realizado um diagnóstico do comportamento dos lançamentos e da inadimplência do imposto no período de 2019 a 2023, considerando tanto a evolução da base tributável quanto a dinâmica de pagamento ao longo do tempo, sob diferentes recortes analíticos.

Na sequência, são apresentados os resultados do modelo de classificação multiclasse, empregado para identificar os fatores associados à adimplência e à inadimplência, bem como os achados da etapa de clusterização, voltada à

identificação de perfis homogêneos de contribuintes inadimplentes. Por fim, o capítulo explora a identificação de padrões atípicos e inconsistências cadastrais, discutindo como as evidências empíricas obtidas podem subsidiar a interpretação dos resultados e orientar potenciais ações da administração tributária.

## **5.1 Diagnóstico do IPTU em São Luís**

O diagnóstico do IPTU em São Luís, apresentado nesta seção, consolida a análise da base financeira e cadastral do município no período de 2019 a 2023, com foco em caracterizar a dimensão do lançamento do tributo, sua evolução recente e os padrões de adimplência observados ao longo do tempo. Inicialmente, examinam-se os lançamentos com valor de IPTU maior que zero e sua distribuição segundo a presença de benefício fiscal, permitindo avaliar a estabilidade da base tributável e a trajetória do montante potencialmente arrecadável no período.

Em seguida, aprofunda-se a análise restrita aos imóveis sem benefício fiscal, por representarem os contribuintes plenamente obrigados ao pagamento, estimando-se a dinâmica de pagamento em janelas temporais sucessivas e os níveis de inadimplência ao final do prazo prescricional. Por fim, a seção adota a perspectiva de série histórica ao integrar informações cadastrais e financeiras em uma classificação de adimplência (adimplente, adimplente parcial e inadimplente), explorando sua associação com atributos do imóvel e indicadores de infraestrutura e serviços urbanos, além de incorporar a dimensão espacial por bairros e setores fiscais, evidenciando disparidades territoriais relevantes para a administração tributária.

### **5.1.1 Análise da base financeira: lançamentos com valor de IPTU maior do que zero**

A análise dos dados do Imposto Predial e Territorial Urbano (IPTU) do município de São Luís, no período de 2019 a 2023, evidencia um cenário de relativa estabilidade com relação ao quantitativo de imóveis com lançamento do tributo. Conforme tabela 3, ao longo dos cinco anos analisados, observa-se um crescimento modesto, da ordem de aproximadamente 2,1%, no número de imóveis únicos com valor de IPTU lançado maior do que zero, indicando uma expansão limitada da base cadastral tributável no período.

**Tabela 3:** Detalhamento dos lançamentos cujo valor de IPTU foi maior que zero

ANO	QUANTIDADE DE LANÇAMENTOS DE IPTU COM "VALOR_IPTU" > 0	QUANTIDADE DE LANÇAMENTOS DE IPTU COM "VALOR_IPTU" > 0 E SEM BENEFÍCIO FISCAL	QUANTIDADE DE LANÇAMENTOS DE IPTU COM "VALOR_IPTU" > 0 E COM BENEFÍCIO FISCAL
2019	363.932	311.998	51.934
2020	366.390	314.258	52.132
2021	368.498	316.411	52.087
2022	369.624	321.564	48.060
2023	371.423	323.757	47.666

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A desagregação dos dados por condição de benefício fiscal revela, adicionalmente, uma tendência de redução no número de lançamentos com benefício, especialmente a partir de 2021. Em 2019, os lançamentos com benefício fiscal totalizavam 51.934 registros, enquanto em 2023 esse número foi reduzido para 47.666. Em termos relativos, também se observa um decréscimo da participação de imóveis beneficiados, de 14,3% para 12,8% do total de lançamentos no período analisado. Esse comportamento pode indicar revisões na política de concessão de benefícios, alterações nos critérios de elegibilidade ou aprimoramentos nos mecanismos de controle cadastral e fiscal.

Por outro lado, os lançamentos sem benefício fiscal apresentaram crescimento contínuo ao longo do período, reforçando a hipótese de um leve fortalecimento da base efetivamente tributável.

Em paralelo, conforme tabela 4, verifica-se que os valores lançados apresentaram uma trajetória de expansão mais significativa: em 2019, o valor total lançado (incluindo imóveis com benefício fiscal) foi de R\$ 202,69 milhões, a R\$ 259,03 milhões em 2023, um crescimento nominal de 27,8%. É importante ressaltar que esse aumento se refere basicamente à atualização monetária anual dos valores imobiliários que é feita com base na variação do Índice Nacional de Preços ao Consumidor Amplo (IPCA), na forma do artigo 170 do Código Tributário do Município de São Luís.

**Tabela 4:** Detalhamento dos valores lançados de IPTU

ANO	SOMA DOS LANÇAMENTOS DE IPTU	SOMA DOS LANÇAMENTOS DE IPTU SEM BENEFÍCIO FISCAL	SOMA DOS LANÇAMENTOS DE IPTU COM BENEFÍCIO FISCAL
2019	R\$ 202.693.060,31	R\$ 175.721.874,94	R\$ 26.971.185,37
2020	R\$ 208.834.188,51	R\$ 181.374.457,28	R\$ 27.459.731,23
2021	R\$ 217.450.697,15	R\$ 188.827.888,48	R\$ 28.622.808,67
2022	R\$ 241.507.047,94	R\$ 210.356.458,63	R\$ 31.150.589,31
2023	R\$ 259.027.297,27	R\$ 227.260.601,86	R\$ 31.766.695,41

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Um ponto metodológico importante deve ser destacado: o *dataset* disponibilizado registra apenas se o imóvel possui ou não possui benefício fiscal, sem detalhar o tipo de benefício ou a proporção de redução aplicada. Isso significa que imóveis registrados como “com benefício” podem abranger tanto isenções totais quanto parciais. Portanto, no tratamento da base de dados, quando o imóvel apresentava indicação de benefício fiscal, o respectivo valor total lançado foi considerado, neste trabalho, como integrante do montante dos imóveis com benefício fiscal. Reconhece-se que essa classificação pode incluir situações de isenção parcial (por exemplo, de 50%), o que implica em possíveis distorções em relação ao valor efetivamente dispensado. Ainda assim, a opção metodológica adotada fundamenta-se na delimitação do objeto de estudo: a análise da inadimplência do IPTU. Nessa perspectiva, a mensuração precisa dos valores referentes aos imóveis com benefício fiscal não se constitui como foco central da pesquisa, uma vez que imóveis beneficiados não representam o núcleo do problema de inadimplência que se busca investigar. Assim, embora tais distorções devam ser registradas, elas não comprometem a finalidade do trabalho, que é compreender o comportamento dos contribuintes plenamente obrigados ao pagamento e que, apesar disso, deixam de recolher o tributo devido.

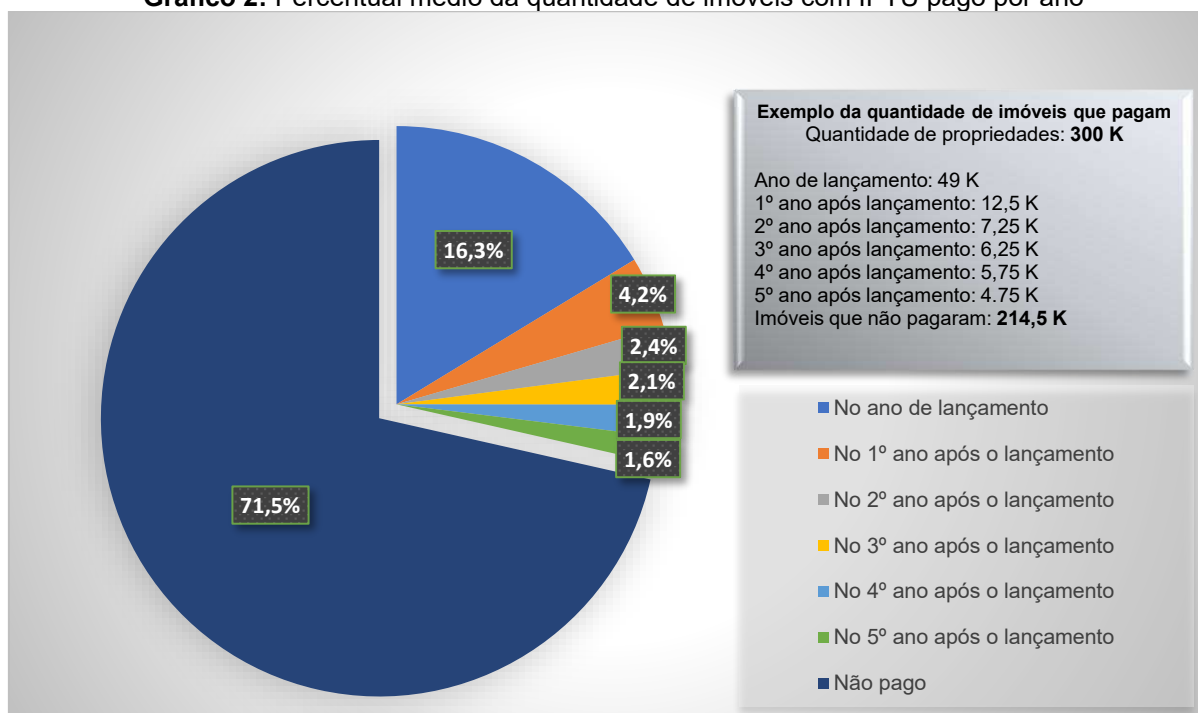
Mesmo com essa limitação, os dados revelam que o valor lançado de imóveis com benefício cresceu de R\$ 26,97 milhões em 2019 para R\$ 31,77 milhões em 2023. Entretanto, a participação desses imóveis em relação ao valor total lançado caiu levemente, passando de 13,3% para 12,3%, sinalizando que o crescimento da arrecadação potencial tem se concentrado mais fortemente nos imóveis sem benefício.

### 5.1.2 Análise da base financeira: lançamentos com valor de IPTU maior do que zero e sem benefício fiscal

As análises realizadas neste item referem-se exclusivamente ao universo de imóveis sem benefício fiscal e com valor de IPTU lançado superior a zero, por ano, pois representam os contribuintes plenamente obrigados ao pagamento do imposto e, portanto, aqueles com maior potencial arrecadatório. Esse conjunto corresponde às quantidades apresentadas na penúltima coluna da Tabela 3.

Conforme ilustrado no Gráfico 2, observa-se que, entre os imóveis sem benefício fiscal e com valor de IPTU positivo, em média apenas 16,3% dos contribuintes efetuam o pagamento no próprio ano de lançamento, o que implica que 83,7% não realizam qualquer pagamento nesse primeiro momento. Ao final do quinto ano após o lançamento, período em que, em geral, ocorre a prescrição do crédito tributário, verifica-se que 71,5% dos imóveis permanecem inadimplentes. Em termos práticos, isso significa que, após a prescrição, cerca de 7 em cada 10 proprietários não pagam o IPTU no município de São Luís, evidenciando um elevado nível de inadimplência estrutural.

**Gráfico 2:** Percentual médio da quantidade de imóveis com IPTU pago por ano



**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

É de extrema importância lembrar que, conforme já exposto no item 4.1.2, por escolha metodológica, em função das limitações da base de dados disponibilizada e das características do problema em análise, a existência de qualquer pagamento foi considerada como adimplência. O Gráfico 2 corrobora essa decisão, ao evidenciar que, mesmo sob uma abordagem favorável à adimplência, o percentual de imóveis inadimplentes permanece elevado, reforçando a magnitude do problema da inadimplência do IPTU no município de São Luís.

Ademais, considerando que a base de dados foi obtida em abril de 2025, apenas os lançamentos realizados em 2019 completaram integralmente o período de cinco anos potenciais de pagamento, enquanto os lançamentos de 2023 dispunham de apenas um ano e alguns meses desde o ano de lançamento. Diante dessa assimetria temporal, optou-se por calcular, para cada ano de lançamento, o percentual de imóveis que efetuaram pagamento em cada período subsequente (no ano de lançamento, no primeiro ano após o lançamento, no segundo ano, e assim sucessivamente), e, em seguida, obter a média desses percentuais por período, considerando o conjunto dos cinco anos analisados (2019 a 2023). O percentual de imóveis não pagos foi determinado como a diferença entre 100% e a soma dos percentuais médios de pagamento observados nos diferentes períodos, permitindo uma leitura consolidada do comportamento de adimplência ao longo do tempo. A Tabela 5 apresenta de forma detalhada esses percentuais.

**Tabela 5:** Detalhamento do percentual da quantidade de imóveis com IPTU pago por ano

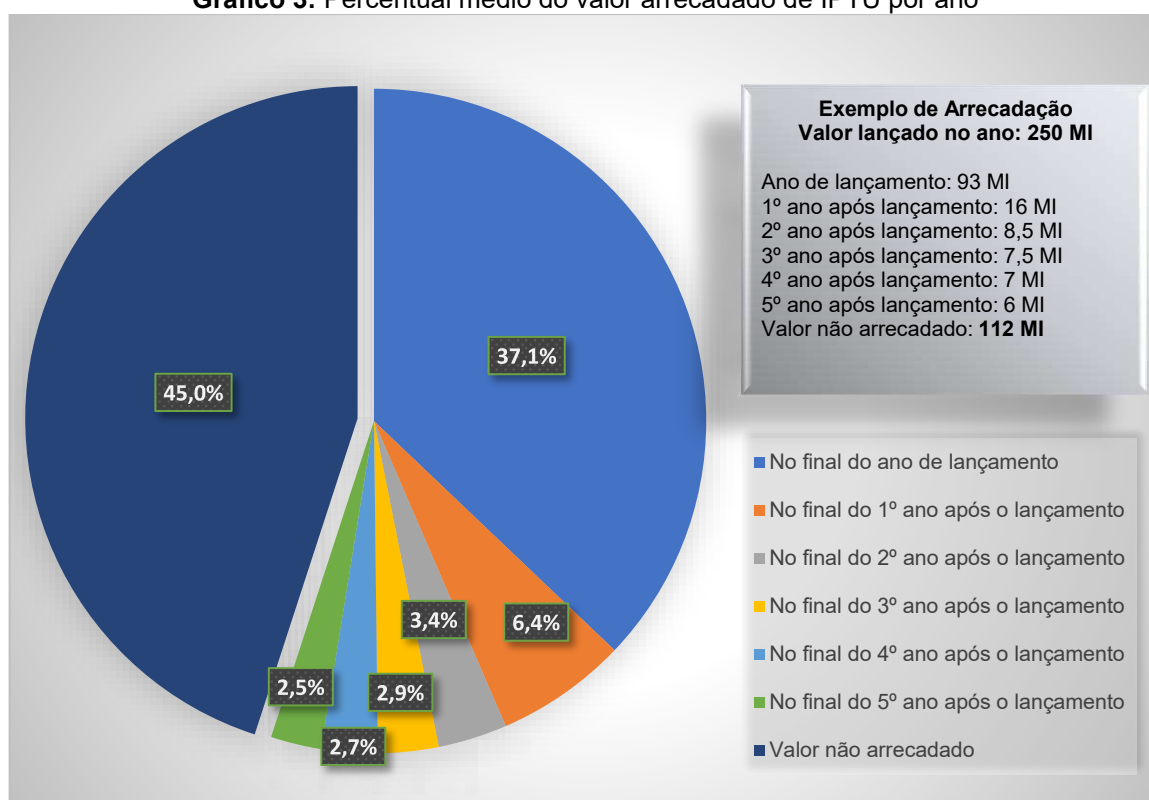
DATA DE REFERÊNCIA DO PAGAMENTO	ANO DE LANÇAMENTO					
	2019	2020	2021	2022	2023	MÉDIA
No ano de lançamento	15,37%	16,15%	16,52%	16,47%	17,03%	16,31%
No 1º ano após o lançamento	4,76%	3,41%	4,25%	4,72%	3,81%	4,19%
No 2º ano após o lançamento	1,80%	2,52%	2,43%	2,96%		2,43%
No 3º ano após o lançamento	1,66%	2,27%	2,36%			2,10%
No 4º ano após o lançamento	1,67%	2,17%				1,92%
No 5º ano após o lançamento	1,59%					1,59%
Perç. quant. não paga (100% - 16,31% - 4,19% - 2,43% - 2,10% - 1,59%) =						71,46%

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Passando à análise dos percentuais em relação ao valor lançado, conforme ilustrado no Gráfico 3, observa-se que, para os imóveis sem benefício fiscal e com valor de IPTU positivo, o percentual médio de arrecadação no próprio ano de

lançamento é de 37,1%, o que implica uma inadimplência de 62,9% nesse mesmo período. Ao final do quinto ano após o lançamento, verifica-se que, em média, 45% do valor total lançado em cada exercício não é arrecadado. Em termos práticos, esse resultado indica que, a cada R\$ 100 milhões lançados, aproximadamente R\$ 45 milhões deixam de ingressar nos cofres públicos após o prazo prescricional, evidenciando perdas significativas para a administração tributária municipal.

**Gráfico 3:** Percentual médio do valor arrecadado de IPTU por ano



**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

As mesmas considerações metodológicas adotadas para o Gráfico 2, relativas à definição de inadimplência e ao cálculo dos percentuais, aplicam-se igualmente ao Gráfico 3. Adicionalmente, ressalta-se que o percentual de inadimplência tende a ser ligeiramente subestimado, uma vez que o valor utilizado como base de cálculo do gráfico foi o valor lançado do IPTU. Assim, sempre que houve qualquer pagamento, ainda que a menor, considerou-se, para fins de cálculo, como se a totalidade do valor lançado de IPTU tivesse sido quitada. Ainda assim, mesmo sob essa abordagem metodológica conservadora e favorável à inadimplência, conforme

já discutido anteriormente, os resultados indicam que o percentual de inadimplência permanece elevado. A tabela 6 detalha como gráfico 3 foi construído.

**Tabela 6:** Detalhamento do percentual dos valores arrecadados de IPTU por ano

DATA DE REFERÊNCIA DO PAGAMENTO	ANO DE LANÇAMENTO					MÉDIA
	2019	2020	2021	2022	2023	
No final do ano de lançamento	35,13%	36,13%	37,66%	37,44%	39,20%	37,11%
No final do 1º ano após o lançamento	7,33%	5,25%	6,46%	7,53%	5,44%	6,40%
No final do 2º ano após o lançamento	2,51%	3,78%	3,32%	3,84%		3,36%
No final do 3º ano após o lançamento	2,21%	3,35%	3,27%			2,94%
No final do 4º ano após o lançamento	2,40%	3,07%				2,73%
No final do 5º ano após o lançamento	2,46%					2,46%
Perc. valor não arrecadado (100% - 37,11% - 6,40% - 3,36% - 2,94% - 2,73% - 2,46%) =						45,00%

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Um aspecto positivo que emerge da análise conjunta dos dois gráficos é que os imóveis que efetuam o pagamento do IPTU tendem a concentrar valores mais elevados do tributo. Explica-se: enquanto apenas 16,3% dos imóveis realizam o pagamento no ano de lançamento, esse grupo é responsável por 37,1% do valor total arrecadado, indicando que os contribuintes adimplentes, em média, estão associados a valores de IPTU mais altos. De forma complementar, observa-se que os contribuintes inadimplentes ao final do quinto ano após o lançamento, que representam 71,5% dos imóveis, correspondem a 45% do valor total não arrecadado. Esse descompasso entre a proporção de imóveis e a proporção de valores indica que, embora a inadimplência seja numericamente predominante, ela se concentra, em grande parte, em imóveis de menor valor venal, ao passo que os imóveis de maior valor revelam maior propensão ao pagamento.

### 5.1.3 Análise da base cadastral e financeira sob a perspectiva da série histórica

A análise deste item e de seus subitens refere-se à base de dados consolidada após, o tratamento dos dados descrito no item 4.2, no qual, tomando-se como base o ano de 2023, integrou-se os dados dos exercícios de 2019 a 2022 e procedeu-se à classificação das inscrições imobiliárias em três categorias: adimplente, adimplente parcial e inadimplente. Consideraram-se adimplentes os imóveis que efetuaram algum pagamento em todos os anos em que houve lançamento de IPTU no período de 2019 a 2023; adimplentes parciais, aqueles que realizaram algum pagamento em pelo menos um dos exercícios, mas não em todos;

e inadimplentes, os imóveis para os quais não foi identificado qualquer pagamento em nenhum dos lançamentos de IPTU ao longo do período analisado.

É importante destacar que a classe adimplente parcial foi criada especificamente para este estudo e está associada ao comportamento histórico dos imóveis no período de 2019 a 2023. Dessa forma, essa classe não existe de maneira anual isolada, mas apenas quando os dados são analisados como uma série temporal, conforme a metodologia adotada e explicada no item 4.2. Por esse motivo, enfatiza-se que os dados não representam a situação específica do ano de 2023, tendo sido utilizados apenas por corresponderem aos valores monetários mais recentes da série histórica e por apresentarem maior consonância com as informações constantes na base cadastral dos imóveis, cuja data-base é abril de 2025.

#### 5.1.3.1 Aspectos fiscais e cadastrais dos imóveis

Este subitem apresenta uma análise descritiva dos principais aspectos fiscais e cadastrais dos imóveis, considerando sua classificação em adimplentes, adimplentes parciais e inadimplentes. As tabelas a seguir exploram a distribuição dessas categorias segundo características como uso, tipologia, padrão construtivo, estrutura e qualidade cadastral, com o objetivo de identificar padrões iniciais e indícios de associação com o comportamento de pagamento.

A Tabela 7 apresenta a distribuição da base financeira do IPTU (ano base 2023) segundo a situação de adimplência, considerando o valor total lançado, a quantidade de inscrições imobiliárias e o valor médio do imposto. Ela mostra forte assimetria entre quantidade de imóveis e valor do IPTU. Embora os imóveis inadimplentes representem a maior parte da base (69,5%), eles concentram valores médios significativamente menores de IPTU. Em contraste, os adimplentes, apesar de numericamente inferiores, apresentam valor médio quase quatro vezes superior, sendo responsáveis por parcela expressiva do valor total lançado.

**Tabela 7:** Resumo dos dados de IPTU em relação à situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	SOMA DO VALOR DE IPTU (ANO BASE 2023)	QUANTIDADE DE INSCRIÇÕES IMOBILIÁRIAS	MÉDIA DO VALOR DE IPTU
Adimplente	R\$ 87.570.781,07	55.792	R\$ 1.569,59
Adimplente Parcial	R\$ 43.118.188,15	38.645	R\$ 1.115,75
Inadimplente	R\$ 89.512.275,50	214.865	R\$ 416,60
<b>Base Completa</b>	<b>R\$ 220.201.244,72</b>	<b>309.302</b>	<b>R\$ 711,93</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Destaca-se o grupo de adimplentes parciais, que, embora não mantenham regularidade de pagamento, realizam pagamentos em alguns exercícios, configurando um segmento com elevado potencial de recuperação de crédito. Esse grupo apresenta valor médio de IPTU superior ao dos inadimplentes e volume financeiro relevante, indicando maior propensão ao pagamento. Esses resultados sugerem que a administração tributária pode adotar ações direcionadas, como comunicações personalizadas, envio de notificações ou protesto em cartório, com maior probabilidade de êxito arrecadatório.

Quanto à validade do CPF/CNPJ do proprietário, a Tabela 8 mostra que entre os adimplentes, 89% dos imóveis possuem CPF/CNPJ válido, percentual que também se mantém elevado entre os adimplentes parciais (83%). Em contraste, entre os inadimplentes, 81% dos registros não apresentam identificação válida. Esses percentuais sugerem uma possível associação entre a qualidade cadastral e o comportamento de pagamento, indicando que essa variável pode ser relevante na etapa de modelagem de classificação supervisionada.

**Tabela 8:** Validade do CPF/CNPJ e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	CPF/CNPJ VÁLIDO	
	NÃO	SIM
Adimplente	11%	89%
Adimplente Parcial	17%	83%
Inadimplente	81%	19%
<b>Base Completa</b>	<b>60%</b>	<b>40%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Tabela 9 apresenta a distribuição dos imóveis segundo o uso do imóvel. Observa-se predominância do uso predial em todas as categorias de adimplência, especialmente entre os adimplentes e adimplentes parciais (95%). Entre os inadimplentes, a participação de imóveis territoriais é relativamente maior (10%), sugerindo maior dificuldade de arrecadação associada a terrenos.

**Tabela 9:** Uso do imóvel e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	USO DO IMÓVEL	
	PREDIAL	TERRITORIAL
Adimplente	95%	5%
Adimplente Parcial	95%	5%

SITUAÇÃO DE ADIMPLÊNCIA	USO DO IMÓVEL	
	PREDIAL	TERRITORIAL
Inadimplente	90%	10%
<b>Base Completa</b>	<b>92%</b>	<b>8%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Em relação a presença ou não do imóvel em condomínio, a Tabela 10 revela que a presença em condomínios é significativamente maior entre os adimplentes (65%) e adimplentes parciais (58%), enquanto os inadimplentes concentram-se majoritariamente em imóveis fora de condomínios (72%). Esse padrão sugere que a organização condominial pode estar associada a maior regularidade de pagamento, possivelmente em razão de melhor qualidade cadastral, maior valor do imóvel ou maior capacidade contributiva.

**Tabela 10:** Condomínio e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	CONDOMÍNIO	
	NÃO	SIM
Adimplente	35%	65%
Adimplente Parcial	42%	58%
Inadimplente	72%	28%
<b>Base Completa</b>	<b>61%</b>	<b>39%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Quanto à destinação do imóvel, consoante tabela 11, verifica-se que os imóveis comerciais apresentam maior participação entre os adimplentes (23%) e adimplentes parciais (21%), enquanto representam apenas 8% dos inadimplentes, indicando maior propensão ao pagamento nesse segmento. Em contraste, os imóveis residenciais concentram a inadimplência (80%), refletindo seu peso majoritário na base e possível maior sensibilidade econômica.

**Tabela 11:** Destinação do imóvel e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	DESTINAÇÃO DO IMÓVEL			
	COMERCIAL	OUTRA	RESIDENCIAL	TERRITORIAL
Adimplente	23%	5%	66%	5%
Adimplente Parcial	21%	5%	69%	5%
Inadimplente	8%	2%	80%	10%
<b>Base Completa</b>	<b>12%</b>	<b>3%</b>	<b>76%</b>	<b>8%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Relacionado à conservação do imóvel, a tabela 12 revela que os imóveis adimplentes e adimplentes parciais se concentram predominantemente em padrões de conservação boa ou ótima, enquanto os inadimplentes estão majoritariamente associados ao padrão regular (50,6%). A baixa participação de imóveis em estado ótimo entre os inadimplentes (3,1%) indica uma relação entre qualidade do imóvel e capacidade de pagamento.

**Tabela 12:** Conservação do imóvel e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	CONSERVAÇÃO DO IMÓVEL				
	BOA	ÓTIMA	PRECÁRIA	REGULAR	TERRITORIAL
Adimplente	46,6%	32,1%	0,4%	15,5%	5,3%
Adimplente Parcial	47,1%	25,7%	0,6%	21,4%	5,1%
Inadimplente	33,7%	3,1%	2,8%	50,6%	9,9%
<b>Base Completa</b>	<b>37,7%</b>	<b>11,2%</b>	<b>2,1%</b>	<b>40,6%</b>	<b>8,5%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A principal informação da Tabela 13, referente à estrutura do imóvel, é a baixa participação de construções em concreto entre os inadimplentes, em contraste com sua elevada presença entre os adimplentes e adimplentes parciais, o que sugere uma associação positiva entre uma melhor estrutura e a adimplência.

**Tabela 13:** Estrutura do imóvel e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	ESTRUTURA DO IMÓVEL							
	ALVENARIA	CONCRETO	MADEIRA	METÁLICA	MISTA ALV/CONC	MISTA ALV/MAD	TAIPA	TERRITORIAL
Adimplente	13,77%	47,02%	0,55%	0,30%	32,71%	0,35%	0,01%	5,30%
Adimplente Parcial	15,51%	37,74%	0,66%	0,35%	40,08%	0,51%	0,01%	5,14%
Inadimplente	15,74%	6,96%	1,96%	0,05%	64,19%	0,73%	0,50%	9,86%
<b>Base Completa</b>	<b>15,36%</b>	<b>18,03%</b>	<b>1,54%</b>	<b>0,13%</b>	<b>55,50%</b>	<b>0,64%</b>	<b>0,35%</b>	<b>8,45%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Em relação ao padrão construtivo do imóvel, a tabela 14 aponta que os adimplentes e adimplentes parciais se concentram principalmente em imóveis de padrão construtivo médio ou alto, enquanto os inadimplentes apresentam concentração em imóveis de padrão baixo e médio e participação mínima de padrão alto (1%).

**Tabela 14:** Padrão construtivo e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	PADRÃO CONSTRUTIVO			
	ALTO	MÉDIO	BAIXO	TERRITORIAL
Adimplente	15%	72%	8%	5%
Adimplente Parcial	11%	73%	12%	5%
Inadimplente	1%	43%	46%	10%
<b>Base Completa</b>	<b>5%</b>	<b>52%</b>	<b>35%</b>	<b>8%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A tabela 15, relativa à tipologia do imóvel, evidencia que os inadimplentes se concentram majoritariamente em casas, enquanto os adimplentes e adimplentes parciais apresentam maior diversidade tipológica, com destaque para apartamentos, casas e lojas/salas/conjuntos.

**Tabela 15:** Tipologia do imóvel e situação de adimplência

TIPOLOGIA DO IMÓVEL	SITUAÇÃO DE ADIMPLÊNCIA			BASE COMPLETA
	ADIMPLENTE	ADIMPLENTE PARCIAL	INADIMPLENTE	
APTO FRENTE	32,72%	25,78%	5,16%	12,71%
APTO FUNDOS	1,83%	1,52%	0,26%	0,70%
APTO TERREO	0,22%	0,26%	0,13%	0,16%
BARRACO	0,00%	0,01%	0,20%	0,14%
CASA	28,80%	38,25%	73,73%	61,19%
COB METALICA	0,21%	0,25%	0,04%	0,09%
COM C/ RESIDENCIA	0,15%	0,25%	0,26%	0,24%
DEPOSITO	0,16%	0,18%	0,10%	0,12%
EDIFICACAO COMPLEMENTAR	3,57%	3,89%	1,04%	1,85%
GALPAO	0,85%	0,87%	0,27%	0,45%
GARAGEM	2,14%	1,06%	0,30%	0,73%
INDUSTRIA	0,04%	0,01%	0,01%	0,01%
LOJA/SALA/CONJUNTO	23,14%	21,49%	7,76%	12,25%
SOBRADO	0,04%	0,05%	0,03%	0,04%
SOBRADO COLONIAL	0,11%	0,15%	0,06%	0,08%
TELHEIRO	0,65%	0,68%	0,36%	0,45%
TEMPLO	0,09%	0,17%	0,42%	0,33%
TERRITORIAL	5,30%	5,14%	9,86%	8,45%

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

### 5.1.3.2 Características de infraestrutura e serviços urbanos dos imóveis

Este subitem apresenta uma análise descritiva das condições de infraestrutura e dos serviços urbanos associados aos imóveis, considerando sua classificação em adimplentes, adimplentes parciais e inadimplentes, com o objetivo de identificar padrões iniciais na distribuição desses atributos segundo a situação de adimplência.

É importante destacar que esta análise reflete exclusivamente as informações constantes na base de dados da administração tributária, a qual pode apresentar defasagens ou desatualizações cadastrais em alguns registros. Dessa forma, determinados parâmetros analisados podem não refletir integralmente a realidade atual do município de São Luís, devendo os resultados ser interpretados à luz dessa informação.

A Tabela 16 mostra a distribuição dos imóveis segundo a conservação da via. Observa-se que a maioria dos adimplentes (83%) e adimplentes parciais (78%) está localizada em vias conservadas, enquanto entre os inadimplentes predomina a ausência de conservação (55%).

**Tabela 16:** Conservação da via e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	CONSERVAÇÃO DA VIA	
	NÃO TEM	TEM
Adimplente	17%	83%
Adimplente Parcial	22%	78%
Inadimplente	55%	45%
<b>Base Completa</b>	<b>44%</b>	<b>56%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Quanto à existência de limpeza pública, a Tabela 17 evidencia que a maioria dos adimplentes (67%) e adimplentes parciais (61%) está em áreas atendidas pelo serviço, enquanto entre os inadimplentes predomina a ausência.

**Tabela 17:** Limpeza pública e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	LIMPEZA PÚBLICA	
	NÃO TEM	TEM
Adimplente	33%	67%
Adimplente Parcial	39%	61%
Inadimplente	66%	34%
<b>Base Completa</b>	<b>57%</b>	<b>43%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Tabela 18 apresenta a distribuição dos imóveis segundo as condições de acesso ao transporte. Observa-se que os inadimplentes se concentram mais frequentemente em áreas sem oferta de transporte (58%), enquanto os adimplentes apresentam maior participação em locais com acesso direto (30%).

**Tabela 18:** Acesso a transporte e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	TRANSPORTE		
	DIRETO	PRÓXIMO	NÃO TEM
Adimplente	30%	26%	44%
Adimplente Parcial	27%	28%	45%
Inadimplente	13%	29%	58%
<b>Base Completa</b>	<b>18%</b>	<b>28%</b>	<b>54%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Em relação ao tipo de coleta de lixo, conforme Tabela 19, observa-se que os adimplentes e adimplentes parciais apresentam maior presença de coleta diária, enquanto os inadimplentes concentram-se predominantemente em áreas com coleta alternada ou inexistente (87%).

**Tabela 19:** Coleta de lixo e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	COLETA DE LIXO		
	DIÁRIA	ALTERNADA	NÃO TEM
Adimplente	38%	59%	3%
Adimplente Parcial	31%	65%	4%
Inadimplente	13%	78%	9%
<b>Base Completa</b>	<b>20%</b>	<b>73%</b>	<b>7%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Tabela 20 apresenta a distribuição dos imóveis segundo as condições de arborização. Ela evidencia que, em todas as classes de adimplência, predomina a ausência de arborização, com percentuais elevados sem arborização. Ainda assim, chama atenção que a presença de mais de três árvores é significativamente maior entre os adimplentes (8,9%) e adimplentes parciais (7,3%) do que entre os inadimplentes (1,3%).

**Tabela 20:** Arborização e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	ARBORIZAÇÃO			
	UMA ÁRVORE	ATÉ TRES ÁRVORES	MAIS DE TRÊS	NÃO TEM
Adimplente	14,4%	1,0%	8,9%	75,7%
Adimplente Parcial	13,8%	0,8%	7,3%	78,0%
Inadimplente	12,9%	0,3%	1,3%	85,5%
<b>Base Completa</b>	<b>13,3%</b>	<b>0,5%</b>	<b>3,5%</b>	<b>82,8%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Quanto à existência ou não de meio-fio, consoante tabela 21, nota-se maior presença desse elemento urbano entre os adimplentes (80%) e adimplentes parciais (77%), enquanto entre os inadimplentes a ausência é relativamente mais frequente (42%).

**Tabela 21:** Existência de meio-fio e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	MEIO-FIO	
	NÃO TEM	TEM
Adimplente	20%	80%
Adimplente Parcial	23%	77%
Inadimplente	42%	58%
<b>Base Completa</b>	<b>36%</b>	<b>64%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Tabela 22 detalha a distribuição dos imóveis segundo o tipo de pavimentação da via. Verifica-se predominância de asfalto em todas as classes, especialmente entre adimplentes (92,8%) e adimplentes parciais (91,1%). Já entre os inadimplentes, a participação de vias sem pavimentação é superior.

**Tabela 22:** Pavimentação e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	PAVIMENTAÇÃO						
	ANTEPO	ASFALTO	CONCRETO	LAJOTA	PARALELEPÍPEDO	PEDRA IRREGULAR	NÃO TEM
Adimplente	1,36%	92,84%	0,28%	0,04%	2,40%	0,32%	2,75%
Adimplente Parcial	1,42%	91,13%	0,35%	0,08%	2,28%	0,67%	4,06%
Inadimplente	2,29%	70,62%	1,83%	0,12%	5,76%	2,13%	17,25%
<b>Base Completa</b>	<b>2,01%</b>	<b>77,19%</b>	<b>1,37%</b>	<b>0,10%</b>	<b>4,72%</b>	<b>1,62%</b>	<b>12,98%</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Em relação à existência de iluminação pública, tabela 23, constata-se elevada presença do serviço em todas as classes, com percentuais superiores a 90%, especialmente entre adimplentes (96%) e adimplentes parciais (95%).

**Tabela 23:** Iluminação pública e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	ILUMINAÇÃO PÚBLICA	
	NÃO TEM	TEM
Adimplente	4%	96%
Adimplente Parcial	5%	95%
Inadimplente	9%	91%
<b>Base Completa</b>	<b>8%</b>	<b>92%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Já quanto a existência de drenagem, segundo tabela 24, verifica-se que a ausência de drenagem é predominante em todas as classes, porém é mais acentuada entre os inadimplentes (80%).

**Tabela 24:** Drenagem e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	DRENAGEM	
	NÃO TEM	TEM
Adimplente	61%	39%
Adimplente Parcial	66%	34%
Inadimplente	80%	20%
<b>Base Completa</b>	<b>75%</b>	<b>25%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Segundo tabela 25, observa-se que a presença de calçada é majoritária entre os adimplentes (69%) e adimplentes parciais (61%), enquanto entre os inadimplentes predomina a ausência desse elemento urbano (71%).

**Tabela 25:** Existência de calçada e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	CALÇADA	
	NÃO TEM	TEM
Adimplente	31%	69%
Adimplente Parcial	39%	61%
Inadimplente	71%	29%
<b>Base Completa</b>	<b>59%</b>	<b>41%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Tabela 26 apresenta a distribuição dos imóveis segundo o tipo de rede telefônica. Observa-se predominância da rede aérea em todas as classes, com diferenças relativamente pequenas entre os grupos. Ainda assim, a ausência de rede telefônica é mais frequente entre os inadimplentes (6%) do que entre os adimplentes e adimplentes parciais (3%), enquanto a participação de rede subterrânea é maior entre os adimplentes.

**Tabela 26:** Rede telefônica e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	REDE TELEFÔNICA		
	AÉREA	SUBTERRÂNEA	NÃO TEM
Adimplente	87%	11%	3%
Adimplente Parcial	89%	8%	3%
Inadimplente	89%	6%	6%
<b>Base Completa</b>	<b>88%</b>	<b>7%</b>	<b>5%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Quanto ao tipo de rede elétrica, a Tabela 27 mostra que os adimplentes e adimplentes parciais concentram-se majoritariamente em locais de redes mistas (baixa/alta tensão), enquanto entre os inadimplentes predomina a baixa tensão (53,6%) e é mais elevada a ausência de rede elétrica (4,5%).

**Tabela 27:** Rede elétrica e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	REDE ELÉTRICA				
	ALTA TENSÃO	BAIXA TENSÃO	BAIXA/ALTA TENSÃO	LIGACÃO IRREGULAR	NÃO TEM
Adimplente	11,29%	26,93%	59,44%	0,02%	2,32%
Adimplente Parcial	9,82%	32,77%	54,60%	0,02%	2,79%
Inadimplente	6,01%	53,63%	35,70%	0,14%	4,52%
<b>Base Completa</b>	<b>7,44%</b>	<b>46,21%</b>	<b>42,34%</b>	<b>0,10%</b>	<b>3,91%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Ainda sobre a tabela 27, observa-se um aspecto relevante: a presença de relativos percentuais de imóveis classificados como atendidos exclusivamente por rede elétrica de alta tensão, especialmente entre os adimplentes e adimplentes parciais. A rede de alta tensão não é destinada ao uso cotidiano dos imóveis, uma vez que o consumo regular requer rede de baixa tensão. Ademais, conforme a tabela de Pesos e Parâmetros para Infraestrutura e Serviços da Planta Genérica de Valores, apresentada no item 4.3, a rede elétrica de alta tensão não recebe pontuação,

justamente por não ser diretamente acessível aos imóveis. Esse resultado pode indicar inconsistências ou erros cadastrais na base da administração tributária, com potencial impacto na subavaliação da infraestrutura do imóvel e, conseqüentemente, na redução do valor do IPTU lançado.

A Tabela 28 apresenta a distribuição dos imóveis segundo a existência de rede de esgoto. Contata-se que a presença do serviço é majoritária entre os adimplentes e adimplentes parciais, enquanto entre os inadimplentes a ausência de rede é significativamente maior quando comparada com as outras duas classes.

**Tabela 28:** Rede de esgoto e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	REDE DE ESGOTO	
	NÃO TEM	TEM
Adimplente	14%	86%
Adimplente Parcial	17%	83%
Inadimplente	51%	49%
<b>Base Completa</b>	<b>40%</b>	<b>60%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Por fim, em relação à existência de rede de abastecimento de água, conforme Tabela 29, observa-se elevada cobertura do serviço em todas as classes, com percentuais superiores a 90%, especialmente entre os adimplentes e adimplentes parciais (95%). Ainda que a diferença seja moderada, verifica-se uma maior ausência de rede de água entre os inadimplentes (10%).

**Tabela 29:** Rede de água e situação de adimplência

SITUAÇÃO DE ADIMPLÊNCIA	REDE DE ÁGUA	
	NÃO TEM	TEM
Adimplente	5%	95%
Adimplente Parcial	5%	95%
Inadimplente	10%	90%
<b>Base Completa</b>	<b>8%</b>	<b>92%</b>

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

### 5.1.3.3 Aspectos fiscais sob a ótica espacial: setores e bairros

Este subitem introduz uma análise sob a ótica espacial do IPTU, com foco nos bairros e setores fiscais. A Tabela 30 apresenta, entre os bairros com mais de 500 imóveis cadastrados, aqueles com os maiores valores médios de IPTU. Observa-se

que esses bairros estão majoritariamente localizados em áreas centrais ou litorâneas, caracterizadas por imóveis de padrão construtivo mais elevado.

**Tabela 30:** Média do valor de IPTU por bairro (Top 20)

<b>BAIRRO</b>	<b>MÉDIA DO VALOR DE IPTU (ANO BASE 2023)</b>
PONTA D AREIA	R\$ 2.861,55
CALHAU	R\$ 2.109,73
SAO MARCOS	R\$ 2.078,48
JARACATY	R\$ 1.984,60
PONTA DO FAROL	R\$ 1.903,54
JARDIM RENASCENCA	R\$ 1.801,63
VILA MARANHAO	R\$ 1.598,47
ITAPIRACO	R\$ 1.466,41
OLHO D AGUA	R\$ 1.418,04
COHAMA	R\$ 1.411,38
RENASCENCA	R\$ 1.338,95
PARQUE ATENAS	R\$ 1.302,61
VINHAI I	R\$ 1.297,69
RECANTO DOS VINHAIS	R\$ 1.168,55
CENTRO	R\$ 1.136,75
CHACARA BRASIL	R\$ 1.112,83
TURU	R\$ 1.100,94
COHAJAP	R\$ 1.059,62
PARQUE PINDORAMA	R\$ 1.025,45
VINHAI S	R\$ 1.020,99

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A estratificação territorial do lançamento do IPTU reforça as disparidades internas do município de São Luís. Conforme Tabela 31 a seguir, os dez bairros com maiores valores totais lançados foram Calhau, Renascença, Olho d'Água, Ponta d'Areia, Turu, Centro, Cohama, Itaqui, Pedrinhas e Ponta do Farol. Em conjunto, esses bairros somaram mais de R\$ 115 milhões em IPTU lançado no ano base de 2023, o que representa mais de 50% do valor total lançado da base analisada, cujo valor total contabilizado foi um pouco mais de R\$ 220 milhões, conforme discutido no item 5.1.3.1. Trata-se, majoritariamente, de áreas com maior valorização imobiliária e imóveis de melhor padrão construtivo, com exceção dos bairros Itaqui e Pedrinhas, cujos elevados valores estão associados à destinação industrial.

**Tabela 31:** Valor lançado de IPTU por bairro e situação de adimplência

BAIRRO	SITUAÇÃO DE ADIMPLÊNCIA			BASE COMPLETA
	ADIMPLENTE	ADIMPLENTE PARCIAL	INADIMPLENTE	
CALHAU	R\$ 9.941.073,87	R\$ 6.173.107,33	R\$ 5.976.853,73	R\$ 22.091.034,93
RENASCENCA	R\$ 13.915.456,42	R\$ 4.621.396,70	R\$ 3.369.704,31	R\$ 21.906.557,43
OLHO D AGUA	R\$ 5.550.580,43	R\$ 3.539.791,08	R\$ 7.065.324,92	R\$ 16.155.696,43
PONTA D AREIA	R\$ 9.154.033,89	R\$ 3.627.238,39	R\$ 2.631.034,44	R\$ 15.412.306,72
TURU	R\$ 4.593.167,83	R\$ 3.832.986,83	R\$ 5.542.570,94	R\$ 13.968.725,60
CENTRO	R\$ 3.330.211,57	R\$ 1.686.712,23	R\$ 3.567.804,19	R\$ 8.584.727,99
COHAMA	R\$ 2.015.372,40	R\$ 1.118.326,24	R\$ 1.543.611,40	R\$ 4.677.310,04
ITAQUI	R\$ 3.836.798,81	R\$ 83.070,52	R\$ 375.491,05	R\$ 4.295.360,38
PEDRINHAS	R\$ 3.816.852,61	R\$ 89.796,34	R\$ 302.807,45	R\$ 4.209.456,40
PONTA DO FAROL	R\$ 2.398.598,21	R\$ 986.047,81	R\$ 679.407,62	R\$ 4.064.053,64

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A análise da coluna referente aos adimplentes na Tabela 31 indica que esses bairros apresentam, em geral, valores expressivos de arrecadação efetiva. Entretanto, merece destaque o fato de que aproximadamente R\$ 57 milhões concentram-se nas categorias de adimplente parcial e inadimplente, correspondendo a cerca de 50% do valor total lançado nesse conjunto de bairros. Esse resultado revela um nicho relevante e estratégico para atuação da administração tributária, com elevado potencial de recuperação de receita.

Ao considerar a Tabela 32 a seguir, que apresenta a quantidade de imóveis associada a esses mesmos bairros, observa-se que eles totalizam 69.723 imóveis, o que corresponde a 22,5% da base analisada (309.302 imóveis). Quando se analisam apenas os adimplentes parciais e inadimplentes, esse grupo soma 41.741 imóveis, representando 13,5% da base analisada, não obstante, conforme discutido nos parágrafos anteriores, esses grupos de imóveis está associado a valores financeiros elevados, o que reforça a oportunidade de ganho arrecadatório. Nesse contexto, os adimplentes parciais ganham especial relevância, uma vez que alternam comportamentos de pagamento e, devido aos altos valores de IPTU, configuram um público prioritário para ações direcionadas, com potencial para arrecadação de anos anteriores. Por fim, destaca-se novamente o caso dos bairros Itaqui e Pedrinhas, que, apesar do baixo número de imóveis, apresentam valores lançados elevados, refletindo sua vocação industrial.

**Tabela 32:** Quantidade de lançamentos de IPTU por bairro e situação de adimplência

BAIRRO	SITUAÇÃO DE ADIMPLÊNCIA			BASE COMPLETA
	ADIMPLENTE	ADIMPLENTE PARCIAL	INADIMPLENTE	
CALHAU	4.295	2.982	3.194	10.471
RENASCENCA	9.006	3.558	3.797	16.361
OLHO D AGUA	2.319	1.665	7.409	11.393
PONTA D AREIA	3.108	1.332	946	5.386
TURU	3.934	3.055	5.699	12.688
CENTRO	2.352	1.577	3.623	7.552
COHAMA	1.342	952	1.020	3.314
ITAQUI	40	9	37	86
PEDRINHAS	134	92	111	337
PONTA DO FAROL	1.452	507	176	2.135

**Fonte:** Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

Conforme tabela 33, a análise dos setores fiscais evidencia um padrão espacial consistente na média do valor do IPTU. Os setores com maiores médias de IPTU concentram-se, em geral, em áreas litorâneas ou centrais, refletindo maior valorização imobiliária e melhor padrão construtivo. A principal exceção é o Setor 16, que apresenta a maior média em razão de abranger o polo industrial, com imóveis de elevado valor venal. Em contraste, os setores com menores médias estão associados a áreas periféricas, mais distantes da faixa litorânea e dos núcleos centrais, ou mais próximas da zona rural, indicando menor dinamismo imobiliário e base tributária menos robusta.

**Tabela 33:** Média do valor de IPTU por setor fiscal

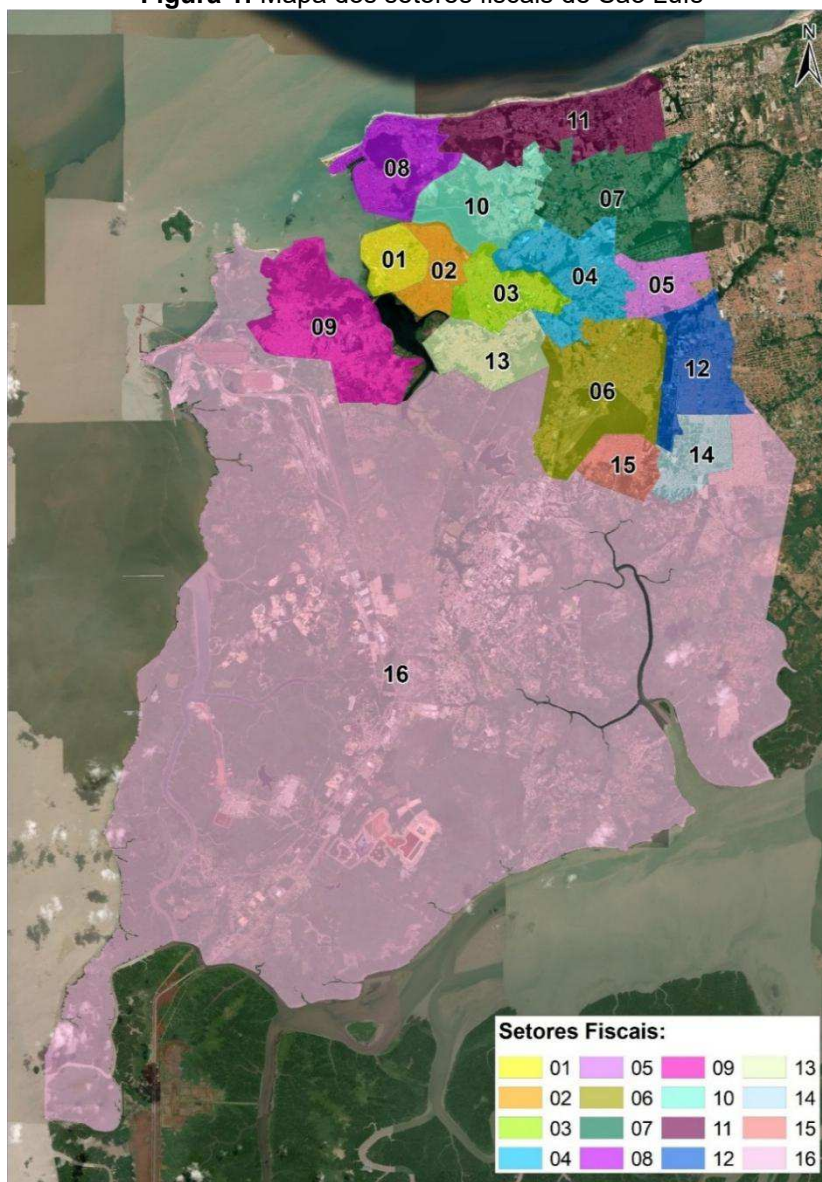
SETOR FISCAL	MÉDIA DO VALOR DE IPTU (ANO BASE 2023)
1	R\$ 876,49
2	R\$ 254,89
3	R\$ 207,83
4	R\$ 438,13
5	R\$ 421,80
6	R\$ 468,69
7	R\$ 1,042,98
8	R\$ 1,458,97
9	R\$ 160,32
10	R\$ 1,140,62
11	R\$ 1,599,77
12	R\$ 249,45

SETOR FISCAL	MÉDIA DO VALOR DE IPTU (ANO BASE 2023)
13	R\$ 232,65
14	R\$ 204,71
15	R\$ 138,85
16	R\$ 2,506,94
<b>Base Completa</b>	<b>R\$ 711,93</b>

Fonte: Elaborado pelo autor. Dados Primários: SEMFAZ-SL.

A Figura 1 apresenta a delimitação dos setores fiscais no mapa de São Luís, facilitando a visualização espacial e a interpretação dos padrões observados na análise dos valores médios de IPTU apresentados na tabela anterior.

Figura 1: Mapa dos setores fiscais de São Luís



Fonte: SEMFAZ-SL.

## 5.2 Modelo Supervisionado: Classificador Multiclasse

O algoritmo utilizado para a identificação dos fatores associados à inadimplência do IPTU foi o *XGBoost*. Trata-se de um algoritmo amplamente empregado em problemas de classificação e previsão em dados tabulares, destacando-se pela combinação entre alta precisão, eficiência computacional, bom desempenho em bases de dados esparsas e elevada capacidade de customização por meio do ajuste de hiperparâmetros. Esse conjunto de características mostrou-se particularmente adequado às especificidades do problema e da base de dados analisada.

Em análises preliminares, também foi avaliado o algoritmo *CatBoost*, que dispõe de recursos nativos para o tratamento de variáveis categóricas. Contudo, observou-se que o tempo de processamento do *CatBoost* foi significativamente superior ao do *XGBoost*, aspecto relevante diante da necessidade de testar múltiplas configurações de modelos e hiperparâmetros. Ademais, os resultados preliminares obtidos com o *CatBoost* mostraram-se inferiores aos alcançados pelo *XGBoost*, o que reforçou a escolha deste último para a etapa de modelagem.

Conforme já apresentado no item 4.2, a variável alvo foi definida a partir de três classes distintas:

- a) Adimplente: houve algum pagamento em todos os anos em que ocorreu lançamento de IPTU ao longo da série histórica de 2019 a 2023
- b) Adimplente parcial: não houve pagamento algum em pelo menos um dos anos em que ocorreu lançamento de IPTU na série histórica de 2019 a 2023;
- c) Inadimplente: não houve qualquer pagamento nos anos em que ocorreu lançamento de IPTU ao longo da série histórica de 2019 a 2023.

Conforme também já detalhado no item 4, as variáveis explicativas utilizadas no modelo englobam atributos cadastrais, funcionais, urbanísticos e econômicos dos imóveis, incluindo variáveis numéricas (como área construída, área do terreno e valor do IPTU), variáveis *booleanas* (infraestrutura urbana e qualidade cadastral) e variáveis categóricas (tipologia, destinação, padrão construtivo, entre outras). A variável de identificação do imóvel foi removida do conjunto de dados, por não possuir valor preditivo.

O conjunto de dados foi dividido de forma estratificada, preservando a proporção das classes: 80% dos dados foram destinados ao conjunto de treino mais

validação e 20% foram reservados como conjunto de teste, utilizado apenas na etapa final de avaliação. No conjunto de treino e validação, foi aplicada validação cruzada estratificada em 5 *folds*, garantindo maior robustez na estimativa do desempenho do modelo.

O modelo *XGBoost* foi configurado para classificação multiclasse (*multi:softprob*). Devido ao forte desbalanceamento entre as classes da variável alvo, com predominância da classe inadimplente, foram utilizados pesos de classe do tipo “*balanced*”, calculados estatisticamente a partir da frequência das classes no conjunto de treino e validação.

Esses pesos foram incorporados ao treinamento do modelo por meio de pesos por amostra (*sample\_weight*), penalizando de forma mais intensa os erros nas classes minoritárias (adimplente e adimplente parcial). Como consequência, houve uma redução da acurácia global, mas uma melhora significativa nas métricas *F1*, especialmente: *F1-macro*, que avalia o desempenho médio entre as classes sem considerar seu tamanho, e *F1-weighted*, que pondera o desempenho pelo suporte de cada classe.

Esse comportamento é esperado e desejável em problemas desbalanceados, pois indica maior capacidade do modelo em reconhecer corretamente as classes minoritárias, em detrimento de uma leve perda na métrica de acurácia.

Após a realização das validações, o modelo que apresentou melhor desempenho foi configurado com os seguintes hiperparâmetros:

- a) *n\_estimators* = 700
- b) *learning\_rate* = 0,03
- c) *max\_depth* = 6

Na validação cruzada (*5-fold*), utilizando pesos balanceados, o modelo apresentou, em média:

- a) Acurácia (validação): 0,7428
- b) *F1-macro* (validação): 0,6119
- c) *F1-weighted* (validação): 0,7659

Esses resultados indicam bom equilíbrio entre desempenho global e capacidade discriminativa entre as classes.

Após a definição dos hiperparâmetros, o modelo foi treinado novamente utilizando todo o conjunto de treino mais validação e avaliado no conjunto de teste, obtendo:

a) Acurácia: 0,7430

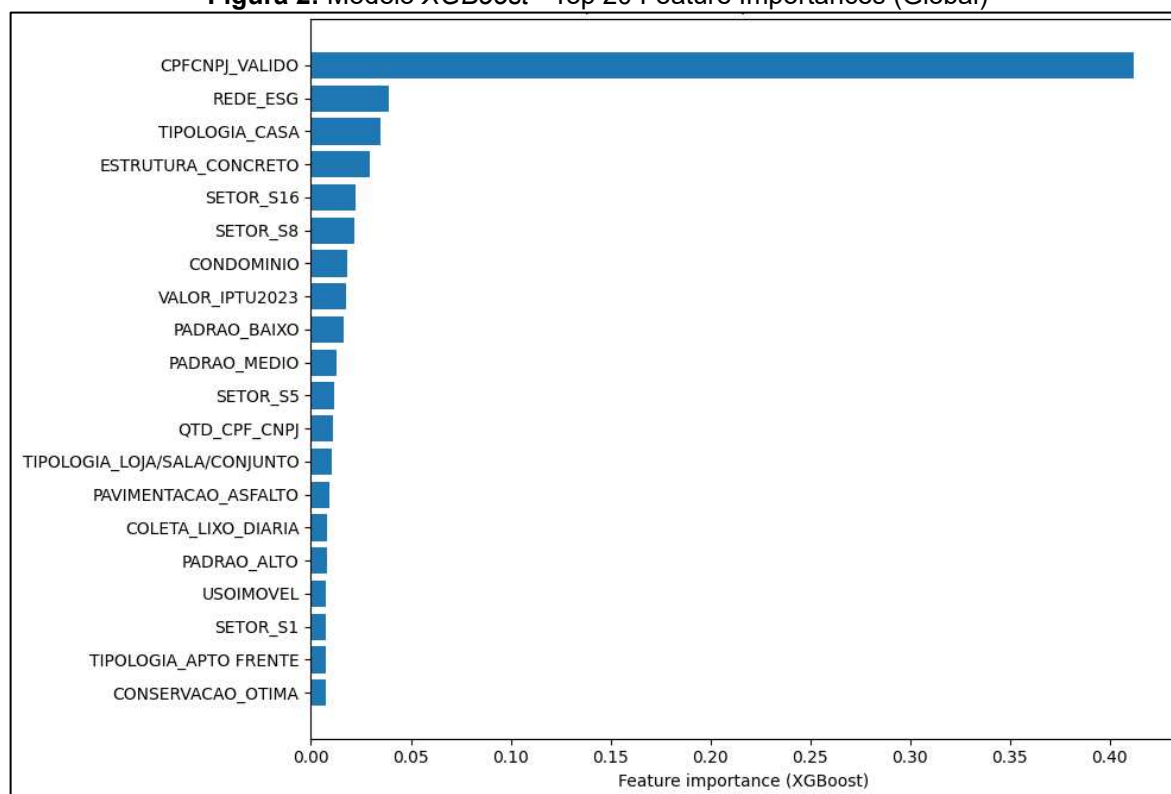
b) *F1-macro*: 0,6135

c) *F1-weighted*: 0,7662

A matriz de confusão e o relatório de classificação evidenciam que o modelo apresenta elevado desempenho na identificação da classe inadimplente, ao mesmo tempo em que melhora substancialmente a capacidade de identificar imóveis adimplentes e adimplentes parciais, em comparação aos testes realizados nos modelos sem balanceamento.

Observa-se que o modelo apresentou desempenho inferior na classe adimplente parcial, especialmente em termos de precisão e *F1-score*. Esse resultado, no entanto, é esperado e justificável, uma vez que essa classe possui natureza intermediária, concentrando características tanto da classe adimplente quanto da classe inadimplente. Em termos práticos, trata-se de um grupo de transição, no qual os padrões de comportamento de pagamento não são claramente definidos, o que dificulta a separação nítida por parte do modelo. Assim, acredita-se que o desempenho modesto nessa classe não indica uma limitação metodológica, mas reflete a própria complexidade e sobreposição de características presentes nos dados.

A análise de importância global das variáveis, obtida a partir do modelo *XGBoost*, evidencia os atributos que mais contribuíram para a classificação da situação de adimplência dos imóveis. Conforme Figura 2, a variável CPF/CNPJ válido destaca-se de forma expressiva como o principal fator explicativo do modelo, indicando a forte associação entre a qualidade do cadastro do contribuinte e o comportamento de pagamento do IPTU. Em seguida, seguem variáveis relacionadas à infraestrutura urbana e características construtivas do imóvel, como a presença de rede de esgoto, tipologia residencial, estrutura em concreto e padrão construtivo, além de atributos locais representados por setores específicos.

**Figura 2: Modelo XGBoost - Top 20 Feature Importances (Global)**

**Fonte:** Plotagem do Modelo XGBoost.

Destaca-se que a variável valor do IPTU de 2023, no modelo, embora relevante, apresenta importância relativa inferior às variáveis cadastrais e urbanísticas, sugerindo que a inadimplência está mais fortemente associada a aspectos de formalização, infraestrutura e uso do imóvel do que exclusivamente ao valor monetário do tributo. Esses resultados reforçam a relevância de políticas voltadas à qualificação cadastral e à gestão territorial como instrumentos centrais para a redução da inadimplência.

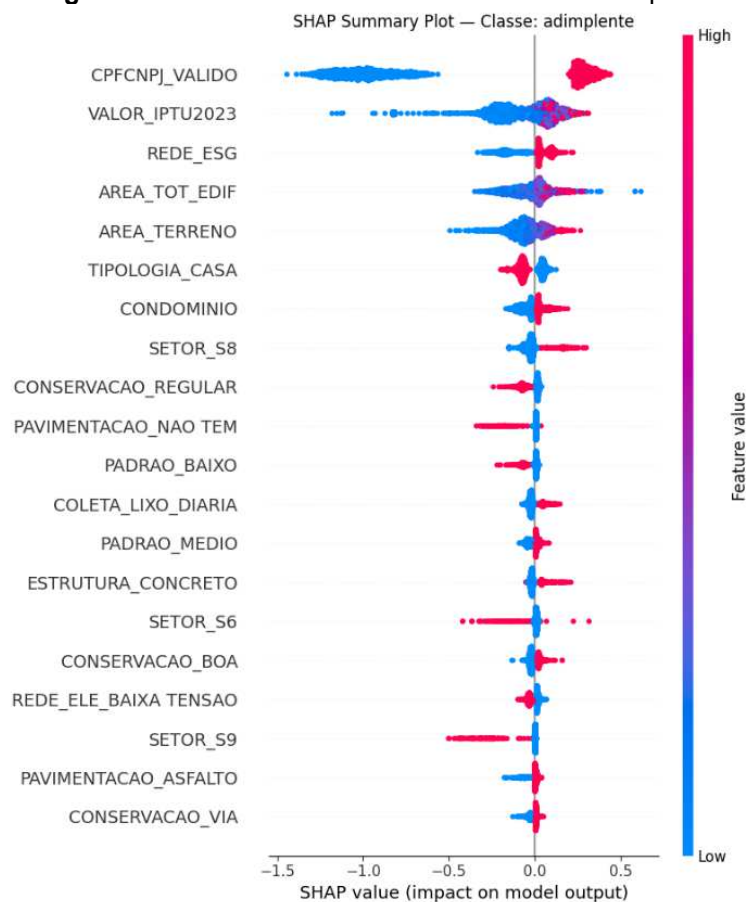
A importância das variáveis (*feature importance*) representa uma medida do quanto cada atributo contribui para as decisões do modelo durante o processo de classificação. No caso do XGBoost, essa importância é calculada com base na contribuição das variáveis nas divisões das árvores de decisão, considerando o quanto cada variável reduz a função de perda do modelo ao longo do treinamento. Em termos práticos, variáveis que são utilizadas com maior frequência e que promovem maiores reduções de erro nas divisões dos nós das árvores tendem a apresentar valores mais elevados de importância.

Os valores apresentados na figura 2 correspondem à importância normalizada das variáveis, de modo que a soma das importâncias de todas as variáveis é igual a 1. Assim, um valor próximo de 0,4, como observado para a variável CPF/CNPJ válido, indica que aproximadamente 40% da importância total do modelo está associada a essa variável, evidenciando seu papel predominante na classificação da inadimplência. As demais variáveis apresentam importâncias relativas menores, refletindo contribuições complementares ao processo decisório do modelo.

Conforme já discutido, a *feature importance* global do *XGBoost* indica quais variáveis foram mais utilizadas pelo modelo ao longo do treinamento, considerando principalmente a redução da função de perda promovida por cada atributo nas divisões das árvores. Trata-se de uma medida agregada e global, útil para identificar as variáveis mais relevantes no modelo como um todo. No entanto, essa abordagem não informa a direção do efeito (se a variável aumenta ou reduz a probabilidade de determinada classe) nem como esse impacto varia entre as classes.

Nesse contexto, a utilização do *SHAP* (*SHapley Additive exPlanations*) permite uma análise mais detalhada e interpretável. O *SHAP* calcula, para cada observação e para cada classe, quanto cada variável contribui positiva ou negativamente para a predição. Assim, além da relevância, o *SHAP* fornece sinal (positivo ou negativo) e intensidade do impacto, possibilitando compreender o comportamento do modelo de forma mais transparente e específica por classe.

Conforme análise da figura 3, para a classe adimplente, observa-se que a variável CPF/CNPJ válido exerce o maior impacto positivo na predição, indicando que a presença de identificação cadastral regular aumenta significativamente a probabilidade de o imóvel ser classificado como adimplente. Nota-se, ainda, a concentração de valores mais elevados de IPTU contribuindo positivamente para essa classe, assim como melhores condições de infraestrutura urbana, tais como a presença de rede de esgoto, algum tipo de pavimentação e coleta diária de lixo.

**Figura 3:** SHAP do Modelo XGBoost - Classe adimplente

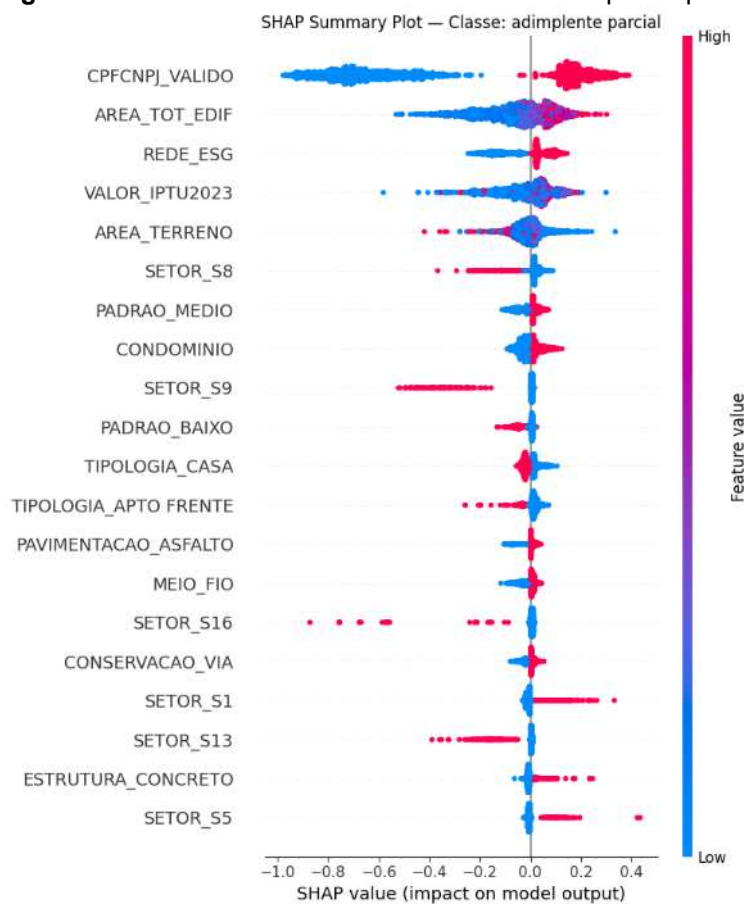
**Fonte:** Plotagem do Modelo XGBoost.

Continuando a análise da figura 3, variáveis relacionadas à maior área construída e à área do terreno também apresentam contribuição positiva para a adimplência. Em relação às características do imóvel, a tipologia casa reduz a probabilidade de adimplência, enquanto a localização em condomínio atua no sentido oposto, aumentando essa probabilidade. No que se refere à dimensão locacional, a presença no setor fiscal 8 eleva a probabilidade de adimplência, ao passo que os setores 6 e 9 estão associados à redução dessa probabilidade.

Por fim, a figura 3 revela que aspectos ligados ao estado de conservação e padrão construtivo também se mostram relevantes: imóveis com conservação regular e padrão baixo tendem a reduzir a probabilidade de adimplência, enquanto a conservação boa exerce efeito positivo. De forma geral, essas evidências revelam que a adimplência está associada, sobretudo, à melhor qualidade dos serviços e da infraestrutura urbana, bem como a características cadastrais e construtivas mais favoráveis dos imóveis.

Na classe adimplente parcial, o *SHAP*, figura 4 a seguir, evidencia um padrão mais difuso e intermediário, coerente com a natureza dessa categoria. A variável CPF/CNPJ válido permanece relevante para a predição, assim como maiores áreas construídas, que contribuem positivamente para essa classe. Por outro lado, as variáveis área do terreno e valor do IPTU apresentam impactos mais dispersos, sem um comportamento claramente definido.

**Figura 4:** SHAP do Modelo XGBoost - Classe adimplente parcial



**Fonte:** Plotagem do Modelo XGBoost.

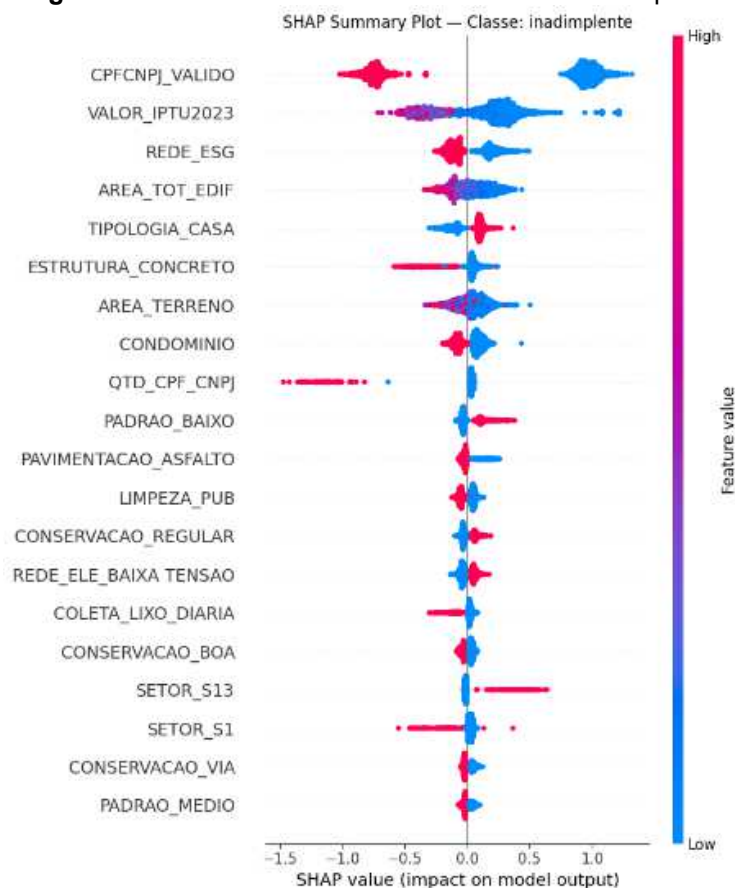
No que se refere à infraestrutura e aos serviços urbanos, a figura 4 mostra que a presença de rede de esgoto exerce contribuição positiva para a classificação como adimplente parcial. Adicionalmente, em termos locais, a presença nos setores fiscais 8, 9 e 13 está associada à redução da probabilidade de pertencimento a essa classe, enquanto a localização no setor 1 atua no sentido oposto, aumentando essa probabilidade. Quanto às características construtivas, o padrão médio e a

presença de imóveis em condomínio favorecem essa classe, ao passo que o padrão baixo atua no sentido contrário.

De modo geral, em relação à classe adimplente parcial, observam-se contribuições mistas, indicando que imóveis com boas características urbanas podem, ainda assim, apresentar comportamento irregular de pagamento. Esse padrão confirma que a classe adimplente parcial representa uma zona de transição, com sobreposição de atributos das classes adimplente e inadimplente.

De acordo com a figura 5, para a classe inadimplente, observa-se um comportamento claramente oposto ao da classe adimplente. A ausência de CPF/CNPJ válido exerce forte impacto no aumento da probabilidade de inadimplência. Na mesma esteira, a maior quantidade de CPF/CNPJ cadastrados por imóvel, indicativa de múltiplos proprietários, reduz a probabilidade de inadimplência, conforme evidenciado pelo *SHAP*. Esse resultado é condizente com a realidade tributária, uma vez que a existência de mais responsáveis legais pelo imóvel amplia as possibilidades de cobrança do imposto e tende a favorecer a regularização do débito.

**Figura 5:** *SHAP* do Modelo *XGBoost* - Classe inadimplente



Fonte: Plotagem do Modelo *XGBoost*.

Adicionalmente, valores mais baixos de IPTU, menores áreas edificadas e de terreno e níveis reduzidos de infraestrutura e serviços urbanos estão fortemente associados à inadimplência, destacando-se a ausência de: rede de esgoto, limpeza pública, pavimentação asfáltica e coleta diária de lixo. Em relação às características do imóvel, a tipologia casa está associada ao aumento da probabilidade de inadimplência, enquanto a localização em condomínio atua no sentido oposto.

Ainda acerca da figura 5, verifica-se que aspectos construtivos e de conservação do imóvel também se mostram relevantes: imóveis com padrão baixo e estado de conservação regular tendem a elevar a probabilidade de inadimplência, ao passo que a conservação boa contribui para sua redução. No componente locacional, a presença no setor fiscal 13 está associada ao aumento da inadimplência, enquanto a localização no setor 1 contribui para sua diminuição.

Em conjunto, esses resultados reforçam que a inadimplência do IPTU está associada a menores níveis de infraestrutura e serviços urbanos, bem como a fragilidades na qualidade cadastral, evidenciando a importância de políticas integradas de qualificação cadastral e melhoria urbana.

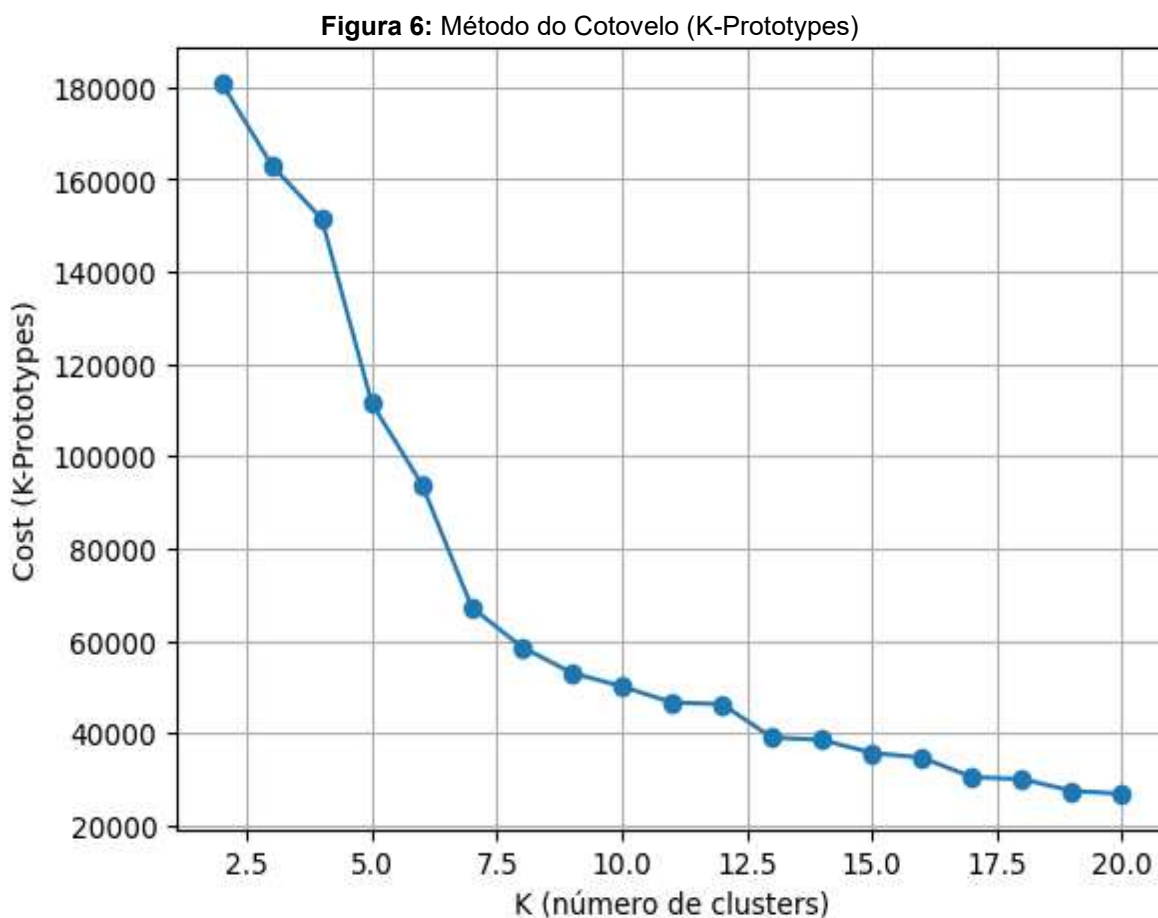
### 5.3 Clusterização

Nesta etapa do trabalho, foi realizada a análise de clusterização, explicada no item 3.3.4. O algoritmo adotado foi o *K-prototypes*, por ser adequado a bases de dados que combinam variáveis numéricas e categóricas, além de apresentar um bom custo computacional. O algoritmo HDBSCAN, baseado em densidade, foi testado de forma preliminar, contudo, não produziu agrupamentos que possibilitassem a extração de padrões relevantes para o problema em estudo, além de demandar maior tempo de processamento.

Conforme foi apresentado no item 4.3, o *dataset* utilizado para esta etapa possuía as colunas Inscrição Fictícia, Uso do Imóvel (0-Territorial e 1-Predial), Condomínio (0-NÃO e 1-SIM), Destinação (Comercial, Outra, Residencial e Territorial), CPF/CNPJ válido (0-NÃO e 1-SIM), Pontuação de Serviços/Infraestrutura (Variando de 0 a 92) e Valor do IPTU do ano de 2023 (Maior do que zero).

A Figura 6 apresenta a aplicação do método do cotovelo ao algoritmo *K-Prototypes*, utilizado como análise inicial (*baseline*) para a definição do número de

clusters. Esse procedimento foi realizado a partir de uma amostra correspondente a 50% da base de dados, com a utilização de cinco inicializações do algoritmo, em função do elevado custo e do tempo computacional associados à execução. Observa-se uma redução acentuada do custo intragrupo até aproximadamente 7 ou 8 clusters, seguida por um comportamento mais suavizado da curva, indicando retornos decrescentes com o aumento de K.



**Fonte:** Plotagem do Modelo *K-Prototypes*.

Apesar de o método do cotovelo fornecer uma referência inicial para a escolha do número de clusters, conforme discutido anteriormente, a etapa de clusterização não deve se restringir a critérios puramente métricos. Assim, foram realizadas análises exploratórias dos agrupamentos obtidos para diferentes configurações de K, variando de 5 a 13 clusters, com o objetivo de avaliar a consistência interna, a separabilidade e, sobretudo, a interpretabilidade dos clusters à luz das características cadastrais e financeiras dos imóveis.

A partir dessas análises, a configuração com 12 clusters mostrou-se a mais adequada, por apresentar melhor equilíbrio entre a distribuição dos imóveis entre os grupos e a capacidade de capturar padrões relevantes e interpretáveis, justificando sua adoção final neste estudo, conforme será apresentado a seguir.

Passa-se agora a discorrer acerca dos parâmetros utilizados na execução do modelo escolhido. Os parâmetros *n\_init* e *max\_iter* foram mantidos em seus valores padrão do algoritmo, 10 e 100, respectivamente, assegurando múltiplas inicializações para reduzir a sensibilidade à escolha inicial dos protótipos e limitando o processo iterativo de forma a garantir a convergência com custo computacional controlado.

O parâmetro *gamma* também foi mantido em seu valor padrão, sendo calculado automaticamente a partir dos dados, o que no final resultou em 0,5. Esse valor atribui maior peso às variáveis numéricas no cálculo da dissimilaridade, o que se mostra adequado ao contexto do estudo, uma vez que atributos como a pontuação de serviços e infraestrutura do imóvel e o valor do IPTU de 2023 são centrais para a caracterização dos agrupamentos.

O método de inicialização *init = "Cao"* foi mantido em seu valor padrão, por proporcionar uma inicialização mais estável dos protótipos em dados mistos. O CAO inicial é utilizado como referência para avaliar a convergência do algoritmo e para comparar as diferentes inicializações realizadas pelo parâmetro *n\_init*.

Por fim, foi fixada uma semente aleatória (*random\_state = 42*) com o objetivo de garantir a reprodutibilidade dos resultados, assegurando que o processo de inicialização dos protótipos e os agrupamentos obtidos possam ser reproduzidos em execuções futuras do modelo.

A Tabela 34 apresenta a caracterização dos *clusters* obtidos a partir da aplicação do algoritmo *K-prototypes* sobre a base de imóveis inadimplentes. Para cada agrupamento, são descritas as principais características cadastrais, funcionais e econômicas dos imóveis. A última linha da tabela representa o conjunto completo da base de inadimplentes analisada e é apresentada como referência para a comparação entre os *clusters*.

**Tabela 34:** Caracterização dos *clusters* obtidos a partir do algoritmo *K-prototypes*

<i>Cluster</i>	Imóvel Territorial (%)	Imóvel em Condomínio (%)	Tem CPF/CNPJ válido (%)	Destinação Residencial (%)	Destinação Comercial/ Outra (%)	Pontuação de Serviços/ Infraestrutura (Min: 0 / Máx: 92)	Soma Valor IPTU 2023	Quantidade de Imóveis	Média Valor IPTU 2023
0	0%	0%	12%	96%	4%	59,26	R\$ 11.688.648,53	48.536	R\$ 240,82
1	35%	13%	40%	28%	38%	68,26	R\$ 14.100.785,53	1.646	R\$ 8.566,70
2	40%	9%	55%	19%	41%	66,44	R\$ 6.756.470,78	216	R\$ 31.279,96
3	34%	7%	59%	17%	48%	68,31	R\$ 3.262.621,09	29	R\$ 112.504,18
4	100%	0%	16%	0%	0%	60,56	R\$ 5.142.697,60	10.687	R\$ 481,21
5	0%	100%	15%	94%	6%	62,98	R\$ 5.232.611,60	27.288	R\$ 191,76
6	7%	16%	10%	88%	5%	42,13	R\$ 7.021.722,75	37.916	R\$ 185,19
7	2%	1%	14%	93%	5%	73,40	R\$ 15.746.197,66	43.686	R\$ 360,44
8	40%	10%	13%	57%	3%	16,88	R\$ 3.955.561,02	14.776	R\$ 267,70
9	1%	84%	79%	84%	15%	79,13	R\$ 10.246.430,99	19.861	R\$ 515,91
10	0%	89%	22%	0%	100%	70,84	R\$ 4.971.248,17	10.221	R\$ 486,38
11	33%	33%	100%	0%	67%	42,67	R\$ 1.387.279,78	3	R\$ 462.426,59
Base Total	10%	28%	19%	80%	10%	59,20	R\$ 89.512.275,50	214.865	R\$ 416,60

**Fonte:** Elaborado pelo autor.

Como os *clusters* 1, 2, 3 e 11 apresentam reduzido número de imóveis, optou-se por agrupá-los em um único *cluster*. Destaca-se que, embora representem *outliers*, esses agrupamentos concentram informações relevantes, pois correspondem a imóveis que não se enquadram nos principais padrões identificados na base de dados, o que justifica uma análise diferenciada. Ressalta-se, ainda, que a presença desses *outliers* foi observada de forma consistente em todos os testes realizados com diferentes números de *clusters*, reforçando a robustez desse achado, conforme será detalhado a seguir.

A Tabela 35 apresenta a reorganização dos *clusters* finais, com destaque para suas principais características cadastrais, funcionais e econômicas, bem como para os indicadores associados à arrecadação do IPTU. Após a junção dos agrupamentos com reduzido número de imóveis, conforme no parágrafo anterior, a base passou a ser representada por 9 *clusters*, identificados pelas letras “A” a “I”. A ordenação dos *clusters* não é aleatória, tendo sido definida com o objetivo de subsidiar a priorização das ações da administração tributária municipal.

Tabela 35: Clusters reorganizados a partir de suas características

Cluster Inicial	Cluster Final	Características	Imóvel Territorial (%)	Imóvel em Condomínio (%)	Tem CPF/CNPJ válido (%)	Destinação Comercial/ Outra (%)	Pontuação de Serviços/ Infraestrutura (Min: 0 / Máx: 92)	Soma Valor IPTU 2023	Quant. de Imóveis	Média Valor IPTU 2023
1, 2, 3 e 11	A	- Maior média e soma do valor de IPTU - Poucos imóveis - Alta pontuação de serviços/infraestrutura	35%	12%	42%	38%	68,01	R\$ 25.507.157,18	1.894	R\$ 13.467,35
9	B	- Alta média do valor de IPTU - Maior pontuação de serviços/infraestrutura - Maior percentual de imóveis que possuem CPF/CNPJ válido - Alto percentual de imóveis em condomínio	1%	84%	79%	15%	79,13	R\$ 10.246.430,99	19.861	R\$ 515,91
10	C	- Alto percentual de imóveis em condomínio - Todos os imóveis possuem destinação comercial/outra - Alta média do valor de IPTU - Alta pontuação de serviços/infraestrutura	0%	89%	22%	100%	70,84	R\$ 4.971.248,17	10.221	R\$ 486,38
4	D	- Todos os imóveis são territoriais - Alta média do valor de IPTU	100%	0%	16%	0%	60,56	R\$ 5.142.697,60	10.687	R\$ 481,21
5	E	- Todos os imóveis pertencem a condomínio - Pontuação de serviços/infraestrutura mediana - Baixo valor médio de IPTU	0%	100%	15%	6%	62,98	R\$ 5.232.611,60	27.288	R\$ 191,76
7	F	- Alta pontuação de serviços/infraestrutura - Alta valor da soma de IPTU - Alta quantidade de imóveis - Valor médio do IPTU em nível intermediário	2%	1%	14%	5%	73,40	R\$ 15.746.197,66	43.686	R\$ 360,44
0	G	- Alto valor da soma de IPTU - Alta quantidade de imóveis - Pontuação de serviços/infraestrutura mediana	0%	0%	12%	4%	59,26	R\$ 11.688.648,53	48.536	R\$ 240,82
8	H	- Menor pontuação de serviços/infraestrutura - Quantidade de imóveis intermediária - Valor médio do IPTU em nível intermediário - Percentual de imóveis territoriais intermediário	40%	10%	13%	3%	16,88	R\$ 3.955.561,02	14.776	R\$ 267,70
6	I	- Menor valor da média de IPTU - Muitos imóveis - Baixa pontuação de serviços/infraestrutura	7%	16%	10%	5%	42,13	R\$ 7.021.722,75	37.916	R\$ 185,19

Fonte: Elaborado pelo autor.

A sequência adotada considera, prioritariamente, a média do valor do IPTU, por refletir diretamente o potencial arrecadatório; em seguida, a existência de CPF/CNPJ válido, indicador relevante de qualidade cadastral e de maior viabilidade de cobrança; a presença de imóveis em condomínio, associada à maior formalização e concentração de unidades; a pontuação de serviços e infraestrutura, como *proxy* da valorização urbana; e, por fim, a proporção de imóveis territoriais, especialmente relevante por seu enquadramento nos instrumentos de política urbana.

Nesse contexto, os imóveis territoriais assumem papel estratégico, uma vez que se relacionam diretamente com os dispositivos previstos no § 4º do art. 182 da Constituição Federal de 1988 e no art. 359 do Novo Código Tributário do Município de São Luís, os quais estabelecem mecanismos voltados ao cumprimento da função social da propriedade urbana, incluindo a aplicação do IPTU Progressivo no Tempo, bem como a obrigatoriedade do adequado aproveitamento do solo urbano não edificado, subutilizado ou não utilizado.

Conforme pode-se observar, a ordem de prioridade sugerida para atuação da administração tributária foi A, B, C, D, E, F, G, H e I, definida principalmente a partir do potencial arrecadatório e da viabilidade operacional para ações em massa, considerando aspectos como valor médio de IPTU, qualidade cadastral e facilidade de identificação dos responsáveis pelos imóveis.

No entanto, caso a administração opte por uma abordagem com maior ênfase social e urbanística, essa ordem de prioridade pode ser ajustada. Em especial, os *clusters* E e F podem ter suas posições invertidas, uma vez que o *cluster* F apresenta a segunda maior pontuação de serviços e infraestrutura, indicando áreas urbanas mais consolidadas.

Da mesma forma, a ordem dos *clusters* H e I pode ser invertida em estratégias voltadas à redução de desigualdades territoriais, tendo em vista que o *cluster* H apresenta a menor pontuação de serviços e infraestrutura entre os agrupamentos.

Ressalta-se que o *cluster* E foi priorizado na ordenação original por concentrar 100% dos imóveis em condomínios, característica que favorece ações de cobrança coletiva e reduz custos administrativos, uma vez que facilita a localização dos proprietários e a realização de notificações, inclusive por meio de comunicação com síndicos. Já o *cluster* H foi priorizado em função de apresentar maior valor médio de IPTU associado a uma menor quantidade de imóveis, o que representa, do ponto de vista arrecadatório, um melhor custo-benefício para a administração tributária.

Uma informação adicional de grande relevância é que os *clusters* A, B, C e D concentram mais de 50% do valor total da base de inadimplentes analisada, o que corresponde a R\$ 45.867.533,94, no ano de 2023. Esse montante refere-se a menos de 20% (42.663 inscrições) da quantidade total de imóveis inadimplentes (214.865) e a aproximadamente 11% da base total de imóveis analisada (383.507), evidenciando elevada concentração do potencial arrecadatório em um grupo relativamente reduzido de contribuintes.

Ressalta-se que os inadimplentes aqui considerados são aqueles que não efetuaram qualquer pagamento de IPTU no período de 2019 a 2023, caracterizando-se, portanto, como devedores contumazes. Nesse contexto, o valor acumulado para um período de cinco anos é por volta de R\$ 230 milhões, considerando apenas os contribuintes pertencentes aos *clusters* A, B, C e D.

Essa projeção parte do pressuposto de que, em razão do histórico de inadimplência observado no período analisado, grande parte desses imóveis também não realizou pagamentos nos exercícios de 2024 e 2025, bem como não deverá efetuar em 2026, exercícios para os quais ainda não se configurou a prescrição do crédito tributário, sendo possível a realização da cobrança pela Administração Tributária.

#### **5.4 Identificação de padrões atípicos e inconsistências cadastrais**

Não foi aplicado um algoritmo específico para detecção de anomalias, uma vez que o padrão de inadimplência observado em São Luís indica um contexto no qual fatores estruturais e sistêmicos tendem a prevalecer sobre ocorrências pontuais ou isoladas. Conforme evidenciado nos itens anteriores, inconsistências cadastrais, como a ausência de CPF ou CNPJ válido, manifestam-se de forma recorrente e disseminada na base de dados, não configurando, portanto, exceções individuais.

Ainda assim, a aplicação dos modelos de *XGBoost* e *K-Prototypes*, embora não concebida com o objetivo primário de detecção de anomalias, possibilitou a identificação de padrões atípicos de comportamento em relação ao conjunto majoritário de contribuintes. Esses resultados constituem um subsídio relevante para a formulação de ações mais direcionadas pela administração tributária, ao permitir a identificação de imóveis ou contribuintes cujo perfil diverge do comportamento esperado dentro de seus respectivos grupos.

No caso do *XGBoost*, o algoritmo realiza a classificação dos imóveis nas classes adimplente, adimplente parcial e inadimplente, com base nas probabilidades estimadas de pertencimento a cada uma dessas categorias. Como exemplo ilustrativo, para a inscrição fictícia nº 266313 (II266313), cujo rótulo verdadeiro era adimplente, o modelo a classificou como inadimplente, atribuindo as seguintes probabilidades: 0,0077 para adimplente, 0,0130 para adimplente parcial e 0,9794 para inadimplente.

Observa-se que a diferença entre a probabilidade associada à classe predita (inadimplente) e a segunda classe mais provável (adimplente parcial) foi de aproximadamente 0,97, indicando um elevado grau de confiança do modelo nessa predição, ainda que divergente do rótulo observado. No caso específico dessa inscrição, destacam-se características como a ausência de CPF/CNPJ válido, padrão construtivo baixo, conservação regular e tipologia casa, atributos que, conforme identificado nas análises anteriores, apresentam forte associação com o perfil de inadimplência. Ressalta-se, adicionalmente, que a segunda classe mais provável atribuída pelo modelo não correspondia ao rótulo verdadeiro do imóvel, reforçando a discrepância entre a classificação probabilística e a observação empírica.

A partir dessa abordagem, as probabilidades estimadas pelo *XGBoost* podem ser exploradas como um instrumento auxiliar para a identificação de possíveis inconsistências ou comportamentos atípicos, especialmente nos casos em que o modelo atribui elevada confiança a uma classe distinta daquela observada. Dessa forma, considerando as inscrições cuja diferença entre a probabilidade da classe predita e a segunda mais provável foi superior a 0,5, verificou-se que 1.111 imóveis foram classificados como inadimplentes pelo modelo, embora apresentassem rótulo verdadeiro adimplente. De forma análoga, 127 inscrições foram classificadas como adimplentes enquanto seu rótulo verdadeiro indicava inadimplência.

Divergências entre a classe predita pelo modelo e o rótulo observado, especialmente quando associadas a elevados níveis de confiança, podem ser utilizadas pela administração tributária como um critério técnico de apoio à priorização de ações administrativas. Esses casos podem orientar a verificação de inconsistências cadastrais, a atualização de informações do contribuinte, bem como estratégias diferenciadas de acompanhamento, comunicação ou negociação.

Já os resultados da aplicação do algoritmo não supervisionado, conforme discutido anteriormente, revelaram a formação de quatro *clusters* caracterizados por elevados valores de IPTU, concentrados em um número reduzido de imóveis. Nesses

agrupamentos, a média do valor do IPTU atingiu R\$ 13.467,35, montante significativamente superior à média geral da base analisada. Considerando que esses imóveis não apresentaram qualquer registro de pagamento ao longo da série histórica de 2019 a 2023, tais casos merecem atenção especial, uma vez que representam um potencial arrecadatório significativo para o município.

Além disso, 7.492 inscrições que apresentaram valores de IPTU ausentes ou igual a zero no exercício de 2023, conforme indicado na Tabela 1, item 4.1, demandam verificação por parte da administração tributária, uma vez que podem refletir inconsistências cadastrais.

Outrossim, imóveis que oscilaram entre situações com e sem benefício fiscal ao longo da série analisada, sendo essa condição passível de identificação por meio de filtros aplicados às colunas `qtd iptu` e `qtd_bnf_SIM` da planilha de pré-processamento descrita no item 4.2.

Ainda com base nessa mesma planilha de pré-processamento, identificaram-se 1.300 inscrições excluídas da análise em razão da ausência de informações relativas à infraestrutura e aos serviços urbanos, ou da inexistência de dados sobre a variável `DESTINACAO` ou da presença de valores nulos ou iguais a zero para `AREA_TERRENO`. Tais casos também merecem atenção específica, por indicarem potenciais falhas ou lacunas no cadastro imobiliário municipal.

## 6 CONCLUSÃO

Os resultados desta pesquisa evidenciam que a inadimplência do IPTU no município de São Luís apresenta caráter estrutural e persistente, associada a fatores cadastrais, urbanísticos, construtivos e territoriais, mais do que exclusivamente ao valor monetário do tributo.

Entre os contribuintes sem benefício fiscal, a análise descritiva da base financeira e cadastral revelou elevados níveis de inadimplência ao longo da série histórica de 2019 a 2023, com significativa perda arrecadatória ao final do prazo prescricional.

A aplicação do modelo supervisionado *XGBoost* permitiu identificar, de forma robusta e interpretável, os principais fatores associados à adimplência e à inadimplência do IPTU. Destacou-se a qualidade cadastral, em especial a validade do CPF/CNPJ do proprietário, como o principal determinante do comportamento de

pagamento, seguida por variáveis relacionadas à infraestrutura urbana, às características construtivas dos imóveis e à localização espacial.

A utilização de técnicas explicativas, como a análise de importância de variáveis e o *SHAP*, ampliou a transparência do modelo e reforçou a compreensão dos mecanismos subjacentes às previsões, evidenciando que fragilidades cadastrais e *déficits* de infraestrutura estão fortemente associados à inadimplência.

De forma complementar, a etapa de clusterização com o algoritmo *K-prototypes* permitiu estruturar a heterogeneidade dos imóveis inadimplentes em grupos interpretáveis, revelando elevada concentração do potencial arrecadatório em poucos *clusters*, compostos por imóveis de maior valor médio de IPTU e maior viabilidade operacional de cobrança. A ordenação desses agrupamentos forneceu subsídios objetivos para a priorização de ações da administração tributária, demonstrando que estratégias orientadas por dados podem ampliar a eficiência da arrecadação sem a necessidade de elevação de alíquotas.

Adicionalmente, a identificação de padrões atípicos e inconsistências cadastrais, como divergências classificatórias de alta confiança, valores de IPTU ausentes ou zerados e registros incompletos, reforçou a utilidade prática dos modelos como instrumentos de apoio à qualificação do cadastro imobiliário.

Do ponto de vista científico, a pesquisa contribui ao integrar métodos de ciência de dados e aprendizado de máquina à análise fiscal municipal, suprimindo uma lacuna ainda pouco explorada na literatura, especialmente no contexto de tributos imobiliários em países em desenvolvimento. A abordagem metodológica adotada, combinando modelos supervisionados, não supervisionados e técnicas explicativas, mostra-se replicável em outros municípios brasileiros e em contextos internacionais, ampliando o alcance e a relevância da investigação.

No plano social e institucional, os resultados oferecem subsídios para a formulação de políticas tributárias mais equitativas e eficazes, ao permitir a focalização de ações administrativas, a correção de distorções cadastrais e o fortalecimento da justiça fiscal. Ao orientar a atuação do poder público com base em evidências, o estudo contribui para o aprimoramento da gestão urbana e para a ampliação da capacidade de investimento em infraestrutura e serviços, especialmente em áreas mais vulneráveis.

Por fim, no aspecto econômico e tecnológico, a pesquisa demonstra o potencial das ferramentas computacionais e dos algoritmos de aprendizado de

máquina como vetores de modernização da administração pública, promovendo ganhos arrecadatórios sustentáveis e reforçando o papel da informação como elemento central no desenvolvimento urbano e fiscal.

Como desdobramento desta pesquisa, estudos futuros podem aprofundar a análise da inadimplência do IPTU por meio da incorporação de novas informações cadastrais e fiscais, bem como da ampliação do horizonte temporal para exercícios posteriores a 2023. A utilização de abordagens adicionais de aprendizado de máquina pode contribuir para capturar mudanças no comportamento de pagamento ao longo do tempo e para refinar a identificação de padrões complexos.

Adicionalmente, técnicas de inteligência artificial apresentam elevado potencial para a melhoria da qualidade cadastral, especialmente por meio do cruzamento automatizado da base do IPTU com dados de outras instituições. A integração de dados espaciais mais detalhados, incluindo imagens de sensoriamento remoto e informações urbanísticas georreferenciadas, pode enriquecer as análises e fortalecer a compreensão das relações entre inadimplência tributária, organização do território e desigualdades socioespaciais.

## REFERÊNCIAS BIBLIOGRÁFICAS

AFONSO, J. R. R.; ARAÚJO, E.; Nóbrega, M. **IPTU no Brasil: um diagnóstico abrangente**. Instituto Brasiliense de Direito Público (IDP), 2009.

ALVARENGA, Lucas. **Fiscalização tributária: uma análise dos autos de infração da SEFAZ-DF**. 2025. 105 f. Dissertação (Programa Stricto Sensu em Governança, Tecnologia e Inovação) - Universidade Católica de Brasília, Brasília, 2025.

ALVES, Elioenai L. G.; VIANA, Luiz Zairo Bastos; SANTOS, Fabio; FRANCO, Wellington; PINHEIRO, Vlândia. Leveraging Property Tax Collection Through an Unsupervised Model for Taxpayer Qualification. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL (ENIAC), 20. 2023, Belo Horizonte/MG. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2023 . p. 1142-1156. ISSN 2763-9061. DOI: <https://doi.org/10.5753/eniac.2023.234637>.

ARAÚJO, Douglas da Silva; GUIMARÃES, Patrícia Borba Vilar; COSTA, Ademir Araújo da. **A implantação de cidades inteligentes no Nordeste brasileiro: um breve diagnóstico / The implementation of intelligent cities in the brazilian northeast: a brief diagnosis**. Revista de Direito da Cidade, [S. l.], v. 12, n. 2, p. 1084–1104, 2020. DOI: 10.12957/rdc.2020.39957. Disponível em: <https://www.e-publicacoes.uerj.br/rdc/article/view/39957>. Acesso em: 28 set. 2025.

BACH, M. P. Data Mining Applications in Public Organizations. **Proceedings of the 25th International Conference on Information Technology Interfaces**, Zagreb, jul. 2003. Disponível em: <[https://www.researchgate.net/profile/Mirjana-Pejic-Bach/publication/4031742\\_Data\\_mining\\_applications\\_in\\_public\\_organizations/links/54d788f50cf2970e4e7398e0/Data-miningapplications-in-public-organizations.pdf](https://www.researchgate.net/profile/Mirjana-Pejic-Bach/publication/4031742_Data_mining_applications_in_public_organizations/links/54d788f50cf2970e4e7398e0/Data-miningapplications-in-public-organizations.pdf)>. Acesso em: 31 mar. 2025

BARTASSON FERREIRA ROSA, B.; PEREIRA, V. S. .; PENEDO, A. S. T. . **Dívida ativa de IPTU do Município de Uberlândia/MG: a região de residência influencia a inadimplência? does the region of residence influence default?. REUNIR**

Revista de Administração Contabilidade e Sustentabilidade, [S. l.], v. 13, n. 1, p. 95-111, 2023. DOI: 10.18696/reunir.v13i1.1413. Disponível em: <https://reunir.revistas.ufcg.edu.br/index.php/uacc/article/view/1413>. Acesso em: 28 set. 2025.

BERNARDES, José Flávio Felix. **Modelando a inadimplência do IPTU no Município de Fortaleza**. 2020. 42f. Dissertação (Mestrado em Economia do Setor Público) - Faculdade de Economia, Administração, Atuária e Contabilidade - FEAAC, Programa de Economia Profissional - PEP, Universidade Federal do Ceará - UFC, Fortaleza (CE), 2020.

BORGES, D. F.; SOUZA, R. M. S. de. **Iniquidade administrativa do imposto predial e territorial urbano no município de Natal**. Revista de Administração Pública, Rio de Janeiro, v. 38, n. 1, p. 57 a 78, 2004. Disponível em: <https://periodicos.fgv.br/rap/article/view/6527>. Acesso em: 28 set. 2025.

BORGES, Leandro & Almeida, Mauricio. (2024). **Ciência de dados: uma análise de conteúdo da produção científica em Ciência da informação**. Biblionline. 20. 46-66. 10.22478/ufpb.1809-4775.2024v20n3.69948.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/Constituicao/Constituicao.htm](http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm). Acesso em: 28 set. 2025.

BRASIL. Lei nº 5.172 de 25 de outubro de 1966. **Código Tributário Nacional**. Brasília, DF, [2025]. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/leis/l5172compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/l5172compilado.htm). Acesso em: 28 set. 2025.

CARVALHO JÚNIOR, P. H. B. de. **A administração tributária do IPTU e seu impacto na efetivação do estatuto da cidade**. In: Costa, M. A. O Estatuto da cidade e a habitat III: um balanço de quinze anos da política urbana no Brasil e a nova agenda urbana. Brasília: Ipea, p. 207-231, 2016.

CARVALHO JÚNIOR, Pedro Humberto Bruno. **O papel da inadimplência na heterogeneidade da arrecadação do IPTU no Brasil**. Boletim Regional, Urbano e Ambiental, Brasília, n. 18, jan.-jun. 2018. Disponível em: [https://repositorio.ipea.gov.br/bitstream/11058/8458/1/BRU\\_n18.pdf](https://repositorio.ipea.gov.br/bitstream/11058/8458/1/BRU_n18.pdf).

CASTRO, K. P. de; AFONSO, J. R. R. **IPTU: avaliação de potencial e utilização sob a ótica da teoria dos conjuntos fuzzy**. Revista de Administração Pública, Rio de Janeiro, v. 51, n. 5, p. 828–853, 2017. Disponível em: <https://periodicos.fgv.br/rap/article/view/72427>. Acesso em: 28 set. 2025.

CHAVEZ, L.; SALINAS, J. **Segmentación de los alumnos ingresantes a una universidad pública aplicando el algoritmo K-prototype**. Tierra nuestra. 15. 10-21. 2021. 10.21704/rtn.v15i2.1825.

CHENNIAPPANADAR, S. K. .; GNANAMURTHY, S. .; SAKTHIVELU, V. K. .; KALIAPPAN, V. K. . **A Supervised Machine Learning Based Intrusion Detection Model for Detecting Cyber-Attacks Against Computer System**. International Journal of Communication Networks and Information Security (IJCNIS), [S. l.], v. 14, n. 3, p. 16–25, 2022. DOI: 10.17762/ijcnis.v14i3.5567. Disponível em: <https://www.ijcnis.org/index.php/ijcnis/article/view/5567>. Acesso em: 28 sep. 2025.

CHEN, Tianqi; GUESTRIN, Carlos. **XGBoost: A Scalable Tree Boosting System**. 785-794. 2016. DOI: 10.1145/2939672.2939785.

COSTA, A. P.; PECINI, A. C.; FUKUMI TSUNODA, D. **A descoberta de padrões por meio da mineração de dados no IPTU de Curitiba**. Revista do Serviço Público, [S. l.], v. 72, n. 4, p. 753-778, 2021. Disponível em: <https://revista.enap.gov.br/index.php/RSP/article/view/6063>. Acesso em: 28 set. 2025.

CROCCO, P.F., et. al. **“Clusterização de Dados Abertos em Oncologia Usando Técnicas de Aprendizado de Máquina: um estudo preliminar sobre recidiva de câncer de mama”**. Jornada de Iniciação Científica do CTI Renato Archer, XXV JICC

– 2023.URL: <https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/jicc/xxv-jicc-2023/pdf/jicc-2023-paper-25.pdf>

DE CESARE, C. M.; FERNANDES, C. E.; OLIVEIRA, G. S.; CARVALHO JÚNIOR, P. H. B. **IPTU e a PEC 110/2019 (Reforma Tributária): teoria, prática, insensatez e retrocesso**. Revista da ESDM, Porto Alegre, v. 6, n. 11, p. 72-93, 2020. Disponível em: <<http://revista.esdm.com.br/index.php/esdm/article/view/134/106>>. Acesso em: 31 mar. 2025.

DOS ANJOS, Pedro Germano; PINHEIRO, Marcelo Teles Silva. **A implementação da inteligência artificial (IA) na fiscalização tributária: inovações disruptivas para eficiência na arrecadação do IPTU**. Revista Tributária e de Finanças Públicas – RTRIB, n. 159, 2023. Disponível em: <https://rtrib.abdt.org.br/index.php/rftp/article/view/719>. Acesso em: 28 set. 2025.

Ezeife E, Kokogho E, Odio PE, Adeyanju MO. **The future of tax technology in the United States: A conceptual framework for AI-driven tax transformation**. International Journal of Multidisciplinary Research and Growth Evaluation. 2021;2(1):542-551. <https://doi.org/10.54660/IJMRGE.2021.2.1.542-551>

FERREIRA, Fabiano Rodrigues. **Metodologia para reconhecimento de perfil cognitivo: uma abordagem computacional pautada pela teoria das inteligências múltiplas**. 2015. Tese (Doutorado em Engenharia de Computação) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2015. doi:10.11606/T.3.2016.tde-21062016-142027. Acesso em: 2026-02-05.

FRANCO, Wellington; ALVES, Elioenai; SOUSA, Fábio; BASTOS, Zairo; PINHEIRO, Vlândia. **An Intelligent Model for Generating Indications of Tax Gap in Service Companies**. In: BRAZILIAN WORKSHOP ON ARTIFICIAL INTELLIGENCE IN FINANCE (BWAIF), 2. , 2023, João Pessoa/PB. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2023 . p. 61-72. DOI: <https://doi.org/10.5753/bwaif.2023.230149>.

GERING, Silvia Maria Parmeggiani; PINTO, Nelson Guilherme Machado; VIEIRA, Kelmara Mendes. **Determinantes da Inadimplência do IPTU no Município de Santa Maria/RS: DETERMINANTS OF THE DEFAULT OF THE IPTU IN THE MUNICIPALITY OF SANTA MARIA/RS.** Desenvolvimento em Questão, [S. l.], v. 19, n. 55, p. 115–131, 2021. DOI: 10.21527/2237-6453.2021.55.10665. Disponível em: <https://www.revistas.unijui.edu.br/index.php/desenvolvimentoemquestao/article/view/10665>. Acesso em: 28 set. 2025.

GUIMARÃES, Tiago Rodrigues. **Análise dos fatores que contribuem para a recuperação dos créditos de ICMS inscritos em dívida ativa no estado do Rio de Janeiro.** 2023. 76 f. Dissertação (Mestrado em Controladoria e Gestão Pública) - Faculdade de Administração e Finanças, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

IBGE: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2022: População e Domicílios: Resultados Preliminares: Atualização da População dos Municípios.** Rio de Janeiro: IBGE, 2023. Disponível em: <https://censo2022.ibge.gov.br>. Acesso em: 30 mar. 2025.

Khalid, Sara & prieto-alhambra, Daniel. (2019). **Machine Learning for Feature Selection and Cluster Analysis in Drug Utilisation Research.** Current Epidemiology Reports. 6. 10.1007/s40471-019-00211-7.

Khan, Danish, et al. "**Application of GIS for Property Tax Assessment: A Case Study of Shillong Municipal Board**" Journal of Landscape Ecology, vol. 0, no. 0, Czech Society for Landscape Ecology, 2025, <https://doi.org/10.2478/jlecol-2026-0003>

Lee, Y. H., & Kim, E. (2024). **Deep Learning-Based Delinquent Taxpayer Prediction: A Scientific Administrative Approach.** KSII Transactions on Internet and Information Systems, 18(1), 30–45. <https://doi.org/10.3837/tiis.2024.01.003>

Lozano, Augusto & Cezar, Augusto & Lozano, Garcia & Ippolito, André. (2021). **Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo.**

MARCELO, Diego Asensi Tavares. **Parametrização de despesas municipais e detecção de anomalias.** Rio de Janeiro, 2021. 75 p. Dissertação de Mestrado em Economia Empresarial e Finanças - Fundação Getúlio Vargas, Rio de Janeiro, 2021.

MATOS, M. T. de; CONDURÚ, M. T.; BENCHIMOL, A. C. **Interseções na produção científica da ciência da informação e ciência de dados.** Acervo, [S. l.], v. 35, n. 2, p. 1–18, 2022. DOI: 10.64729/an.acervo.v35i2.1804. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1804>. Acesso em: 28 set. 2025.

MATOS, Renata Assis de. **Comparação de metodologias de análise de agrupamentos na presença de variáveis categóricas e contínuas;** Dissertação (Mestrado em Estatística) – Programa de Pós- Graduação em Estatística, UFMG, Belo Horizonte, 2007.

MENG, Y. et al. What Makes an Online Review More Helpful: **An Interpretation Framework Using XGBoost and SHAP Values.** Journal of Theoretical and Applied Electronic Commerce Research, 16(3), 466-490. 2021.  
<https://doi.org/10.3390/jtaer16030029>

Morettin, Pedro A.; Singer, Julio M.. (2020), '**Introdução à ciência de dados fundamentos e aplicações versão parcial preliminar.**', Departamento de Estatística Universidade de São Paulo.

MÜLLER , César Santini; DA SILVEIRA , Daniel Gomes Batista; CADEMARTORI , Luiz Henrique Urquhart. **NOVO FEDERALISMO FISCAL BRASILEIRO.** ARACÊ , [S. l.], v. 7, n. 8, p. e7328 , 2025. DOI: 10.56238/arev7n8-152. Disponível em: <https://periodicos.newsciencepubl.com/arace/article/view/7328>. Acesso em: 28 set. 2025.

Nogueira, Ana & Gama, João & Ferreira, Carlos. (2020). **Improving Prediction with Causal Probabilistic Variables**. 10.1007/978-3-030-44584-3\_30.

OLIVEIRA, Antônio Augusto Ferreira de. **Avaliação em massa com modelos de aprendizado de máquina aplicados aos terrenos urbanos do Município de Fortaleza**. 2020. 79f. Dissertação (Mestrado em Economia do Setor Público) - Faculdade de Economia, Administração, Atuária e Contabilidade - FEAC, Programa de Economia Profissional - PEP, Universidade Federal do Ceará - UFC, Fortaleza (CE), 2020.

Oliveira, Vinícius & Chaim, Ricardo & Weigang, Li & Neto, Sergio & Filho, Geraldo. (2021). **Towards a Smart Identification of Tax Default Risk with Machine Learning**. 422-429. 10.5220/0010712200003058.

PACHECO, Bornieque Brister Marcovitz; DISCONZI, Marcelo Salton. **Ciência de Dados: Enfoque no Desafio do Processamento**. Res., Soc. Dev. 2019; 8(11):e128111444 ISSN 2525-3409 | DOI: <http://dx.doi.org/10.33448/rsd-v8i11.1444>, [s. l.]. Acesso em: 28 set. 2025.

Pamisetty, Vamsee. (2025). **AI-Driven Decision Support for Taxation and Unclaimed Property Management : Enhancing Efficiency through Big Data and Cloud Integration**. SSRN Electronic Journal. 10.2139/ssrn.5250776.

PICCIRILLI, Tiago Levergger. **Mineração de dados aplicada a classificação dos contribuintes do ISS**. 2013. 134 f. Dissertação (Mestrado em Engenharia) - Pontifícia Universidade Católica de Goiás, GOIÂNIA, 2013.

Rodrigues, Luiz Alberto. **Cidades inteligentes em perspectiva: sem gente a cidade é uma triste ficção** / Luiz Alberto Rodrigues. – 1. ed. – Santo André: Coopacesso, 2022. 256 p.

SÃO LUÍS. Lei Nº 6289 DE 28/12/2017. **Novo Código Tributário do Município de São Luís**. Disponível em: <https://www.legisweb.com.br/legislacao/?id=357666>. Acesso em: 28 set. 2025.

Shu, X., & Ye, Y. (2023). **Knowledge Discovery: Methods from data mining and machine learning**. SocialScienceResearch, 110.  
<https://doi.org/10.1016/j.ssresearch.2022.102817>

SOARES, Gustavo Fleury. **Ciência de dados aplicada à Auditoria Interna**. Revista da CGU, [S. l.], v. 12, n. 22, p. 196–208, 2020.  
DOI: 10.36428/revistadacgu.v12i22.195. Disponível em: [https://revista.cgu.gov.br/Revista\\_da\\_CGU/article/view/195](https://revista.cgu.gov.br/Revista_da_CGU/article/view/195). Acesso em: 28 set. 2025.

Varsano, R. **O imposto predial e territorial urbano: receita, equidade e adequação aos municípios**. Pesq. Plan. Econ. Rio de Janeiro, p. 581-662, 1977.  
Disponível em: <<http://repositorio.ipea.gov.br/handle/11058/6868?mode=full>>.  
Acesso em: 31 mar. 2025.

VETTORATO, Gustavo; AFONSO, José Roberto Rodrigues; FUCK, Luciano Felício. **TRANSPARÊNCIA TRIBUTÁRIA E EFICIÊNCIA ARRECADATÓRIA: ANÁLISE DE DADOS EMPÍRICOS**. Revista de Direito Brasileira, Florianopolis, Brasil, v. 27, n. 10, p. 92–111, 2021. DOI: 10.26668/IndexLawJournals/2358-1352/2020.v27i10.6179.  
Disponível em: <https://www.indexlaw.org/index.php/rdb/article/view/6179>. Acesso em: 28 set. 2025.

Wang, J; Hu, X.; Zhu, D. **Data mining applications in public administration**. In: Rahman, H. Social and Political Implications of Data Mining: Knowledge Management in E-Government. IGI Global, cap. 18. p. 302-314, 2009.



*Emitido em 08/03/2026*

**DOCUMENTOS COMPROBATÓRIOS Nº 266/2026 - DPARQ (11.14.68.07.05)**

**(Nº do Protocolo: NÃO PROTOCOLADO)**

*(Assinado digitalmente em 12/03/2026 17:16)*

**CRISTIANE DE JESUS PEREIRA GASPAR**

*SECRETARIO III*

*866500*

Para verificar a autenticidade deste documento entre em <https://sis.sig.uema.br/documentos/> informando seu número:  
**266**, ano: **2026**, tipo: **DOCUMENTOS COMPROBATÓRIOS**, data de emissão: **12/03/2026** e o código de  
verificação: **327831e34d**

