

A descoberta de padrões por meio da mineração de dados no IPTU de Curitiba

Ana Paula Costa

André Custódio Pecini

Denise Fukumi Tsunoda

Universidade Federal do Paraná (UFPR), Curitiba – PR, Brasil

O presente artigo detalha o uso de mineração de dados sobre os dados cadastrais e de pagamento do Imposto Predial e Territorial Urbano (IPTU) do município de Curitiba, referente aos meses de janeiro a junho de 2020. Com esse propósito, examina-se uma base de dados construída em planilha *Excel*, relacionando o pagamento de tal imposto a diversos atributos referentes às características dos imóveis ou dos terrenos não edificados. Para efetuar a análise, utiliza-se um algoritmo de aprendizado não supervisionado para descoberta de regras de associação (*Apriori*) no software *Weka* (*Waikato Environment for Knowledge Analysis*). Os resultados indicam que ao associar os diferentes atributos de cadastro e pagamento, o pagamento do tributo é feito de forma parcelada, independentemente do padrão de construção do imóvel ou bairro. Outro resultado obtido foi que a inadimplência é maior em imóveis residenciais de madeira ou alvenaria, de padrão de acabamento: simples e média simples. Dessa forma, diante dos resultados obtidos, sugere-se que sejam realizadas ações por parte da gestão pública municipal no sentido de incentivar o uso do débito automático e o pagamento do tributo à vista pelos contribuintes, o que seria muito importante para o fluxo de caixa municipal e poderia minimizar a inadimplência do pagamento deste tributo.

Palavras-chave: descoberta de conhecimento em bases de dados, mineração de dados, Imposto Predial e Territorial Urbano, Curitiba

DOI: <https://doi.org/10.21874/rsp.v72.i4.6063> | ISSN: 0034-9240 | E-ISSN: 2357-8017

[Submetido: 27 de julho de 2021. Aprovado: 23 de setembro de 2021.]



El descubrimiento de patrones a través de la minería de datos en el IPTU de Curitiba

Este artículo detalla el uso de la minería de datos sobre los datos de registro y pago del Impuesto sobre la Propiedad Urbana y Territorio (IPTU) en la ciudad de Curitiba, para los meses de enero a junio de 2020. Para ello, se examina una base de datos construida en una hoja de cálculo Excel, relacionando el pago de dicho impuesto a diversos atributos relacionados con las características de las propiedades o de los terrenos no edificados. Para realizar el análisis se utiliza un algoritmo de aprendizaje no supervisado para el descubrimiento de reglas de asociación (Apriori) en el software Weka (Waikato Environment for Knowledge Analysis). Los resultados indican que, al asociar los diferentes atributos de registro y pago, el pago del impuesto se realiza en cuotas, independientemente del estándar de construcción del inmueble o barrio. Otro resultado obtenido fue que la morosidad es mayor en las propiedades residenciales de madera o mampostería, con un patrón de acabado: simple y promedio simple. Así, dados los resultados obtenidos, se sugiere que la administración pública municipal adopte acciones para incentivar el uso del débito automático y el pago del impuesto en efectivo por parte de los contribuyentes, lo que sería muy importante para el flujo de caja y podría minimizar la morosidad en el pago de este impuesto.

Palabras clave: descubrimiento de conocimiento en bases de datos, minería de datos, Impuesto Territorial y Urbano, Curitiba

The discovery of patterns through data mining at IPTU in Curitiba

This article details the use of data mining on registration data and payment of the Urban Property and Land Tax (IPTU) in the city of Curitiba, for the months of January to June 2020. For this purpose, a database built in an Excel spreadsheet is examined, relating the payment of such tax to various attributes related to the characteristics of properties or unbuilt land. To carry out the analysis, an unsupervised learning algorithm for discovery of association rules (Apriori) in the Weka software (Waikato Environment for Knowledge Analysis) is used. The results indicate that by associating the different attributes of registration and payment, the payment of the tax is made in installments, regardless of the pattern of construction of the property or neighborhood. Another result obtained was that delinquency is higher in residential properties made of wood or masonry, with a finishing pattern: simple and simple average. Thus, given the results obtained, it is suggested that actions be taken by the municipal public administration to encourage the use of automatic debit and the payment of tax in cash by taxpayers, which would be very important for the cash flow and could minimize default in the payment of this tax.

Keywords: knowledge discovery in databases, data mining, Land and Urban Tax, Curitiba

1 Introdução

A competência tributária conferida pela Constituição Federal¹ (CF) de 1988 permitiu que os entes federativos instituíssem em seus territórios determinados tributos.

O Imposto Predial e Territorial Urbano (IPTU) também está disposto nos artigos 32, 33 e 34 do Código Tributário Nacional (CTN), que regula o Sistema Tributário Nacional, estabelecendo as diretrizes gerais do IPTU, como fato gerador, base de cálculo e contribuinte do imposto. Seu fato gerador é a propriedade, o domínio útil ou a posse de bem imóvel por natureza ou por acessão física, localizado na zona urbana do município (BRASIL, 1966). A base de cálculo é o valor venal, ou seja, o valor de venda dos imóveis, e o contribuinte do imposto é o proprietário do imóvel, o titular do seu domínio útil, ou o seu possuidor a qualquer título (BRASIL, 1966).

Segundo Varsano (1977), 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, posto que possui pelo menos três vantagens: (i) não é objeto de conflitos tributários com demais municípios; (ii) o contribuinte do imposto só pode deixar de pagá-lo se transferir a propriedade e, conseqüentemente, o tributo; (iii) como o valor das propriedades é relativamente estável, a receita advinda do imposto não fica extremamente sujeita às oscilações econômicas. Esta última vantagem é também citada por De Cesare *et al.* (2020), que, mesmo diante da crise econômica ocasionada pela Covid-19, o IPTU se mostra uma fonte de recursos significativa e constante para assegurar a manutenção dos serviços públicos básicos.

Por outro lado, conforme apontado por Afonso, Araújo, Nóbrega (2009), o IPTU não constitui a principal fonte de arrecadação dos municípios. Isso se dá por dois motivos principais. Primeiro, pela escassez de recursos humanos e técnicos imprescindíveis para criar e manter uma estrutura administrativa com o objetivo de obter o potencial máximo de arrecadação do tributo. Em segundo lugar, decorrente da pressão realizada pelos contribuintes para que o valor venal das suas propriedades não seja alterado, causando uma maior vulnerabilidade política nos municípios (AFONSO; ARAÚJO; NÓBREGA, 2009). Nessa perspectiva, De Cesare (2005) destaca que o Imposto Sobre Serviços de Qualquer Natureza (ISS) é o mais importante dos impostos para a geração de recursos próprios dos municípios. Em seguida, o IPTU, que tem como característica seu fluxo estável e previsível de arrecadação, independentemente do ciclo econômico.

¹ Constituição Federal, art. 156, incisos I, II e III.

Ao realizar um levantamento sobre a representatividade média dos impostos municipais na arrecadação da receita tributária, de 2010 a 2019, com as capitais brasileiras, utilizando os dados obtidos no site “Compara Brasil”, o resultado vai ao encontro da literatura ora analisada, posto que em todas as capitais, excetuando-se Campo Grande, a arrecadação do IPTU fica em segundo lugar quando comparada aos demais tributos municipais na arrecadação dos municípios. Em Curitiba, a arrecadação média do IPTU entre 2010 e 2019 representou 23,4% da receita tributária, ou seja, possui um peso significativo para os cofres públicos, ficando atrás somente do ISS, que corresponde a uma média de 47% da arrecadação tributária. Juntos, eles representam 70,4% das receitas tributárias (COMPARA BRASIL, 2021).

Há diversos desafios para a arrecadação do IPTU. Afonso, Araújo, Nóbrega (2009) destacam que muitas expansões e reformas, bem como a construção de imóveis de forma ilegal não são registradas nos cadastros das prefeituras. Outro item que merece destaque é a dificuldade encontrada pelos municípios em avaliar a valoração dos imóveis, que em algumas vezes resultam em avaliações desiguais para imóveis de características semelhantes (AFONSO; ARAÚJO; NÓBREGA, 2009). Outro desafio relevante é a inadimplência, que pode ser considerada um dos principais motivos da baixa eficiência arrecadatória desse tributo no Brasil (CARVALHO JUNIOR, 2018). Além disso, De Cesare (2012) aponta que a presença de uma legislação clara sobre as obrigações e mecanismos tributários do imposto predial e territorial urbano é essencial, pois “quando a lei é muito complexa ou ambígua, ou quando muda frequentemente, a incerteza leva ao atraso nos pagamentos e incentiva a evasão fiscal.” (DE CESARE, 2012, p. 6). Carvalho Junior (2016) destaca que a inadimplência do pagamento de impostos no Brasil tende a ser alta, o que se verifica também no IPTU. Afirma ainda que essa inadimplência se deve a quatro principais motivos. O primeiro motivo surge quando o terreno não possui área edificada. Dessa forma, o proprietário não habita neste terreno, e conseqüentemente a identificação do proprietário se torna problemática. Como segundo motivo, aponta que as áreas cadastradas de amplos terrenos por muitas vezes são loteamentos irregulares, com o cadastro totalmente desatualizado. Terceiro, os terrenos sem área edificada tendem a ter alíquotas maiores de IPTU, e por sua vez, valores maiores do referido imposto, o que poderia induzir a inadimplência. E, por fim, existe o comportamento especulativo dos proprietários desses terrenos, que “costumam esperar receber incentivos fiscais, renegociações de dívida ativa de forma vantajosa, desapropriações lucrativas, alterações urbanísticas e apreciação rápida do valor da terra” (CARVALHO JUNIOR, 2016, p. 224).

Considerando-se que parte desses problemas são causados pela falta de informação acerca da situação dos imóveis e terrenos, ou da quitação do tributo, a administração pública pode utilizar técnicas de mineração de dados para a extração de conhecimento, que podem subsidiar os órgãos nas tomadas de decisões de forma mais ágil e com maior confiança e, por conseguinte, promover melhorias nos serviços (GALVÃO; MARIN, 2009). A mineração de dados é capaz de realizar classificações e associações entre registros de bancos de dados com grande volume de dados e, assim, permite que sejam traçados planos específicos para resolver cada um dos casos identificados.

Os municípios têm cada vez mais coletado e armazenado grandes bancos de dados de forma eletrônica, e isso atualmente está sendo muito valorizado, pois existem muitas informações que estão “ocultas” nessas bases de dados. Frequentemente, essas informações não estão disponíveis por dois principais motivos: ausência de ferramentas apropriadas para a extração de informações e o grande volume de dados que, muitas vezes, está além da capacidade do ser humano de análise e extração de relações significativas (REZENDE *et al.*, 2003).

Conforme exposto por Wang, Hu e Zhu (2009), a mineração de dados no setor público começou a ser utilizada como uma ferramenta para a detecção de fraudes e desperdícios, mas atualmente tem avançado para os campos de medição e melhoria de desempenho de programas e políticas públicas. Bach (2003) elenca as áreas do setor público que mais utilizam a aplicação de mineração de dados. Essas podem ser divididas em: finanças e economia, saúde, justiça criminal e segurança, trabalho e bem-estar social, educação e transporte. Detalhando as áreas, Bach (2003) enfatiza ainda nas áreas de finanças e economia, o uso de modelos preditivos de arrecadação e padrões de inadimplência de contribuintes, além de verificar a eficiência dos programas de empréstimos. Também descreve a aplicação desses modelos com sucesso em outros setores, como saúde, justiça criminal, trabalho, bem-estar social, educação e transporte.

Diante do exposto, o objetivo desse estudo é aplicar uma técnica de aprendizado não supervisionado para descoberta de regras de associação (*Apriori*), com o fito de extrair o conhecimento da base de dados do IPTU de 2020 do município de Curitiba. Esse processo é composto por diversas etapas, que vão desde a obtenção da base de dados e a padronização das centenas de milhares de registros que o compõem, a verificação dos atributos disponíveis, até a aplicação de regras estatísticas de forma automatizada a fim de produzir conjuntos e identificar relações entre esses registros. Por fim, descrevemos os resultados obtidos na aplicação do referido processo, com especial atenção às regras

encontradas pelo processo de mineração de dados que estão relacionadas à inadimplência ou à forma de pagamento do tributo. Esses poderão ser utilizados posteriormente para subsidiar decisões por parte do governo municipal na promoção de campanhas de conscientização da população, reavaliação das regras que ditam as alíquotas do imposto, modalidades de descontos e isenção a partir das características resultantes do processamento dos microdados.

O trabalho encontra-se organizado da seguinte forma: a primeira parte, introdução apresentada; a segunda parte detalha o funcionamento do IPTU em Curitiba; já a terceira parte apresenta o funcionamento do processo de descoberta de conhecimento em banco de dados, com os principais conceitos que embasam este estudo; enquanto a quarta parte apresenta os dados do IPTU que foram utilizados. A quinta parte traz a realização do pré-processamento e a transformação dos dados, para, na sexta parte, explicar sobre o algoritmo utilizado. Os resultados são apresentados na sétima parte; e, por fim, a oitava e última parte apresenta as considerações finais e as limitações da pesquisa.

2 IPTU em Curitiba

Nesta seção, apresentam-se dados e estatísticas do IPTU na cidade de Curitiba a fim de estabelecer a base para a avaliação de suas características, absolutas e relativas a outras cidades. Este quadro também tem o objetivo de contextualizar o trabalho de mineração de dados e seus resultados.

De acordo com dados do IBGE e do Lincoln Institute of Land Policy, estima-se que em 2019 havia 63.154 terrenos, 498.918 imóveis residenciais e 175.949 não-residenciais em Curitiba, resultado em uma distribuição das propriedades de 8,56% para terrenos, 67,60% para imóveis residenciais e 23,84% para não-residenciais.²

A respeito do cadastramento imobiliário, a Prefeitura de Curitiba dispõe do Decreto nº 1.123, de 22 de outubro de 2007, que regulamenta a composição e atualização do cadastro imobiliário. Esse decreto declara que o sujeito passivo do IPTU deve manter atualizados os seguintes dados do cadastro imobiliário: nome do proprietário, titular do domínio útil, possuidor a qualquer título ou adquirente; número de inscrição no Cadastro de Pessoas Físicas (CPF); área construída, padrão de construção e utilização do imóvel; e, endereço para correspondência. De acordo com o art. 4º desse decreto, caso não haja a

² Embora a nota (2) do Dicionário presente na planilha do IBGE afirme que “(2) A diferença da soma dos itens residenciais e não residenciais com relação ao item total trata-se do quantitativo de terrenos ou imóveis mistos (residenciais e não residenciais) que foram considerados apenas no total.” (Ibge, 2019), o total de Curitiba corresponde exatamente à soma dos dois tipos de imóveis, o que se repete com todos os outros municípios na base.

atualização do cadastro, o contribuinte ficará sujeito ao pagamento de multa (CURITIBA, Prefeitura Municipal, 2007).

A Emenda Constitucional nº 29, de 13 de setembro de 2000, instituiu a possibilidade de cobrança do IPTU progressivo de acordo com o valor venal do imóvel e de sua localização e uso (BRASIL, 2000). A última atualização da Planta Genérica de Valores da cidade de Curitiba ocorreu em 2017, por meio da Lei Complementar nº 105, que estabelece os parâmetros para a cobrança do imposto relativo aos exercícios de 2018 a 2021 (CURITIBA, Prefeitura Municipal, 2017).

O primeiro elemento constitutivo do cálculo do valor venal do imóvel é o do metro quadrado construído, conforme Tabela 1 abaixo.

Tabela 1 | Valor da construção na Planta Genérica de Valores (exercício 2018)

VALORES DE M DE CONSTRUÇÃO				
	TIPO DE CONSTRUÇÃO	N.º PAV	VL. M² RES. (R\$)	VL. M² COM. (R\$)
1	Alvenaria simples	1 a 3	560,00	560,00
1	Alvenaria simples	4 ou +	1.080,00	1.134,00
21	Alv. média simples	1 a 3	784,00	784,00
22	Alvenaria média	1 a 3	960,00	960,00
22	Alvenaria média	4 ou +	1.480,00	1.554,00
3	Alvenaria fina	1 a 3	1.200,00	1.200,00
3	Alvenaria fina	4 ou +	1.600,00	1.680,00
4	Mista simples	QQUER	352,00	352,00
5	Mista média	QQUER	352,00	352,00
6	Madeira simples	QQUER	256,00	256,00
7	Madeira média	QQUER	256,00	256,00
8	Galpão alvenaria	QQUER	400,00	400,00
9	Galpão madeira	QQUER	192,00	192,00
10	Telheiro mad/alv.	QQUER	200,00	200,00
11	Alvenaria luxo	1 a 3	1.600,00	1.600,00
11	Alvenaria luxo	4 ou +	2.056,00	2.160,00
12	Madeira luxo	QQUER	712,00	712,00

Fonte: Curitiba. Prefeitura Municipal. (2017, p. 3).

Há variação substancial de valores, com o metro quadrado da construção em alvenaria luxo avaliado em mais de 8 vezes o valor da construção em madeira simples e média. Além das características da construção, há variação substancial do imposto de acordo com o logradouro, de acordo com extensa lista que consta no anexo II da lei complementar supracitada. Sobre o valor venal, aplicam-se alíquotas que variam de 0,20% a 1,10% para imóveis residenciais, de 0,35% a 1,80% para imóveis não-residenciais e de 1,00% a 3,00% para imóveis territoriais.³ É digno de nota que o Instituto de Pesquisa e Planejamento de Curitiba (IPPUC) oferece uma ferramenta online na qual o cidadão pode consultar a evolução do valor venal dos imóveis desde o ano 2000.⁴

De acordo com a Secretaria Municipal de Planejamento, Finanças e Orçamento de Curitiba (CURITIBA, sem data), há uma série de imóveis que gozam de isenção, redução ou imunidade à cobrança do IPTU. Entre aqueles que possuem isenção, estão:

1. Proprietários de imóveis simples (...) isenção concedida automaticamente pela prefeitura aos proprietários de imóveis de até 70,00m² de área construída, padrão simples de acabamento e valor venal de até R\$ 140.000,00 e edificações construídas pela COHAB, inclusive aquelas em parceria com a iniciativa privada – Minha Casa Minha Vida;
2. Clubes filiados à Federação Paranaense de Futebol Amador;
3. Ex-combatentes na II Guerra Mundial;
4. Incentivo ao esporte - entidades civis sem fins lucrativos, inclusive os clubes sociais. (CURITIBA. Secretaria Municipal de Planejamento, Finanças e Orçamento, sem data, não p.).

A redução do imposto é concedida a:

1. Imóveis considerados como patrimônio histórico-cultural;
2. Proprietários ou possuidores de terrenos integrantes do Setor Especial de Áreas Verdes;
3. Terrenos ocupados com atividade econômica primária;
4. Pessoa idosa. (CURITIBA. Secretaria Municipal de Planejamento, Finanças e Orçamento, sem data, não p.).

³ A tabela se encontra em: <<https://www.curitiba.pr.gov.br/conteudo/tabela-do-iptu/368>>. Acesso em: 11 jul. 2021.

⁴ Disponível em: <<http://geoapp.ippuc.org.br/plantagenericaidevalores/>>. Acesso em: 11 jul. 2021.

Têm direito à imunidade:

1. Os imóveis integrantes do patrimônio da União, Estados, Municípios, Distrito Federal, autarquias e fundações instituídas e mantidas pelo poder público;
2. Os templos de qualquer culto;
3. Os imóveis integrantes do patrimônio dos partidos políticos, inclusive suas fundações; do patrimônio das entidades sindicais dos trabalhadores; das instituições de educação e assistência social, sem fins lucrativos e atendidos os requisitos da lei (CF, artigo 150, VI, c); e das instituições de Educação e de Assistência Social. (CURITIBA. Secretaria Municipal de Planejamento, Finanças e Orçamento, sem data, não p.).

Segundo Carvalho Júnior (2018), as diferentes modalidades de isenção contemplam 14% do cadastro imobiliário da cidade (dados de 2012). A taxa de adimplência na cidade alcança os 84% (dados de 2014), colocando-a na 6ª posição de menor inadimplência entre as 53 cidades pesquisadas pelo autor.

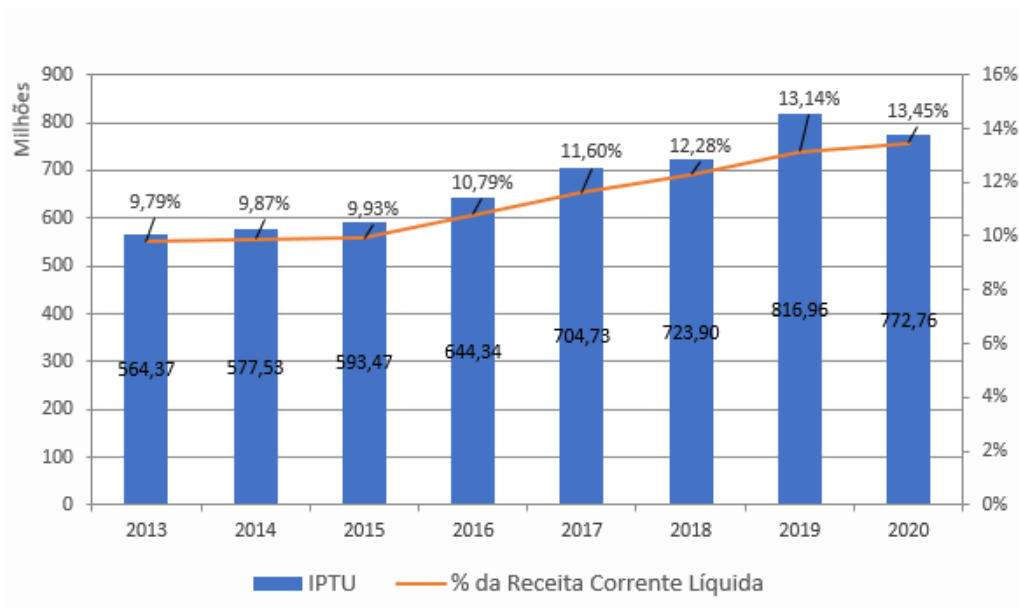
O pagamento do imposto em parcela única até o dia 10 de fevereiro garantiu um desconto de 4% aos contribuintes no ano de 2021. A outra opção oferecida pela prefeitura foi o pagamento em dez parcelas mensais. Excepcionalmente por causa da pandemia, também foi lançado um programa de refinanciamento do imposto de 2020 em 36 parcelas (CURITIBA. Prefeitura Municipal 2021).

O Gráfico 1 abaixo mostra a evolução recente da receita com IPTU na cidade, inflacionada pelo IPCA a valores de 12/2020. Nota-se que a arrecadação representou um crescimento real de 36,88% entre 2013 e 2020, de pouco mais de R\$ 564 milhões para mais de R\$ 772 milhões (valores de dez/2020). Como percentual da receita corrente líquida do município, o IPTU aumentou de 9,79% em 2013 para 13,45% em 2020. O destaque negativo é o ano de 2020, quando a receita com o imposto caiu. Essa queda de arrecadação pode ser atribuída à crise causada pela pandemia do novo coronavírus, que afetou severamente a população em todos os estratos socioeconômicos, mas notadamente teve efeito na taxa de desemprego e, portanto, na renda da população mais pobre.⁵ Contudo,

⁵ A taxa de desemprego no primeiro trimestre de 2020, quando a pandemia tomou proporções alarmantes no país, foi de 12,2%, contra 11% no 4º trimestre de 2019, mas abaixo dos 12,7% do 1º trimestre de 2019. A partir do 2º trimestre, a taxa sobe para 13,3% e desde então tem viés ascendente até atingir 14,7% no 1º trimestre de 2021 e estabilizar em 14,1% no 2º trimestre deste ano, último dado disponível. (Agência Ibge Notícias, 2021).

a representatividade do IPTU ainda maior nas receitas do município indica que outras fontes de renda sofreram redução ainda mais significativa, corroborando sua estabilidade face à instabilidade econômica, conforme seção anterior.

Gráfico 1 | Evolução recente das receitas de Curitiba



Fonte: elaborado pelos autores (2021).⁶

3 Método

O processo de descoberta de conhecimento em banco de dados, chamado de *Knowledge Discovery in Databases* (KDD) foi definido, em 1996, por Fayyad em conjunto com outros pesquisadores, como sendo: “o processo não trivial de identificar padrões de dados válidos, novos, potencialmente úteis e, em última análise, compreensíveis” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). O processo de descoberta de conhecimento em banco de dados foi estruturado pelos autores em cinco etapas: seleção dos dados, pré-processamento, transformação, mineração de dados (aplicação de algoritmos) e avaliação ou validação do conhecimento.

⁶ Dados obtidos no site do TCE/PR. Valores inflacionados pelo IPCA para 12/2020. Receita com IPTU desconsidera juros, multas e dívida ativa.

Isto posto, constitui o primeiro passo selecionar a base de dados que será utilizada para o processo de descoberta de conhecimento. Após, deve ser realizado o pré-processamento, com a limpeza nos dados, como por exemplo, seleção e construção de atributos e preenchimento de valores faltantes, pois regularmente os dados são encontrados com muitas inconsistências (FONSECA; NAMEN, 2016). Essa é uma etapa fundamental do processo, uma vez que tem o propósito de melhorar a qualidade dos dados estatisticamente, no intuito de extrair padrões, e de modo a não interferir no resultado após a aplicação do algoritmo.

A fase de transformação dos dados é necessária para atender à necessidade do algoritmo que será usado na etapa de mineração, e pode ser feita de várias maneiras, como por exemplo: (i) transformação direta, que substitui valores numéricos por categóricos, como por exemplo sexo: 1 → M e 0 → F; (ii) discretização, que transforma atributos contínuos em atributos categóricos, simplificando os dados; e, (iii) binarização, que transforma os valores categóricos em valores representados por 0 ou 1, como por exemplo, o estado civil: solteiro = 1 e casado = 0 (FILHO, 2009).

Em seguida, passa-se à etapa de mineração dos dados, que segundo Berry e Linoff⁷ (1997 *apud* COSTA; BERNARDINI; FILHO, 2014), consiste na “exploração e análise, através de meios automáticos ou semiautomáticos, de grandes quantidades de dados com o objetivo de descobrir padrões e regras significativas.

E, por fim, realiza-se a etapa de avaliação, que objetiva a interpretação dos padrões que resultaram da mineração. Com o conhecimento obtido, ele serve como apoio para a tomada de decisões.

4 Seleção dos dados

A fim de atingir o objetivo proposto neste trabalho, foram utilizadas duas bases de dados, uma contendo os cadastros dos imóveis de Curitiba (Base_IPTU_Cadastro) do ano de 2020, e outra contendo os lançamentos de pagamento do tributo (Base_IPTU_Pagamento), até junho de 2020. As bases, que estavam em formato do *Microsoft Excel 365*, foram compiladas por meio do *Microsoft Access 365*, resultando na base que foi utilizada neste estudo (Base_IPTU_2020). Esta base de dados apresenta as variáveis descritas no Quadro 1.

⁷ Berry, M. J. A.; Linoff, G. Data mining techniques: For marketing, sales, and customer support. New York: John Wiley & Sons, 1997.

Quadro 1 | Atributos e descrição

Atributo	Descrição
Regional nome	Nome da regional que o bairro pertence
Bairro nome	Nome do bairro relativo ao imóvel ou terreno
Natureza	“Predial” para o imóvel com construção e “territorial” para os imóveis sem área edificada
Espécie	Indica a espécie do lançamento
Construção utilização	Descreve a finalidade do imóvel
Pavimento	Faixa de número de pavimentos
Construção padrão	Padrão que designa a qualidade da construção
Construção material	Material utilizado na construção do imóvel
Pagamento antecipado	Pagamento antecipado do imposto com desconto
Débito automático	Utilização ou não do débito automático para o pagamento do imposto
Pagamento	Forma de pagamento

Fonte: elaborado pelos autores (2020).

5 Pré-processamento e transformação dos dados

Com a finalidade de preparar a base de dados para a mineração, foram excluídos os dados contidos na base que não possuíam a informação do pagamento (22.611 linhas) e os dados que não apresentaram o nome do bairro (107 linhas). As linhas que estavam com o campo “Regional nome” ausente foram completadas de acordo com o bairro cadastrado nos demais campos ora preenchidos. Após a exclusão e a correção dos dados ausentes, realizaram-se as transformações de alguns campos (atributos). Primeiro, o atributo “Pagamento” continha dados em valores monetários - “Reais”, que foram excluídos e substituídos por “pagamento à vista”, “pagamento parcelado” e “não pago”.

No atributo “Pavimento” foi utilizada a metodologia de discretização equidistante, que “divide a faixa de valores entre v_{min} e v_{max} em k intervalos de mesmo tamanho. Deste modo, cada intervalo tem tamanho $t = (v_{max}-v_{min})/k$, e os pontos que dividem os intervalos são $v_{min} + t$, $v_{min} + 2t$... $v_{min} + (k-1)t$, onde k é predefinido pelo usuário.”

(CATLETT⁸, 1991 *apud* SILVA, 2004). As três faixas resultantes foram: “1 a 16”, “17 a 31” e “32 ou mais”. Saliendo que, o atributo “pavimento” possui o campo “não utiliza o campo”, o qual se refere ao imóvel que possui natureza territorial e, portanto, não possui área edificada. Esse processo foi necessário, dado que o algoritmo *Apriori* exige dos atributos numéricos a sua substituição por atributos categóricos (FACELLI *et al.*, 2011).

Em suma, a base de dados passou a apresentar 11 atributos em um total de 915.034 observações. O Quadro 2 apresenta os atributos e seus conteúdos.

Quadro 2 | Atributos e seu conteúdo

Atributo	Conteúdo (exemplos de preenchimento)
Regional nome	Ex: Matriz, Pinheirinho, Cajuru, Boa Vista, Santa Felicidade
Bairro nome	Ex: Mercês, Umbará, Seminário, Alto da Glória, Campina do Siqueira
Natureza	Predial, Territorial
Espécie	Ex: Normal, Condomínio horizontal, Condomínio vertical
Construção utilização	Ex: Comercial, Residencial, Escolar, Templos
Pavimento	1 a 16, 17 a 31, 32 ou mais, Não utiliza o campo
Construção padrão	Ex: Galpão, Simples, Luxo
Construção material	Ex: Alvenaria, Madeira
Pagamento antecipado	Sim, Não
Débito automático	Sim, Não
Pagamento	Pagamento à vista, Pagamento parcelado, Não pago

Fonte: elaborado pelos autores (2020).

6 Mineração dos dados

Continuando com o processo de descoberta de conhecimento, avança-se para a etapa de mineração dos dados, na qual ocorre a extração do conhecimento implícito na base de dados. A utilização de algoritmos de aprendizado de máquina no processo de mineração de dados pode ser classificada em aprendizado preditivo e aprendizado descritivo.

⁸ Catlett, J. On changing continuous attributes into ordered discrete attributes. In: Proceedings of the European working session on learning, Porto, Portugal, March, 1991. p. 164-178.

Segundo Costa, Bernardini e Filho (2014), os algoritmos preditivos têm como objetivo construir um estimador dado um conjunto de exemplos rotulados. Se o conjunto ao qual os rótulos dos dados pertencem for um conjunto infinito e ordenado de valores, o problema será de regressão, e o estimador é denominado como regressor. Entretanto, se o conjunto ao qual os rótulos dos dados pertencem for um conjunto finito e não ordenado de valores, o problema será de classificação, e o estimador é denominado como classificador. Ademais, o aprendizado preditivo é igualmente chamado de aprendizagem supervisionada.

Ainda de acordo com Costa, Bernardini, Filho (2014), o aprendizado descritivo envolve tarefas de identificação de informações pertinentes nos dados sem um elemento externo para nortear o processo de aprendizado. Essas tarefas podem ser divididas em três tipos: (i) sumarização, que tem por objetivo descobrir uma descrição mais simples e compacta dos dados; (ii) associação, que tem como finalidade a busca de padrões frequentes de associações entre os atributos de um determinado conjunto de dados; e, (iii) agrupamento, cujo propósito é realizar a identificação de grupos nos dados de acordo com a proximidade entre os objetos. Outrossim, o aprendizado descritivo é também chamado de aprendizagem não supervisionada.

Neste trabalho foi utilizado o algoritmo pertencente à aprendizagem não supervisionada, o *Apriori*. Ele foi proposto por Agrawal, Imielinski e Swami, em 1993, (COSTA; BERNARDINI; FILHO, 2014), e baseia-se na busca por padrões indicadores do relacionamento entre conjuntos de itens. Por conseguinte, o objetivo de se realizar mineração de dados com regras de associação é encontrar todos os conjuntos de itens que frequentemente ocorrem de forma simultânea em determinada base de dados e formar regras a partir destes conjuntos. O *Apriori* é considerado um dos algoritmos com maior utilização na aplicação de descoberta de regras de associação, porque executa inúmeras leituras nas bases de dados, sendo eficaz para trabalhar com muitos atributos (COSTA; BERNARDINI; FILHO, 2014).

As regras de associação são apresentadas da seguinte forma: $X \rightarrow Y$ (lê-se: se X então Y), no qual X é o antecedente da regra, e Y o conseqüente da regra. Ademais, X e Y são dois *itemsets* (total de itens) distintos em determinada base de dados. Os autores Carvalho, Sampaio e Mongiovi (1999) citam como exemplo: “ao encontrar a seguinte associação {mercúrio, gaze, esparadrapo} \rightarrow {algodão} (0,78), significando que 78% dos clientes que compram mercúrio, gaze, esparadrapo também compram algodão”. Desse modo, o gerente do estabelecimento pode colocar esses produtos próximos nas prateleiras, apli-

car algum desconto ao vender esses produtos em conjunto, e até mesmo compreender o motivo de uma queda nas vendas de determinados produtos.

Ainda de acordo com Carvalho, Sampaio e Mongiovi (1999), o algoritmo *Apriori* realiza a mineração em duas etapas: geração e poda. Na geração, é realizada uma varredura sobre os dados, com o objetivo de gerar todos os conjuntos de combinações de valores possíveis. Após, com a poda, são analisadas apenas as combinações com uma frequência maior que um valor mínimo pré-determinado, denominado suporte. O suporte é definido como:

$$\text{Suporte}(X) = \frac{N^{\circ} \text{ de registros do arquivo que contêm os elementos do conjunto } X}{N^{\circ} \text{ total de registros do arquivo}}$$

Assim, ao descobrir regras de associação, faz-se necessário definir um suporte mínimo para as regras.

Em seguida, as regras são produzidas a partir dos conjuntos de combinações e é utilizada outra medida para seleção de regras que foram apontadas como relevantes: a confiança da regra, definida pela equação abaixo:

$$\text{Confiança}(R) = \frac{N^{\circ} \text{ de registros com } X \text{ e } Y}{N^{\circ} \text{ de registros com } X}$$

Dessa forma, a confiança, além de ser um parâmetro de seleção das regras, também é utilizada para avaliar a qualidade das regras produzidas (COSTA; BERNARDINI; FILHO, 2014).

Outra medida importante é o *lift*, também chamado de coeficiente de interesse, que mede a dependência entre o antecedente e consequente da regra. O *lift* de uma regra é dado pela seguinte equação:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confiança}(X \rightarrow Y)}{\text{Suporte}(Y)} = \frac{\text{Suporte}(X \cup Y)}{\text{Suporte}(X) \times \text{Suporte}(Y)}$$

Quanto maior o valor do *lift*, mais interessante a regra, dado que a dependência é maior entre os itens que a compõe. Portanto, se o valor do *lift* de uma regra for menor que 1, então X e Y são negativamente dependentes; se o valor for igual a 1, X e Y são independentes, e se o valor do *lift* for maior que 1, então X e Y são positivamente dependentes (ANSELMO, 2017).

7 Resultados

Utilizando o aplicativo Weka (*Waikato Environment for Knowledge Analysis*) e aplicando o algoritmo de associação *Apriori* na base de dados (Base_IPTU_2020), as melhores regras encontradas pelo algoritmo com suporte de 10% e confiança de 90% estão apresentadas no Quadro 3.

Quadro 3 | Resultados do algoritmo *Apriori*

(i)	Construção material = Alvenaria → natureza = PREDIAL conf: (1) lift: (1.05)
(ii)	Pavimento = 1 a 16 → natureza = PREDIAL conf: (1) lift: (1.05)
(iii)	Construção material = Alvenaria débito automático = NÃO → natureza = PREDIAL conf: (1) lift: (1.05)
(iv)	Pavimento = 1 a 16 débito automático = NÃO → natureza = PREDIAL conf: (1) lift: (1.05)
(v)	Débito automático = NÃO → natureza = PREDIAL conf: (0.96) lift: (1)
(vi)	Pavimento = 1 a 16 → débito automático = NÃO conf: (0.93) lift: (1)
(vii)	Natureza = PREDIAL pavimento = 1 a 16 → débito automático = NÃO conf: (0.93) lift: (1)
(viii)	Pavimento = 1 a 16 → natureza = PREDIAL débito automático = NÃO conf: (0.93) lift: (1.05)
(ix)	Natureza = PREDIAL → débito automático = NÃO conf: (0.93) lift: (1)
(x)	Construção material = Alvenaria → débito automático = NÃO conf: (0.93) lift: (1)

Fonte: elaborado pelos autores (2020).

As quatro primeiras regras encontradas apresentaram a confiança de 100%: (i) se o material utilizado na construção for alvenaria, então a natureza é predial; (ii) se o imóvel tiver de 1 a 16 pavimentos, então o imóvel é de natureza predial; (iii) se o material utilizado na construção for alvenaria e não tiver deixado o imposto em débito automático, então o imóvel é de natureza predial; (iv) se o imóvel tiver de 1 a 16 pavimentos e não tiver deixado o imposto em débito automático, então o imóvel é de natureza predial.

Algumas regras encontradas podem não apresentar utilidade no subsídio de informações à gestão municipal, como por exemplo as regras (i) e (ii) acima, pois se há material utilizado na construção e há pavimentos, é evidente que a natureza será exclusivamente predial, ou seja, não poderá ser territorial. Isso apenas confirma que o algoritmo gerou regras a partir de uma base de dados coesa.

No entanto, usando atributos como “pavimento”, “natureza”, “débito automático” e “construção material” surge como consequente da regra a não utilização do débito automático do tributo pelos contribuintes.

No Quadro 4, abaixo, foi considerada a mesma base de dados, porém, utilizando somente os atributos “Bairro nome”, “Natureza”, “Construção padrão”, “Débito automático” e “Pagamento”, com o suporte de 10% e confiança de 90%. A fim de proporcionar a descoberta de conhecimento especificamente vinculado ao pagamento do imposto, “Débito automático” foi selecionado como atributo meta do algoritmo (por meio da alteração do parâmetro “car”⁹ para “true”¹⁰). Assim, os registros são classificados pelos outros atributos – denominados previsores – a partir desta meta ou classe e consegue-se gerar conhecimento sobre grupos que são formados em função da forma de pagamento.

Quadro 4 | Resultados do algoritmo *Apriori* utilizando os atributos “Bairro nome”, “Natureza”, “Construção padrão”, “Débito automático” e “Pagamento”

Pagamento = Pagamento à vista → débito automático = NÃO conf: (0.98)
Natureza = PREDIAL pagamento = Pagamento à vista → débito automático = NÃO conf: (0.98)
Construção padrão = MEDIA pagamento = Pagamento à vista → débito automático = NÃO conf: (0.97)
Natureza = PREDIAL construção padrão = MEDIA pagamento = Pagamento à vista → débito automático = NÃO conf: (0.97)
Construção padrão = SIMPLES → débito automático = NÃO conf: (0.93)
Natureza = PREDIAL construção padrão = SIMPLES → débito automático = NÃO conf: (0.93)
Natureza = PREDIAL → débito automático = NÃO conf: (0.93)
Construção padrão = MEDIA → débito automático = NÃO conf: (0.92)
Natureza = PREDIAL construção padrão = MEDIA → débito automático = NÃO conf: (0.92)
Construção padrão = SIMPLES pagamento = Pagamento parcelado → débito automático = NÃO conf: (0.9)

Fonte: elaborado pelos autores (2020).

As regras apresentadas acima refletem que o débito automático não foi utilizado em nenhuma das opções. Se excetuarmos as quatro primeiras regras que não demonstram descobertas, e sim corroboram um fato (se o contribuinte faz o pagamento à vista, ele não utiliza o débito automático para cobrança), as demais regras são particularmente interessantes para pesquisa sobre inadimplência nos grupos de construção padrão sim-

⁹ Quando o parâmetro “car” é habilitado, as regras de associação de classe são exploradas em vez das regras de associação gerais.

¹⁰ True, em português, verdadeiro.

ples e média, a fim de se verificar se o incentivo ao débito automático pode aumentar o índice de pagamento do tributo.

No Quadro 5 apresentado na sequência, também considerando a mesma base de dados, foram selecionados os atributos “Bairro nome”, “Construção padrão”, e “Pagamento”, com o suporte de 1% e confiança de 5%. Abaixo, as regras geradas pelo algoritmo que merecem destaque:

Quadro 5 | Resultados do algoritmo *Apriori* utilizando os atributos “Bairro nome”, “Construção padrão” e “Pagamento”

Bairro nome = CENTRO construção padrão = SIMPLES → pagamento = Pagamento parcelado conf: (0.77) lift: (1.4)
Bairro nome = CENTRO pagamento = Pagamento à vista → construção padrão = MEDIA 20129 conf: (0.75) lift: (2.06)
Bairro nome = SÍTIO CERCADO pagamento = Pagamento parcelado → construção padrão = SIMPLES conf: (0.74) lift: (1.93)
Bairro nome = SÍTIO CERCADO → construção padrão = SIMPLES conf: (0.71) lift: (1.85)
Bairro nome = VILA IZABEL → pagamento = Pagamento parcelado conf: (0.7) lift: (1.28)
Bairro nome = REBOUÇAS → pagamento = Pagamento parcelado conf: (0.69) lift: (1.25)
Bairro nome = CRISTO REI → construção padrão = MEDIA conf: (0.66) lift: (1.81)
Bairro nome = CENTRO → construção padrão = MEDIA conf: (0.64) lift: (1.74)
Bairro nome = CAMPO COMPRIDO → pagamento = Pagamento parcelado conf: (0.64) lift: (1.17)
Bairro nome = BIGORRILHO → pagamento = Pagamento parcelado conf: (0.62) lift: (1.12)
Bairro nome = PORTÃO → pagamento = Pagamento parcelado conf: (0.61) lift: (1.11)
Construção padrão = SIMPLES → pagamento = Pagamento parcelado conf: (0.61) lift: (1.11)
Construção padrão = FINA → pagamento = Pagamento parcelado conf: (0.6) lift: (1.1)
Bairro nome = BOA VISTA → pagamento = Pagamento parcelado conf: (0.59) lift: (1.08)
Bairro nome = CAPÃO RASO → pagamento = Pagamento parcelado conf: (0.59) lift: (1.08)
Bairro nome = ÁGUA VERDE → pagamento = Pagamento parcelado conf: (0.58) lift: (1.07)
Bairro nome = CIDADE INDUSTRIAL → construção padrão = SIMPLES conf: (0.58) lift: (1.52)
Bairro nome = TATUQUARA → construção padrão = SIMPLES conf: (0.58) lift: (1.51)
Bairro nome = PINHEIRINHO → pagamento = Pagamento parcelado conf: (0.57) lift: (1.03)
Bairro nome = BATEL → pagamento = Pagamento parcelado conf: (0.55) lift: (1.01)

Fonte: elaborado pelos autores (2020).

Depreende-se das regras geradas acima que, se o bairro for “Centro” e o imóvel for considerado um padrão “simples”, então o pagamento é realizado de forma “parcelada”, apresentando uma confiança de 77%. Mas, se ainda no bairro “Centro”, o pagamento for feito “à vista”, então o padrão de construção do imóvel se classifica como “média”, com 75% de confiança.

O bairro “Sítio Cercado” aparece em duas regras, sendo a primeira: se o bairro for “Sítio Cercado” e o pagamento do tributo “parcelado”, então o padrão de construção do imóvel é “simples”. A segunda regra implica que, se o bairro é o “Sítio Cercado”, então o padrão de construção é “simples”. Ademais, outras regras trazem como consequente o padrão “simples” para os bairros “Cidade Industrial” e “Tatuquara”. Isso demonstra que os bairros que são localizados nas zonas periféricas possuem um padrão menor de complexidade construtiva. Outra regra reforça com 61% de confiança que, se o padrão da construção for “simples”, então o pagamento é realizado de forma “parcelada”.

Por outro lado, os bairros “Cristo Rei” e “Centro” apresentaram como consequentes de regra o padrão de construção do imóvel como “médio”.

Conforme demonstrado nos Quadros 3 e 4, o Quadro 5 também apresenta várias regras em que o pagamento do tributo foi realizado de forma parcelada, principalmente nos seguintes bairros: “Vila Izabel”, “Rebouças”, “Campo Comprido”, “Bigorrião”, “Portão”, “Boa Vista”, “Capão Raso”, “Centro”, “Água Verde”, “Pinheirinho”, “Batel”.

Outra regra que merece destaque é que, se o padrão da construção do imóvel for “fina”, então o pagamento também é “parcelado”, com confiança de 60%.

A fim de detectar um padrão para os contribuintes inadimplentes, passou-se a utilizar somente a base de dados com os contribuintes que não efetuaram o pagamento do tributo, selecionando da base de dados o campo “Pagamento” como “Não pago”, e mantendo os atributos: “Regional nome”, “Bairro nome”, “Construção utilização”, “Construção padrão” e “Construção material”. Os resultados com suporte de 10% e confiança de 90%, estão apresentados no Quadro 6:

Quadro 6 | Resultados do algoritmo *Apriori* selecionando o atributo “Pagamento” como “Não pago”

Construção padrão = MEDIA SIMPLES → construção material = Alvenaria conf: (1) lift: (1.43)
Construção utilização = Residencial construção padrão = MEDIA SIMPLES → construção material = Alvenaria conf: (1) lift: (1.43)
Construção padrão = SIMPLES construção material = Madeira → construção utilização = Residencial conf: (0.97) lift: (1.23)
Construção material = Madeira → construção utilização = Residencial conf: (0.96) lift: (1.23)
Construção utilização = Residencial construção material = Madeira → construção padrão = SIMPLES conf: (0.95) lift: (1.84)
Construção material = Madeira → construção padrão = SIMPLES conf: (0.94) lift: (1.83)
Bairro nome = CIDADE INDUSTRIAL → regional nome = A.R. CIC conf: (0.93) lift: (8.57)
Regional nome = A.R. CIC → bairro nome = CIDADE INDUSTRIAL conf: (0.92) lift: (8.57)
Construção padrão = SIMPLES → construção utilização = Residencial conf: (0.91) lift: (1.17)
Construção material = Madeira → construção utilização = Residencial construção padrão = SIMPLES conf: (0.91) lift: (1.93)

Fonte: elaborado pelos autores (2020).

Constata-se das regras geradas acima que os padrões de construção “simples” e “média simples”, e material de construção de “madeira” ou “alvenaria”, e sendo o imóvel de “utilização residencial”, são características dos imóveis dos contribuintes mais inadimplentes.

Além disso, merece destaque a “Regional CIC” (Cidade Industrial de Curitiba) que engloba o bairro “Cidade Industrial”. Este se sobressai como bairro que deixa de pagar o tributo, pois é o único bairro que aparece nas regras quando utilizamos o atributo “pagamento” como “não pago”. De acordo com o Ippuc (Instituto de Pesquisa e Planejamento Urbano de Curitiba), a Cidade Industrial é o bairro mais populoso da capital paranaense, e conforme apontado no Censo 2010, 172 mil pessoas moravam nesse bairro. Destaca-se ainda que o bairro apresenta uma das menores rendas do município, de

acordo com dados de 2010, que são os mais recentes disponíveis no sistema infocuritiba, do Instituto de Pesquisa e Planejamento Urbano de Curitiba.¹¹

Diante dessa descoberta, pode-se inferir que a inadimplência acontece principalmente nos imóveis que apresentam características de construção mais modestas e em localidade de moradia de famílias de baixa renda.

8 Considerações finais

A utilização da mineração de dados no setor público possibilita a identificação de regras e descoberta de padrões que podem não ser identificadas pelas ferramentas tradicionais de análise de dados, como as consultas realizadas por meio de ferramentas de *Business Intelligence* e relatórios gerenciais.

Além disso, a mineração de dados oferece possibilidades à administração pública para otimizar as suas decisões, baseando-se em tendências gerais extraídas de dados históricos. Esse conhecimento adquirido pela extração de dados pode servir como um instrumento para uma melhor governança e, também, como um meio para sustentar o conhecimento organizacional (WANG; HU; ZHU, 2009).

Diante dos resultados ora apresentados, as duas principais conclusões extraídas são:

1. independentemente do padrão de construção do imóvel ou bairro, o pagamento é feito de forma “parcelada”. Sugere-se que sejam realizadas ações por parte da gestão pública municipal no sentido de incentivar o uso do débito automático pelos contribuintes e campanhas com incentivos para o pagamento do tributo à vista, especialmente naqueles identificados com padrão de construção “fina”, que se supõe não pertencer a famílias com baixo poder aquisitivo, que seriam potencialmente capazes de efetuar o pagamento do tributo em parcela única. O pagamento à vista teria impacto positivo no fluxo de caixa municipal e auxiliaria na amenização da inadimplência do pagamento deste tributo.
2. a inadimplência é maior em imóveis residenciais de “madeira” ou “alvenaria”, com padrão “simples” e média “simples” e em bairro com uma das menores

¹¹ Pesquisa realizada no website do infocuritiba, do Instituto de Pesquisa e Planejamento Urbano de Curitiba, disponível em: <<http://infocuritiba.ippuc.org.br>> em 18 set. 2021, por “Rendimento nominal mensal médio das pessoas de 10 anos ou mais de idade - Total (Real)” em todos os bairros, com valores de 2010 (únicos disponíveis). O bairro Cidade Industrial teve a 65ª. renda mais baixa dos 75 bairros listados, R\$779,67. O bairro com maior renda foi o Batel (R\$ 4.140,70) e aquele com menor renda, Caximba (R\$ 559,51).

rendas médias da cidade, o que sugere uma correlação com a população de renda mais baixa e, conseqüentemente, menor capacidade contributiva. Essa conclusão corrobora o potencial papel do IPTU e do cadastro de propriedades na administração do município. Uma das possibilidades é realizar a integração de bases de dados – neste caso específico, entre cadastro de propriedades e o CadÚnico (Cadastro Único para Programas Sociais do governo federal)¹², a fim de promover mais insumos à identificação de vulnerabilidades diversas em subgrupos da população. Outro potencial desse resultado é reforçar a importância do papel da administração tributária para outros programas de governo. Por exemplo, a cobertura de propriedades informais, que traria diversas vantagens, desde a regulação do mercado de propriedades, passando pela redução da informalidade (pois em propriedades sem nenhum registro formal, o comprovante do IPTU serviria como registro da propriedade em nome do cidadão) e até mesmo como forma de aumentar o conhecimento do poder público sobre certas áreas, auxiliando na alocação de serviços e equipamentos públicos à população, conforme apontam De Cesare e Smolka (2006).

Uma sugestão adicional, que teria impacto tanto para contribuintes de alta quanto de baixa renda, seria a publicação de boletins anuais com a participação percentual do imposto nas receitas arrecadadas pelo município, relacionando alguns serviços ou obras que foram realizadas no período, produção de cartilhas e materiais educativos ou mesmo publicitários, para que o cidadão fique consciente da sua responsabilidade e importância do pagamento do tributo para o custeio das despesas públicas. Essa atitude “poderia ajudar a desenvolver uma cultura fiscal positiva, tornando os cidadãos mais conscientes da sua responsabilidade e do custo dos serviços públicos”. (DE CESARE, 2012, p. 8)

Além das sugestões apresentadas acima, recomenda-se a realização de trabalhos frequentes, *in loco* ou por meio de georreferenciamento, para que o cadastro se mantenha sempre atualizado a fim de identificar com mais agilidade/facilidade o contribuinte.

Ressalta-se que a limitação deste artigo reside no fato de que a base de dados dos pagamentos continha informações somente até junho de 2020, o que limitou a análise

¹² Embora seja um cadastro do governo federal, o CadÚnico é gerido pela Fundação de Ação Social (FAS), que é “o órgão público responsável pela gestão da assistência social em Curitiba, atuando de forma integrada a órgãos governamentais e instituições não governamentais, que compõem a rede socioassistencial do município”, conforme apresentado no site da fundação. Disponível em: <<https://fas.curitiba.pr.gov.br/conteudo.aspx?id=44>>. Acesso em: 18 set. 2021.

acerca dos pagamentos totais (de todas as parcelas) dos contribuintes. Para estudos futuros sugere-se a utilização da base completa, ou seja, ano base de pagamentos completo.

Por fim, propõe-se testar a aplicação de outras técnicas de mineração de dados no intuito de adquirir novos conhecimentos, pois esse trabalho ficou restrito a uma determinada técnica de mineração de dados, a associação, devido principalmente à natureza da questão a ser resolvida.

Referências

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

AGÊNCIA IBGE NOTÍCIAS. *Desemprego recua para 14,1% no 2º tri, mas ainda atinge 14,4 milhões de pessoas*. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/31480-desemprego-recua-para-14-1-no-2-tri-mas-ainda-atinge-14-4-milhoes-de-pessoas>. Acesso em: 18 set. 2021.

ANSELMO, F. C. G. *Regras de Associação – Market Basket Analysis Itens Frequentes e Itens Raros*. 2017. 82 f. Dissertação (Mestrado em Modelação, Análise de Dados e Sistemas de Apoio à Decisão) – Faculdade de Economia da Universidade do Porto, Porto (Portugal), 2017. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/107693/2/218342.pdf>. Acesso em: 29 mar. 2021.

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-mining-applications-in-public-organizations.pdf>. Acesso em: 02 mar. 2021.

BRASIL. *Lei nº 5.172, de 25 de outubro de 1966*. Dispõe sobre o Sistema Tributário Nacional e institui normas gerais de direito tributário aplicáveis à União, Estados e Municípios. Brasília, DF. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l5172compilado.htm>. Acesso em: 12 dez. 2020.

BRASIL. [Constituição (1988)]. *Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal, 1988.

BRASIL. *Emenda Constitucional nº 29, de 13 de setembro de 2000*. Altera os arts. 34, 35, 156, 160, 167 e 198 da Constituição Federal e acrescenta artigo ao Ato das Disposições Constitucionais Transitórias, para assegurar os recursos mínimos para o financiamento das ações e serviços públicos de saúde. Brasília, DF. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc29.htm>. Acesso em: 18 set. 2021.

CARVALHO, J. V.; SAMPAIO, M. C.; MONGIOVI, G. (1999). *Utilização de técnicas de “Data Mining” para o reconhecimento de caracteres manuscritos*. Disponível em: <<http://www.inf.ufsc.br/>>

sbbd99/anais/SBBD-Completo/20.pdf>. Acesso em: 05 dez. 2020.

CARVALHO JUNIOR, 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, P. H. B. de. O papel da inadimplência na heterogeneidade da arrecadação do IPTU no Brasil. *Boletim regional, urbano e ambiental*. Ipea, nº 18, jan-jun, 2018. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/8473/1/BRU_n18_Papel.pdf>. Acesso em: 16 set. 2021.

CARVALHO JÚNIOR, P. H. B. de. *Panorama do IPTU: um retrato da administração tributária em 53 cidades selecionadas*. Texto para discussão. Brasília: Rio de Janeiro: Ipea, 2018.

COSTA, J. J.; BERNARDINI, F. C.; FILHO, J. V. (2014). A mineração de dados e a qualidade de conhecimentos extraídos dos boletins de ocorrência das rodovias federais brasileiras. *Revista AtoZ*, Curitiba, 3(2), p. 139-157, jul./dez. 2014. Disponível em: <<https://revistas.ufpr.br/atoz/article/view/41346>>. Acesso em: 10 dez. 2020.

COMPARA BRASIL. *Consulta Municípios*. Disponível em: <<http://comparabrasil.com/municipios/paginas/modulo1.aspx>>. Acesso em: 26 fev. 2021.

CURITIBA. *Desconto para pagamento à vista do IPTU termina nesta quarta*. 01/02/2021. Disponível em: <<https://www.curitiba.pr.gov.br/noticias/desconto-para-pagamento-a-vista-do-iptu-termina-nesta-quarta/57806>>. Acesso em: 11 jul. 2021.

CURITIBA. *Decreto nº 1.123, de 22 de outubro de 2007*. Regulamenta a composição e atualização do cadastro imobiliário. Disponível em: <<http://mid.curitiba.pr.gov.br/2010/00083502.pdf>>. Acesso em: 18 set. 2021.

CURITIBA. *Lei complementar nº 105*. Aprova a Planta Genérica de Valores, para efeitos de lançamento e cobrança do Imposto Predial e Territorial Urbano relativo aos exercícios de 2018, 2019, 2020 e 2021 e dá outras providências. 8 de dezembro de 2017. Disponível em: <<http://leismunicipais.com.br/a/pr/c/curitiba/lei-complementar/2017/10/105/lei-complementar-n-105-2017-aprova-a-planta-generica-de-valores-para-efeitos-de-lancamento-e-cobranca-do-imposto-predial-e-territorial-urbano-relativo-ao-exercicio-de-2018-2019-2020-e-2021-e-da-outras-providencias>>. Acesso em: 09 jul. 2021.

CURITIBA. Secretaria Municipal de Planejamento, Finanças e Orçamento. *Isenção e redução*. Sem data. Disponível em: <<https://www.curitiba.pr.gov.br/conteudo/isencao-e-reducao/373>>. Acesso em: 19 set. 2021.

DE CESARE; SMOLKA. Property Taxation and Informality: Challenges for Latin America. *Land Lines*. 2016. Disponível em: <<https://www.lincolnst.edu/publications/articles/property-taxation-informality>>. Acesso em: 18 set. 2021.

DE CESARE, C. M. Melhorar o desempenho do Imposto sobre a Propriedade Imobiliária na América Latina. *Foco em políticas fundiárias*, Lincoln Institute of Land Policy. Cambridge (EUA), 2012. Disponível em: <https://www.lincolnst.edu/sites/default/files/pubfiles/melhorar-desempenho-imposto-propriedade-imobiliaria-full_0.pdf>. Acesso em: 20 jun. 2021.

DE CESARE, C. *O Cadastro como instrumento de política fiscal*. In: ERBA, D. A.; OLIVEIRA, F. L. de; LIMA JR., P. de N. (orgs.). *Cadastro Multifinalitário como Instrumento de Política Fiscal e Urbana*. Rio de Janeiro, 2005. Disponível em: <<https://www.capacidades.gov.br/biblioteca/detalhar/id/78/titulo/cadastro-multifinalitario-como-instrumento-de-politica-fiscal-e-urbana>>. Acesso em: 17 set. 2021.

DE CESARE, C. M.; FERNANDES, C. E.; OLIVEIRA, G. S.; CARVALHO JUNIOR, 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: 19 jul. 2021.

FACELLI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. *Inteligência artificial: uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC, 2011.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge discovery and data mining: towards a unifying framework. *KDD-96 Proceedings*, 1996.

FILHO, R. T. F. *Uma aplicação de mineração de dados ao programa Bolsa Escola da prefeitura da cidade do Recife*. 2009. 113 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife -(PE), 2009. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/2440/1/arquivo3328_1.pdf>. Acesso em: 03 dez. 2020.

FONSECA, S. O.; NAMEN, A. A. Mineração em bases de dados do INEP: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. *Educação em Revista*, v. 32, n. 1, 2016, Belo Horizonte - MG. Disponível em: <https://www.scielo.br/scielo.php?pid=s0102-46982016000100133&script=sci_abstract&tlng=pt>. Acesso em: 02 dez. 2020.

GALVÃO, N. D.; MARIN, H. F. Técnica de mineração de dados: uma revisão da literatura. *Acta Paulista de Enfermagem*, São Paulo, v. 22, n. 5, p. 686-690, abr. 2009. Disponível em: <<http://www.scielo.br/pdf/ape/v22n5/14.pdf>>. Acesso em: 02 mar. 2021.

IBGE. *Pesquisa de informações básicas municipais* – base de dados de 2019. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/educacao/10586-pesquisa-de-informacoes-basicas-municipais.html?=&t=downloads>>. Acesso em: 09 jul. 2021.

LINCOLN INSTITUTE OF LAND POLICY. *Composition of property recorded in the cadastre by property-type (%)*. 2019. Disponível em: <<https://www.lincolninst.edu/sites/default/files/sources/data/indicators-ptla-eng-2017-03-3-3-1.xlsx>>. Acesso em: 11 jul. 2021.

PREFEITURA DE CURITIBA. *CIC tem mais moradores que cidades como Guarapuava e Paranaguá*. Disponível em: <<https://www.curitiba.pr.gov.br/noticias/cic-tem-mais-moradores-que-cidades-como-guarapuava-e-paranagua/42472>>. Acesso em: 02 mar. 2021.

REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; PAULA, M. D. *Mineração de dados*. In: REZENDE, S.O. (org.). *Sistemas inteligentes: fundamentos e aplicações*. São Paulo: Manole, 2003.

SECRETARIA MUNICIPAL DE PLANEJAMENTO, FINANÇAS E ORÇAMENTO DE CURITIBA. *IPTU - Isenção e redução*. Sem data. Disponível em: <<https://www.curitiba.pr.gov.br/conteudo/isencao-e-reducao/373>>. Acesso em: 11 jul. 2021.

UNIVERSITY OF WAIKATO. *Weka 3: data mining with open source machine learning*. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 01 nov. 2020.

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: 02 dez. 2020.

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.

Ana Paula Costa

 <https://orcid.org/0000-0002-7014-7353>

Mestranda em Economia pela Universidade Federal do Paraná (UFPR). Bacharel em Ciências Econômicas pela Universidade Federal do Paraná (UFPR).

E-mail: anacosta@ufpr.br

André Custódio Pecini

 <https://orcid.org/0000-0002-1226-2038>

Doutor e Mestre em Comunicação e Cultura, e Bacharel em Comunicação Social - Publicidade, pela Universidade Federal do Rio de Janeiro (UFRJ). Em estágio pós-doutoral na Universidade Federal do Paraná (UFPR). Pós-graduado em Administração pelo Insper.

E-mail: pecini@ufpr.br

Denise Fukumi Tsunoda

 <https://orcid.org/0000-0002-5663-4534>

Doutora e Mestra em Engenharia Elétrica e Informática Industrial pela Universidade Tecnológica Federal do Paraná (UTFPR). Professora e Bacharel em Informática na Universidade Federal do Paraná (UFPR), no curso de graduação e pós-graduação em Gestão da Informação e no Mestrado Profissional em Economia.

E-mail: dtsunoda@ufpr.br