

Leveraging Property Tax Collection Through an Unsupervised Model for Taxpayer Qualification

Elioenai L. G. Alves¹, Luiz Zairo Bastos Viana², Fabio Santos¹,
Wellington Franco², Vladia Pinheiro¹

¹Universidade de Fortaleza (UNIFOR) – Programa de Pos-Graduao em Informatica Aplicada
Fortaleza – CE – Brasil.

²Universidade Federal do Ceara (UFC) – Campus Crateus
Crateus – CE – Brasil

l.oenaialves@gmail.com, zairo.vianahd@alu.ufc.br,
fabiosantos@edu.unifor.br, wellington@crateus.ufc.br,
vladiacelia@unifor.br

Abstract. *The Urban Property and Land Tax (IPTU), a municipal tax and an important financial support for public services, has been experiencing a decrease in its collection in recent years. In this context, many city governments have been seeking measures to improve communication and the collection system, and, understanding the profile of each taxpayer/property is crucial to optimize the resources allocated to collection and reduce default rates. This study proposes an unsupervised Machine Learning model to assess the qualification of taxpayers of property tax, such as IPTU, based on payment history and tax debts. The results of the model trained with data from the city of Fortaleza, compared to the current baseline system in use in that city, showed promising performance with an F1-Score of 92.30%.*

Resumo. *O Imposto Predial e Territorial Urbano (IPTU), tributo de competencia municipal e importante suporte financeiro de servios publicos, vem sofrendo diminuio na sua arrecadao nos ultimos anos. Neste contexto, muitas prefeituras tem buscado medidas para melhorar a comunicao e o sistema de cobrana, e, conhecer o perfil de cada contribuinte/imovel e fundamental para otimizar os recursos destinados a cobrana e diminuir a inadimplencia. O presente trabalho propoe um modelo de Machine Learning nao supervisionado para qualificar os contribuintes de impostos imobiliarios como o IPTU, a partir do historico de pagamentos e dos debitos do imposto. Os resultados do modelo treinado com dados da cidade de Fortaleza, comparados com o sistema baseline em uso hoje nessa cidade, indicaram desempenho promissor com F1-Score de 92,30%.*

1. Introduo

O Imposto Predial e Territorial Urbano (IPTU), tributo de competencia municipal, incide sobre a propriedade de imoveis localizados nas areas urbanas dos municpios [Afonso et al. 2013] e e uma importante fonte de receita para os municpios de todo Brasil, contribuindo para o funcionamento de servios publicos essenciais, como saude e

educação, e para a manutenção de ruas, praças, limpeza da cidade, etc., necessários para a melhoria da vida e bem-estar nas cidades. Em São Paulo, a receita advinda do IPTU corresponde a 33% da receita de impostos próprios, e em Fortaleza corresponde a 28%. No entanto, é um imposto que vem sofrendo queda na arrecadação nos últimos anos, por vários motivos, dentre os quais podemos citar: diminuição na renda dos contribuintes; falta de percepção dos benefícios e da aplicação do imposto pago, na gestão municipal; e falta de fiscalização [Cupertino et al. 2014]. Mais recentemente, estudos em ciência comportamental, relatam que há espaço para melhoria da arrecadação do IPTU se a comunicação com os contribuintes for mais adequada ao perfil dos mesmos¹.

Para a notificação e cobrança do IPTU, muitas prefeituras tem utilizado uma abordagem conhecida como “régua de cobrança” [Barbosa Júnior 2020]. Essa ferramenta, originalmente empregada no mundo corporativo, permite que as empresas estabeleçam procedimentos sistemáticos de cobrança, com datas específicas para cada tipo de ação a ser realizada. Isso inclui o envio de cartas, e-mails e mensagens (SMS e WhatsApp) e até contato telefônico, para alertar sobre valores a pagar e em atraso. Portanto, conhecer o perfil de cada contribuinte e imóvel é fundamental para melhorar a interação e diminuir a inadimplência. Além disso, cada modalidade de comunicação e cobrança - cartas, e-mails, WhatsApp, SMS e até ligação telefônica - tem seus custos associados, requisitos e limitações. No caso específico do envio de cartas, por ser, ainda, a modalidade mais usada e a mais cara, a identificação precisa dos contribuintes que terão maior probabilidade de pagar é fundamental para a alocação eficiente dos recursos, evitando envio de cartas desnecessárias. De outro lado, conhecer os “maus pagadores” (contribuintes e imóveis com menor histórico de pagamento do IPTU), possibilitará uma comunicação mais elucidativa, com informações sobre bens públicos, normas sociais, intimidação, simplificação, taxa média de pagamento no bairro, etc [Sobrinho and de Alencar 2019].

Neste sentido, o presente trabalho propõe um modelo baseado em algoritmos de aprendizado de máquina não-supervisionados, que visa aprender e qualificar perfis de contribuintes e imóveis, a partir do histórico de pagamentos do IPTU e suas variações (pagamento via cota única, com desconto, no prazo, parcelado, em dívida ativa, etc). Nossa hipótese de investigação é que um sistema inteligente construído conforme o modelo proposto contribuirá para o envio mais eficaz de notificações, comunicações e cartas de cobrança, além de evitar a seleção manual e *ad-hoc* dos bons ou dos maus pagadores.

Para avaliar o modelo proposto, foi desenvolvido um sistema para Régua de Cobrança do IPTU de 2023 da cidade de Fortaleza, com base nos dados históricos de pagamentos de 2018 até 2022, e foi realizada análise comparativa com o modelo baseline utilizado pela Secretaria Municipal das Finanças de Fortaleza (SEFIN) para cobrança do IPTU em 2022. Os resultados das análises indicaram desempenho promissor do modelo proposto.

O presente trabalho está organizado em cinco seções distintas. Na Seção 2, são apresentados os estudos relacionados a este trabalho. Em seguida, na Seção 3, será detalhada a análise dos dados de arrecadação do IPTU da cidade de Fortaleza, a 4ª capital mais populosa do Brasil, os quais são típicos de uma grande cidade brasileira. Na Seção

¹https://repositorio.enap.gov.br/bitstream/1/5220/1/\%28011\%29lab_Como\%20aumentar\%20o\%20pagamento\%20de\%20impostos\%20em\%20atraso\%3F.pdf

4, é proposto o Modelo para Qualificação de Perfis de Contribuintes do IPTU. Na Seção 5, são discutidos os experimentos realizados e a análise dos resultados. Por fim, a conclusão do trabalho e possíveis direções para futuras investigações são apresentadas.

2. Trabalhos Relacionados

Na literatura existem diversos trabalhos que exploram o uso de técnicas de Inteligência Artificial no setor financeiro, abordando tarefas como a detecção de fraudes e segmentação de clientes. Nesse contexto, o trabalho de [Beltran 2019] é relevante, pois utiliza diversas técnicas de *clustering*, como K-means, DBSCAN com diferentes métricas e um modelo híbrido, para detectar e prevenir ataques cibernéticos e fraudes financeiras. Foram escolhidos K-means e DBSCAN, devido à sua capacidade de agrupar os dados em um alto número de clusters [Pedregosa et al. 2011]. Os experimentos foram conduzidos com o conjunto de dados “*Credit Card Fraud Detection*”² disponível no Kaggle. Os resultados mostraram que o K-means teve uma taxa de detecção de fraudes de 72,76% e uma taxa de detecção de transações normais de 52,57%. O DBSCAN obteve melhores resultados com as métricas Euclidiana e Manhattan, com taxas de detecção de fraudes acima de 85% e taxas de transações normais de até 26%. No entanto, a combinação dos dois algoritmos teve um desempenho inferior. Apesar disso, ambos os algoritmos mostraram eficiência na detecção de fraudes, ressaltando a importância de explorar e aprimorar as técnicas de Inteligência Artificial para melhorar a detecção de fraudes no setor financeiro.

Além disso, um estudo adicional utilizou o algoritmo K-means para corrigir a estrutura do imposto de renda de pessoas físicas. O objetivo do trabalho [de Souza Barbosa 2022] era alcançar uma distribuição mais igualitária e aumentar a arrecadação através desse imposto. O estudo envolveu a extração e manipulação de dados de uma base tributária, seguido do tratamento e filtragem dos dados. O algoritmo K-means foi aplicado para agrupar os dados com base em centroides, identificando grupos de contribuintes com características semelhantes. Em seguida, um processo de otimização definiu as alíquotas de imposto com base nas novas faixas de renda estabelecidas. Os resultados indicaram a necessidade de uma nova alocação de alíquotas e faixas para a estrutura do imposto de renda, promovendo uma distribuição mais justa e equitativa dos encargos fiscais e aumentando a arrecadação de forma eficiente.

O estudo realizado por [Barbosa Júnior 2020] pretende apresentar um novo modelo de cobrança baseado na régua de cobrança na cidade de Fortaleza. Segundo [Ávila et al. 2015], a economia comportamental se baseia nas contribuições da neurociência e de outras ciências humanas e sociais para identificar o perfil do contribuinte ideal. Para isso, foram desenvolvidas ferramentas específicas, como o *Business Intelligence (BI)*, para a gestão e controle da inadimplência, permitindo a definição da melhor estratégia de comunicação e abordagem ao munícipe para que ele regularize sua situação fiscal. Os resultados obtidos no trabalho mostraram-se excelentes, superando o desempenho médio do mercado de recuperação de crédito tributário. Enquanto o mercado corporativo trabalha com taxas entre 4.50% e 6.50%, as ações realizadas pela CATRI/SEFIN representaram apenas 1,38% do montante recuperado. Em conclusão, o estudo evidencia a importância da régua de cobrança e da aplicação da Economia Comportamental no processo de recuperação de créditos tributários. Nosso trabalho, tem relação direta com

²<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

a estratégia delineada por [Barbosa Júnior 2020]. Devido ao fato de que essa abordagem de cobrança, visa empregar diversos canais de comunicação com o cliente, visando a regularização de seus débitos pendentes. No nosso caso, ao identificarmos os indivíduos que não efetuam o pagamento com o envio da carta, conseguimos reduzir os investimentos associados ao envio de correspondências e otimizar a estratégia de cobrança. Dessa forma, abrimos caminho para a exploração de outras estratégias de cobranças.

Com base nestes trabalhos relacionados, tem-se evidências que a utilização de técnicas de Inteligência Artificial, incluindo métodos de *clustering* como o K-means e o emprego de métodos como régua de cobrança são eficientes tanto para a detecção de fraudes no setor financeiro como na correção da estrutura do imposto de renda para pessoas físicas, além de melhorar a eficácia das estratégias de cobrança. Os resultados promissores dos estudos, aqui referenciados, incentivaram a aplicação de métodos para o agrupamento dos dados do IPTU de Fortaleza, visando identificar de forma sistemática e eficaz os perfis dos contribuintes.

3. Análise do Conjunto de Dados relacionados ao IPTU no Município de Fortaleza

Como padrão de referência, foram analisados os dados de imóveis, propriedade, pagamentos, débitos e modalidades de cobrança do IPTU da cidade de Fortaleza, respeitando-se os preceitos da Lei Geral de Proteção de Dados (LGPD)³. Fortaleza é, segundo dados do último Censo Demográfico 2022⁴, a 4ª. cidade mais populosa do Brasil e possui algumas características interessantes. Além de ser a capital com maior densidade demográfica (2.428.678 hab por 312,353 km²)⁵ é uma das cidades mais socialmente desiguais do mundo, sendo classificada como a quinta cidade mais desigual do mundo em 2016 [Xavier 2016].

Os conjuntos de dados fornecidos pela equipe de arrecadação da Secretaria Municipal das Finanças de Fortaleza (SEFIN) continha 5.196.107 registros, referentes aos anos de 2018 a 2023, com informações sobre o imóvel, proprietário (identificação anonimizada), lançamento, pagamentos e débitos do IPTU, bem como sobre o envio de cartas e outras modalidades de comunicação, para cada ano. A Tabela 1 apresenta uma lista dos principais atributos descritivos de cada registro dos conjuntos de dados, nomeado neste trabalho de DSIPTU_FOR.

Fortaleza, no ano de 2022, possui aproximadamente 802 mil imóveis distribuídos para 363 mil proprietários diferentes, dos quais, 353 mil são pessoas físicas. Dos 801.691 imóveis, 562.487 são imóveis residenciais, 163.673 são não residenciais e 75.531 são territoriais e a grande maioria (630 mil) são imóveis de pessoa física, como pode ser visualizado nos gráficos à esquerda da Figura 1. Percebe-se, portanto, alta concentração dos imóveis em apenas 21% da população adulta de Fortaleza (em torno de 1,654 milhões de pessoas).

No gráfico à direita da Figura 1, tem-se a distribuição dos imóveis quanto a

³https://www.planalto.gov.br/ccivil/_03/_ato2015-2018/2018/lei/113709.htm

⁴<https://censo2022.ibge.gov.br/>

⁵https://pt.wikipedia.org/wiki/Lista_de_munic%C3%A9pios_do_Brasil_por_densidade_demogr%C3%A1fica

Tabela 1. Principais atributos do conjunto de dados original DSIPTU_FOR, sobre o IPTU de Fortaleza.

ATRIBUTO	DESCRIÇÃO
NUMEXERCICIO	Exercício a que se refere o tributo
IDCONTRIBUINTE	identificacao do contribuinte resultado do processo de anonimização.
FLAGTEMRESTRICAO	Indica se contribuinte possui restrição no cadastro. Pode assumir três valores: “Sem restrição”; “Com restrição Não Impeditiva”; “Com pelo menos uma restrição impeditiva”
NUMINSCRICAO	Número de inscrição do imóvel.
DESCEMAIL	Email de correspondência do sujeito passivo do imóvel.
DESCTELEFONECELULAR	Telefone de correspondência do sujeito passivo do imóvel.
VALIMPOSTOCALCULADO	Valor do imposto calculado de acordo com as características do imóvel.
VALIMPOSTOBENEFICIO	Valor do benefício, quando houver, que o imóvel possui.
VALIMPOSTODEVIDO	Valor do imposto devido. É a diferença entre o valor do imposto calculado e o valor do benefício.
DESCSITUACAODEBITO	Situação atual do débito. Pode ter cinco possibilidades: Aberto: ainda existe, para o exercício, débito em aberto do imóvel. Extinto: débito do imóvel referente ao exercício foi pago. Imunidade: o imóvel possui o benefício de imunidade. Isenção: o imóvel possui o benefício de isenção Suspensão: débito suspenso devido a parcelamento e outros
VALPRINCIPAL	Valor principal do débito.
VALABERTO	Valor em aberto do imposto. Inclui parcelas vencidas e a vencer.
VALPARCELASVENCIDAS	Valor total das parcelas que já venceram.
QTDPARCELAS	Quantidade de parcelas em que o débito foi dividido.
QTDPARCELASPAGAS	Quantidade de parcelas pagas.
QTDPARCELASVENCIDAS	Quantidade de parcelas vencidas.
VALLANCADO	Valor lançado do imposto no referido exercício.
VALDESCONTO	Valor do desconto obtido no pagamento em cota única.
VALPRINCIPALPAGO	Valor principal do imposto pago.
VALJURO	Valor pago com juros.
VALMULTA	Valor pago em multa.
VALTOTALPAGO	Valor total pago. É o somatório do valor principal e de juros e multa.
FLAGCOTAUNICA	Informa se o débito foi pago em cota única.
FLAGDIVIDAATIVA	Informa se o débito do imóvel está inscrito em dívida ativa. Assume o valor 1, se inscrito em dívida ativa e 0 caso contrário.
QTDNOTIFICAOEMAIL	Quantidade de notificações por email enviadas referente ao imóvel.
QTD CARTA	Quantidade de notificações por carta enviadas referente ao imóvel.
QTDWHATTSAPP	Quantidade de notificações por mensagem de whatsapp enviadas referente ao imóvel.

Fonte: Elaborada pelos autores.

situação do débito do IPTU, nos anos de 2018 a 2022. Os débitos podem assumir os seguintes estados: Aberto, Extinto, em Imunidade, Isento e Suspensão. Importante observar que gradativamente, a cada ano, vem aumentando o número de imóveis que terminam o ano com débito de IPTU “Aberto” (não pago ou pago parcialmente), ou seja, a inadimplência vem aumentando nos últimos anos na cidade de Fortaleza. Por exemplo, no final do ano de 2018, 130 mil imóveis apresentavam débito do IPTU em aberto, e, no final de 2022, este número aumentou para 235 mil imóveis. Importante considerar que, neste período, ocorreu a pandemia da COVID-19 e a arrecadação de impostos, em geral, sofreu queda expressiva em todo o país.

Em relação ao envio de comunicações e mensagens aos contribuintes, observou-se que, dos 363 mil proprietários de imóveis, presentes no DSIPTU_FOR, apenas 112.947

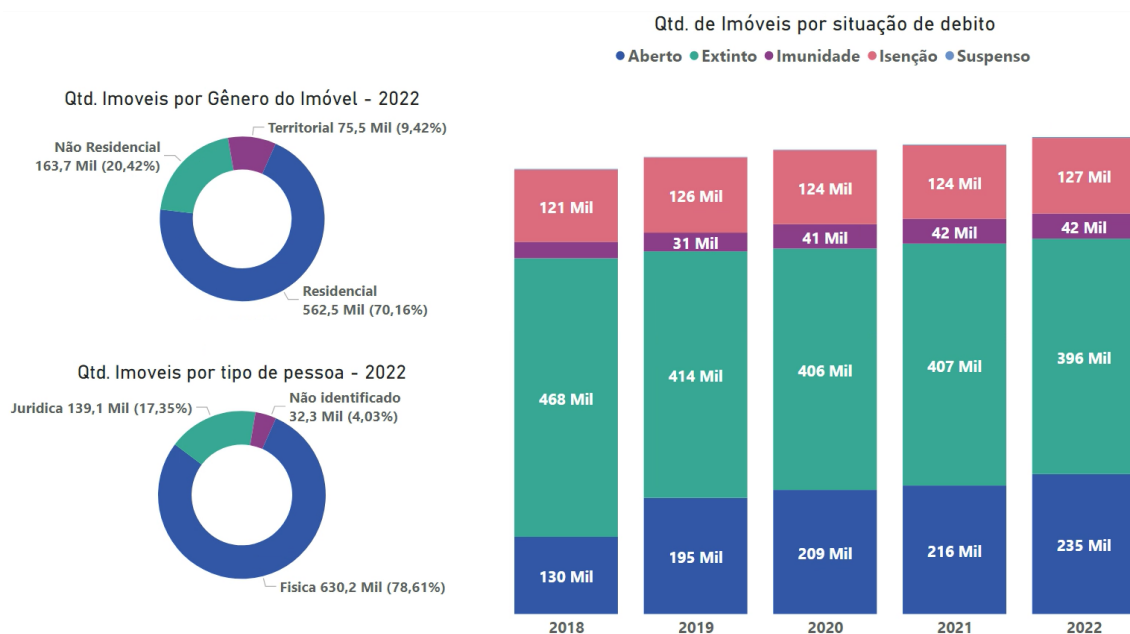


Figura 1. Descrição dos dados de Imóveis, propriedade e débito de IPTU da cidade de Fortaleza (DSIPTU_FOR).

Fonte: Elaborada pelos autores.

possuem o número de celular cadastrado e 159.802 possuem o e-mail cadastrado. Considerando essas informações, conclui-se que o principal método de envio de notificações e cobrança é o envio do DAM IPTU pelo Correios, ainda que seja o mais caro (o envio básico de uma carta de até 20g custa R\$ 2,45).

Foi analisado o envio de cartas via Correios nos últimos anos, conforme apresentado na Figura 2. No ano de 2022, houve um aumento significativo no número de cartas enviadas em comparação com o ano de 2021 (46% de aumento). Considerando que o custo de envio básico de cartas pelo Correios de R\$2,45, estima-se que a SEFIN tenha investido aproximadamente R\$ 2 milhões nessa modalidade de cobrança no ano de 2022.

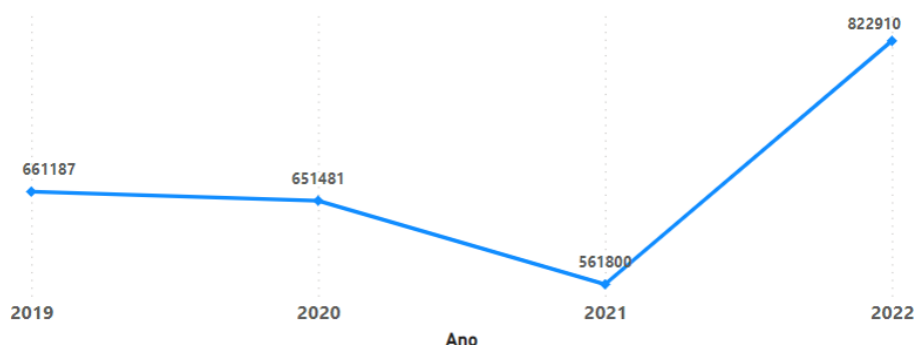


Figura 2. Quantidade de cartas, enviada via Correios, de 2019 a 2022, aos contribuintes de IPTU da cidade de Fortaleza.

Em outra análise relacionada ao envio recorrente de cartas de cobrança, conforme ilustrado na Figura 3, pode-se observar que o ano de 2022 apresentou o maior número de

repetições de envio para um mesmo imóvel. A maioria dos imóveis (299.708) recebeu apenas uma carta, enquanto 51.825 imóveis não receberam nenhuma. No entanto, mais de 200 mil imóveis receberam mais de duas cartas, sendo que 75.098 deles receberam três cartas. Estes números foram bem maiores que seus correlatos do ano de 2021.

Comparando estas informações com a situação dos débitos ao final de 2022, percebe-se que o aumento no número de cartas enviadas não gerou um aumento da adimplência. Uma hipótese a ser investigada é se direcionarmos melhor o envio de cartas por perfil de contribuinte, teremos melhor resposta em termos de arrecadação. No pior caso, em que não haja dependência entre estes eventos aleatórios - o pagamento do imposto e o recebimento de cartas de cobrança - o município, pelo menos terá investido menos.

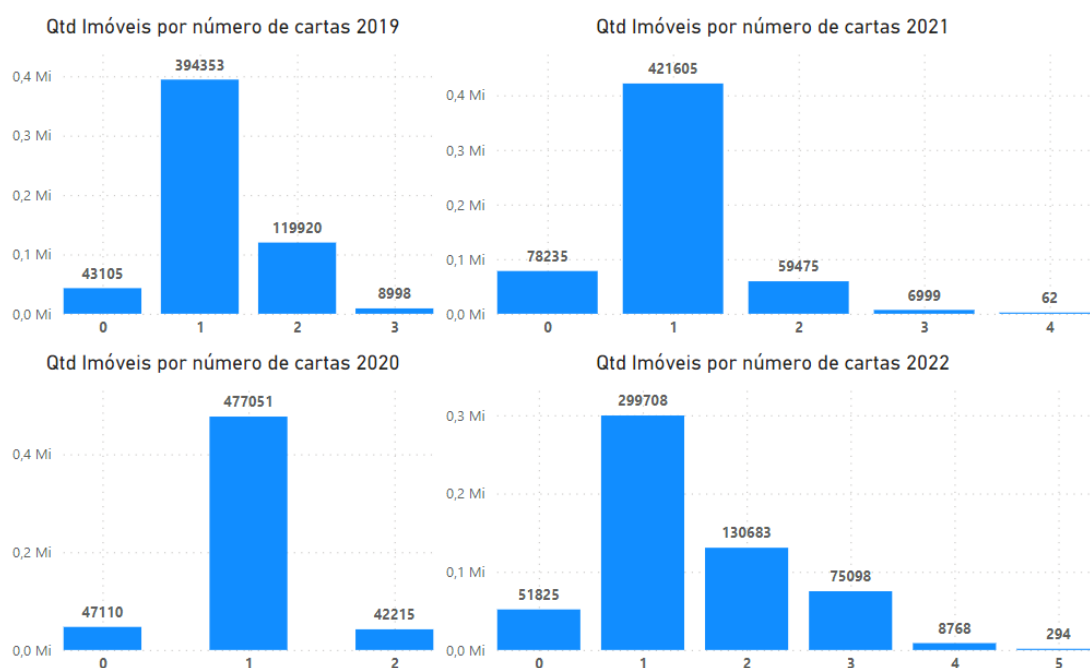


Figura 3. Quantidade de imóveis, por quantidade de cartas recebidas via Correios, nos anos de 2019 a 2022

Neste sentido, o objetivo desse trabalho é propor um modelo para qualificação do perfil do contribuinte/imóvel, com base no histórico e características de pagamentos e situação do débito de IPTU de cada contribuinte/imóvel. Uma aplicação prática do modelo a ser desenvolvido é otimizar o envio de cartas de cobrança, visando aumento da adimplência do IPTU.

4. Modelo para Qualificação de Perfil de Contribuinte do IPTU

4.1. Modelagem do Conjunto de Dados para Treinamento

Na modelagem do conjunto de dados para treinamento do modelo de aprendizagem automática, um passo fundamental é a seleção do objeto da análise e a seleção das *features* (ou características) do objeto a ser qualificado. A primeira decisão foi considerar o contribuinte/imóvel como objeto a ser qualificado, pois foi observado que o comportamento

fiscal de um contribuinte em relação aos seus imóveis (quando este possui mais de um imóvel) pode diferir. Ou seja, um contribuinte pode ser “bom pagador” do IPTU do imóvel em que reside e ser “mau pagador” do IPTU de seus outros dois imóveis, por exemplo, que adquiriu para investimento. Uma vantagem de qualificar o comportamento do pagamento do contribuinte no nível de cada imóvel de sua propriedade é que será possível agregar e inferir a qualificação do contribuinte por regras heurísticas. O contrário não seria possível.

Foi realizada esta modelagem no conjunto de dados original DSIPTU_FOR. Como descrito na seção anterior, o conjunto de dados fornecido pela SEFIN continha 5.196.107 linhas, referentes aos anos de 2018 a 2023, e 60 colunas contendo diversas informações sobre o imóvel, o contribuinte proprietário, e o IPTU relacionado ao imóvel. A maioria das informações presentes no *dataset* eram de natureza cadastral da pessoa (proprietário) e do imóvel (nome do contribuinte, CPF, número da inscrição, email, telefone, endereço, data de construção do imóvel, latitude, longitude, descrição de uso do imóvel, etc.), e não foram utilizados como entrada do modelo por não terem relação direta com a cobrança. Dessa forma, para compor o *dataset* de treinamento foram selecionadas as colunas do DSIPTU_FOR relacionadas às seguintes *features*: situação do débito, pagamento em cota única, débito em dívida ativa do município, valor principal do IPTU, valor em aberto do IPTU, e quantidade de parcelas vencidas do IPTU. Uma rotina de pré-processamento gerou as *features* com seus respectivos valores em (*labels*), para cada exercício (2018 a 2022), a seguir explanadas:

- **Débito Aberto** - representa a situação de débito do imóvel com valor 0 (débito não está aberto) e 1 (débito em aberto - não pago ou pago parcialmente);
- **Cota Única** - indica se o contribuinte pagou o IPTU em única cota (valor 1) ou em pagamento parcelado (valor 0);
- **Dívida Ativa** - indica se o imóvel possui o débito do IPTU inscrito em dívida ativa (valor 1) ou não (valor 0);
- **Tipo do imóvel** - informa se o imóvel é residencial (valor 0), não residencial (valor 1) ou territorial (valor 2);
- **Tipo de pessoa** - representa o tipo do proprietário do imóvel - pessoa física (valor 0), pessoa jurídica (valor 1) ou não identificado (valor 2);
- **Pagamento atrasado** - indica se o pagamento do IPTU foi realizado em dia (valor 0) ou com atraso (valor 1).

Adicionalmente, foi necessário desenvolver e aplicar uma rotina de pós-tratamento dos dados para selecionar apenas os imóveis com débito em aberto e extinto e remoção de linhas duplicadas. Em seguida, foram selecionados os imóveis com informações em todos os anos, de 2018 a 2022, visando a completude da análise histórica. Ao final desse processo, o conjunto de dados de treinamento continha 566.376 imóveis distintos. A Figura 4 apresenta um extrato dos dados de treinamento.

4.2. Seleção do Modelo de Aprendizagem Automática

Para a realizar a qualificação do perfil de contribuintes/imóveis, utilizou-se de abordagem não supervisionada por meio do algoritmo de *Machine Learning* (ML) K-means, da biblioteca Sklearn⁶. K-means é amplamente utilizado em análises e mineração de dados por

⁶<https://scikit-learn.org/stable/>

Debito_Aberto_2022	Cota_Unica_2022	Divida_ativa_2022	Tipo_imovel_2022	Tipo_Pessoa_2022	pag_atrasado_2022	Debito_Aberto_2021	Cota_Unica_2021	Divida_ativa_2021	Tipo_imovel_2021
1	0	0	0	0	0	1	1	0	1
0	1	0	0	0	0	0	0	1	0
1	0	0	0	0	0	1	1	0	0
1	0	0	0	0	0	1	0	1	0
0	1	0	1	0	0	0	0	1	0
...
0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	1	0

Figura 4. Exemplo dos dados de treinamento do Modelo de Qualificação de Perfil de Contribuinte de IPTU, no caso da cidade de Fortaleza.

ser um método eficaz de aprendizado não supervisionado. Sua abordagem flexível não exige que o esquema de cada cluster seja definido com antecedência, tornando-o particularmente adequado para a tarefa de agrupamento de dados [Cardoso et al. 2008].

Uma primeira etapa é determinar o número apropriado de clusters (K) a ser usado para agrupar os dados. Isso pode ser feito usando técnicas de validação interna ou externa, como o método do cotovelo ou o índice de validação externa [Cardoso et al. 2008]. Dependendo do número de clusters selecionados, o algoritmo K-means realiza o agrupamento dos dados, atribuindo cada objeto a um cluster específico com base em sua semelhança com outros dados. Normalmente, a similaridade é medida usando a distância euclidiana entre os pontos de dados [Arthur and Vassilvitskii 2007]. O objetivo principal do K-means é dividir um conjunto de dados em grupos homogêneos, maximizando a similaridade dentro do grupo e minimizando a similaridade inter-grupo [Ochi et al. 2004].

Para determinar a quantidade ideal K de *clusters* e identificar a melhor qualificação para os contribuintes/imóveis, realizaram-se quatro experimentos utilizando 3, 4, 5 e 6 *clusters*, com base no Método *Elbow Curve*. Em seguida, especialistas da SEFIN analisaram as características de cada *cluster* em cada um dos quatro experimentos: situação de débito, situação de dívida ativa, pagamento atrasado e pagamento em cota única. Foram selecionados o 2o. experimento ($k = 4$) e o 4o. experimento ($k = 6$) como os mais promissores.

Na Figura 5 são apresentadas as características do 2o. experimento com $k = 4$ clusters. Neste, os cluster 0 e 3 foram qualificados como “Bom Pagador” e o cluster 1 como “Mau Pagador”. Porém, para o cluster 2, com 140 mil imóveis, não foi possível determinar uma qualificação exata, pois houve bastante interseção com outros clusters. Teoricamente, ele deveria ser qualificado como “Mau Pagador”, pois apresentou imóveis com débito em aberto, na dívida ativa ou não, e com o pagamento em atraso. Porém, durante uma análise detalhada de cada variável, os especialistas concluíram que o cluster 2 não conseguiu qualificar bem os imóveis, mesclando “Bom Pagador” com imóveis “Mau Pagador”. O principal motivo disso ter acontecido foram “Bons Pagadores” com pagamento em atraso.

O 4o. experimento, com $k = 6$ *clusters*, foi o que melhor qualificou os dados em estudo, além de atribuir permitir uma melhor qualificação que representasse cada *cluster*. A Figura 6 apresenta os resultados da análise de cada característica para os 6 clusters do 4o experimento e as respectivas qualificações:

- **“Bom Pagador”** - em sua grande maioria, caracterizam-se por não possuir débito

Cluster	Cota_Unica	Deb. Aberto	Div.Ativa	Pag.Atrasado	Qtd_Imóveis	Qualificação
0	Sim/Não	Não	Não	Não	118k	Bom_pagador
1	Não	Sim	Sim	Sim	73k	Mau_pagador
2	Não	Sim	Sim/Não	Sim	140k	???
3	Sim/Não	Não	Não	Não	133k	Bom_pagador

Figura 5. Principais características de cada *cluster* do 2o. experimento e a respectiva “Qualificação”, no caso dos dados de IPTU de Fortaleza

Cluster	Pag. Atrasado	Div. Ativa	Deb. Aberto	Cota_Unica	Qtd_Imóveis	Qualificação
0	Não/Sim	Não/Sim	Não	Não	42k	Possível_Mau_pagador
1	Sim	Sim	Sim	Não	117k	Mau_pagador
2	Não	Não	Não	Sim	131k	Bom_pagador
3	Não	Não	Não	Sim	189k	Bom_pagador
4	Sim	Sim	Sim	Não	65k	Mau_pagador
5	Sim	Não	Sim	Não	20k	Possível_Bom_pagador

Figura 6. Principais características de cada *cluster* do 4o. experimento e a respectiva “Qualificação”, no caso dos dados de IPTU de Fortaleza

em aberto, não ter sido inscrito na dívida ativa nos últimos anos analisados, não está com o pagamento em atraso e terem efetuado o pagamento em cota única.

- **“Mau Pagador”** - definem-se assim por, em sua grande maioria, possuir débito em aberto, terem sido inscritos na dívida ativa em alguns dos últimos anos analisados, estarem com o pagamento em atraso e não efetuarem o pagamento em cota única.
- **“Possível Bom Pagador”** - caracterizam-se assim por, em sua grande maioria, possuírem o débito em aberto, não terem sido inscrito na dívida ativa nos últimos anos analisados, estarem com o pagamento em atraso e não efetuarem o pagamento em cota única.
- **“Possível Mau Pagador”** - em sua grande maioria, não possuem débito em aberto, foram inscritos na dívida ativa em alguns dos últimos anos analisados, sim e não para pagamento em atraso e não efetuaram o pagamento em cota única.

Na Figura 7 é apresentada a distribuição dos contribuintes/imóveis do conjunto de treinamento do IPTU de Fortaleza por Qualificação. Observa-se que a maioria dos contribuintes/imóveis (320 mil) foram classificados como “Bom Pagador”, seguido do número de “Mau Pagador” = 183.630 mil contribuintes/imóveis.

5. Avaliação Experimental

Para avaliar o modelo de qualificação do perfil do contribuinte/imóvel de IPTU, aqui proposto, foi desenvolvido um sistema para Régua de Cobrança do IPTU de 2023 da cidade de Fortaleza. Neste experimento, serão selecionados, para envio da 1ª notificação de lançamento e DAM IPTU, somente os contribuintes/imóveis qualificados como “Bom Pagador”, numa abordagem conservadora, e comparando com o sistema baseline utilizado pela SEFIN para cobrança do IPTU em 2022. Neste caso, tomou-se como referência os

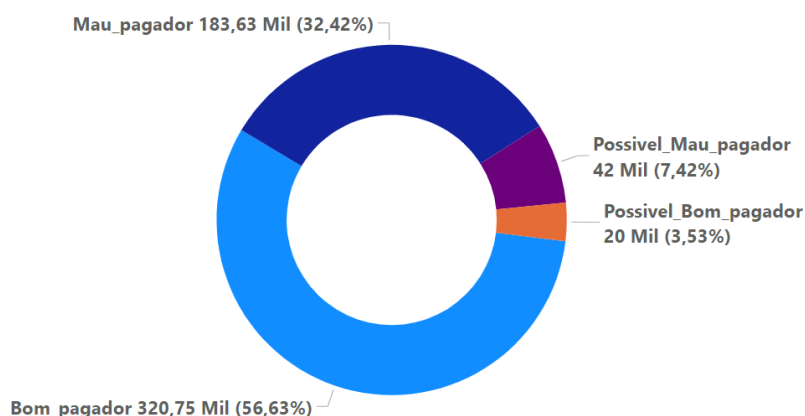


Figura 7. Qualificação dos Contribuintes/Imóveis do IPTU de Fortaleza, constantes do conjunto de treinamento do modelo proposto.

imóveis que receberam apenas uma carta no ano de 2022, provavelmente a primeira carta de notificação do lançamento e DAM IPTU de 2022.

Como já mencionado, em 2022 foram enviados 822.910 cartas e desse montante 299.708 mil imóveis receberam apenas a 1ª carta. Numa primeira análise, este conjunto foi qualificado conforme resultado do modelo treinado para Fortaleza. No primeiro gráfico da Figura 8, é apresentada a quantidade de imóveis por cada qualificação. Tem-se que 81,85% dos imóveis (245.322) que receberam apenas uma carta foram qualificados pelo modelo proposto como “Bom Pagador”. Ainda no lado direito da Figura 8, apresentam-se os gastos em R\$ com o envio das cartas por cada tipo de contribuinte/imóvel, no total de R\$ 734 mil. Interessante observar que o sistema baseline da SEFIN em 2022 enviou 48.923 cartas para contribuintes/imóveis considerados “Mau Pagador” ou “Possível Mau Pagador”, iniciando desperdício de R\$ 120 mil, segundo o modelo proposto.

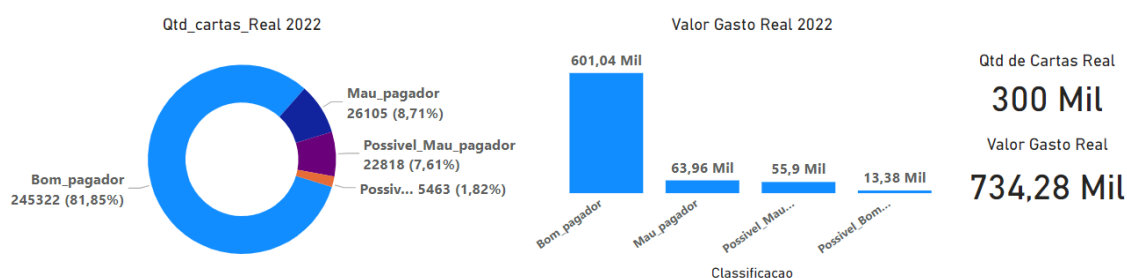


Figura 8. Análise do envio real de cartas de IPTU no ano de 2022 em Fortaleza, segundo o modelo proposto neste trabalho.

A Figura 9 apresenta um diagrama com dois grupos selecionados para envio da 1ª notificação de lançamento de IPTU, de acordo com as abordagens a seguir:

- **“Baseline SEFIN 2022 (em azul)”** - contribuintes/imóveis selecionados manualmente e com critérios ad-hoc pelos especialistas em cobrança e arrecadação da SEFIN em 2022;
- **“Modelo Proposto (em amarelo)”** - contribuintes/imóveis qualificados como

“Bom pagador” pelo modelo proposto, treinado com os dados históricos do IPTU de Fortaleza de 2018 a 2022.

No diagrama da Figura 9, tem-se que o Baseline SEFIN selecionou 299.708 imóveis e o Modelo proposto selecionou 320.746 imóveis para receber a 1ª DAM IPTU. Destes, 245.322 imóveis estão nos dois grupos, ou seja, são VP (verdadeiros-positivos), implicando numa precisão do modelo proposto = 76,48%. De outro lado, 54.386 imóveis não foram selecionados pelo modelo e foram pelo baseline SEFIN, importando numa cobertura de 81,85%. Portanto, o modelo proposto neste trabalho, treinado com os dados de IPTU de Fortaleza, apresentou F1-Score = 79%.

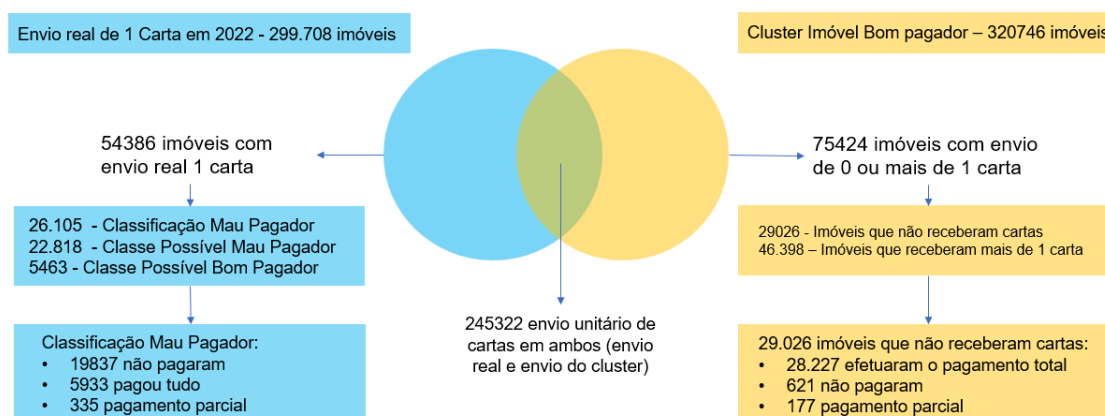


Figura 9. Diagrama Comparativo dos grupos selecionados para envio da 1ª notificação de lançamento e cobrança de IPTU, pelas abordagens Baseline SEFIN (azul) e Modelo proposto (amarelo).

Fonte: Elaborada pelos autores.

A seguir apresentamos análises detalhadas dos possíveis casos de erro do modelo proposto - casos FP (falso-positivo) e FN (falso-negativo).

A Figura 10 apresenta uma análise mais detalhada dos 54.386 imóveis, considerados FN, com informações relacionadas à variável de pagamento do IPTU. Os rótulos utilizados são 0, 1 e 2, que indicam respectivamente se o IPTU do imóvel não foi pago (0), se houve pagamento integral (1) ou se houve pagamento parcial (2). Essa análise mais granular nos permite fazer algumas inferências em relação aos casos FN. A seguir, detalharemos as análises:

1. Para a classe “Mau Pagador”, verifica-se que 19.837 contribuintes não pagaram nada do imposto devido, 5.933 efetuaram o pagamento total e 335 pagaram ao menos 1 parcela, ou seja, dos 26.105 imóveis classificados como “Mau Pagador” pelo modelo, 19.837 realmente foram maus pagadores;
2. Para os imóveis do grupo “Possível Mau Pagador”, classificados dessa forma por terem sido incluídos no cadastro de dívida ativa nos anos anteriores, 138 não pagaram nada de IPTU, e 22.536 efetuaram o pagamento completo, quase a totalidade, ou seja, o envio da carta pode ter provocado o contribuinte a efetuar o pagamento;
3. Na classe “Possível Bom Pagador”, a grande maioria, 4.290, apesar de não terem histórico de inscrição na dívida ativa, por algum motivo não efetuaram o pagamento.

A partir do comportamento real dos contribuintes, somando-se os que não efetuaram pagamento de nenhum valor de IPTU em 2022 (19.837 (do grupo “Mau Pagador” + 138 (do grupo “Possível Mau Pagador”) + 4.290 (do grupo “Possível Bom Pagador”), tem-se que 24.265 foram classificados para não receberem carta em 2022 pelo modelo proposto e realmente não efetuaram pagamento (mesmo tendo recebido carta na realidade). Neste caso, avaliamos que os casos reais de FN do modelo foram 30.121 imóveis (54.386 - 24.265), elevando a cobertura do modelo para 89,94%.

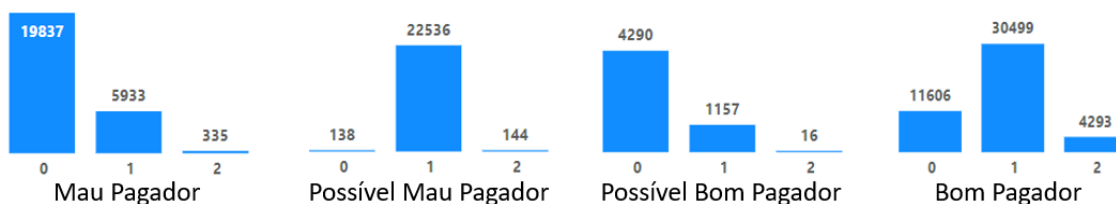


Figura 10. Situação de pagamento real do IPTU de 2022 em Fortaleza, para cada qualificação gerada pelo modelo proposto.

Com relação aos casos de FP, temos um grupo de 75.424 imóveis classificados como “Bom Pagador”, para os quais seria adequado o envio da 1ª carta de cobrança pelo modelo proposto. Conforme Figura 11, constata-se que, desse grupo, 29.026 não receberam cartas e 46.398 receberam mais de uma carta. Dos que não receberam nenhuma carta, 28.227 efetuaram pagamento total, e dos que receberam mais de uma carta, 30.499 efetuaram pagamento total do IPTU em 2022. Portanto, dos 75.424 casos de FP, temos que 58.726 apresentaram, de fato, comportamento de bons pagadores. Logo, a precisão real do modelo proposto foi = 94,79%. Por fim, o modelo proposto, após análise do comportamento real durante o exercício de 2022, apresenta F1-Score = 92,30%.

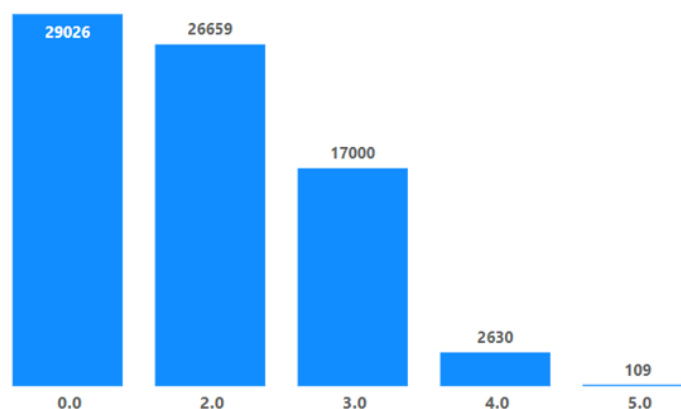


Figura 11. Distribuição dos imóveis Falsos-positivos do modelo proposto (75.424) em relação à quantidade de cartas que receberam em 2022 pelo sistema Baseline SEFIN.

Tendo em vista as análises apresentadas, do ponto de vista financeiro, o modelo proposto de envio da 1ª carta de DAM IPTU para os bons pagadores representa um aumento de custo de R\$ 51 mil. No entanto, ao contabilizar o total arrecadado em ambos os grupos, o grupo Baseline SEFIN, com 299.708 imóveis, importou numa arrecadação

de R\$431 milhões de IPTU em 2022; e o grupo do Modelo proposto (qualificação “Bom Pagador”), com 320.746 imóveis, arrecadou R\$506 milhões, iniciando, portanto, a efetividade do modelo proposto de qualificação de contribuintes para fins de notificação e cobrança do IPTU.

6. Considerações Finais

Neste trabalho, propomos um modelo baseado em aprendizado de máquina não supervisionado para qualificar perfis de contribuintes e imóveis utilizando o histórico de pagamentos do IPTU. Os principais objetivos do modelo é otimizar o envio de notificações, comunicações e cartas de cobrança e aumentar a eficiência na arrecadação de impostos, na medida em que fornece uma análise mais precisa dos padrões de pagamento e evita a seleção manual e *ad-hoc* de bons ou maus pagadores.

Para validar este trabalho, foi desenvolvido um sistema de Régua de Cobrança do IPTU para o ano de 2023 na cidade de Fortaleza. Utilizamos dados históricos de pagamentos do IPTU nos anos de 2018 a 2022 e realizamos uma análise comparativa com o modelo *baseline*, usado na Secretaria Municipal das Finanças de Fortaleza (SEFIN) para a cobrança do IPTU em 2022. Os experimentos realizados demonstraram um desempenho promissor do modelo treinado. Com a abordagem de seleção de bons pagadores para envio da 1a. carta com a notificação e lançamento do IPTU, o modelo obteve F1-Score de 92,30% e um aumento estimado da arrecadação da ordem de R\$ 75 milhões.

Como trabalhos futuros, em 2023 e 2024 os resultados do modelo proposto neste trabalho e implementado na SEFIN serão coletados para suportar análises estatísticas e de correlação entre as variáveis e os eventos aleatórios de cobrança e pagamento de impostos. Adicionalmente, pretende-se avançar na aplicação de outros algoritmos de aprendizado de máquina, utilizar dados socioeconômicos atualizados do IBGE, visando aprimorar ainda mais a definição dos perfis, além de usar o modelo para direcionar experimentos em ciência comportamental.

Referências

- Afonso, J. R. R., Araújo, E. A., and Nóbrega, M. A. R. d. (2013). O iptu no brasil: um diagnóstico abrangente.
- Arthur, D. and Vassilvitskii, S. (2007). K-means++ the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035.
- Ávila, F., Bianchi, A. M., and Motta, L. T. (2015). *Guia de economia comportamental e experimental*. EconomiaComportamental.org.
- Barbosa Júnior, N. L. (2020). Economia comportamental e régua de cobrança para recuperação de créditos tributários na cidade de fortaleza.
- Beltran, R. D. (2019). Detecção de fraudes bancárias utilizando métodos de clustering.
- Cardoso, G. d. S., Nascimento, V. L., de Oliveira Nunes, E., and Neto, E. d. S. M. (2008). Clusterização k-means: Uma proposta de melhoria.
- Cupertino, S. A., Faroni, W., Durante, M. O., and Abrantes, L. A. (2014). Análise da inadimplência na arrecadação do iptu no município de viçosa–minas gerais. *Gestão Pública: Práticas e Desafios*, 5(2).

- de Souza Barbosa, P. V. (2022). Justiça social e eficiência para o cálculo do imposto de renda de pessoa física. *BIUS-Boletim Informativo Unimotrisaúde em Sociogerontologia*, 32(26):1–13.
- Ochi, L. S., Dias, C. R., and Soares, S. S. F. (2004). Clusterização em mineração de dados. *Instituto de Computação-Universidade Federal Fluminense-Niterói*, 1:46.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.
- Sobrinho, A. L. and de Alencar, F. A. G. (2019). Análise do esforço fiscal na arrecadação do iptu para os municípios do estado do ceará utilizando modelo de fronteira estocástica. *Revista Controle-Doutrina e Artigos*, 17(1):170–200.
- Xavier, L. N. (2016). Fortaleza da desigualdade e violência: Geopolítica do medo e anomia social como fator de produção da violência concentrada e da sensação de insegurança. *Conpedi Law Review*, 2(1):112–130.