

idp

idn

MESTRADO PROFISSIONAL

EM ECONOMIA

**MACHINE LEARNING E A PREVISÃO DE PREÇOS
DE TERRENOS EM BRASÍLIA**

WILLAMY MAMEDE DA SILVA DIAS

Brasília-DF, 2023

WILLAMY MAMEDE DA SILVA DIAS

**MACHINE LEARNING E A PREVISÃO DE PREÇOS
DE TERRENOS EM BRASÍLIA**

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Economia, Políticas Públicas e Desenvolvimento, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como parte dos requisitos para obtenção do título de Mestre.

Orientador

Professor Doutor Leonardo Monteiro Monasterio

Brasília-DF 2023

WILLAMY MAMEDE DA SILVA DIAS

MACHINE LEARNING E A PREVISÃO DE PREÇOS DE TERRENOS EM BRASÍLIA

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Economia, Políticas Públicas e Desenvolvimento, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como parte dos requisitos para obtenção do título de Mestre.

Aprovado em 14 / 09 / 2023

Banca Examinadora

Prof. Dr. Leonardo Monteiro Monastério - Orientador

Prof. Dr. Guilherme Mendes Resende

Prof. Dr. João da Costa Pantoja

-
- D541 Dias, Willamy Mamede da Silva
- Machine Learning e a previsão de preços de terrenos em Brasília / Willamy Mamede da Silva Dias. – Brasília: IDP, 2023.
- 69 p.
- Inclui bibliografia.
- Trabalho de Conclusão de Curso (Tese/Dissertação) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa – IDP, Curso de Mestrado Profissional em Economia, Brasília, 2023.
- Orientador: Prof. Dr. Leonardo Monteiro Monasterio.
1. Precificação de terreno. 2. Aprendizado de máquina. 3. Modelo hedônico. 4. Terracap. I. Título.
- CDD: 332

Ficha catalográfica elaborada pela Biblioteca Ministro Moreira Alves
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

AGRADECIMENTOS

A Deus, pelo dom da vida e por sua infinita graça e misericórdia, que tem transbordado em minha vida.

A minha esposa Vanessa, meus filhos Arthur e Bruna, que tiveram paciência e compreensão pelos momentos de isolamento.

A Terracap, que financiou este empreendimento acadêmico.

Ao meu orientador, Leonardo Monastério, pelos ensinamentos compartilhados ao longo do trabalho.

RESUMO

Precificação de imóveis foi, por muito tempo, uma atividade restrita a especialistas do setor imobiliário. Muitas vezes o solo urbano não recebe o devido dimensionamento do seu valor, ora supervalorizado, ora subvalorizado, causando distorções e perdas econômicas. A tecnologia de sistemas autônomos com aprendizado de máquina é capaz de prever e aprender sobre o comportamento do consumidor. Partindo de uma regressão linear múltipla e adotando um modelo de precificação hedônico que considerou muitas variáveis intrínsecas e extrínsecas de imóveis, foi possível prever com certo nível de acurácia os preços prováveis de negociação dos terrenos comercializados no período de 2000 a 2020 pela Companhia Imobiliária de Brasília - Terracap, para as mais variadas tipologias arquitetônicas. Esse modelo foi construído especificamente para auxiliar a Terracap a determinar valores para os terrenos ofertados em licitação pública no Distrito Federal. A precificação hedônica se mostrou uma boa preditora de valores, com resultados muito próximos dos valores efetivamente transacionados em licitação pública no período avaliado.

Palavras-chaves: precificação de terreno, aprendizado de máquina, modelo hedônico, Terracap.

ABSTRACT

Real estate pricing has long been an activity restricted to real estate sector experts. Often, urban land does not receive the proper dimensioning of its value, sometimes overvalued, sometimes undervalued, causing distortions and economic losses. Autonomous systems technology with machine learning is capable of predicting and learning about consumer behavior. Starting from a multiple linear regression and adopting a hedonic pricing model that considered many intrinsic and extrinsic variables of properties, it was possible to predict with a certain level of accuracy the likely trading prices of parcels marketed from 2000 to 2020 by the Real Estate Company of Brasília - Terracap, for the most varied architectural typologies. This model was specifically built to assist Terracap in determining values for the lands offered in public bidding in the Federal District. Hedonic pricing proved to be a good predictor of values, with results very close to the values effectively transacted in public bidding during the evaluated period.

Keywords: Land valuation, machine learning, hedonic price model, Terracap.

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AIC	Critério de Informação de Akaike
BIC	Critério de Informação Bayesiano
COD	Coefficiente de Dispersão da Mediana
DF	Distrito Federal
FAT	Fundo de Amparo do Trabalhador
GWR	Regressão Geograficamente Ponderada
IDP	Instituto Brasileiro de Direito Público
LIK	Função de Verossimilhança
LM	Multiplicador de Lagrange
MSE	Erro Quadrático Médio
MWR	Regressão de Janelas Móveis
Med R	Mediana das Razões de Avaliação
NBR	Norma Brasileira
NOVACAP	Companhia Urbanizadora da Nova Capital
OLS	Regressão dos Mínimos Quadrados Ordinários
POI	Pontos de interesse
PRD	Diferencial Relativo ao Preço
RGP	Regressão Geograficamente Ponderada
RMS	Raiz do Valor Quadrático Médio
RMSE	Raiz Quadrada do Erro Quadrático Médio
TERRACAP	Companhia Imobiliária de Brasília
UTM	Universal Transversa de Mercator

LISTA DE ILUSTRAÇÕES

Figura 1

Distribuição espacial da ocupação urbana da cidade de Brasília.

.....31

LISTA DE TABELAS

Tabela 1

Evolução dos investimentos em infraestrutura realizados pela Terracap (em Milhares de Reais)

.....14

Tabela 2

Número de imóveis cadastrados na Terracap por Administração Regional

.....34

Tabela 3

Variáveis locacionais e de acessibilidade

.....40

Tabela 4

Relação das Variáveis independentes utilizadas no modelo - Físicas

.....40

Tabela 5

Relação das Variáveis independentes utilizadas no modelo - Locacionais

.....41

Tabela 6

Relação das Variáveis independentes utilizadas no modelo - Financeiras, temporais e negociais

.....42

Tabela 7

Estatística descritiva das variáveis do modelo de dados

.....42

Tabela 8

Ranking dos modelos treinados segundo o R^2

.....50

Tabela 9

Ranking das principais características em termos de importância de permutação

.....53

LISTA DE GRÁFICOS

Gráfico 1

Histograma da variável dependente

39

Gráfico 2

Plotagem dos valores reais x previstos do modelo Hist Gradient Boosting Regressor

52

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO DE LITERATURA	18
2.1	Precificação de Imóveis	18
2.2	Influência Espacial no Preço de Imóveis.....	19
2.3	Modelo de Preços Hedônicos.....	20
2.4	Estimativas e Modelos.....	26
3	CARACTERÍSTICAS IMOBILIÁRIAS DO DISTRITO FEDERAL	31
3.1	Mercado Imobiliário no DF.....	32
3.2	A Companhia Imobiliária de Brasília - Terracap.....	33
4	DADOS E PROCESSAMENTO	37
4.1	Variáveis incluídas no modelo.....	38
5	METODOLOGIA	44
5.1	Algoritmos de Aprendizado de Máquina.....	46
6	RESULTADOS E DISCUSSÃO	49
7	CONCLUSÃO	55
	Referências	58
	Apêndice	65

1



1

INTRODUÇÃO

Em geral, a precificação de imóveis no Brasil é feita caso a caso com base na experiência e conhecimento de corretores ou outros especialistas. Brasília segue esse padrão, mas a precificação dos imóveis é particularmente complexa devido ao papel significativo que o Governo do Distrito Federal (GDF) desempenha no sistema imobiliário, possuindo vastas extensões de terrenos e oferecendo lotes e imóveis em licitações públicas. Ao mesmo tempo em que o GDF busca otimizar seus recursos, oferecendo lotes e imóveis por um preço condizente com o valor de mercado, há o aspecto social que visa democratizar o acesso da casa própria aos brasilienses.

Essa promoção do setor imobiliário é realizada pela Agência de Desenvolvimento do Distrito Federal (Terracap), uma empresa pública incumbida de gerenciar terras públicas de maneira sustentável, com responsabilidade social, ambiental e econômica. A Terracap impulsiona negócios e oferece empreendimentos imobiliários no DF.

Observa-se na Tabela 1 os valores dos investimentos em infraestrutura realizados pela Terracap entre 2011 e 2021, cabe destacar que foi feito o ajuste deflacionário dos valores pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA) do período. No recorte dos últimos cinco anos (2017-2021) houve um aumento gradual nos valores, sugerindo uma tendência. Considerando todo o período (2011 a 2021) os valores oscilaram, tendo ficado na média aproximada de 1,12 bilhão de reais por ano, refletindo as variações e tendências observadas durante todo o período.

Tabela 1: Evolução dos investimentos em infraestrutura realizados pela Terracap (em milhões de reais).

Ano	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Investimento	1.452	1.785	1.810	2.340	728	873	698	796	752	912	961

Fonte: TERRACAP, 2022, deflacionados pelo IPCA do período.

Com vistas a promover uma ferramenta que auxilie a Terracap a precificar os terrenos de forma sistêmica, automatizada e com aprendizado de máquina, o presente estudo busca oferecer um modelo de dados para precificação hedônica dos valores de mercado dos terrenos comercializados no Distrito Federal.

Nos últimos 10 anos, 133 estudos sobre previsão de preços de imóveis utilizando técnicas de aprendizado de máquina foram publicados, de acordo com buscas no site Scielo e Web of Science. Esses trabalhos exploram diversos mercados imobiliários, incluindo cidades na Bélgica, na Indonésia e nos Estados Unidos. No entanto, nenhum desses estudos focou no mercado imobiliário brasileiro.

O modelo de precificação hedônico se baseia no prazer ou na felicidade que um consumidor apresenta na escolha de um imóvel e que depende do nível de atributos que o bem adquirido possui (Besanko et al., 2006). Dessa forma, a ferramenta proposta irá mapear variáveis de pontos de interesse que possam estar associadas ao modelo e que serão testadas acerca de sua relevância para explicação dos resultados.

Além disso, para se medir a precisão do modelo, será apresentada uma comparação entre a previsão dos preços dos terrenos utilizando a modelagem hedônica, em contraste com os preços avaliados pela área técnica da Terracap e com os preços efetivamente transacionados dos terrenos vendidos por licitação pública no ano de 2021.

Ao tentar determinar o valor de mercado de ativos imobiliários em larga escala, é necessário incorporar ao modelo variáveis que retratam as características únicas dos imóveis que os tornam atraentes ou indesejáveis para o mercado consumidor. Para tanto, o modelo utilizou-se de técnicas estatísticas variadas, dentre elas foi usada a regressão múltipla na qual a variável dependente na regressão é o preço do produto. Já as variáveis independentes são preditoras que medem a presença e a extensão de diferentes atributos do produto.

O estudo buscou investigar o impacto das variáveis independentes relacionadas à localização e à acessibilidade dos imóveis. Ao associar essas variáveis ao modelo de precificação hedônica, busca-se aprimorar a capacidade de predição dos preços. Diversos algoritmos de aprendizado de máquina serão utilizados para

identificar o modelo que melhor descreve a relação entre as características dos terrenos e seu valor de mercado.

Os resultados indicam que os valores obtidos pelo modelo de precificação hedônico se aproximaram mais dos valores efetivamente transacionados pela Terracap. Ou seja, o modelo desenvolvido conseguiu explicar mais de 90% da variação dos valores reais de compra e venda dos terrenos em relação às variáveis disponíveis no banco de dados.

Portanto, há indícios de que o modelo de precificação hedônico pode traduzir com mais precisão o devido valor de um imóvel na percepção do consumidor e estimar um preço correspondente que se aproximará da “quantia mais provável pela qual se negociaria voluntária e conscientemente um bem, em uma data de referência, dentro das condições do mercado vigente” (ABNT, 2019).

2

2

REVISÃO DE LITERATURA

2.1 Precificação de Imóveis

O monitoramento oportuno e preciso da distribuição de preços do solo urbano, bem como de suas mudanças e dinâmicas, é crucial para a compreensão científica das tendências do mercado imobiliário. A partir desse conhecimento gerado, é possível propor medidas ou tomar decisões sobre o equilíbrio entre oferta e demanda de terrenos, sendo uma ferramenta de grande valia para os atores do mercado e para a sociedade como um todo (ZHANG, 2021).

No entanto, atribuir um preço a um bem imóvel foi, por certo tempo, tarefa exclusiva dos especialistas do mercado imobiliário, dada a complexidade e número de características de um terreno, casa ou apartamento. Ciarlini (2020) reconhece a heterogeneidade do bem imóvel, lembrando que a avaliação patrimonial está diretamente relacionada à identificação, análise e quantificação das características intrínsecas de cada um dos bens (DUNESE; JONES, 1998).

Malpezzi & Mayo (1985) afirmam que havia pouca informação sobre o comportamento do mercado imobiliário disponível para os formuladores de políticas nos países em desenvolvimento. Eles argumentam que a literatura nos países desenvolvidos tem crescido porque, para intervir com eficiência nos mercados imobiliários, os governos exigem conhecimento detalhado sobre eles. Esses autores enfatizaram que a literatura associada aos mercados imobiliários avançaria nos países em desenvolvimento se houvesse uma maior preocupação dos governos locais em coordenar de forma mais eficaz os mercados imobiliários.

Passados alguns anos da constatação de Malpezzi e Mayo (1985), Zilli (2020) afirma, citando Moura et al. (2004), que no Brasil praticamente todos os municípios de médio e grande porte possuem

seus sistemas cadastrais de imóveis, em algum grau, incompletos e ineficientes. Já Jahanshiri et al. (2011) ressalta que a avaliação em massa de propriedades é importante para cálculo de imposto sobre a propriedade predial, construção de índices de preços e compreensão da dinâmica do mercado. Além disso, os autores citam e analisam várias maneiras e modelos que podem ser utilizados nessa avaliação em massa, entre eles, a precificação hedônica.

Souza (2022) desenvolveu uma avaliação no município de Lauro Müller/SC que apontou para a utilidade pública da avaliação em massa do valor dos imóveis a fim de promover o recolhimento de diferentes impostos territoriais. Os autores desenvolveram uma Planta de Valores Genéricos atualizada, a partir de valores venais obtidos conjuntamente a variáveis independentes, caracterizadoras do espaço urbano. Isso proporcionou um sistema arrecadatário seguro para a definição dos valores tributários, refletindo os valores venais de imóveis de acordo com a realidade do mercado imobiliário municipal.

Seabra, Silveira Neto e Menezes (2016) afirmam que a imobilidade espacial, ou seja, a localização do imóvel está diretamente relacionada à determinação de novos condicionantes para o entendimento da valoração de qualquer propriedade imobiliária nas economias de mercado. Concluindo com a afirmação de Rosen (1974) que, dada sua localização espacial, a imobilidade faz com que as características intrínsecas do mesmo não possam ser dissociadas da influência do meio no qual estão inseridos.

2.2 Influência Espacial no Preço de Imóveis

Hornburg (2009) menciona que em diversas pesquisas de avaliação em massa de imóveis urbanos têm sido incorporadas às metodologias avaliativas técnicas que permitem identificar a influência que a dependência espacial exerce sobre a formação dos preços, bem como a influência que um imóvel exerce sobre seus vizinhos, consolidada na literatura como adjacência.

Moreira de Aguiar et al. (2014) investigou as determinantes dos preços de apartamentos em Belo Horizonte/MG. Os modelos hierárquicos espaciais utilizados pelo autor indicaram que variáveis locais, como violência urbana, infraestrutura e serviços, explicaram mais de 75% da variabilidade remanescente dos preços.

Bottero (2018) afirma que um grande e crescente corpo de literatura investigou a influência de externalidades indesejadas nos

preços das casas. Recentemente, um modelo proposto por Cordera et al. (2018) estimou a presença de relações espaciais entre os valores imobiliários e a presença de uma área industrial na província de Taranto, Itália. No mesmo sentido, foi realizado um modelo espacial em Nantes, França, para verificar os efeitos da poluição do ar e da exposição ao ruído nos preços das casas (Boennec & Salladarré, 2017).

Já Dziauddin e Idris (2017) utilizaram a regressão geograficamente ponderada (GWD) em seu estudo para estimar os efeitos dos atributos de localização nos valores das propriedades residenciais em Kuala Lumpur, na Malásia. Os resultados apontaram que os valores das propriedades residenciais são determinados, principalmente, pelos atributos físicos (estruturais) da propriedade, e que a proximidade com os atributos locais contribui marginalmente.

De acordo com Campos e Almeida (2018) é preciso superar a limitação dos modelos microfundamentados tradicionais quando se busca estudar os fatores espaciais capazes de afetar o preço dos imóveis residenciais. Diversos autores apontaram a questão das amenidades como elementos essenciais, além das qualidades inerentes ao espaço geográfico, ambas capazes de alterar o vetor dos preços dos imóveis, (Alonso, 1964; Muth, 1967, 1972; Mills, 1972 e Brueckner et al. 1999).

Wei (2021), por fim, afirma que a *big data* pode, inevitavelmente, levar a uma revolução no modo de pesquisa ao citar a afirmação de Pei et al., (2020) que a essência da *big data* é a amostragem de cobertura total, e a amostragem “completa” é muito maior do que a amostragem intencional de pequenos dados tradicionais. Isto é, a tecnologia de processamento de dados é suficiente para alimentar os modelos com informações suficientes para que se obtenha projeções confiáveis dada a quantidade e qualidade dos dados.

2.3 Modelo de Preços Hedônicos

O Modelo de Preços Hedônicos (Hedonic Price Model) baseia-se na ideia de que os imóveis são bens heterogêneos (Rosen, 1974) e que a utilidade das famílias responde às características do imóvel, intrínsecas e extrínsecas (Campos, 2018). Para Bottero (2018) o valor de mercado é influenciado pela valoração do conjunto de atributos: locais, estruturais, temporais, geográficos e ambientais. Cada característica tem um preço implícito incorporado no preço de venda; a primeira é revelada apenas a partir de valores observados –

preferências reveladas – de produtos diferenciados com uma quantidade específica de cada atributo.

A abordagem de preços hedônicos, que preconiza o preço do bem como uma função de suas características, teve início com os estudos de Lancaster (1966) e Rosen (1974), muito embora Houthakker (1952) foi quem deu os primeiros passos ao incorporar a qualidade do produto na avaliação de preferência do consumidor (Oliveira, 2019). De acordo com Ceccato (2017), uma das variáveis fundamentais do modelo de preços hedônicos está relacionada à segurança da área do imóvel, especialmente para a realidade brasileira.

Dessa forma, o modelo estima a disposição dos indivíduos para pagar por uma redução da insegurança a partir de uma amostra com valores de preços de imóveis. A autora afirma que, no Brasil, “o crime (ou falta de segurança) tende a ser um fator mais importante na formação dos preços (de venda ou aluguel das propriedades) do que aquele mostrado em estudos baseados em cidades da América do Norte ou Europa” (CECCATO, 2017, p. 11).

Desde a década de 1990, a maioria dos estudos com dados brasileiros aplicou modelos hedônicos estimados com técnicas de regressão linear (Aguirre & Faria, 1997; Aguirre & Macedo, 1997; Moraes & Cruz, 2003; Paixão, 2009, 2010; Rondon & Andrade, 2005; Sousa Filho & Arraes, 2004; Teixeira & Serra, 2006).

Muitos analisaram apenas a influência das características dos imóveis e da localização nos preços em diferentes municípios e regiões metropolitanas do Brasil. Outros trabalhos discutiram o efeito de determinadas características locais sobre os preços dos imóveis, como projetos de saneamento em São Paulo (Aguirre & Faria, 1997); poluição do ar em Brasília (Batalhone, Nogueira, & Mueller, 2002); e construção do Rodoanel Metropolitano de São Paulo (Maciel & Biderman, 2010). Além disso, um ponto enfatizado na literatura brasileira é a influência da criminalidade e da sensação de segurança nos preços dos imóveis urbanos (Paixão, 2009; Rondon & Andrade, 2005; Teixeira & Serra, 2006).

Contudo, conforme aponta Ceccato (2017) não é tão simples dimensionar a influência de uma amenidade no preço dos imóveis. A autora ressalta que a pesquisa de preços hedônicos faz inferências com base no valores de venda dos imóveis. Logo em seguida, calcula-se a demanda baseada em certas características da habitação, por exemplo, o número de quartos; acrescenta-se à predição outros dados espaciais

de localização do bairro e das amenidades, por exemplo, a acessibilidade aos serviços e transportes.

Veloso (2019) cita Furtado (2011) ao explicar que a divisão dos preços em vetores de características é necessária, pois não existem mercados explícitos para as amenidades urbanas, como por exemplo qualidade do ar, trânsito e áreas verdes. Portanto, a incorporação dos preços hedônicos aos modelos de precificação pode captar os efeitos dessas amenidades, sejam positivas ou negativas, e incorporá-las ao preço dos imóveis que serão afetados por elas. Afirma que, segundo Cheshire e Sheppard (2004), independentemente da direção adotada, quanto mais completa for a especificação da função de preços hedônicos, melhores serão os resultados obtidos.

Nesse ponto, o trabalho de Fávero et al. (2008) concluiu que a precificação hedônica apresenta implicações estratégicas e políticas, no que diz respeito à definição de projetos, ao suprimento de recursos e serviços e até mesmo à definição de cargas tributárias diferenciadas em função de cada localidade. Já que os valores e benefícios de um determinado imóvel dependem dos componentes do “pacote” de atributos, deve-se, nas fases de prospecção, projeto e planejamento, levar em conta as prioridades associadas a cada local, consistentes com as preferências daqueles que poderão vir a ser os pretendentes para a compra. Com isto, pode-se inclusive determinar os níveis mínimos de recursos e tecnologias dispensados para cada tipo de construção, em função de seus atributos intrínsecos e extrínsecos (FÁVERO et. al, 2008, p. 92)

Diversos autores brasileiros se debruçaram sobre a precificação hedônica aplicada ao setor imobiliário: Hermann & Haddad (2005) e Batalhone et al. (2002) buscaram entender como um conjunto de variáveis ambientais impactam os preços em São Paulo e Brasília, respectivamente. No caso de Hermann & Haddad (2005, p. 267), os achados iniciais refutaram a hipótese primária e não constataram que a poluição pudesse influir na formação de preços. Em contrapartida, os autores averiguaram que a presença de estações de trem, áreas verdes e zonas estritamente residenciais contribuem para a elevação do aluguel, enquanto a criminalidade reduz o seu valor.

A pesquisa de Guilherme et al. (2014) destaca a importância dos atributos extrínsecos na valoração dos bens imóveis. O estudo conclui que as amenidades ambientais são altamente valorizadas pelos proprietários de imóveis, exercendo um impacto substancial nos

preços desses bens. De acordo com a pesquisa realizada na Inglaterra, uma mudança do percentil inferior de 1% para o melhor percentil de 1% pode representar um valor adicional de cerca de £100.000, ou £3.700 por ano considerando as amenidades ambientais. Isso reforça a relevância dos atributos ambientais na determinação do valor dos imóveis.

Já Batalhone et al. (2002) identificaram uma forte correlação da poluição do ar nos imóveis em um setor de Brasília, o que demonstra a especificidade da localidade e do público consumidor. Embora a principal variável analisada pelos autores tenha sido o mau cheiro proveniente de uma estação de tratamento de esgoto, ligeiramente diferente da delineação da variável 'poluição do ar' em Hermann & Haddad (2005), a conclusão de Batalhone et al. (2002) confirmou a hipótese inicial da regressão hedônica, de que a qualidade do ar tem influência no valor do apartamento. Os autores puderam verificar que os apartamentos localizados nas proximidades da estação de tratamento de esgoto possuíam preços menores, em relação aos similares localizados em áreas mais distantes.

No Brasil, o modelo de preços hedônicos também tem sido aplicado para se mensurar diversos aspectos do mercado imobiliário. Além dos supracitados que se debruçaram sobre um conjunto de variáveis ambientais, como Hermann & Haddad (2005) para São Paulo, Albuquerque et al. (2007), para Recife, e Batalhone et al. (2002), para Brasília. Outros estudos focaram em aspectos mais específicos: Teixeira & Serra (2006) testam o impacto da criminalidade urbana para a cidade de Curitiba; Amrein (2010) teve como foco a qualidade das escolas no município de São Paulo; Aguirre & Faria (1997) e Faria (2008) utilizaram o modelo para avaliar políticas públicas.

Fávero et al. (2008) avaliaram uma amostra de 1860 apartamentos residenciais, lançados em 2004, na Região Metropolitana de São Paulo. A principal hipótese dos autores foi direcionada para o perfil sócio-demográfico. A avaliação foi feita em distritos de renda baixa, média e alta da Região Metropolitana de São Paulo. Por meio da utilização da especificação logarítmica em equações de dois estágios de Rosen, o modelo verificou quais atributos mais interferem nas condições de oferta e demanda dos imóveis e comparou a importância relativa de cada um quando se varia o perfil sócio-demográfico.

Mais recentemente, outros autores brasileiros enriqueceram a literatura nacional, como Seabra et. al (2015), Moreira (2020) e Oliveira

(2019). No caso de Seabra et al. (2015) os pesquisadores investigaram a influência das amenidades sobre o valor dos imóveis residenciais em Recife. As características próprias de uma cidade litorânea foi fator chave para o modelo de preços hedônicos com correção para diferentes tipos de dependência espacial utilizada pelos autores. Os resultados achados indicaram que “vista ao mar” e “proximidade ao rio” são as amenidades com efeito positivo no valor dos imóveis, enquanto a proximidade às Zonas Especiais de Interesse Social, aos metrô e aos locais violentos, afetou negativamente.

Moreira (2020) construiu um modelo hedônico espacial de avaliação em massa de imóveis para a cidade de Fortaleza, Ceará, com suporte em técnicas capazes de explorar potenciais relações não-lineares da autocorrelação espacial. Na listagem, observa-se que estudos sobre os efeitos da localização, seja através dos modelos espaciais ou por meio de variáveis geográficas, não empregam modelos não-lineares, como as redes neurais artificiais ou *Random Forest*.

Oliveira (2019) mensura uma externalidade associada à localização que é o Impacto da Criminalidade nos Preços dos Imóveis Residenciais, qual seja, a variação do preço dos imóveis de acordo com a taxa de homicídio de determinada localidade. Suas estimações se apoiam em modelos econométricos do tipo preços hedônicos. Em uma amostra obtida entre 2009 e 2018, com recursos de *machine learning* e totalizando mais de cem mil imóveis da cidade de Fortaleza, foi constatado o impacto adverso da criminalidade sobre o preço dos imóveis, bem como a compensação quando se considera o fato de o imóvel estar ou não em condomínio.

A partir da conclusão de Oliveira (2019), a elasticidade em relação à criminalidade foi sempre negativa e se elevou substancialmente nos imóveis de alta renda, ao passo que a localização em condomínio gera impacto positivo e majorado nos imóveis de alta renda. Em números, a perda média de valor com a criminalidade pode ultrapassar 40%, já nos imóveis de alto valor a perda supera 67%, e os imóveis em condomínio valorizam em média 39%, chegando a 47% nos imóveis de maiores valores.

Paixão (2015) buscou montar um índice de preços para a cidade de Belo Horizonte/MG. Utilizando o Modelo Hedônico, o autor achou melhor adequação ao mercado de apartamentos. Além disso, as variáveis mais relevantes para a determinação do preço foram: área do

imóvel, padrão de acabamento, distância ao centro e hierarquia socioespacial.

Dias e Raiher (2018) adotaram um modelo de preços hedônicos, estimado por meio de uma Regressão Ponderada Geograficamente (RPG) para realizar uma análise espacial do preço do solo urbano em Curitiba, pelo método de diferenças em diferenças (diff-diff). O resultado apontou que os equipamentos culturais, a distância do centro da cidade e as estruturas ambientais causaram, em sua maioria, um impacto negativo para o preço dos lotes, e que o valor inicial dos lotes, a expansão comercial e os equipamentos de segurança foram fatores positivos para o crescimento dos preços durante o período.

Em âmbito internacional, Gamble e Downing (1982) utilizaram como variável dependente o valor de venda da moradia em algumas cidades americanas e como exemplos de variáveis independentes a distância em relação à usina, o número de aquecedores, a área do terreno, a quantidade de garagens cobertas, etc. Outro exemplo é o trabalho de Zeiss e Atwater (1989) que regrediu o valor da venda da moradia de algumas cidades americanas por variáveis como a qualidade da água, a existência ou não de lareira, o número de quartos, a distância do tratamento de lixo, entre outras.

Com base na premissa de que os preços dos imóveis refletem o valor implícito de suas características subjacentes, sejam elas estruturais, locacionais, ambientais ou relacionadas ao bairro, Gibbons (2014), estudou as valorações marginais ou "preços implícitos" dos atributos individuais da habitação, analisando regressões dos preços de venda de habitações com base nos atributos componentes da casa vendida, sendo demonstrado eficaz não apenas na avaliação da qualidade ambiental, mas também em outras amenidades e características do bairro que podem influenciar a decisão de compra de um imóvel.

Zhang (2021) explora em seu artigo uma simulação da distribuição de preços de terrenos residenciais urbanos com base em dados de várias fontes e aprendizado integrado. Tomando a cidade de Wuhan como exemplo, o autor seleciona variáveis geográficas que caracterizam a vizinhança e localização de terrenos, demonstrando a distribuição heterogênea de fatores individuais, como fatores de vizinhança, fatores de localização e proporções de parcelas. Esses fatores têm um impacto significativo no padrão espacial dos preços dos terrenos residenciais urbanos, quanto mais próspero o serviço

comercial e mais conveniente o transporte público, geralmente significa melhor localização do terreno e maior o preço do terreno residencial comparado às variáveis relacionadas ao meio-ambiente, educação, assistência médica.

Wei (2021) estudou o uso da tecnologia da *big data* em seu artigo. O autor resumiu 124 estudos que investigaram o uso de *big data* para otimizar a avaliação de imóveis por meio do modelo de preço hedônico (HPM), listou os recursos de *big data* e métodos-chave amplamente utilizados no campo de avaliação imobiliária. Observando que os recursos mais usados foram: a internet, o sensoriamento remoto e a Internet das coisas (IoT). Vale destacar a tecnologia de raspagem de informação (*web crawler*) como o método de aquisição de dados mais importante, seguido pelo pré-processamento de dados, modelagem espacial e análise espacial do sistema de informações geográficas (GIS). Nesse estudo, os métodos de aprendizado de máquina foram introduzidos com sucesso no HPM devido aos recursos de *big data* imobiliário, embora seja uma tecnologia em evolução, ela proporcionou uma maior precisão de avaliação.

Segundo Fellipa (2022) os algoritmos de aprendizagem de máquinas são divididos em 3 categorias amplas: aprendizagem supervisionada, aprendizado sem supervisão e aprendizado de reforço. A aprendizagem supervisionada é utilizada nos casos em que uma propriedade (rótulo) está disponível para um determinado conjunto de dados (conjunto de treinamento), mas está faltando e precisa ser previsto para outras instâncias. A aprendizagem não supervisionada é útil nos casos em que o desafio é descobrir relações implícitas em um determinado conjunto de dados não marcados (os itens não são pré-atribuídos).

2.4 Estimativas e Modelos

Para Nascimento (2021) na avaliação de imóveis urbanos têm utilizado mais comumente o método comparativo direto de dados de mercado com tratamento de dados por regressão linear, que consiste em identificar a tendência da formação preço do imóvel por meio do tratamento científico das características dos elementos comparáveis constituintes da amostragem de dados obtidos no mercado.

A análise por modelo de regressão utiliza equação matemática que representa a relação linear, ou seja, uma linha reta entre duas variáveis, comumente chamadas de variável independente (x) e

variável dependente (y). Mesmo utilizando transformações matemáticas, na maioria dos casos, a relação entre as variáveis independente e dependente não é exata, portanto, é necessário levar em consideração a ausência de outras variáveis no modelo e acrescentar o erro aleatório designada pela letra “e” (RAMOS, 2021).

Para Zhang (2021) o grande desafio da análise é a complexa relação não linear entre os preços dos terrenos e vários fatores potenciais de influência, como natureza, economia, política e população, fazendo com que a simulação fina da distribuição espacial dos preços seja mais complexa.

Zilli (2020) utilizou a Regressão Geograficamente Ponderada (GWR) para avaliação em massa de apartamento nos bairros Centro, Agrônômica e Trindade, no município de Florianópolis/SC, a partir de uma amostra de trabalho contendo 190 dados e uma amostra de validação contendo 35 dados de apartamentos. Os resultados do autor demonstram a eficácia da modelagem GWR em reduzir os efeitos espaciais e superioridade para os indicadores de qualidade usuais do ajuste do modelo. Assim, Zilli (2020) conclui que o uso do método de regressão geograficamente ponderada (GWR) é uma abordagem melhor do que outros métodos para examinar o efeito dos atributos locais nos valores de propriedades residenciais.

Won Kim et al. (2003) voltaram sua atenção para a aplicação conjunta de modelos econométricos espaciais e valoração ambiental. Os autores desenvolveram e estimaram um modelo de habitação econômico-espacial para a área metropolitana de Seul, Coreia do Sul, com o objetivo de medir o valor marginal das melhorias na qualidade do ar, especificamente nas concentrações de dióxido de enxofre (SO₂) e nitrogênio (NO_x).

Jahanshiri et al. (2021) realizaram uma análise individualizada de cada modelo em seu estudo de revisão. Eles concluíram que a velocidade, consistência e precisão da avaliação em massa são desafios significativos enfrentados pelas comunidades de avaliação. Os problemas mais comuns associados à Análise de Regressão Múltipla (MRA) são a negligência dos efeitos espaciais no modelo. A dependência espacial, que se refere à influência de propriedades próximas umas das outras, é um fator importante que deve ser considerado de alguma forma no modelo.

Esses efeitos espaciais não são consistentes em todas as regiões, resultando em variabilidade espacial, heterogeneidade espacial ou

segmentação de mercado. A finalidade dos modelos espaciais é otimizar a influência de propriedades vizinhas mais próximas e, ao mesmo tempo, reduzir o impacto de propriedades mais distantes na previsão de imóveis não comercializados.

Os modelos de regressão espaciais (SRA) têm como objetivo aprimorar o MRA ao incorporar componentes de dependência espacial, empregando pesos de conectividade tanto na variável de resposta quanto nos erros de regressão. Por outro lado, a krigagem geoestatística busca estabelecer uma abordagem de previsão que capitaliza a distribuição geográfica dos preços e sua relação com as características dos imóveis adjacentes.

Os modelos locais buscam segmentar regiões com base em janelas específicas para prever propriedades não comercializadas. O método de Regressão de Janela Móvel (MWR) confere peso igual aos vizinhos que impactam uma propriedade nas janelas.

Já a Regressão Geograficamente Ponderada (GWR) utiliza pesos que variam conforme a localização, sendo útil para detectar heterogeneidade espacial (JAHANSHIRI, 2023, p. 29). A técnica GWR, desenvolvida por Brunsdon, Fotheringham e Charlton (1996), modela processos que variam regionalmente, ajustando um modelo de regressão para cada ponto e ponderando outras observações pela distância (Zilli, 2020).

Campos e Almeida (2018) se valeram do Método Hierárquico Linear Espacial para analisar os efeitos implícitos da vizinhança e das adjacências no município de São Paulo. Os resultados, após a primeira decomposição, demonstraram que 90,13% dos preços dos imóveis em São Paulo podem ser entendidos como efeito implícito (intra imóvel) e 6,76% como efeito adjacência (efeito entre imóveis).

Já Macedo (1997), utilizou instrumentos de econometria espacial como testes de diagnósticos para a verificação da presença de dependência espacial e heterogeneidade que provêm os meios de identificação de efeitos de adjacência (transbordamento) na determinação dos preços de imóveis em Belo Horizonte/MG.

Julio (2021) baseou-se em Morenoff (2003) e Campos (2014) ao trabalhar um modelo de dois níveis para análise do mercado imobiliário de São Paulo entre 1985-2012, destacando a divisão espacial dos submercados no primeiro nível e inclusão no segundo nível de variáveis

explicativas defasadas espacialmente para capturar os impactos de transbordamento das características das vizinhanças.

O estudo testou um número de formas funcionais alternativas para MPH - na especificação padrão, bem como naquela com efeitos espaciais - usando a transformação Box-Cox das variáveis analisadas. O autor obteve resultados que evidenciam o transbordamento (adjacência) como uma importante fonte de variação nos preços das unidades residenciais analisadas, e que a especificação (dupla) logarítmica se adequa melhor, em termos de ajustamento estatístico, à descrição da variabilidade dos preços das unidades relativamente a seus respectivos atributos (MACEDO, 1997).

3



3

**CARACTERÍSTICAS IMOBILIÁRIAS DO
DISTRITO FEDERAL**

De acordo com o Censo de 2022, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a população do DF é de mais de 2,8 milhões de habitantes e possui a maior renda per capita do país, cerca de R\$ 2.300,00 reais mensais. Isso explica, em parte, os elevados preços de imóveis e aluguéis.

Os brasilienses estão distribuídos em cerca de 30 núcleos urbanos, as Regiões Administrativas (RAs). Cada RA possui uma dinâmica social e econômica própria, mas muitas delas são dependentes do Plano Piloto, região central de Brasília composta pela Asa Sul, Asa Norte, Setor Militar Urbano (SMU), Noroeste, Granja do Torto, Vila Planalto e Vila Telebrasília.

Figura 1: Distribuição espacial da ocupação urbana de cidade de Brasília



Fonte: SEDUH, 2022.

3.1 Mercado Imobiliário no DF

A observação do mercado imobiliário é estratégica para todos os agentes do mercado imobiliário, sendo utilizado para decidir pela oportunidade de determinado tipo de investimento ou que tipo de produto deve ser produzido para atender uma demanda estabelecida.

Sob outra abordagem, essa informação se reveste de interesse coletivo, pois o bem analisado é comumente dado em garantia para empréstimos de longo prazo. Isso forma extensas carteiras (estoques) de dívidas, com juros subsidiados por políticas governamentais, que utilizam recursos do Fundo de Amparo do Trabalhador (FAT). Em situações específicas os desdobramentos dessas dívidas podem influenciar, sobremaneira, no contexto macroeconômico do país.

O avanço das técnicas e ferramentas de geotecnologia e geoestatística têm proporcionado um maior domínio das informações acerca do espaço geográfico, seja pela popularização do uso de mapas temáticos, de imagens de satélites, e até mesmo de informações dos pontos de interesse agregadas de forma colaborativa e incremental. A observação do mercado imobiliário utiliza diversos indicadores que tentam capturar a preferência individual por determinadas características dos imóveis, sendo que, aquelas ligadas a aspectos locais e de acesso só podem ser identificadas se o modelo tiver especificamente essa capacidade.

A informação acerca da preferência do mercado é indispensável para dar suporte a tomada de decisão das empresas que atuam com empreendimentos imobiliários, seja produzindo lotes urbanizados para suprir às atividades humanas, seja adquirindo lotes para edificar e comercializar. O preço do solo urbanizado passa de receita, no primeiro momento, para, posteriormente, tornar-se custo no âmbito da modelagem econômico-financeira.

O uso de métodos estatísticos para determinação do valor de mercado de um imóvel é consagrado pela literatura e pelas normas técnicas que regulam a temática, à exemplo da Norma Brasileira (NBR) de número 14.653 preconiza a avaliação geral de bens. Na parte dois dessa norma, (NBR 14.653-2) é disposta a Avaliação de Imóveis Urbanos com diversos métodos estatísticos de precificação e análises (ABNT, 2011).

A peculiaridade do mercado imobiliário desenvolvido na Capital Federal é a atuação da Companhia Imobiliária de Brasília (Terracap),

que é uma empresa pública integrante da administração indireta do Distrito Federal.

3.2 A Companhia Imobiliária de Brasília - Terracap

A Terracap teve sua origem no Departamento Imobiliário da Companhia Urbanizadora da Nova Capital - NOVACAP -, empresa criada pelo Presidente Juscelino Kubitschek para administrar a construção de Brasília. Naquela época era o setor da NOVACAP responsável pela utilização, aquisição, administração, aluguéis, concessão de direito real de uso, concessão de uso oneroso, parcelamento do solo, cessão de uso, disposição, incorporação, oneração ou alienação de bens, mediante contraprestação de serviços. Além disso, a NOVACAP realizava diretamente obras e serviços de infraestrutura, de urbanização, edificações e viárias outras atividades correlatas das terras públicas do DF.

A Terracap foi instituída por meio da Lei 5.861, de 12 de dezembro de 1972, mas só entrou em operação em 14 de agosto de 1973. Na composição da empresa, mais da metade (51%) do capital social pertencem ao DF e o restante (49%) pertence à União. Na constituição societária da empresa foi aportado o capital na forma de patrimônio equivalente a 338.337 hectares de terras, além da responsabilidade de administrar esses terrenos em prol da sociedade. Nesse escopo, compete a Terracap responder pela regularização fundiária, inclusive com poderes para desapropriação, comercialização, arrendamento ou doação das terras do Governo do Distrito Federal.

Apenas no ano de 2020, a Terracap apresentou o seguinte panorama: foram alienados, através de processo licitatório, um total de 583 terrenos situados em diferentes setores e com destinações distintas. O montante total obtido com estas vendas foi de aproximadamente R\$ 676 milhões.

O presente trabalho se beneficiou da informação de 320.288 imóveis constituídos pela Terracap, distribuídos nas Administrações Regionais conforme apresentado na Tabela 1. Os dados se referem às cinco décadas de atuação da empresa com grande quantidade de dados geográficos disponibilizados pelos diversos órgãos governamentais do DF, além daqueles disponibilizados pela iniciativa privada por meio de ferramentas colaborativas.

Embora nem todos os imóveis da companhia tenham sido comercializados na janela temporal delimitada para o estudo, ainda assim, foram vendidos 19.814 imóveis em licitação pública para 17.441 compradores distintos desde o ano de 1989. Isso, por si só, constitui um grande patrimônio da sociedade de Brasília, tanto imobiliário, quanto informacional.

Tabela 2: Número de imóveis cadastrados na Terracap por Administração Regional.

Unidades de Tomadas de Decisão	Número de móveis
RA-I - Brasília	12966
RA-II - Gama	14046
RA-III - Taguatinga	25373
RA-IV - Brazlândia	17151
RA-IX - Ceilândia	10966
RA-V - Sobradinho	9305
RA-VI - Planaltina	74452
RA-VII - Paranoá	3070
RA-VIII - Núcleo Bandeirante	19095
RA-X - Guará	7012
RA-XII - Samambaia	11644
RA-XIII - Santa Maria	11226
RA-XIV - São Sebastião	26984
RA-XIX - Candangolândia	48473
RA-XV - Recanto das Emas	2661
RA-XVII - Riacho Fundo	10
RA-XVIII - Lago Norte	6773
RA-XXIX - Setor De Indústria E Abastecimento	3116
RA-XXV - Setor Complem. De Indústria E Abastecimento	8040
RA-XXX - Vicente Pires	133
RA-XXXIII - Arniqueira	2423

Fonte: elaborado pelo autor.

A partir desses dados, foi possível tratar as variáveis de localização e acessibilidade no intuito de aprimorar a capacidade preditiva dos modelos de precificação hedônica. No entanto, é essencial empregar técnicas de econometria espacial para abordar a autocorrelação e a heterogeneidade espaciais. Métodos como defasagem espacial, erro espacial e regressão geograficamente ponderada são particularmente relevantes nesse contexto.

Esse trabalho foi desenhado com o objetivo precípua de criar um modelo de dados para precificação hedônica dos valores de mercado dos lotes comercializados pela Terracap em Brasília aplicando algoritmos de aprendizado de máquina e regressão linear.

Foi desenvolvido um modelo de dados geográfico multi-fonte, que se baseou nos preços dos terrenos comercializados pela Terracap. Foram coletados dados de variáveis de pontos de interesse relevantes,

que foram testados quanto à sua capacidade de explicação do modelo. Em seguida, a precisão das previsões de preços dos terrenos foi comparada entre diferentes algoritmos de aprendizado de máquina. Esses resultados foram contrastados com os preços avaliados pela área técnica da Terracap, bem como com os preços reais dos terrenos vendidos por licitação pública no ano de 2022.

Esse estudo está restrito à cidade de Brasília, especialmente às transações imobiliárias realizadas pela Terracap por meio de licitação pública. O processo de venda é feito na modalidade de propostas fechadas, na qual é estabelecido um valor mínimo pela Terracap para cada imóvel. É necessário que cada licitante deposite uma caução no valor equivalente a 5% do preço mínimo para poder concorrer. O licitante que apresentar a maior proposta será o vencedor. Esse leilão é realizado sem que os concorrentes tenham ciência dos valores ofertados por outros concorrentes, ou seja, as propostas fechadas são inseridas em uma urna e abertas em sessão pública marcada, só então as propostas são lidas e é declarada a oferta ganhadora.

O procedimento para que a Terracap estabeleça esse valor mínimo requer uma atividade técnica preliminar de avaliação do terreno, que é realizada de acordo com a norma NBR 14653¹. Dividida em várias partes, essa norma aborda aspectos específicos da avaliação de imóveis, incluindo imóveis urbanos, rurais, empreendimentos, máquinas, equipamentos, instalações, bens industriais e recursos naturais e ambientais. Após essa avaliação específica, o menor preço para cada imóvel é estabelecido e divulgado em um edital.

A partir da divulgação desses editais e das compras efetivamente realizadas, delineou-se a amostra deste trabalho, caracterizando-a como não probabilística, por ser composta por todas as observações do banco de dados corporativo da empresa objeto de estudo. Nesse sentido, o grau de detalhamento das observações levou em consideração a disponibilidade de informação em relação às variáveis observadas e o processamento dos agrupamentos dos dados se fez necessário para construir uma relação de hierarquia espacial.

¹ A norma ABNT NBR 14653 é a referência técnica brasileira para a avaliação de imóveis. É dividida em partes que aborda aspectos específicos da avaliação, sendo: imóveis urbanos (NBR 14653-2), rurais (NBR 14653-3), empreendimentos (NBR 14653-4), máquinas, equipamentos, instalações e bens industriais (NBR 14653-5), e recursos naturais e ambientais (NBR 14653-6).



4

4

DADOS E PROCESSAMENTO

Os dados referentes aos lotes, frequentemente denominados como terrenos ou projeções, especialmente quando circundados por área pública e destinados à construção de edifícios, foram obtidos do banco de dados corporativo da Terracap.

O recorte temporal adotado para a coleta da amostra abrange um período de duas décadas, iniciando-se em janeiro de 2000 e estendendo-se até dezembro de 2022, a escolha foi motivada pela relevância de analisar as tendências e padrões de precificação de terrenos ao longo de um período significativo que reflete as dinâmicas e evoluções do mercado imobiliário. O intervalo das observações apresenta dados contínuos, garantindo uma representação abrangente e consistente das transações.

As informações geográficas de localização e acessibilidades foram adquiridas de fontes públicas abertas, organizados pela Secretaria de Estado de Desenvolvimento Urbano e Habitação - SEDUH - na plataforma GEOPORTAL², organizadas em pontos de interesse de negócios, educação, transporte, parques, cursos d'água, saúde pública e privada.

O processamento das informações geográficas foi realizado no *python* utilizando a biblioteca *geopandas*³ como ferramenta de Geoprocessamento capaz de realizar o processamento com um desempenho satisfatório, levando-se em consideração o volume de informações adquiridas que se associa ao modelo de dados para uso nos algoritmos de aprendizado de máquina.

Várias formas de processamento foram aplicadas aos dados originais, visando à exclusão de observações que não continham todas as variáveis adequadamente preenchidas. Este procedimento não apenas limitou o número de observações conforme descrito anteriormente, mas também assegurou que todos os modelos de aprendizado de máquina pudessem produzir respostas válidas.

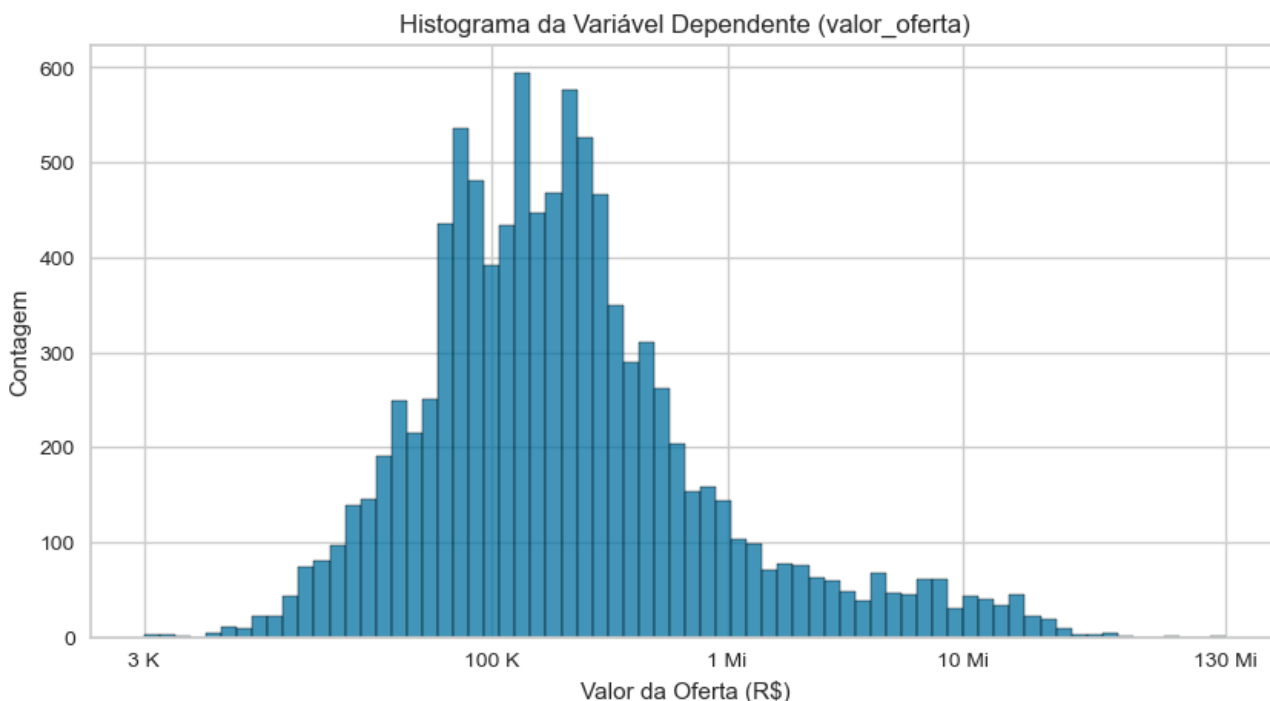
² <https://www.geoportal.seduh.df.gov.br/geoportal/>

³ <https://geopandas.org/en/stable/>

4.1 Variáveis incluídas no modelo

A variável dependente do modelo é a informação apresentada no campo “valor_oferta” que representa o valor⁴ pelo qual a venda do terreno foi realizada, para construção do modelo, os lotes vendidos em agrupados foram identificados com a variável “vendido_em_grupo” atribuído o valor 1, tendo o valor de venda encontrado a partir da divisão do valor da venda do grupo ponderado pela área. Considerando as 9.959 observações, a variável dependente tem uma média de R\$ 916.170,69 e um desvio padrão de R\$ 3.237.944,42. O valor mínimo registrado é de R\$ 3.375,00, enquanto o máximo é de R\$ 129.670.000,00.

Gráfico 1: Histograma da variável Dependente.



Para a seleção do conjunto de variáveis independentes, que sintetizam os atributos normativos, físicos e espaciais dos imóveis, foi realizada uma coleta específica dos dados na base de dados corporativa da Terracap, focando nas informações relativas aos imóveis comercializados.

É importante salientar que parte dessas características foi obtida por meio de técnicas de geoprocessamento. As informações correlatas estão detalhadamente catalogadas nas Tabelas 3 e 4.

⁴ Valor monetário nominal deflacionado até o ano de 2023 pelo índice de preços INPC.

Tabela 3: Variáveis locacionais e de acessibilidade.

Amenidades	Indicador
Comerciais	Densidade do núcleo do shopping Densidade do kernel do supermercado Densidade central da instalação viva
Educacionais	Densidade/Distância à Centros universitários Densidade/Distância a Escolas ensino médio Densidade/Distância a Escolas ensino primário/creches
Médica/saúde	Densidade/Distância nuclear hospitalar/clínicas
Transporte	Densidade/Distância ao ponto de ônibus Densidade/Distância a estações de metrô
Centralidades	Densidade/Distância ao centro de atividade hierarquizada (Local, Setorial, Regional) Distância ao corpo hídrico mais próximo (lago/riacho)

Fonte: adaptado de ZHANG, 2021.

Tabela 4: Relação das Variáveis independentes utilizadas no modelo - Físicas.

Variável	Descrição
ra	Região administrativa no qual o imóvel está inserida
setor	Termo equivalente à bairro
max_condicao	Atributo anterior do lote no banco de dados
situacao	Descrição da ocupação do lote ('6 - CERCADO', '3 - OCUPADO', '1 - VAGO', '2 - OBSTRUÍDO', '5 - EM CONSTRUÇÃO', '4 - CONSTRUÍDO', '0 - SITUACAO NAO CADASTRADA')
tipologia	Caracterização genérica dos usos permitidos na legislação urbanística
destinacao	Caracterização específica do uso planejado no projeto urbanístico originário
ref_referencia	Caracterização da forma do imóvel
pos_posicao	Caracterização do lote quanto a posição do lote (ISOLADO, ESQUINA, MEIO DA QUADRA)
area_max	Área máxima de construção
area	Área da superfície do terreno
ocupacao_valor	taxa de ocupação permitida para o lote
medlado	Medida do lado do lote
medfrente	Medida da frente do lote
iptu_st	Flag do banco que identifica uso do solo passível de tributação
area_cons	área máxima de construção
nr_index_aproveitamento	Índice de aproveitamento na legislação do solo
valor_coef_aproveitamento_area	Coefficiente de aproveitamento na legislação do solo

Fonte: elaborado pelo autor.

Tabela 5: Relação das Variáveis independentes utilizadas no modelo - Locacionais.

Variável	Descrição
area_tombada_contagem	Descreve se está em área tombada (Plano Piloto)
diretrizes_urbanisticas_contagem	Proximidade com projeto Especial de urbanismo da SEDUH
distancia_de_400_metros_maisperto	Distância de 400m em relação Estações e Terminais
distancia_de_600_metros_distancia	Distância de 600m em relação Estações e Terminais
distancia_de_600_metros_maisperto	Identificação da Estações/Terminais mais perto
espacos_comunitarios_contagem	Contagem de espaços comunitários no raio de 1,5km
intervencoes_ciclovias_programa_nos_urbanos_contagem	Contagem das ciclovias, ciclovias compartilhadas, ciclofaixa e ciclorrotas em implantação no raio de 1,5km
linhas_de_transportes_de_alta_densidade_contagem	Número de vias servidas de transporte de média e alta capacidade, sendo eles Trem, Metrô, Veículos Leves sobre Trilhos (VLT), Veículos Leves sobre Pneus (VLP)/Bus Rapid Transit (BRT), corredores de ônibus, e vias com alta densidade de viagens de transporte público coletivo por ônibus no raio de 1,5km
mapa_1c_unidades_de_planejamento_territorial_contagem	Contagem das zonas de planejamento dentro do raio de 1,5km
mapa_1c_unidades_de_planejamento_territorial_maisperto	Identificação da Zona de planejamento mais perto (Urbana, Rural e de Proteção Integral)
mobiliario_esporte_e_lazer_contagem	Contagem dos mobiliários de esporte e lazer implantados em área pública no raio de 1,5km
mobiliario_esporte_e_lazer_maisperto	Identificação dos mobiliários de esporte e lazer implantados em área pública mais próximo
sistema_ciclovionario_contagem	Contagem de ciclovias, ciclovias compartilhadas, ciclofaixa e ciclorrotas no raio de 1,5km
pontes_viadutos_e_passarelas_maisperto	Identificação das obras de engenharia mais perto
populacao_abrangida_pela_infraestrutura_ciclovitaria_maisperto	Número da população abrangida pela intervenção ciclovárias (em milhares)
rede_estrutural_de_transporte_coletivo_contagem	Contagem de elementos da rede estrutural de transporte coletivo no raio de 1,5km
regioes_administrativas_maisperto	Identificação da administrações regionais mais próxima
regulaziracao_contagem	Contagem dos projetos de regularização fundiária no raio de 1,5km
rodovia_maisperto	Identificação da rodovia mais próxima

zoneamento_das_diretrizes_urbanisticas_maisperto	Identificação da Diretrizes Urbanísticas mais próxima
--	---

Fonte: elaborado pelo autor.

Tabela 6: Relação das Variáveis independentes utilizadas no modelo - Financeiras, temporais e negociais.

Variável	Descrição
vendido_em_grupo	Flag identificador da forma de venda do imóvel individualizado ou em grupo
vend_data_data	data da venda do imóvel
index_atualizacao	Index relativo ao valor do INPC do mês da venda

Fonte: elaborado pelo autor.

Tabela 7: Estatística descritiva das variáveis do modelo de dados.

Variável	unique	freq	mean	std	min	max
vendido_em_grupo	2	8541	-	-	-	-
valor_oferta	-	-	916,17 0,68	3.237,9 44,42	3.375 ,0	129.670,00 0,0
area	-	-	716,89	1929,16	4	75471,5
area_max	-	-	1779,7 2	4496,6 2	4	181131,6
ra	19	298 7	-	-	-	-
setor	164	1284	-	-	-	-
max_condicao	-	-	43,78	39,25	4	403
situacao	8	647 5	-	-	-	-
tipologia	40	1970	-	-	-	-
destinacao	192	1970	-	-	-	-
ref_referencia	3	874 0	-	-	-	-
pos_posicao	3	6198	-	-	-	-
ocupacao_valor	-	-	86,13	18,57	0	100
medfrente	-	-	15,53	18,61	0	500
vend_data_data	-	-	15808, 64	2186,92	3470	19662
area_cons	-	-	1728,5 8	31894,2 2	4	3163500
valor_coef_aproveitamento_area	-	-	2,65	1,22	0	22,11
index_atualizacao	237	205	-	-	-	-
espacos_comunitarios_contagem	-	-	22,58	10,79	0	74
distancia_de_600_metros_maisperto	-	-	9,47	10,39	0	41
distancia_de_600_metros_distancia	-	-	354,5	455,68	0	1500
mobiliario_esporte_e_lazer_maisperto	-	-	114,56	193,62	0	2442
sistema_cicloviario_contagem	-	-	20,16	18,74	0	103
zoneamento_das_diretrizes_urbanisticas_maisperto	-	-	40,91	51,12	0	192
linhas_de_transportes_de_alta_densidade_contagem	13	2941	-	-	-	-
pontes_viadutos_e_passarelas_maisperto	-	-	121,79	170,17	0	733
populacao_abrangida_pela_infraestrutura_cicloviaria_maisperto	-	-	108,87	211,63	0	4071
mapa_1c_unidades_de_planejamento_territorial_contagem	4	669 4	-	-	-	-
mapa_1c_unidades_de_planejamento_territorial_maisperto	6	365 9	-	-	-	-
area_tombada_contagem	2	895 4	-	-	-	-
regioes_administrativas_maisperto	23	236 0	-	-	-	-

rede_estrutural_de_transporte_coletivo_contagem	20	2160	-	-	-	-
intervencoes_cicloviarias_programa_nos_urbanos_contagem	-	-	8,62	11,92	0	57
regulaziracao_contagem	5	4357	-	-	-	-
diretrizes_urbanisticas_contagem	7	3946	-	-	-	-
rodovia_maisperto	93	704	-	-	-	-

Fonte: elaborado pelo autor.

Conforme apresentado na Tabela 7, a pesquisa engloba um total de 9.959 imóveis vendidos em licitação pública desde 1989. Os preços foram deflacionados para remover os efeitos da inflação ao longo dos anos. O valor total das vendas, após o ajuste deflacionário, ultrapassou R\$ 9,1 bilhões.



5

5

METODOLOGIA

Utilizou-se a linguagem *python*⁵ e R⁶ para tratar os dados obtidos e bibliotecas de geoprocessamento (GIS) para compilação, generalização e extração das variáveis relativas à localização e acessibilidade e de algoritmos de aprendizado de máquina em linguagem para precificação dos terrenos e comparação quantitativa dos resultados.

As características dos imóveis geram um preço implícito incorporado no preço de venda, a observação das características possibilita identificar o quanto individualmente cada uma contribui para a formação do todo. Pode-se formalizar a função hedônica P que é determinada por diferentes atributos, conforme representado na equação (3.1):

$$P = f(CI, LA, A, WP, WCI, WLA, WA) \quad (3.1)$$

Onde, CI - Características intrínsecas (área, destinação, potencial construtivo), LA - Localização e acessibilidade (distância ao centro, ponto de ônibus/metrô), A - amenidades (contexto socioeconômico, segurança, parques), WCI - Características intrínsecas defasadas espacialmente (área, destinação, potencial construtivo), WLA - defasadas espacialmente das Localização e acessibilidade (distância ao centro, ponto de ônibus/metrô), WA - amenidades defasadas espacialmente (contexto socioeconômico, segurança, parques) e WP preço defasado espacialmente.

Cabe pontuar que a influência das amenidades (A) nos preços dos imóveis de sua unidade espacial (setor/bairro), é chamado de efeito vizinhança (EV). Quando este efeito ultrapassa para a unidade espacial adjacente, tem-se o efeito adjacência do distrito (EA-D) das amenidades adjacentes (WA). Ocorre também a influência das características estruturais e de acessibilidade dos imóveis (CI + LA) pelos canais denominados de efeito implícito (EI) e efeito adjacência do imóvel (EA-I) – identificado pelas características estruturais e de acessibilidade dos vizinhos (WCI + WLA). Isto é, os efeitos EI e EV capturam o impacto das características intra imóvel e intra distrito, os

⁵ <https://www.python.org/>

⁶ <https://www.r-project.org>

efeitos EA-I e EA-D incorporam os efeitos de transbordamento espacial entre imóveis e entre distritos, respectivamente (CAMPOS, 2018).

No modelo empregado, incorporou-se a variável linear de tempo denominada "index_atualizacao" com o intuito de controlar tendências latentes que poderiam exercer influência sobre os preços dos terrenos. Uso da variável tem como propósito capturar flutuações sistemáticas nos valores ao longo do período que não são atribuíveis às demais variáveis presentes no modelo. A inserção desta variável temporal permite isolar o impacto de outras variáveis independentes sobre o preço, assegurando uma maior precisão do modelo e conferindo maior robustez às estimativas obtidas.

Uma vez selecionadas as camadas de interesse e as variáveis foi feito intenso processamento para limpar a base de dados. As camadas de informações espaciais passaram por um geoprocessamento para captura da influência da vizinhança dos imóveis vendidos relacionadas à sua proximidade, mensuração da distância euclidiana à feição mais próxima, identificação do ponto de interesse mais próximo e contagem do número de feições no raio de 1.500 metros.

As características intrínsecas dos terrenos são ligadas à própria terra, abrangendo elementos como o tamanho do lote, a forma do lote, a topografia (por exemplo, se é plano ou inclinado), a qualidade do solo, a presença de recursos naturais e a disponibilidade de serviços públicos. As características locais estão associadas à localização geográfica do terreno, estas incluem a proximidade a centros urbanos, a acessibilidade a infraestruturas como estradas e transporte público, a proximidade a serviços como escolas e hospitais, a qualidade do bairro e a vista do terreno.

Para identificar a data em que a venda foi realizada, empregou-se uma variável de índice de tempo que representou a progressão temporal, permitindo a análise da evolução das vendas no período estudado. Os valores foram deflacionados pelo Índice Nacional de Preços ao Consumidor (INPC).

Outro aspecto relevante foi a identificação nos dados da prática da venda de terrenos agrupados, ou seja, a colocação à venda dos terrenos em conjunto, resultante da estratégia de venda que envolve a combinação de diversos terrenos lindeiros em uma única transação. Isso implicou na necessidade de se ponderar o valor de venda desses terrenos pela área.

5.1 Algoritmos de Aprendizado de Máquina

Foram aplicados aos modelos de dados 40 algoritmos de aprendizado de máquina disponíveis nas bibliotecas de scikit-learn⁷, LightGBM⁸ e xgboost⁹. Tais bibliotecas são ferramentas poderosas para aprendizado de máquina em Python, cada uma oferecendo uma variedade de modelos.

A biblioteca Scikit-learn fornece uma ampla gama de algoritmos, incluindo Regressão Linear, Regressão Logística, Árvores de Decisão, Random Forest, Gradient Boosting, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, Redes Neurais (MLP), além de técnicas de Clustering como K-Means e DBSCAN, e métodos de redução de dimensionalidade como PCA e T-SNE.

O LightGBM, abreviação de "Light Gradient Boosting Machine", é uma implementação avançada do algoritmo Gradient Boosting Decision Tree (GBDT). Distingue-se por sua alta eficiência e velocidade, sendo particularmente adequado para grandes conjuntos de dados. Uma de suas características notáveis é a capacidade de lidar com valores ausentes, otimizando automaticamente sua gestão durante o treinamento. Além disso, o LightGBM introduz técnicas inovadoras, como Gradient-based One-Side Sampling (GOSS) e Exclusive Feature Bundling (EFB), que melhoram a eficiência do treinamento. Inspirado em outras implementações populares de GBDT, o LightGBM tem sido amplamente adotado em diversas aplicações de aprendizado de máquina devido à sua robustez e desempenho superior.

A Árvore de Regressão Baseada em Histograma de Gradient Boosting (Histogram-based Gradient Boosting Regression Tree) é uma implementação do scikit-learn, uma biblioteca popular de aprendizado de máquina em Python, inspirada no LightGBM, e que se destaca por sua eficiência em grandes conjuntos de dados ($n_{\text{samples}} \geq 10.000$). A escolha do modelo dependeu da resposta estatística dos modelos face ao problema apresentado, e cada modelo com seus pontos fortes e fracos foi avaliado por sua resposta.

Os resultados das regressões foram testados quanto à eficácia da modelagem dos algoritmos de aprendizado de máquina treinados, comumente avaliados quanto ao coeficiente de determinação (3.3.1)- R^2

⁷ <https://scikit-learn.org/>

⁸ <https://github.com/microsoft/LightGBM>

⁹ <https://xgboost.readthedocs.io/en/stable/index.html>

- e erro absoluto médio (3.3.2) - MAE, e o erro quadrático médio (3.3.3) - RMSE.

$$R^2 (y_i, \tilde{y}_i) = 1 - (\Sigma (y_i - \tilde{y}_i) / \Sigma (y_i - \hat{y}_i)) \quad (3.3.1)$$

$$MAE (y_i, \tilde{y}_i) = \frac{1}{n} \times \Sigma |Y_i - \hat{Y}_i| \quad (3.3.2)$$

$$RMSE (y_i, \tilde{y}_i) = \sqrt{\frac{\Sigma (y_i - \tilde{y}_i)^2}{n}} \quad (3.3.3)$$



6

6

RESULTADOS E DISCUSSÃO

Foram consideradas três abordagens para a variável dependente: o uso direto, a transformação pelo logaritmo natural do preço de venda e a transformação anterior ponderada pela área. Após a realização dos testes estatísticos dos modelos a formar da variável dependente que apresentou melhor resultado foi o log (valor_de_oferta).

O ponto de partida da análise foi a comparação com o modelo de regressão linear (LinearRegression) baseado na técnica do Mínimos Quadrados Ordinários (OLS), com a variável dependente ajustada a função logarítmica e as variáveis independentes sem transformações. obtivemos a seguinte respostas estatísticas R-quadrado de 0.63 e um RMSE de 0.86, indicando que 63% da variabilidade na variável dependente é explicada pelas variáveis independentes, considerando a penalidade pelo número de preditores no modelo.

A Tabela 8 apresenta os resultados dos algoritmos aplicados aos modelos de dados com o maior R^2 ajustado.

Tabela 8: Ranking dos modelos treinados segundo o R^2 .

Model	R-Squared	RMSE
HistGradientBoostingRegressor	0,92	0,39
LGBMRegressor	0,92	0,40
XGBRegressor	0,92	0,40
ExtraTreesRegressor	0,91	0,42
BaggingRegressor	0,90	0,45
GradientBoostingRegressor	0,90	0,45
SVR	0,87	0,52
NuSVR	0,87	0,52
MLPRegressor	0,86	0,53
DecisionTreeRegressor	0,83	0,58
ExtraTreeRegressor	0,81	0,61
KNeighborsRegressor	0,77	0,67
AdaBoostRegressor	0,74	0,72
LassoLarsIC	0,63	0,86
BayesianRidge	0,63	0,86
RidgeCV	0,63	0,86
Ridge	0,63	0,86
TransformedTargetRegressor	0,63	0,86

LinearRegression	0,63	0,86
ElasticNetCV	0,62	0,87
LassoCV	0,62	0,87
LassoLarsCV	0,62	0,87
LarsCV	0,61	0,88
Lars	0,61	0,89
PoissonRegressor	0,58	0,91
OrthogonalMatchingPursuitCV	0,53	0,97
TweedieRegressor	0,52	0,98
GammaRegressor	0,48	1,02
HuberRegressor	0,46	1,04
LinearSVR	0,41	1,09
OrthogonalMatchingPursuit	0,40	1,10
ElasticNet	0,09	1,35

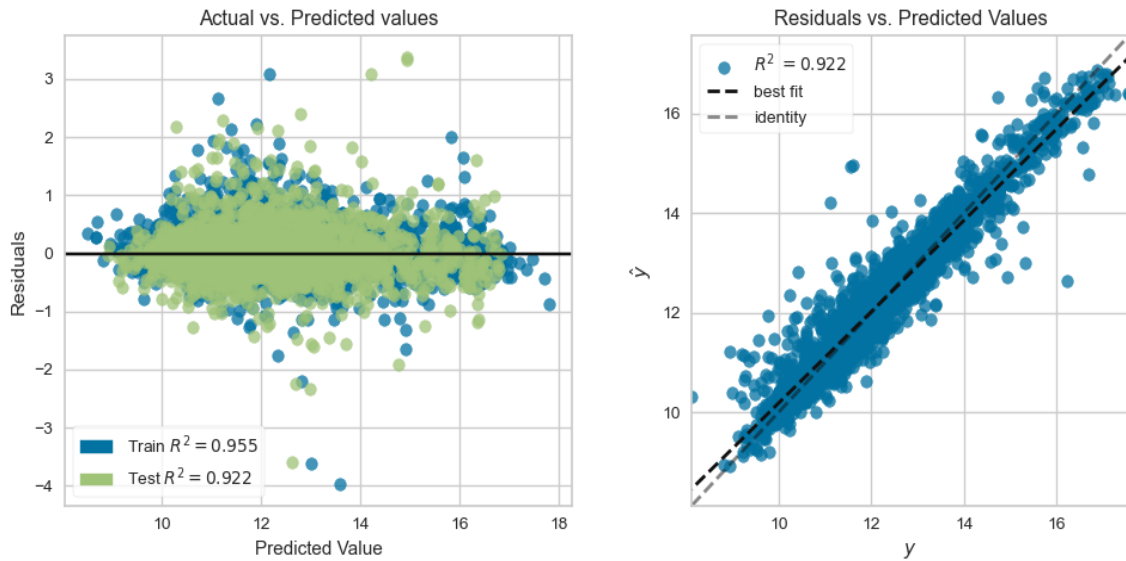
Fonte: elaborado pelo autor.

Após a fase de pré-processamento e treinamento do modelo, procedeu-se à avaliação do desempenho do modelo em uma amostra de teste. Esta amostra foi selecionada aleatoriamente do conjunto de dados e não foi utilizada durante a fase de treinamento, proporcionando uma avaliação imparcial do desempenho do modelo. Foi empregada a função “train_test_split”, componente da biblioteca de aprendizado de máquina e ciência de dados scikit-learn. A função dividiu o banco de dados em dois grupos, o conjunto de treinamento e teste, na proporção 70% / 30%.

A divergência observada nos resultados dos algoritmos de aprendizado de máquina é uma combinação de fatores relacionados aos dados, à natureza intrínseca dos algoritmos, à configuração dos hiperparâmetros e ao processo de treinamento. Ao trabalhar com os algoritmos é essencial entender esses fatores e considerar que a estratégia de particionamento dos dados para a validação cruzada e outros tratamentos dos dados pode influenciar nos resultados e desempenho dos modelos.

O modelo de aprendizado de máquina HistGradientBoostingRegressor ao analisar os dados do mercado imobiliário, considerando a variável dependente transformada pelo logaritmo do valor da oferta ($\log(\text{Valor_Oferta})$), obteve os resultados bem promissores, utilizando os dados das observações do conjunto de teste, apresentou o coeficiente de determinação (R-Squared), que leva em conta o número de preditores no modelo, de 0.922, indicando que o modelo é capaz de explicar 92.3% da variabilidade no logaritmo do valor da oferta.

Gráfico 2: Plotagem dos valores reais x previstos do modelo HistGradientBoostingRegressor.
HistGradientBoostingRegressor $y = \log(\text{Valor_Oferta})$



Fonte: elaborado pelo autor.

Adicionalmente, o erro quadrático médio da raiz (RMSE) do modelo foi de 0.396. O RMSE é uma métrica que quantifica a diferença média entre os valores previstos pelo modelo e os valores reais, sendo que o valor do RMSE, indica que o modelo tem um bom desempenho na previsão do logaritmo do valor da oferta.

Além disso, calculou-se o coeficiente de determinação R^2 das previsões, que foi de 0.9214, cabe observar que, pela própria dinâmica no processo de aprendizado, cada realização gerou um valor diferente na mesma ordem de grandeza. Este valor indica que o modelo foi capaz de explicar aproximadamente 92% da flutuação na variável dependente “valor_oferta”. Isso sugere que o modelo é altamente eficaz em capturar a relação entre as características das propriedades e seu valor de oferta.

A análise dos resíduos, que são as diferenças entre os valores reais e previstos, sugere que o modelo HistGradientBoostingRegressor é adequado para prever o valor de oferta de propriedades. Os resíduos estão distribuídos aleatoriamente ao redor da linha $y=0$, sem padrões aparentes, o que é um bom indicativo.

Embora o modelo tenha apresentado um bom desempenho geral, esses resultados indicam que há espaço para melhorias, especialmente na previsão do valor de oferta para propriedades de valor mais alto, visto que para este grupo existem menos observações. Isso poderia ser explorado em trabalhos futuros.

Ao aplicarmos uma técnica de permutação para medida da importância das características do modelo, permutando a presença dos valores dessa característica e medindo o quanto a métrica de desempenho do modelo (neste caso, o R^2) é impactada. Características que são importantes para a previsão terão uma grande diminuição na métrica de desempenho quando seus valores são permutados, enquanto características que não são importantes terão pouca ou nenhuma mudança na métrica de desempenho.

Na tabela 8, é possível constatar que o “index_atualizacao” (0,447) e a “area” (0,006) e suas formas derivadas (area_max, area_cons) entre outras características estão entre as mais importantes para a previsão do modelo, tais como: vendido_em_grupo, ra, area_tombada_contagem e rede_estrutural_de_transporte_coletivo_contagem, variando em valores escalares de 0.052 à 0.011.

Tabela 9: Ranking das principais características em termos de importância de permutação.

Característica	Importância de Permutação
index_atualizacao	0,447
area_cons	0,244
area_max	0,204
area	0,063
vendido_em_grupo	0,052
ra	0,038
area_tombada_contagem	0,019
rede_estrutural_de_transporte_coletivo_contagem	0,011
valor_coef_aproveitamento_area	0,009
vend_data_data	0,007
pontes_viadutos_e_passarelas_maisperto	0,006
pos_posicao	0,006
rodovia_maisperto	0,004
populacao_abrangida_pela_infraestrutura_cicloviaria_maisperto	0,004
espacos_comunitarios_contagem	0,004
destinacao	0,004
mapa_lc_unidades_de_planejamento_territorial_maisperto	0,003

Fonte: Gerado pelo autor.

Um aspecto de destaque em nossa análise é a constatação da influência da variável “vendido_em_grupo” (0,052) para o desempenho preditivo do modelo, revelando que a prática da venda de terrenos

agrupados é perceptível na modelagem. Isso sugere que a estratégia comercial tem implicações diretas na forma como os terrenos são precificados pelo mercado. A presença desta variável no modelo expõe a correlação entre a venda conjunta de terrenos e seu valor de oferta, reforçando a ideia de que as dinâmicas de mercado e as estratégias de venda estão intrinsecamente ligadas e podem afetar os preços dos terrenos de maneira significativa.



7

7

CONCLUSÃO

A aplicação de algoritmos de aprendizado de máquina disponíveis nas bibliotecas *python*, em modelos de dados estruturados demonstrou ser uma abordagem eficaz e eficiente para a construção de modelos hedônicos de precificação de imóveis no DF. Em particular, o modelo HistGradientBoostingRegressor exibiu um desempenho notável na previsão do logaritmo do valor da oferta no conjunto de dados do mercado imobiliário.

A pesquisa empregou modelos avançados de regressão, tomando como base de referência a regressão linear obteve como modelo mais ajustado a Árvore de Regressão Baseada em Histograma (HistGradientBoostingRegressor), que se destacou com um R-quadrado de 0.922, explicando 92.3% da variabilidade no valor da oferta. A análise dos resíduos confirmou a adequação do modelo, embora haja espaço para refinamento em propriedades de alto valor.

Com os avanços obtidos nesta pesquisa, agora dispõe-se de um método automatizado de estimação para o mercado imobiliário. Este método, baseado em técnicas avançadas de aprendizado de máquina, demonstrou ser superior à tradicional regressão linear. A capacidade de processar e interpretar grandes volumes de dados de forma eficiente e precisa torna essa abordagem uma ferramenta valiosa para profissionais e tomadores de decisão no setor imobiliário, proporcionando estimativas mais acuradas e *insights* mais profundos sobre as dinâmicas do mercado.

Em trabalhos futuros, será essencial avaliar a seleção de feições do modelo, considerando aspectos como a importância estatística das feições, a simplificação e interpretabilidade do modelo, a minimização do risco de overfitting, a integração do conhecimento de domínio, a otimização do custo computacional, a análise de correlações entre feições e a reavaliação contínua do desempenho do modelo após qualquer modificação. Esta abordagem permitirá uma otimização mais refinada e eficiente de modelos a serem desenvolvidos.

Esses resultados reforçam a eficácia do HistGradientBoostingRegressor como uma ferramenta valiosa para a análise do mercado imobiliário. A precisão de 92% obtida com o

algoritmo LGBMRegressor, juntamente com o RMSE de 0,39 unidades na previsão da variável dependente, evidencia o alto grau de ajuste do modelo aos dados.

Para trabalhos futuros, a modelagem desenvolvida neste estudo poderá ser integrada aos sistemas corporativos da empresa, permitindo a previsão do valor venal estimado para cada novo registro no banco de dados. Estratégias para complementação de dados das novas observações podem incluir a imputação por previsão com algoritmos de aprendizado específico e o uso de geoprocessamento para adicionar novas camadas temáticas.

Além disso, à medida que novas vendas são realizadas, os valores obtidos podem ser utilizados dinamicamente para atualizar a resposta do modelo de aprendizado treinado, permitindo que o modelo se adapte às mudanças no mercado imobiliário.



REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

- AGUIAR, M. M.; SIMÕES, R.; GOLGHER, A. B. Housing market analysis using a hierarchical–spatial approach: the case of Belo Horizonte, Minas Gerais, Brazil. **Regional Studies, Regional Science, Taylor & Francis**, v. 1, n. 1, p. 116–137, 2014.
- ALMEIDA, E. **Econometria Espacial Aplicada**. Campinas-SP: Alínea, 2012.
- ARBIA, G. **A primer for spatial econometrics: with applications in R**. AIAA, 2014. 248 p.
- BATALHONE, S. A.; NOGUEIRA, J. M.; MUELLER, B. P. M. **Economics of air pollution: hedonic price model and smell consequences of sewage treatment plants in urban areas**. Brasília, DF: Universidade de Brasília, 2002. Disponível em: <http://chula.livocity.com/econ/Micro/Charit/Hedonic.pdf>. Acesso em 03 jan. 2023.
- BOENNEC, R. L.; SALLADARRÉ, F. The impact of air pollution and noise on the real estate market. The case of the 2013 European Green Capital: Nantes, France. **Ecological Economics**, v. 138, p. 82-89, 2017.
- BOTTERO, M.; BRAVI, M.; GASCA, E.; DELL'ANNA, F.; MONDINI, G. Valuing buildings energy efficiency through Hedonic Prices Method: are spatial effects relevant? **Journal Valori e Valutazioni**, n. 21, 2018.
- CAMPOS, R. B. A.; ALMEIDA, E. S. d. Decomposição espacial nos preços de imóveis residenciais no município de São Paulo. **Estudos Econômicos (São Paulo)**, SciELO Brasil, v. 48, n. 1, p. 5–38, 2018.
- CECCATO, V. Segurança e mercado imobiliário urbano. In: **Novas perspectivas analíticas e interpretativas da Ciência geográfica no atual contexto do sistema mundo**. Campinas: Paco Editorial, 2017.
- CESHIRE, P.; SHEPPARD, S. Capitalising the Value of Free Schools: The Impact of Supply Characteristics and Uncertainty. **The Economic Journal**, v. 114, p. 397-424, 2004.
- CIARLINI, M. G. C. **Precificação hedônica de imóveis locáveis no Recife** - 2019. 2020.
- CORDERA, R.; CHIARAZZO, E.; OTTOMANELLIM., DELL'OLIO, L.; IBEAS, A. The impact of undesirable externalities on residential property values:

spatial regressive models and an empirical study. **Transport Policy**, In press, 2018.

DANTAS, R. A. **Engenharia de Avaliações: Uma introdução à metodologia científica**. 3. ed. São Paulo: Pini, 2012. 255 p.

DIAS, A. A.; RAIHER, A. P. **Políticas fundiárias e cidades inteligentes: uma análise espacial do preço do solo urbano em Curitiba**. 2018.

DUNSE, N.; JONES, C. A hedonic price model of office rents. **Journal of Property Valuation & Investment**, v. 16, n. 3, p. 297-312, 1998.

FÁVERO, L. P. L.; BELFIORE, P. P.; LIMA, G. A. S. F. de. Modelos de precificação hedônica de imóveis residenciais na região metropolitana de São Paulo: uma abordagem sob as perspectivas da demanda e da oferta. **Estudos Econômicos**, São Paulo, v.38, n.1, p. 73-96, jan./mar., 2008.

FILIPA, S. **Algoritmos de aprendizagem de máquina que os engenheiros precisam de conhecer**. 2022. Disponível em: <https://www.cetax.com.br/blog/0-algoritmos-de-aprendizagem-de-maquina-que-os-engenheiros-precisam-de-conhecer/>. Acesso em 30 mar. 2022.

GIBBONS, S.; MOURATO, S.; RESENDE, G. M. The amenity value of English nature: a hedonic price approach. **Environmental and Resource Economics**, v. 57, n. 2, p. 175-196, 2014.

GOLDSTEIN, H. **Multilevel Statistical Models**. London: Arnold, 1995.

HERMANN, B. M.; HADDAD, E. A. Mercado imobiliário e amenidades urbanas: a view through the window. **Estudos Econômicos**, v.35, n.2, p. 237-269, 2005.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **PNAD Contínua - Pesquisa Nacional por Amostra de Domicílios Contínua**. IBGE, 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html?=&t=destaques>. Acesso em 04 jan. 2023.

JAHANSHIRI, E.; BUYONG, T.; SHARIFF, A. R. M. AA review of property mass valuation models. **Pertanika Journal of Science & Technology**, 29(2), 1549-1568. Disponível em: <https://doi.org/10.47836/pjst.29.2.14>. 2021. Acesso em 03 jan. 2023.

JULIO, R. **Modelo de preços hedônicos espacial: análise dos mercados implícitos para os imóveis residenciais de Porto Alegre**. 2021.

KAN, B.; PU, L.; XU, C. Fatores determinantes da heterogeneidade espacial do preço do terreno residencial no centro de Nanjing com base no modelo GWR. **Geografia Econômica**, v. 39, n. 3, p. 100-107, 2019.

LANCASTER, K. A new approach to consumer's theory. **Journal of Political Economy**, v.74, p. 132-157, 1966.

LE BOENNEC, R.; SALLADARRÉ, F. The impact of air pollution and noise on the real estate market. The case of the 2013 European Green Capital: Nantes, France. **Ecological Economics**, v. 138, p. 82-89, 2017. Disponível em: <https://ideas.repec.org/a/eee/ecolec/v138y2017icp82-89.html>. Acesso em 03 jan. 2023.

LEE, W.; KIM, N.; CHOI, Y. H.; KIM, Y. S.; LEE, B. D. Previsão baseada em aprendizado de máquina do valor dos edifícios. **Trans. KSII Internet Inf. Sistema (TIIS)**, v. 12, p. 3966-3991, 2018.

LIPORONI, A. S. Avaliação em massa com ênfase em planta de valores. In: ALONSO, N. R. P. (Coord). **Engenharia de Avaliações**. 2. ed. São Paulo: PINI, 2013.

MA, J.; CHENG, J. C. P.; JIANG, F.; CHEN, W.; ZHANG, J. Analisando fatores determinantes dos valores da terra em escala urbana com base em big data e técnicas de aprendizado de máquina não linear. **Política de Uso do Solo**, v. 94, 104537, 2020.

MACEDO, P. B. R. et al. **Hedonic price models with spatial effects: an application to the housing market of Belo Horizonte**, Brazil. Cedeplar/Face/UFMG, 1996. Disponível em: <https://bibliotecadigital.fgv.br/ojs/index.php/rbe/article/download/723/8081/0>. Acesso em 03 jan. 2023.

MCCLUSKEY, W. J.; DEDDIS, W. G.; MANNIS, A.; McBURNEY, D.; BORST, R. A. Interactive application of computer assisted mass appraisal and geographic information systems. **Journal of Property Valuation and Investment**, v. 15, n. 5, p. 448-465, 1997.

MCMILLEN, D. P.; MCDONALD, J. F. Land values in a newly zoned city. **Review of Economics and Statistics**, MIT Press, v. 84, n. 1, p. 62-72, 2002.

MCMILLEN, D. P.; REDFEARN, C. L. Estimation and hypothesis testing for nonparametric hedonic house price functions. **Journal of Regional Science**, Wiley Online Library, v. 50, n. 3, p. 712–733, 2010.

MEI, Y.; ZHAO, X.; LIN, L.; GAO, L. Capitalização da vegetação verde urbana em um mercado imobiliário com má qualidade ambiental: Evidências de Pequim. J. **Plano Urbanístico**. Dev., v. 144, 05018011, 2018.

MOREIRA DE AGUIAR, M.; SIMÕES, R.; BRAZ GOLGHER, A. Housing market analysis using a hierarchical–spatial approach: the case of Belo Horizonte, Minas Gerais, Brazil. **Regional Studies, Regional Science**, v. 1, n. 1, p. 116-137, 2014. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/21681376.2014.934391>. Acesso em 03 jan. 2023.

MOREIRA, F. F. **Modelo hedônico espacial para avaliação em massa de imóveis de Fortaleza**. 2020.

MOSSIN, J. Equilibrium in a asset market. **Econometria**, v. 34 , n . 4, p. 768-783, 1966.

NASCIMENTO, W. E. M. L. **O tratamento de dados na avaliação de imóveis urbanos por meio de regressão linear** – Goiânia, 2021.

OECD, et al. Hedonic Regression Method. **Handbook on Residential Property Price Indices (RPPIs)**, 2013.

OLIVEIRA, D. X. A. **Os impactos da criminalidade em uma análise teórica, empírica e espacial**. Fortaleza. 2019.

PAIXÃO, L. A. R. O impacto da violência no preço dos imóveis comerciais de belo horizonte: Uma abordagem hedônica. **Revista De Economia Aplicada**, v. 13, p. 125–152, 2009.

_____, L. A. R. Índice de preços hedônicos para imóveis: uma análise para o município de Belo Horizonte. **Economia Aplicada**, v. 19, p. 5-29, 2015. Disponível em: <https://www.scielo.br/j/ecoa/a/8CNnDy9Mww43PRkpp6fXgvz/abstract/?lang=pt>. Acesso em 03 jan. 2023.

RAMOS, R. **Regressão Linear Simples: O Que é? Para Que Serve? Como Funciona?** 2020. Disponível em: <https://oestatistico.com.br/regressao-linearsimples/>. Acesso em 18 jul. 2021.

RAUDENBUSH, S. W.; Bryk, A. S. **Hierarchical linear models: applications and data analysis methods**. Second edition, Thousands Oaks, CA: Sage Publications, Inc., 2002.

ROSEN, S. Hedonic prices and implicit markets: production differentiation in pure competition. **Journal of Political Economy**, v. 82, n. 1, p. 34-55, 1974.

SALGADO, V. M. **Curso de Avaliação de Imóveis: Metodologia Científica. INEAA Educação**. Goiás, 2017. 68 p.

SANGANI, D.; Erickson, K.; Hasan, M. A. Prevendo o erro de estimativa de Zillow usando regressão linear e aumento de gradiente. In **Proceedings of the 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)**, Orlando, FL, EUA, 22 a 25 de outubro de 2017; págs. 530–534.

SANTOS, B. R. dos; **Modelos de Regressão Quantílica**. Dissertação de mestrado Universidade de São Paulo, São Paulo, 2012.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

SEABRA, D.; SILVEIRA NETO, R.; MENEZES, T. Amenidades urbanas e valor das residências: Uma análise empírica para a cidade de Recife. **Economia Aplicada**, v.20, n.1, p. 143-169, 2016.

SEABRA, D. M.; NETO, R. da M. S.; DE MENEZES, T. A. Amenidades urbanas e valor das residências: uma análise empírica para a cidade do Recife. **Economia Aplicada**, v. 20, n. 1, p. 143-169, 2016. Disponível em: <https://www.revistas.usp.br/ecoa/issue/download/8431/594#page=147>. Acesso em 04 jan. 2023.

SHEPPARD, S. **Hedonic Analysis of Housing Markets. Handbook of Regional and Urban Economics**. Amsterdam, pp. 1.595-1.636, 1999.

SOUZA, Á. S. de et al. **Construção de variáveis independentes caracterizadoras do espaço urbano e qualificadoras do mercado imobiliário de Lauro Müller?** SC. 2022. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/243713/PTRA0111-D.pdf?sequence=-1>. Acesso em 03 jan. 2023.

SRIVASTAVA, S. Weka: Uma ferramenta para pré-processamento de dados, classificação, ensemble, clustering e mineração de regras de associação [J]. **Revista Internacional de Aplicações de Computador**, 2014 , 88 (10): 26 - 29 .

UBERTI, M. S. **Avaliação em massa de imóveis rurais através de modelagem clássica, espacial e geoestatística**. 2016. 151 f. Tese (Doutorado em Ciência, Tecnologia e Inovação) – Universidade Federal Rural do Rio de Janeiro (UFRRJ), Rio de Janeiro, 2016.

VELOSO, J. F. A.; FREIRE, T. de J. S.; BEMERGUY, D. A.; GALVÃO, M. S. Abordagem de preços hedônicos para precificação de lotes: estudo de caso para o bairro do Taquari. **Terracap**. 2017.

WEI, C.; FU, M.; WANG, L.; YANG, H.; TANG, F.; XIONG, Y. The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of Big Data. **Land** 2022, 11, 334. <https://doi.org/10.3390/land11030334>

WON KIM, C. W.; PHIPPS, T. T.; ANSELIN, L. Measuring the benefits of air quality improvement: a spatial hedonic approach. **Journal of Environmental Economics and Management**, v. 45, p. 24-39, 2003.

WON, J.; LEE, J. Investigating how the rents of small urban houses are determined: Using spatial hedonic modeling for urban residential housing in Seoul. **MDPI Sustainability**, v.10, n.31, 2018.

XU, L. **Pesquisa sobre avaliação de lotes de preços de imóveis com base em rede neural profunda**. Dissertação de mestrado, Universidade de Tecnologia de Harbin, Harbin, China, 2018.

YANG, X.; ZHANG, Y.; LI, X.; YU, Y.; WANG, L. A comparative analysis of the impacts of objective versus subjective neighborhood environment on physical, mental, and social health. **Health and Place**, v. 58, 102168, 2019.

ZHAO, Z.; CHEN, Y.; WANG, Z. A deep learning ensemble approach for crude oil price forecasting. **Energy Economics**, v. 66, p. 9-16, 2017.

ZHU, H.; LI, R. Y. M.; LI, S. Modelling property price patterns in Hong Kong: a review and empirical study. **Property Management**, v. 34, n. 1, p. 5-28, 2016.



APÊNDICE

APÊNDICE

Apêndice 1

Lista temática adquirida no site GEOPORTAL e o número de observações em cada uma delas.

Feição geográfica	Contagem de ocorrências
escolas	1196
áreas de interesse ambiental	22
estações e terminais	43
quadras	4700
espaços culturais	404
espaços comunitários	2225
distância de 600 metros	43
mobiliário esporte e lazer	2458
sistema cicloviário	1930
mapa 5 densidades demográficas	1
densidade urbana	4
estação de metro	29
zoneamento das diretrizes urbanísticas	130
linhas de transportes de alta densidade	96
zoneamento do distrito federal	78
pontes viadutos e passarelas	764
população abrangida pela infraestrutura cicloviária	4454
unidades de conservação e parques	94
mapa 1c unidades de planejamento territorial	6
área tombada	1
evolução urbana	37
regiões administrativas	32
mapa 6 áreas econômicas	54
parques urbanos	33
equipamentos de segurança	319
equipamentos de saúde	131
feiras livres	249
metro	3
rede estrutural de transporte coletivo	240
intervenções cicloviárias programa nos urbanos	278
área econômica	54
limite do df	1
ciclovias	1749
distância de 400 metros	43
área de proteção de manancial	26
regularização	99
setor	271
unidades de conservação gestão ibram	72
diretrizes urbanísticas	50
rodovia	497



Apêndice 2

Código fonte disponível no repositório: <https://github.com/willamydias/mestrado>





idn

Bo
pro
cit
ref
Nos
são

idp

A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO