

Machine Learning e Mercado Imobiliário: Usos Alternativos à Precificação de Imóveis

ALAN MARQUES MIRANDA LEAL (*)

1 Introdução

A precificação correta de imóveis é um tema largamente estudado e implementado dentro da Econometria, principalmente via estimação de preços hedônicos, e outros métodos de *machine learning*, com seus modelos de previsão de valores de forma intuitiva e *black box*, que produzem ainda assim resultados interessantes. O presente texto, contudo, não almeja fazer uma revisão da literatura sobre métodos diversos de precificação de imóveis; ele vem num estágio anterior à precificação em si, qualificando o processo de definição de mercado no qual determinado imóvel está sendo negociado. Um problema para o agente coletor de impostos imobiliários diz respeito à sua (in-)capacidade de identificar mudanças no mercado de imóveis em uma determinada região. Naturalmente, vetores de urbanização já conhecidos, como estações de metrô (São Paulo) ou *shopping centers* (Belo Horizonte), podem ser bons indicativos de que um imóvel do tipo casa está sendo negociado em mercado outro que sua tipologia declarada. Isso ocorreria principalmente em imóveis do tipo casas, barracões e galpões

sendo negociados em mercados de terreno (lotes) e não no seu próprio mercado. Para o formulador de políticas e o coletor de impostos, saber em qual mercado um determinado imóvel está sendo negociado é crucial para que a estimação de seu valor seja a mais fidedigna. Uma estimação acurada do preço de um imóvel implica uma coleta de impostos imobiliários justa e clara para o ente público e também para todos os contribuintes. Além dessa Introdução, o presente texto se divide numa seção de métodos, explicando as possibilidades metodológicas desse exercício. A seção seguinte explica como esse método foi aplicado para o município de Belo Horizonte (MG). Por fim, uma seção final conclui com os principais achados deste texto.

2 Método: Modelos de Classificação

O problema pré-precificação de imóveis que se apresenta é aquele de definir se um novo imóvel recebido pelo ente governamental como recém-vendido foi negociado no seu próprio mercado ou no mercado de terrenos. Intuitivamente, um imóvel negociado em mercado de terrenos tende a ser precifi-

cado pelo tamanho – área – do imóvel, junto com seu potencial construtivo, isto é, o coeficiente de aproveitamento permitido pelo zoneamento na cidade. Por outro lado, no caso de uma casa, características como números, tamanho e tipos dos cômodos importam, assim como seu nível de acabamento. Subjacente a essas questões também se encontra o aspecto espacial. A casa recém-vendida se encontra adjacente a várias outras casas, isto é, a região é residencial. Ou então a casa é cercada por inúmeros prédios e destoa do resto da paisagem urbana naquela localidade. Aspectos locais comuns, como acesso a transporte público, equipamentos de saúde, culturais e ambientais e de outros tipos tendem a impactar também o preço de um imóvel e esse impacto não necessariamente será neutro em relação ao mercado de negociação do imóvel. O argumento aqui é o seguinte: por exemplo, uma estação de metrô próxima a uma casa que continuará sendo casa terá impacto diferenciado de preço ao de uma casa que será transformada num prédio.

Essa é uma tarefa de uso bastante comum de métodos do *Machine*

Learning, qual seja, uma tarefa de classificação binária. Quando um novo imóvel é vendido na cidade, para que essa venda seja propriamente realizada, é necessário haver registro perante a prefeitura local. A prefeitura conhece várias características do imóvel, tem seu próprio preço estimado para o imóvel, sobre o qual baseia-se sua tributação, e acolhe o novo preço de venda desse imóvel no mercado. A prefeitura precisa reestimar seu preço de mercado de posse dessa nova informação e ela esbarra na seguinte e crucial questão: o imóvel está sendo precificado mais por características do imóvel (mercado próprio) ou mais por características e terreno (mercado de lotes)?

Se a prefeitura possuir um histórico prévio de imóveis negociados e que depois foram substituídos por outro perfil construtivo, então ela pode empreender um exercício de previsão de mercados com base nessa informação.

Os modelos disponíveis são diversos; segue uma breve lista de alguns deles, que podem ser cogitados a depender de qual é o interesse do pesquisador (maior acurácia ou maior interpretabilidade):

- LOGIT: esse modelo, também tradicionalmente usado na Econometria, consiste em modelagem do valor esperado de uma variável binária condicional em variáveis explicativas, conside-

rando o *link* da função sigmoide logística.

- SVM: esse modelo constrói um hiperplano separador das observações num espaço multidimensional. O espaço tem $K+1$ dimensões, em que K é o número de variáveis preditoras e a dimensão extra dá conta justamente da variável a ser classificada.
- Árvores de decisão: criam-se árvores de decisão, baseadas em valores de outras variáveis, para prever o valor de uma variável de interesse.
- *Boosted Trees*: o *boosted tree* é um avanço da árvore de decisão na medida em que a árvore de decisão final consiste numa ponderação das árvores de decisão que mais acertaram isoladamente. A lógica por trás desse método consiste em fazer com que árvores mais bem-sucedidas na tarefa de classificação tenham maior peso nas regras de decisão que o algoritmo indicará na sua árvore final.

Considerando essa seleção de modelos disponíveis para a tarefa de classificação visada, o próximo passo consiste na validação do modelo. Para tanto, métricas tais como a acurácia, área sobre a curva e o coeficiente de correlação de Matthews parecem bastante

razoáveis em apontar os melhores modelos de previsão. Via de regra, a listagem anterior dos modelos usuais da tarefa de classificação possui acurácia em ordem crescente, isto é, o modelo LOGIT menos preciso em suas previsões que o modelo de *boosted trees*. A interpretabilidade, por outro lado, tem ordem decrescente entre os modelos presentes na listagem anterior. Logo, um *trade-off* entre interpretabilidade e capacidade preditiva se estabelece nessa tarefa de classificação. A escolha do modelo deve considerar aspectos legais e políticos relacionados a quais são os objetivos do uso desse modelo, e a necessidade de seu uso efetivo precisa ou não estar prevista em legislação apropriada. De toda forma, caso o modelo de classificação dos imóveis tenha uma característica apenas intermediária no processo de precificação dos imóveis, isso não chega a configurar um problema. Naturalmente, um modelo ruim de classificação dos imóveis poderia contaminar de alguma forma a precificação dos imóveis pelo ente público, introduzindo vieses diversos em sua estimação.

3 Classificando Imóveis em Belo Horizonte (MG)

A Prefeitura de Belo Horizonte, responsável por definir alíquotas e coletas relacionadas aos imóveis nas suas cercanias na forma do IPTU (Imposto Predial e Territorial Urbano) e do ITBI (Imposto sobre a

Transmissão de Bens Imóveis) é um ente interessado em estimativas mais precisas para os valores dos imóveis da cidade. Contudo, a problemática levantada no presente texto pode contribuir por reduzir a acurácia de suas previsões. Logo, um modelo de classificação bem estabelecido e testado se faz bastante apropriado. Em termos práticos, para a estimação desse modelo utilizaram-se todos os imóveis vendidos na cidade previamente, assim como aqueles que sofreram alterações de seus tipos construtivos. A variável de interesse assumiria o valor 1, caso o imóvel negociado tivesse sofrido alteração de tipo construtivo em algum momento ulterior à sua venda, e 0, caso contrário. Usualmente, para imóveis recém-vendidos, há um descasamento entre o momento da venda e quando alguma possível alteração de seu tipo construtivo fosse declarada para a prefeitura. De toda forma, a *rationale* consiste em construir um modelo de classificação dos imóveis com base em mudanças de tipo construtivo históricas e a partir daí ter uma boa estimativa para o que estava acontecendo com imóveis recém-vendidos no mercado de imóveis de Belo Horizonte.

De posse do modelo bem estabelecido e testado, o método de *bagged trees* foi escolhido como modelo final. Na prática, assim que um novo imóvel fosse vendido na cidade e houvesse a emissão de guia para pagamento do ITBI, os técnicos da Prefeitura de Belo Horizonte teriam capacidade de, com base no modelo de classificação, verificar se aquele imóvel tinha sido precificado no mercado próprio (casas, galpões e barracões) ou mercado de terrenos (lotes). Ao ente público, interessava se antecipar a essa precificação diferenciada do imóvel de forma a conferir justiça, clareza e precisão na arrecadação tributária relacionada ao mercado imobiliário da cidade.

4 Conclusão

Este breve texto teve por interesse apresentar e contextualizar a aplicação de um modelo de *Machine Learning* pelo setor público na área do mercado imobiliário. O enfoque da abordagem foi expor a problemática que requereu o uso desses métodos e as possíveis soluções usadas para abordar o problema. Espera-se que outros municípios se apoderem dos *insights* aqui expostos para aumentar a eficiência e eficácia de arrecadação tributária, no que concerne aos imóveis.

(*) *Doutorando em Teoria Econômica na FEA/USP.*
(E-mail: prof@alanleal-econ.com).