

APLICAÇÕES DE *MACHINE LEARNING* PARA VALORAÇÃO DE EMPREENDIMENTOS RESIDENCIAIS DE BASE IMOBILIÁRIA NA CIDADE DE SÃO PAULO

Profª. Dra. Abla Maria Proência Akkari Osso e Especialista Thays Carmonha Damin

O Machine Learning (ML) é uma vertente da Inteligência Artificial (IA), termo que se tornou uma verdadeira "*buzzword*" e está cada vez mais integrado ao nosso cotidiano.

Se você deixou a TV ligada em casa e pediu à ALEXA para desligá-la, ou se utilizou o chat GPT para corrigir um texto, ou ainda se programou seu carro para estacionar sozinho devido à dificuldade ou preguiça, você já fez uso da IA.

Nesta carta, exploraremos as influências dessa tecnologia no mercado de Real Estate, focando especificamente na valoração de empreendimentos de base imobiliária (EBIs) residenciais. Apresentaremos a metodologia e os resultados que permitiram avaliar o nível de engajamento e aderência do ML em uma amostra representativa desse nicho, destacando suas implicações e potenciais.

1. Mas o que é a IA, o ML e quais as suas influências no Real Estate?

Começando por explicar o termo, a IA foi definida como "a ciência e engenharia de fazer máquinas inteligentes"¹, em 1955, pelo professor John McCarthy. Já o ML acaba sendo a parte da IA que estuda como os agentes computacionais podem melhorar a percepção, conhecimento ou ações das máquinas com base em experiência ou dados.

O ML se baseia em ciência da computação, estatística, psicologia, neurociência, economia e teoria de controle, um ramo da engenharia que lida com o uso de respostas ou feedbacks de um sistema para influenciar o seu comportamento a fim de se treinar e atingir um objetivo desejado, diminuindo erros.

Levando em conta que o mercado imobiliário é marcado por decisões complexas, com diversas variáveis e fontes de dados, o ML acaba contribuindo na melhor leitura desse conjunto de variáveis, consideradas independentes, e que influenciam o valor de um imóvel (e.g. atributos físicos, de localização, infraestrutura, serviços no enclave de inserção urbana e outros fatores relevantes). Os modelos de ML têm capacidade, se

¹ MANNING, C. (2020). Artificial Intelligence Definitions. Palo Alto: Stanford University. HAI (Human-Centered Artificial Intelligence).

treinados a partir dos dados disponíveis, de acessar uma compreensão mais profunda das relações não-lineares entre essas variáveis, contribuindo para uma avaliação mais precisa da rentabilidade de ativos imobiliários, por exemplo.

Para ilustrar o potencial dessa tecnologia, um dos exemplos apresentados no trabalho que deu origem a essa carta é a pesquisa de Potrawa, que focou seus esforços em modernizar a metodologia hedônica tradicional de valoração, com seus modelos simples de regressão linear, a fim de compreender a percepção dos investidores sobre o valor das características dos imóveis em Roterdã e descobrir relações complexas que não poderiam ser capturadas por modelos convencionais.

Os dados utilizados no estudo de Potrawa foram coletados do *Journal of Business Research*, que incluíam informações sobre as características das propriedades (como o preço, a área de estar, o número de quartos, o tipo de casa, a presença de garagem, banheiro, elevador, entre outros) e descrições das propriedades, bem como informações sobre o tempo de viagem a pé, de bicicleta e o transporte público até o centro da cidade. Além disso, também foram analisadas imagens das propriedades, coletadas de sites de aluguel e do Google Maps. Essas imagens foram usadas para obter variáveis adicionais, como a presença de uma vista da cidade, a proximidade de áreas verdes ou corpos d'água, e a qualidade da insolação. Essas variáveis foram extraídas usando técnicas de reconhecimento de imagem, como *Convolutional Neural Networks*.

Os resultados do estudo de Potrawa mostraram que a aplicação de técnicas de ML aumentou em 25% a precisão dos modelos de preços hedônicos. Identificando que o preço marginal da área habitável para propriedades alugadas é apenas linear por partes, sendo que cada metro quadrado adicional entre alguns grupos de faixa de metragem representa um aumento ou diminuição do valor percebido, ou seja, daquilo que pode ser convertido em preço pago pelo ativo.

Adicionalmente, com o uso de covariáveis extraídas de imagens de satélite foi descoberto que o valor dos atributos de habitação depende não apenas de sua própria quantidade e qualidade, mas também da quantidade de outros atributos. Um resultado interessante foi o valor da vista panorâmica de Roterdã, que foi estimado em mais de 100 euros por mês, para propriedades grandes e bem localizadas.

Como esse exemplo, o trabalho de monografia a que se baseia essa carta também apresenta outros estudos de caso a partir de uma ampla bibliografia. A fim de dar visibilidade sobre aplicações dos algoritmos de ML para a valoração de EBIs, esses estudos foram divididos em dois grupos: (1) Avaliação de desempenho de modelos de ML para definição de valor de venda de um ativo residencial, testando quais os

algoritmos mais eficazes para, após treinados, atingirem valores próximos aqueles praticados no mercado, e (2) Avaliação de modelo para descobrir relações complexas que não poderiam ser capturadas por modelos convencionais. Esses estudos de caso foram selecionados tanto pela diversidade quanto às fontes de dados e localização geográfica de cada exemplo como também pela variedade de algoritmos selecionados.

2. Quais são algumas das vantagens e desafios para adoção do ML no mercado de valoração de EBIs?

Considerando que o principal objetivo do empreendedor de EBIs é obter renda por meio da exploração do espaço físico, a decisão de se investir é, em grande parte, apoiada quase exclusivamente na relação entre aluguel praticado e preço da transação, que são diretamente associados ao poder de troca do bem e, portanto, no valor que este bem será transacionado.

Levando em conta que o EBI possui um ciclo de vida longo, existe uma série de fatores que vão influenciar o seu valor no momento da transação, desde características físicas, de localização e qualidade do imóvel, até conceitos mais abstratos, como oferta e demanda, taxas de juros e economia em geral. Nesse sentido, essa diversidade de agentes que moldam a decisão do investidor, que é o tomador do risco, pode dificultar o prognóstico do *Market Tradable Value* (MTV) para arbitragem do preço do ativo e do fluxo de recebimento de resultados provenientes da operação do empreendimento (RODi - Resultado Operacional Disponível).

Alguns sustentam que há evidências de que a aplicação de modelos de ML apresentaria menos erros de estimação quando comparados a outros modelos de valoração imobiliária e isso ocorre em grande parte devido a capacidade de capturar relacionamentos não lineares de atributos que possuem influência sobre o valor percebido do imóvel. Além disso, os modelos de ML podem considerar não apenas os atributos estruturados como também as informações não estruturadas, como descrições de propriedades e imagens.

Ainda assim, segundo algumas observações que foram identificadas a partir de artigos científicos, a percepção geral aponta que o setor imobiliário é relutante em adotar mudanças rapidamente. Algumas entrevistas realizadas com acadêmicos e profissionais do setor, inclusive, comentam sobre a hesitação em adotar novas tecnologias por parte das empresas do setor imobiliário, tanto devido a atritos culturais quanto pelo desejo de não compartilhar dados por conta de vantagens competitivas. No Brasil, principalmente devido ao fato de que as transações acontecem em sua

maioria no mercado privado, sem a necessidade de tornar público o dado do valor transacionado, existe um nível de protecionismo que acaba dificultando a transição para mudanças tecnológicas.

Dentre as entrevistas aplicadas para esse trabalho, as principais oportunidades e desafios identificados pelas empresas desse nicho foram, [1] a digitalização das documentações em cartório, que tornará possível acessar dados de transação para compra e venda de imóveis e, conseqüentemente, influenciará na qualidade dos dados que treinam os algoritmos tonando as estimativas mais próximas ao *Market Tradable Value* e [2] a busca por uma solução que unifique dados de locação diante de um mercado que foi identificado por uma das empresas como “fragmentado”.

3. Recorte, Objetivo e Metodologia

A fim de identificar o nível de engajamento e adoção de empresas focadas na área de valoração de empreendimentos de EBI residencial, esta carta busca apresentar os resultados identificados através da realização de entrevistas de 8 empresas representativas nesse setor, em São Paulo – Baseado no trabalho de Damin (2024).

O objetivo da pesquisa que deu origem a esta carta foi o de avaliar como o ML tem sido aplicado nas empresas ou qual a opinião geral para a adoção dessa tecnologia em sua empresa ou no setor. Deste modo, buscou-se, atingir uma dinâmica em que seria possível detectar qual o nível de aderência às tecnologias de ML e IA que as empresas entrevistadas se encontram; identificando se seu uso já estava presente nas empresas ou se estava num estágio inicial, ou seja, quando existe a reconhecimento sobre a existência da tecnologia e de seu potencial, porém sem abertura para aplicação no ambiente profissional.

A metodologia do estudo apresentado foi a SURVEY, que é baseada num instrumento de pesquisa, normalmente um questionário, a fim de obter dados ou informações sobre características, ações ou opiniões de determinado grupo de pessoas, que representam uma população-alvo.

O questionário aplicado nas entrevistas buscou responder duas questões centrais:

A. Como as tecnologias de IA e ML estão sendo aplicadas no setor residencial de base imobiliária?

B. Quais oportunidades concretas se apresentam para a adoção de IA e ML no setor residencial de base imobiliária?

Para abordar essas questões, foram feitas 8 perguntas, com o propósito de avaliar a aderência das empresas selecionadas quanto ao conhecimento e a aplicação das tecnologias em questão. Além disso, as entrevistas buscaram identificar como essas tecnologias já estão sendo implementadas nas empresas dos entrevistados e explorar as possibilidades futuras que elas podem oferecer.

Dentre as empresas selecionadas, foi feita uma divisão em 3 categorias: C (Empresas focadas na comercialização de ativos residenciais para renda), LS (Empresas focadas no desenvolvimento, administração e operação de residenciais de locação de longa estadia) e BD (Empresas focadas no oferecimento de inteligência com base em dados para o setor residencial para renda). Sendo cada empresa nomeada conforme seu grupo: C1, C2, C3, C4, LS1, LS1, LS3 e BD1.

Finalmente, é importante ressaltar que, na escolha dos entrevistados, foram priorizados profissionais ocupando cargos de liderança nas empresas selecionadas e, em algumas exceções, profissionais indicados por pessoas em tais cargos. Estes líderes possuem uma visão ampla e estratégica, compreendendo não apenas a atual implementação de tecnologias como ML e IA, mas também vislumbrando oportunidades para aprimorar sistemas e processos internos. Destacando indivíduos com influência e conhecimento abrangente, a abordagem adotada busca obter uma perspectiva mais profunda sobre as estratégias em vigor, os desafios enfrentados e as projeções futuras dessas empresas no contexto dinâmico do mercado imobiliário. Essa metodologia visa não apenas compreender o estado atual, mas também antecipar e explorar as potenciais melhorias e inovações que essas lideranças enxergam para o futuro de suas organizações.

4. O questionário

O questionário seguiu duas seções, explicadas a seguir:

I. SEÇÃO 1: FOCO EM APLICAÇÕES NA EMPRESA

Nesta seção, o objetivo foi identificar se houve implementações de ML e IA nas empresas representadas pelos entrevistados. Compostas por cinco perguntas, a seção abrange (a) especificação sobre quais tipos de dados e fontes de dados a empresa utiliza, (b) exemplos de maneiras que a empresa utiliza esses dados e (c) caso exista uso de ML, como esse uso está sendo aplicado (quais algoritmos e quais as áreas em valoração).

Como não houve filtragem prévia para selecionar as empresas que utilizam ML e AI de alguma maneira, possibilitando, assim, respostas afirmativas ou negativas à pergunta sobre implementação dessas tecnologias. Apesar disso, a continuidade da seção, ao indagar sobre os tipos de dados utilizados, busca-se extrair respostas que podem direcionar a *insights* para a identificação de potenciais benefícios associados ao emprego dessas tecnologias.

II. SEÇÃO 2: FOCO EM OPINIÃO GERAL DE MERCADO E IDENTIFICAÇÃO DE OPORTUNIDADES

Esta seção, foi composta por três perguntas, visando compreender a visão dos entrevistados em relação às oportunidades que podem aprimorar as práticas no mercado imobiliário. Inicialmente, o entrevistado foi questionado sobre se está ciente de outras maneiras pelas quais o ML está sendo aplicado no nicho em questão e, em seguida, mais especificamente, sobre quais as áreas, dentro do mercado de valoração desses ativos, eles acreditam que existe a oportunidade de aplicar as técnicas de ML. Finalmente, de maneira mais genérica, a fim de captar respostas de entrevistados menos familiarizados com a tecnologia, mas que podem estar num estágio inicial de utilização de dados, perguntou-se quais tipos de dados eles acreditam estarem faltando ou serem úteis para que enfim o ML pudesse ser aplicado.

5. **Análise dos dados obtidos através das entrevistas**

A fim de estruturar as conclusões obtidas para definição de qual o nível de aderência do uso de tecnologias de ML na amostragem de empresas selecionadas, criou-se uma divisão em 4 níveis:

- Nível 1. Empresas que já utilizam dados para processos da empresa em geral;
- Nível 2. Empresas que conhecem o conceito de IA e ML;
- Nível 3. Empresas que veem potenciais de aplicação de tecnologias de IA e ML no seu nicho de EBI residencial ou em outros nichos dentro de Real Estate, sendo capazes de enxergar oportunidades;
- Nível 4. Empresas que já aplicam ML.

Dentro da amostra, tivemos os seguintes resultados:

Nível 1. Empresas que já utilizam dados para processos da empresa em geral

(8 das 8 empresas atingiram o Nível 1)

Todas as empresas entrevistadas declaram utilizar dados em suas operações, indicando a importância tanto de [1] dados estruturados (por exemplo, planilhas e bancos de dados relacionais), quanto de [2] dados não estruturados (por exemplo, texto na web, documentos de texto não formatados, mídias sociais, áudios e imagens), em que “texto na web” foi o mais selecionado, com 7 dentre 8 votos, seguido de “imagens”, com 3 votos e “documentos não formatados” com 2 votos.

Esse resultado indica que existe grande potencial para o aproveitamento de dados não-estruturados, que ainda são pouco utilizados. Portanto, entendendo que o uso de dados é o primeiro passo para avaliar como processar esses dados por meio de tecnologias como ML, por exemplo, todas as empresas atingiram o Nível 1 de aderência.

Nível 2. Empresas que conhecem o conceito de IA e ML

(8 das 8 empresas atingiram o Nível 2)

Com base nas respostas à pergunta 1, todas as empresas entrevistadas demonstraram um entendimento preciso de ML e IA, alinhado com as definições deste estudo. Isso sugere que esses conceitos não são estranhos a eles, o que pode indicar a possibilidade para que, futuramente, oportunidades de implementação dessas tecnologias sejam identificadas em seus respectivos setores.

Nível 3. Empresas que veem potenciais de aplicação de tecnologias de IA e ML no seu nicho de EBI residencial ou em outros nichos dentro de Real Estate, sendo capazes de enxergar oportunidades

(6 das 8 empresas atingiram o Nível 3)

Todas as empresas assinalaram alternativas que indicavam ou terem conhecimento de outras aplicações dentro no nicho de valoração de EBIs residenciais que estivessem fora da empresa em que trabalham ou serem capazes de identificar oportunidades para que tecnologias de ML pudessem ser aplicadas em seu nicho, tanto em relação a tipos de dados quanto a áreas específicas no mercado de valoração de EBIs residenciais. Isso indica que o conceito de ML tem seu valor percebido entre as empresas.

Entretanto, somente 6 das 8 empresas foram capazes de dissertar sobre suas escolhas ao menos em uma das alternativas, indicando maior conhecimento sobre como efetivamente o ML poderia contribuir para seu nicho. Para aquelas empresas que conseguiram justificar sua resposta com detalhes, considerou-se que foi atingido o Nível 3 de aderência. Para aquelas que, (1) não conseguiram justificar ou exemplificar sua resposta, (2) tentaram justificar sua resposta, apontando alguns exemplos, sem muitos detalhes, ou (3) não apontaram maneiras específicas de aplicação, mas destacou pontos relevantes à área em estudo (eg. desafios, deficiências etc.) considerou-se que não foi atingido o Nível 3.

Nível 4. Empresas que já aplicam ML

(5 das 8 empresas atingiram o Nível 4)

Apesar de 6 dos 8 entrevistados terem indicado aplicar ML dentro de suas empresas, a empresa C3 não conseguiu indicar quais os algoritmos utilizados. Dentre as empresas, a C4 é a que utiliza maior quantidade de algoritmos, com 7 algoritmos, 6 deles listados na alternativa de múltipla escolha e 1 (Large Language Models (LLMs)) adicionado pelo entrevistado. BD1 é a segunda empresa com maior uso de algoritmos, com 5, seguida pela LS1 e LS2, com 3 algoritmos, e, finalmente, pela C2, com 2 algoritmos.

Dentre as aplicações de tais algoritmos, foram destacados a coleta de dados, inclusive com a utilização de web scrapers, a valoração automatizada, com indicação de imóveis similares com base imóvel modelo e identificação de objetos em imagens/fotos de satélite.

6. Resultados

Os resultados obtidos através do trabalho exposto nessa carta indicaram que:

[A] a ML melhora a precisão da avaliação imobiliária e permite identificar relações complexas que não são possíveis por modelos tradicionais;

[B] 62% das empresas da amostragem entrevistada já utilizam ML e 75% veem potencial para o uso de tecnologias de IA e ML em seu segmento, sendo que, desse universo, 100% já utiliza dados para processos da empresa em geral e já estão familiarizados com o conceito de ML e, adicionalmente;

[C] a barreira documental de acesso a dados de transação no mercado brasileiro é um dos principais desafios identificados pelos entrevistados para que o uso das tecnologias

de ML possa ser mais difundido no mercado imobiliário de EBIs residenciais.

7. Conclusão

As entrevistas realizadas desempenharam um papel fundamental, permitindo a identificação de camadas profundas de conhecimento por parte dos representantes de cada empresa. Elas proporcionaram uma compreensão mais holística dos conceitos, aplicações e oportunidades que essas tecnologias podem oferecer na indústria imobiliária e serviram como uma amostra do nível de aderência das tecnologias de ML e IA no mercado de EBIs em São Paulo.

Ao longo deste trabalho, ficou evidente que os resultados apontam para a relevante melhoria na precisão alcançada por resultados obtidos por meio da aplicação de Machine Learning (ML) e Inteligência Artificial (IA) no mercado imobiliário, especialmente na área de valoração. No entanto, apesar desses avanços, diversas barreiras impedem a plena adoção dessa tecnologia no mercado imobiliário como um todo. Uma das principais barreiras é o desafio relacionado ao retorno financeiro, que ainda não se mostra tão significativo em comparação com as melhorias no processo, principalmente em áreas mais analíticas e regulamentadas. Em outras palavras, os modelos tradicionais ainda desempenham um papel eficaz neste momento.

Em conclusão, o potencial transformador da tecnologia de ML e IA é claro no mercado de Real Estate e, especialmente, na valoração. Apesar das atuais barreiras, os benefícios estratégicos e a evolução constante dessas tecnologias indicam um futuro promissor para a sua implementação generalizada na indústria imobiliária. Sendo assim, voltando a questão inicial, quem acreditar que a IA é um *buzzword* e logo irá passar estará completamente enganado. Essa tecnologia está ganhando cada vez mais maturidade e vantagem competitiva a todos aqueles que estão investidos nas pesquisas, e o setor de Real Estate não foge a essa tendência.