

## **XII CONGRESSO DE INICIAÇÃO CIENTÍFICA DO IFSP ITAPETININGA**

Itapetininga, 19, 20 e 21 de maio de 2026

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo

*Campus Itapetininga*

### **ANÁLISE DE SÉRIES TEMPORAIS MULTIVARIADAS DO MERCADO IMOBILIÁRIO COM MODELOS ESTATÍSTICOS E REDES NEURAIS**

Pedro de Moraes Machado – PIBIC/IFSP<sup>1</sup>

Prof. Dr. Tardelli Ronan Coelho Stekel- IFSP<sup>2</sup>

#### **Introdução**

A necessidade de moradia é algo inerente a qualquer ser humano, sendo um fator determinante para a qualidade de vida. Esta condição gera uma divisão clara no mercado, de pessoas que buscam apenas imóveis para a satisfação de moradia e os que desejam uma fonte de investimento. Perante a essa ambivalência, é irrevogável que a formação dos preços tornam-se um desafio complexo, pois agora se torna crucial entender como os dados econômicos funcionam para prever as tendências do mercado imobiliário, já que o preço é influenciado por um conjunto de externalidades geográficas e econômicas (ARRAES; SOUSA FILHO, 2008). Destarte, com a alta disponibilidade de dados públicos atualmente se abrem novas portas para desenvolver modelos para entender o mercado de maneira precisa, possibilitando o uso de métodos científicos para modelar esses fenômenos. Data a complexidade na composição desses preços, as abordagens estatísticas tradicionais podem muitas das vezes apresentarem limitações por conta do alto volume de variáveis e relações não lineares presentes no mercado, desde variáveis de preços à localização dos imóveis. Nesse caso, levanta-se a ideia do uso de Aprendizado de Máquina (Machine Learning), sendo ele um campo da Inteligência Artificial no qual permite que os computadores aprendem padrões a partir de séries temporais, um conjunto de dados registrados sequencialmente em intervalos regulares de tempo, possibilitando entender os padrões do mercado. Como demonstra Helena (2023), a injeção dessas tecnologias levanta a oportunidade de transformar as bases de dados heterogêneas em informações com significado, precificando de forma assertiva e identificando as tendências que métodos convencionais poderiam negligenciar. Portanto, a automatização da coleta conjunta do processamento faz com que o caminho da compreensão do mercado seja mais preciso. Considerando o cenário exposto, o objetivo deste trabalho é estudar a precificação de propriedades, na cidade de Jacareí, e definir uma metodologia consistente para realizar a modelagem de preços que permita prever os preços ao longo dos anos. Para operacionalizar essa proposta, o trabalho foi desenhado conforme o pipeline ilustrado na Figura 1, a sistematização do processo possui três fases fundamentais, a ingestão automatizada via Web Scraping, uma atividade que consiste na automatização por bots que coletam dados específicos de um certo campo, o tratamento com engenharia de atributos e, por fim, a análise da performance do modelo preditivo por meio de métricas de validação.

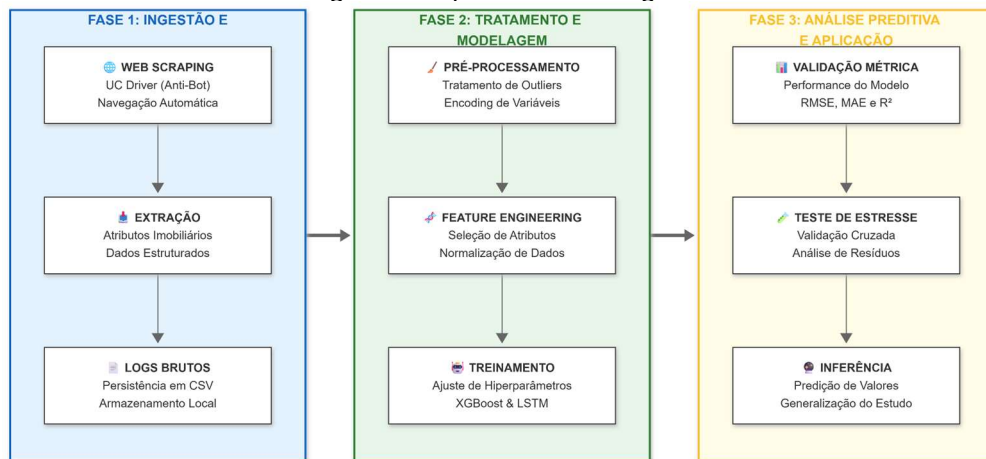
---

<sup>1</sup>Estudante do curso de Análise e Desenvolvimento de Sistemas, IFSP Campus Jacareí – Jacareí/SP. machado.moraes@aluno.ifsp.edu.br. ORCID: <https://orcid.org/0009-0002-9702-3163>.

<sup>2</sup>Doutor em Geofísica Espacial pelo INPE. Professor do IFSP – Jacareí/SP. stekel@ifsp.edu.br. ORCID: <https://orcid.org/0000-0002-3135-3656>.

**XII CONGRESSO DE INICIAÇÃO CIENTÍFICA DO IFSP ITAPETININGA**  
Itapetininga, 19, 20 e 21 de maio de 2026  
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo  
Campus Itapetininga

Figura 1 - Pipeline Metodológico.



Fonte: A autoria própria, 2026.

**Objetivo**

Este trabalho tem como objetivo principal investigar e implementar um sistema analítico automatizado para a modelagem de dados do mercado imobiliário de Jacareí, utilizando técnicas de Aprendizado de Máquina para a identificação de padrões de preços. A pesquisa visa mapear e coletar dados públicos via Web Scraping, garantindo assim a integridade das informações, levando em consideração a recorrência para captação de tais dados, para gerar uma série temporal válida ao longo do tempo e, garantindo a integridade das informações por meio de limpeza e engenharia de dados. O foco central é a validação por meio da acurácia das métricas do sistema.

**Metodologia**

A metodologia desta pesquisa fundamenta-se em uma abordagem quantitativa e computacional que utiliza um fluxo de engenharia de dados para prever preços imobiliários em Jacareí, São Paulo, cidade escolhida por sua relevância logística e expansão industrial. O processo inicia-se com a coleta automatizada de dados no portal Zap Imóveis, utilizando ferramentas que simulam o comportamento humano para contornar bloqueios digitais e garantir a captura de informações essenciais, como metragem e valor, mesmo em páginas de carregamento dinâmico. Após a coleta, os dados passam por uma limpeza rigorosa com o uso de expressões regulares para extrair valores numéricos de textos complexos, além de serem enriquecidos com indicadores econômicos externos, como o índice oficial de preços de venda de imóveis e taxas de inflação obtidas via sistemas de interface de programação de aplicações (API - que são pontes de comunicação entre diferentes softwares) de órgãos como o Banco Central. A análise estatística resultante prepara o terreno para a fase de predição, que abandona modelos matemáticos tradicionais em favor de Redes Neurais de Memória de Longo Prazo (LSTM - uma arquitetura de inteligência artificial capaz de aprender padrões em sequências de tempo), escolhidas por sua habilidade superior em entender ciclos econômicos e variações históricas de oferta. Para garantir a precisão, o sistema é testado com métricas que avaliam o erro médio das previsões, como a Raiz do Erro Quadrático Médio (RMSE - uma medida que dá maior peso a erros grandes para evitar desvios graves) e o Erro Médio Absoluto (MAE - que indica a

## XII CONGRESSO DE INICIAÇÃO CIENTÍFICA DO IFSP ITAPETININGA

Itapetininga, 19, 20 e 21 de maio de 2026

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo

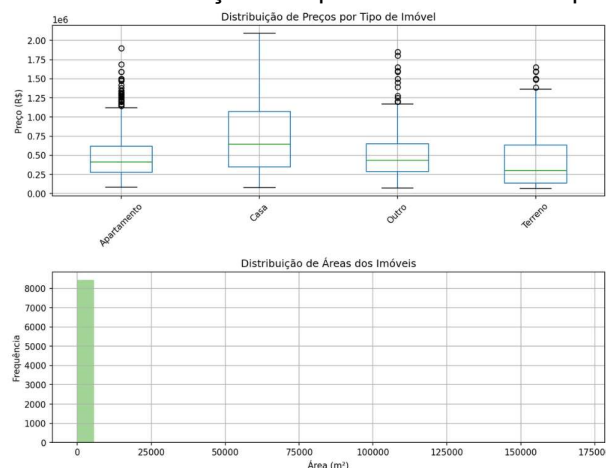
Câmpus Itapetininga

diferença média simples entre o valor real e o previsto), resultando em uma ferramenta robusta para reduzir a incerteza no mercado imobiliário.

### Resultados

A aplicação do pipeline de engenharia de dados viabilizou a consolidação de uma base estruturada do mercado imobiliário de Jacareí. Observa-se a predominância do segmento residencial, com destaque para apartamentos (44,6%) e casas (34,0%), seguidos por terrenos (11,6%) e imóveis comerciais (6,1%). Quanto ao comportamento estatístico das variáveis, as distribuições de preço de venda e área útil apresentam uma forte assimetria positiva, conforme ilustrado na Figura 2. Os histogramas e gráficos de caixa (boxplots) evidenciam que a maior concentração do mercado reside em faixas de valores e metragens menores, porém com a presença de outliers significativos que representam propriedades de alto padrão. Essa diferença reforça a necessidade de tratamentos de escala e normalização antes da fase de modelagem, já que isso garante que valores extremos não fiquem enviesados no aprendizado do algoritmo. A relação entre o valor do imóvel e sua dimensão física foi analisada por meio do diagrama de dispersão. Embora exista uma correlação positiva clara, nota-se uma elevada heterocedasticidade, onde a dispersão dos preços aumenta consideravelmente em imóveis de maior metragem. Esse padrão revela que o preço não é dado de forma linear pela área, já que pode ser influenciado por externalidades qualitativas e locacionais, conforme postulado por Arraes e Sousa Filho (2008). Por fim, a matriz de correlação exibida na Figura 3 traz revelações fundamentais sobre a interação do mercado com a economia macro. Enquanto os atributos intrínsecos (como número de banheiros e vagas) possuem correlações fortes com o preço, os indicadores macroeconômicos como Selic, IPCA e Dólar apresentam coeficientes próximos a zero no recorte estático analisado. Este resultado demonstra que o impacto da macroeconomia nos preços imobiliários não é imediato, apresentando efeitos defasados no tempo. Tal característica justifica a superioridade das Redes Neurais LSTM, defendida por Brito, Felzemburgh e Jardim (2022), uma vez que essa arquitetura é capaz de reter memória de longo prazo e capturar essas dependências temporais complexas que modelos lineares simples negligenciam.

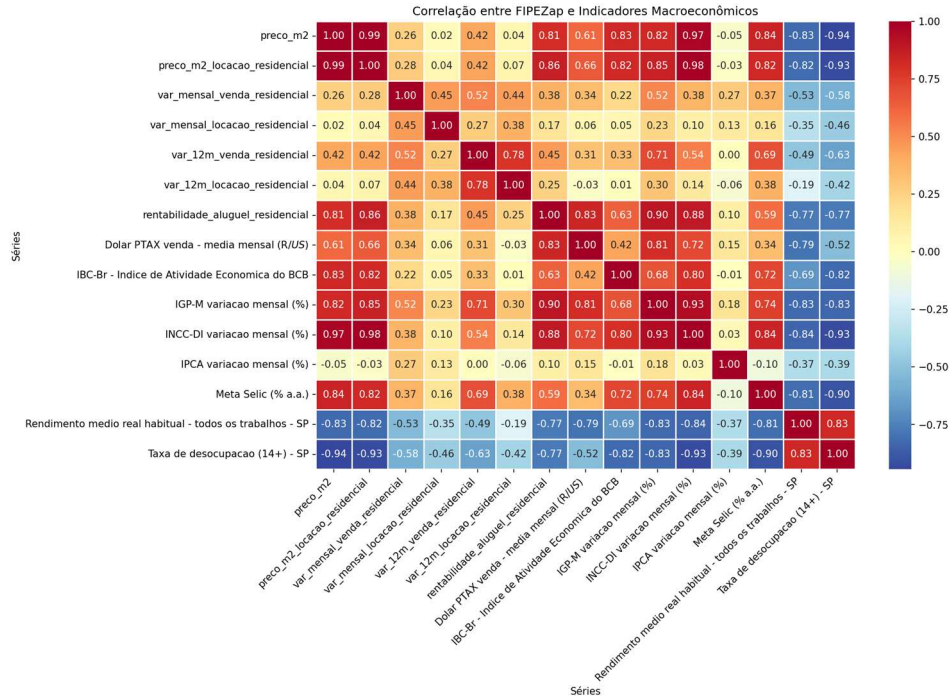
Figura 2 - Análise de distribuição e dispersão estatística de preços e áreas.



Fonte: Autoria própria, 2026.

**XII CONGRESSO DE INICIAÇÃO CIENTÍFICA DO IFSP ITAPETININGA**  
Itapetininga, 19, 20 e 21 de maio de 2026  
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo  
*Campus Itapetininga*

Figura 3 - Matriz de correlação entre atributos do imóvel e indicadores macroeconômicos.



Fonte: Autoria própria, 2026.

**Conclusão**

Destarte, a implementação do pipeline de dados para o mercado de Jacareí demonstra que a automação via Web Scraping e o tratamento de dados são etapas fundamentais para a análise imobiliária. Dito isso, os resultados exploratórios evidenciaram uma elevada heterocedasticidade e uma correlação não linear entre os preços e os indicadores macroeconômicos no curto prazo. Os dados obtidos demonstram que a hipótese de que modelos estatísticos tradicionais não são suficientes para apreender a complexidade do mercado local, justificando tecnicamente a adoção de redes neurais LSTM para lidar com os efeitos defasados da economia. Como passos futuros, o projeto irá focar no refinamento destes modelos preditivos e na expansão da base de dados para a Região Metropolitana do Vale do Paraíba, visando reduzir a assimetria de informação e fornecer estimativas de preços mais precisas e científicas.

**Referências**

ARRAES, Ronaldo A.; SOUSA FILHO, Edmar de. Externalidades e formação de preços no mercado imobiliário urbano brasileiro: um estudo de caso. *Economia Aplicada*, São Paulo, v. 12, n. 2, p. 289-319, abr./jun. 2008.

BRITO, Bruno Leão de; FELZEMBURGH, Maurício; JARDIM, Rei Carlos. Previsão de preços de imóveis utilizando redes neurais recorrentes do tipo LSTM. [S. l.: s. n.], 2022.

HELENA, Flávio de Falcão e. Oportunidades no mercado imobiliário com aplicação de modelos de aprendizagem de máquina: um estudo de caso em São Paulo, Brasil. 2023. 100 f. Dissertação (Mestrado em Ciências) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2023.

KISHI, Ricardo Yukihiro; STEKEL, Tardelli Coelho. A precificação de imóveis e a identificação de tendências no mercado imobiliário. [S. l.: s. n.], 2024.