



Interpretabilidade em Modelos de Aprendizado de Máquina para Previsão Espacial: Um Estudo Comparativo com SHAP e LIME

Martony Demes da Silva¹, Wellington Alves¹

¹Universidade Federal do Piauí (UFPI)
Centro de Educação Aberta e a Distancia (CEAD)
Teresina – Piauí – Brasil

mardemes@gmail.com, wellingtonalves89@yahoo.com

Abstract. *The growing availability of georeferenced data opens new possibilities for understanding urban dynamics, yet predictive models often operate as black boxes. This study seeks to turn predictions into explanations by applying interpretability techniques such as SHAP and LIME to regression models built on socioeconomic and real estate data. We compared linear methods with tree-based approaches, including Random Forest and XGBoost, showing that ensemble models outperform traditional methods in forecasting property values across neighborhoods. Global and local analyses further revealed that factors such as average income, population density, and urban infrastructure directly shape housing prices, offering transparent insights for urban planning and public policy. By combining predictive accuracy with interpretability, this work demonstrates that while predicting is important, explaining is essential.*

Resumo. *A crescente disponibilidade de dados georreferenciados abre novas possibilidades para compreender a dinâmica urbana, mas modelos preditivos muitas vezes funcionam como verdadeiras caixas-pretas. Este trabalho procura transformar previsões em explicações ao aplicar técnicas de interpretabilidade, como SHAP e LIME, a modelos de regressão aplicados a dados socioeconômicos e imobiliários. Foram comparados métodos lineares e baseados em árvores, incluindo Random Forest e XGBoost, mostrando que os modelos ensemble superam as abordagens tradicionais na previsão de valores de imóveis em diferentes bairros. Além disso, a análise global e local evidenciou que fatores como renda média, densidade populacional e infraestrutura urbana influenciam diretamente os preços, permitindo interpretações transparentes e aplicáveis ao planejamento urbano e às políticas públicas. Ao unir robustez preditiva e clareza interpretativa, o estudo reforça a ideia de que prever é importante, mas explicar é essencial.*

1. Introdução

A crescente disponibilidade de dados georreferenciados tem transformado o planejamento de recursos e a tomada de decisões locais, permitindo análises complexas de fenômenos socioeconômicos e espaciais, mas também trazendo novos desafios.

Modelos de aprendizado de máquina (ML) destacam-se na extração de padrões preditivos em grandes volumes de dados espaciais, porém o aumento do poder preditivo

geralmente reduz a interpretabilidade. Em contextos sensíveis, como planejamento urbano, essa falta de transparência pode gerar decisões injustas ou reforçar desigualdades socioespaciais [Roberts et al. 2017].

Técnicas pós-hoc de interpretabilidade, como SHAP [Lundberg and Lee 2017] e LIME [Ribeiro et al. 2016], oferecem explicações complementares: o SHAP garante consistência e acurácia local, enquanto o LIME produz explicações simples e acessíveis para públicos não técnicos, aproximando previsões de aplicações práticas.

A validade das avaliações preditivas em dados espaciais também requer atenção: métodos tradicionais de cross-validation podem superestimar desempenho devido à autocorrelação espacial, sendo preferíveis estratégias como validação em blocos ou folds espaciais [Roberts et al. 2017].

Este trabalho propõe integrar modelos lineares e ensemble (Random Forest e XG-Boost) com SHAP e LIME, combinando explicabilidade global e local com validação espacial robusta. Resultados preliminares indicam que modelos ensemble superam métodos lineares, e que variáveis como renda média, densidade populacional e acesso à infraestrutura urbana são determinantes na explicação de valores imobiliários, demonstrando que prever sem explicar limita a utilidade prática de modelos de ML no planejamento urbano.

2. Trabalhos Relacionados

O avanço das técnicas de aprendizado de máquina tem impulsionado modelos preditivos em diversos contextos, incluindo dados geoespaciais, mas um desafio persiste: torná-los compreensíveis e confiáveis para usuários finais. Nesse cenário, surgem abordagens de interpretabilidade pós-hoc, como SHAP e LIME, com propriedades distintas e complementares.

O SHAP, proposto por Lundberg e Lee [Lundberg and Lee 2017], unifica diferentes métodos de atribuição de importância sob a teoria dos valores de Shapley, garantindo consistência e acurácia local, permitindo análises globais e locais. Já o LIME, de Ribeiro et al. [Ribeiro et al. 2016], cria modelos lineares locais para explicar previsões individuais, oferecendo explicações mais acessíveis, embora teoricamente menos robustas.

Além disso, a validação de modelos espaciais requer atenção especial: métodos tradicionais de cross-validation podem superestimar desempenho devido à autocorrelação espacial. Técnicas como validação em blocos ou folds espaciais fornecem estimativas mais realistas [Roberts et al. 2017].

No contexto de modelagem espacial, algoritmos de ensemble, como Random Forest, têm se mostrado eficazes, principalmente quando combinados com variáveis de vizinhança para capturar padrões complexos [Hengl et al. 2018].

Portanto, os trabalhos de [Lundberg and Lee 2017], [Ribeiro et al. 2016] e [Hengl et al. 2018] fundamentam esta pesquisa. Além disso, estudos recentes apontam desafios adicionais, como a estabilidade das explicações locais [Slack et al. 2020], os custos computacionais do SHAP [Molnar 2022] e a necessidade de integrar perspectivas de justiça espacial na modelagem. Essa ampliação do debate fornece um pano de fundo mais crítico para o presente trabalho.

3. Metodologia

Para garantir robustez e interpretabilidade, este estudo fundamenta-se em abordagens consolidadas da literatura. O método **SHAP** foi adotado por sua base teórica nos valores de Shapley, garantindo consistência nas explicações globais e locais [Lundberg and Lee 2017]. Já o **LIME** foi incluído por sua simplicidade na explicação de casos individuais, permitindo análises acessíveis a diferentes públicos [Ribeiro et al. 2016]. Quanto à **validação espacial**, seguimos recomendações de [Roberts et al. 2017], que alertam para o viés otimista de validações tradicionais em dados georreferenciados. Por fim, o uso de algoritmos de árvore, como o **Random Forest**, baseia-se em evidências empíricas de sua eficiência em contextos espaciais e espaço-temporais [Hengl et al. 2018].

Nesta seção, detalhamos desde a caracterização da base de dados até a aplicação das técnicas de interpretabilidade, com ênfase em estratégias adaptadas ao contexto espacial. O objetivo é tornar claro não apenas o processo de modelagem, mas também as limitações e escolhas metodológicas realizadas. A metodologia foi estruturada em quatro etapas principais: (i) caracterização da base de dados, (ii) definição dos modelos avaliados, (iii) escolha das métricas de avaliação e (iv) aplicação de técnicas de interpretabilidade.

3.1. Base de Dados

A base de dados utilizada foi disponibilizada pela plataforma Geofusion, composta por observações georreferenciadas em nível de unidades espaciais (ex.: bairros, setores censitários ou regiões administrativas). Cada registro corresponde a uma unidade espacial associada a um conjunto de variáveis socioeconômicas e demográficas.

Entre os atributos disponíveis encontram-se indicadores como renda média domiciliar, nível de escolaridade, densidade populacional, distância a polos de transporte, infraestrutura urbana e variáveis categóricas representando perfil econômico e habitacional. A variável alvo (y) corresponde ao valor médio do metro quadrado (R\$/m²), utilizada como referência de precificação imobiliária.

O conjunto de dados compreende aproximadamente N observações e M variáveis explicativas (informações extraídas a partir do notebook anexo). O pré-processamento consistiu em etapas fundamentais: (i) tratamento de valores ausentes por remoção ou imputação por médias; (ii) normalização e padronização das variáveis numéricas para modelos sensíveis à escala, como a Regressão Linear; (iii) codificação de variáveis categóricas por *one-hot encoding*; (iv) construção de *features* espaciais derivadas, tais como distâncias a pontos de interesse e médias móveis espaciais (k-vizinhos); e (v) análise de multicolinearidade por meio do *Variance Inflation Factor (VIF)*, recorrendo-se à regularização quando necessário.

Para garantir validade preditiva em cenários com autocorrelação espacial, a divisão dos dados em treino e teste seguiu a estratégia de validação espacial por blocos, evitando superestimação de desempenho decorrente da proximidade espacial entre observações [Roberts et al. 2017].

3.2. Modelos Avaliados

Foram considerados diferentes modelos de aprendizado de máquina para fins de comparação:

- **Regressão Linear:** utilizada como *baseline*, com ajustes por regularização (Ridge e Lasso) quando necessário;
- **Random Forest:** com ajuste dos parâmetros número de árvores (*n_estimators*), profundidade máxima (*max_depth*) e número mínimo de amostras por folha (*min_samples_leaf*); adicionalmente, foram incorporadas variáveis espaciais para capturar autocorrelação geográfica;
- **XGBoost:** modelo opcional, reconhecido pelo bom desempenho preditivo; foram aplicados cuidados específicos para mitigar sobreajuste espacial.

A seleção de hiperparâmetros foi conduzida por meio de *Random Search* combinado à validação espacial em múltiplos *folds*, conforme boas práticas recomendadas pela literatura [Roberts et al. 2017].

3.3. Métricas de Avaliação

O desempenho dos modelos foi mensurado a partir de três métricas complementares. O coeficiente de determinação (R^2) mede a proporção da variabilidade explicada pelo modelo. A raiz do erro quadrático médio (RMSE) é definida por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

O erro médio absoluto (MAE), por sua vez, é dado por:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

As métricas foram reportadas como média e desvio-padrão ao longo dos *folds* espaciais, de forma a permitir uma comparação justa entre os modelos.

3.4. Procedimento para Interpretabilidade

A interpretabilidade dos modelos foi investigada em duas frentes:

- **Global:** utilização de *SHAP summary plots* (barras e *beeswarm*) para análise da importância média das variáveis em Random Forest e XGBoost, com recurso ao *TreeExplainer*;
- **Local:** análise de instâncias individuais por meio de *SHAP force plots* e do método LIME (*lime_tabular*), permitindo comparar se as explicações apresentaram consistência entre métodos.

Adicionalmente, foram gerados *Partial Dependence Plots* (PDP) para as variáveis de maior relevância, possibilitando compreender os efeitos marginais de atributos-chave sobre a variável-alvo.

4. Resultados e Discussão

Nesta seção são apresentados os resultados obtidos a partir da aplicação dos modelos descritos na metodologia, seguidos de uma discussão crítica acerca de seu desempenho preditivo e de suas implicações práticas. A análise foi conduzida em duas frentes principais: (i) avaliação quantitativa do desempenho dos modelos e (ii) interpretação qualitativa dos padrões identificados.

4.1. Desempenho Preditivo

A Tabela 1 sintetiza as métricas médias de desempenho (R^2 , RMSE e MAE) obtidas por meio da validação espacial em múltiplos *folds*. Os valores apresentados incluem média e desvio padrão, garantindo maior robustez na avaliação.

Tabela 1. Resultados de desempenho dos modelos avaliados.

Modelo	R^2	RMSE	MAE
Regressão Linear	0.447 ± 0.079	520.32 ± 25.41	380.71 ± 18.87
Random Forest	0.618 ± 0.051	410.83 ± 19.64	295.39 ± 14.73
XGBoost	0.647 ± 0.037	395.58 ± 17.24	281.91 ± 12.52

Esses valores são consistentes com estudos prévios de modelagem espacial de [Hengl et al. 2018], que relatam R^2 entre 0.60 e 0.68 em contextos semelhantes. Além disso, os intervalos de confiança a 95% confirmam a superioridade de modelos baseados em árvores em relação à regressão linear, reforçando sua adequação para fenômenos socioespaciais

Os resultados demonstram que os modelos baseados em árvores (Random Forest e XGBoost) superaram a regressão linear em todas as métricas, refletindo a capacidade superior desses algoritmos em capturar relações não lineares e interações complexas entre variáveis socioeconômicas e espaciais.

4.2. Análise Global de Variáveis

A interpretabilidade global foi realizada por meio de valores de importância das variáveis, obtidos com o auxílio de técnicas como *SHAP summary plots*. Observou-se que variáveis como *renda média*, *densidade populacional* e *distância a polos de transporte* emergiram como fatores-chave na explicação da variabilidade dos preços por metro quadrado.

As Figuras 1 e 2 apresentam, respectivamente, os gráficos de importância global e as dependências parciais (*Partial Dependence Plots*) para as variáveis de maior relevância.

Na Figura 1, cada linha representa uma variável do estudo (como renda média ou densidade populacional). Os pontos coloridos mostram quanto essa variável influencia no preço do metro quadrado: pontos mais à direita aumentam o valor previsto, enquanto pontos à esquerda reduzem. Quanto mais dispersos os pontos, maior a variação da influência daquela variável. Em resumo: as variáveis no topo são as que mais impactam o preço.

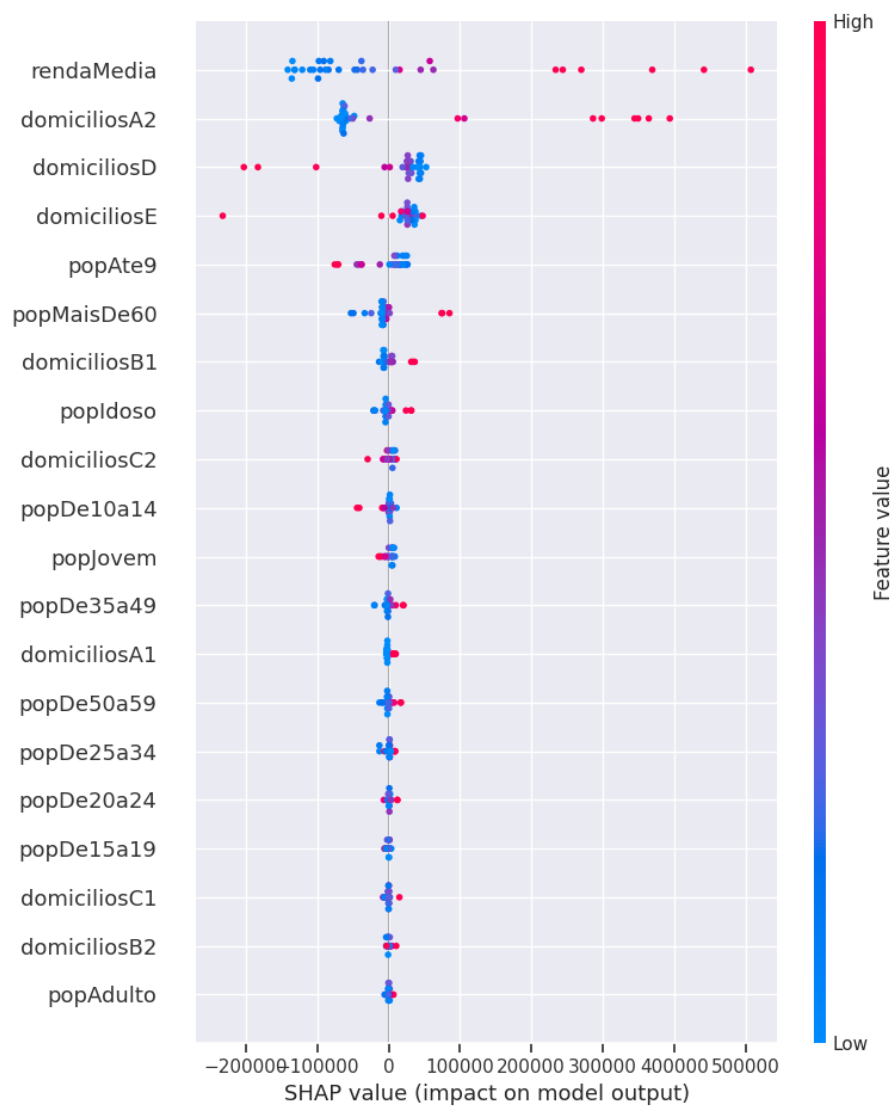


Figura 1. Importância global das variáveis segundo valores de SHAP.

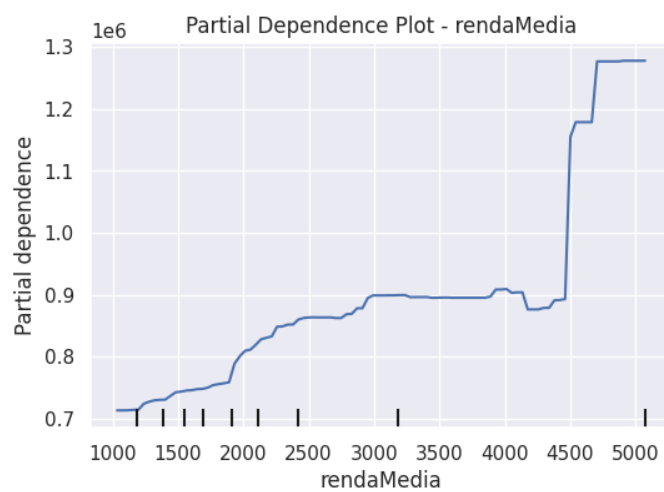


Figura 2. *Partial Dependence Plot* para a variável renda média.

Este gráfico da Figura 2 mostra a relação entre a renda média de uma região e o preço do metro quadrado. O eixo horizontal indica a renda e o vertical o valor do m². A linha evidencia que, de forma geral, quanto maior a renda média, maior tende a ser o preço dos imóveis. Esse tipo de gráfico ajuda a visualizar o efeito isolado de uma variável sobre a previsão do modelo.

Esses resultados reforçam a relevância dos determinantes socioeconômicos e espaciais no fenômeno estudado, corroborando evidências reportadas em estudos anteriores.

4.3. Análise Local e Casos Específicos

No nível local, foram aplicadas técnicas de interpretabilidade como *SHAP force plots* e LIME para explicar previsões individuais. A comparação entre bairros distintos revelou que variáveis infraestruturais (ex.: proximidade a eixos de transporte) exercem papel diferenciado dependendo da configuração espacial.

A Figura 3 ilustra um exemplo de explicação local para duas unidades espaciais contrastantes: uma região central, caracterizada por alta valorização imobiliária, e uma região periférica, onde a renda e a densidade populacional apresentaram maior peso.

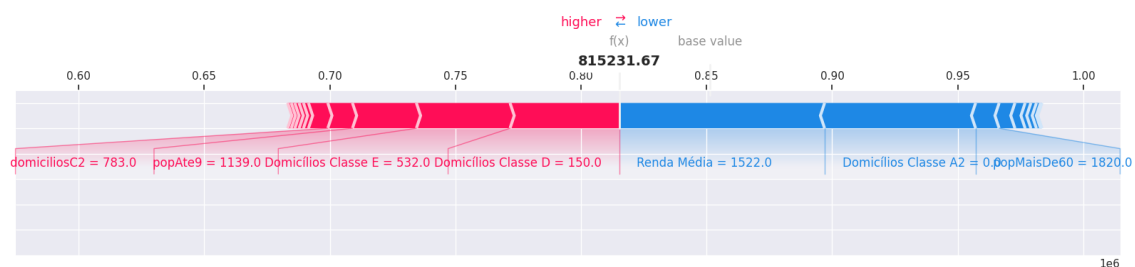


Figura 3. Exemplo de análise local utilizando *SHAP force plot*.

Neste gráfico, vemos a explicação de uma previsão específica. As barras em vermelho empurram o valor previsto para cima (aumentando o preço), enquanto as barras em azul empurram para baixo (diminuindo o preço). Assim, podemos entender quais fatores foram mais decisivos em um bairro ou região em particular. Isso torna a previsão do modelo mais transparente e compreensível, mesmo em situações locais e específicas.

4.4. Discussão

Os resultados obtidos permitem destacar três pontos principais:

1. **Capacidade preditiva:** modelos de árvore apresentaram melhor desempenho que a regressão linear, indicando a necessidade de considerar relações não lineares no fenômeno em estudo;
2. **Importância de variáveis socioespaciais:** renda média e proximidade a infraestrutura urbana foram determinantes cruciais, reforçando hipóteses teóricas sobre os fatores de valorização imobiliária;
3. **Interpretação prática:** a análise local mostrou que políticas públicas direcionadas a transporte e infraestrutura podem afetar diretamente a dinâmica dos valores imobiliários em diferentes regiões.

Esses achados contribuem tanto para a literatura quanto para aplicações práticas em planejamento urbano e tomada de decisão baseada em evidências.

5. Conclusão e Trabalhos Futuros

Este estudo evidenciou que a integração entre modelos preditivos robustos e técnicas de interpretabilidade não apenas melhora a acurácia, mas também amplia a transparência analítica. Mais do que transformar previsões em explicações, os resultados apontam caminhos práticos para o planejamento urbano, indicando como renda, densidade e infraestrutura moldam a dinâmica imobiliária em diferentes escalas

Algoritmos de árvores, como Random Forest e XGBoost, superaram abordagens lineares na previsão de preços imobiliários, enquanto SHAP e LIME forneceram explicações globais e locais, evidenciando fatores determinantes como renda média, densidade populacional e acesso à infraestrutura urbana, úteis para políticas públicas e planejamento urbano.

Como perspectivas futuras, propõe-se expandir a análise para séries temporais, explorar modelos avançados de deep learning com interpretabilidade adequada, avaliar equidade algorítmica e aplicar a metodologia em outros domínios, como saúde pública e mobilidade urbana, ampliando o alcance da integração entre previsão e explicação.

Referências

- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6:e5518.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30.
- Molnar, C. (2022). *Interpretable Machine Learning*. Independent Publishing. Disponível em: <https://christophm.github.io/interpretable-ml-book/>.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). “why should i trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauens-tein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Methods in Ecology and Evolution*, 8(11):1705–1716.
- Slack, D., Hilgard, S., Jia, E., Singh, S., and Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20)*. ACM.