

FELIPE MONTEIRO DE SÁ BARRETO

**MODELOS HEDÔNICOS ESPACIAIS E REDES NEURAIS ARTIFICIAIS NA
AVALIAÇÃO IMOBILIÁRIA EM MASSA: UMA ANÁLISE COMPARATIVA**

Monografia apresentada ao Departamento de
Economia como requisito parcial para a
obtenção do título de Bacharel em Ciências
Econômicas.

Orientador: Prof. Dr. Danilo Camargo Igliori

SÃO PAULO/SP

AGRADECIMENTOS

Primeiramente, agradeço a Deus e a Jesus Cristo, meu Senhor e Salvador.

Agradeço também aos meus pais e a minha irmã, que me deram suporte e forças durante toda essa jornada.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	V
RESUMO.....	VI
1 INTRODUÇÃO	8
1.1 JUSTIFICATIVA	8
1.2 OBJETIVO.....	9
2 REVISÃO DE LITERATURA	11
3 MATERIAL E MÉTODO	17
3.1 MODELOS	17
3.2 ANÁLISE ESTATÍSTICA	21
4 RESULTADOS.....	24
5 DISCUSSÃO	28
6 CONCLUSÕES	30
REFERÊNCIAS.....	31

LISTA DE ILUSTRAÇÕES

TABELA 2.1	ARTIGOS COMPARATIVOS	12
FIGURA 3.1.1	REPRESENTAÇÃO DA TOPOLOGIA DAS OBSERVAÇÕES	18
FIGURA 3.1.2	MODELO DE REDE NEURAL	20
FIGURA 3.2.1	NÚMERO DE OBSERVAÇÕES POR CEP	21
TABELA 3.2.1	ANÁLISE DESCRITIVA	22
FIGURA 4.1	RESULTADO DA REGRESSÃO	24
TABELA 4.1	MEDIDAS DE PERFORMANCE	26

RESUMO

MODELOS HEDÔNICOS ESPACIAIS E REDES NEURAIS ARTIFICIAIS NA AVALIAÇÃO IMOBILIÁRIA EM MASSA: UMA ANÁLISE COMPARATIVA

Objetivo: O presente projeto visa contribuir com a discussão sobre a métodos de precificação de ativos imobiliários em massa, replicando a comparação entre o modelo hedônico espacial e o modelo de rede neural artificial MLP, Perceptron Multicamadas treinada por retropropagação. **Material e Método:** Dados referentes ao preço, área útil, vagas de carro, banheiros, quartos, suítes, condomínio mensal, de anúncios de 10.073 apartamentos residenciais na cidade de Campinas/SP em maio de 2022. Aplicação de modelo ANN MLP e GMM 2SLS. **Resultados:** Medidas de performance RMSE (GMM – 0,2817; ANN – 0,2864), MAE (GMM- 0,2193; ANN – 0,2225), RAE (GMM - 0,0200; ANN – 0,0204), MAPE (GMM – 0,0168; ANN – 0,0172), Median (GMM – 0,1797; ANN – 0,1848) e R^2 (GMM – 0,8246; ANN – 0,8121) **Conclusão:** O modelo hedônico espacial superou o modelo de redes neurais artificiais, apresentando melhor capacidade preditiva.

Descritores: Avaliação imobiliária em massa, Redes neurais artificiais, Modelos Hedônicos, Econometria Espacial

INTRODUÇÃO

1 INTRODUÇÃO

1.1 JUSTIFICATIVA

A precificação de ativos imobiliários é de interesse de todos os setores da economia e, portanto, é um campo amplamente estudado. É necessária para avaliar os custos e os benefícios da propriedade imobiliária. Sua avaliação é requerida, e muitas vezes realizada, por agentes imobiliários, peritos, fiscais, brokers, empreendimentos, investidores e gerentes de fundos, bancos, hipotecas, analistas de mercado. Na maioria dos países industrializados, o imóvel próprio é o maior componente do portfólio de uma família. (SELIM, 2009)

Como explicitado por Zurada, Levitan e Guan (2011), preços errôneos reduziram a confiança nos balanços das instituições financeiras. A necessidade de avaliações não viesadas, objetivas e sistemáticas, portanto, se mostrou essencial após a crise financeira de 2008. Muitos pesquisadores buscaram explicações sobre os mecanismos dessa bolha imobiliária, sendo que vários estudos foram conduzidos atentando-se para a formação dos valores da terra para entender os componentes temporais e geográficos de tamanha volatilidade. (Krause; BITTER, 2012)

A variação de preços dos imóveis é de interesse tanto de indivíduos quanto do governo, posto que influencia as condições socioeconômicas, a economia doméstica, a arrecadação do governo e as contas públicas. (SELIM, 2009)

O campo acadêmico de precificação de propriedades engloba diversas disciplinas, tais como economia, geografia, urbanismo, design, administração, finanças e estatísticas, assim como departamentos e escolas especializadas no mercado de imóveis. Portanto, avanços na área são sempre fluidos, com movimentos em diversas direções simultâneas, impulsionados ora pelas disciplinas mencionadas, ora pelo setor comercial subjacente. (KRAUSE; BITTER, 2012)

Inúmeros estudos foram realizados para avaliar a performance dos diversos modelos de precificação em massa. Apesar disso, é difícil obter conclusões significativas, visto que os trabalhos tiveram resultados divergentes, por variações no tamanho das amostras, nos números e tipos de características, no uso de parâmetros de avaliação e na aplicação de diferentes modelos. (ZURADA; LEVITAN; GUAN, 2011)

1.2 OBJETIVO

O objetivo deste trabalho é realizar uma análise comparativa entre dois modelos de precificação imobiliária em massa: um modelo de precificação hedônica espacial e um modelo baseado em rede neural artificial.

A hipótese básica do trabalho é de que modelos baseados em redes neurais artificiais (ANN – Artificial Neural Networks) têm melhor capacidade preditiva do que o modelo de precificação hedônica. Isso porque as ANN podem avaliar complexas relações não-lineares, sem as restrições paramétricas e suposições dos modelos hedônicos de regressão múltipla, como linearidade (MCCLUSKEY et al., 2012), identificação de formas funcionais, seleção de variáveis independentes, heterocedasticidade, outliers, dependências e descontinuidade. (SELIM, 2009)

2 REVISÃO DE LITERATURA

2 REVISÃO DE LITERATURA

Pagourtzi et al. (2003) agrupam os métodos de precificação, primeiramente, em duas categorias: tradicionais e avançados. Os métodos tradicionais dependem de algum tipo de comparação para avaliar o valor de mercado, seja diretamente ou através de várias observações, utilizando um modelo de regressão. Já os métodos avançados modelam a racionalidade dos participantes do mercado para estimar o ponto de troca.

Como exemplo de métodos, temos:

1. Métodos tradicionais:
 - a. Método comparativo
 - b. Método de investimento/renda
 - c. Método de lucro
 - d. Método residual
 - e. Método de custo
 - f. Método de regressão múltipla
 - g. Método de regressão stepwise
2. Métodos avançados:
 - a. Rede neural artificial
 - b. Precificação hedônica
 - c. Análise espacial
 - d. Lógica difusa
 - e. Modelo autorregressivo integrado de médias móveis (ARIMA)

Wang e Li (2019) realizaram uma revisão de literatura sistemática, encontrando 104 artigos entre 2000 e 2018.

O avanço da capacidade computacional e de armazenamento de informação nas últimas décadas fez surgir e ampliar o uso de técnicas como Big Data e Inteligência Artificial. Em conjunto com esse avanço, houve um aumento do interesse em métodos de precificação em massa, principalmente a partir de 2010, como demonstram Wang e Li (2019). Os autores encontraram três grandes tendências: modelos IA, modelos SIG (Sistema de informação geográfica) e modelos mistos.

Vários estudos foram realizados com a intenção de comparar o uso da Inteligência Artificial com métodos tradicionais de regressão e métodos de precificação hedônica. No entanto, Zurada, Levitan e Guan (2011) apontam que as conclusões destes trabalhos foram mistas, já que algumas demonstraram melhores resultados a partir de modelos IA, enquanto outros não observaram diferenças relevantes. Alguns poucos ainda apresentaram resultados negativos. Os autores ainda expõem que é difícil a comparação entre os resultados, por variações no tamanho das amostras, nos números e tipos de características e no uso de parâmetros de avaliação.

Tendo os resultados anteriores, Zurada, Levitan e Guan (2011) apresentam um artigo comparando quatro métodos baseados em regressão (tradicional, Support Vector Machines, regressão aditiva, árvores M5P) e três métodos baseados em Inteligência Artificial (redes neurais, redes de função de base radial, raciocínio baseado em memória), utilizando uma base de dados de 16.366 observações e 18 variáveis. Sua conclusão foi de que os métodos de regressão não tradicionais foram superiores aos demais com dados homogêneos, e que os métodos IA tiveram uma boa performance com dados mais heterogêneos em determinados cenários.

Selim (2009) realizou uma comparação direta entre o modelo hedônico e o modelo de redes neurais artificiais sobre um conjunto de dados de vendas de imóveis residenciais na Turquia, em que a ANN se mostrou uma melhor alternativa.

McCluskey et al. (2012) tabelaram os artigos comparativos anteriores por eles observados e seus resultados:

Tabela 2.1 - Artigos comparativos

Artigo	Número de propriedades	Técnica de modelagem alternativa	de Performance da acurácia da predição da ANN
Lin e Mohan (2011)	33.342	Regressão	Superior
Peterson e Flanagan (2009)	46.467	Regressão	Superior
Limsombunchai et al. (2004)	200	Regressão	Superior

McCluskey e Borst (1997)	1.495	Regressão	Não superior
Worzala et al. (1995)	217	Regressão	Não superior
Tay e Ho (1992)	833	Regressão	Superior
Do e Grudnitski (1992)	105	Regressão	Superior

Fonte: McCluskey, Davis, Haran, McCord, & McIlhatton (2012)

Entretanto, essas comparações foram feitas entre métodos que ignoram o espaço geográfico e sua influência. O uso da expressão “Location, location, location!” no mercado imobiliário norte-americano resume bem a necessidade de se considerar a localização nos modelos de precificação de imóveis. Contudo, incorporar essa característica não é de simples execução. Problemas como dependência espacial, heterogeneidade espacial, fenômenos anisotrópicos e efeitos de fronteira tornam os modelos monocêntricos da economia urbana (que consideram um declínio não-linear simples de acordo com a distância do local até a região comercial central). De fato, são complexas relações policentricas, geograficamente heterogêneas e com distintas regras de uso da terra que influenciam a formação de preços. (KRAUSE; BITTER, 2012)

Em 1970, Jean Paelinck cunhou o termo econometria espacial, ao se referir aos aspectos metodológicos que incorporavam dependências em modelos econométricos multirregionais em seções cruzadas. A partir de então, houve um aumento significativo de interesse pela análise espacial, impulsionada principalmente pelos avanços teóricos e tecnológicos (GIS – Geographic Information Systems), culminando na criação da “Spatial Econometrics Association” em 2006. (ANSELIN, 2007)

O desenvolvimento da área foi separado por Anselin (2007) em três períodos distintos: pré-condições para o crescimento, decolagem e estado estacionário. O primeiro período, entre as décadas de 70 e 80, é caracterizado por um interesse em autocorrelações espaciais residuais, especificação de modelos (modelos de lag espacial e de erro espacial), e estimações simples em modelos de regressão linear

espacial (modelos de máxima verossimilhança e variáveis instrumentais. Durante os anos 90, o campo alcançou a maturidade, através de provas formais rigorosas das propriedades dos estimadores e das estatísticas de teste, da introdução de novas abordagens (estatísticas LM, estimações GMM, técnicas Bayesianas), da consideração de dados em painel e de escolhas discretas, assim como o aumento do uso e da acessibilidade de ferramentas computacionais. Após os anos 2000, a econometria espacial alcançou certa estabilidade, com um grande número de artigos, tanto teóricos quanto práticos, sendo publicados nas maiores revistas científicas especializadas em econometria.

Métodos de econometria espacial dão especial atenção para a violação do modelo amostral através das similaridades causadas pela proximidade geográfica. Esse novo campo pode ser descrito por cinco características fundamentais: (ARBIA, 2006)

1. O papel da interdependência espacial;
2. A assimetria das relações espaciais;
3. A importância de fatores explicativos em outros espaços;
4. A diferenciação em interações *ex-ante* e *ex-post*;
5. A modelagem explícita do espaço;

Estudos iniciais por Goodman, de 1978, e Li e Brown, de 1980, examinaram a questão da interdependência espacial na escala intrabairro. Dublin, em 1988, e Can e Megbolugbe, em 1997, consideraram os impactos específicos de uma propriedade na dependência espacial no mercado. (KRAUSE; BITTER, 2012)

A especificação do modelo de lag espacial, que permite que os valores de observações geograficamente próximas afetem o valor da variável dependente, e do modelo de erro espacial, que lida com a autocorrelação com os termos de erro (violação de uma das suposições do modelo OLS), permitiu que análises mais elaboradas fossem realizadas.

Diversos estudos consideraram a diferença entre modelos OLS simples e modelos com dependência espacial. Brasington e Haurin (2006) notaram que o impacto de gastos públicos em educação no preço de imóveis residenciais na região era maior ao serem estimados através de modelos de lag espacial ao invés de modelos OLS tendo bairro como variáveis descritivas. Cohen e Coughlin (2008) observaram que os modelos tradicionais subestimavam o impacto do barulho

causado por aeroportos no preço dos imóveis próximos, ao comparar modelos de lag espacial, modelos de erro espacial e modelos OLS. Kuethe (2012) conclui que o resultado de uma regressão OLS mostra que uma variável que indica a diversidade de usos da terra tem efeitos positivo sobre o preço do imóvel e que quando se considera um modelo espacial esse impacto positivo não é observado. Isso mostra que o uso de modelos que omitem essa característica afetam negativamente os resultados obtidos. (KRAUSE; BITTER, 2012)

3 MATERIAL E MÉTODO

3 MATERIAL E MÉTODO

3.1 MODELOS

Preços hedônicos podem ser definidos como: “the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them¹” (ROSEN, 1974)

A forma funcional do modelo hedônico é semi-logarítmica, recomendada pela literatura, e que pode ser interpretada como a variação percentual do preço diretamente atribuída à variação da característica respectiva. (SELIM, 2009)

$$\ln P = \beta x + u$$

Onde P é o preço do imóvel, β é a matriz de coeficientes, x é a matriz das variáveis independentes e u é o vetor dos termos de erro.

Ao modelo hedônico tradicional adiciona-se o conceito de campo aleatório, que foi definido por Arbia (2006) como:

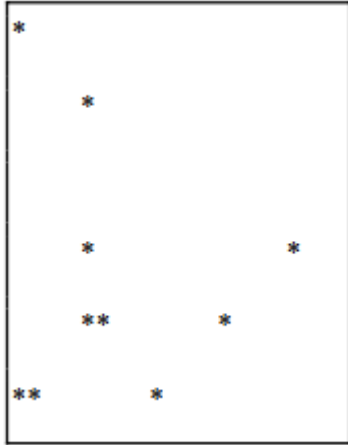
Definição 1. Seja $(\Omega, B, P(.))$ um espaço de probabilidade, com Ω representando o espaço amostral, B o conjunto de Borel associado e $P(.)$ a medida de probabilidade. Além disso, seja S um conjunto não vazio em R^2 , e defina-se a função $X(., s)$, com $X(., s)$ tal que: $\Omega \times S \rightarrow R \times S$. A sequência ordenada de variáveis aleatórias $\{X(., s), s \in S\} = \{X(s), s \in S\}$, indexadas em respeito a s , é chamada de *processo aleatório espacial* ou *campo aleatório*.

Duas características definem um campo aleatório: a topologia das observações e a estrutura de dependência entre elas.

¹ “Preços hedônicos são os preços implícitos dos atributos e são revelados aos agentes econômicos através dos preços observados de produtos diferenciados e das quantidades específicas de características associadas a eles.”

No caso em que a localização geográfica é representada por coordenadas geográficas, como latitude e longitude, a topologia das observações (ou a natureza do índice s) pode ser representada por coordenadas de pontos em R^2 .

FIGURA 3.1.1 REPRESENTAÇÃO DA TOPOLOGIA DAS OBSERVAÇÕES



FONTE: ARBIA, 2006

Como essas medidas são contínuas, essa topologia pode ser completamente especificada pela distância euclidiana:

$$d_{ij} = d(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{(\mathbf{s}_i - \mathbf{s}_j)^T (\mathbf{s}_i - \mathbf{s}_j)}$$

Com d_{ij} a distância entre o ponto $\mathbf{s}_i$ e o ponto $\mathbf{s}_j$. (Arbia, 2006)

A estrutura de dependência pode ser representada pela definição de vizinhança e de pesos:

Definição 2. Vizinho mais próximo (*Nearest Neighbor*). Dois pontos $\mathbf{s}_i$ e $\mathbf{s}_j$ são ditos vizinhos se $d_{ij} = \min(d_{ik}) \forall i, k$. (ARBIA, 2006)

Pode-se expandir essa definição para a escolha dos k -ésimos vizinhos mais próximos (K-NN), e definir $N(i)$ como o conjunto de todos os vizinhos de $\mathbf{s}_i$ e η_i como a cardinalidade desse conjunto.

A expressão para uma matriz de pesos generalizada W é:

$$w_{ij} = g[d_{ij}]$$

Com g sendo a função inversa da distância. (ARBIA, 2006)

Definição 3. Dado um campo aleatório $\{X(\mathbf{s}), \mathbf{s} \in \mathcal{S}\} = \{X(\mathbf{s}_1); X(\mathbf{s}_2); \dots, X(\mathbf{s}_n)\}$, e dada a topologia induzida pela definição de vizinhança $N(i)$, o valor de defasagem espacial da variável aleatória $X(\mathbf{s}_i)$ é definido como: (Arbia, 2006)

$$L[X(\mathbf{s}_i)] = \frac{1}{\eta_i} \sum_{\mathbf{s}_j \in N(i)} X(\mathbf{s}_j)$$

A base do modelo aplicado é o modelo SAR-SAR (*mixed regressive spatial autoregressive model with a spatial autoregressive error term*), que pode ser descrito como:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

Onde $\mathbf{y}$ um vetor $n \times 1$ de observações da variável dependente, $\mathbf{W}$ é uma matriz $n \times n$ de operador *lag* espacial e $\mathbf{W} \mathbf{y}$ como o termo de *lag* espacial com o parâmetro espacial autorregressivo ρ , $\mathbf{X}$ como uma matriz $n \times k$ de observações de variáveis independentes exógenas com o vetor $k \times 1$ de coeficientes $\boldsymbol{\beta}$, e um vetor $n \times 1$ de erros. (ANSELIN, 2011)

A estimação é feita através do método dos momentos generalizado (GMM), realizada em várias etapas, usando o *two stage least squares* (2SLS).

Usando nós para simular neurônios, redes neurais são modelos de inteligência artificial que tentam replicar o funcionamento de aprendizado do cérebro humano (PAGOURTZI et al., 2003). Um Perceptron Multicamadas consiste em três componentes primários:

1. Camada de entrada
2. Camadas ocultas
3. Camada de saída

A camada de entrada somente propaga os seus valores para as camadas ocultas. As camadas ocultas realizam dois processos: a combinação linear e a função de ativação.

A combinação linear é responsável pela soma da multiplicação dos pesos de cada entrada pelos seus respectivos valores:

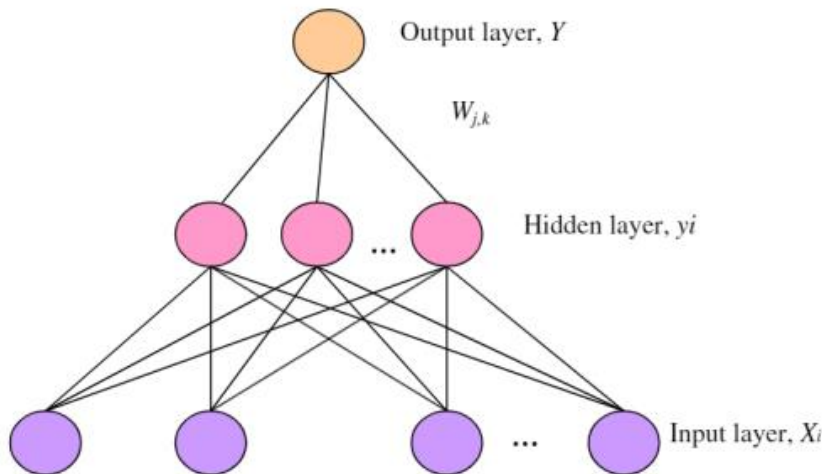
$$Y_j = \sum_j^n X_i W_{ij}$$

Onde X_i são os valores de entrada, W_{ij} são os pesos de cada entrada i em cada um dos nós j .

A função de ativação define a ativação de saída do nó em termos do seu nível de ativação interno. Essa ativação pode ser de várias formas, sendo a mais utilizada a função sigmoideal (PAGOURTZI et al., 2003):

$$Y_T = \frac{1}{1 + e^{-y}}$$

Figura 3.1.2 Modelo de rede neural



Fonte: (Selim, 2009)

Os pesos são ajustados através do método de backpropagation, pela iteração de cada nó j da camada s :

$$w_{i,j}^s(k+1) = w_{i,j}^s(k) - \alpha \frac{\partial F}{\partial w_{i,j}^s}$$

Utilizando o método do gradiente, que minimiza a função de perda:

$$F =_w \min \sum (t - a)^2$$

Onde t é o valor da variável dependente e a é o valor predito pela rede. (PETERSON; FLANAGAN, 2009)

Os modelos ANN e hedônicos foram implementados em Python, com uso das bibliotecas PySal e Scikit-learn.

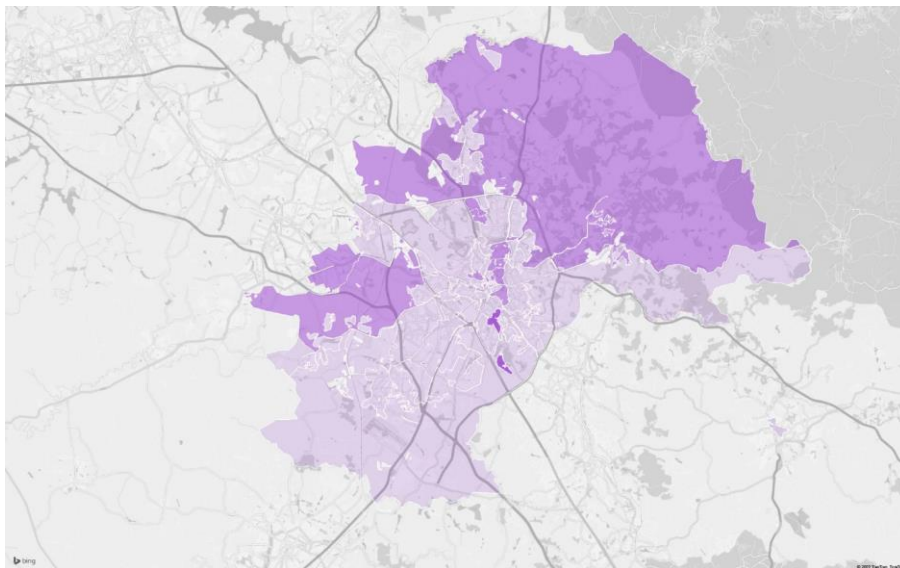
No modelo hedônico espacial, os vizinhos $N(i)$ foram selecionados com base nos 20 vizinhos mais próximos de cada observação (KNN(20)) e gerada a matriz de pesos W .

O modelo de ANN utilizado foi o *Multi-layer Perceptron regressor* com retropropagação de erros, sendo sua estrutura definida por uma camada oculta composta por 100 nós.

3.2 ANÁLISE ESTATÍSTICA

Os dados utilizados se referem à 10.073 apartamentos residenciais anunciados dentro da cidade de Campinas na plataforma online zapimoveis.com.br, ativos no dia 13 de maio de 2022, e que continham informações de localização (nome da rua e Código de Endereçamento Postal - CEP). A partir dos dados de localização, foram aproximados, por centro geométrico, a latitude e a longitude de cada imóvel, utilizando o serviço de *geocoding* do *Google Cloud*.

FIGURA 1.2.1 NÚMERO DE OBSERVAÇÕES POR CEP



Fonte: Elaborado pelo autor

A variável independente é o preço anunciado, sendo consideradas variáveis dependentes: área útil do imóvel, tamanho da garagem (número de carros), número de banheiros, número de quartos, número de suítes, valor mensal do condomínio, assim como os dados de latitude e longitude.

Tabela 3.2.1 Análise Descritiva

	Área útil	Quartos	Banheiros	Suítes	Vagas	Condomínio	Preço
Média	86,13	2,39	2,05	0,79	1,39	R\$ 654,09	R\$ 560.787,79
Desvio Padrão	58,60	0,78	1,20	0,91	0,85	R\$ 604,76	R\$ 588.743,93
Mín.	20	0	1	0	0	R\$ -	R\$ 90.000,00
25%	55	2	1	0	1	R\$ 351,00	R\$ 260.000,00
50%	70	2	2	1	1	R\$ 505,00	R\$ 390.000,00
75%	96	3	2	1	2	R\$ 750,00	R\$ 630.000,00
máx.	782	6	9	6	9	R\$ 8.800,00	R\$ 10.541.000,00

RESULTADOS

4 RESULTADOS

Figura 4.1 Resultado da regressão

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES

```
-----
Data set           :      unknown
Weights matrix     :      unknown
Dependent Variable :      ln(price)
Mean dependent var :      12.9650
S.D. dependent var :      0.6729
Pseudo R-squared   :      0.8657
Spatial Pseudo R-squared: 0.8247
Number of Observations:      10073
Number of Variables :      8
Degrees of Freedom  :      10065
-----
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	7.4416465	0.0882049	84.3677039	0.0000000
usableAreas	0.0020850	0.0000875	23.8311444	0.0000000
bedrooms	0.1056814	0.0042849	24.6637200	0.0000000
bathrooms	0.0907325	0.0045979	19.7333079	0.0000000
suites	0.1154025	0.0052269	22.0786345	0.0000000
parkingSpaces	0.1511444	0.0045769	33.0231990	0.0000000
monthlyCondoFee	-0.0000208	0.0000071	-2.9408438	0.0032732
W_ln(price)	0.3559413	0.0070826	50.2559325	0.0000000

```
-----
Instrumented: W_ln(price)
Instruments: W_bathrooms, W_bedrooms, W_monthlyCondoFee, W_parkingSpaces,
             W_suites, W_usableAreas
-----
```

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Anselin-Kelejian Test	1	1555.500	0.0000

Todas as variáveis independentes mostraram-se estatisticamente significantes no nível de 95% de significância, apesar da pequena contribuição negativa do valor do condomínio sobre o preço do imóvel. De maior interesse é o peso da influência dos instrumentos espaciais. Se os 20 imóveis mais próximos do apartamento focal dobrassem de preço, somente esse fato geraria um aumento do valor do apartamento em 35,59%.

O diagnóstico de Anselin-Kelejian confirma essa dependência.

Ambos os modelos utilizam as mesmas variáveis independentes - área útil do imóvel, tamanho da garagem, número de banheiros, número de quartos, número de suítes, valor mensal do condomínio, e latitude e longitude, tendo como variável dependente o logaritmo natural do preço.

Pela natureza fechada dos métodos de ANN, não é possível inferir a relação das variáveis, o seu impacto e a sua significância dentro do modelo.

Portanto, para comparar os resultados dos modelos, 6 medidas de performance foram utilizadas: Root Mean-squared error (RMSE), Mean Absolute Error (MAE), Root Mean-Squared Log Error (RAE), R^2 , Mean Absolute Percentage Error (MAPE) e Median Absolute Error.

Root Mean-squared Error é uma medida correspondente a raiz do valor esperado do erro quadrático. Formalmente:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2}$$

Mean Absolute Error é a medida correspondente ao valor esperado do módulo do erro. Formalmente:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i|$$

Root Mean-Squared Log Error corresponde a raiz do valor esperado do quadrado do logarítimo do erro. Formalmente:

$$RMSLE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (\ln(1 + y_i) - \ln(1 + \hat{y}_i))^2}$$

É importante notar que o RMSLE penaliza mais fortemente valores subestimados do que valores sobrestimados.

Mean Absolute Percentage Error é sensível a valores relativos de erros, sendo independente da escala da variável explicada:

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)} \right)$$

Onde ϵ é um valor positivo arbitrariamente pequeno, utilizado para evitar divisões por zero quando y_i tem esse valor.

Median Absolute Error é uma medida robusta a valores atípicas, calculada a partir da mediana de todos os módulos da diferença entre valores preditos e observados. Mais especificamente,

$$MedAE(y, \hat{y}) = \text{mediana}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$

O R^2 é o coeficiente de determinação, que representa o quanto da variância pode ser explicada pelas variáveis independentes do modelo.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Tabela 4.1 Medidas de performance

	GMM	ANN
RMSE	0,2817	0,2864
MAE	0,2193	0,2225
RAE	0,0200	0,0204
R^2	0,8246	0,8121
MAPE	0,0168	0,0172
Median	0,1797	0,1848

Os resultados das medidas RMSE, MAE, RAE, MAPE e Median indicam que os erros de previsão do modelo GMM foram sempre menores do que os erros do modelo de rede neural artificial, independentemente de como foram medidas as diferenças entre valores observados e valores preditos.

De mesmo modo, o R^2 dos modelos indicam que a variância do modelo GMM é mais bem explicada pelas variáveis independentes do que a variância do modelo ANN.

Contudo, apesar de ser possível ranquear os dois modelos a partir dos resultados obtidos, é importante notar que as medidas de performance de ambos se encontram bem próximas, com diferença nas medidas de desvio entre 0,0004 e 0,0051.

5 DISCUSSÃO

O modelo hedônico espacial (GMM 2SLS) superou o modelo de redes neurais artificiais MLP em todas as medidas apresentadas. O tamanho significativo da amostra (10.073 apartamentos) produz vários graus de liberdade para os testes estatísticos e para a variabilidade da amostra de treino e da amostra de teste do modelo de rede neural artificial, permitindo que os resultados sejam robustos.

Contudo, algumas ressalvas devem ser feitas. Primeiramente, o modelo hedônico é mais especializado, pois foi desenvolvido especificamente para lidar com relações de dependência espacial. Em oposição, o modelo MLP é uma ferramenta generalista. Outra questão que pode ter impactado o desempenho do modelo de redes neurais é a otimização da rede, que é um assunto complexo e que não foi tratado no presente trabalho. Portanto, sem que haja um grande trabalho de especificação, que também concerne o modelo GMM, e de otimização, o modelo de inteligência artificial mostra-se subótimo.

Nada obstante, os resultados mostram que o estudo de métodos econométricos espaciais, um campo que não é novo, mas que se encontra em fases iniciais se comparado os métodos de séries temporais, rendeu bons resultados preditivos, o que atesta o seu desenvolvimento e a sua importância na inclusão das similaridades provenientes da proximidade espacial.

Este estudo tinha como proposta contribuir com novos dados, localizados, para a discussão. A aplicação de modelos de Inteligência Artificial mais avançados e mais específicos poderia avançar a discussão, assim como a introdução da avaliação da otimização de tais métodos.

6 CONCLUSÕES

6 CONCLUSÕES

Considerando os resultados obtidos, pode-se concluir que o modelo hedônico espacial (GMM 2SLS) superou o modelo de redes neurais artificiais MLP em todas as medidas de performance apresentadas, conquanto os valores sejam próximos. Isso vai de encontro à hipótese inicialmente apresentada, ou seja, que modelos de inteligência artificial teriam melhor capacidade preditiva do que modelos hedônicos tradicionais.

REFERÊNCIAS

- ANSELIN, Luc. Spatial econometrics in RSUE: Retrospect and prospect. **Regional Science and Urban Economics**, v.37, n.4, p. 450-456, 2007
- ANSELIN, Luc. GMM estimation of spatial error autocorrelation with and without heteroskedasticity. **Note (GeoData Center, Arizona State University, 2011)**, 2011.
- ARBIA, Giuseppe. **Spatial econometrics: statistical foundations and applications to regional convergence**. Springer Science & Business Media, 2006.
- KAUKO, Tom. On current neural network applications involving spatial modelling of property prices. **Journal of housing and the built environment**, v.18, n. 2, p. 159-181, 2003.
- KRAUSE, Andy L.; BITTER, Christopher. Spatial econometrics, land values and sustainability: Trends in real estate valuation research. **Cities**, v. 29, p. S19-S25, 2012.
- MCCLUSKEY, William et al. The potential of artificial neural networks in mass appraisal: the case revisited. **Journal of Financial Management of Property and Construction**, 2012.
- PAGOURTZI, Elli et al. Real estate appraisal: a review of valuation methods. **Journal of Property Investment & Finance**, v.21, n. 4, p 383-401, 2003.
- PETERSON, Steven; FLANAGAN, Alber. Neural network hedonic pricing models in mass real estate appraisal. **Journal of real estate research**, v. 31, n. 2, p. 147-164, 2009.
- ROSEN, S. Hedonic prices and implicit markets: product differentiation in pure competition. **Journal of political economy**, v. 82, n. 1, p 34-55, 1974.
- SELIM, Hasan. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. **Expert systems with Applications**, v. 36, n. 2, p. 2843-2852, 2009.
- WANG, Daikun; LI, Victor Jing. Mass appraisal models of real estate in the 21st century: A systematic literature review. **Sustainability**, v. 11, n. 24, p. 7006, 2019.
- ZURADA, Jozef; LEVITAN, Alan; GUAN, Jian. A comparison of regression and artificial intelligence methods in a mass appraisal context. **Journal of Real Estate Research**, v. 33, n. 3, p. 349-388, 2011.