

**ANÁLISE COMPARATIVA DE MÉTODOS INFERÊNCIA ESTATÍSTICA
TRADICIONAL, INFERÊNCIA ESPACIAL E REDES NEURAIS
ARTIFICIAIS. ESTUDO DE CASO.**

AUTORES

Carlos Alberto Peruzzo Trivelloni

Maria Florencia Rodríguez

Luis Calderón

REALIZAÇÃO

RESUMO

Este trabalho apresenta uma análise comparativa entre os resultados obtidos por diferentes métodos: inferência estatística tradicional por mínimos quadrados, inferência espacial conjuntamente com geoestatística e Redes Neurais Artificiais (RNA). O estudo é aplicado sobre uma amostra de 380 imóveis de tipo apartamento, pertencentes a um bairro da cidade de Montevidéu, Uruguai. Os resultados de estimação obtidos pelos diferentes métodos mostram que as Redes Neurais Artificiais podem ter um desempenho superior aos métodos de inferência tradicional e espacial, conseguindo um ajuste mais próximo aos dados de mercado. Os coeficientes de determinação R^2 são calculados e comparados para todos os modelos e também é realizada uma análise detalhada dos resíduos de cada modelo, comparando-se a distribuição e grau dispersão dos mesmos. Em todos os casos estudados os modelos de RNA obtiveram melhores resultados que os modelos inferenciais.

Palavras-chave: Regressão Múltipla, Regressão Espacial, Geoestatística, Redes Neurais Artificiais.

RESUMEN

Este trabajo presenta un análisis comparativo entre los resultados obtenidos por diferentes métodos: inferencia estadística tradicional por mínimos cuadrados, inferencia espacial conjuntamente con geoestadística y Redes Neuronales Artificiales (RNA). El estudio es realizado sobre una muestra de 380 inmuebles de tipo apartamento, pertenecientes a un barrio de la ciudad de Montevideo, Uruguay. Los resultados de estimación obtenidos por los diferentes métodos muestran que las Redes Neuronales Artificiales pueden tener un desempeño superior a los métodos de inferencia tradicional y espacial, consiguiendo un ajuste mejor a los datos de mercado. Los coeficientes de determinación R^2 son calculados y comparados para todos los modelos y también es realizado un análisis detallado de los residuos de cada modelo, comparandose la distribución y grado de dispersión de los mismos. En todos los casos estudiados los modelos de RNA obtuvieron mejores resultados que los modelos inferenciales.

Palabras clave: Regresión Múltiple, Regresión Espacial, Geoestadística, Redes Neuronales Artificiales.

REALIZAÇÃO

1. INTRODUÇÃO

A avaliação de imóveis é uma área profissional em constante evolução. Na busca de fornecer estimativas confiáveis do valor das propriedades, diferentes métodos de avaliação tem sido propostos e recomendados. Nos últimos anos, tem havido um interesse crescente na comparação de diversos métodos, visando aprimorar a precisão e confiabilidade das estimativas de valor.

Entre os métodos estudados estão os métodos inferenciais tradicionais, como a regressão linear múltipla, métodos de análise espacial como a regressão espacial e a geoestatística, e também as redes neurais artificiais, entre outros.

Os métodos inferenciais tradicionais são amplamente utilizados na avaliação de imóveis devido à sua facilidade de cálculo e interpretabilidade. Eles se baseiam em modelos estatísticos clássicos, como regressão linear e análise de variância, para estimar o valor dos imóveis. No entanto, esses métodos podem ser limitados pela suposição de linearidade nas relações entre as variáveis e pela falta de consideração da dependência espacial dos valores dos imóveis.

Por outro lado, os métodos espaciais têm ganhado destaque ao levar em conta a influência da localização geográfica na avaliação de imóveis. Esses métodos utilizam técnicas de análise espacial, como os modelos de regressão espacial e a krigagem, para modelar a estrutura espacial do valor dos imóveis. Isso permite estimar padrões e tendências espaciais, resultando em estimativas mais precisas. No entanto, a escolha do modelo espacial adequado e a disponibilidade de dados georreferenciados podem ser desafios importantes.

Mais recentemente, as redes neurais artificiais têm despertado crescente interesse na sua aplicabilidade na área de avaliação de imóveis, devido ao avanço tecnológico e à sua capacidade de capturar relações complexas entre as variáveis. As redes neurais podem modelar não linearidades e interações entre os atributos dos imóveis, melhorando a precisão das estimativas de valor. No entanto, a complexidade das redes neurais e sua limitada capacidade de interpretar o peso de cada atributo na formação do valor dos imóveis podem gerar questionamentos na sua aplicação.

O objetivo deste trabalho é apresentar uma análise comparativa de resultados obtidos pelos diferentes métodos, para uma mesma amostra de dados de mercado imobiliário.

2. REVISÃO DE LITERATURA

2.1. Inferência Estatística

Os métodos de inferência estatística são de uso amplamente conhecido e pesquisado na Engenharia de Avaliações.

Essa abordagem estatística busca fornecer estimativas confiáveis do valor dos imóveis por meio da análise de dados e da aplicação de técnicas estatísticas para a estimação de modelos de valor (Dantas, 1998).

Os métodos inferenciais tradicionais utilizados na avaliação de imóveis incluem técnicas estatísticas como regressão linear, análise de variância e a análise de regressão múltipla. Essas técnicas são aplicadas para modelar e quantificar as relações entre as variáveis que influenciam o valor dos imóveis, como as características construtivas e de localização, entre outros fatores relevantes (Associação Brasileira de Normas Técnicas, NBR 14653).

Através da inferência estatística tradicional, é possível realizar testes de hipóteses, calcular intervalos de confiança e avaliar a significância estatística das variáveis utilizadas no modelo. Essas análises estatísticas contribuem para a objetividade e transparência do processo de avaliação, permitindo uma interpretação clara e fundamentada dos resultados (Dantas, 1998).

No entanto, é importante destacar que o uso da inferência estatística tradicional na avaliação de imóveis não está isento de uma série de desafios. Limitações como a suposição de linearidade nas relações entre as variáveis, a dependência espacial não considerada e a necessidade de uma amostra representativa podem impactar a precisão das estimativas.

2.2. Métodos Espaciais

O uso de métodos espaciais na avaliação de imóveis tem sido proposto para considerar os efeitos espaciais e de localização para melhorar a precisão e a confiabilidade das estimativas de valor. Esses métodos levam em consideração a influência da localização geográfica na determinação do valor dos imóveis, reconhecendo que propriedades próximas umas das outras tendem a apresentar valores semelhantes (Anselin, 1999a).

Entre os métodos espaciais utilizados, destacam-se os modelos de regressão espacial e os modelos geoestatísticos.

Os modelos de regressão espacial permitem a inclusão de variáveis espaciais, como preditores no modelo de avaliação, utilizando a distância entre os imóveis como um fator de influência fundamental (Anselin, 2001). Essas informações espaciais ajudam a capturar padrões e tendências locais, melhorando a acurácia das estimativas de valor. Os modelos de regressão espacial mais conhecidos e utilizados são o Modelo de Defasagem Espacial da variável dependente, e o Modelo do Erro Espacial que considera a autocorrelação entre resíduos (Anselin, 2002).

Por sua vez, os modelos geoestatísticos são particularmente úteis quando há uma dependência espacial dos valores dos imóveis. Eles permitem a modelagem da estrutura de dependência espacial dos dados, considerando a autocorrelação entre os valores dos imóveis em diferentes locais. Isso significa que os valores dos imóveis mais próximos geograficamente tendem a estar mais correlacionados do

que aqueles mais distantes. A krigagem, uma técnica comum na geoestatística, pode ser usada para interpolar os valores dos imóveis em locais não amostrados, melhorando a precisão das estimativas (Chica Olmo, 1994)

Uma das principais vantagens dos métodos espaciais é a capacidade de incorporar informações geográficas relevantes na avaliação de imóveis, levando em consideração a localização e a proximidade com outros imóveis ou pontos de interesse. Isso pode levar a estimativas mais precisas e realistas, especialmente em áreas com alta heterogeneidade espacial. Além disso, esses métodos ajudam a identificar padrões espaciais e clusters de valores, fornecendo insights adicionais sobre o mercado imobiliário local (Cano Guervós, 1999).

No entanto, é importante mencionar algumas limitações dos métodos espaciais na avaliação de imóveis. Um desafio é a necessidade de dados georreferenciados de boa qualidade, com informações precisas sobre a localização dos imóveis. Nem sempre esses dados estão disponíveis de forma abrangente e confiável. Além disso, a escolha adequada do tipo de modelo espacial, inferencial ou geoestatístico, apresenta outros desafios. Os métodos de geoestatística foram desenvolvidos para serem aplicados em dados homogêneos na sua composição intrínseca, tendo a sua localização espacial como única variável a ser modelada, sendo assim a geoestatística chamada de Teoria das Variáveis Regionalizadas. Por isso na avaliação de imóveis, onde existe heterogeneidade intrínseca dos imóveis, além da variabilidade espacial, é conveniente usar ambos métodos conjuntamente, a geoestatística pode ser utilizada em combinação com técnicas de inferência espacial (Peruzzo Trivelloni, 2005)

Apesar das limitações mencionadas, os métodos espaciais têm se mostrado uma ferramenta valiosa na avaliação de imóveis, permitindo uma abordagem mais abrangente para estimar o valor das propriedades. Ao analisar a localização geográfica e a dependência espacial dos valores, esses métodos podem complementar e enriquecer a inferência tradicional, contribuindo para estimativas mais precisas na avaliação de imóveis

2.3. Redes Neurais Artificiais

As redes neurais artificiais (RNAs) têm ganhado destaque na Engenharia de Avaliações como uma abordagem promissora para a estimativa de valores de imóveis. Inspiradas no funcionamento do cérebro humano, as RNAs são modelos computacionais compostos por neurônios interconectados, que processam informações e aprendem a partir dos dados disponíveis.

Essas redes são capazes de aprender padrões complexos e não-lineares dos dados, adaptando-se e ajustando seus pesos sinápticos durante o treinamento. Através de várias camadas de neurônios interligados, as RNAs são capazes de realizar a estimação do valor com base nas características dos imóveis (Preciado, 2015).

Na Engenharia de avaliações, as RNAs são aplicadas para estimar o valor dos imóveis com base em um conjunto de variáveis de entrada, como tamanho, localização, características construtivas e dados socioeconômicos relevantes. Essas redes são treinadas usando conjuntos de dados onde os valores dos imóveis são conhecidos, permitindo que a RNA aprenda a relação entre as variáveis de entrada e o valor dos imóveis (Pelli Neto, 2006).

Uma das principais vantagens das RNAs na Engenharia de Avaliações é sua capacidade de lidar com relações complexas e não-lineares entre as variáveis de entrada e o valor dos imóveis. Elas podem capturar nuances e interações entre os atributos dos imóveis, levando a estimativas mais precisas e detalhadas.

Além disso, as RNAs têm a capacidade de lidar com grandes volumes de dados e de variáveis explicativas, o que é especialmente relevante em um mercado imobiliário dinâmico.

Especialmente as RNA podem ser de grande utilidade quando se trata de avaliações em massa, onde a heterogeneidade é característica fundamental dos dados e os métodos inferenciais tradicionais mostram poder explicativo limitado, sendo também de pouca aplicabilidade devido à falta de significância estatística das variáveis que precisam ser utilizadas.

No entanto, é importante destacar algumas limitações das RNAs na Engenharia de Avaliações. A interpretação dos resultados pode ser um desafio, já que as RNAs são consideradas "caixas pretas", ou seja, a diferença dos modelos de regressão que podem ser interpretados de forma objetiva e clara, nas RNA não é possível visualizar de forma simples como chegam às estimativas de valor. Além disso, as RNAs requerem um conjunto de dados adequado e representativo para o treinamento, e o processo de treinamento pode ser computacionalmente intensivo (Preciado, 2015).

3. METODOLOGIA

O presente trabalho apresenta uma comparação dos resultados obtidos utilizando três métodos diferentes para uma mesma amostra de mercado:

1. Inferência estatística tradicional, modelo de mínimos quadrados ordinário (MQO)
2. Uso de modelos espaciais, regressão espacial e geoestatística, para analisar e modelar a estrutura espacial dos dados.
3. Uso de Redes Neurais Artificiais para estimar o valor dos imóveis.

3.1 Área de estudo

A área de estudo corresponde ao bairro Pocitos, na cidade de Montevideu, Uruguai.

O bairro Pocitos é um bairro principalmente residencial e considerado um dos bairros com melhor índice de renda per capita da cidade. É um bairro onde predomina a construção de prédios de apartamentos, com bom acesso a todos os serviços urbanos. Por estar próximo do mar e da Avenida Rambla, que contorna a orla da cidade no Rio da Prata, o principal fator de valorização por localização é a proximidade à Avenida Rambla. Outros possíveis pólos de valorização na região são dois Shopping Center localizados próximos aos dois extremos da área de estudo, o Montevideo Shopping Center e o Shopping Punta Carretas.

A delimitação da área de estudo está definida pelos imóveis localizados no bairro Pocitos a uma distância máxima de até 700 metros da Rambla).

Os dados de mercado utilizados para compor a amostra correspondem a vendas realizadas durante os anos de 2018 e 2019 para imóveis de tipo apartamento.

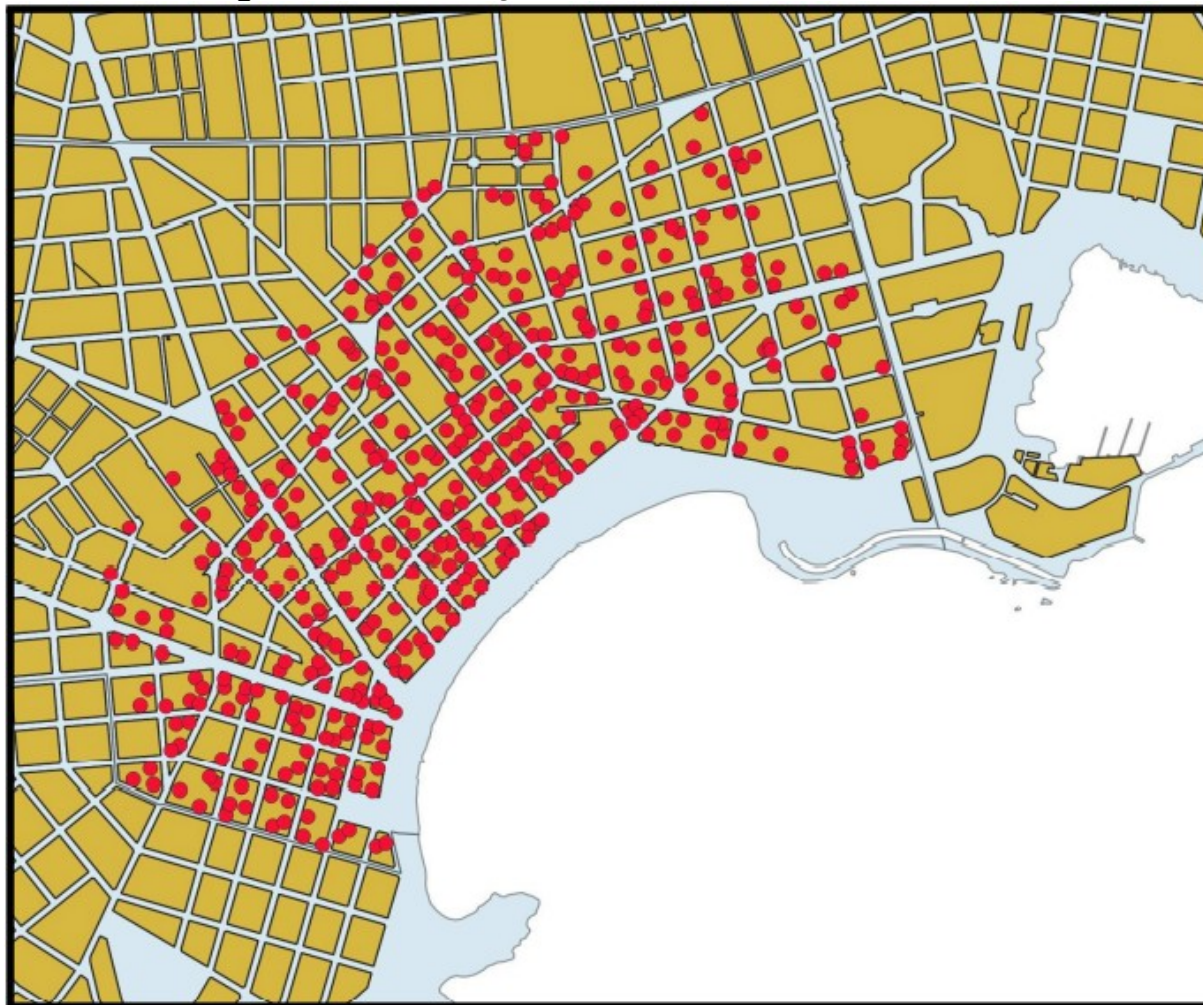
Os dados das transações declaradas desse período foram analisados e dessa análise foi realizada um descarte daqueles que poderiam estar declarados abaixo do valor real de mercado. Para isto foram descartados aqueles valores que apresentaram Valor Unitário inferior a 1.500,00 U\$/m².

Além dessa depuração foi também selecionado apenas um imóvel por face de quadra e quando havia mais de um apartamento no mesmo prédio foi considerado aquele do andar mais baixo para diminuir a heterogeneidade da amostra.

As variáveis construtivas dos imóveis considerados no estudo foram retiradas do Cadastro Nacional de Imóveis (Dirección General de Catastro), que disponibiliza acesso via internet para acessar as características intrínsecas de cada imóvel.

A amostra está composta por 380 apartamentos e sua distribuição geográfica pode ser visualizada na Figura 1.

Figura 1 – Distribuição da amostra na área de estudo



Fonte: elaborado por alunos da disciplina Avaliações 3 – Fing – Udelar (2022)

As variáveis utilizadas nesse trabalho são as seguintes:

Valor Unitário do imóvel, medido em U\$/m², obtido a partir do Valor de Venda do imóvel, como variável dependente.

Superfície ou Área Total do apartamento, medida em m².

Categoria Construtiva, codificada com valores numéricos variando desde o valor 4 (Categoria Média-Econômica) até o valor 8 (Categoria Muito Boa).

Estado de Conservação, codificada variando desde o valor 1 (Ruim) até 7 (Excelente).

Altura, correspondendo ao andar do imóvel, medida em metros.

Idade, correspondendo a idade do imóvel, medida em anos.

Distância Avenida, correspondendo à distância à Avenida Rambla, medida em metros.

Distância Shopping, correspondendo à menor das distâncias aos dois Shoppings próximos à área de estudo, medida em metros.

3.2 Aplicação da Inferência Estatística

Como variável dependente foi definido o Valor Unitário (VU), e também foi testado o logaritmo neperiano do Valor Unitário. O melhor resultado nos testes de regressão foi obtido em todos os casos para a variável dependente VU.

As variáveis Categoria Construtiva e Estado de Conservação foram testadas de duas formas: como variáveis ordinais com os códigos numéricos escolhidos, e também separando cada código das variáveis como variáveis dicotômicas, associando uma variável dicotômica para cada valor delas. Em ambas variáveis os modelos de inferência tiveram melhores resultados para a variável considerada como variável ordinal.

A variável Idade foi considerada na forma lineal e também na forma polinômica de grau 2, sendo que o melhor resultado foi para a variável linear.

A variável Distância Avenida foi testada de três formas diferentes: por um lado, como variável linear, usando diretamente a distância em metros. Por outro lado foi testada como variável polinômica, considerando polinômios de segundo e terceiro grau. Em terceiro lugar a variável foi redefinida como um conjunto de variáveis dicotômicas, dividindo os dados da amostra segundo as distâncias considerando intervalos a cada 100 metros. Entre as três formas de definição da variável, o melhor resultado foi obtido com a variável diretamente na sua forma linear.

A variável Distância Shopping não se mostrou significativa na regressão.

Foram testadas também as interações entre todas as variáveis independentes, mas nenhuma interação foi estatisticamente significativa.

O modelo de inferência estatística final foi o representado na Figura 2.

Figura 2 – Regressão Múltipla. Resumo dos resultados

1	RESUMO DOS RESULTADOS					
2						
3	<i>Estatística de regressão</i>					
4	R múltiplo	0,53634				
5	R-Quadrado	0,287661				
6	R-quadrado aj	0,276202				
7	Erro padrão	498,6258				
8	Observações	380				
9						
10	ANOVA					
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
12	Regressão	6	37450042	6241674	25,1045	5,04203E-25
13	Resíduo	373	92738143	248627,7		
14	Total	379	1,3E+08			
15						
16		<i>Coefficiente: erro padrão</i>		<i>Stat t</i>	<i>valor-P</i>	
17	Interseção	3036,171	218,6986	13,8829	1,28E-35	
18	AREA	-6,43717	0,724394	-8,88629	2,73E-17	
19	Categoria	38,59189	32,41855	1,190426	0,234636	
20	EDAD	-4,42142	1,410278	-3,13514	0,001854	
21	Alturas	10,65236	3,161202	3,36972	0,000831	
22	Est. Conservac	56,2397	23,50041	2,393136	0,017199	
23	distance	-0,53977	0,138549	-3,89587	0,000116	

Fonte: elaborado pelos autores

O resultado da regressão mostra um coeficiente de determinação R² baixo, de apenas 0.288. As variáveis independentes se apresentam com sinais coerentes e níveis de significância aceitáveis, exceto a Distância ao Shopping que não se mostrou significativa e por isso foi retirada do modelo final.

O baixo valor do coeficiente de determinação mostra que a especificação do modelo não é completa. Podemos dizer que a variável Distância à Rambla como única variável explicativa de localização não é suficiente, pois existe uma estrutura de localização mais complexa que apenas a distância à Rambla, pelo qual outros fatores de localização deveriam ser considerados para melhorar o ajuste do modelo. Esta possibilidade será estudada por meio de modelos espaciais, regressão espacial e geoestatística.

3.3 Modelos Espaciais

REALIZAÇÃO

Para testar autocorrelação espacial e testar os modelos de regressão espacial foi utilizado o software gratuito Geoda (Anselin, 2003)

No Geoda foram criadas matrizes de pesos espaciais W para três critérios de vizinhança diferentes: distância máxima de vizinhos de 300, 400 e 500 metros. Com estas matrizes de pesos espaciais foram testadas as regressões espaciais da Variável Dependente e do Erro espacial. Foi testada também uma variável espacializada com a variável independente Categoria, por meio da multiplicação da matriz W com a variável Categoria (WxCat). Esta variável espacial, que pode ser interpretada, como todo produto da matriz W com qualquer variável, como a média da variável Categoria na vizinhança de cada imóvel, se mostrou significativa nos modelos de regressão.

O melhor modelo de regressão espacial foi o Modelo do Erro Espacial, que se mostra na Figura 3.

Figura 3 – Modelo do Erro Espacial: resumo de resultados

REGRESSION				
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION				
Data set	: Pocitos-Geoda 379 dados			
Spatial Weight	: Pocitos-Geoda 379 dados 300m			
Dependent Variable	: ValorUnitario			
Number of Observations:	379			
Mean dependent var	: 2690,796834	Number of Variables	: 8	
S.D. dependent var	: 585,921581	Degrees of Freedom	: 371	
Lag coeff. (Lambda)	: -0,746375			
R-squared	: 0,355324	R-squared (BUSE)	: -	
Sq. Correlation	: -	Log likelihood	: -2870,979536	
Sigma-square	: 221320	Akaike info criterion	: 5757,96	
S.E of regression	: 470,446	Schwarz criterion	: 5789,46	
Variable	Coefficient	Std. Error	z-value	Probability
CONSTANT	-2284,53	649,574	-3,51696	0,00044
Superficie	-6,4653	0,655935	-9,85662	0,00000
DistanciaAv	-0,124736	0,103499	-1,20519	0,22813
Edad	-4,12478	1,33866	-3,08128	0,00206
Alturas	10,765	2,95456	3,64353	0,00027
EstadoCons	68,3721	21,3224	3,20658	0,00134
W400_cat	1004,12	114,701	8,75416	0,00000
DistanciaSh	-0,0945396	0,0333735	-2,83277	0,00461
LAMBDA	-0,746375	0,351862	-2,12121	0,03390

Fonte: elaborado pelos autores

Pode ser observado que o Coeficiente de Determinação R² nesse modelo é 0.342.

Portanto representa um poder explicativo maior que o modelo de inferência estatística tradicional que tem R² = 0,287, aumentando em 5% o poder explicativo.

3.4 Combinação de Regressão Espacial e Geoestatística

Para utilizar o método geoestatístico do krigagem é necessário homogeneizar previamente os dados em relação a todas as variáveis construtivas ou intrínsecas dos imóveis, deixando apenas de fora as variáveis de localização.

De acordo a metodologia apresentada em Peruzzo Trivelloni (2005), o valor de localização pode ser estimado com uma nova variável obtida pela aplicação combinada de ambos métodos de análise espacial, através de duas etapas:

1. Aplicação do Modelo de Regressão Espacial do Erro considerando como variáveis independentes apenas as variáveis construtivas dos imóveis.

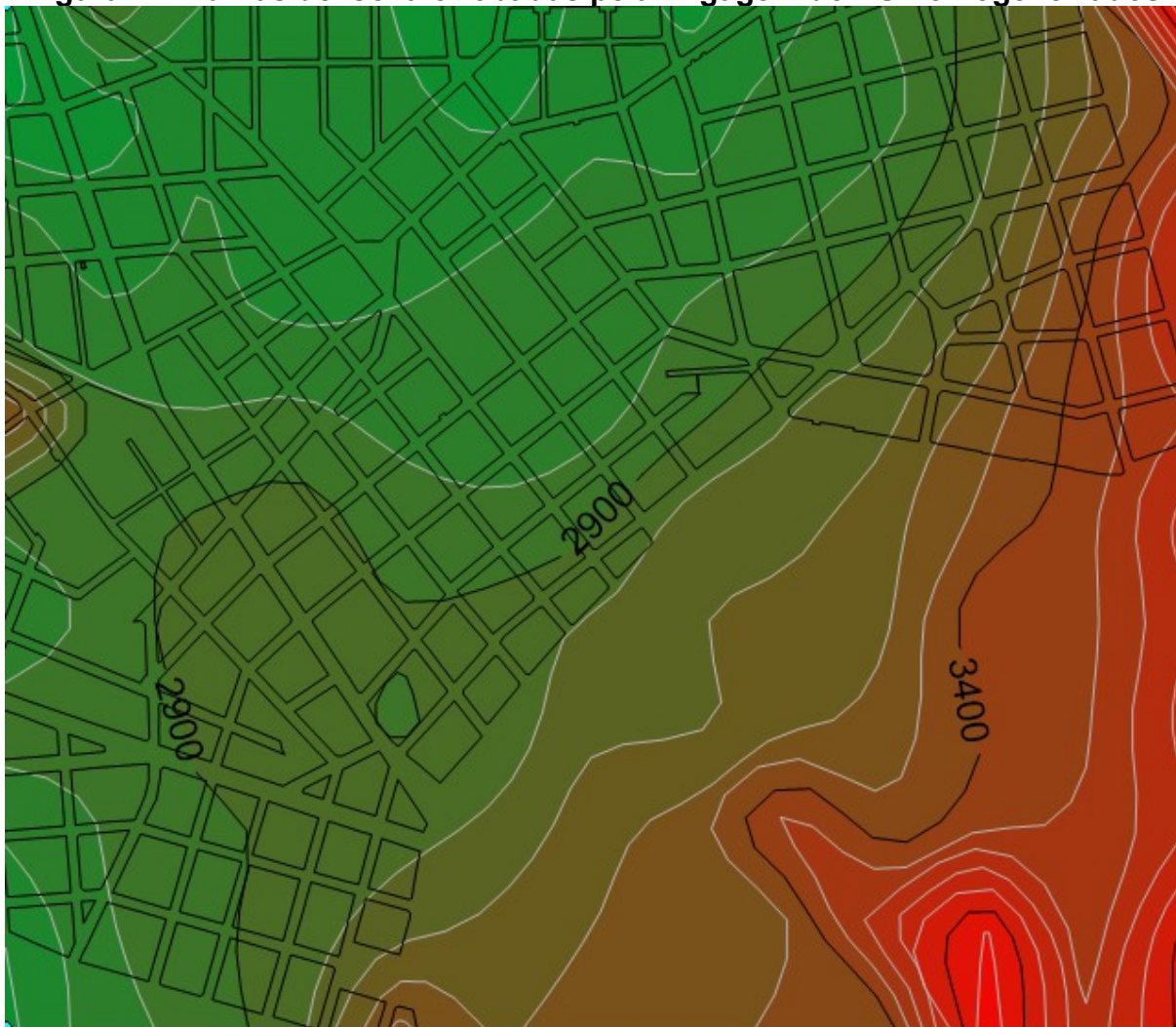
2. Os VU dos imóveis são homogeneizados a partir do resultado do Modelo do Erro Espacial, e este novo VU homogeneizado, como variável regionalizada, é analisado e modelado pelo método de interpolação geoestatística de krigagem.

A superfície de krigagem assim obtida é definida como uma Variável de Localização, VL. Esta variável VL é então utilizada junto com as outras variáveis construtivas num novo modelo clássico de regressão.

Aplicando a metodologia antes mencionada obtemos uma nova variável VL que representa o valor de localização na área de estudo considerada, que pode ser observada na Figura 4.

Na Figura 4 se observam as linhas de isovalor e que os valores de localização aumentam com a proximidade da Avenida Rambla, como era esperado, e também aumentam em outras duas regiões do bairro: uma delas próxima da Avenida Luis Alberto de Herrera e a outra próxima da Avenida Brasil, duas avenidas importantes do Bairro, as duas com saída direta para a Avenida Rambla.

Figura 4 – Curvas de isovalor obtidas pela krigagem do VU homogeneizados



Fonte: elaborado pelos autores

Assim, a análise geoestatística mostra, de forma objetiva, a partir da informação contida nos dados da amostra, quais são as regiões que apresentam maior ou menor valorização na área de estudo considerada. Vemos que a Distância a Avenida Rambla não é, efetivamente, a única variável importante de localização.

Uma vez definida a nova variável de localização, VL, aplicamos novamente o modelo de Regressão Múltipla tradicional, considerando como variáveis explicativas as variáveis construtivas e a nova variável VL obtida, e os resultados da nova regressão são os que mostra a Figura 5.

Figura 5 – Regressão Múltipla usando a variável espacial de localização VL

REGRESSION					

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION					
Data set	:	Pocitos 379datos VL final			
Dependent Variable	:	ValorUnitario			
Number of Observations:	:	379			
Mean dependent var	:	2690.8	Number of Variables	:	6
S.D. dependent var	:	585.922	Degrees of Freedom	:	373
R-squared	:	0.351699	F-statistic	:	40.47
Adjusted R-squared	:	0.343008	Prob(F-statistic)	:	3.22672e-33
Sum squared residual:	:	8.43519e+07	Log likelihood	:	-2871.09
Sigma-square	:	226145	Akaike info criterion	:	5754.17
S.E. of regression	:	475.547	Schwarz criterion	:	5777.8
Sigma-square ML	:	222564			
S.E of regression ML:	:	471.767			

Variable	Coefficient	Std.Error	t-Statistic	Probability	

CONSTANT	-282.285	453.081	-0.623035	0.53364	
Superficie	-6.51066	0.653649	-9.96048	0.00000	
Edad	-4.22422	1.34457	-3.14169	0.00181	
Alturas	10.5428	2.99801	3.51659	0.00049	
EstadoCons	57.875	21.5492	2.68572	0.00756	
VL	1.16401	0.154367	7.54055	0.00000	

Fonte: elaborado pelos autores

Na Figura 5 se observa que o melhor Modelo de Regressão utilizando como variáveis explicativas as variáveis construtivas e a variável de localização VL, obtida por análise espacial, tem um Coeficiente de Determinação de 0.352.

3.5 Redes Neurais Artificiais (RNA)

Para a aplicação das RNA foi definida uma arquitetura básica da rede como sendo um modelo com as seguintes capas:

- A capa de entrada, com as variáveis pesquisadas para os dados da amostra: Area, Categoria, Estado de Conservação, Idade, Altura, Distância à Rambla e Distância a Shopping.
- 3 capas ocultas que serão testadas com diferentes números de neurônios artificiais
- A capa de saída, constituída por um único neurônio correspondendo ao Valor Unitário do imóvel.

Como função de ativação foi escolhida nestes testes a função Relu, que é uma das que tem obtido melhores resultados em pesquisas anteriores.

O software utilizado para o treinamento da rede foi no site Google Colabs, que permite programar uma rede de acordo com a pesquisa, utilizando uma capacidade de processamento adequada aos nossos requerimentos.

Os testes para cada definição de uma RNA são aplicados dividindo a amostra em duas partes: uma das partes é utilizada para treinamento da rede, e a outra parte é utilizada para teste do aprendizado.

O número de ciclos de aprendizado utilizado para treinar cada modelo definido foi de 500 ciclos.

A divisão da amostra numa amostra de treinamento e outra de prova foi testada em duas formas: uma foi a divisão 70/30, sendo 70% da amostra utilizada para treinamento da rede e 30% para o teste dos resultados. A outra foi a divisão 80/20, sendo 80% da amostra total usada para treinamento e o 20% restante para teste de resultados.

Foram testados diversos modelos de RNA diferenciados pelo número de neurônios de cada capa e pela divisão da amostra como aparece na Figura 6.

Figura 6 – Modelos de RNA testados

Modelo	Neurônios	Treino/Teste
1	10,10,10	80/20
2	10,10,10	70/30
3	20,10,10	80/20
4	20,10,10	70/30
5	20,20,10	80/20
6	20,20,10	70/30
7	30,30,30	70/30

Fonte: elaborado pelos autores

Para cada um dos modelos definidos de RNA foi realizado o treinamento da rede e posteriormente foram calculados os valores de saída, ou seja, o VU estimado pela RNA. A partir deste valor estimado foi calculado o valor do Coeficiente de Determinação R² de cada modelo de RNA e também foi calculado o erro para cada observação, através da diferença do VU observado e VU estimado.

Os resultados obtidos para os coeficientes R² de cada um dos modelos de RNA pode ser observado na Figura 7.

Figura 7 – Resultados de R2 obtidos para os modelos RNA

Modelo	Neurônios	Treino/Teste	R2
1	10,10,10	80/20	0,54
2	10,10,10	70/30	0,63
3	20,10,10	80/20	0,67
4	20,10,10	70/30	0,69
5	20,20,10	80/20	0,71
6	20,20,10	70/30	0,69
7	30,30,30	70/30	0,85

Fonte: elaborado pelos autores

Como se observa na Figura 7 todos os modelos de RNA obtiveram valores de R2 muito superiores aos valores de R2 obtidos pelo modelo de regressão tradicional (0.29) e pelos modelos de análise espacial (0.34 e 0.35).

3.6 Análise e comparação de resíduos obtidos pelos três modelos

Outra forma de analisar os resultados obtidos pelos modelos inferencial clássico, inferencial com análise espacial e melhor modelo de RNA testado, foi calcular os resíduos obtidos por cada um dos métodos, tanto da regressão tradicional, as análises espaciais e o melhor modelo de RNA.

A partir deles foram calculados os resíduos relativos, ou seja, a divisão entre (valor do resíduo)/(Valor Unitário). Desta maneira o objetivo foi comparar os resíduos como uma porcentagem de erro, para que fossem melhor comparáveis.

E a partir do cálculo de resíduos percentuais, foram tomados os valores absolutos deles, para poder medir o grau de afastamento e dispersão desses erros em relação aos Valores Unitários da amostra de mercado.

Uma vez obtidos estes valores absolutos foram calculadas média e desvio padrão destes erros percentuais em valor absoluto entre os seguintes 3 modelos:

1. Modelo de inferência tradicional
2. Modelo de inferência com análise espacial.
3. Modelo de RNA número 7, com 3 camadas de 30 neurônios.

Os resultados podem ser observados na Figura 8.

Figura 8 – Média e Desvio padrão dos resíduos relativos dos três modelos

Valor absoluto de (Resíduos/VU)	MQO	Regr Esp+GeoE	RNA
Média	0,147	0,143	0,069
Desvio padrão	0,142	0,134	0,100

Fonte: elaboração própria dos autores

Da análise da Figura 8 se observa que a média de erros percentuais em valor absoluto para os modelos inferenciais é de 14,7% e 14,3% respectivamente, enquanto que para o modelo de RNA é de 6,9%.

Da mesma forma a dispersão dos resíduos do modelo de RNA também é menor, sendo o valor do seu desvio padrão de 0,100 enquanto os modelos inferenciais tem desvio padrão de 0,142 e 0,134 respectivamente.

3.7 Análise da média dos erros em relação com variáveis específicas.

O valor absoluto dos erros percentuais (Resíduo/VU) obtidos por cada método foi comparado também para cada um dos métodos, dividindo a amostra em faixas segundo cada uma das variáveis Distância a Avenida Rambla, Área do imóvel, Categoria do imóvel e Valor Unitário, como se observa na Figura 9.

Figura 9 – Média de erros por faixa das variáveis

Médias do valor absoluto de Resíduos/VU			
Distancia a Avenida Rambla (m)	MQO	Regr Esp+GeoE	RNA
<50	0,150	0,134	0,044
50-150	0,131	0,125	0,071
150-300	0,155	0,148	0,066
300-500	0,140	0,141	0,060
>500	0,159	0,159	0,077
Area do imóvel (m2)	MQO	Regr Esp+GeoE	RNA
<50	0,184	0,170	0,071
50-100	0,139	0,143	0,069
100-150	0,127	0,118	0,057
>150	0,147	0,142	0,051
Categoria do imóvel	MQO	Regr Esp+GeoE	RNA
4	0,159	0,156	0,048
5	0,146	0,144	0,069
6	0,123	0,141	0,068
7	0,164	0,144	0,058
8	0,164	0,109	0,027
Valor Unitário (U\$/m2)	MQO	Regr Esp+GeoE	RNA
<2000	0,386	0,364	0,099
2000-2500	0,151	0,145	0,062
2500-3000	0,068	0,079	0,058
>3000	0,146	0,135	0,069

Fonte: elaborado pelos autores

Pode se observar na Figura 9 que o modelo RNA foi superior aos modelos de Inferência Clássica e Modelo Espacial em todas as faixas para cada uma das variáveis consideradas.

No caso da variável Valor Unitário do imóvel, os modelos inferenciais mostraram uma média de erros menores para a faixa de VU entre 2500 e 3000 U\$/m2, de apenas 6.8%, em relação com as outras faixas de VU. Ao mesmo tempo, para a faixa de VU menor que 2000 U\$/m2 a média de erros percentuais foi muito maior, de 38.6%, em relação à média geral. Se observa assim que os modelos inferenciais não estimaram com a mesma qualidade as 4 faixas mencionadas de VU. O modelo RNA, ao contrário, obteve média de erros muito semelhantes entre as 4 faixas de VU consideradas, variando entre 5.8% e 9.9%. Isto mostrou um comportamento mais homogêneo nas estimativas realizadas pela RNA sobre os dois modelos inferenciais considerados.

4. CONCLUSÕES

Como conclusões das análises realizadas podemos destacar as seguintes:

1. Em relação ao modelo de Regressão tradicional, os modelos de Regressão Espacial obtiveram resultados com maior poder explicativo, com um R^2 de 0,34 para o Modelo do Erro Espacial e 0,35 para o Modelo contendo uma variável de localização definida por geoestatística, em comparação ao R^2 de 0,29 do modelo tradicional de regressão por mínimos quadrados.

2. Estes valores baixos do R^2 indicam que variáveis importantes na formação do valor não foram consideradas nos modelos. Os modelos espaciais, ao considerar os efeitos espaciais de vizinhança e a consequente autocorrelação espacial nos dados, conseguiram melhorar os resultados, aumentando em 5% e 6% o poder explicativo dos modelos. Mesmo assim, o R^2 dos modelos espaciais continua sendo baixo.

3. Os modelos de RNA obtiveram resultados superiores em relação aos modelos inferenciais tradicional e espacial. Os valores de R^2 obtidos com os modelos de RNA resultaram substantivamente superiores, para todas as arquiteturas de redes consideradas, de 3 capas ocultas e número de neurônios por capa variando entre 10 e 30.

4. Nos modelos com maior número de neurônios por capa os resultados de estimação da rede foram claramente superiores, tanto no referente ao valor de R^2 obtido quanto à análise detalhada dos resíduos da estimação.

5. O R^2 da RNA foi substancialmente maior (0.85) frente aos R^2 dos modelos inferenciais tradicional e espaciais (0.29, 0.34 e 0.35).

6. As estimações dos valores unitários realizadas pela RNA obtiveram um ajuste mais próximo aos valores de mercado observados: a média dos valores absolutos dos resíduos da RNA foi de 6,9%, frente a 14,7% do modelo de MQO e 14,3% do modelo espacial combinado.

7. A dispersão das estimações obtidas pela RNA (0.100) foi menor que as obtidas pelos métodos inferencial clássico (0.142) e métodos espaciais (0.134).

8. Se analisamos a média dos erros de cada modelo comparativamente com a distribuição das variáveis Distância a Rambla, Area, Categoria e Valor Unitário, o modelo de RNA estimou com maior homogeneidade que os modelos inferenciais em todas as faixas de distribuição dessas variáveis. Especialmente no caso da variável Valor Unitário observamos que a média dos erros obtidas pelos modelos inferenciais não foi homogênea, enquanto que a RNA teve um comportamento similar para todas as faixas de valor.

9. Ao contrário dos modelos inferenciais, o modelo da RNA não permite analisar individualmente a influência de cada variável no valor final, sendo esta uma fragilidade do método das RNA, principalmente quando a justificação das estimativas obtidas é necessária. As relações matemáticas entre todos os neurônios da rede neuronal, ajustados através do treinamento da rede, são de interpretação complexa, ao contrário do que acontece nos modelos de regressão múltipla.

10. O fato de não poder analisar a influência de cada variável na estimação do valor pela RNA não diminui a importância da comprovação prática que as RNA podem fornecer estimativas muito mais precisas de valor que os métodos inferenciais tradicionais.

11. As RNA podem ser uma ferramenta muito poderosa especialmente no caso das avaliações em massa de imóveis, onde a heterogeneidade e complexidade espacial dos fatores de valorização não resultam facilmente modelados pelos métodos inferenciais tradicionais.

5. RECOMENDAÇÕES

O uso de RNA na Engenharia de Avaliações é uma área que ainda precisa de mais testes para se compreender melhor as suas potencialidades e limitações na nossa área de atividade profissional.

É fundamental continuarmos desenvolvendo pesquisas sobre as possibilidades das novas ferramentas de Inteligência Artificial (IA) e publicando os resultados para que toda a comunidade da Engenharia de Avaliações possa ter melhores elementos de análise e de escolha.

As RNA são apenas uma das áreas da Inteligência Artificial que está em plena evolução e desenvolvimento. Também existem outras tecnologias e ferramentas de IA sendo desenvolvidas. Estamos na presença de uma verdadeira revolução que avança rapidamente em todas as áreas profissionais e da vida humana em geral.

Os profissionais da Engenharia de Avaliações precisamos continuar atentos e abertos aos desenvolvimentos tecnológicos nessa área, aprofundando nossos conhecimentos para poder acompanhar e avaliar corretamente as possibilidades que a Inteligência Artificial oferece para a nossa profissão.

REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L. **GeoDa 0.9 User's Guide**. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, 2003.

ANSELIN, L. **Spatial Econometrics**. Discussion paper. Bruton Center, School of Social Sciences, University of Texas at Dallas, 1999a.

ANSELIN, L. **The Future of Spatial Analysis in the Social Sciences**. Geographics Information Sciences 5, 1999b.

ANSELIN, L. **Spatial Externalities, Spatial Multipliers and Spatial Econometrics**. Specialist Meeting on Spatial Externalities, Santa Barbara, California, 2001.

ANSELIN, L. **Under the Hood. Issues in the Specification and Interpretations of Spatial Regression Models**, Agricultural Economics, vol.27, Issue 3, pp 247-267, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-1. Avaliação de bens. Parte 1: Procedimentos gerais**. Rio de Janeiro, 2001.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2. Avaliação de bens. Parte 2: Imóveis urbanos**. Rio de Janeiro, 2004

BAPTISTELLA, M., BASTOS CUNICO, L.H., ARS STEINER, M.T. **O uso de redes neurais na Engenharia de Avaliações: determinação dos valores venais de imóveis urbanos**. Revista Ciências Exatas e Naturais, Vol. 9, No.2, 2007.

CANO GUERVÓS, R. **Aproximación al Valor de la Vivienda. Aplicación a la Ciudad de Granada**. Ed. Universidad de Granada, 1999.

Chávez Barturén, D. A. **Aplicación de redes neuronales artificiales para la valorización de los departamentos habitacionales de la ciudad de Lima em 2020**. Tesis, Universidad Señor de Sipán, Perú, 2022.

CHICA OLMO, J. **Teoría de las Variables Regionalizadas. Aplicación en Economía Espacial y Valoración Inmobiliaria**. Ed. Universidad de Granada, 1994.

CHICA OLMO, J. **Spatial Estimation of Housing Prices and Locational Rents**. Urban Studies, Vol. 32, N. 8, pag. 1331-1344, 1995.

DANTAS, R. A. **Engenharia de avaliações: uma introdução à metodologia científica**. Ed. Pini, 1998.

DANTAS, R. A. **Modelos Espaciais aplicados ao Mercado Habitacional. Um Estudo de Caso para a Cidade do Recife**. Tese de Doutorado em Economia. Universidade Federal de Pernambuco. Recife, 2003.

DANTAS, R. A., MAGALHÃES, A. M., ROCHA, F. J. S. **La importancia de la regresión espacial en la tasación inmobiliaria**. I Congreso Internacional en Tasación y Valoración. Valencia, 2002.

DANTAS, R. A., MAGALHÃES, A. M., VERGOLINO, J. R. de O. **Modelos Espaciais aplicados ao Mercado de apartamentos do Recife**. XII COBREAP Congresso Brasileiro de Engenharia de Avaliações e Perícias, Anais. Belo Horizonte, 2003a.

DANTAS, R. A., SÁ, L. A.C. M., PORTUGAL, J. L. **Elaboração de Plantas de Valores sob a ótica da inferência espacial**. XII COBREAP Congresso Brasileiro de Engenharia de Avaliações e Perícias, Anais. Belo Horizonte, 2003b.

NUÑEZ TABALES, J.M., REY CARMONA, F.J., CARIDAD Y OCERIN, J.M. **Redes neuronales aplicadas a la valoración de locales comerciales**. Informes de la Construcción, Vol. 69, 2017.

PELLI NETO. A. **Redes Neurais Artificiais aplicadas as avaliações em massa. Estudo de caso para a cidade de Belo Horizonte/MG**. Dissertação de Mestrado, Universidade Federal de Minas Gerais, 2006.

PERUZZO TRIVELLONI, C. A. **Comparação entre Métodos Inferenciais e Inteligência Artificial (Redes Neurais Artificiais)**. III Seminario Sulamericano de Avaliações e Perícias de Engenharia, Balneario Camboriu, 2023.

PERUZZO TRIVELLONI, C. A. **Geoestadística y Regresión Espacial aplicada a los análisis de los inmuebles: conceptos y aplicaciones**. Primer Congreso Internacional de Avaluaciones y Expropiaciones del Uruguay, Punta del Este, 2014.

PERUZZO TRIVELLONI, C. A. **Comparação de Modelos Inferenciais Tradicionais e Espaciais utilizando diferentes variáveis de localização**. XIV Congresso Brasileiro de Engenharia de Avaliações e Perícias, Salvador, 2007.

PERUZZO TRIVELLONI, C. A.; HOCHHEIM, N. **O valor de localização dos imóveis: determinação por métodos de Análise Espacial**. XIII Congresso Brasileiro de Engenharia de Avaliações e Perícias e XXII Congresso Panamericano de Valuación, Fortaleza, 2006.

PERUZZO TRIVELLONI, C. A. **Método para determinação do valor de localização com uso de técnicas inferenciais e geoestatísticas na avaliação em massa de imóveis**. Tese de Doutorado, Universidade Federal de Santa Catarina, 2005.

PRECIADO CARRILLO, J. C. **Redes neuronales artificiales, una aplicación al mercado inmobiliario habitacional de segmento medio de la ciudad de Morelia Michoacán, México**. Revista Nicolaita de Estudios Económicos, Vol. XIV, No.1, 2019.

PRECIADO CARRILLO, J. C. **Valoración de inmuebles urbanos: comparativa modelo de regresión multivariable versus redes neuronales artificiales para la ciudad de Morelia, Michoacán, México**. Tesis Doctoral, Universidad Politécnica de Valencia, 2015.