

Análise comparativa de métodos: inferência estatística tradicional, inferência espacial e redes neurais artificiais. Estudo de caso.

Carlos Peruzzo

Florencia Rodriguez

Luis Calderón



11 A 15

SETEMBRO | 2023

SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

Análise comparativa

- Inferência estatística tradicional
 - Inferência espacial e geoestatística
 - Redes Neurais Artificiais
-
- Estudo de caso: 379 apartamentos localizados no Bairro Pocitos, Montevideu, Uruguai.

Método 1. Inferência estatística

- Inferência Estatística Tradicional
 - Equação de Regressão Múltipla
 - Estimação por Mínimos quadrados ordinários.
 - Permite modelar e quantificar as relações entre variáveis explicativas e a variável explicada (valor do imóvel)
 - Aplica testes de hipótese para aceitação de variáveis no modelo (significância do modelo e das variáveis), assim como condições necessárias para os resíduos do modelo.
 - O Modelo de Valor obtido permite uma interpretação clara e fundamentada dos resultados.
 - Limitações: suposição de linearidade e independência das variáveis, no considerar relação espacial das variáveis (autocorrelação espacial nos resíduos e nos valores).

Método 2. Modelos Espaciais

- Considerar os efeitos espaciais e de localização.
- Considerar a localização geográfica no modelo e as interações em função da distância entre observações
- Modelos de Regressão Espacial (Modelo da variável dependente espacialmente defasada e modelo do erro espacialmente defasado)
- Modelos geoestatísticos (interpolação por krigagem)
- Combinação dos métodos regressão e krigagem
- Limitações: contar com suficientes dados georreferenciados, tratar adequadamente efeitos de heterogeneidade (características construtivas) e os efeitos de localização (krigagem requer variável homogênea).

Método 3. Redes Neurais Artificiais

- Modelos computacionais compostos por um conjunto de “neurônios artificiais” interconectados que processam informações de entrada e saída buscando as relações entre as variáveis de entrada e a variável de saída (valor).
- O “treinamento” da RNA significa ajustar equações matemáticas de acordo a parâmetros definidos pelo avaliador.
- Ajuste dos parâmetros de todas as equações para cada neurônio são realizadas para minimizar o erro final de estimação.
- Estrutura geral de uma RNA:
 - Capa de neurônios de entrada (variáveis disponíveis a serem utilizadas)
 - Capas intermediárias de neurônios
 - Capa de saída (resultado final: valor)

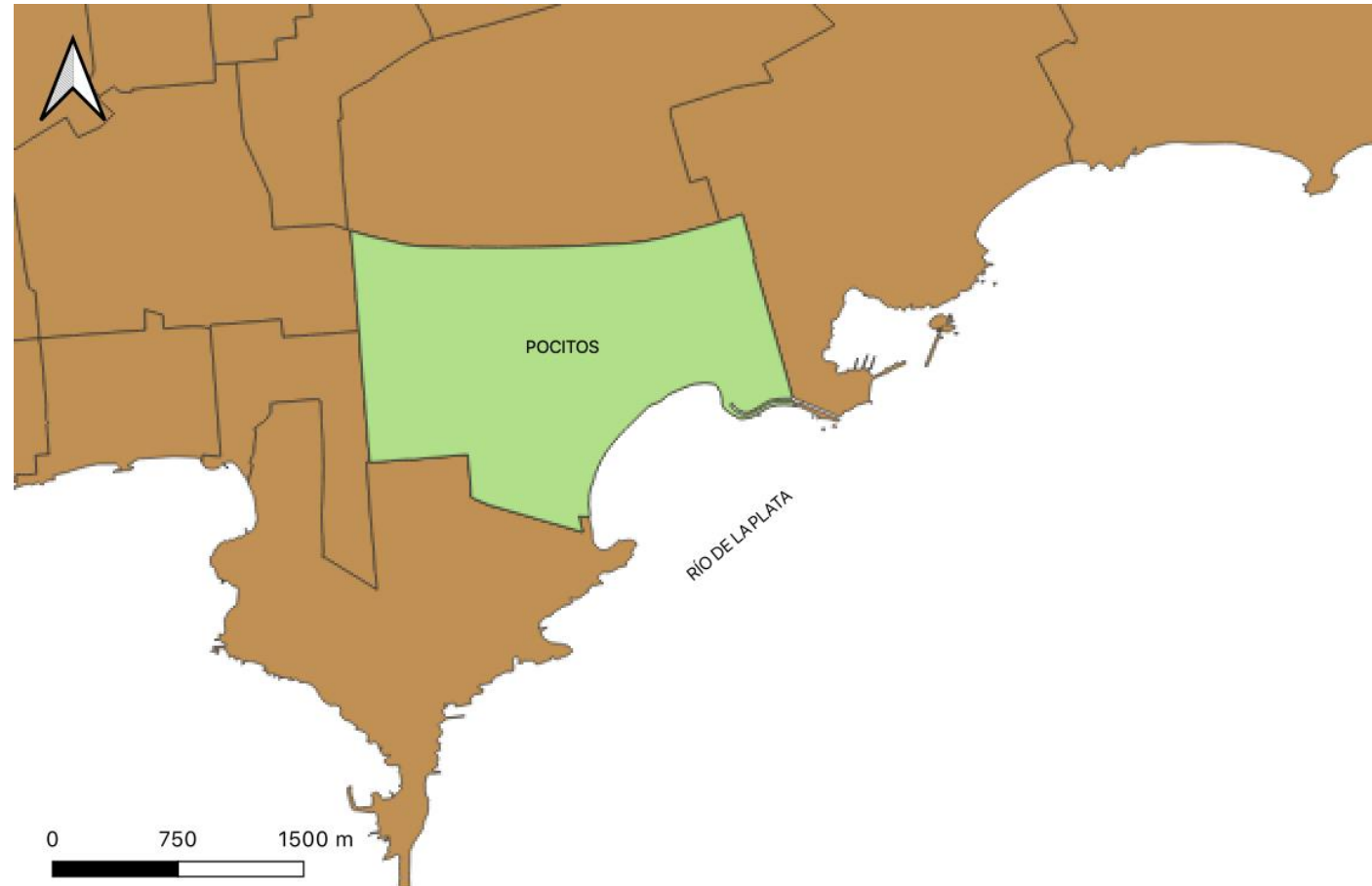
Redes Neurais Artificiais

- Vantagens:
 - capacidade de lidar com relações complexas e não lineares entre as variáveis de entrada e de saída (valor)
 - Capacidade de lidar com grandes volumes de dados e de variáveis explicativas
 - Avaliações em massa: capacidade de lidar com a heterogeneidade dos dados, onde métodos inferenciais tradicionais mostram poder explicativo limitado.
 - Atualmente estão disponíveis diversos softwares e até existe a possibilidade de processamento e modelagem online
- Desvantagens ou desafios:
 - Modelo matemático extremamente complexo quando a RNA tem um número importante de “neurônios”, que vira praticamente uma “caixa preta”
 - Requer número de dados suficiente para separar amostra em dados de treinamento e dados de teste da RNA
 - Requer eventualmente capacidade computacional intensa.

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

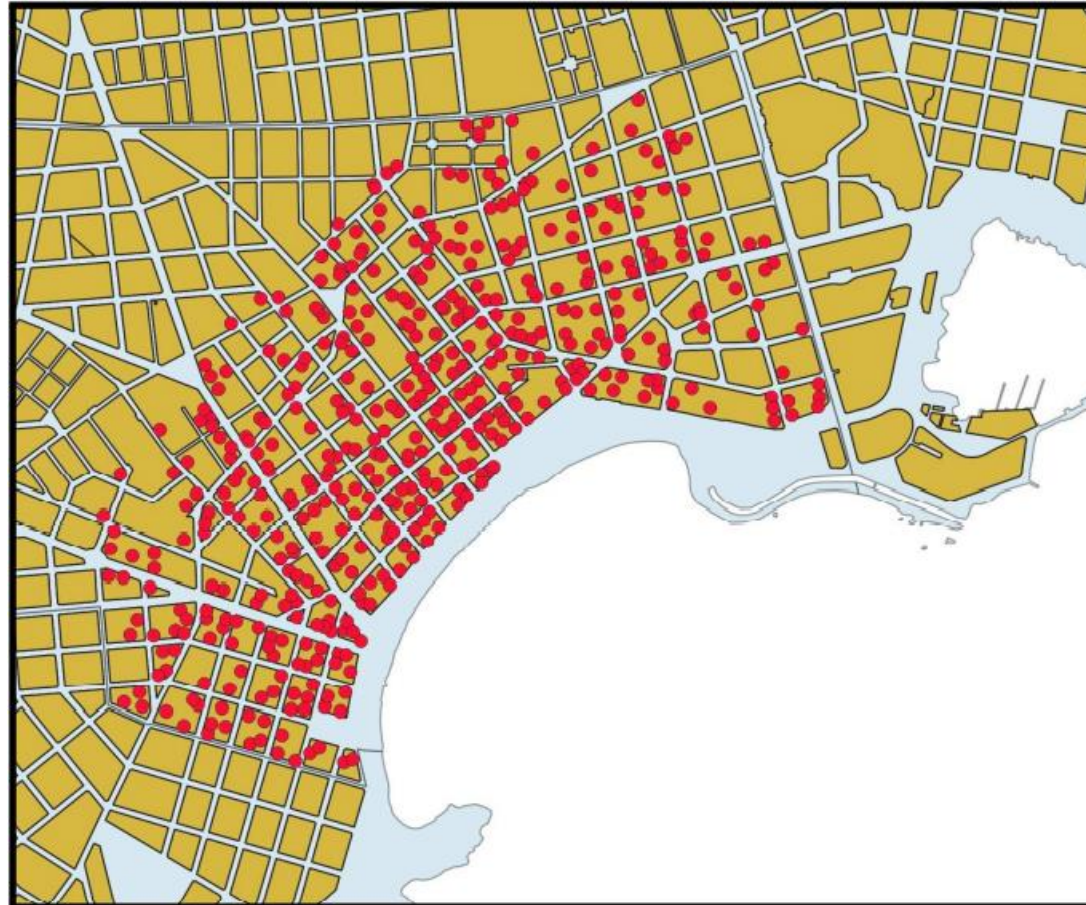
Area de estudo: Pocitos, Montevideo, Uruguay



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Amostra selecionada: 279 apartamentos localizados até 700m da Rambla



Informações utilizadas

- Valor de Venda de Transações realizadas (em U\$S) (registradas na DGR);
- Variáveis descritivas (do Cadastro Nacional - DNC)
 - Area (m2);
 - Categoria construtiva (de 1 Muito Economico até 9 Muito Boa);
 - Estado de conservação (de 1 Ruim até 7 Muito Bom)
 - Idade da construção (anos);
 - Andar (m);
- Localización distancias a polos de valorização (por geoprocessamento):
 - Localização do imóvel (coordenadas UTM);
 - Distância a Avenida Rambla (m); (selecionado por buffer até 700m)
 - Distância Shopping (m);

1. Inferência estatística

- Variável dependente: Valor Unitário (U\$S/m²)
- Variáveis Categoria e Estado de Conservação foram testadas como ordinais e também como variáveis dicotômicas separando os diferentes códigos
- Variável Idade testada em forma linear e polinômica 2º grau
- Variável Distancia a Rambla testada como:
 - Variável continua linear
 - Variável polinômica 2º grau
 - Variáveis dicotômicas: intervalos de 100m
- Interações entre as variáveis independentes

Resultados Regressão Múltipla

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 de 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	
24						

Resultados

- Modelo coerente:
 - todas as variáveis explicativas apresentam sinais coerentes e
 - níveis de significância aceitáveis exceto a variável Distancia a Shopping;
- Baixo poder explicativo do modelo: $R^2 = 0,29$
- Outros fatores de localização deveriam ser considerados para obter um modelo mais explicativo e representativo (ruas comerciais, avenidas).
- Variáveis explicativas do valor faltantes (vagas garagem, projeto arquitetônico)

2. Modelos de Regressão Espacial

- Software: Geoda.
- Efeitos de vizinhança: testadas matrizes de pesos espaciais para 300m, 400m e 500m de distância entre observações
- Testados modelos de regressão espacial da variável dependente e modelo do erro espacial.
- Testada como variável explicativa o produto $W \cdot \text{Categoria}$ (a variável Categoria espacialmente defasada) representa o valor médio da variável Categoria para os imóveis localizados na vizinhança dos imóveis.

Resultado Modelo do Erro Espacial

REGRESSION

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set : Pocitos-Geoda 379 datos
Spatial Weight : Pocitos-Geoda 379 datos 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

Resultados

- Todas as variáveis independentes significativas e sinais coerentes, inclusive Distância a Shopping.
- Variável [W*Cat] resulta altamente significativa, captando o efeito espacial da categoria na vizinhança dos imóveis.
- O poder explicativo do modelo é melhor que o de Inferência Estatística tradicional; porém continua sendo baixo: $R^2 = 0,36$
- Próximo modelo espacial: uso de geoestatística (interpolação por krigagem)

2b. Modelo espacial combinado: regressão espacial e krigagem

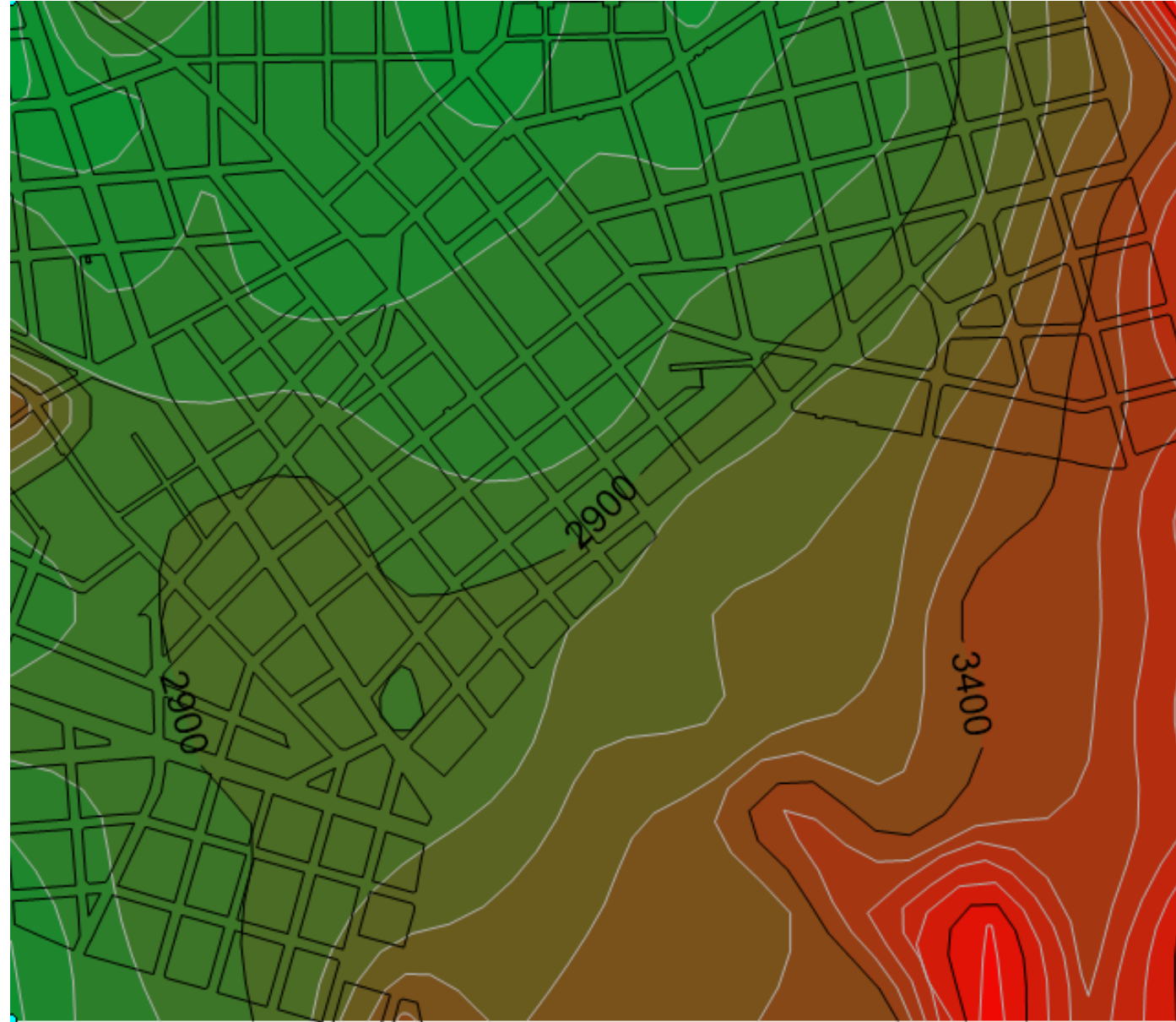
- Para utilizar o método geoestatístico de krigagem é necessário homogeneizar previamente o VU respeito a todas as variáveis construtivas (não localizativas)
- Para homogeneizar VU é utilizado o Modelo Espacial do Erro, considerando como variáveis explicativas só as construtivas sem considerar nenhuma variável de localização.
- O valor VU homogeneizado é interpolado por krigagem com o qual se obtém um índice que representa o Valor de Localização VL.
- A nova variável VL é utilizada no Modelo Clássico de Regressão.

Superfície e Curvas de isovalor VL por krigagem



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



Modelo Clássico com a variável VL

REGRESSION

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : Datos_Pocitos_379_VL pruebas varias Wcons

Dependent Variable : ValorUnitario

Number of Observations: 379

Mean dependent var : 2689,38 Number of Variables : 7

S.D. dependent var : 584,642 Degrees of Freedom : 372

R-squared : 0,360070 F-statistic : 34,8856

Adjusted R-squared : 0,349749 Prob(F-statistic) : 2,02599e-33

Sum squared residual: 8,28996e+07 Log likelihood : -2867,79

Sigma-square : 222848 Akaike info criterion : 5749,59

S.E. of regression : 472,068 Schwarz criterion : 5777,15

Sigma-square ML : 218732

S.E of regression ML: 467,688

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-1314,2	580,768	-2,26287	0,02422
Superficie	-7,0279	0,665258	-10,5642	0,00000
Edad	-4,04395	1,33827	-3,02178	0,00269
Alturas	11,1091	2,98711	3,71901	0,00023
EstadoCons	63,442	21,5423	2,945	0,00343
W_cat_300	434,596	112,763	3,85408	0,00014
VL	0,698246	0,178482	3,91215	0,00011

Resultados dos Modelos Espaciais

- Modelo de Regressão Espacial do Erro: $R^2=0,36$
- Modelo Clássico com VL (geoestatística): $R^2=0,36$
- Os modelos espaciais tem coeficientes coerentes e com níveis de significância aceitáveis.
- Especialmente a variável de localização VL resulta altamente significativa e coerente com a valorização no conjunto da região.
- Os modelos espaciais melhoraram levemente o poder explicativo perante o Modelo Clássico puro.
- Porém, continuam apresentando um baixo poder explicativo da variável VU.

3. Redes Neurais Artificiais RNA

- Estruturas de RNA utilizadas nessa análise:
 - 1 Capa de entrada com 7 variáveis ou neurônios artificiais: Area, Categoria, Estado de Conservação, Idade, Altura, Distância a Rambla e Distância a Shopping.
 - 3 capas ocultas que foram testadas com diferentes números de neurônios artificiais, variando de 10 até 30.
 - 1 Capa de saída com a estimação do Valor Unitário.

3. Redes Neurais Artificiais RNA

- A amostra total é separada em duas partes: uma parte para treinar a RNA e outra parte para testar o grau de aprendizado. Foram testadas duas formas diferentes de separar a amostra: 70/30 e 80/20;
- Função de ativação RELU que tem mostrado bons resultados para em pesquisas anteriores para estudos de estimação de valores;
- Software: programação no site Google Colabs (livre)

Estruturas de RNA testadas

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

Resultados de R^2 para as 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

Resultados das RNA

- O poder explicativo dos modelos das RNA se mostra superior em todos os casos analisados que o Modelo Clássico de Regressão e Modelos Espaciais.
- R^2 equivalente das RNA foi desde 0,54 (Modelo 1) até 0,85 (Modelo 7)

Comparação dos métodos

- Parâmetros utilizados para comparação entre modelos:
 - 1. R^2
 - 2. Análise dos resíduos: se utiliza o Erro Percentual (Resíduo/VU) em valor absoluto
 - Valor da Média
 - Valor do Desvio padrão
- Entre os modelos de RNA testados foi escolhido para comparação o Modelo 7.

Resíduos relativos (em valor absoluto) Médias e desvios padrão

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

Comparativo dos erros relativos dos métodos em relação a diferentes 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

Analizando os erros Percentuais em relação a Diferentes faixas de valor Para algumas variáveis, O resultado é:

Os erros percentuais das Estimções por RNA Se mostraram Consistentemente melhores Em todas as faixas de valor

Variáveis analisadas:

- Distância a Rambla
- Area do Imóvel
- Categoria

Resultados

- Os resultados de R^2 e grau de ajuste dos valores estimados aos valores observados nos modelos de RNA foi claramente superior que nos modelos Clássico e Espacial.
- Ambos modelos inferenciais apresentaram valores menores nos R^2 , indicando possíveis problemas de especificação dos modelos: provavelmente pela falta de análise de interações não lineares entre variáveis construtivas e entre elas e variáveis de localização.
- Testes realizados a posteriori, utilizando o Modelo Clássico e buscando interações significativas entre variáveis explicativas, mostraram significância de algumas interações, mas aumentaram pouco o poder explicativo do modelo.

Exemplos de Interações significativas entre variáveis explicativas

-Cat
-Cat²
-W_Cat
-W_Cat * VL

-Altura
-Primcua
*Altura

-VL*DR50m

>>10/09/2023 22:09:16

REGRESSION

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : Datos_Pocitos_379_VL pruebas varias Wcons

Dependent Variable : ValorUnitario

Number of Observations: 379

Mean dependent var : 2689,38 Number of Variables : 12

S.D. dependent var : 584,642 Degrees of Freedom : 367

R-squared : 0,383481 F-statistic : 20,7525

Adjusted R-squared : 0,365002 Prob(F-statistic) : 1,24666e-32

Sum squared residual: 7,98668e+07 Log likelihood : -2860,73

Sigma-square : 217621 Akaike info criterion : 5745,46

S.E. of regression : 466,498 Schwarz criterion : 5792,71

Sigma-square ML : 210730

S.E of regression ML: 459,054

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-899,033	1270,19	-0,707796	0,47952
Superficie	-11,7522	2,37126	-4,9561	0,00000
Sup2	0,0226761	0,0114021	1,98876	0,04747
Categoria	663,536	347,307	1,91052	0,05685
Cat2	-55,6545	29,3237	-1,89794	0,05849
Edad	-4,20935	1,34091	-3,13918	0,00183
Alturas	9,95505	3,19854	3,11237	0,00200
Cons2	6,78797	2,37304	2,86045	0,00447
W_cat_300	140,931	182,138	0,773755	0,43957
W_Cat300xVL	0,1021	0,0324769	3,14378	0,00180
Primcuadra_alt	6,41315	5,60507	1,14417	0,25330
VLxDR50	0,037173	0,0350286	1,06122	0,29929

Conclusões e recomendações

- Os resultados obtidos com as RNA se mostraram superiores aos obtidos pelos métodos inferenciais tradicionais, com erros de estimação percentuais bastante menores.
- Isto pode ser especialmente relevante no caso de Avaliações em Massa, onde a heterogeneidade de dados e limitações nas variáveis dificulta obter modelos inferenciais suficientemente explicativos.
- A hipótese que as RNA conseguem aprender padrões complexos e não lineares de interação entre as variáveis parece ser altamente plausível, precisando contudo de novas pesquisas

Conclusões e recomendações

- Como desvantagem, ao contrário do Modelo Clássico, as RNA não permitem analisar individualmente a influência de cada variável no valor final.
- Porém, os resultados claramente superiores obtidos com as RNA ameritam novas pesquisas, assim como analisar diferentes arquiteturas de RNA em relação as aqui apresentadas.
- Como sugestão para próximos trabalhos, usar a RNA já treinada para estimar o valor para diversos tipos de imóveis para as diferentes faces de quadra da área de estudo.

Reflexões finais

- A chamada Inteligência Artificial é uma revolução em pleno e acelerado desenvolvimento, que vem sendo aplicada cada vez mais em diferentes áreas e aspectos da nossas atividades sociais e profissionais;
- As RNA são apenas uma das várias ferramentas da IA, existem outras que também podem ser estudadas e testadas em Engenharia de Avaliações.
- Os profissionais de Engenharia de Avaliações precisamos estar atentos e acompanhando aos novos desenvolvimentos em IA, para poder utilizar aqueles que podem colaborar e melhorar a nossa prática avaliatória.

