



# 4º SEMINÁRIO NACIONAL DESAPROPRIAÇÃO E REASSENTAMENTO

2025



**DNIT**  
DEPARTAMENTO  
NACIONAL DE  
INFRAESTRUTURA  
DE TRANSPORTES

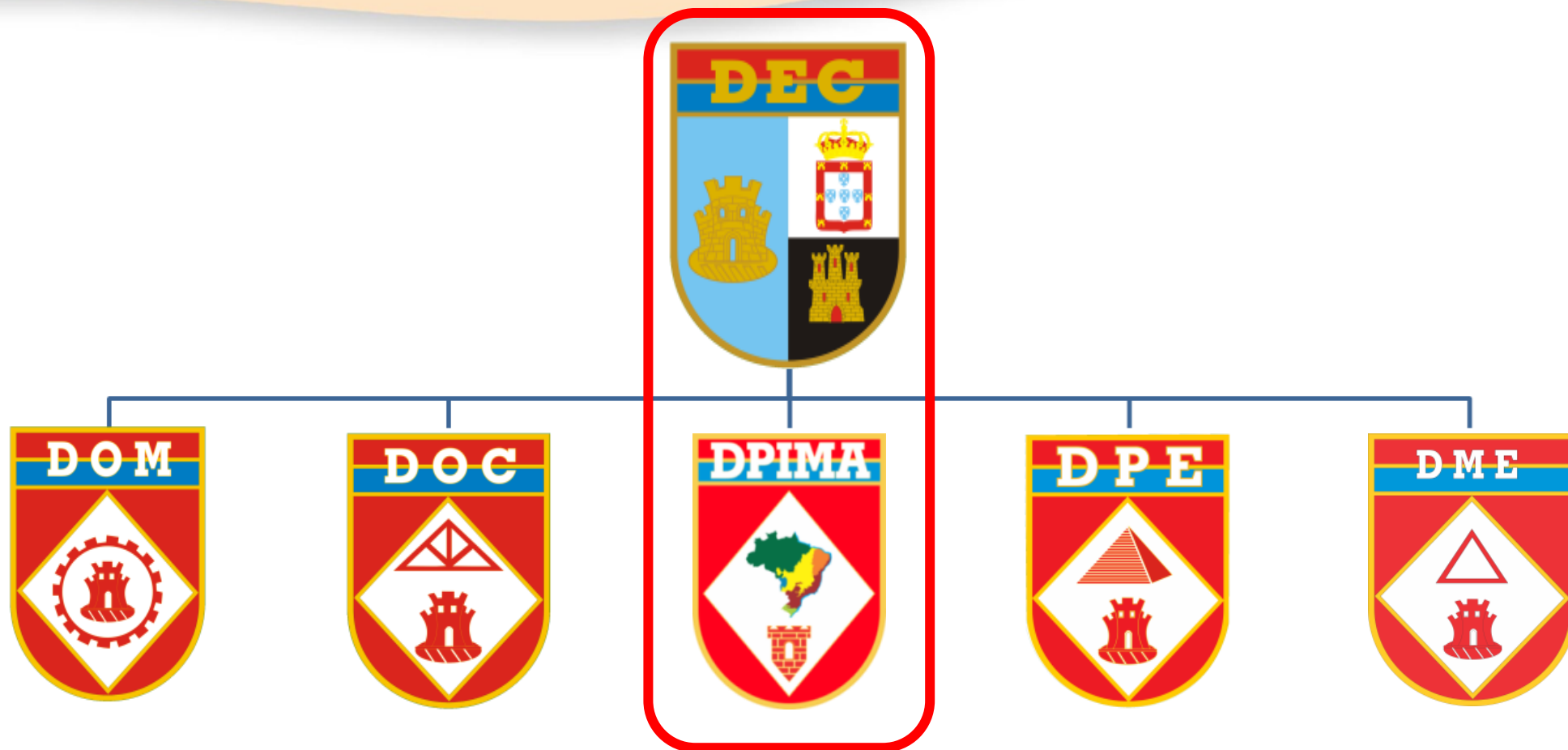
MINISTÉRIO DOS  
TRANSPORTES



# **AVALIAÇÃO DE IMÓVEIS URBANOS E RURAIS COM USO DE APRENDIZADO DE MÁQUINA**

Estudos de Caso

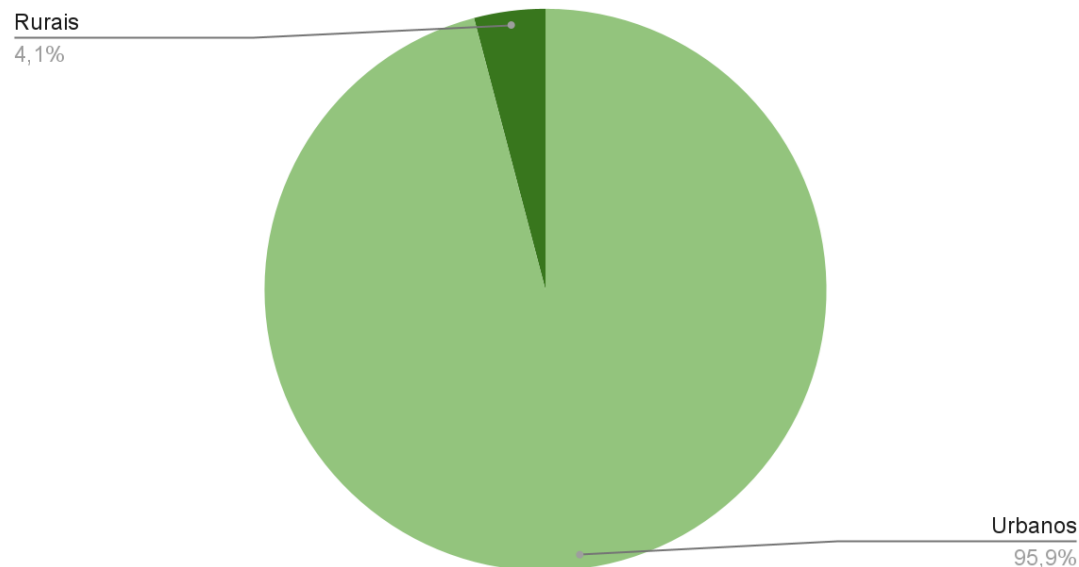
# CONTEXTO



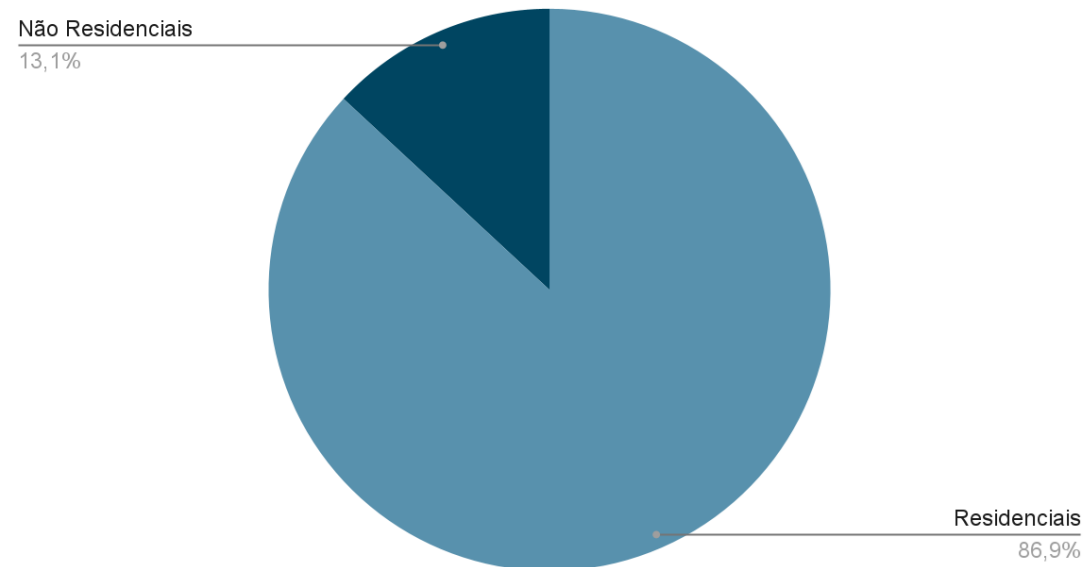
# CONTEXTO

- O EB possui **mais de 20 mil** parcelas imobiliárias da União sob sua administração. A União, **mais de 700 mil**.

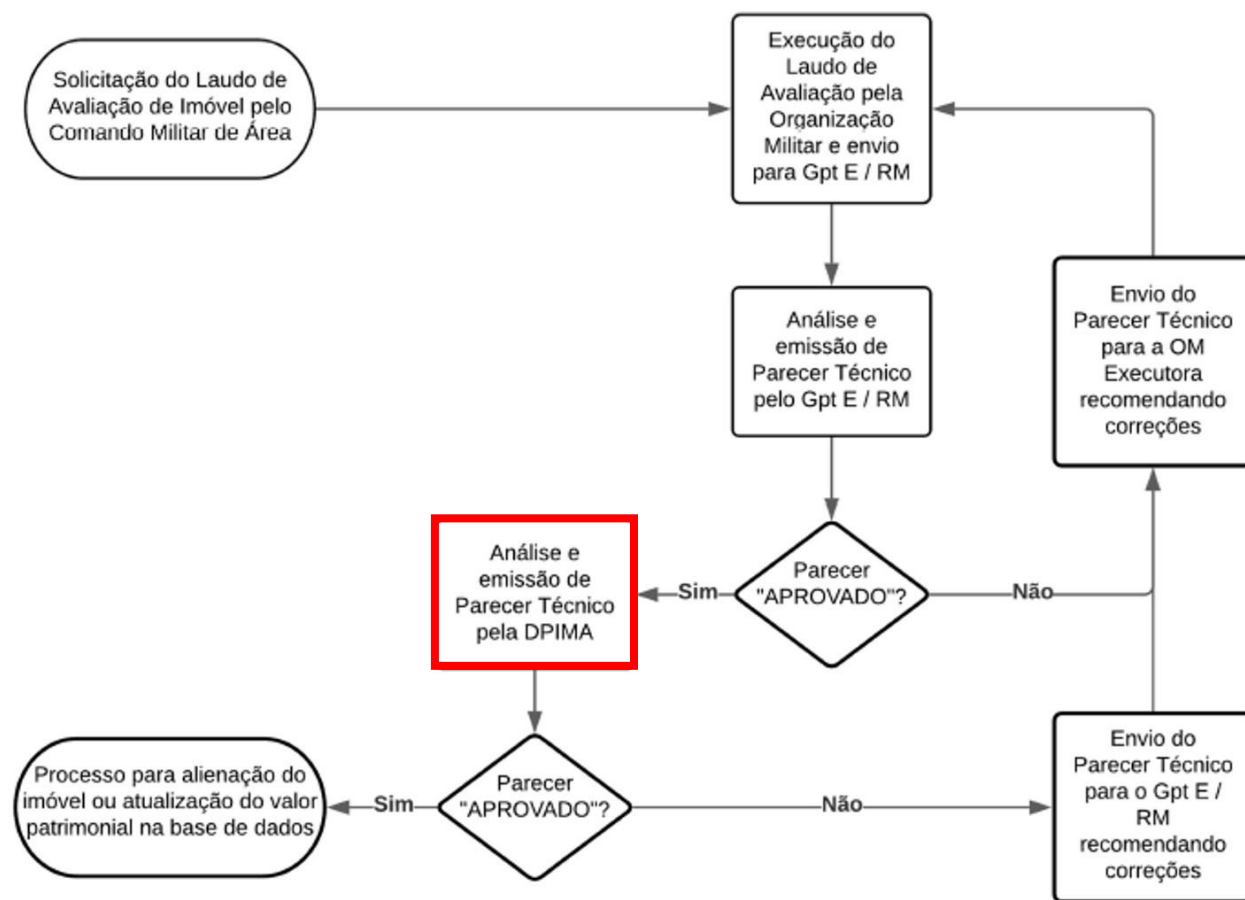
População (aprox. 21500 unidades imobiliárias)



População (aprox. 21500 unidades imobiliárias)



# CONTEXTO





# CONTEXTO

- Como propor modelos capazes de construir **estimativas de valor** dos imóveis administrados pelo EB com razoável aptidão à generalização?

Trabalho	$R^2 > 57\%$	Abrangência nível País	Modelagem linear	Modelagem não linear	Ajuste de hiperparâmetros	Componentes espaciais	Mais de uma vocação de imóvel	Interpretabilidade
Dantas et al. 2010 [3]	✓		✓			✓		✓
Park e Bae 2015 [4]	✓		✓	✓	✓			
Kiely e Bastian 2020 [1]			✓	✓	✓	✓	✓	
Hagenauer e Helbich 2022 [2]		✓	✓	✓		✓		
Tchuente e Nyawa 2022 [5]	✓	✓	✓	✓	✓	✓		
Solução proposta	✓	✓	✓	✓	✓	✓	✓	✓

# BASES DE DADOS



Fonte: EB /  
SPU / IBGE /  
GitHub Prof.  
Dr. Bernardo  
Furtado.

Fonte	Qtde de instâncias urbanas com valor conhecido	Participação relativa
EB	258	5,6%
SPU	4340	94,4%
EB e SPU	4598	100,0%

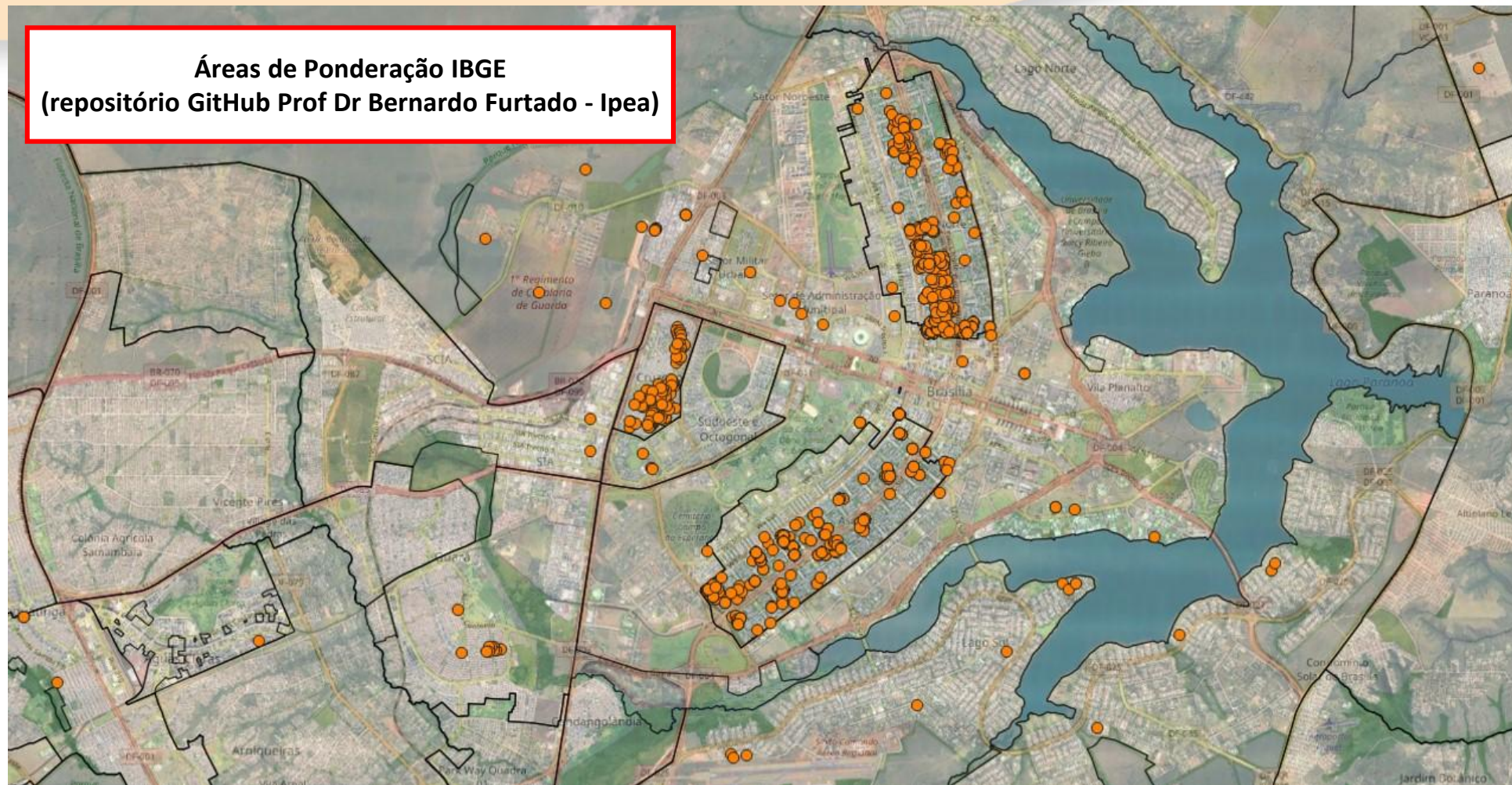
# METODOLOGIA

- **Pré-processamento:** geocodificação, união dos arquivos gerados (incluindo elementos geográficos IBGE e Google), **codificação binária** de variáveis categóricas nominais.
- **Análise exploratória:** verificação gráfica e seleção de atributos (com SEM).
- Aplicação de **testes de normalidade** e de **transformações matemáticas** sobre os atributos.
- Validação cruzada: **10 folds**.
- **Testes intradomínio:** treinamento e teste em base **EB/SPU (Cenário A)**.
- **Testes intradomínio:** treinamento em base **EB/SPU** e teste em base **EB (Cenário B)**.
- **Comparação** dos modelos entre si (lineares e não lineares).
- Métricas de avaliação utilizadas: **coeficiente de determinação ( $R^2$ )** e **raiz do erro quadrático médio ( $RMSE$ )**.
- Interpretabilidade com **valor de Shapley** (Teoria dos Jogos).



# EXPERIMENTOS REALIZADOS

**Áreas de Ponderação IBGE**  
(repositório GitHub Prof Dr Bernardo Furtado - Ipea)



# EXPERIMENTOS REALIZADOS

## IMÓVEIS URBANOS

data zap<sup>+</sup>



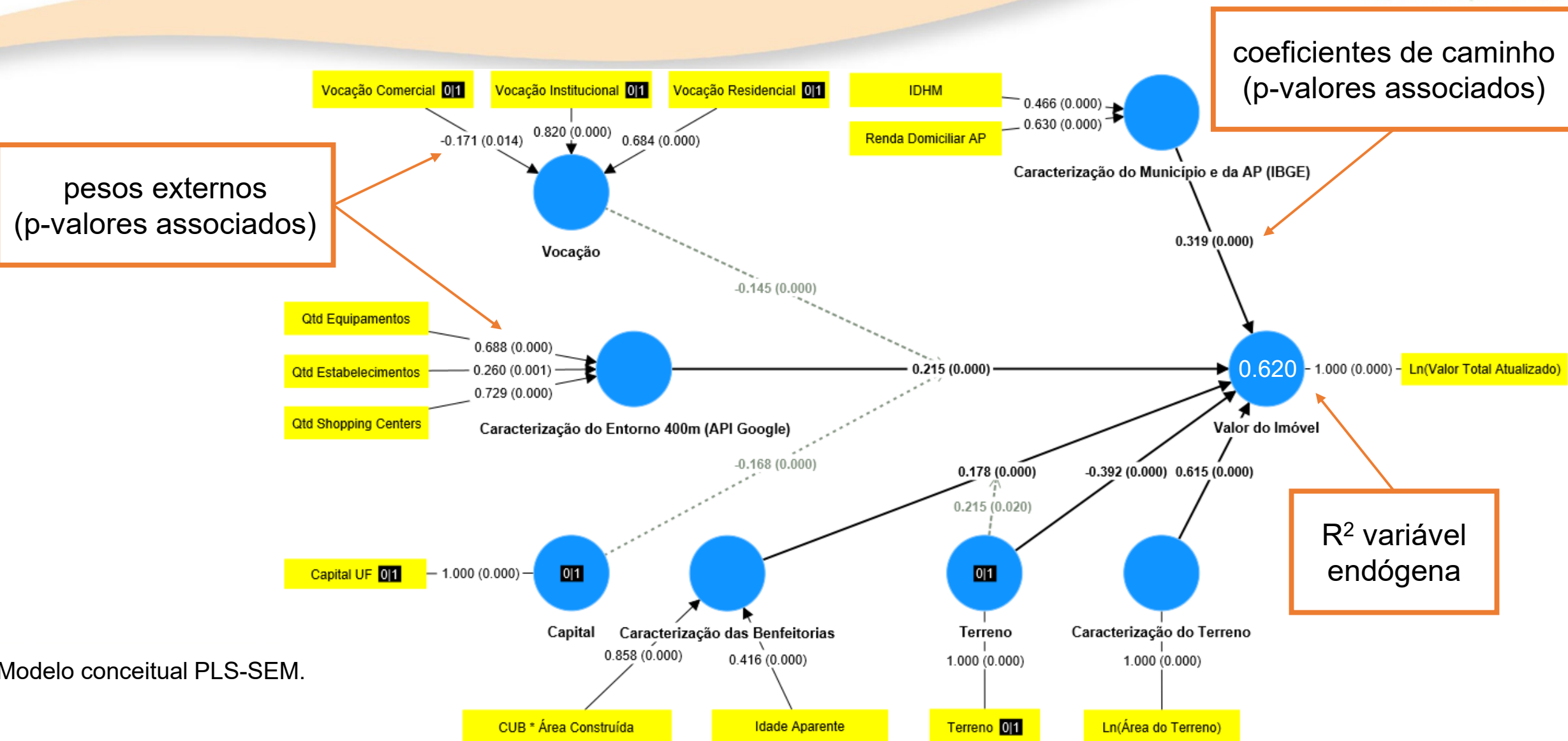
Google Places API

**Pontos de interesse**  
em raio de 400 m de cada imóvel:

- **agregado equipamentos**
- atrações turísticas
- delegacias
- escolas
- hospitais
- parques
- universidades
- aeroportos
- estações de metrô
- estações de trem
- estações de VLT
- paradas de ônibus
- **agregado estabelecimentos**
- cafeterias
- lojas
- padarias
- restaurantes
- **shopping centers**
- supermercados



# EXPERIMENTOS REALIZADOS

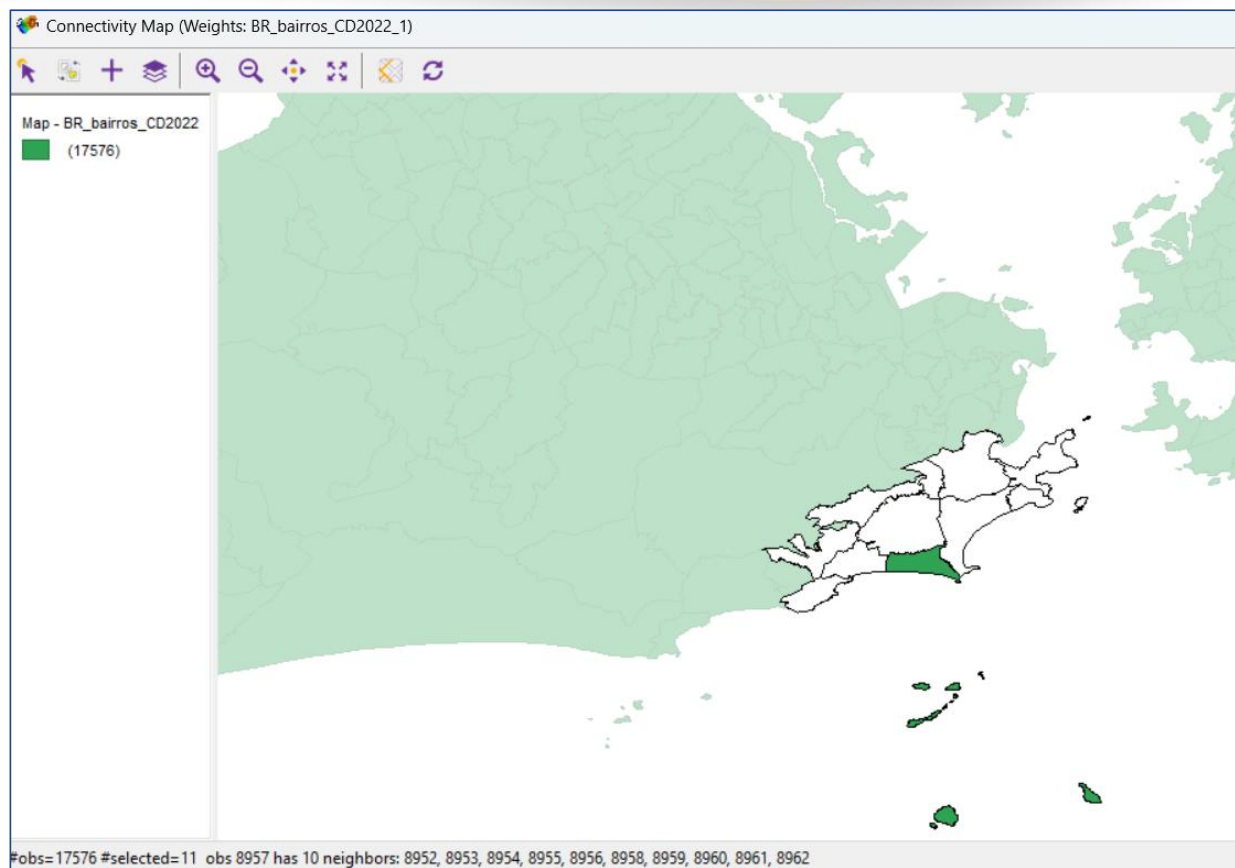


# EXPERIMENTOS REALIZADOS

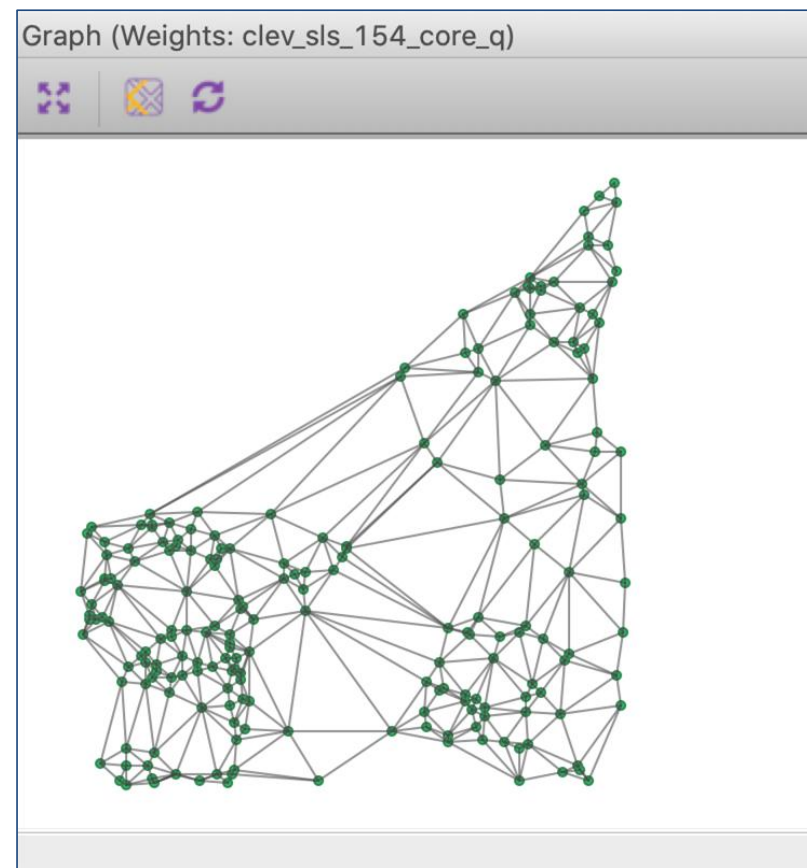
- Atributos **urbanos** efetivamente utilizados em cada nível de modelagem:

Variável	Nível de Modelagem
<i>Capital UF</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>
<i>Vocação do Imóvel</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>
<i>Índice de Desenvolvimento Humano Municipal</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>
<i>Ln(Área do Terreno)</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>
<i>Custo Unitário Básico * Área Construída</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>
<i>Idade Aparente</i>	U <sub>2</sub> , U <sub>3</sub>
<i>Renda Domiciliar AP</i>	U <sub>3</sub>
<i>Pontos de Interesse API Google Places</i>	U <sub>3</sub>
<i>Ln(Valor Total Atualizado)</i>	U <sub>1</sub> , U <sub>2</sub> , U <sub>3</sub>

# EXPERIMENTOS REALIZADOS



Mapa de conectividade construído para a região do Rio de Janeiro.



Grafo utilizado para construção de matrizes de vizinhança.



# EXPERIMENTOS REALIZADOS

## IMÓVEIS URBANOS

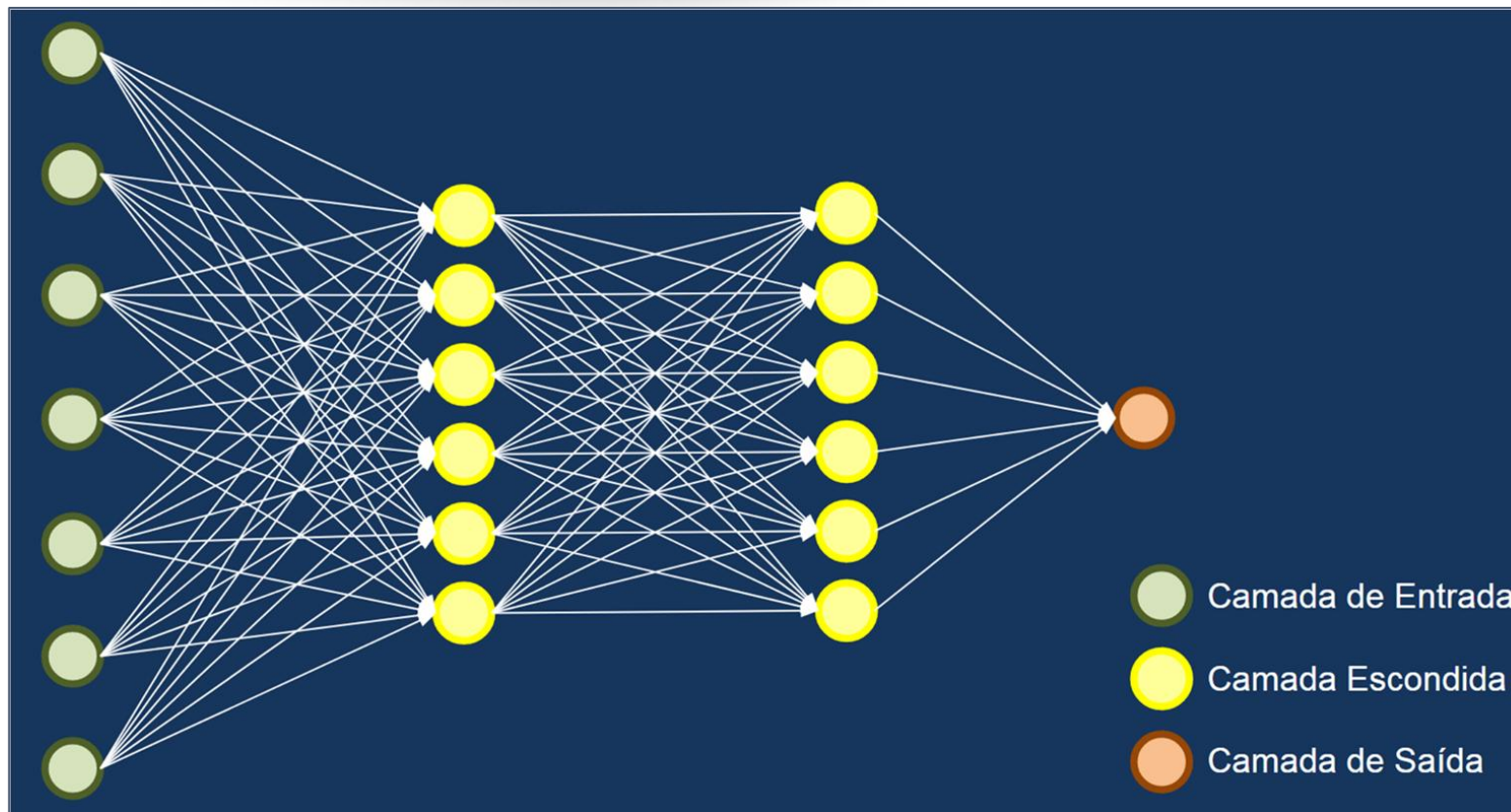
Parâmetro	<i>SGDRegressor</i> linear	<i>ANN MLPRegressor</i> não linear	<i>XGBRegressor</i> não linear
Fração de Treinamento	53,3%	53,3%	53,3%
Fração de Validação	13,3%	13,3	13,3%
Fração de Teste	33,3%	33,3	33,3%
Tipo de taxa de aprendizagem	constante	constante	constante
Valor da taxa de aprendizagem	0,0001	0,01	0,3
Função de custo	erro quadrático	erro quadrático	erro quadrático
Termo de regularização	$L2$	$L2$ (multiplicador $\alpha = 0,1$ )	$L1$ ( $\alpha = 5$ ) e $L2$ ( $\lambda = 1$ )
Arquitetura básica	não se aplica	(5, 5) <sup>a</sup> em $U_1$ , (6, 6) em $U_2$ e (8, 8) em $U_3$	$max\_depth = 5$

<sup>a</sup> camadas ocultas da rede neural artificial



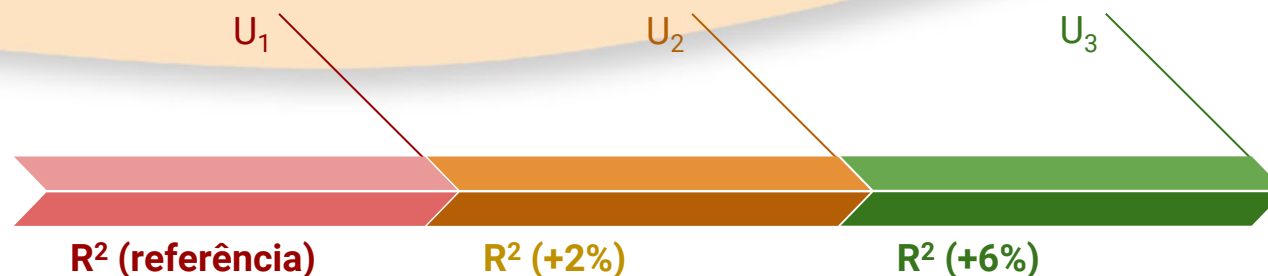
Treinamento - Validação - Teste

# EXPERIMENTOS REALIZADOS



Arquitetura de rede neural artificial (RNA) construída pela DPIMA.

# RESULTADOS OBTIDOS



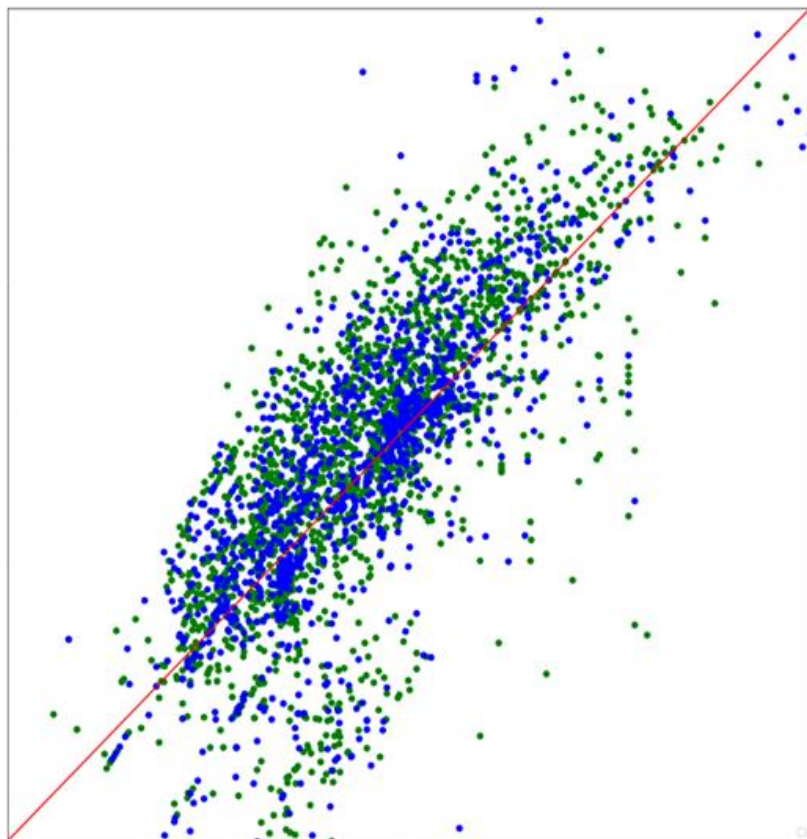
Modelos U <sub>3</sub> (Cenário)	$R^2$ médio 10 <i>folds</i>	dp $R^2$ 10 <i>folds</i>	$R^2$ <i>train</i>	$R^2$ <i>test</i>	<i>RMSE</i> <i>train</i>	<i>RMSE</i> <i>test</i>
OLS (A)	NA <sup>a</sup>	NA <sup>a</sup>	61,9%	52,3%	1,28	1,73
Regressão Espacial (A)	NA <sup>a</sup>	NA <sup>a</sup>	63,4%	55,7%	1,26	1,39
<i>SGDRegressor</i> linear (A)	61,2%	0,047	61,9%	61,1%	1,27	1,31
<i>SGDRegressor</i> linear (B)	60,7%	0,044	61,7%	50,3%	1,29	1,35
ANN não linear (A)	73,8%	0,048	74,8%	69,3%	1,04	1,17
ANN não linear (B)	71,8%	0,041	73,8%	71,3%	1,06	1,02
<i>XGBRegressor</i> não linear (A)	86,2%	0,019	95,7%	83,6%	0,43	0,85
<i>XGBRegressor</i> não linear (B)	86,7%	0,035	94,8%	80,6%	0,47	0,84

<sup>a</sup> não se aplica

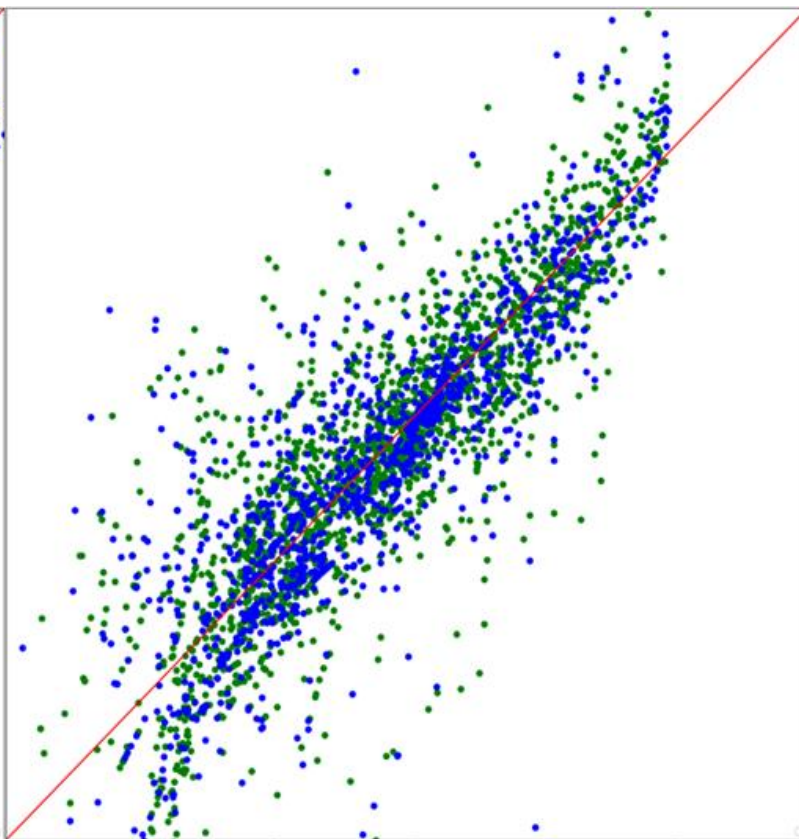


# RESULTADOS OBTIDOS

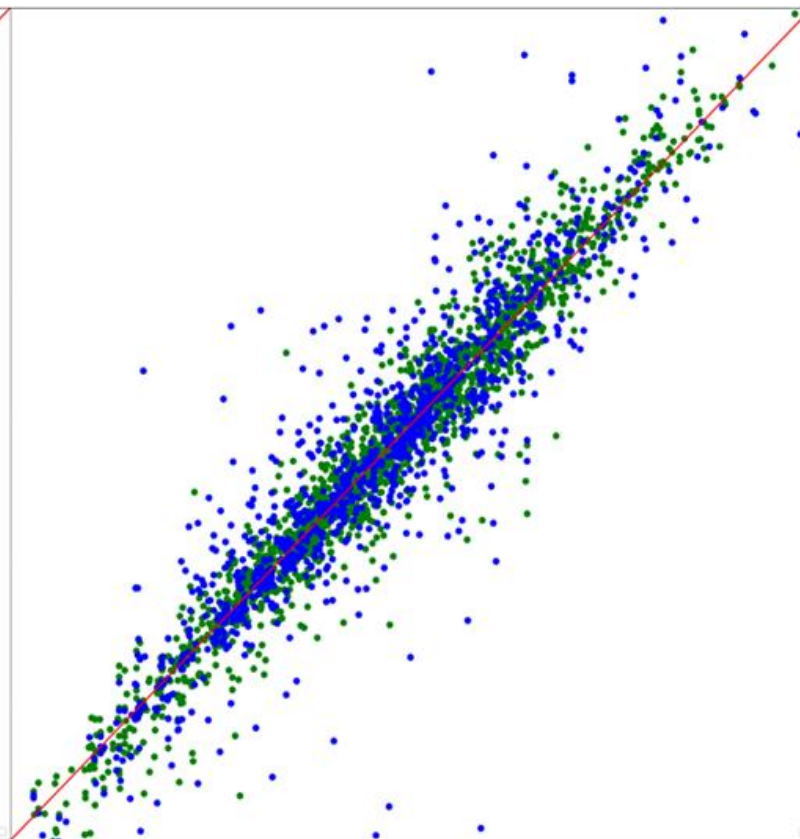
U<sub>3</sub> *SGDRegressor* (Cenário A)



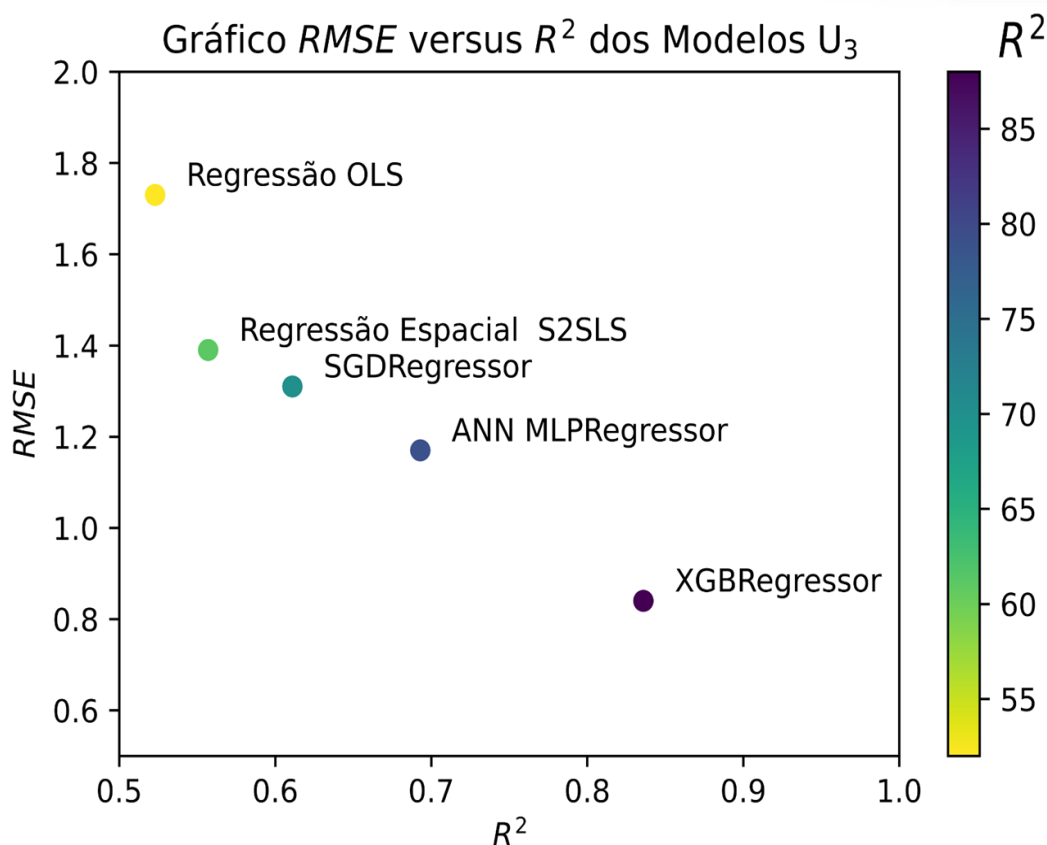
U<sub>3</sub> *ANN MLPRegressor* (Cenário A)



U<sub>3</sub> *XGBRegressor* (Cenário A)



# RESULTADOS OBTIDOS

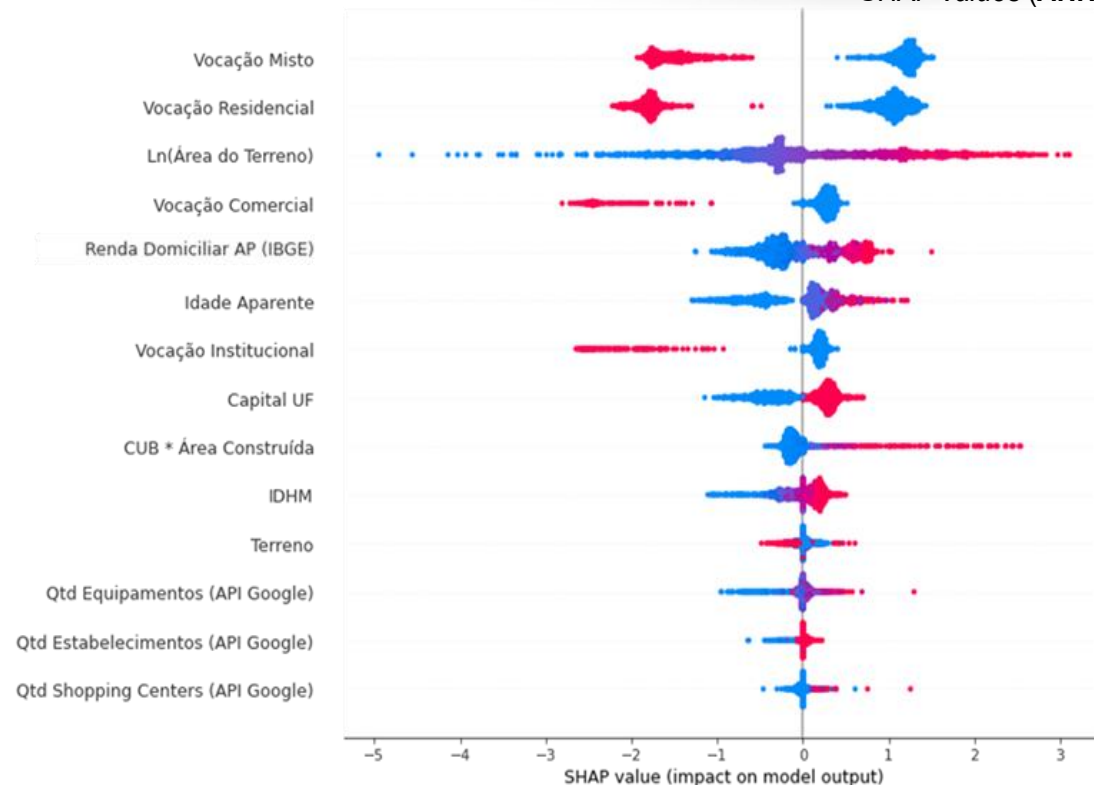


Modelo	Limite IC Inferior	Limite IC Superior	Amplitude IC
<i>ANN MPLRegressor</i> $U_1$	-51,0%	+109,0%	160,0%
<i>ANN MPLRegressor</i> $U_2$	-50,9%	+107,9%	158,8%
<i>ANN MPLRegressor</i> $U_3$	-51,1%	+98,8%	149,9%
<i>XGBRegressor</i> $U_1$	-43,9%	+79,8%	123,7%
<i>XGBRegressor</i> $U_2$	-43,8%	+79,6%	123,4%
<i>XGBRegressor</i> $U_3$	-41,9%	+73,3%	115,2%



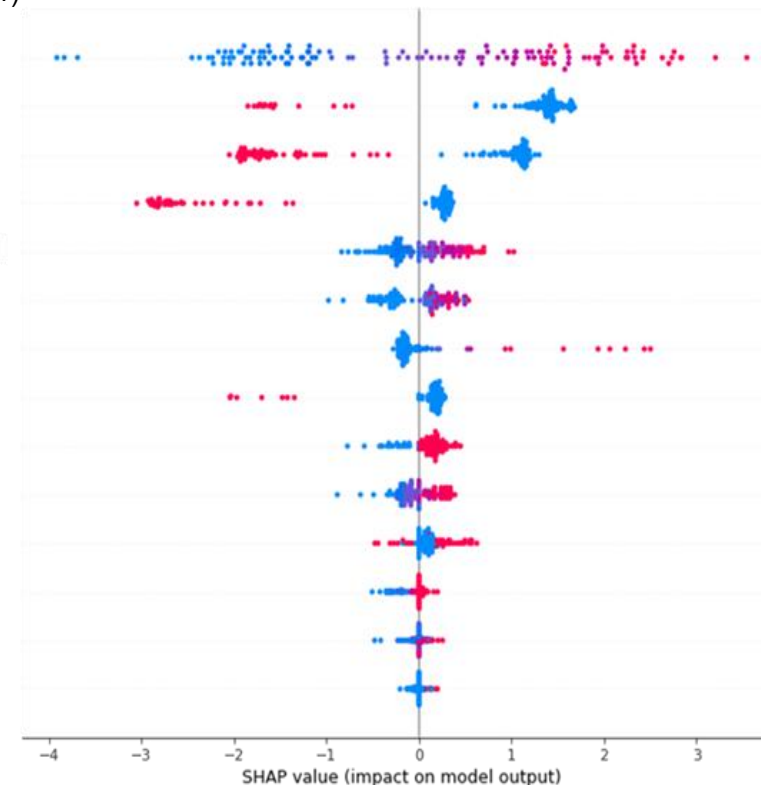
# RESULTADOS OBTIDOS

SHAP values (*ANN MLPRegressor* não linear)



Cenário A

menor



Cenário B

maior



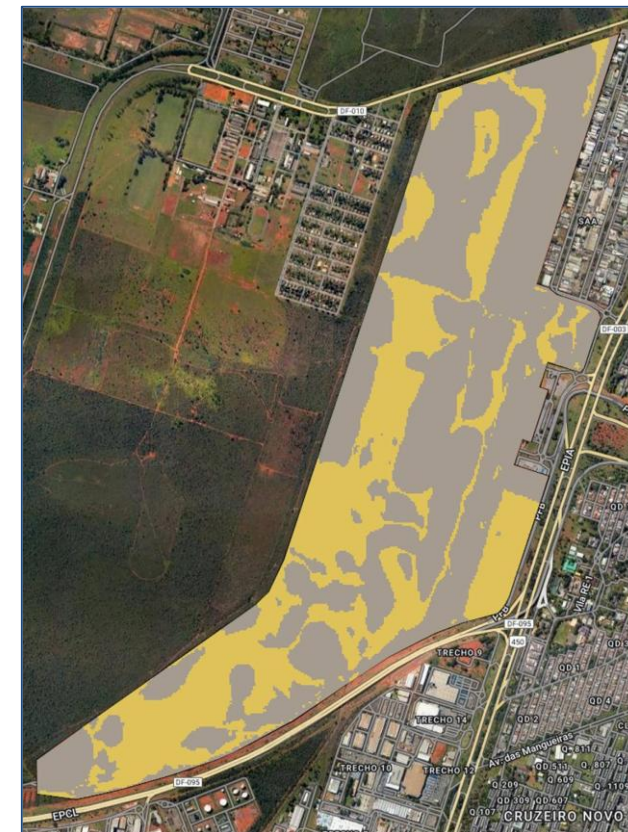
# RESULTADOS OBTIDOS



MAIO 2022



JULHO 2022

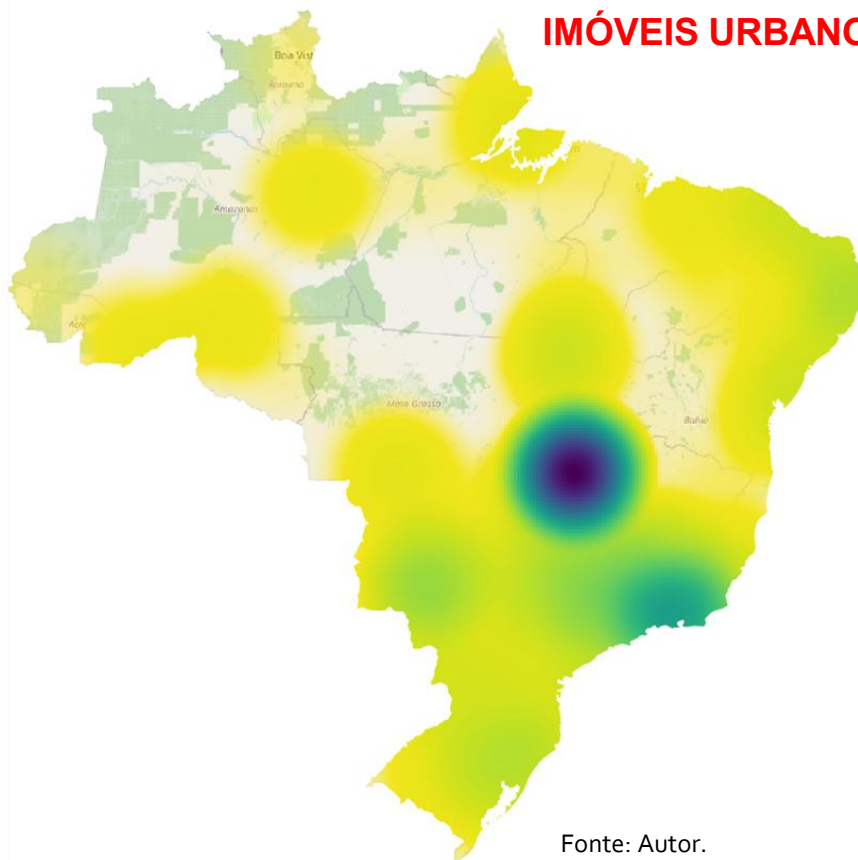


+ 2x



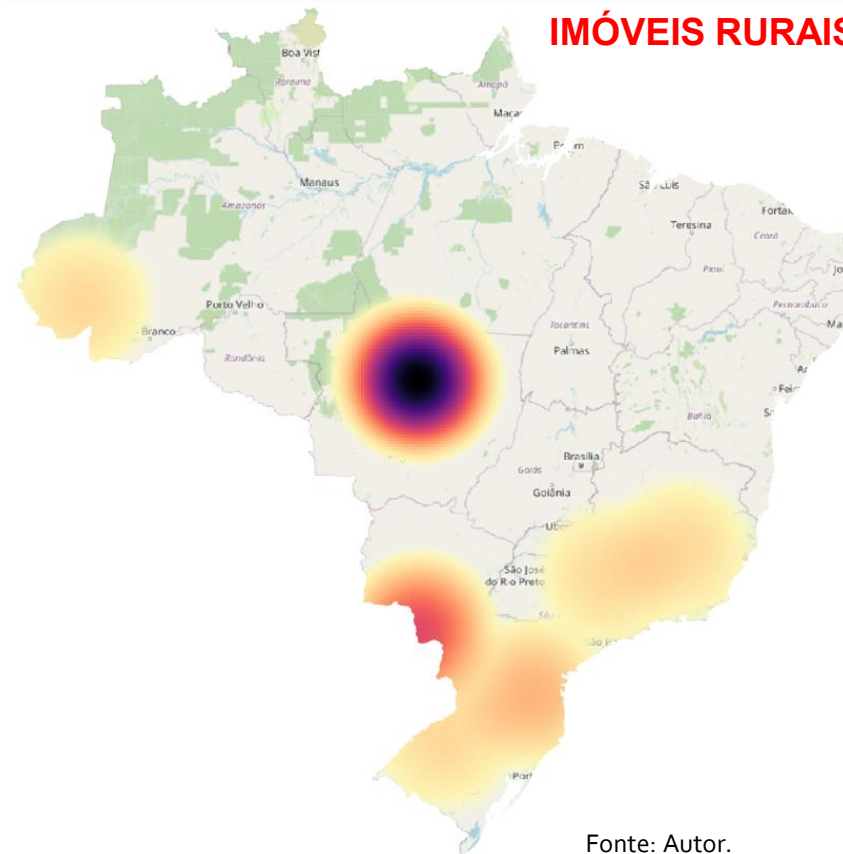
# CONSIDERAÇÕES FINAIS

**IMÓVEIS URBANOS**



Fonte: Autor.

**IMÓVEIS RURAIS**



Fonte: Autor.

# OBRIGADO!