

An aerial photograph of the Perth city skyline, featuring numerous high-rise buildings and skyscrapers. In the foreground, there is a large body of water with a curved bridge and a marina area with several boats. The text "PERTH HOUSE PRICING" is overlaid in the center of the image.

PERTH HOUSE PRICING



Alex Rodrigues de Oliveira

Leila Fabiola Ferreira

Mercado Imobiliário

Apresentou maior crescimento de lançamentos dos últimos 10 anos – Câmara Brasileira da Indústria da Construção (CBIC)

Crescimento de 12% no Brasil, em relação ao ano anterior – Associação de Dirigentes de Empresas do Mercado Imobiliário (Ademi)

O Valor Geral de Vendas (VGV) tem previsão de encerrar o ano em R\$ 99 bilhões – Associação de Dirigentes de Empresas do Mercado Imobiliário (Ademi)

Os preços dos imóveis australianos devem apresentar um pico de 10% no próximo ano – Commonwealth Bank

Objetivos Centrais



- Utilizar diferentes métodos de *ensemble* das *Machine Learnings* para melhorar as predições já existente sobre o preço dos imóveis (*House Pricing*).
- Comparar os métodos através das métricas de coeficiente de determinação (R^2) e erro médio quadrático (MSE) procurando obter os melhores resultados para os mesmos.
- Fazer uma aplicação (*deploy*) simples para os modelos treinados.



Sumário



ETL (Extração, Transformação e Carregamento) - Slide 5



EDA (Análise Exploratória de Dados) - Slide 7



Feature Engineering (Manipulação de *features*) - Slide 13



Modelagem das *Machine Learnings* - Slide 14



Deploy dos modelos (Aplicação) - Slide 22

ETL ([GitHub](#))

Dataset original - [Perth House Prices \(Kaggle\)](#)

	ADDRESS	SUBURB	PRICE	BEDROOMS	BATHROOMS	GARAGE	LAND_AREA	FLOOR_AREA	BUILD_YEAR	CBD_DIST	NEAREST_STN	NEAREST_STN_DIST	DATE_SOLD	POSTCODE	LATITUDE	LONGITUDE	NEAREST_SCH	NEAREST_SCH_DIST	NEAREST_SCH_RANK
0	1 Acorn Place	South Lake	565000	4	2	2.0	600	160	2003.0	18300	Cockburn Central Station	1800	09-2018\r	6164	-32.115900	115.842450	LAKELAND SENIOR HIGH SCHOOL	0.828339	NaN
1	1 Addis Way	Wandi	365000	3	2	2.0	351	139	2013.0	26900	Kwinana Station	4900	02-2019\r	6167	-32.193470	115.859553	ATWELL COLLEGE	5.524324	129.0
2	1 Ainsley Court	Camillo	287000	3	1	1.0	719	86	1979.0	22600	Challis Station	1900	06-2015\r	6111	-32.120578	115.993579	KELMSCOTT SENIOR HIGH SCHOOL	1.649178	113.0
3	1 Albert Street	Bellevue	255000	2	1	2.0	651	59	1953.0	17900	Midland Station	3600	07-2018\r	6056	-31.900547	116.038009	SWAN VIEW SENIOR HIGH SCHOOL	1.571401	NaN
4	1 Aman Place	Lockridge	325000	4	1	2.0	466	131	1998.0	11200	Bassendean Station	2000	11-2016\r	6054	-31.885790	115.947780	KIARA COLLEGE	1.514922	NaN

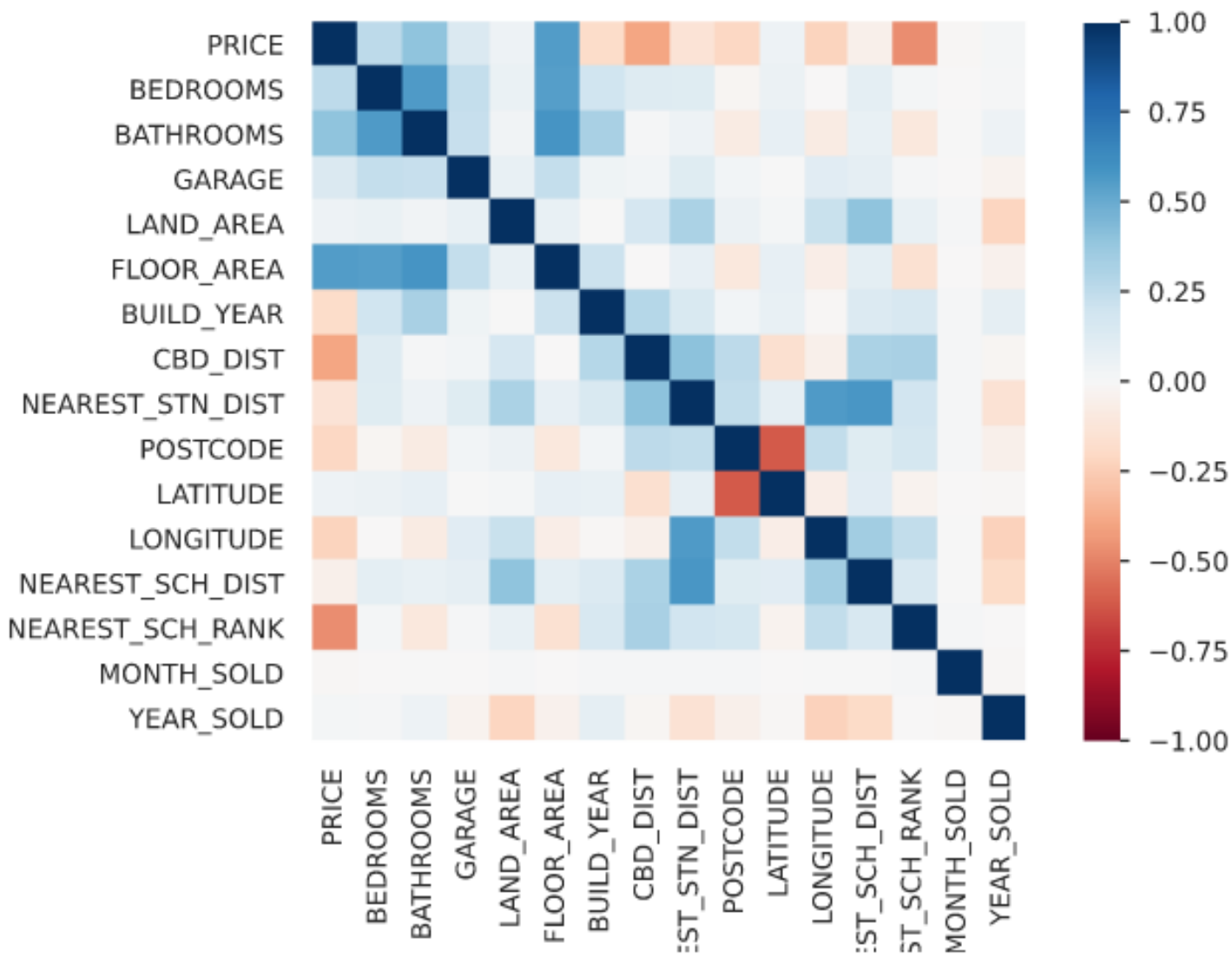
Transformações aplicadas

```
df['BUILD_YEAR'] = pd.to_datetime(df['BUILD_YEAR'], format='%Y')
df['BUILD_YEAR'] = pd.to_datetime(df['BUILD_YEAR']).dt.strftime('%Y')
df['DATE_SOLD'] = df['DATE_SOLD'].map(lambda x: x.rstrip('\r'))
df['MONTH_SOLD'] = pd.to_datetime(df.DATE_SOLD).dt.strftime('%m')
df['YEAR_SOLD'] = pd.to_datetime(df.DATE_SOLD).dt.strftime('%Y')
df.drop('DATE_SOLD', axis='columns', inplace=True)
df = df[df['BATHROOMS']<15]
df = df[df['GARAGE']<=10]
df = df[df['FLOOR_AREA']>=50]
df = df[df['LAND_AREA']<=900000]
```

ETL (GitHub)

Dataset transformado – ETL finalizado

	ADDRESS	SUBURB	PRICE	BEDROOMS	BATHROOMS	GARAGE	LAND_AREA	FLOOR_AREA	BUILD_YEAR	CBD_DIST	NEAREST_STN	NEAREST_STN_DIST	POSTCODE	LATITUDE	LONGITUDE	NEAREST_SCH	NEAREST_SCH_DIST	NEAREST_SCH_RANK	MONTH_SOLD	YEAR_SOLD
0	1 Addis Way	Wandi	365000	3	2	2.0	351	139	2013	26900	Kwinana Station	4900	6167	-32.193470	115.859553	ATWELL COLLEGE	5.524324	129.0	02	2019
1	1 Ainsley Court	Camillo	287000	3	1	1.0	719	86	1979	22600	Challis Station	1900	6111	-32.120578	115.993579	KELMSCOTT SENIOR HIGH SCHOOL	1.649178	113.0	06	2015
2	1 Arundel Street	Bayswater	685000	3	2	8.0	552	126	1999	5900	Bayswater Station	508	6053	-31.917880	115.907050	CHISHOLM CATHOLIC COLLEGE	0.936243	29.0	10	2019
3	1 Ashcott Gate	Butler	367500	3	2	2.0	398	158	2003	36300	Butler Station	2100	6036	-31.654280	115.702200	BUTLER COLLEGE	0.680843	39.0	11	2018
4	1 Ashendon Boulevard	Hammond Park	535000	4	2	4.0	704	247	2002	23100	Cockburn Central Station	3900	6164	-32.159590	115.849480	ATWELL COLLEGE	2.220643	129.0	07	2019



EDA ([GitHub](#))

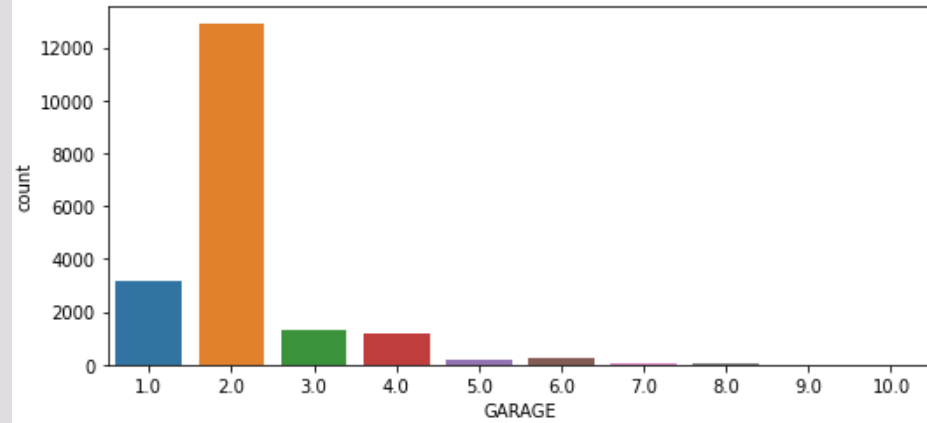
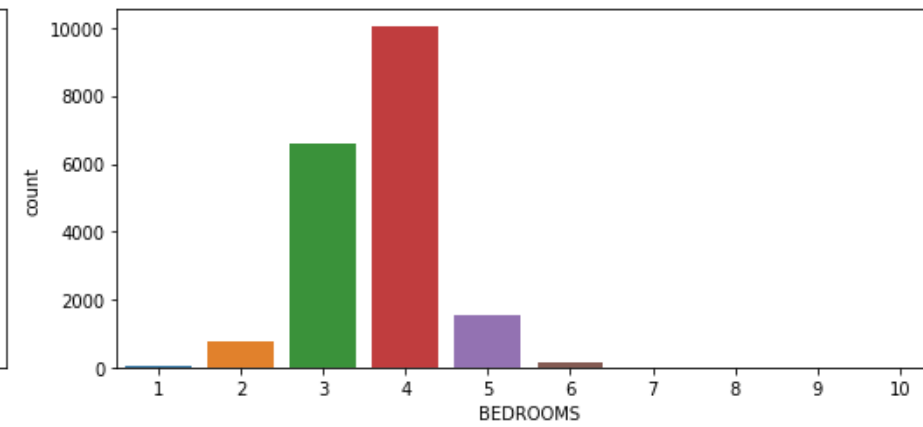
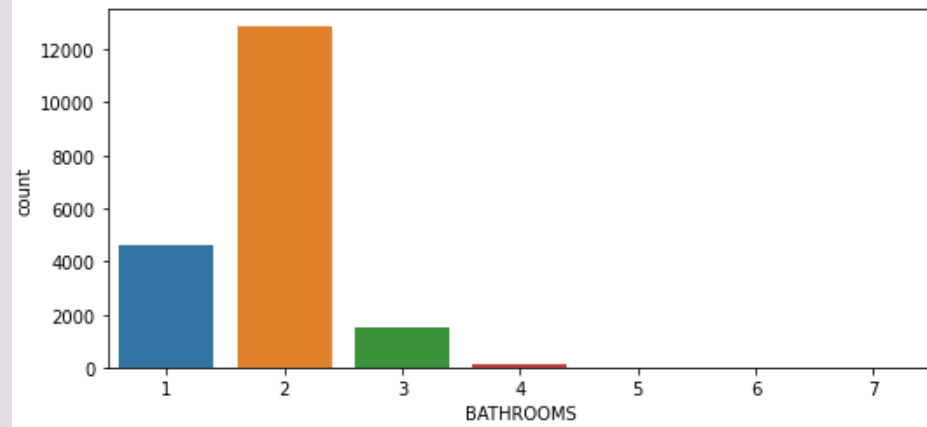
Correlação de Pearson

FLOOR_AREA e BATHROOMS
apresentam as maiores correlações positivas.

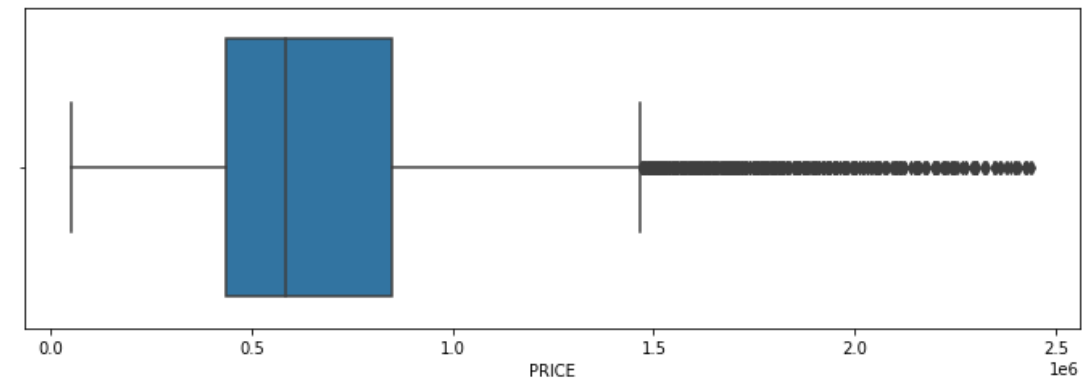
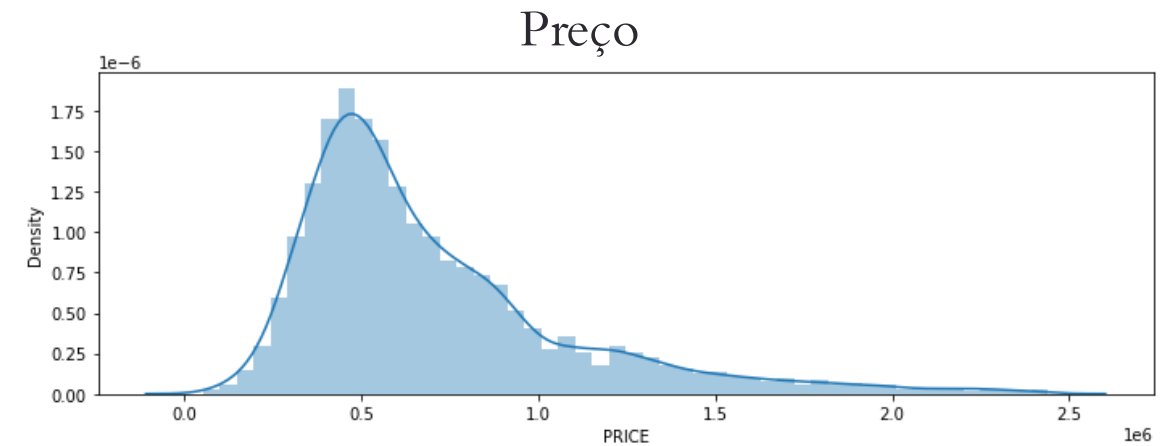
NEAREST_SCH_RANK e CBD_DIST
apresentam as maiores correlações negativas.

LAND_AREA e MONTH_SOLD parece não
apresentar correlação.

EDA (GitHub)

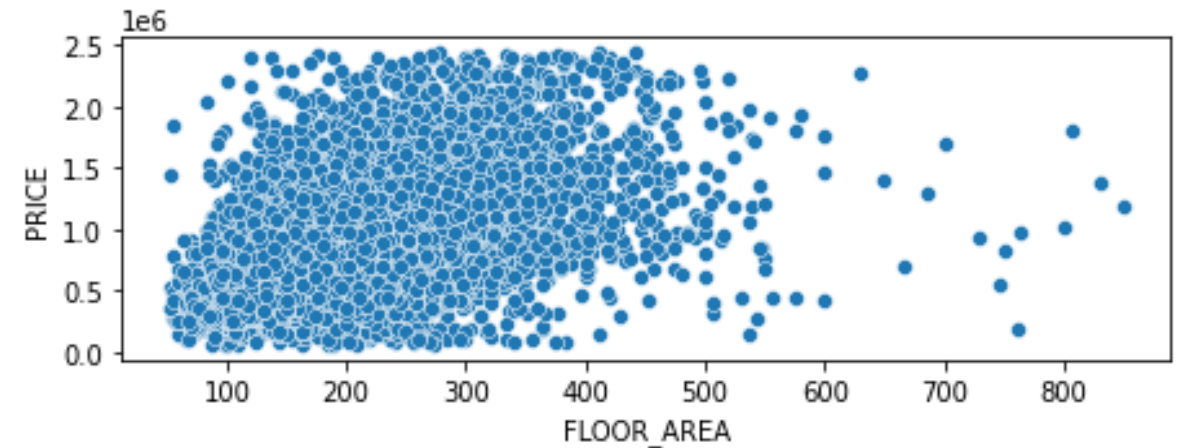
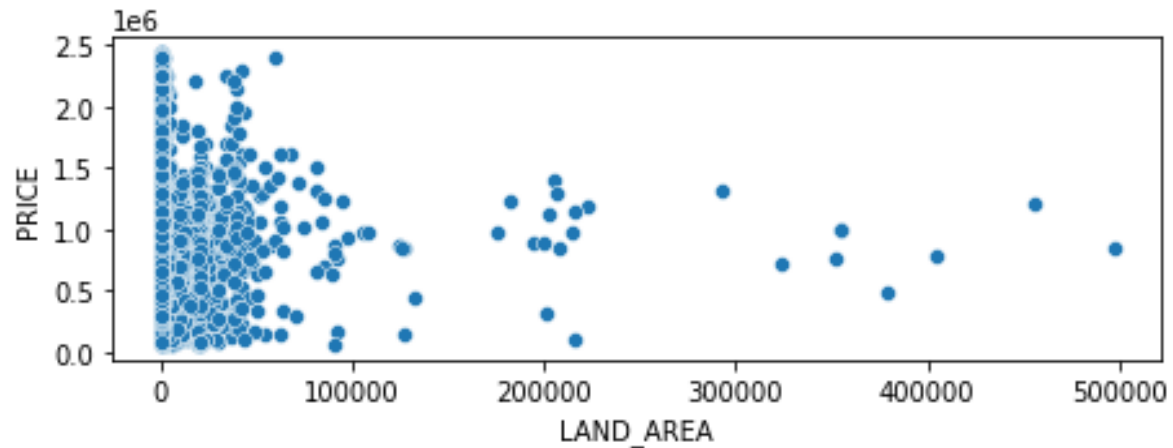


Banheiros, Quartos e Vagas de Garagem



EDA ([GitHub](#))

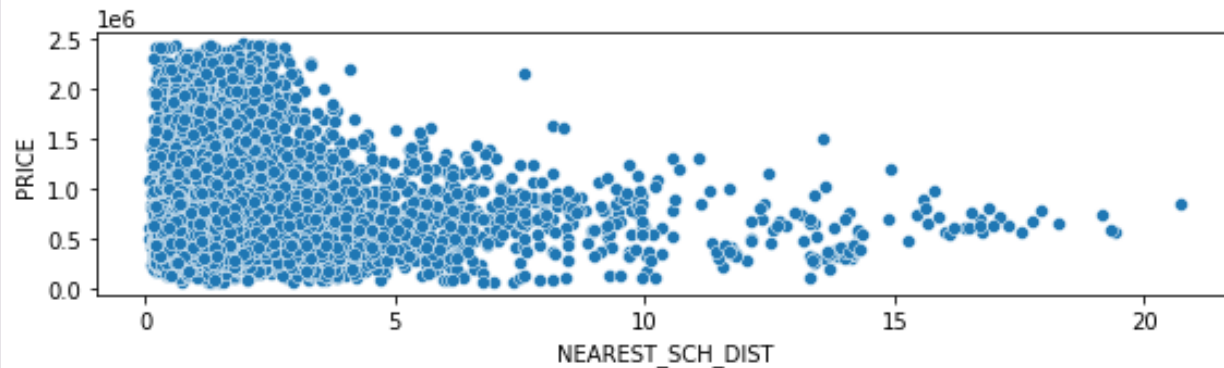
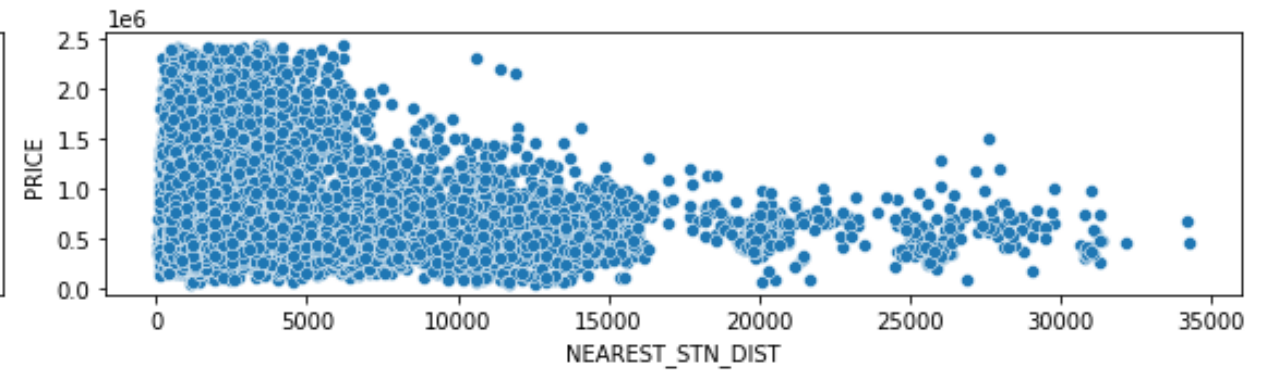
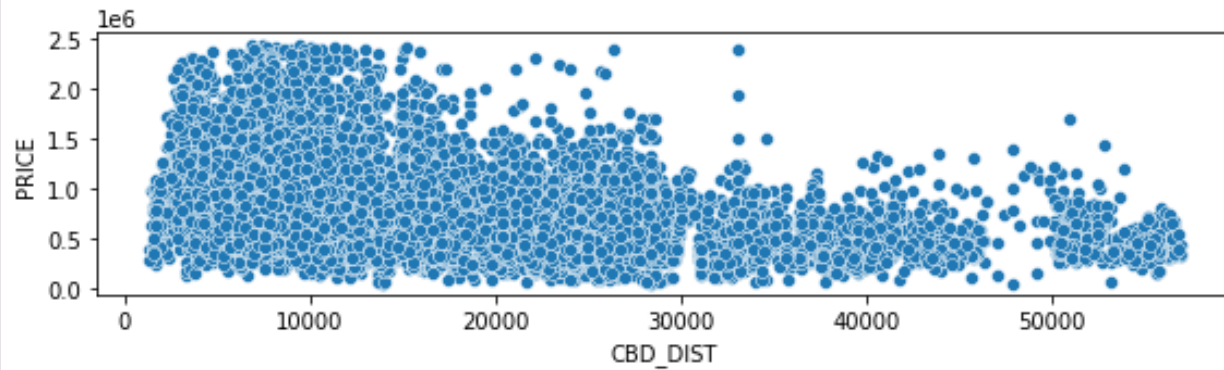
Área de terreno e Área construída



As áreas não apresentam comportamento que pareça ser muito decisivo para a definição de preço.

EDA (GitHub)

Distância para o Centro da cidade, Estação de trem e Escola mais próximos



As distâncias mostram uma correlação inversa ao preço das casas, tendendo a serem mais decisivas para o valor.

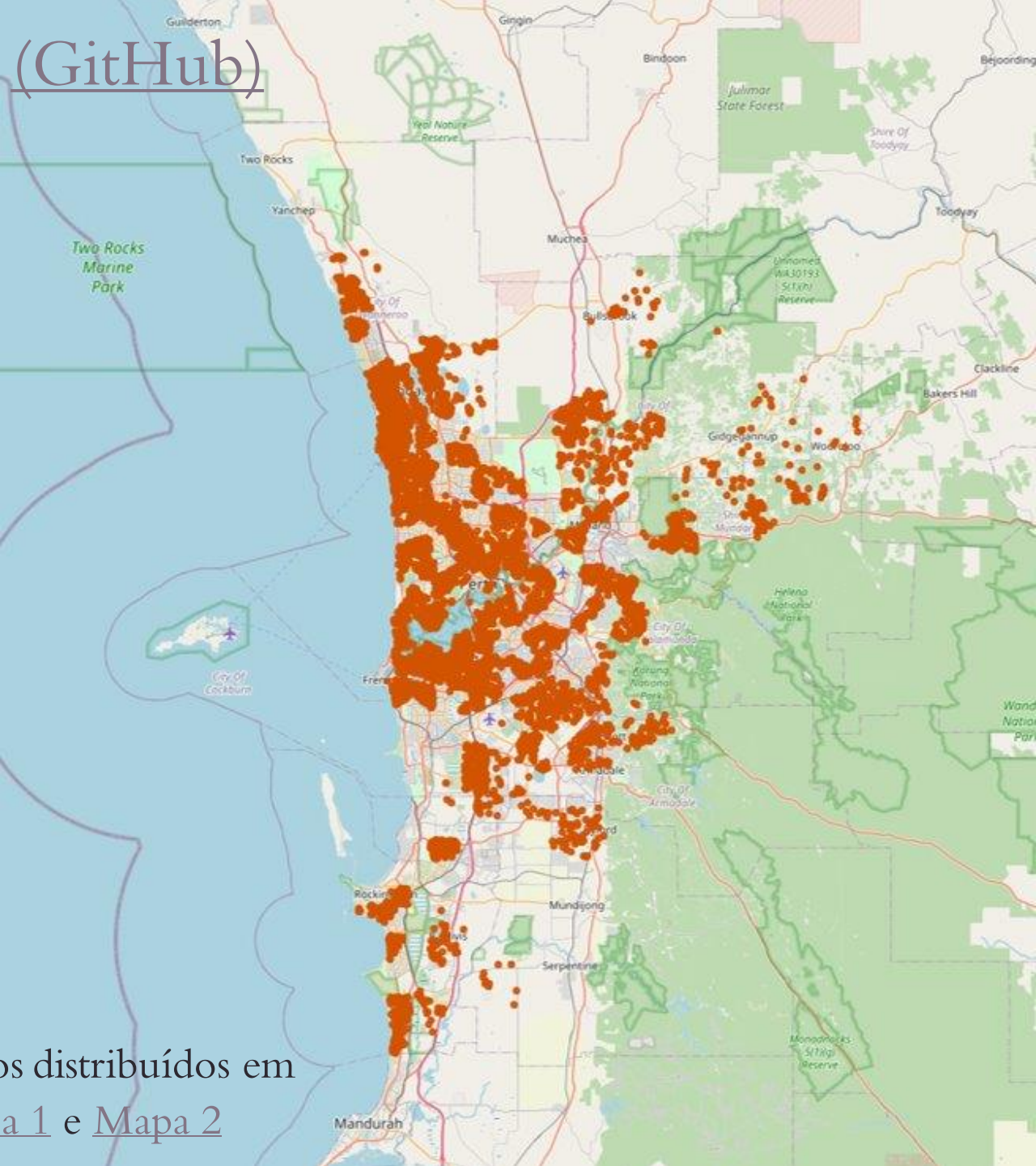
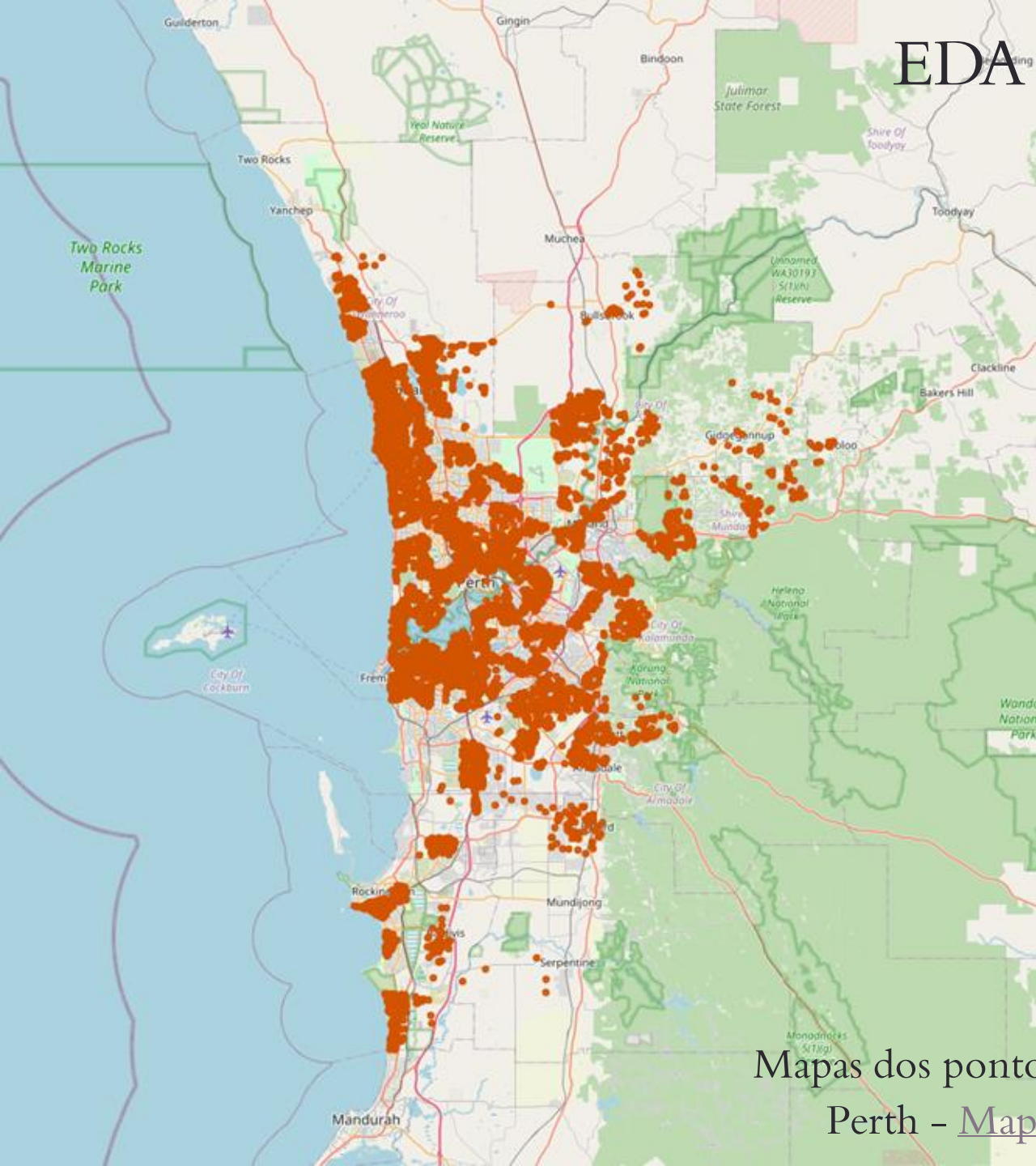
EDA ([GitHub](#))

Aumento dos preços por ano

Os preços aumentam significativamente após as Olimpíadas de Sidney (2000) e o aumento global do preço das commodities (2003–2004), devido à mineração australiana.



EDA (GitHub)



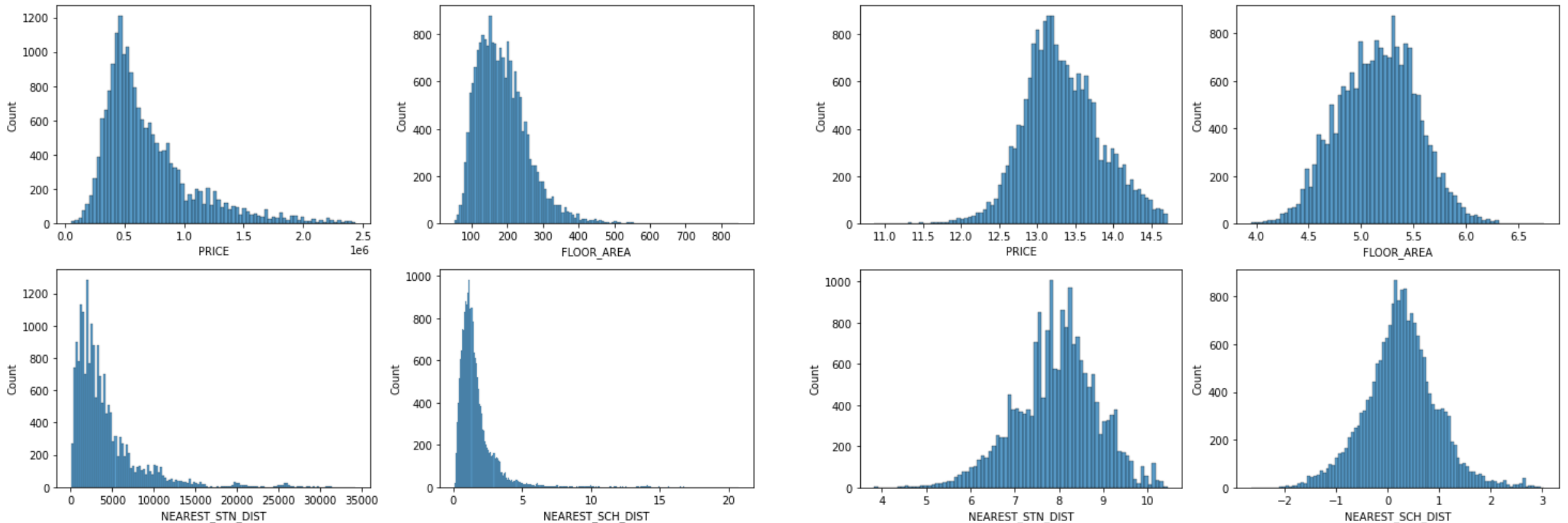
Mapas dos pontos distribuídos em
Perth – [Mapa 1](#) e [Mapa 2](#)

Feature Engineering ([GitHub](#))

Aplicação de log para lidar com distribuição enviesada;

Mudança da coluna SUBURB pela média dos preços de cada distrito;

Retirada dos demais dados categóricos.



Machine Learnings – Ensemble Learnings

Bagging – aprende os *weak learners* de maneira independente uns dos outros, e os combina seguindo um processo médio determinístico/preditivo;

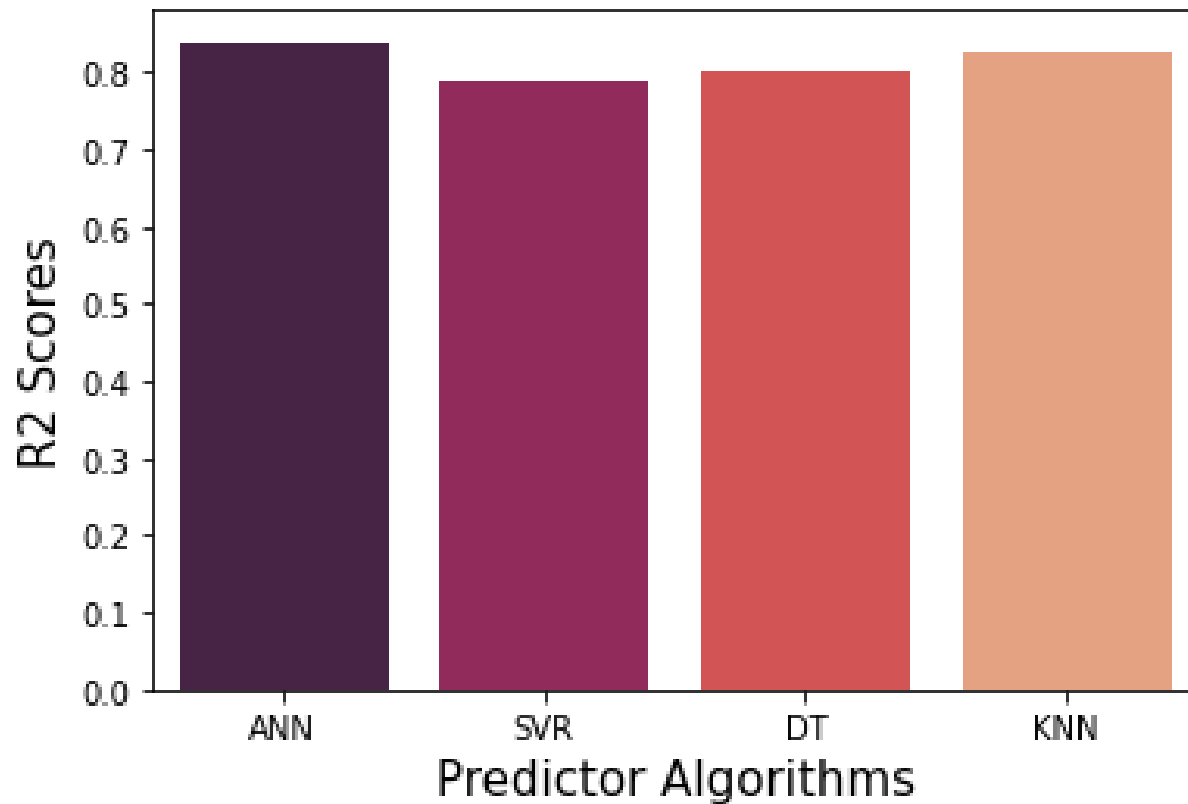
Boosting – aprende os *weak learners* sequencialmente, adaptando-se e os combinando com uma estratégia determinística/preditiva;

Stacking – aprende *weak learners* em paralelo e os combina num metamodelo para produzir uma previsão baseada nas previsões dos diferentes weak learners.

Stacking (GitHub)

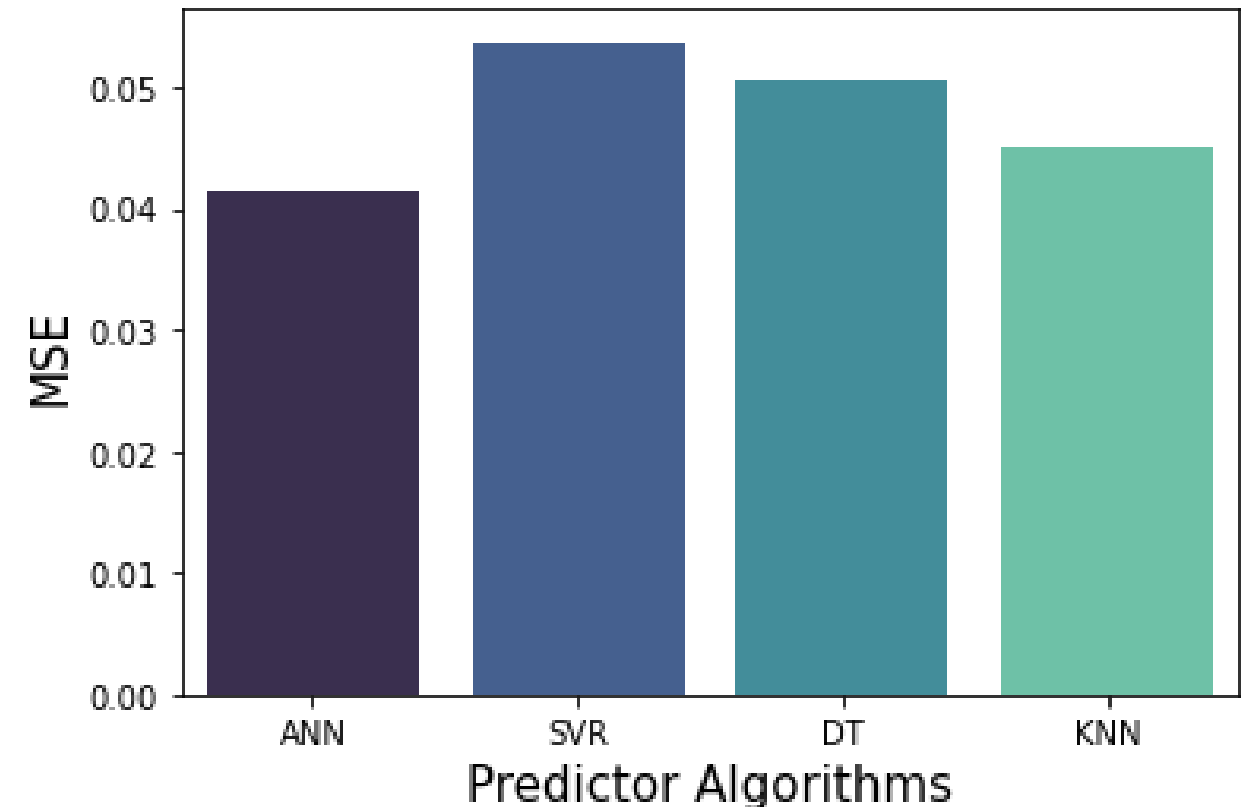
FS – k=3

`['SUBURB', 'FLOOR_AREA', 'NEAREST_SCH_RANK']`



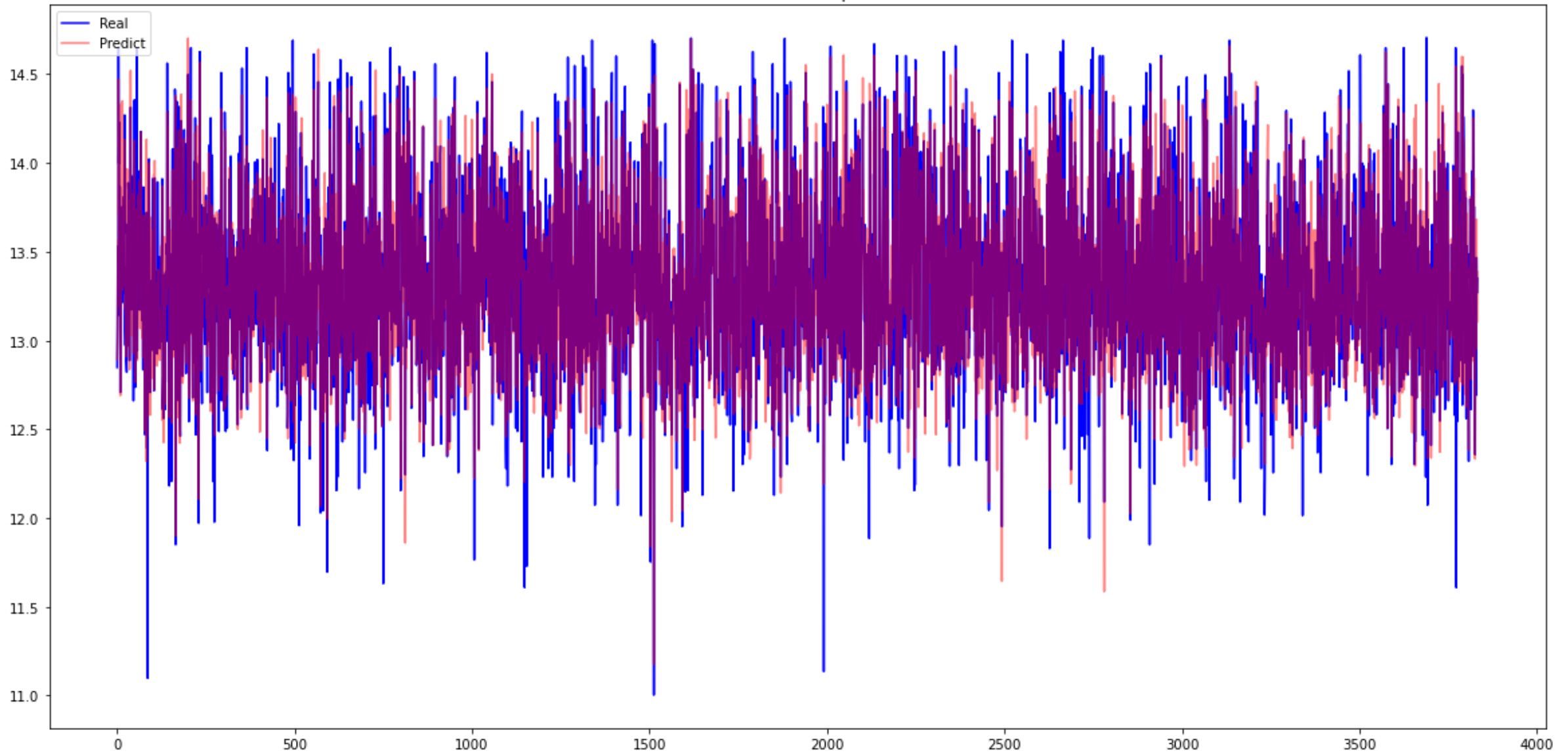
Modelos aplicados:

- Redes Neurais;
- *SVM*;
- Árvore de decisão;
- *KNN*;
- Regressão Linear.



Stacking (GitHub)

Real x Predict comparison

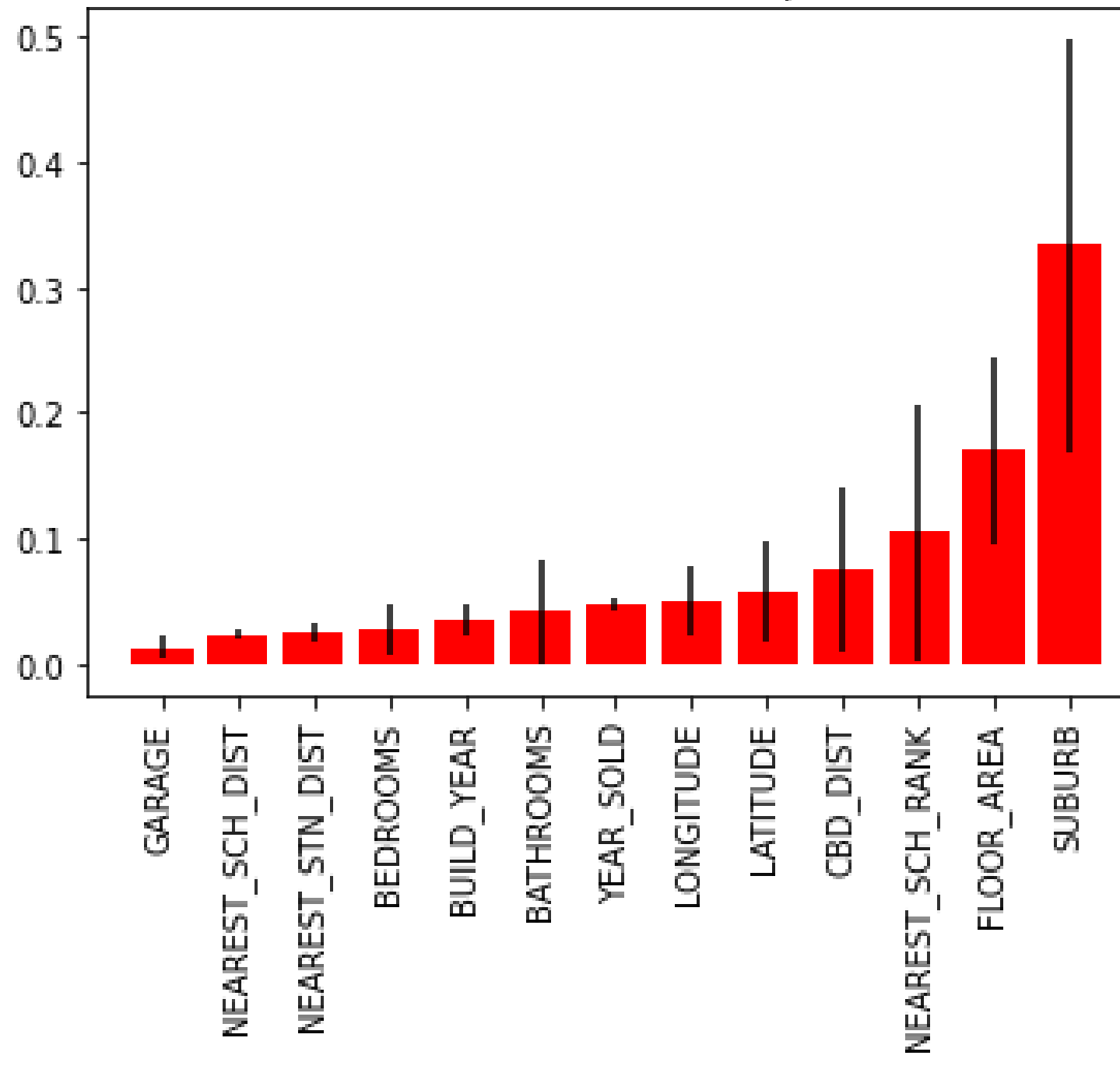


BAGGING (GITHUB)



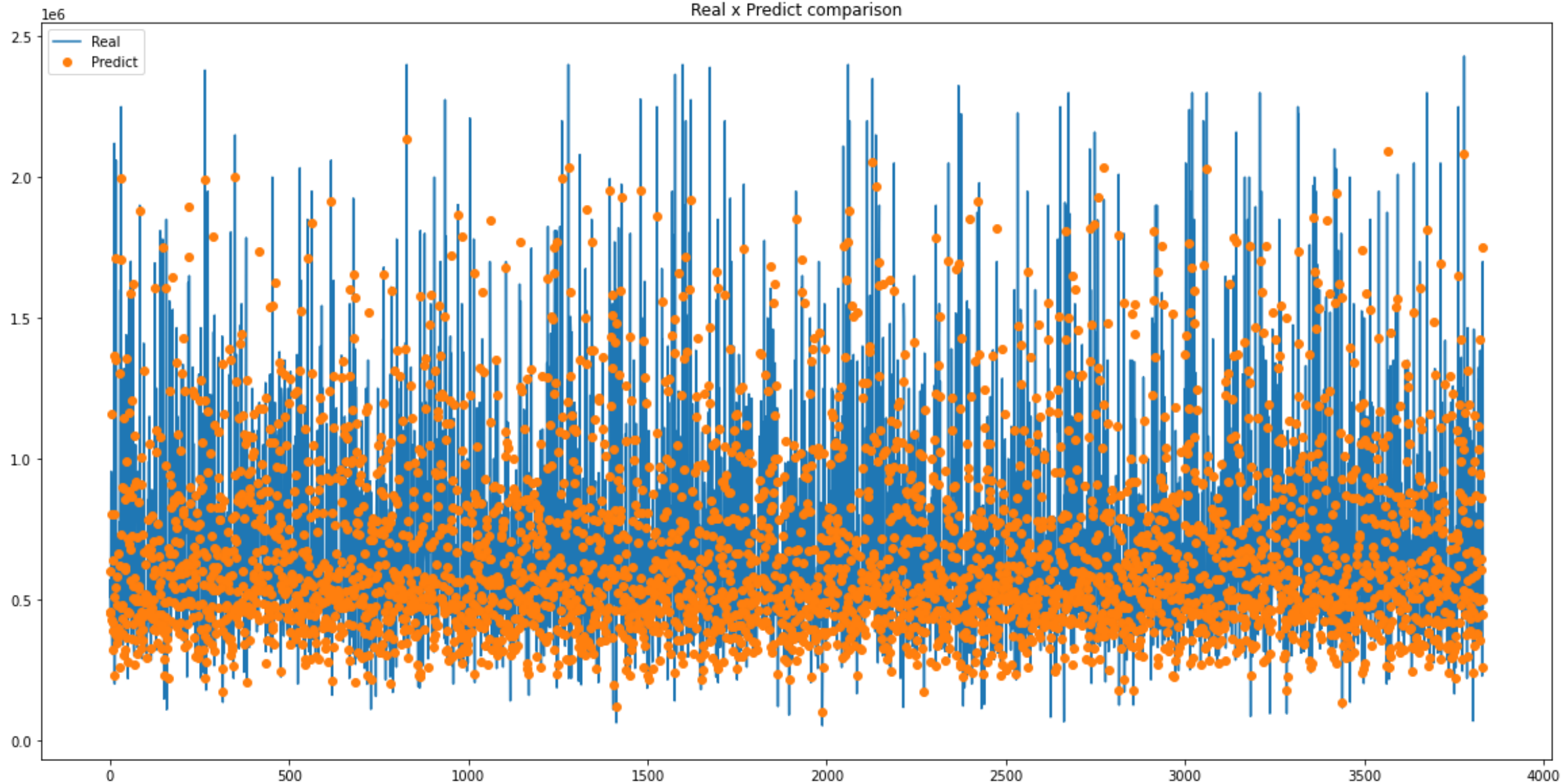
Modelo aplicado – Random Forest

Random Forest Feature Importance



Bagging (GitHub)

Real x Predict comparison

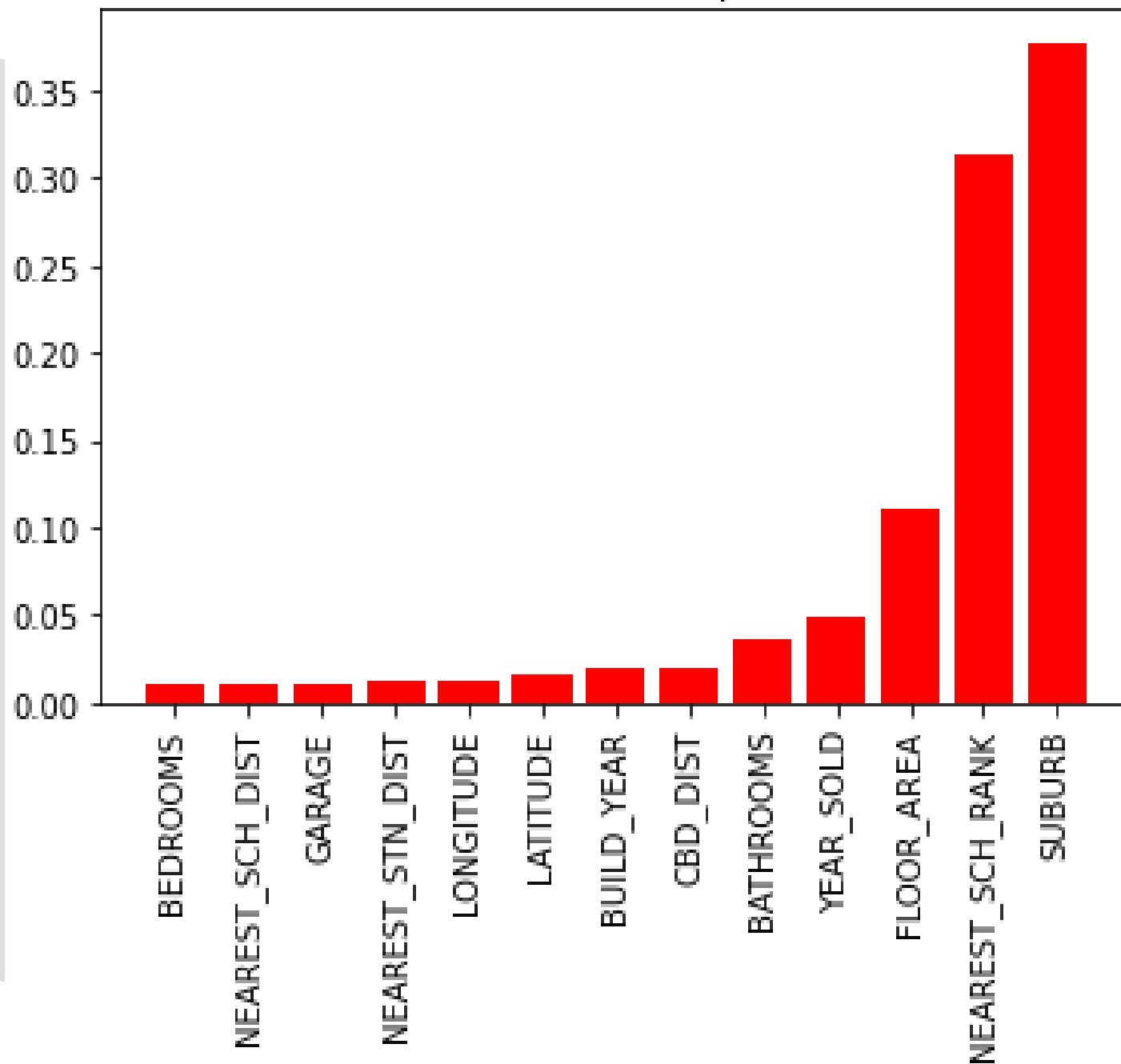


BOOSTING (GITHUB)



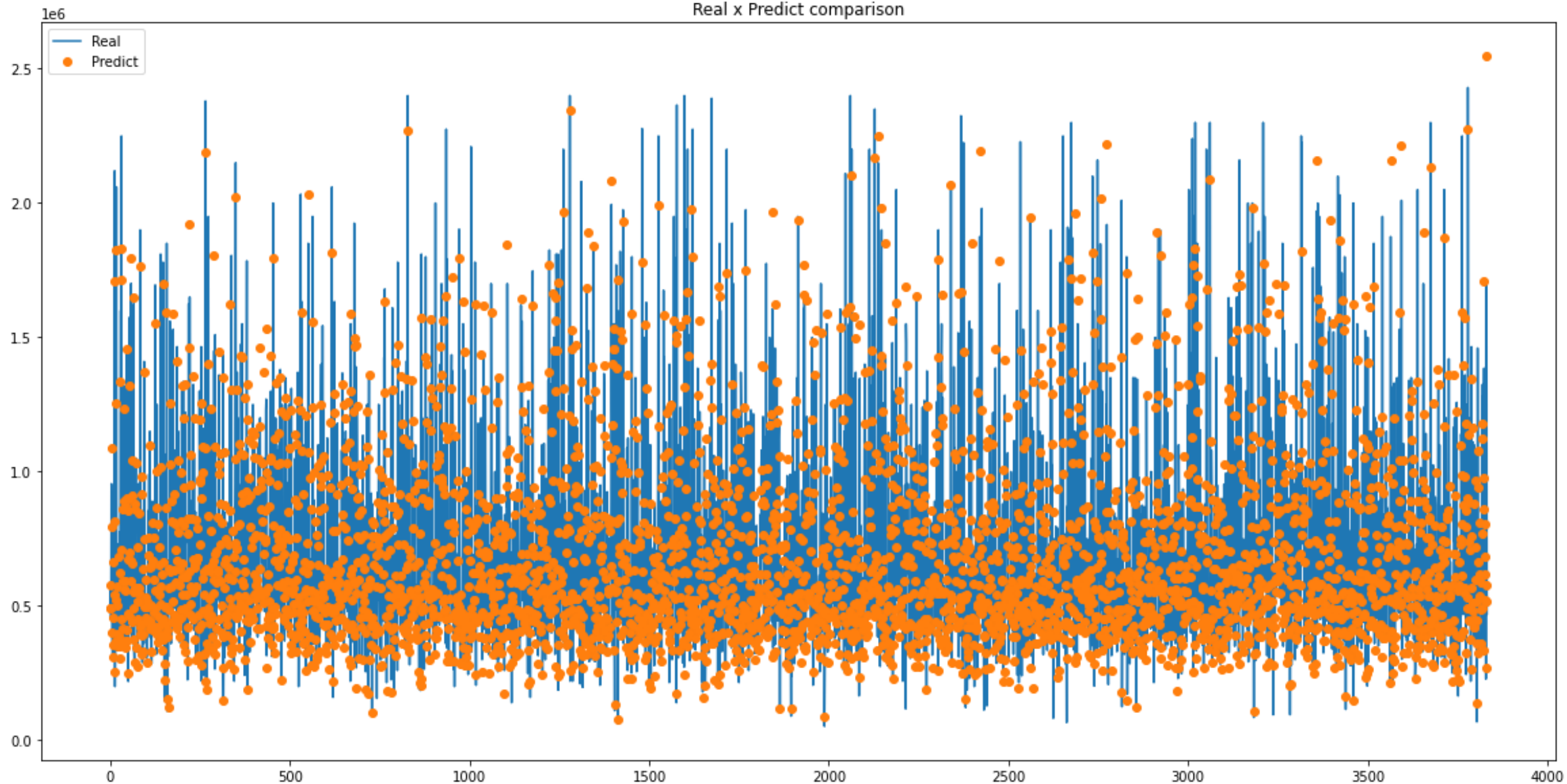
Modelo aplicado – Extreme
Gradient Boosting

XGBoost Feature Importance

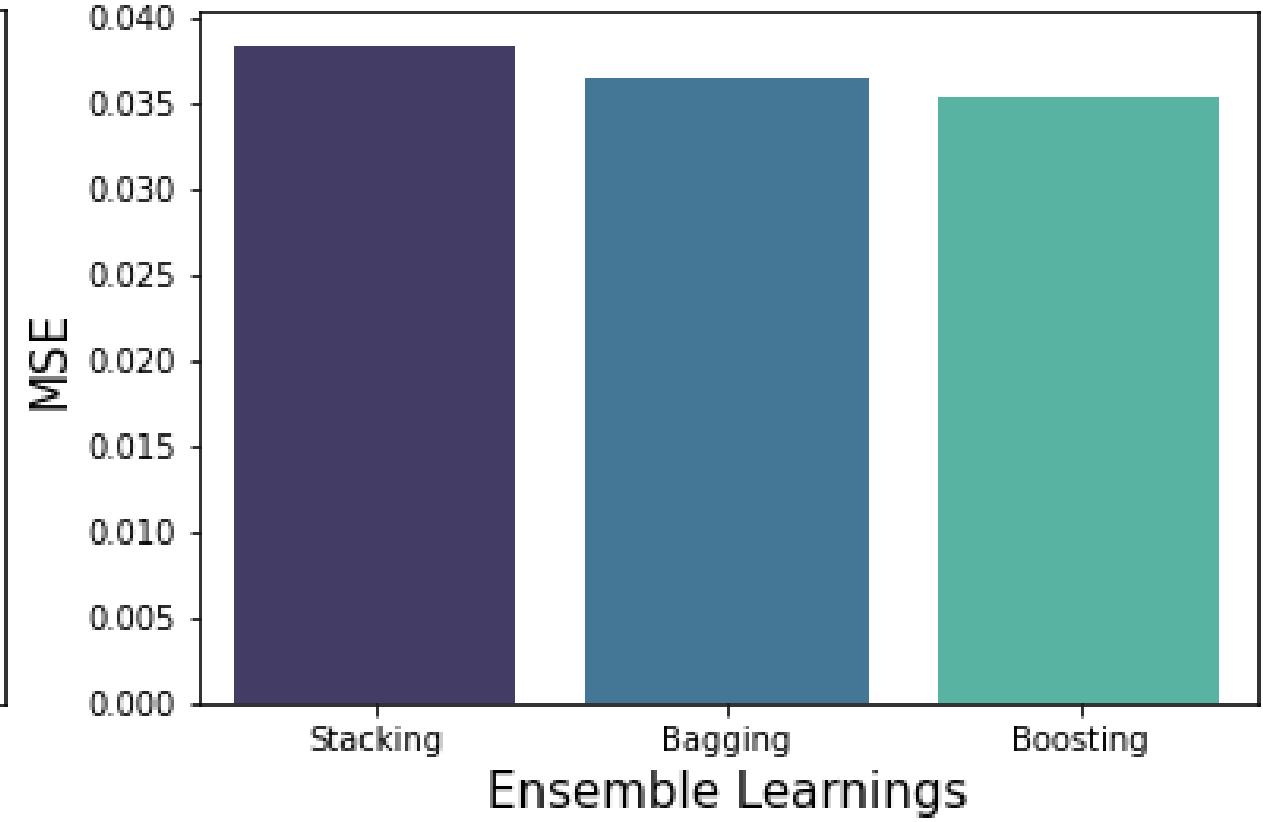
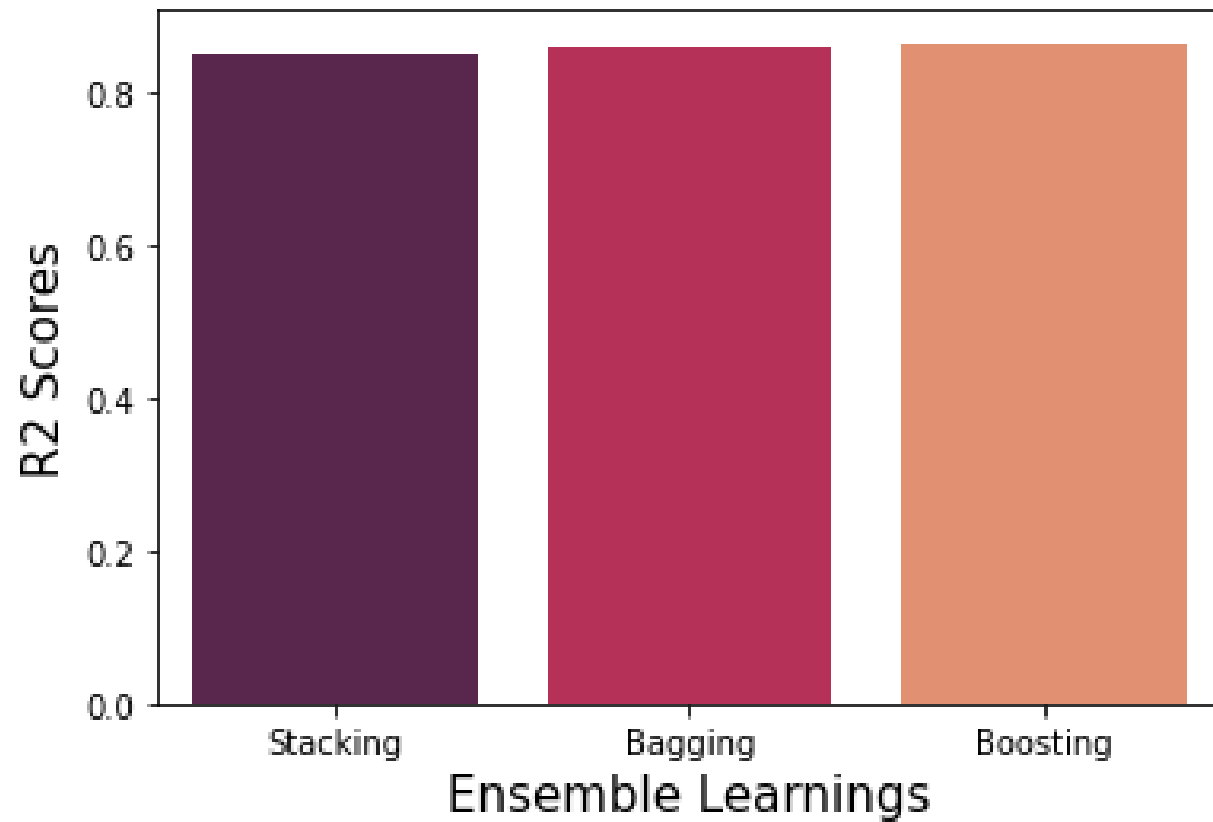


Boosting (GitHub)

Real x Predict comparison



R^2 e MSE dos Ensemble Learnings



	R2	MSE
Stacking	0.850600	0.038409
Bagging	0.859343	0.036571
Boosting	0.864239	0.035298

Deploy (App)

Streamlit – framework de código aberto, criado para colocar em funcionamento os projetos de Machine Learning, sem a necessidade de ferramentas de front-end ou de deploy de aplicações.



×

Choose the ensemble

XGBoost (Boosting) ▾

Suburb mean prices

90000.0

170000.0 1896586.0

Number of Bedrooms

1

1 10

Number of Bathrooms

1

1 7

Number of garages

1

1 10

Floor Area Size - m2

52

52 849

Build Year

1870

1870 2017

Downtown distance

1300

1300 56900

Nearest Station distance

46

46 34300

▾

House Price Sales Prediction: Perth City

Explore different ensemble approaches for Regression



Image Source: <https://www.travelsafe-abroad.com/br/australia/perth/>

About this APP

This application aims to explore 3 different types of regression model ensembles, namely: Boosting, Bagging and Stacking. The predictions made are for home prices in the city of Perth - Australia

Map of Perth

The red points are sold houses in Perth and whose descriptive data was used in this project.

< Manage app

An aerial photograph of a city harbor, likely Sydney, Australia. In the background, a dense skyline of skyscrapers is visible under a clear sky. The middle ground shows a mix of urban development, including buildings, green spaces, and a large stadium. The foreground is dominated by a body of water with a prominent bridge featuring two large arches. Several boats and ferries are docked along the waterfront.

MUITO OBRIGADO!

github.com/alexro6 – linkedin.com/in/alexrodriguesdeoliveira

github.com/leilaff89 – linkedin.com/in/leila-fabiola-ferreira-31675163