

Business Case

ML para gerenciamento de portfólio

Rafaela Medeiros

O desafio



- 1 A partir de uma série histórica, criar modelos de previsão de valor de venda e tempo esperado até ela.
- 2 Avaliar e selecionar o melhor modelo para previsão.
- 3 Aplicar o resultado em um novo banco de dados (Df), gerando variáveis confiáveis para estimação.
- 4 Implementar modelos para alocação de um portfólio de 150 milhões, otimizando o valor esperado do retorno por dia:

$$E_r = \frac{E_{valor} - V_{compra}}{E_{tempo}}$$

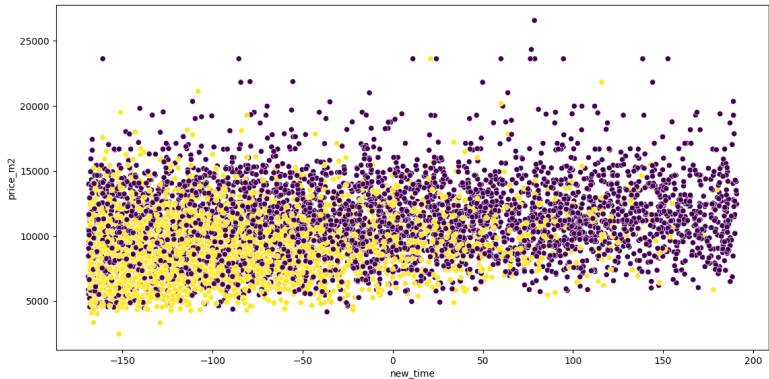
Pressupostos

- Ambos os Dfs espelham movimentos causais.
- O mecanismo gerador é o mesmo.
- Existe um preço ótimo para venda por unidade.

Estratégia

- Análise exploratória.
- Definir uma teoria adequada para a modelagem.
- Preço-venda ótimo encontra-se próximo à mediana do tempo de venda.

Pressupostos





“The problem with confounding signal with noise is that noise is, by definition, unpredictable.” (PRADO, 2019)

Variable	Simulation	Target
rooms	1.61E-16	-9.71E-17
garages	-1.83E-16	2.12E-17
useful area	-4.25E-17	-8.23E-17
latitude	1.11E-13	-3.45E-15
longitude	-1.04E-13	-7.40E-14
value	1.01E-16	5.72E-17
interior quality	7.56E-17	-1.63E-16
time on market	-6.77E-17	–
sold	-7.11E-02	–



Correlações

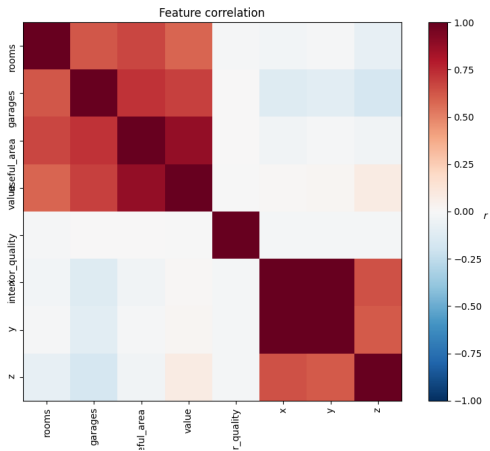


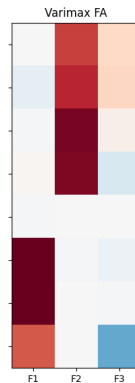
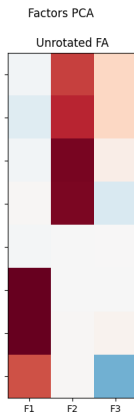
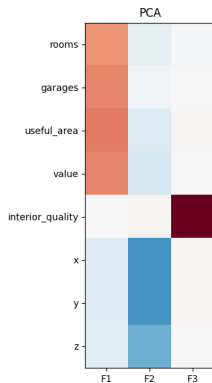


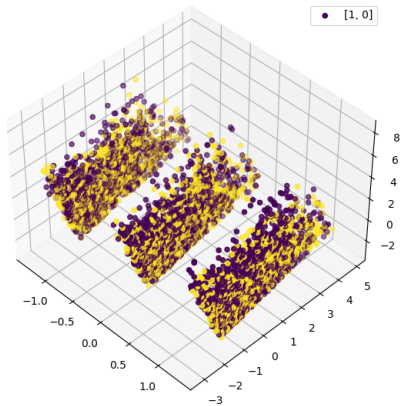
Figura: 01 - ML procurando pelo sinal



Figura: 02 - Overfitting

Fatores - PCA

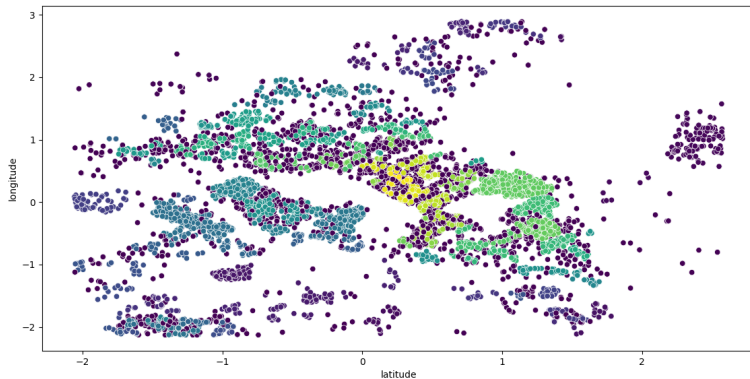




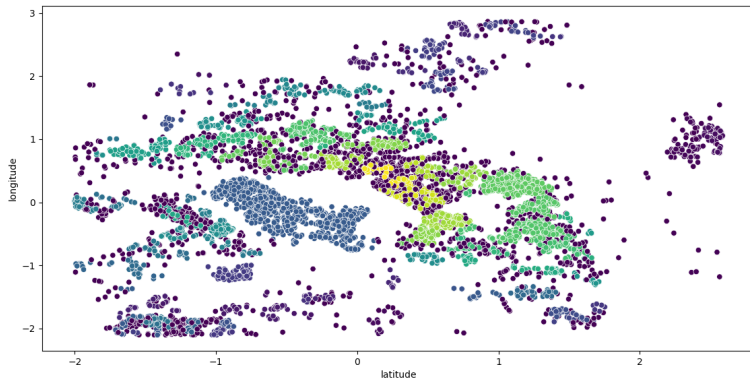
- Pode ser usado para selecionar variáveis.
- Não distingue entre ruído e sinal.

- Três variáveis: estimadores mais precisos e menos overfitting.
- Cluterização: "localização" e "qualidade".
- Estimar tempo usando todas as obs. de simulação.
- Usar Kaplan-Meier para calcular o tempo ponderado pela prob. de venda para simulação.
- Estimar preço apenas para as obs. no intervalo de 90-150 dias.

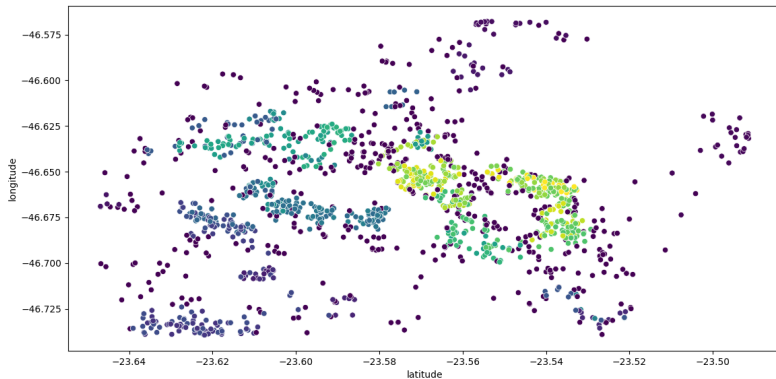
Clusters Simulations - HDBSCAN



Clusters Target - HDBSCAN



Clusters Price - HDBSCAN

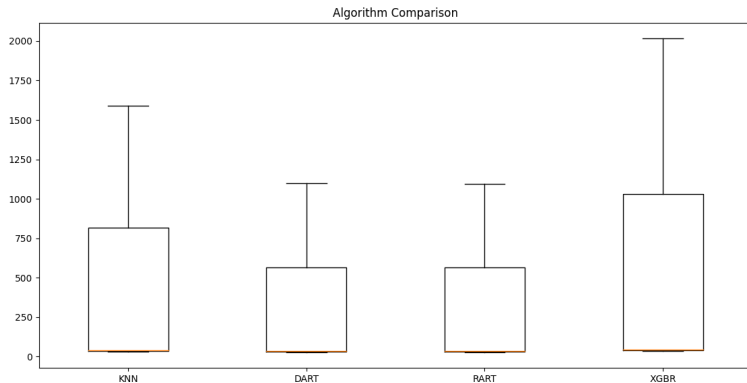


$$Time = area + quality + y$$

$$Price = area + quality + y$$

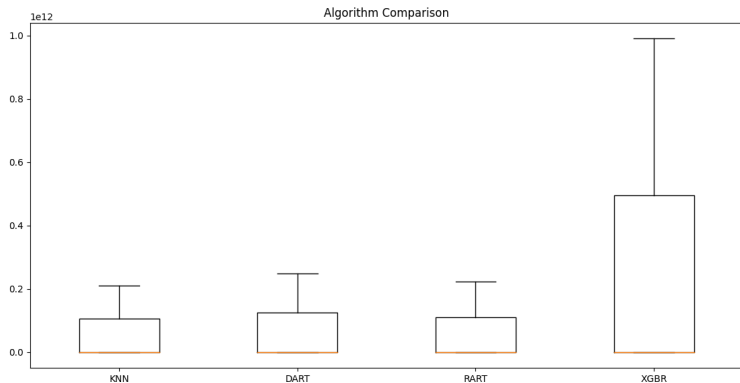
- Kernel Nearest Neighbors, Decision Tree, Random Forest e XgBoost

Comparação - Time



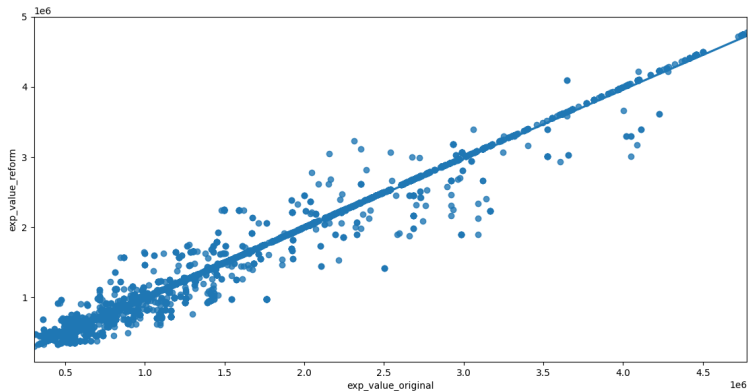
RMSE: 39.89 33.14 33.02 44.90

Comparação - Price

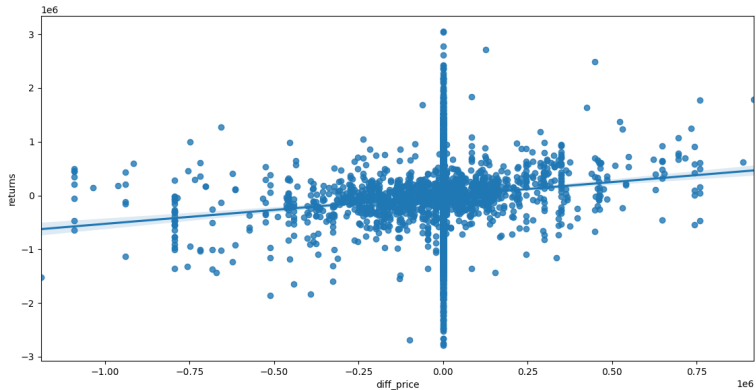


RMSE: 451,605.24 498,262.34 469,276.70 995,176.51

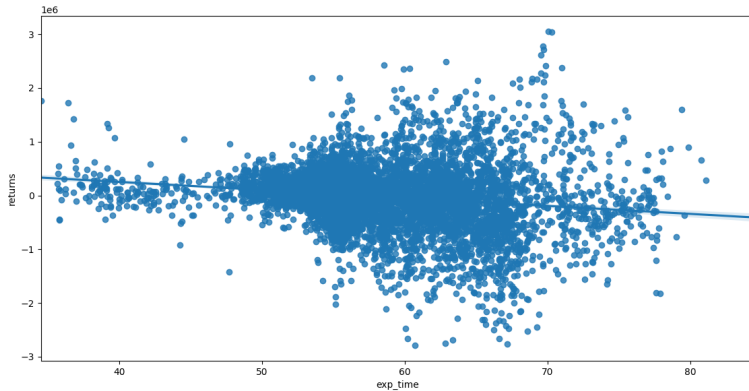
Resultados



Resultados



Resultados



Condições

- 1 Retorno esperado total maior que o RMSE estimado.
- 2 Valor esperado sem reforma maior que o valor de compra.
- 3 Diferença de preço positiva após reforma.
- 4 Retorno/dia positivo

Conj. Selecionado

- 94 unidades
- Valor total de compra
R\$115,640,476.00
- Retorno estimado
R\$ 73,397,985.50
- Tempo médio para venda
58 dias

