



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
PÓS-GRADUAÇÃO *LATO SENSU* EM CIÊNCIA DE DADOS E BIG DATA

ANÁLISE DESCRITIVA E PREDITIVA DOS PREÇOS DOS FUNDOS DE INVESTIMENTO IMOBILIÁRIO NO MERCADO FINANCEIRO - UTILIZANDO MACHINE LEARNING

ALUNO: ANTONIO AURELIANO DOS SANTOS NETO



Etapas de desenvolvimento do TCC

1. Identificação e definição do problema

2. Coleta de dados

3. Processamento/tratamento dos dados

4. Análise/exploração dos dados

5. Criação de modelo de ML

6. Apresentação dos resultados



Identificação e definição do problema – O que é fundo de investimento imobiliário?

O que é fundo de investimento imobiliário?

Fundo de investimento imobiliário (“FII”) é uma comunhão de recursos destinados à aplicação em empreendimentos imobiliários. O FII é constituído sob a forma de um condomínio fechado, sendo dividido em cotas, que representam parcelas ideais do seu patrimônio.



Vantagens?

- Permite ao investidor aplicar em ativos relacionados ao mercado imobiliário sem, de fato, precisar comprar um imóvel.
- Não há a necessidade de desembolsar todo o valor normalmente exigido para investimento em um imóvel.
- Diversificação em diferentes tipos de ativos do mercado imobiliário (ex.: shopping centers, hotéis, residências etc.).
- As receitas geradas pelos imóveis ou ativos detidos pelo fundo são periodicamente distribuídas para os cotistas.
- Aumento nos preços dos imóveis do fundo gera aumento do patrimônio do fundo e, conseqüentemente, valorização do valor das suas cotas.
- Todo o conjunto de tarefas ligadas à administração de um imóvel fica a cargo dos profissionais responsáveis pelo fundo: busca dos imóveis, trâmites de compra e venda, procura de inquilinos, manutenção, impostos etc.



Identificação e definição do problema

Qual é o problema?

- Identificar o melhor modelo de previsão para os preços de fechamento dos fundos imobiliários no mercado de ações – utilizando machine learning.

Quem tem este problema?

- Trabalho de conclusão do curso de Ciência de dados e Big Data.

Por que este problema deve ser resolvido?

- Auxiliar novos investidores na aquisição de cotas de fundos de investimentos imobiliários no mercado brasileiro.

Como saberei que este problema foi resolvido?

- Avaliando qual o modelo teve o melhor desempenho nas previsões dos preços de fechamento dos fundos.



Coleta de dados, processamento e tratamento de dados

- Extração dos fundos de investimentos imobiliários do site fundexplorer, utilizando a técnica web scraping + python.
- Extração das informações das movimentações diárias dos fundos imobiliários no mercado de ações brasileiro, utilizando API Yahoo finance. Período: 01/01/2019 a 30/06/2021.
- Extração das datas de pagamento de rendimentos de cada fundo no site status invest.
- Extração da série temporal para treinamento do modelo preditivo. Período: 02/01/2015 a 30/06/2021.

```
import yfinance as yf
#Importar lib para manipulação de datas
from datetime import datetime

#Transforma o código do fundo do Dataset Lista de FIIs fundexplorer e adiciona ".SA"
cod_fundos = df.cod + '.SA'

#Define o período das operações realizadas no mercado (B3)
dt_inicio = datetime(2019,1,1)
dt_fim = datetime(2021,6,30)
```



Ferramentas e bibliotecas utilizadas



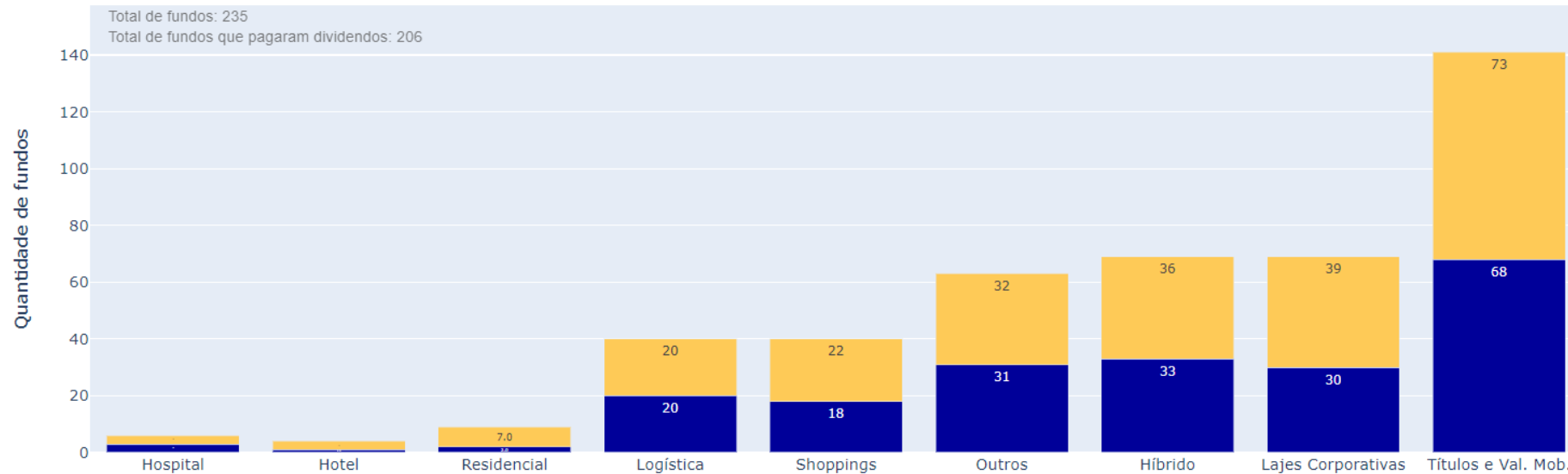
ARIMA



Análise/exploração dos dados

Fundos por setor que pagaram dividendo regularmente no 1º sem. de 2021.

Uma das vantagens dos FIs são os pagamento de dividendos/proventos.



■ Total de fundos
■ Fundos que pagaram Divid.

Fundos por setor que realizaram pagamentos de forma regular, nos últimos 6 meses

Setor	Qtd. Fundos pagadores	Total de fundos por setor	% de pagamento
Residencial	2	7	28,57%
Hotel	1	3	33,33%
Shoppings	18	22	81,82%
Lajes Corporativas	30	39	76,92%
Outros	31	32	96,88%
Titulos e Val. Mob	68	73	93,15%
Híbridos	33	36	91,67%
Logística	20	20	100,00%
Hospital	3	3	100,00%



Análise/exploração dos dados

Ranking dos melhores fundos imobiliários utilizando os critérios abaixo:

- Setores ['Shoppings', 'Lajes Corporativas', 'Logística'] -> qtd_ativos >= 4;
- P/VPA entre 85% e 120%;
- Vacância física deverá ser menor que 15%;
- A liquidez diária deverá ser maior que 20 mil;
- Fundos que pagaram dividendos todos os meses no 1º semestre de 2021. (Verificar a regularidade no pagamento);
- Selecionar os 2 melhores fundos de cada setor.

cod	setor	preco_atual
MXRF11	Híbrido	10.23
HSRE11	Híbrido	97.00
TRXF11	Outros	102.90
HCTR11	Outros	134.99
CPTS11	Títulos e Val. Mob.	99.70
KNIP11	Títulos e Val. Mob.	109.80
XPML11	Shoppings	102.35
VISC11	Shoppings	112.50
BTLG11	Logística	110.64
HGLG11	Logística	169.21
PVBI11	Lajes Corporativas	94.00
VINO11	Lajes Corporativas	60.20



Análise/exploração dos dados

Análise do fundo imobiliário HGLG11 – Preço de Fechamento

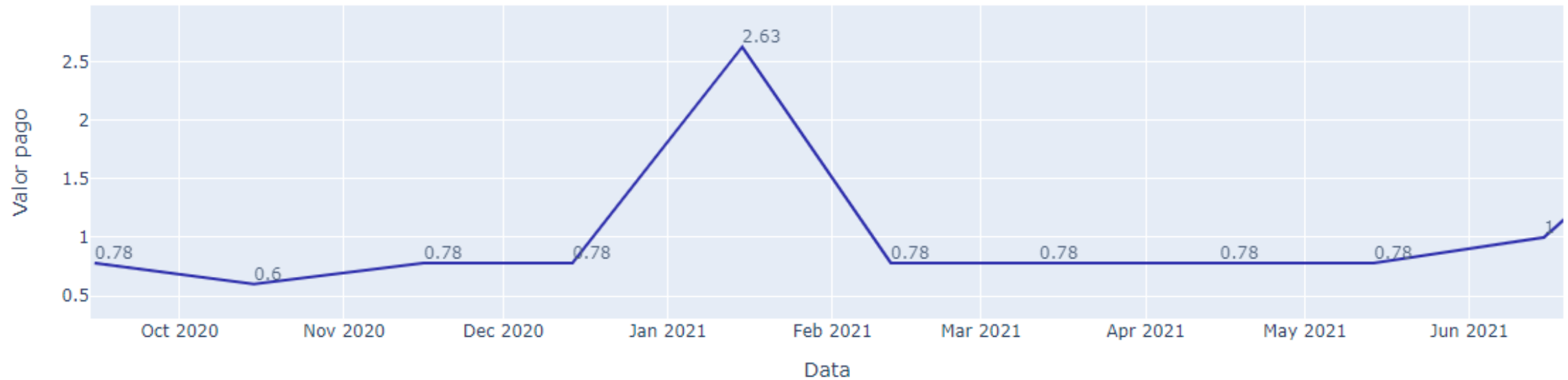
Histórico de preços - HGLG11.SA - Setor: Logística



Análise/exploração dos dados

Análise do fundo imobiliário HGLG11 – Histórico de pagamento

Histórico de pagamento de dividendo por cota do fundo HGLG11.SA Setor: Logística



Criação do modelo de machine learning

- Criação do modelo preditivo utilizando o FII HGLG11.SA.

Como já conhecemos a base e sabemos que a saída será o preço de fechamento, nosso modelo é classificado como **supervisionado**.

Período: 01/01/2015 a 30/06/2021 – Variável: Preço de fechamento

A previsão será de 90 dias a partir da ultima data da base de treinamento.

Modelos utilizados para a predição

- Modelo GRU (rede recorrente fechada)
- Modelo ARIMA (modelo autorregressivo)
- Modelo Prophet (modelo de regressão)

	ds	y
0	2015-01-02	109.099998
1	2015-01-05	111.000000
2	2015-01-06	110.000000
3	2015-01-07	109.000000
4	2015-01-08	108.500000
...
1598	2021-06-23	170.550003
1599	2021-06-24	171.000000
1600	2021-06-25	167.500000
1601	2021-06-28	165.500000
1602	2021-06-29	170.000000
1603 rows × 2 columns		



Avaliação dos resultados

- Diferença entre valor de teste x Previsões dos modelos

	VALOR_TESTE	PREVISAO_ARIMA	PREVISAO_PROPHET	PREVISAO_GRU	DIFF. ARIMA	DIFF. PROPHET	DIFF. GRU
ds							
2021-02-10	178.30	178.40	175.61	178.58	0.10	2.69	0.28
2021-02-11	178.51	178.34	175.81	178.32	0.17	2.70	0.19
2021-02-18	177.89	178.22	176.43	178.52	0.33	1.46	0.63
2021-02-19	178.00	178.11	176.77	177.86	0.11	1.23	0.14
2021-02-22	177.55	177.99	176.93	177.98	0.44	0.62	0.43
...
2021-06-23	170.55	169.72	174.13	170.56	0.83	3.58	0.01
2021-06-24	171.00	169.62	174.30	170.62	1.38	3.30	0.38
2021-06-25	167.50	169.53	174.55	171.16	2.03	7.05	3.66
2021-06-28	165.50	169.44	174.46	167.83	3.94	8.96	2.33
2021-06-29	170.00	169.35	174.37	165.88	0.65	4.37	4.12

90 rows × 7 columns

MÉTRICAS	ARIMA	PROPHET	GRU
MAE	1.78	3.67	1.1
MSE	4.66	20.82	2.17
RMSE	1.33	1.92	1.47
R2	61.0%	56.0%	82.0%

Pode-se observar, dos dados apresentados na Figura acima, que o modelo GRU apresentou, para o conjunto de teste, os menores erros (MAE, MSE e RMSE). No R2 (coeficiente de determinação) o valor do ficou 82%, ou seja, se ajusta melhor à amostra.



Conclusão

- Nos estudos realizados neste trabalho, os fundos dos setores de Hotéis, Shoppings e Lajes Corporativas foram os mais afetados, reflexo do fechamento do comércio e medidas de restrição. Os fundos do setor Hospitalar tiveram destaque devido ao aumento de internações, exames e consultas relacionadas à Covid-19 e os fundos do setor de logística cresceu devido ao aumento nas encomendas realizadas nas lojas virtuais.
- O modelo de machine learning que conseguiu prever com melhor desempenho foi o GRU (Rede Recorrente Fechada), porém deve-se destacar a complexidade em prever preços no mercado de ações, pois envolve mais variáveis do que simplesmente o histórico dos preços, como por exemplo, a pandemia e/ou epidemia.



ARIMA



Obrigado!

