

PREDIÇÃO DE PREÇOS DE AÇÕES DA PETROBRÁS UTILIZANDO A BIBLIOTECA FACEBOOK PROPHET

PETROBRAS STOCK PRICE FORECASTING USING FACEBOOK PROPHET LIBRARY

PREDICCIÓN DEL PRECIO DE LAS ACCIONES DE PETROBRÁS UTILIZANDO LA BIBLIOTECA FACEBOOK PROPHET

Maysa Vitória Diniz Marquez¹
Mirian Souto da Silva Fagundes²
Beatriz Regina Valeiro³
Ramon Abilio⁴

Resumo: Dentre as ações negociadas na Bolsa de Valores brasileira, estão as da Petrobrás. O objetivo deste trabalho foi prever o preço das ações da Petrobrás utilizando histórico de preços e a biblioteca Facebook Prophet. Essa biblioteca tem recursos para análise e previsão de séries temporais. Nos experimentos, de maneira geral, obteve-se um erro percentual médio de 0,92% utilizando-se 15 dias para treinamento do modelo e predizendo-se um dia a frente. Percebeu-se ainda uma possível relação entre a quantidade de dias necessária para o treinamento do modelo e a volatilidade da cotação, sendo necessário mais dias quando a volatilidade for maior.

Palavras-chave: Aprendizado de Máquina. Bolsa de Valores. Análise Técnica. Séries Temporais.

Abstract: Among the shares traded on the Brazilian Stock Exchange are those of Petrobras. The objective of this work was to predict the price of Petrobras shares using price history and the Facebook Prophet library. This library has features for time series analysis and prediction. In the experiments, in general, an average percentage error of 0.92% was obtained using 15 days for model training and predicting one day ahead. A possible relationship between the number of days required for model training and price volatility was also noticed, with more days needed when volatility is higher.

Keywords: Machine Learning. Stock Exchange. Technical Analysis. Time Series.

¹Discente do curso técnico em Redes de Computadores. Instituto Federal de São Paulo (IFSP), campus Boituva. ORCID: <https://orcid.org/0009-0006-2991-1379>. E-mail: maysavdm Marquez@gmail.com

²Discente do curso técnico em Redes de Computadores. Instituto Federal de São Paulo (IFSP), campus Boituva. ORCID: <https://orcid.org/0009-0003-2086-7517>. E-mail: miriannfagundes.lf@gmail.com.

³Discente do curso técnico em Redes de Computadores. Instituto Federal de São Paulo (IFSP), campus Boituva. ORCID: <https://orcid.org/0009-0007-0230-4162>. E-mail: be.valeiro@gmail.com.

⁴Mestre em Ciência da Computação. Universidade Federal de Lavras (UFLA). Docente da área de Informática do Instituto Federal de São Paulo (IFSP), campus Boituva. ORCID: <https://orcid.org/0000-0002-7197-5951>. E-mail: ramon.abilio@ifsp.edu.br.

Resumen: Entre las acciones negociadas en la Bolsa de Valores de Brasil están las de Petrobras. El objetivo de este trabajo fue predecir el precio de las acciones de Petrobras utilizando el historial de precios y la biblioteca Facebook Prophet, que realiza predicción de series temporales. En los experimentos, se obtuvo un porcentaje de error promedio de 0.92% utilizando 15 días para el entrenamiento del modelo y prediciendo un día antes. También se observó una posible relación entre el número de días necesarios para la formación del modelo y la volatilidad de los precios, con más días necesarios cuando la volatilidad es mayor.

Palabras-clave: Aprendizaje automático. Bolsa de Valores. Análisis técnico. Series de tiempo.

Submetido 02/04/2023

Aceito 04/09/2023

Publicado 26/09/2023

Introdução

Investimentos no mercado de ações vêm se tornando relevantes no contexto brasileiro, pois o aumento de disseminação de conteúdos relacionados ao mercado de capitais em sites, ou até mesmo em redes sociais, contribui para que muitas pessoas busquem saber mais sobre o assunto. O investimento nesse tipo de mercado pode ser uma oportunidade de renda extra por oferecer autonomia para seus investidores e ampliar os ganhos conforme os resultados obtidos. É importante que o investidor se atente ao prazo do investimento (curto ou longo) conforme seus objetivos (BRAGA, 2019).

No Brasil, a bolsa responsável pelas negociações na sociedade de capital aberto é a B3 (B3, 2022). A B3 é um ambiente que permite aos investidores realizarem negociações de ativos, oferecendo acesso à compra e à venda de ativos financeiros por meio de instituições autorizadas (CVM, 2020).

Em junho de 2022, existiam cerca de 2.122 empresas listadas na B3, considerando empresas nacionais e internacionais (por meio de *Brazilian Depositary Receipts* Patrocinados - BDR) (B3, 2022). Conforme o Boletim Diário da B3 do dia 14 de junho de 2022, no ano anterior houve uma média diária do volume de negociação de 33.545,34 milhões de Reais (BDI, 2022).

O volume de negociação indica quanto, em Reais, investidores individuais ou institucionais compraram ou venderam em ativos na Bolsa de Valores. Esses investidores podem utilizar diferentes técnicas para analisar o momento de se realizar compras ou vendas, como, por exemplo: Análise Fundamentalista, Análise Técnica e Análise de Sentimentos. Na Análise Fundamentalista, o investimento é baseado na análise da empresa de forma qualitativa, quantitativa e temporal utilizando-se dados financeiros da empresa (MALTA; CAMARGOS, 2016). Na Análise Técnica, tenta-se prever o preço, ou cotação, de uma ação analisando-se o histórico de preço dessa ação e a análise de sentimentos combina processamento de linguagem natural e aprendizado de máquina (SINGH; DUBEY, 2014).

Na literatura, pode-se encontrar trabalhos que aplicaram aprendizado de máquina na predição de preços de ações. Os algoritmos de aprendizado de máquina fazem a análise de grandes quantidades de dados em busca de padrões (CASTRO; FERRARI, 2016). Dessa forma, o aprendizado de máquina pode ser útil na predição de preços de ações devido à quantidade de dados disponíveis, principalmente, quando se utiliza Análise Técnica.

Nesse contexto, o objetivo deste trabalho foi utilizar aprendizado de máquina por meio da biblioteca Facebook Prophet para realizar a predição de preços de ações preferenciais da Petrobrás utilizando o histórico de Preço de Fechamento Ajustado. Como esses dados são séries temporais (PARRAY et al, 2020), essa biblioteca se torna adequada, pois ela faz predições baseadas em séries temporais (FACEBOOK PROPHET, 2022).

Fundamentação Teórica

Uma das formas de se construir patrimônio é realizando investimento em ações de empresas negociadas em bolsa de valores. Ações são títulos que conferem a seus detentores, ou investidores, a participação na sociedade da empresa (B3, 2022). A negociação (compra ou venda) de ações é realizada pelos investidores por meio de corretoras ou distribuidoras de valores mobiliários que possuem registro na CVM (Comissão de Valores Mobiliários) (CVM, 2022).

As corretoras oferecem a seus clientes acesso a plataformas de negociação da Bolsa de Valores por meio do sistema *Home Broker*. Esse sistema permite que investidores enviem ordens de compra e venda de ações. Dessa forma, esse sistema automatiza o atendimento da corretora, pois permite ao cliente colocar para execução ordens de compra e venda de valores mobiliários no mercado à vista (lote-padrão e fracionário) (CVM, 2022).

Ao investir em ações, o acionista deve analisar aspectos como: retorno, prazo, proteção, rentabilidade, liquidez e grau de risco. A rentabilidade está diretamente vinculada ao fator risco (STEFANI et al, 2013). Para tanto, é necessário que o investidor defina o grau de risco que está disposto a assumir (STEFANI et al, 2013).

Além de se analisar o risco, um investidor deve definir se pretende obter renda por meio de dividendos distribuídos pelas empresas, se pretende construir patrimônio com a valorização das ações ou se optará por obter lucros de curto prazo com especulação (LAGIOIA, 2009). Em relação aos dividendos, cabe ressaltar que a empresa pode distribuí-los em dinheiro, ou ela ainda pode bonificar os investidores com novas ações, ou permitir que os acionistas subscrevam novas ações (STEFANI et al, 2013).

Para analisar ações, pode-se utilizar, por exemplo, as análises: técnica, de sentimentos ou fundamentalista. Na análise fundamentalista, realiza-se análises quantitativas, qualitativas e temporais com base em dados financeiros da empresa, do setor de atuação e na avaliação

macroeconômica. Ela visa a avaliação de desempenho das empresas em relação ao mercado financeiro, identificando resultados e obtendo uma hipótese detalhada da empresa (MALTA; CAMARGOS, 2016).

Segundo Lameira (2005), na Análise Técnica, tenta-se prever o preço de uma ação analisando-se o histórico de preços dessa ação. Esta escola vale-se do comportamento dos preços negociados em bolsa utilizando-se, por exemplo, cotação, preço máximo e mínimo, volume de negociação, índices e indicadores e figuras de tendência de mercado. Enquanto a Análise Técnica se baseia em dados numéricos, a análise de sentimentos combina processamento de linguagem natural com mineração de texto e utiliza técnicas de aprendizado de máquina para classificar textos como positivos ou negativos (SINGH; DUBEY, 2014).

Aprendizado de máquina é uma área na qual se desenvolve programas computacionais capazes de melhorar automaticamente seu desempenho por meio da experiência (MITCHELL, 1997). O aprendizado de máquina também é utilizado na Mineração de Dados. A Mineração de Dados consiste em explorar uma base de dados usando algoritmos adequados para obter conhecimento (CASTRO; FERRARI, 2016). O processo de mineração envolve, por exemplo, coleta e pré-processamento de dados, análise e busca de padrões por meio de tarefas como (CASTRO; FERRARI, 2016):

1. Análise descritiva de dados: Essa análise permite investigar a distribuição de frequência, as medidas de centro, variação, medidas de posição relativa e associação dos dados;
2. Predição (classificação e regressão): Predição se refere à construção e ao uso de modelo para estimar um valor ou categoria. A classificação é utilizada para prever valores discretos e a estimação é usada para prever valores contínuos;
3. Análise de grupos: É um processo no qual se agrupa um conjunto de objetos em grupos de objetos similares;
4. Associação: O objetivo dessa tarefa é encontrar relações entre os objetos de uma base de dados.

Essas tarefas têm sido utilizadas para prever preços de ações negociadas em bolsa de valores (SOUZA, 2021) e diferentes técnicas têm sido utilizadas, como: Rede Neural Artificial, Regressão e *Support Vector Machine* (SVM).

Os dados utilizados em análise técnica das ações são séries temporais (PARRAY et al, 2020). Séries temporais são definidas como uma sequência de pontos ordenados e distribuídos

em um intervalo de tempo e, normalmente, combinam componentes como: tendência, sazonalidade e ciclos (JANSEN, 2020; PARRAY et al, 2020).

Séries temporais podem ser analisadas, por exemplo, com modelos ARIMA (*Autoregressive Integrated Moving Average*). Esse modelo é utilizado pela biblioteca Prophet (FACEBOOK PROPHET, 2022). O Prophet é uma biblioteca de código aberto desenvolvida e mantida pelo time de Ciência de Dados do Facebook e é disponibilizada em R e Python (FACEBOOK PROPHET, 2022). Essa biblioteca faz previsões utilizando séries temporais nas quais relações não lineares não são evidentes. De forma automatizada, o Prophet realiza o treinamento de diferentes modelos e disponibiliza, além das previsões, ferramentas para avaliação da qualidade das previsões e geração de gráficos.

Trabalhos Relacionados

Werner, Bisognin e Araújo (2020) fizeram a previsão de volume de negociação das ações ordinárias da empresa brasileira Petrobrás (código na bolsa de valores brasileira PETR3). Eles coletaram o histórico de volume do período de 04 de janeiro de 2010 a 18 de setembro de 2018. Os dados foram utilizados juntamente com as técnicas Rede Neural Recorrente (RNN), modelo *AutoRegressive Fractionally Integrated Moving Average* (ARFIMA) e técnicas de combinações de previsões como média aritmética, variância mínima e por regressão, e para tomada de decisão medidas de acurácia como RMSE (*Root Mean Squared Error* ou Raiz quadrada do Erro Quadrático Médio), MAPE (*Mean Absolute Percentage Error* ou Erro Percentual Absoluto Médio) e U de Theil.

Parray et al (2020) apresentam um estudo sobre a previsão do movimento de preços (tendência de alta ou baixa) das ações da NIFTY 50 da National Stock Exchange of India. Eles utilizaram técnicas de *Support Vector Machine* (SVM), algoritmos de Rede Neural Artificial (RNA) e Regressão logística com uma base constituída por dados coletados no período de 1º de janeiro de 2013 a 31 de dezembro de 2018. Para a previsão do preço de um dia D, os autores utilizaram os dados de D-1, D-2 e D-3 dias para treinamento dos modelos. Os autores utilizaram as medidas de acurácia, erros do tipo I e II, e F1 Score para verificar a qualidade dos resultados obtidos no estudo.

Souza (2021) realizou a classificação de sequências de preços de fechamento entre as classes *sell* e *buy*. O autor utilizou dados do período de 01 de janeiro a 31 de dezembro de 2019

da Petrobrás (PETR4). Ele realizou três experimentos com sequências contendo cinco, dez e vinte últimos dias considerando o dia D. Nesses experimentos, ele utilizou técnicas de Redes Neurais e árvore de decisão avaliando os resultados por meio das medidas de Acurácia e Matriz de Confusão.

Barros (2021) utilizou Redes Neurais Artificiais para predição de preços de fechamento das ações ordinárias da Vale (VALE3). Ele coletou dados do período entre 2 de janeiro de 2017 a 1 de janeiro de 2019, selecionou 12 indicadores técnicos e, em seus experimentos gerou 320 configurações diferentes de RNAs. Como saída das RNAs, o autor obteve a predição dos preços dos próximos cinco dias. Ele utilizou as medidas *Mean Squared Error* (MSE) e Coeficiente de Determinação (R²) para a verificação do desempenho dos modelos de RNA e seleção do melhor modelo (combinação que gerou a menor soma de erros percentuais para d+1, d+2, ..., d+5. Para análise dos resultados, o autor priorizou o erro percentual entre o valor real e o predito em cada dia predito. O melhor modelo gerou erros 1,50% para d+1; 1,66% para d+2; 1,52% para d+3; 1,60% para d+4; e 1,76% para d+5, totalizando um erro de 8,04% na predição dos cinco dias.

Diferente de Werner, Bisognin e Araújo (2020), que fizeram predição de volume de negociações, de Parray et al (2020), que fizeram predição do movimento de preços de empresas pertencentes a um índice estrangeiro utilizando uma quantidade fixa de dias de treinamento dos modelos, e de Souza (2021), que realizou a classificação de sequências de preços em duas classes, neste trabalho foram realizadas predições dos preços das ações de uma empresa brasileira. Em relação a Barros (2021), além das técnicas e medidas diferentes, também utilizou dados de outra empresa brasileira. Por fim, neste trabalho, foram utilizados dados de uma única empresa brasileira, testou-se com qual quantidade de dias de treinamento poderia se obter o menor erro na predição e utilizou-se a métrica RMSE na avaliação dos modelos.

Materiais e Métodos

Neste estudo foi utilizada a linguagem de programação Python v.3.7.13, juntamente com as seguintes bibliotecas: Pandas v.1.3.5, Pandas Datareader v. 0.10, Sci-kit Learn v.1.0.2 e Facebook Prophet v.1.1, e o ambiente online de programação Google Colaboratory (Colab) v.1.

Nos primeiros testes realizados utilizou-se um conjunto de dados com cotações de 10 anos e outro com cotações de 1 ano (considerando como referência o mês de Agosto de 2022) para predizer as cotações dos últimos cinco dias do período das ações preferenciais da Petrobrás

(PETR4). Porém, observou-se um RMSE (*Root Mean Square Error*) de, aproximadamente, R\$ 3,00, que foi considerado um erro grande quando se trata de investimentos em bolsa de valores. Então, optou-se por utilizar dados de três períodos com, aproximadamente, 60 dias cada um (equivalentes a, aproximadamente, 60 pregões realizados no período).

Os dados técnicos (Preços de Abertura, de Fechamento, Mínimo, Máximo, Volume de Negociações e Valor de Fechamento Ajustado) foram obtidos por meio da biblioteca Pandas Datareader, que recebeu como parâmetros: o código da ação “PETR4.SA”, e as datas de início e fim de cada período. Os dados foram coletados nos períodos: i) Período 1: 01/01 a 31/03/2022; ii) Período 2: 01/03 a 31/05/2022; e iii) Período 3: 01/05 a 31/07/2022. Os períodos possuem uma sobreposição de um mês para que a biblioteca utilizada para predições pudesse verificar os padrões iniciais do período.

Com o objetivo de verificar quantos dias seriam necessários para treinar o modelo e obter o menor erro na predição dos preços dos últimos 5 dias de cada período, separou-se os últimos 5 dias para teste e a quantidade de dias do conjunto de treino variou de 5 a N (N = número de dias do período menos os 5 dias separados para teste) com incrementos de 5 dias. Esses conjuntos foram nomeados como T5, T10, T15, T20, T25, T30, T35, T40, T45, T50, T55 e T60. O processo de treino, teste e avaliação do modelo foi realizado para cada conjunto de cada período. Dessa forma, utilizou-se, por exemplo, os 5 dias (T5) imediatamente anteriores ao conjunto de teste para treinar o modelo e fazer a predição do conjunto de testes. Os resultados foram avaliados utilizando-se a métrica RMSE para que fosse possível manter os valores do erro em Real (mesma unidade das cotações). Para comparação com o trabalho de Barros (2021), calculou-se o percentual do erro entre o preço real e o preço predito.

A predição foi realizada utilizando-se a biblioteca Facebook Prophet com seus parâmetros padrão. Não foi necessário tratamento para utilização de dias úteis pelo Prophet, pois os dados coletados são somente de dias úteis. Além disso, a utilização ou não dos feriados brasileiros disponibilizados pelo Prophet não afetaram as predições nos experimentos realizados. Portanto, optou-se por não os incluir.

Resultados e Discussão

Nesta seção, são apresentados e discutidos os resultados de treinamento e teste para conjunto e período. Inicialmente, é apresentado o processo de coleta dos dados e divisão dos

conjuntos. Em seguida, é apresentada uma análise descritiva dos dados coletados e, por último, são discutidos os resultados das previsões utilizando o Facebook Prophet.

Processo Coleta dos Dados e Divisão dos Conjuntos

Os dados de cotação da Petrobrás (PETR4) foram coletados utilizando-se a biblioteca Pandas Datareader. Ao informar o código da ação e um período, essa biblioteca faz o download dos dados técnicos: Preços de Abertura, de Fechamento, Mínimo, Máximo, Volume de Negociações e Valor de Fechamento Ajustado, além da data do pregão.

A Figura 1 apresenta um exemplo de cinco registros obtidos por meio do Pandas Datareader. Essa base de dados possui 7 colunas: i) *Date*: data do pregão; ii) *High*: mostra a maior cotação do dia; iii) *Low*: referencia a menor cotação do dia; iv) *Open*: cotação na abertura do pregão; v) *Close*: cotação no fechamento do pregão; vi) *Volume*: valores correspondentes ao volume de negociações do dia; vii) *Adj Close*: cotação de fechamento ajustada por algum evento (por exemplo: pagamento de proventos e desdobramentos). Por exemplo, no dia 03/01/2022, o maior valor negociado foi R\$ 29,21, o menor valor foi R\$ 28,53, os valores de abertura e fechamento foram R\$ 28,54 e R\$ 29,09, o preço de fechamento ajustado foi R\$ 19,43 e volume de negociação foi R\$ 52.704.700.

Figura 1 - Exemplos de registros da base de dados.

	High	Low	Open	Close	Volume	Adj Close
Date						
2022-01-03	29.219999	28.530001	28.540001	29.090000	52704700	19.430082
2022-01-04	29.400000	28.910000	29.160000	29.200001	51739200	19.503555
2022-01-05	29.270000	27.940001	29.190001	28.070000	78459800	18.748793
2022-01-06	28.650000	27.840000	28.290001	28.049999	61163100	18.735434
2022-01-07	28.290001	27.820000	28.110001	28.180000	47507600	18.822266

Fonte: Dados dos Autores (2022)

A Petrobrás possui ações ordinárias (código PETR3) e ações preferenciais (PETR4), que possuem maior volume de negociações diárias (maior liquidez). Então, para a obtenção dos dados, informou-se o código “PETR4.SA” e as datas de início e fim de três períodos: i) Período 1: 01/01 a 31/03/2022; ii) Período 2: 01/03 a 31/05/2022; e iii) Período 3: 01/05 a 31/07/2022.

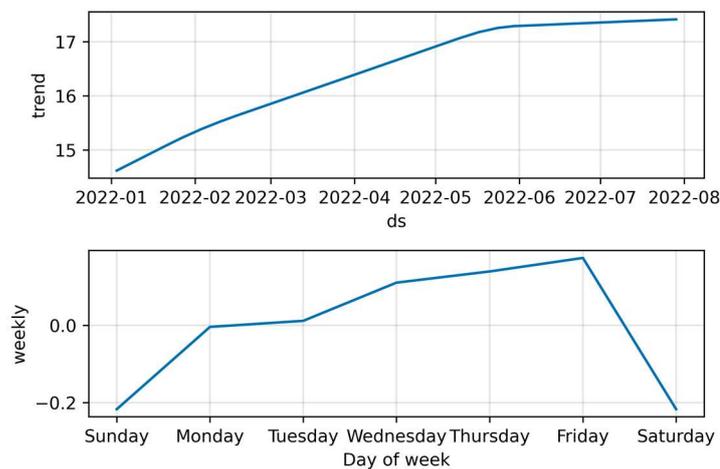
Ao final, obteve-se 61 registros no Período 1, 62 registros no Período 2 e 64 registros no Período 3. Como só existem pregões em dias úteis, a quantidade de pregões (dados obtidos) não coincide com a quantidade exata de dias do período (dias corridos).

Para a predição, foi escolhida a cotação de fechamento ajustada (“Adj Close”) associado a cada data. Como a biblioteca Prophet espera, como entrada, um *dataframe* com duas colunas: i) “ds”: com as datas; e ii) “y”: com os valores, criou-se um novo *dataframe* somente com esses dados. Os valores da coluna “y” são as cotações do atributo “Adj Close” e serão comparadas às cotações preditas.

Análise Descritiva

Utilizando o Prophet para analisar os componentes da série temporal do período completo (01/01 a 31/08/2022), observa-se na Figura 2, que as cotações têm uma tendência geral (*trend*) de alta mais acentuada de janeiro a junho e, de junho a agosto, têm crescimento menos acentuado. Além disso, nota-se também na Figura 2 que a sazonalidade semanal (*weekly*) cresce durante a semana sendo menor nas segundas-feiras e maior nas sextas-feiras.

Figura 2 - Tendência e sazonalidade do período completo (01/01 a 31/08/2022).

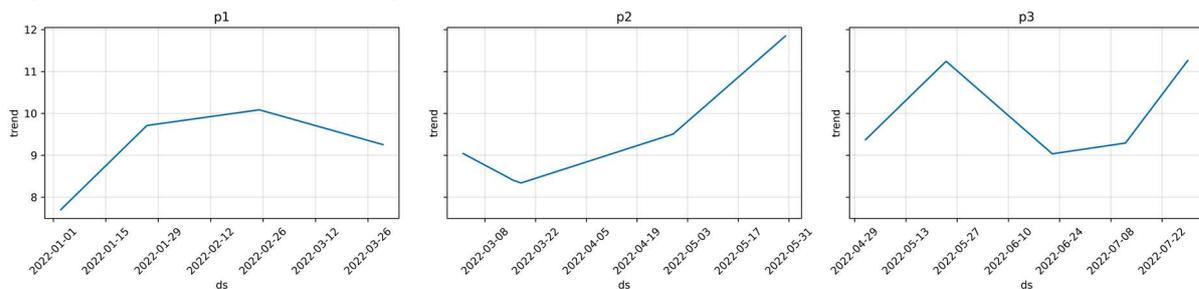


Fonte: Dados dos Autores (2022)

A Figura 3 apresenta a tendência (*trend*) em cada um dos períodos (p1, p2 e p3). Apesar da tendência geral de alta (Figura 2), observa-se na Figura 3 uma tendência de baixa no Período 1 (p1) e tendências de alta nos Períodos 2 (p2) e 3 (p3). A sazonalidade semanal também foi

analisada em cada período, mas ela se mantém estável em todos os dias da semana dos três períodos.

Figura 3 - Tendência em cada período.

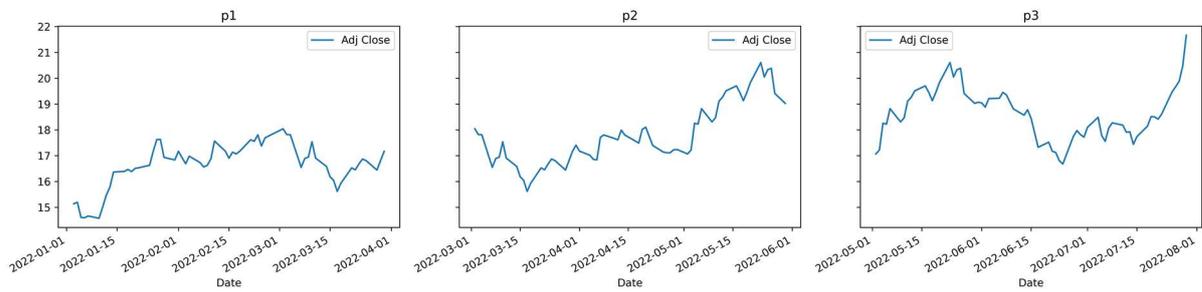


Fonte: Dados dos Autores (2022)

A base de dados, considerando-se os dados dos três períodos, possui 187 registros, nenhum valor faltante e dois atributos: i) ds: data do pregão; e ii) y: cotação (fechamento ajustado) em moeda Real (R\$). De modo estratificado: i) o Período 1 possui 61 registros, valor médio das cotações igual a R\$ 16,63 (desvio padrão de R\$ 0,85) e, os valores mínimo e máximo iguais a R\$ 14,57 e R\$ 18,04; ii) o Período 2 possui 62 registros, o valor médio de R\$ 17,78 com um desvio padrão de R\$ 1,21, e valores mínimo e máximo de R\$ 15,61 e R\$ 20,60; e iii) no Período 3, obteve-se 64 registros com valor médio das cotações de R\$ 18,65, desvio padrão de R\$ 1,01, e valores mínimo e máximo do preço de fechamento ajustado foram R\$ 16,67 e R\$ 21,66.

A variação das cotações em relação aos dias de cada período pode ser observada na Figura 4. Nota-se que, aproximadamente, em 15 dias no início no Período 1, as cotações estavam abaixo de R\$ 15,00 e, após esses 15 dias, as cotações oscilaram entre R\$ 16,00 e R\$ 18,00. No Período 2, nota-se uma tendência de alta nas cotações com valores iniciando em, aproximadamente, R\$ 18,00 e subindo até, aproximadamente, R\$ 19,00 no final do período com um valor máximo de R\$ 20,60 em 26/05/2022. No Período 3, as cotações oscilaram cerca de R\$ 4,60 partindo de, aproximadamente, R\$ 17,00 em 01/05, subindo até R\$ 21,60 em 23/05 e, em seguida, caindo até R\$ 16,67 em 24/06 e voltando a subir até R\$ 21,66 em 29/07.

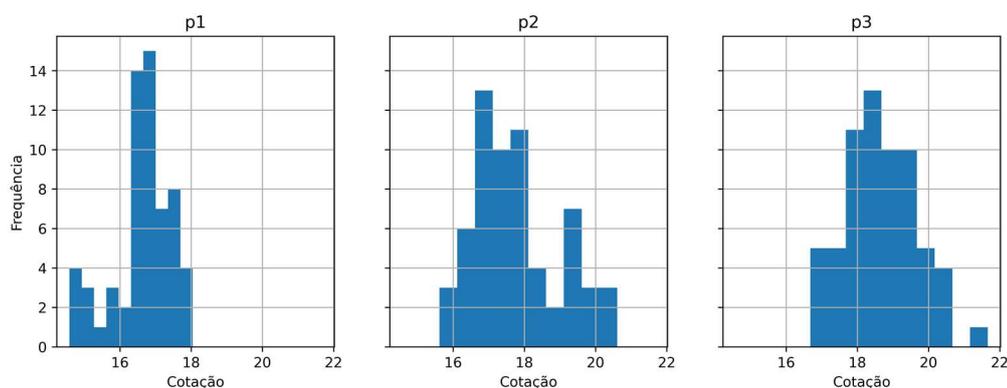
Figura 4 - Variações das cotações nos três períodos.



Fonte: Dados dos Autores (2022)

Após analisar as variações durante o tempo, analisou-se as distribuições de frequência das cotações em cada período. A Figura 5 apresenta histogramas com a distribuição de frequência das cotações para cada período: Período 1 (p1), Período 2 (p2) e Período 3 (p3). Observa-se que, no Período 1, a distribuição é assimétrica à esquerda e as cotações se concentram entre R\$ 16,00 e R\$ 18,00. No entanto, nos períodos seguintes, as cotações migram gradativamente para valores superiores a R\$ 18,00, com uma distribuição assimétrica à direita no Período 2, com valores ainda se concentrando entre R\$ 16,00 e R\$ 18,00, mas com presença de valores superiores a R\$ 18,00, enquanto, no Período 3, a distribuição se apresenta mais simétrica com valores se concentrando entre R\$ 17,00 e R\$ 20,50 e com, apenas, um valor de R\$ 21,66 que se destaca do restante da distribuição.

Figura 5 - Distribuição de frequência das cotações de cada período.

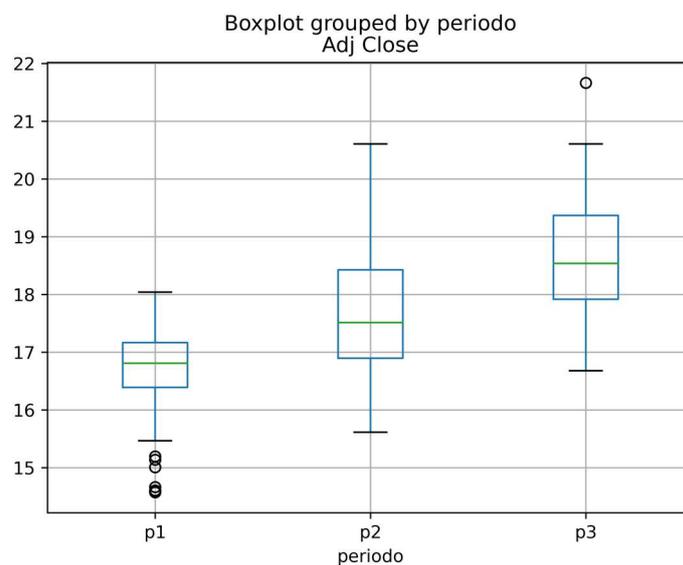


Fonte: Dados dos Autores (2022)

Para analisar a variabilidade das cotações e visualizar possíveis *outliers* (pontos fora da curva), utilizou-se diagramas de caixa (*boxplots*) apresentados na Figura 6. Nota-se, na Figura 6,

que as cotações aumentaram a cada período, pois as caixas estão posicionadas em valores crescentes. O tamanho menor da caixa do Período 1, em relação às demais, indica uma concentração maior das cotações em valores mais próximos, enquanto as caixas maiores indicam maior variação nas cotações. Observa-se também a existência de possíveis *outliers* (pequenos círculos) abaixo da cotação mínima do Período 1 (p1) e acima da cotação máxima do Período 3 (p3). Esse possível *outlier* do Período 3 aparece também no histograma de p3 como valor mais à direita do histograma.

Figura 6 - Boxplot por período.



Fonte: Dados dos Autores (2022)

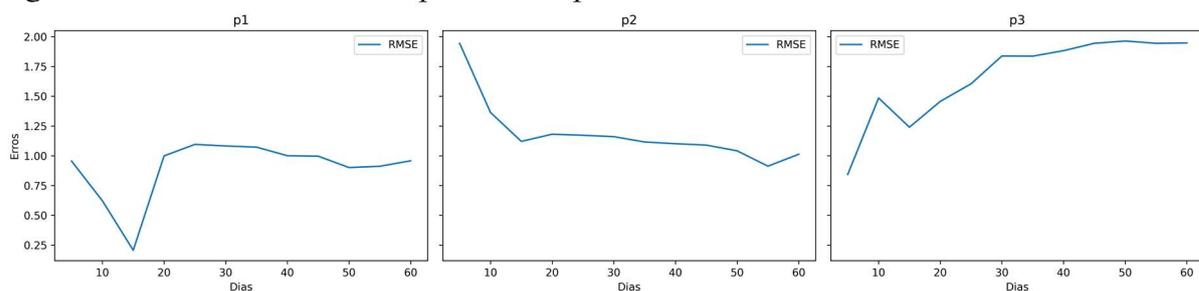
Por meio das análises realizadas, percebe-se que os conjuntos de dados possuem características diferentes entre si e isso permitiu verificar o comportamento das predições em diferentes cenários. Além disso, no mercado de ações, altas variações na cotação estão relacionadas a um maior risco no investimento. Essas variações são calculadas por meio do desvio padrão das cotações. Então, o risco de investimento no Período 1 foi o menor, pois o desvio padrão foi de R\$ 0,85, enquanto o Período 2 foi o de maior risco, pois o desvio padrão foi de R\$ 1,21, seguido do Período 3 com R\$ 1,01. Dessa forma, especuladores poderiam obter retornos ou prejuízos expressivos no Período 2 e no Período 3.

Análise das Predições

Para o treinamento e teste dos modelos, os dados de cada período foram separados em dois conjuntos: i) teste: contém as cotações dos últimos 5 dias do período; e ii) treinamento: subdividido em conjuntos com tamanhos variáveis. Para cada subconjunto de treinamento, foi criado um modelo que foi testado com o conjunto de testes (ds), obtendo-se os valores preditos, chamados de “yhat”. Esses valores foram utilizados com o “y” (valores reais) para o cálculo do RMSE, utilizado para avaliar a qualidade do modelo e os erros de predição.

Os valores do RMSE gerados durante os testes dos modelos foram tabulados e podem ser visualizados na Figura 7 para cada período. Em cada gráfico, tem-se o tamanho dos subconjuntos de treinamento no eixo x e o valor do erro (RMSE) no eixo y. Observando-se a Figura 7, nota-se que, no Período 1 (p1), o RMSE começa em R\$ 0,95, quando se treina o modelo com dados de 5 dias (T5), cai até R\$ 0,20 quando se utiliza o T15, e passa a oscilar ao redor de R\$ 1,00 quando se aumenta a quantidade de dados de treinamento. No Período 2, o RMSE começa em R\$ 1,94 com T5, cai para R\$ 1,12 quando se utiliza o T15, sobe para R\$ 1,18 com T20 e começa a cair lentamente até R\$ 0,91 com T55, quando volta a subir com T60. Para o Período 3, obteve-se os seguintes resultados: com o conjunto T5, o RMSE começa em R\$ 0,84, sobe para R\$ 1,48 com T10, cai até R\$ 1,24 com T15 e sobe até R\$ 1,94 com T60.

Figura 7 - Resultados do RMSE para os três períodos.

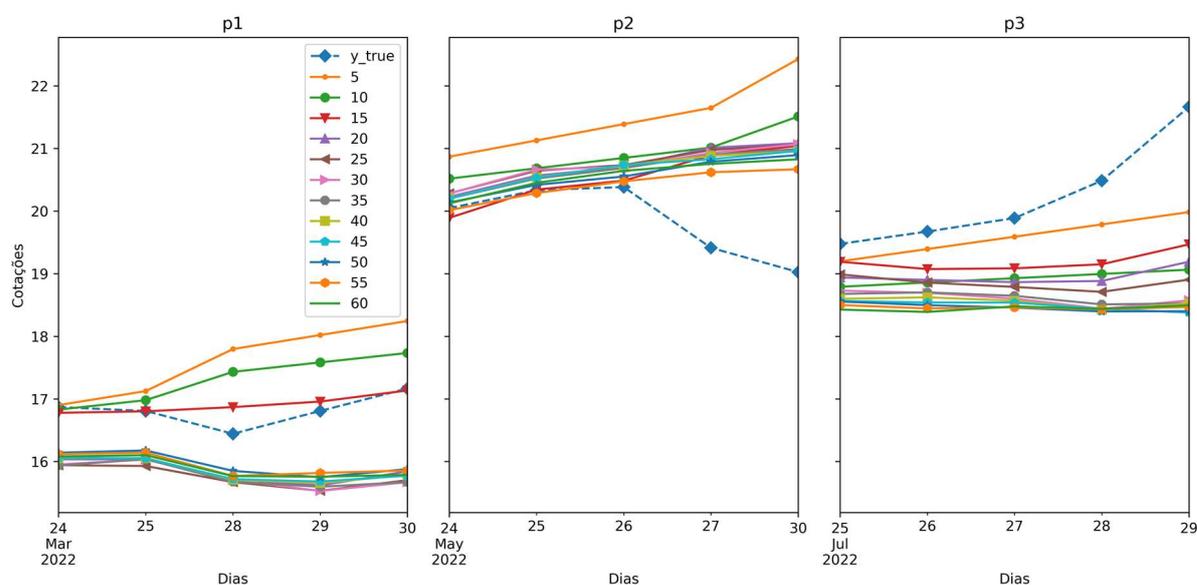


Fonte: Dados dos Autores (2022)

Considerando o desvio padrão e o RMSE de cada período, observa-se que períodos com menor desvio padrão (Período 1 e Período 3) tiveram menores erros quando foram utilizados menos dias de treinamento (15 dias no Período 1 e 5 dias no Período 3). Além disso, períodos com alta volatilidade (maior desvio padrão) têm maiores erros nas predições.

O RMSE representa o erro médio das previsões para os 5 dias de teste. Então, procurou-se comparar dia a dia as cotações reais e as previstas. Os gráficos da Figura 8 mostram as cotações reais e as previstas de cada período. No eixo x, tem-se os dias da semana, no eixo y, tem-se as cotações, e as linhas indicam os subconjuntos de treinamento. As linhas tracejadas na cor azul indicam as cotações reais (y_{true}).

Figura 8 - Cotações Reais e Preditas.



Fonte: Dados dos Autores (2022)

Analisando-se o gráfico do Período 1 (p1) na Figura 8, nota-se que, no dia 24/03, os modelos treinados com T5, T10 e T15 dias praticamente não erraram na previsão. O modelo T15 (linha vermelha com marcadores \square) também obteve alta precisão nos dias 25 e 30/03, mas não acompanhou a queda da cotação do dia 28/03 (uma segunda-feira). Verificando o site do Relações com Investidores⁵, encontrou-se um comunicado ao mercado sobre indicação de novos membros para o conselho diretor da empresa e essa indicação pode ter influenciado negativamente na cotação.

⁵Disponível em: <https://www.investidorpetrobras.com.br/acoes-dividendos-e-dividas/dividendos-e-jcp/>. Acesso em: 10 set. 2022.

No gráfico do Período 2, nota-se que T55 (indicado pela linha laranja com marcadores ) gerou predições com uma maior proximidade em relação aos valores reais nos dias 24, 25 e 26/05, mas não acompanhou as quedas dos dias 27 e 30/05. No Período 3, diferente dos outros períodos, os modelos não conseguiram precisão na predição em nenhum dia. O modelo que mais se aproximou foi o T5, que conseguiu predizer alta na cotação de maneira geral, mas não acompanhou a alta da cotação real que aconteceu a partir de 27/07. Consultando o site de Relação com Investidores da Petrobrás⁶, verificou-se que, no dia 28/07, a empresa declarou que pagaria cerca de R\$ 3,35 (valor bruto) em proventos. Essa declaração pode ter influenciado na alta da cotação.

Os dados indicam uma relação entre a volatilidade e a quantidade de dias necessários para treinar um modelo de predição das cotações, mas é necessário realizar mais experimentos para confirmar essa relação. Percebe-se também que a precisão dos modelos para predizer um ou dois dias a frente é alta, mas decai à medida que se aumenta o número de predições. Além disso, eventos externos podem afetar as cotações reais e os modelos podem não acompanhar essas variações nas cotações. Dessa forma, é prudente evitar predições para, por exemplo, mais de dois dias a frente para evitar a influência de eventos externos.

Barros (2021) chegou às mesmas conclusões em relação à predição de vários dias à frente, pois, em sua estratégia de predição com RNAs, os erros também aumentaram a partir do segundo dia de predição. Comparando o uso do Prophet com as RNAs, enquanto Barros (2021) consumiu cerca de 72h para testar todas as 320 configurações e conseguir um modelo que precisa de 12 atributos de entrada para gerar um erro de 1,5% predizendo um dia a frente, o treinamento e teste do Prophet com os três períodos e 12 conjuntos (36 combinações) e um atributo de entrada, consumiu cerca de 35 segundos, e os melhores modelos de cada período geraram erros de 0,54%, 0,17% e 1,43% para predizer um dia a frente.

Interessante observar que o modelo de Barros (2021) gerou predições com erro maior no segundo dia (1,66%) enquanto os modelos deste trabalho geraram erros menores no segundo dia: 0,03%, 0,15% e 1,41%. A limitação dos modelos deste trabalho está na dependência do monitoramento da volatilidade do mercado para definir a quantidade de dias necessária para o

⁶Disponível em: <https://www.investidorpetrobras.com.br/resultados-e-comunicados/comunicados-ao-mercado/>. Acesso em: 10 set. 2022.

treinamento. De maneira geral, treinar o Prophet com 15 dias para predizer um dia a frente gera um erro percentual médio de 0,92%, menor que o 1,5% de Barros (2021).

Em seu trabalho, Parray et al (2020) fixaram o uso de três dias anteriores ao dia atual para realizar a predição, argumentando que, segundo a Hipótese do Mercado Eficiente, o preço atual da ação reflete as informações mais recentes que o Mercado tem sobre aquela ação. No entanto, observou-se neste trabalho, que a quantidade de dias necessária para treinar um modelo pode depender da volatilidade da ação.

Os resultados da abordagem de predição de preços deste trabalho mostram que, de maneira prática, o uso do Prophet traz como benefícios a necessidade de quantidade pequena de dados para treinamento e teste, e redução do esforço para treinamento e testes de modelos que podem gerar erros aceitáveis na área de investimentos.

Conclusões

Neste trabalho foi realizada a predição de cotações das ações PETR4 utilizando a biblioteca Facebook Prophet, que utiliza o método de séries temporais. Nos primeiros testes de predição, considerando-se que técnicas de análise de séries temporais, normalmente utilizam grandes quantidades de dados, utilizaram-se dados de 10 anos e de 1 ano, mas o valor do RMSE foi considerado alto para a área de investimentos. Portanto, decidiu-se testar a quantidade de dias de treinamento necessária para predizer as cotações dos últimos 5 dias de um período, gerando o menor RMSE. Para os experimentos, importaram-se dados de três períodos com aproximadamente 60 dias úteis cada um, iniciando em janeiro e finalizando em julho de 2022.

Na análise exploratória, foram estudados os componentes da série temporal (tendência e sazonalidade), métricas e atributos relacionados às bases de dados de cada período, ou seja, valores relacionados com quantidade de registros, valores da média e de desvio padrão, valores mínimo e máximo, além das distribuições de frequência e variabilidade das cotações. Dessa forma, a análise exploratória permitiu entender a composição das bases de dados.

Para as predições, utilizou-se a biblioteca Facebook Prophet. O Prophet foi treinado e testado com diferentes subconjuntos de dados pertencentes a cada um dos três períodos, totalizando 36 combinações (3 períodos com 12 subconjuntos). Na avaliação da qualidade dos modelos, observou-se o RMSE obtido nas predições de cada período com cada número de dias

de treinamento. Na análise dos erros, percebeu-se uma possível relação com a volatilidade das cotações.

A volatilidade das cotações foi observada através da análise do desvio padrão de cada período. Considerando-se a quantidade de dias de treinamento com menores RMSE, observou-se também que do Período 1 para o Período 2, o desvio padrão aumentou e o RMSE aumentou e, do Período 2 para o Período 3, houve uma redução no desvio padrão e uma redução no RMSE. Dessa forma, verificou-se uma possível relação entre o desvio padrão e a quantidade de dias necessária para treinar o modelo: quanto mais volátil for a cotação, maior é o desvio padrão e maior é a quantidade de dias para treinamento. No entanto, são necessários mais experimentos para verificar essa correlação.

Ainda em relação à quantidade de dias necessária para treinamento dos modelos para se obter o menor erro, percebeu-se que, de maneira geral para este conjunto de dados, pode-se utilizar 15 dias para treinamento e predição de um ou dois dias a frente, pois o erro pode aumentar quando faz-se predições de vários dias a frente. Por exemplo, o erro percentual médio entre os três períodos quando utiliza-se 15 dias para predizer um dia a frente foi de 0,92% e, para predizer dois dias à frente, foi de 1,03% para as ações preferenciais da Petrobrás, erros menores que 1,5% e 1,66% de outra estratégia encontrada na literatura, que utilizou 12 atributos de entrada e dados diários de 2 anos das ações ordinárias da Vale.

As cotações das ações sofrem, por exemplo, com a situação econômica do país, notícias sobre seus setores de atuação ou mesmo em relação aos membros de sua diretoria. Portanto, a quantidade de dias e os erros gerados devem ser testados com mais conjuntos de dados e também com outras ações para que se possa generalizar os resultados obtidos neste trabalho.

A volatilidade das ações, se explorada com cuidado, pode fazer com que especuladores alcancem resultados positivos. A análise técnica, juntamente com o aprendizado de máquina, pode ser utilizada para prever tendências nas cotações e potencialmente aumentar os lucros na negociação de ações. Portanto, a predição realizada no trabalho pode ser um auxílio nos investimentos, o que também contribui na maximização dos ganhos de capital.

Para trabalhos futuros, sugere-se a realização de estudos utilizando métodos de aprendizado incremental, além da abordagem de outras combinações de períodos, para que assim a relação entre volatilidade e quantidade de dias de treinamento seja melhor observada.

Referências

B3. **Ações**. Disponível em: https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/acoes.htm. Acesso em: 14 abr. 2022.

BARROS, G. C. **Redes Neurais Artificiais para Predição do Preço de VALE3 na Bolsa de Valores Brasileira**. 2021. 61 p. Trabalho de Conclusão de Curso (Graduação) - Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Universidade Federal de São Paulo - Campus Diadema, São Paulo.

CVM. **Como negociar ações?**. Disponível em: https://www.investidor.gov.br/menu/Menu_Investidor/Old/Valores_Mobiliarios/Acoes/Como_negociar_acoes.html. Acesso em: 16 abr. 2022.

FACEBOOK PROPHET. **Previsão em escala**. Disponível em: <https://facebook.github.io/prophet/>. Acesso em: 17 set. 2022.

IGARASHI, W.; VALDEVIESO, G. S.; IGARASHI, D. C. C. Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. **Brazilian Journals of Business**, Curitiba, v. 3, n. 1, p. 470-486, jan./mar. 2021.

JANSEN, S. **Machine Learning for Algorithmic Trading**. 2ed. Birmingham: Packt Publishing, 2020.

LAGIOIA, U. **Fundamentos do mercado de capitais**. 2ed. São Paulo: Atlas, 2009. 212 p.

LAMEIRA, V. **Negócios em bolsas de valores: estratégias para Investimentos**. São Paulo: Alaúde, 2005.

MITCHELL, T. M. **Machine learning**. Nova York: McGraw-Hill, 1997.

PARRAY, I. R.; KHURANA, S. S.; KUMAR, M.; ALTALBE, A. A. Time series data analysis of stock price movement using machine learning techniques. **Soft Computing**, 24, p. 16509-16517, 2020. Disponível em: <https://link.springer.com/article/10.1007/s00500-020-04957-x>. Acesso em: 15 mai. 2022. DOI: <https://doi.org/10.1007/s00500-020-04957-x>.

SILVA, J.C.; ZANINI, M. J.; SAURIN, G. Análise econômico-financeira de sociedades anônimas: decisão para investimento na bolsa de valores. In: Encontro Científico Cultural Interinstitucional, 13º, 2015. **ECCI**. Out. 2015. ISSN 1980-7406. p. 1-10.

SINGH, V.; DUBEY, S. K. Opinion mining and analysis: A literature review. In: International Conference - Confluence The Next Generation Information Technology Summit, 5., 2014, Noida. **Confluence**. p. 232-239.

SOUZA, W. B. C. **Mineração de dados aplicada à previsão de preços de ações utilizando Weka**. 2021. 68 p. Trabalho de Conclusão de Curso (Graduação) - Escola de Ciências Exatas e da Computação, Pontifícia Universidade Católica de Goiás (PUC-GO), Goiânia.

WERNER, L.; BISOGNIN, C.; ARAUJO, C. W. Análise de técnicas de previsão: um estudo de caso para o volume de ações da Petrobras. **Brazilian Journal of Development**, Curitiba, v. 6, n. 1, p.1103-1115, 2020. Disponível em: <https://brazilianjournals.com/ojs/index.php/BRJD/article/view/5979/5334>. Acesso em: 15 mai. 2022. DOI: <https://doi.org/10.34117/bjdv6n1-078>.