

É POSSÍVEL PREVER O PREÇO DO BARRIL DE PETRÓLEO?

Autoras : Joana Enes e Natália Assad

É POSSÍVEL PREVER O PREÇO DO BARRIL DE PETRÓLEO?

O preço do barril de petróleo é de suma importância para a economia mundial. Para se ter uma ideia da importância do setor, no ano de 2019, ele movimentou 3,8% do PIB global, equivalente a US\$ 3,26 trilhão.

À uma tendência de que seu aumento impacte o crescimento do déficit externo gerando um aumento de inflação, desemprego e uma consequente redução do PIB.

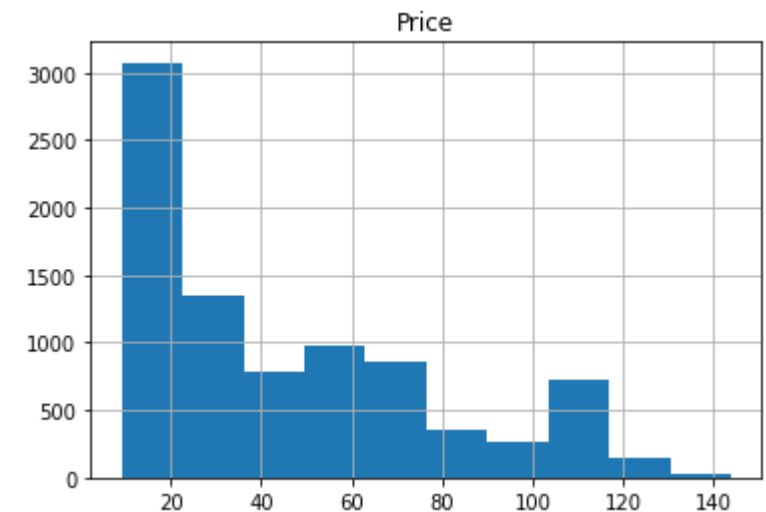
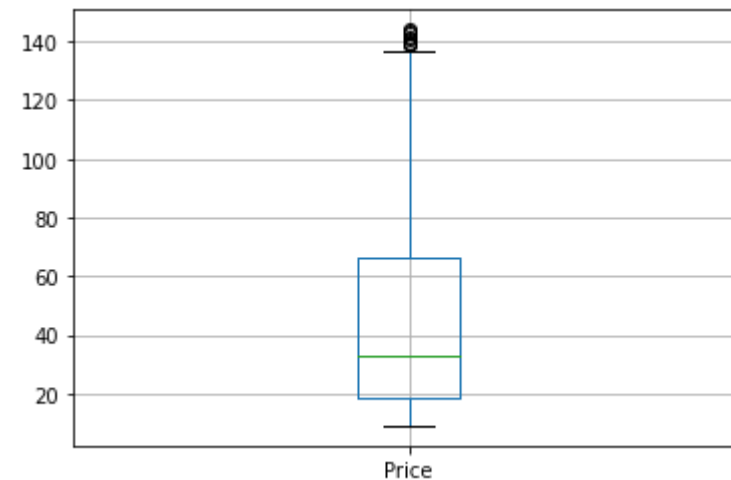
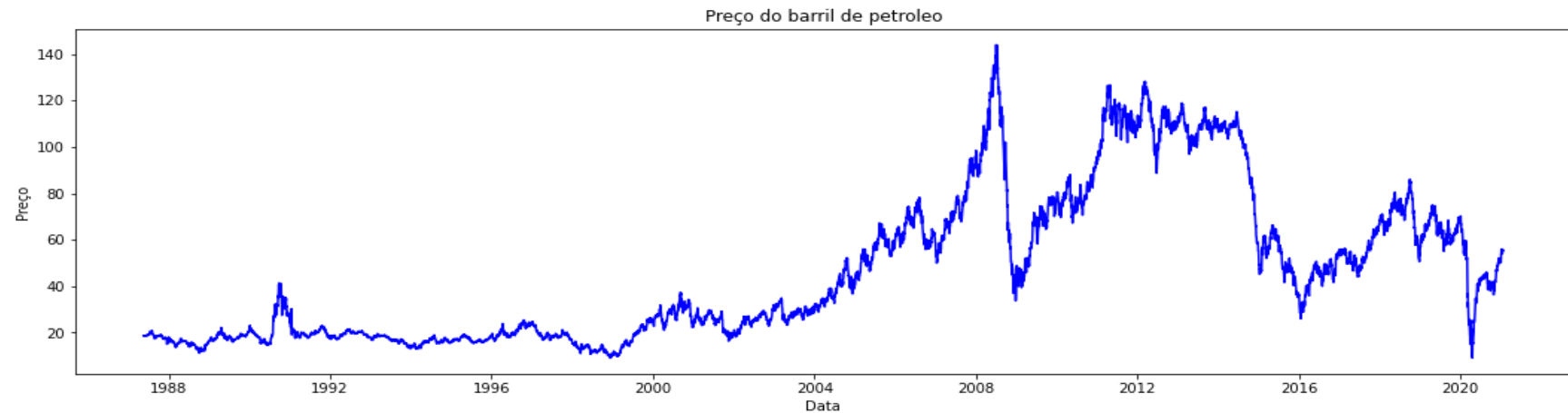
O custo de vida também é drasticamente afetado, pois o valor do barril, traz impacto para o custo dos derivados como gasolina e diesel que são essenciais para o funcionamento, produção e desenvolvimento de atividades essenciais.

Este trabalho tem por objetivo avaliar se é possível prever o preço do barril de petróleo através do uso de algoritmos específicos.

Exploração dos dados:

- Dados extraídos do kaggle;
- Dataset contempla dados diários de 1987 à 2021;
- Trata-se de uma serie temporal, onde o preço está variando de acordo com o tempo;

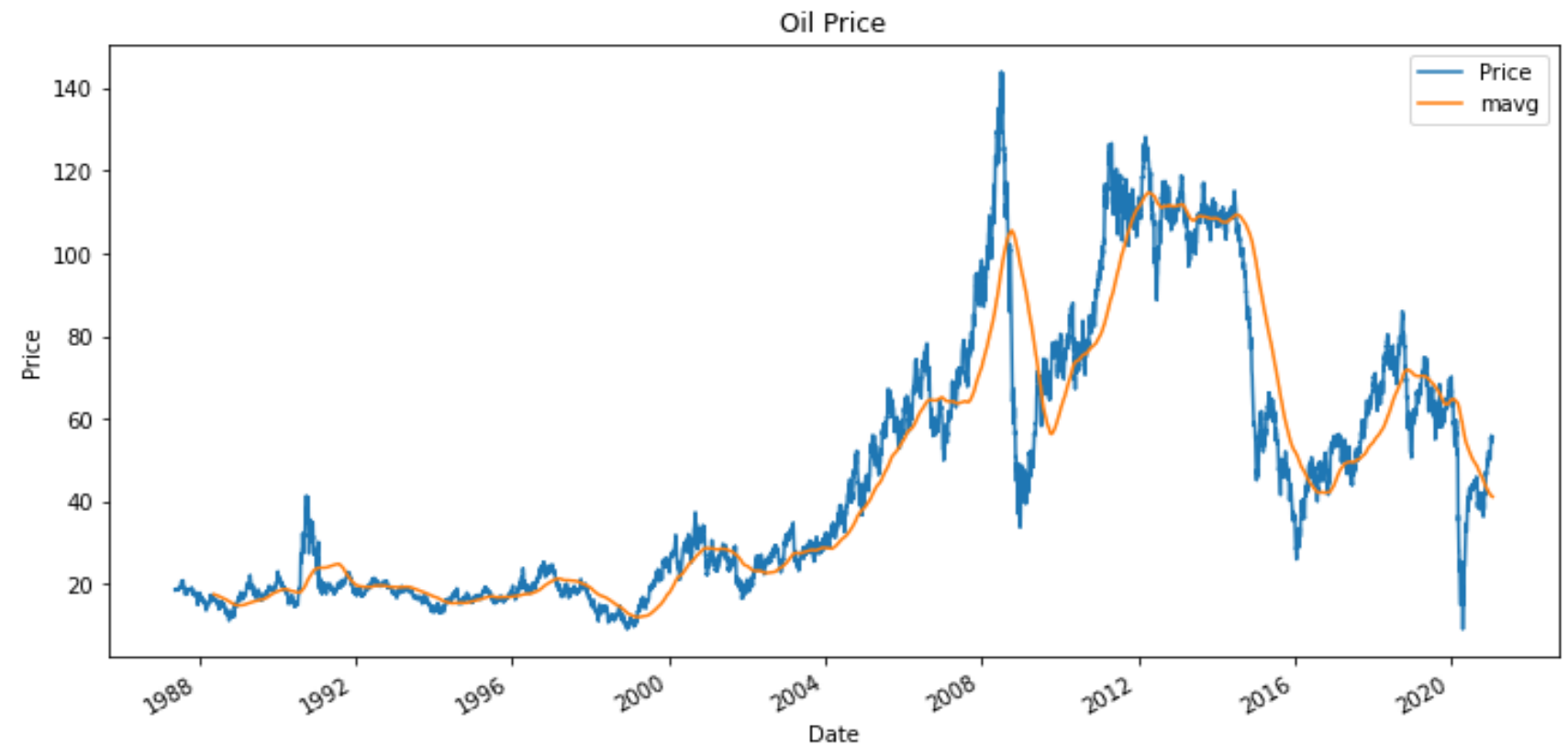
Séries Temporais



Média móvel - janela de 253 dias.

A média móvel nos ajuda a suavizar os dados com muitas flutuações e nos ajuda a ver melhor a tendência de longo prazo dos dados.

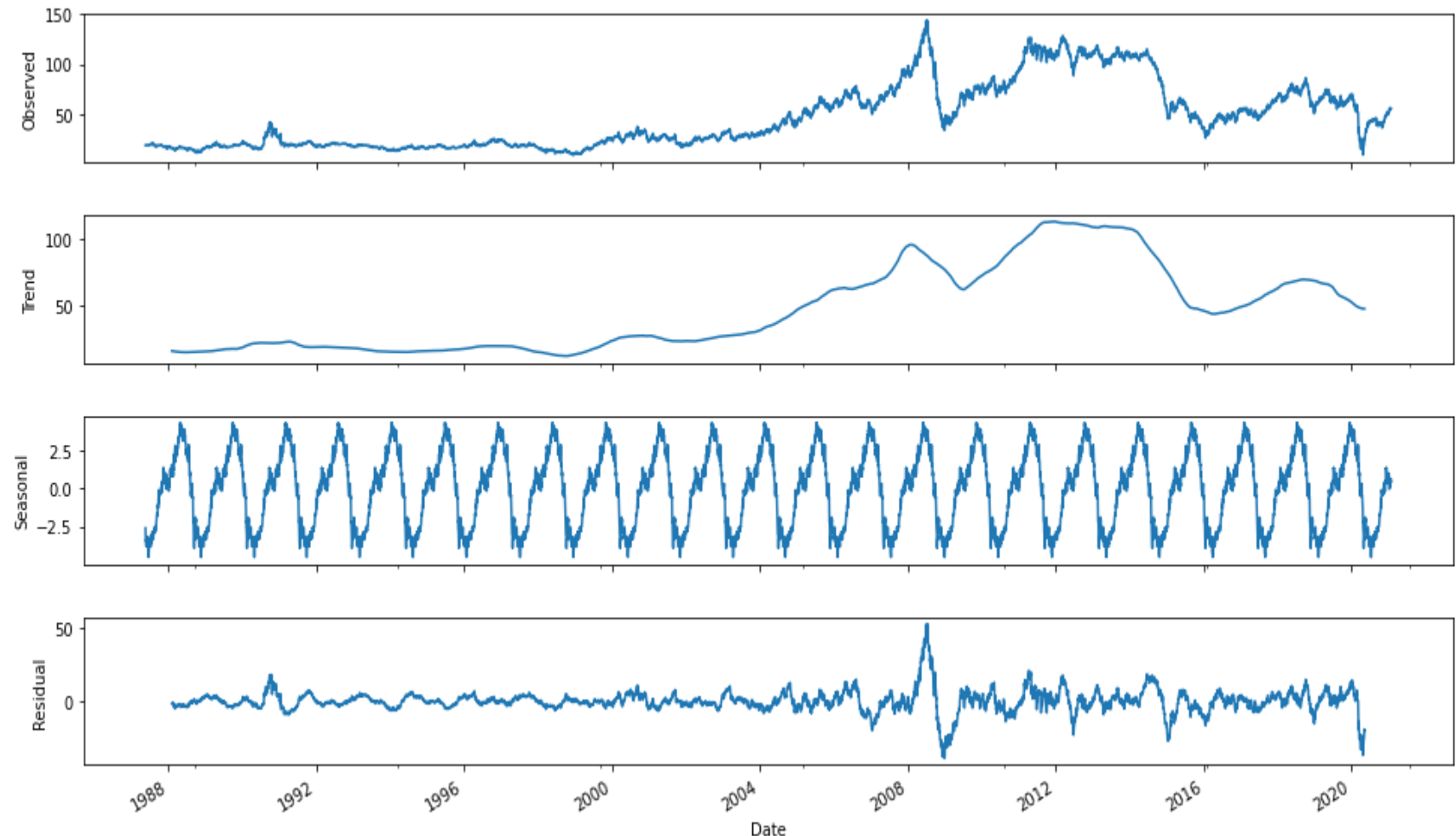
Média Móvel



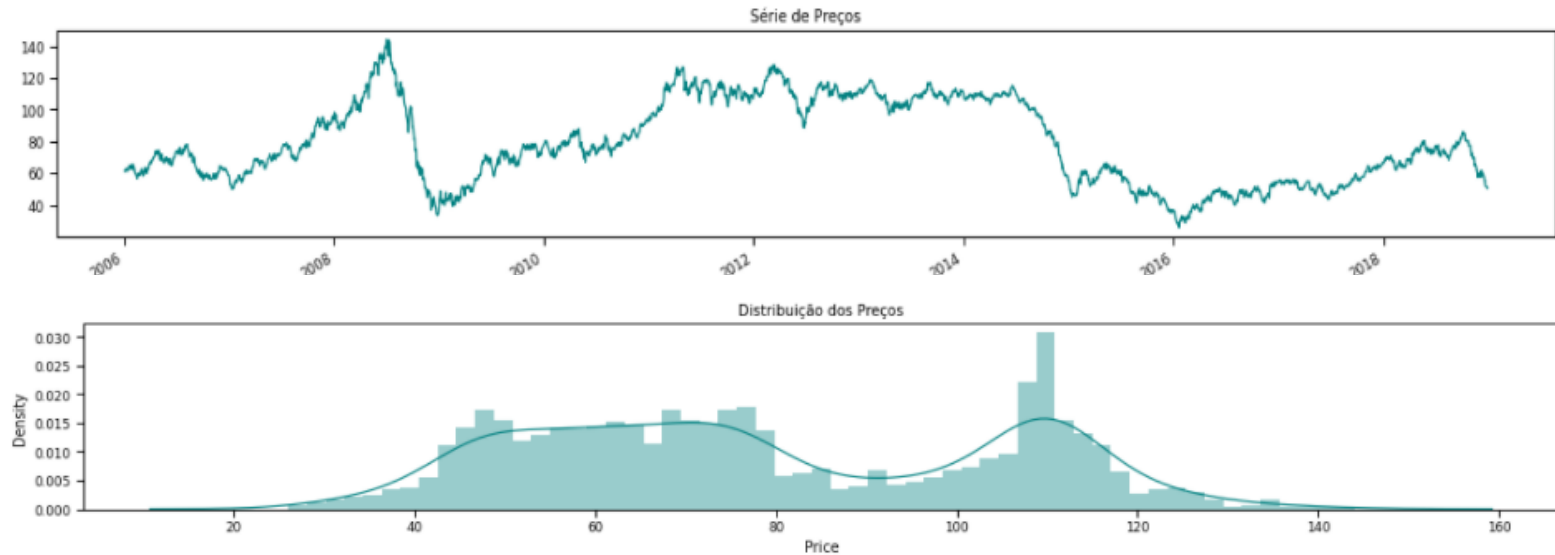
Toda serie temporal possui componentes que precisam ser explorados antes da aplicação de qualquer algoritmo.

Para isso fizemos decomposição da serie temporal conforme segue abaixo:

Componentes da Series Temporal

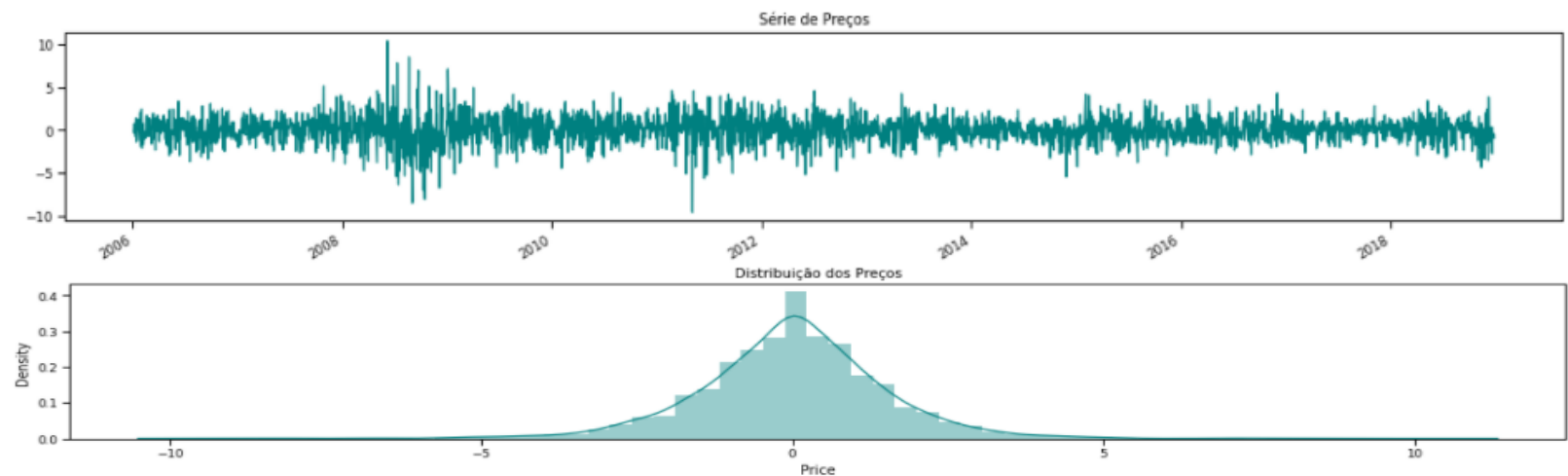


A nossa série tem um perfil não estacionária.



Para alguns algoritmos essa informação é relevante, pois precisaremos ter em mente qual será o esforço para torná-la estacionária.

Para tornar nossa série estacionária utilizamos a técnica de diferenciação, pois iremos trabalhar com ARIMA no decorrer desse estudo.



Estacionaridade

Existem testes específicos que podem confirmar ou não a estacionaridade de uma serie.

O teste que usamos em nosso estudo foi o Dickey-Fuller. Para considerarmos uma serie como estacionária precisamos ter o p-value abaixo 0,05. Seguem os resultados dos testes antes e depois do processo de diferenciação:

Resultados do teste de Dickey-Fuller(*antes da diferenciação*):

Teste Statístico	-1.6212
Valor-P	0.4721
Lags Usados	1.0000
Números de Observações	3281.0000
Valores Críticos (1%)	-3.4323
Valores Críticos (5%)	-2.8624
Valores Críticos (10%)	-2.5672

Resultados do teste de Dickey-Fuller(*após diferenciação*):

Teste Statístico	-55.3968
Valor-P	0.0000
Lags Usados	0.0000
Números de Observações	3281.0000
Valores Críticos (1%)	-3.4323
Valores Críticos (5%)	-2.8624
Valores Críticos (10%)	-2.5672

O p-value foi para 0 no segundo teste , confirmando nossa tese de que nesse momento temos uma serie estacionária.

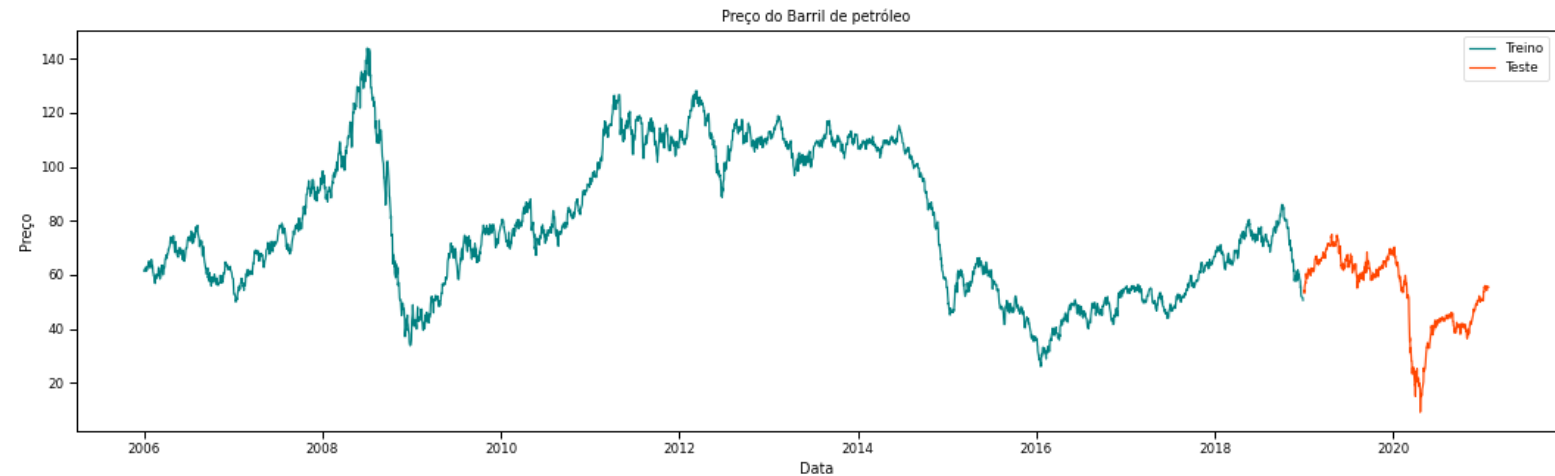
The background of the slide is a photograph of a sunset or sunrise over a field. The sky is filled with orange and yellow clouds. In the foreground, there is a field of crops. A large red rectangle is overlaid on the left side of the image, containing the word "Estacionaridade" in white text.

Estacionaridade

Por se tratar de uma serie temporal, a divisão do dataset entre treino e teste torna-se relevante, visto que a característica temporal precisa ser levada em consideração.

Optamos por seguir com a seguinte divisão:

- **Treino : Dados contidos entre 01/01/2006 até 31/12/2018;**
- **Teste: Dados a partir de 01/01/2019;**



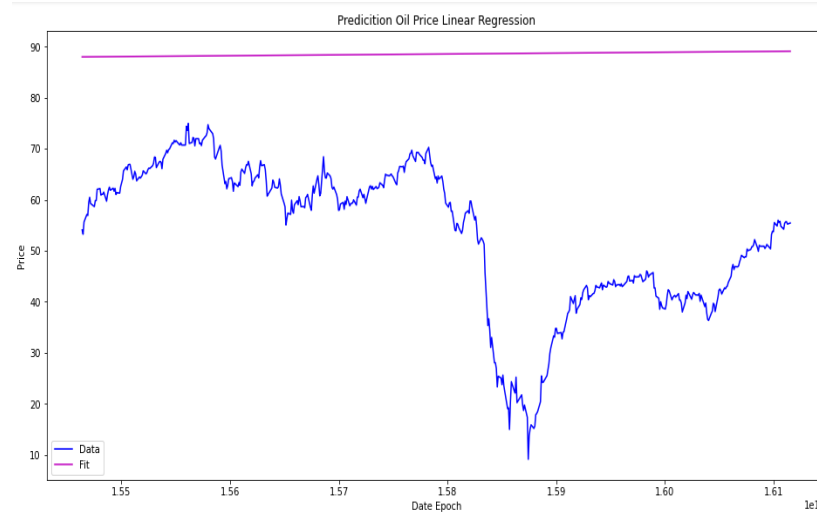
Neste estudo iremos trazer o resultado dos seguintes algoritmos abaixo:

- **Regressão Linear**
- **Support Vector Regression**
- **ARIMA**
- **LSTM**

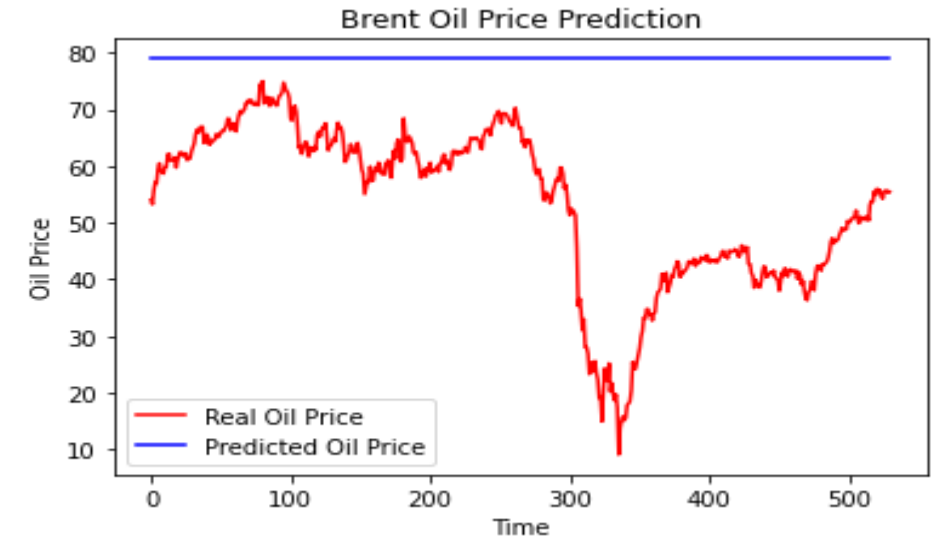
Métodos e Modelos

Os dois modelos de Regressão apresentaram péssimos resultados, confirmando nossa expectativa de que estes modelos não seriam apropriados para este tipo de problema

Regressão Linear & SVR



MSE: 1447.1488923221
RMSE: 38.041410230460265
R2: -6.2836694170232095



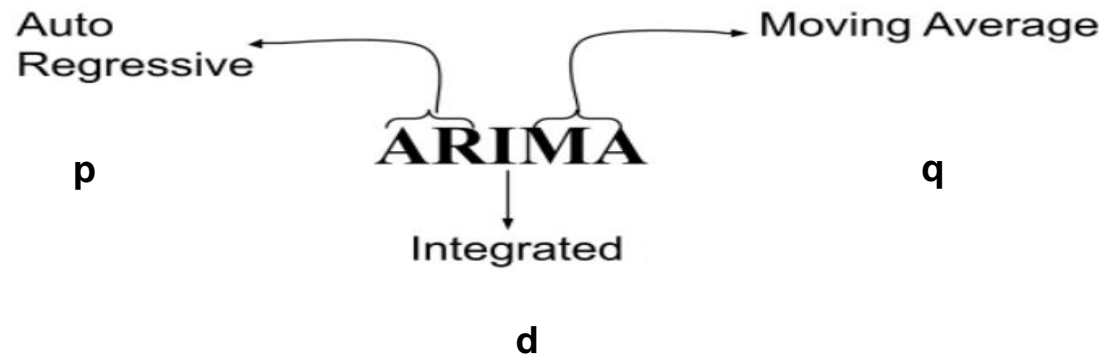
RBf MAE: 25.62350091631097
RBf MSE: 855.2478428496985
RBf RMSE: 29.244620750655983
RBf R2: -3.3045622948606956

ARIMA utiliza dados passados para prever o futuro, usando dois principais recursos: a autocorrelação e médias móveis.

ARIMA possui três componentes - AR (termo autoregressivo), I (termo diferencial) e MA (termo médio móvel).

Para aplicarmos o modelo precisamos definir os parâmetros para cada um desses componentes (p, d, q).

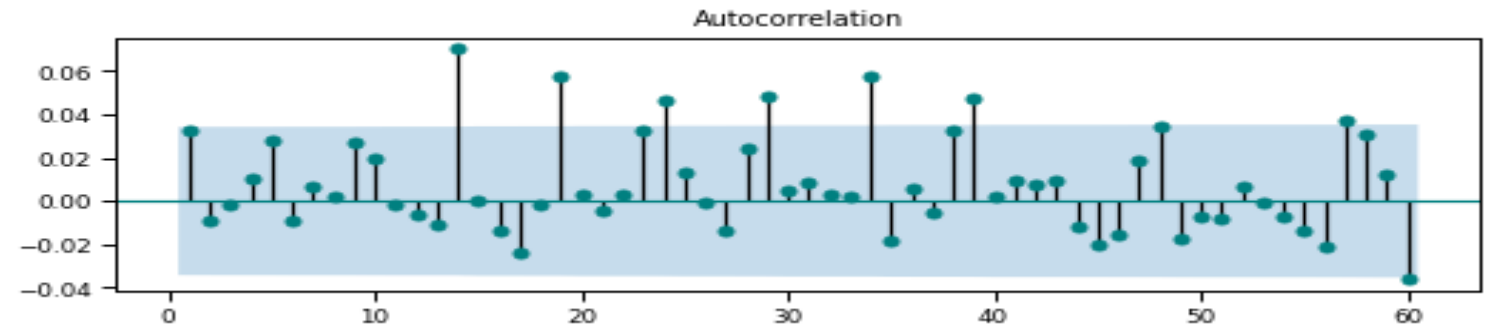
ARIMA



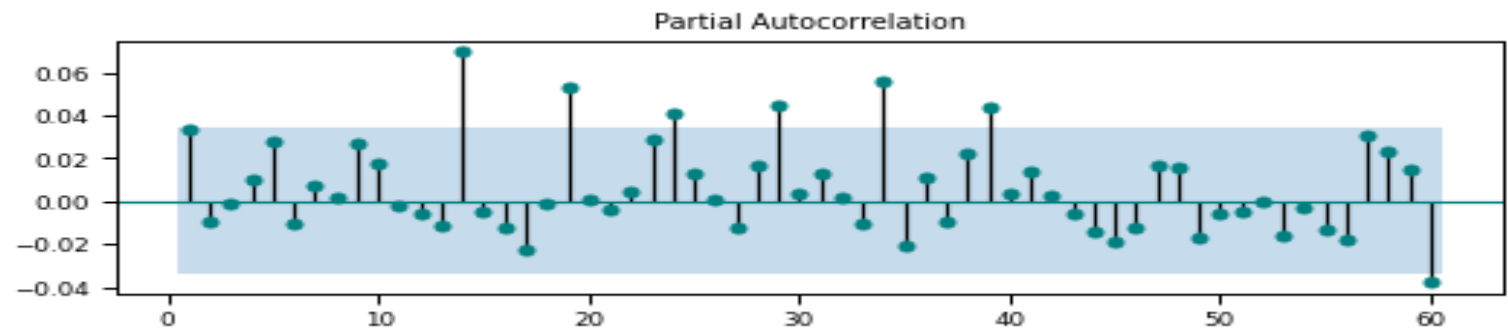
Onde :

- O valor de 'p' é determinado usando o gráfico PACF.
- O valor de 'd' é determinado pelo número de vezes que a operação de diferenciação é executada em série para torná-la estacionária.
- O valor de 'q' é determinado usando o gráfico ACF.

ACF



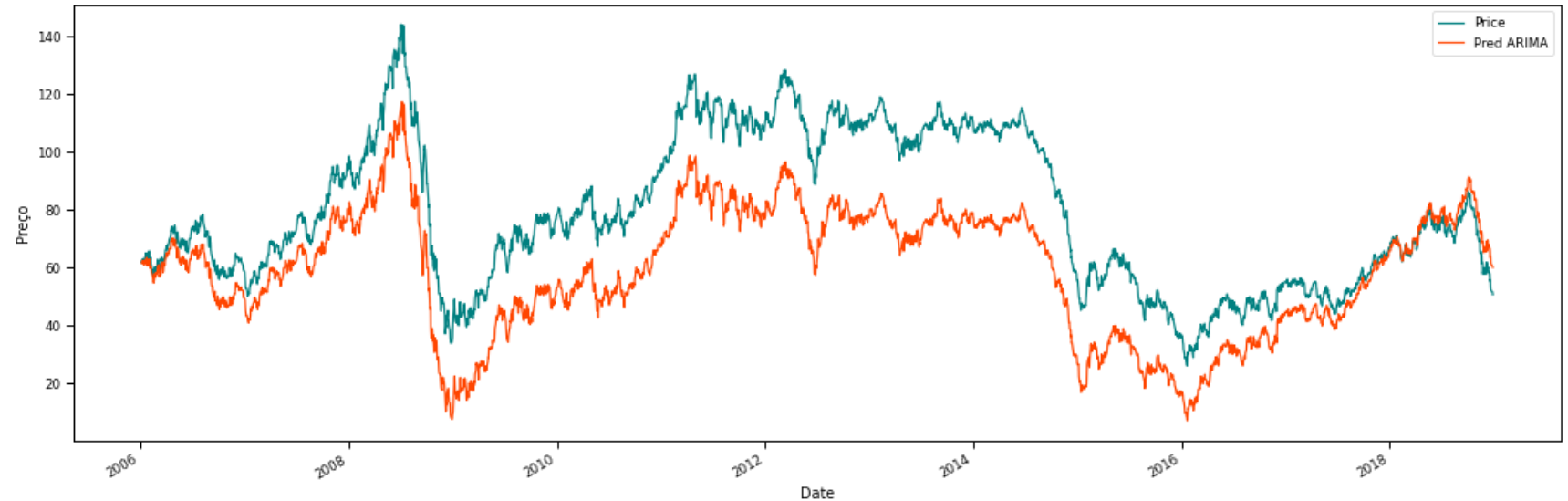
PACF



ARIMA
ACF/PACF

ARIMA Resultado

Modelo aplicado no conjunto de treino:



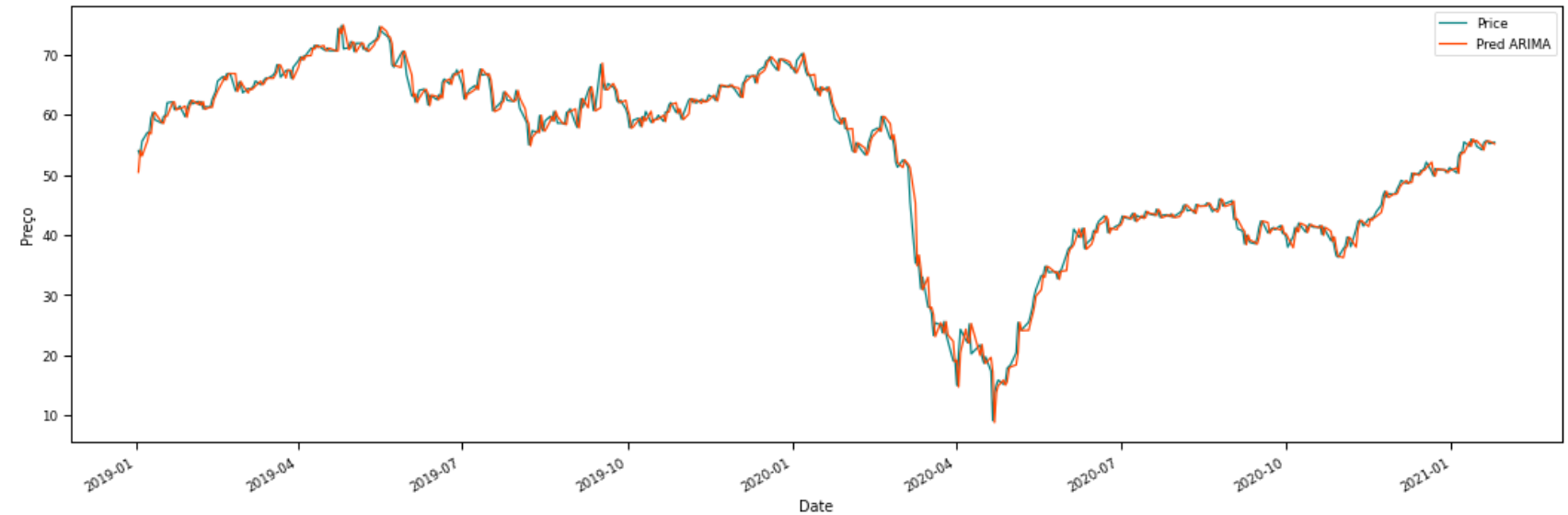
ARIMA MAE: 20.410943402837432

ARIMA MSE: 528.8348472221292

ARIMA RMSE: 22.996409441957002

ARIMA R2: 0.20020958987869608

Modelo aplicado no conjunto de teste:



ARIMA MAE: 1.040692547537601

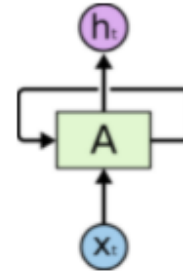
ARIMA MSE: 2.255988882790562

ARIMA RMSE: 1.5019949676315703

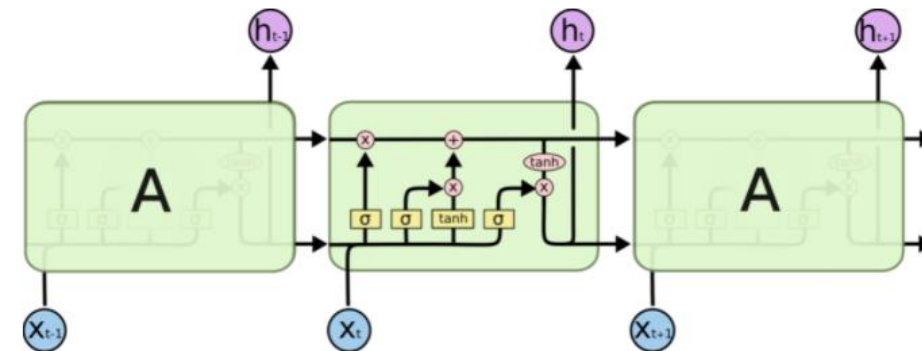
ARIMA R2: 0.9886453444300686

ARIMA
Resultado

O LSTM fará a predição do dia seguinte, com base nos dias anteriores.



Recurrent Neural Networks have loops.



Modelo Sequencial

Dropout

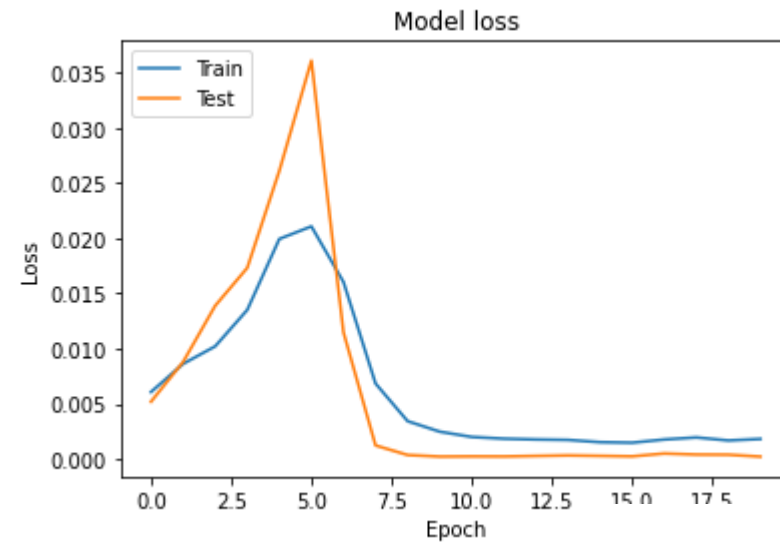
Compilador - otimizador Adam. Função de perda - erro quadrático médio

Definimos nossas épocas como 20 e o tamanho de nosso lote é 10. Também usamos a função de divisão Keras integrada para dividir nossos dados em um treinamento de 70% e um teste de 30%.

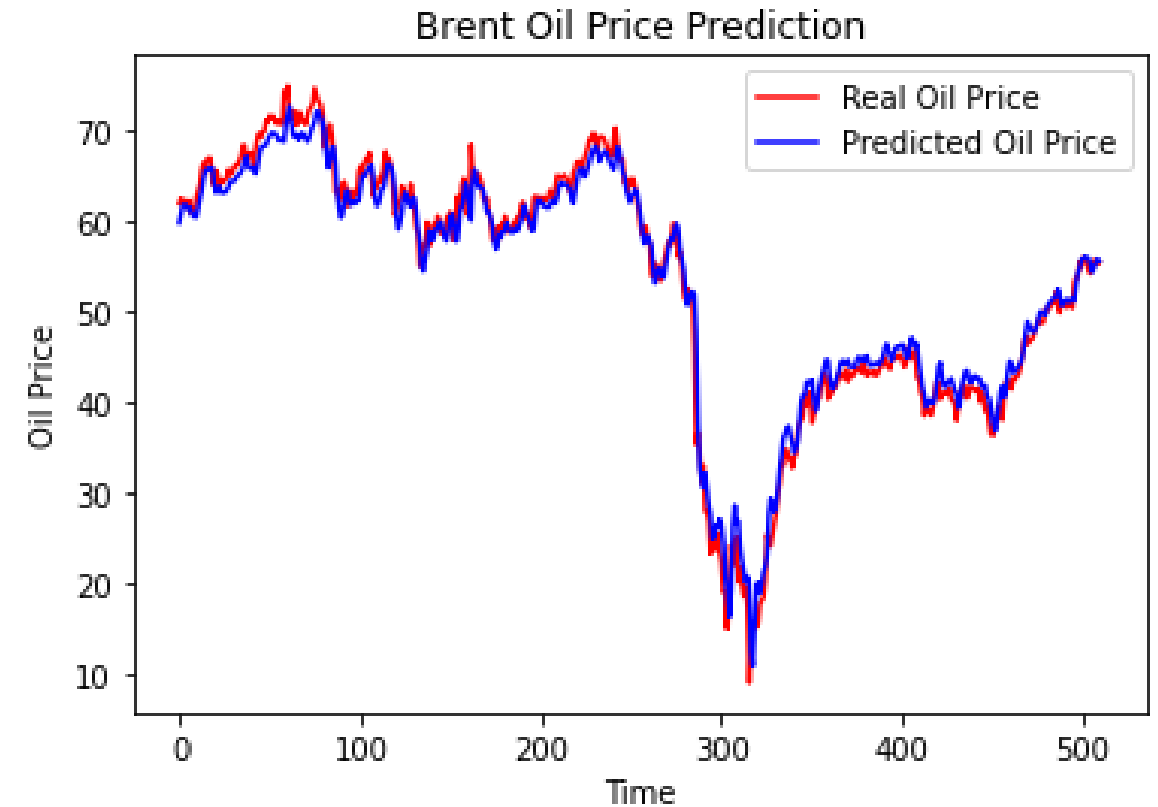
 **Keras**

LSTM

LSTM

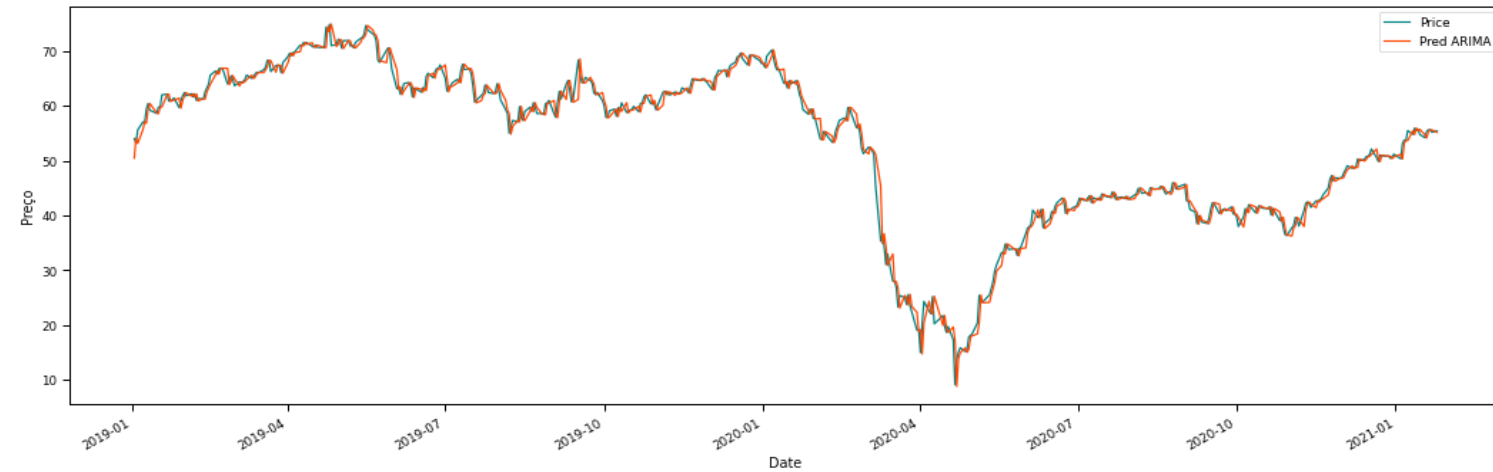


MSE: 4.102318507450047
RMSE: 2.025418106823884
R2: 0.9799737325741295



Resultados

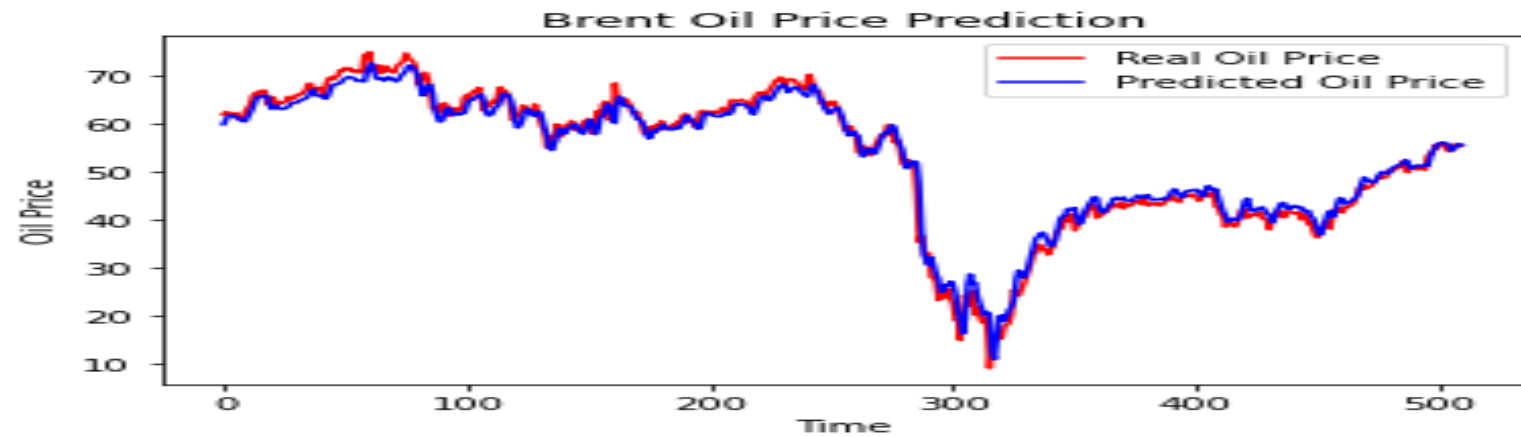
ARIMA



RMSE: 1.5019949676315703

R2: 0.9886453444300686

LSTM



RMSE: 2.025418106823884

R2: 0.9799737325741295

Conclusão



Conclusão



MUITO OBRIGADA!