

FIAP – PÓS TECH - DATA ANALYTICS

FASE 2 – TECH CHALLENGE

TURMA 5DTAT – GRUPO 44

INTEGRANTES

Gabriel Silva Ferreira

Gustavo Duran Domingues

Jhonny Amorim Silva

Lucas Alexander dos Santos

Sandro Semmer

Análise Preditiva da Bolsa de Valores – IBOVESPA

Objetivo

Realizar um modelo preditivo com dados da IBOVESPA a fim de criar uma série temporal e prever diariamente o fechamento da base.

Introdução

O Ibovespa é o principal índice da Bolsa brasileira, criado em 1968, sendo atualizado a cada quatro meses. O índice é calculado com base no desempenho das ações das maiores empresas de capital aberto da Bolsa, como Vale, Petrobras ON e PN, Itaú, Bradesco PN, Banco do Brasil, B3 e Eletrobras ON.

Séries Temporais

Séries temporais tratam de observações de uma variável realizadas em diferentes instantes de tempo (seja diária, mensal ou trimestralmente) e podem ser utilizadas tanto para identificar padrões de comportamento quanto para realizar previsões a partir de dados históricos.

Forecasting

Forecasting (ou previsão) em séries temporais é uma técnica comum no campo de Data Science, capaz de auxiliar as organizações a realizarem previsão de demanda, detecção de anomalias e definição de metas.

Diante dessa complexidade, a utilização de frameworks pode auxiliar na realização de previsões eficientes, mesmo sem possuir profundo conhecimento sobre a série.

Modelo Prophet

É um desses meios de utilização para previsão é o prophet, um framework open-source do Facebook e uma das ferramentas disponíveis capaz de realizar previsões de maneira eficiente e automatizada.

O Prophet é eficiente para lidar com séries temporais que possuam as seguintes características:

- Observações horárias, diárias ou semanais com pelo menos alguns meses (preferencialmente um ano) de histórico;
- Sazonalidades, em escala humana, fortes e destacadas (dia da semana ou período do ano);

- Feriados ou datas importantes previamente conhecidos; e
- Tendência de crescimento não linear que se aproximam de um limite.

Ele utiliza o modelo de séries temporais decomposto com três componentes principais: tendência (g), sazonalidade (s) e feriados (h), combinados na seguinte equação:

$$y(t)=g(t)+s(t)+h(t)+\epsilon t$$

Sendo uma ferramenta simples que possibilita fazer previsões em séries temporais — desde que dentro das limitações do framework e sem que haja a necessidade de possuir conhecimentos profundos.

Quando escolher o modelo Prophet?

Escolha o Prophet em situações de séries temporais com múltiplas sazonalidades, especialmente aquelas de "escala humana"(diárias, semanais, anuais).

É uma opção eficaz quando há feriados ou eventos especiais que afetam os padrões dos dados. O Prophet é particularmente útil para previsões em ambientes de negócios, oferecendo resultados precisos com configurações intuitivas e facilmente interpretáveis.

Uma das principais vantagens do Prophet sobre outros modelos é sua interpretabilidade e a possibilidade de alterar parâmetros. Isso permite ajuste de flexibilidade, especificação de pontos de quebra, adição dos feriados e eventos especiais de maneira simples. A proposta do Prophet é justamente gerar um modelo com parâmetros interpretáveis que podem ser ajustados intuitivamente por quem está analisando os dados.

Modelo ARIMA

O modelo ARIMA (AutoRegressive Integrated Moving Average) é um modelo estatístico utilizado para fazer previsões em séries temporais. Ele combina as técnicas dos modelos autorregressivos (AR) e de média móvel (MA), incluindo também a diferenciação da série temporal para torná-la estacionária. A estacionariedade é importante para que o modelo funcione adequadamente. O modelo ARIMA é amplamente utilizado para prever valores futuros com base em padrões e tendências observados nos dados históricos.

A estacionariedade é uma propriedade importante para esses modelos, significando que as características estatísticas (i.e., média, variância, etc.) de uma série temporal não dependem do tempo que elas foram observadas. Sendo assim, séries que apresentam tendência e sazonalidade, por exemplo, não são estacionárias.

Partindo de uma série estacionária, esses modelos são úteis para quando as observações passadas da série temporal, assim como choques aleatórios defasados, possuem uma influência ou são preditoras dos valores futuros da série.

Biblioteca statsforecast

Statsforecast é um biblioteca desenvolvida pela Nixtla e inclui implementações eficientes de vários algoritmos de previsão, como ARIMA, Theta, Exponential Smoothing, e outros métodos estatísticos. Reunindo modelos estatísticos de previsão de séries temporais mais populares numa interface simples, nos permitindo fazer previsões probabilísticas e de ponto.

Além disso, você não fica limitado a prever uma série de cada vez, ela é feita para ser escalável para prever múltiplas séries ao mesmo tempo.

Modelo Naive

Naive Bayes é um algoritmo de classificação que se baseia no Teorema de Bayes. Ele é considerado "ingênuo" porque assume independência entre os recursos (variáveis independentes) do conjunto de dados. O algoritmo utiliza probabilidades para calcular a probabilidade de um dado pertencer a uma determinada classe.

Seasonal Naive ou Naive Sazonal

É uma extensão do método Naive que faz previsões com base no último valor observado. No caso do SeasonalNaive, além de considerar o último valor observado, ele leva em conta o padrão sazonal da série. Isso significa que ele projeta o próximo valor da série com base no último valor observado no mesmo período de tempo no ciclo sazonal anterior.

Esse método é útil para séries temporais que apresentam um padrão sazonal bem definido, como vendas mensais ou dados climáticos sazonais.

SeasonalWindowAverage ou Média Móvel Sazonal

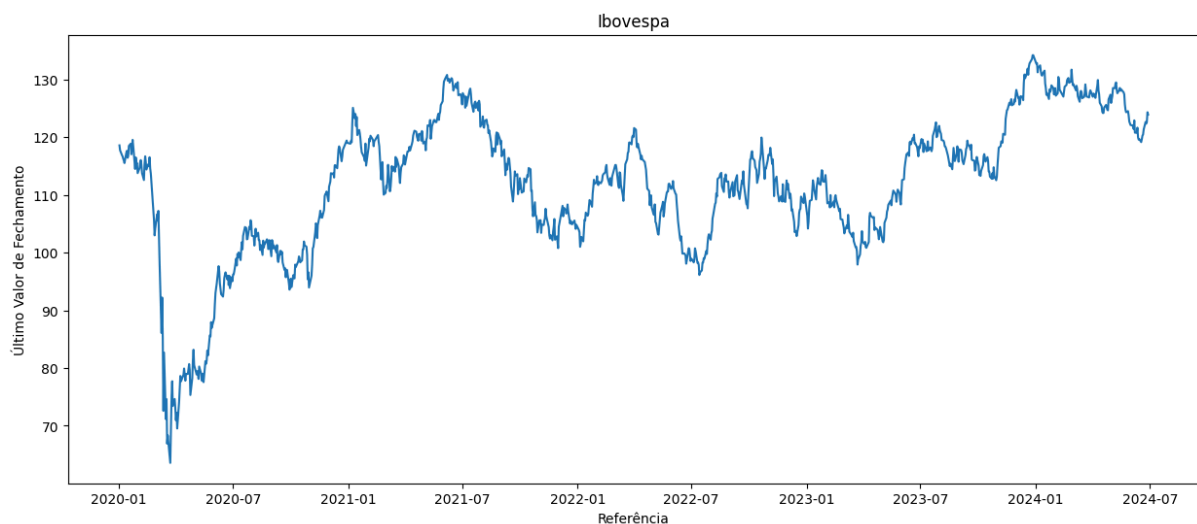
Basicamente, o SeasonalWindowAverage consiste em calcular a média dos valores observados em uma janela de tempo, que corresponde ao mesmo período sazonal no ciclo anterior. Por exemplo, se a série temporal tiver um padrão sazonal mensal, a janela de tempo pode ser de 12 meses, calculando a média dos valores observados nos mesmos meses do ano anterior.

Esse método é útil para identificar tendências e padrões sazonais em séries temporais.

Prevendo os dados

Levando em conta os modelos e meios de predição apresentados, vamos utilizar os dados históricos do site da Investing [Ibovespa \(IBOV\) Histórico de Cotações - Investing.com](https://pt.investing.com/indices/ibovespa) no período de Janeiro/2020 a Junho/2024 para prever diariamente o fechamento da base.

Primeiramente fizemos um simples gráfico de linhas para entender a variação dos valores de fechamento e notamos que o valor de mercado das ações sofrem muita variação ao longo dos dias e meses, causando maiores dúvidas e incertezas para um possível investidor.



Utilizada a base de dados da Investing, no período de Janeiro de 2020 a Junho de 2024, de acordo com o valor diário do histórico de fechamento.

Escolhendo o melhor modelo

Em seguida, precisamos escolher o melhor modelo que tenha os melhores resultados. Para isso, utilizamos algumas métricas responsáveis por medir e analisar erros de modelos, as quais são:

Erro Médio Absoluto - MAE: o erro médio absoluto, MAE (da sigla em inglês Mean Absolute Error), é calculado a partir da média dos erros absolutos, ou seja, utilizamos o módulo de cada erro para evitar a subestimação, isso porque, o valor é menos afetado por pontos especialmente extremos (outliers). Utiliza-se essa medida em séries temporais, pois há casos em que o erro negativo pode zerar o positivo ou dar uma ideia de que o modelo é preciso.

Erro Quadrático Médio - MSE: o erro quadrático médio, MSE (da sigla em inglês Mean Squared Error), é comumente usado para verificar a acurácia de modelos e dá um maior peso aos maiores erros, já que, ao ser calculado, cada erro é elevado ao quadrado individualmente e, após isso, a média desses erros quadráticos é calculada.

Média Percentual Absoluta do Erro - MAPE: esta é outra métrica interessante para usar, geralmente usada em relatórios de gerenciamento, porque o erro é medido como uma porcentagem e assim, é possível fazer comparações entre erros percentuais do modelo entre produtos.

Considerando nossa divisão dos períodos de treino e testes para nossa base (80% para treino e 20% para teste) e aplicando os modelos anteriormente citados, tivemos os seguintes resultados:

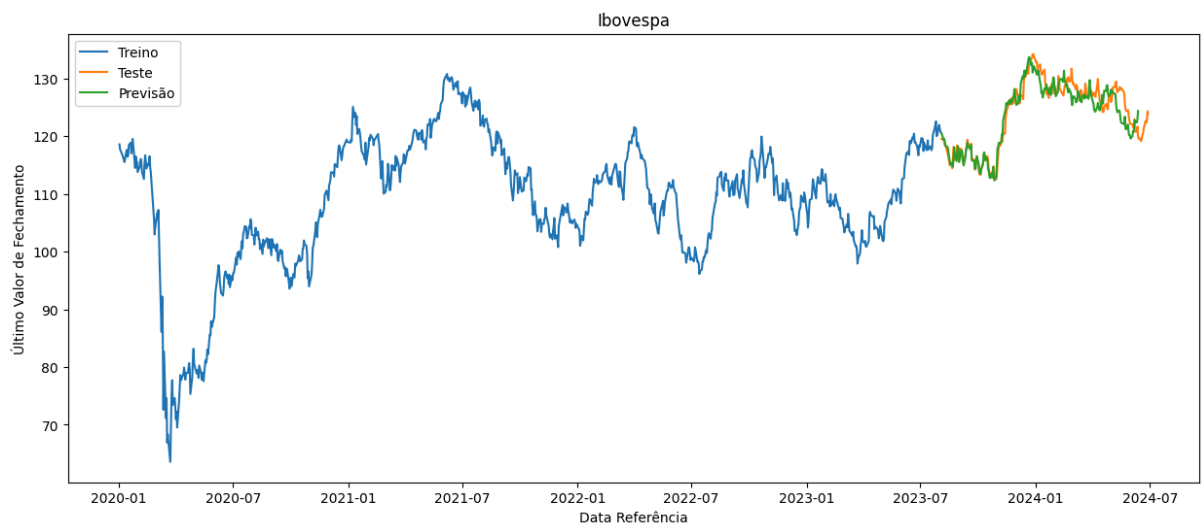
Tabela - Métricas para avaliação de Séries Temporais para os modelos de previsão

model name	MAE	MSE	MAPE
SeasonalWindowAverage	6.438857333319527	55.45746671046026	5.096280173520567
Arima	5.864501866030833	43.33667644495718	4.6770491707404025
Naive	5.849036185128348	43.03125043390979	4.666037808616484
SeasonalNaive	5.706665069852556	40.7560717884223	4.5674147788656985
Prophet	0.8930310914256899	1.242133683723621	0.7228539260827419

Fonte: Dados Históricos da Investing referente a IBOVESPA. Jan.2020 - Jun., 2024.

Conclusão

Com isso concluímos que o modelo **Prophet** foi o que deu os melhores resultados, com menores medições de erro e com percentual maior de acertos. Também se adaptando muito bem a nossa base, com uma predição muito próxima dos valores reais que analisamos, conforme gráfico feito para melhor visualização abaixo:



Análise da predição feita pelo modelo Prophet.

Referências:

ALENCAR, Valquíria. **Time Series Forecasting: Prophet é mesmo um Profeta?** Disponível em: <https://valquiria-c-alencar.medium.com/time-series-forecasting-prophet-é-mesmo-um-profeta-d3356f7b943f>. Acesso em: 25 jul. 2024.

BÔSCOA, Vinícius. **Prophet — Prevendo o futuro em Séries Temporais.** Disponível em: <https://www.viniboscoa.dev/blog/prophet-prevendo-o-futuro-em-series-temporais>. Acesso em: 25 jul. 2024.

BORGES, Mateus. **Análise e Previsão de Séries Temporais via Facebook Prophet.** Disponível em: <https://repositorio.ufscar.br/bitstream/handle/ufscar/19516/TGMPB.pdf?sequence=1&isAllowed=y>. Acesso em: 25 jul. 2024.

CARVALHO, João. **O algoritmo Naive Bayes — descrição e implementação em Python.** Disponível em: <https://joaoclaudionc.medium.com/o-algoritmo-naive-bayes-descrição-e-implementação-em-python-35757ade6b36>. Acesso em: 25 jul. 2024.

FILHO, Mario. **Como Prever Séries Temporais Univariadas Usando Python.** Disponível em: <https://mariofilho.com/como-prever-series-temporais-univariadas-usando-python/#como-instalar-o-statsforecast>. Acesso em: 25 jul. 2024.

FILHO, Mario. **Naive Time Series Forecasting in Python.** Disponível em: <https://forecastegy.com/posts/naive-time-series-forecasting-in-python/>. Acesso em: 25 jul. 2024.

PENTEADO, Karoline. **Métricas de avaliação para séries temporais**. Disponível em: <https://www.alura.com.br/artigos/metricas-de-avaliacao-para-series-temporais>. Acesso em: 28 jul. 2024.

SILVA, Fernando da. **PREVISÃO ECONÔMICA COM MODELOS ARIMA NO PYTHON**. Disponível em: <https://analysemacro.com.br/econometria-e-machine-learning/previsao-economica-com-modelos-arima-no-python/>. Acesso em: 25 jul. 2024.