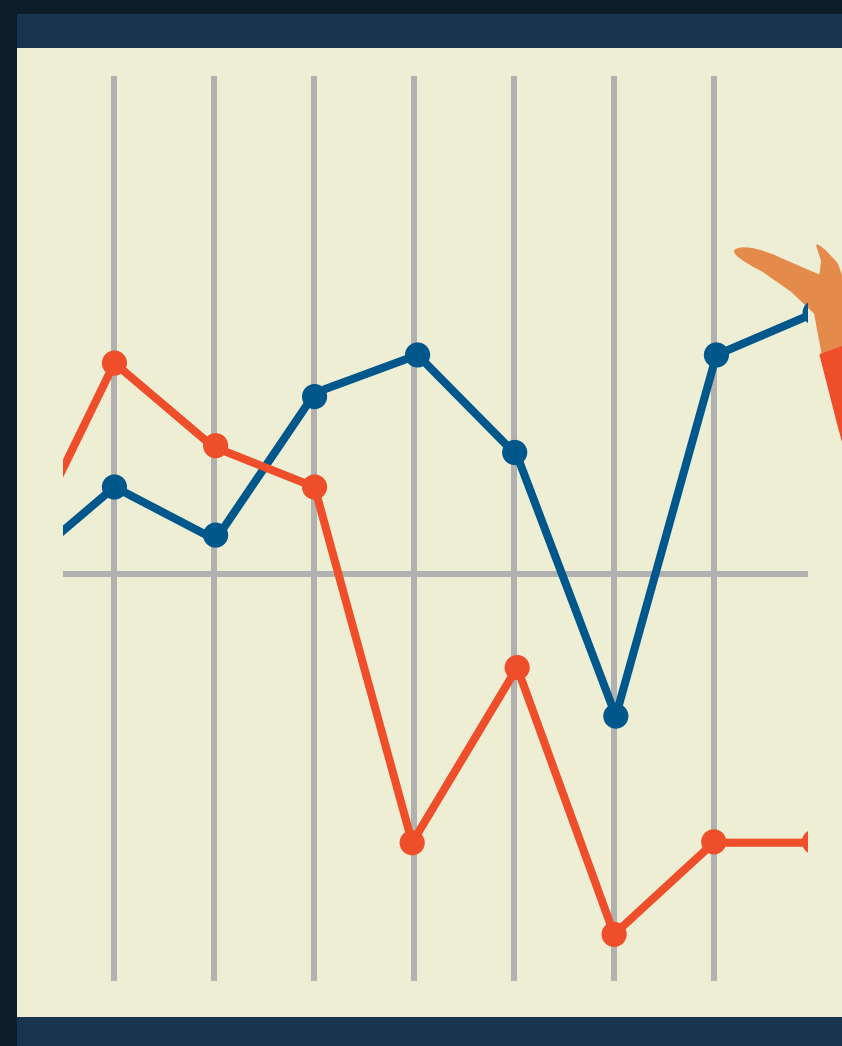


PREDIÇÃO DE SÉRIES TEMPORAIS

Conceitos básicos, preparação de dados e exemplo prático com dados reais.



REALIZAÇÃO:



INSIGHT

APRESENTAÇÃO



Paula Luana

Técnica em informática e graduanda em Ciência da Computação pela UFC - campus Quixadá. Atualmente é bolsista na Lead Dell como Dev Python IA.

AGENDA

1. Apresentação do problema
2. Conceitos básicos de séries temporais.
3. Estruturar dados.
4. Aplicação prática utilizando o ARIMA.

APRESENTAÇÃO DO PROBLEMA

COVID-19 World Vaccination Progress

Daily and Total Vaccination for COVID-19 in the World from Our World in Data

[hwww.kaggle.com/gpreda/covid-world-vaccination-progress](https://www.kaggle.com/gpreda/covid-world-vaccination-progress)

CONCEITOS BÁSICOS DE SÉRIES TEMPORAIS

Definição

Séries temporais são dados coletados em intervalos regulares de tempo. Esses dados possuem dependência com a ordem e com eles podemos observar em que data esses dados decrescem ou crescem. Usamos séries temporais comumente para entender em que dias, meses ou anos, os valores são alterados e porque são alterados.

CONCEITOS BÁSICOS DE SÉRIES TEMPORAIS

Padrões

Sazonalidade: Padrão que se repete.

Ciclos: Padrões que ocorrem eventualmente.

Tendência: Nos ajuda a visualizar e entender qual a tendência dos nossos dados, se eles crescem no decorrer do tempo ou decrescem.

Ruídos: São os dados que não se incluem nos padrões.

ESTRUTURAR DADOS

Série estacionária

Uma série é estacionária quando suas propriedades estatísticas permanecem constantes

- A média da série não deve ser uma função do tempo, mas sim uma constante.
- A variação da série não deve ser função do tempo.
- A covariância do i ésimo termo e do $(i + m)$ ésimo termo não deve ser função do tempo.

ESTRUTURAR DADOS

Teste de Dickey-Fuller

Este é um teste do tipo Hipótese de Raiz Unitária no qual ele tenta determinar o quão forte a série temporal é definida por uma tendência.

Basicamente vamos ter uma tabela que apresentará o que chamamos de valor critico, que é uma espécie de valor de corte. Em nosso exemplo são exibidos níveis de 1%, 5% e 10%. Se o resultado do ADF for menor do que o valor critico para um determinado percentual (5% no nosso caso) e o p-value for significativo (menor que 0.05), a série é considerada estacionária. Do contrário a série é não estacionária.

APLICAÇÃO PRÁTICA UTILIZANDO O ARIMA

Modelo de Média Móvel Integrada Autoregressiva

O ARIMA(AutoRegressive Integrated Moving Average model) utiliza métodos de diferenciação para tornar a série temporal em estacionária(as propriedades estatísticas são constantes), aplica modelos de autoregressão e média móvel. Ou seja, o modelo que é estacionário, sua média, variância e autocovariância são invariáveis em relação ao tempo. O método é adequado para séries temporais univariadas com tendência e sem componentes sazonais. Basicamente, o ARIMA utiliza dados passados para prever o futuro, usando dois principais recursos: a autocorrelação e médias móveis.

APLICAÇÃO PRÁTICA UTILIZANDO O ARIMA

Parâmetros do ARIMA

Na função ARIMA do python, passamos três parâmetros, sendo eles:

p: é a ordem do modelo autoregressivo; Mais especificamente é quantidade de lags(quantidade de valores correlacionados)

d: o número de vezes que as observações serão diferenciadas.

q: é a ordem do modelo de média móvel.

No R há uma função que encontra esses parâmetros para gente. Mas já em python podemos encontrar esses valores através de tentativas na mão até encontrar um melhor valor(o que é muito custoso) ou observando o gráfico de Autocorrelação e o de Autocorrelação Parcial.

MÃOS NA MASSA

LINKS

link do tutorial no github:

Acesse os dados do problema: www.kaggle.com/gpreda/covid-world-vaccination-progress