



Rafael Rodrigues Marquesi <rafael.4marcos.nt@gmail.com>

#7DaysOfCode - Ciência de Dados 3/7: Forecasting

Paulo Vasconcellos <paulo.vasconcellos@7daysofcode.io>

Sun, Apr 24, 2022 at 7:03 AM

Reply-To: reply-caelum57945.activehosted.1375.8450.1043955@d32a.emsend3.com

To: Rafael Rodrigues Marquesi <rafael.4marcos.nt@gmail.com>



Olá, Rafael Rodrigues Marquesi.

É hora de encarar o terceiro dia de desafio!

E como você já viu, a tarefa hoje é: **Forecasting**. Vou te explicar melhor o que é isso.

O trabalho de um Cientista de Dados geralmente gira em torno de duas principais frentes: **descrição** (entender o que aconteceu e por que aconteceu) e **predição** (prever o que pode acontecer e o que deve ser feito).

É provável que ontem você tenha aplicado algumas dessas técnicas descritivas, utilizando análise de dados e visualizações para entender o que aconteceu.

Neste caso, você analisou o que aconteceu com os gastos dos senadores.

Para as tarefas de predição, são utilizadas técnicas de Machine Learning e ferramentas estatísticas para prever o futuro. E é sobre elas que falaremos hoje.

Forecasting é uma das técnicas mais conhecidas e

utilizadas por Data Scientists para prever indicadores de negócios.

Por exemplo: uma área de Finanças pode querer que você preveja o faturamento da empresa; ou então seu time de Marketing pode querer prever quantos usuários irão se cadastrar na plataforma nos próximos meses.

No desafio de hoje, eu quero te provocar a criar o seu próprio modelo de Forecasting!

Será que você consegue criar um modelo que irá prever quanto os senadores vão gastar nos próximos três meses?

Você ainda utilizará o dataset do CEAPS que utilizou ontem, mas será muito importante para o seu modelo que você tenha mais de um ano de dados.

Os dados do CEAPS são divididos por ano, então seria interessante pegar pelo menos quatro anos e juntar em um grande dataset (isso fará seu modelo mais preciso).

Mas, se você não tem tempo pra isso, não tem problema. **Eu já fiz um dataset para esse desafio, que você pode baixar aqui.** Esse dataset contém dados de 2018 a 2021 do CEAPS, e nele você encontrará duas colunas:

- **DS**, que representa a data do reembolso;
- **Y**, que é a soma de reembolsos dos senadores naquele dia.

DICA

Tente começar suas previsões de forma simples.

Uma dica de modelo inicial (também chamado de *baseline*) é utilizar médias.

Por exemplo, para os próximos três meses, você pode dizer que o valor de "y" será a média dos últimos três meses.

Isso vai te dar uma base de referência para quando você criar modelos mais complexos e quiser saber se são melhores que uma abordagem simples.

Existem muitas técnicas que podem ser utilizadas em problemas de Forecasting.

Uma solução muito famosa é o Facebook Prophet: uma biblioteca de previsão criada pelo Meta, disponível pra Python e R.

Ela é muito poderosa e simples de usar, e pode ser uma boa escolha para esse problema. Você também pode tentar usar algoritmos de Machine Learning com Regressão Linear do Scikit-learn, ou técnicas como ARIMA. Experimente várias abordagens e escolha a melhor, mas **tente sempre começar da mais simples para a mais complexa.**

Para avaliar seu modelo, você vai querer utilizar alguma métrica. Para problemas de Forecasting, **você pode utilizar métricas de problemas de regressão, como RMSE e MAPE.**

EXTRA

Time Series Forecasting é um tema que me interessa muito, e que já fiz muitos conteúdos sobre o assunto.

Se você quiser saber o que são séries temporais, leia esse

artigo que escrevi sobre.

Eu também já escrevi sobre como você pode criar suas primeiras previsões utilizando o famoso ARIMA (Autoregressive Integrated Moving Average) e outras técnicas. Caso queira saber mais sobre o tema e aprender como você pode aplicá-los nesse desafio, leia o artigo aqui.

E, como mais uma fonte de conteúdo, recomendo assistir essa palestra que dei falando sobre o tema de forma bem introdutória também.

Bom trabalho!

Paulo Vasconcellos

Cientista de Dados e Alura Star

alura

Enviado para: rafael.4marcos.nt@gmail.com

[Não quero mais receber os Desafios](#)

Alura, Rua Vergueiro 3185, São Paulo - SP, 04101-300, Brasil