

Bootcamp IGTI: Analista de Machine Learning

Trabalho Prático

Módulo 2	Modelos Preditivos e Séries Temporais
-----------------	--

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Realizar a análise de uma base de dados real em Python.

Enunciado

O uso de bicicletas como um meio de transporte ganhou muita força nos últimos anos, seja por questões ambientais, de saúde ou até mesmo infraestrutura de trânsito. Para incentivar o uso, cidades em todo o mundo têm implementado programas de compartilhamento de bicicleta. Nesses sistemas, as bicicletas são retiradas e devolvidas em quiosques automatizados espalhados em diversos pontos da cidade.

As plataformas de compartilhamento de bicicletas costumam coletar diversos tipos de dados, entre eles: duração da viagem, localização inicial e final do percurso, entre outros. Esses dados coletados pela plataforma, em conjunto com informações sobre o clima, trânsito e relevo, por exemplo, tornam possível uma análise mais robusta de compartilhamento de bicicletas.

A seguir, um descritivo dos dados coletados:

- **rec_id**: índice do registro de locação;
- **datetime**: data;
- **season**: estação do ano (1: inverno, 2: primavera, 3: verão, 4: outono). Relativo ao hemisfério norte;

- **year**: ano (0: 2011, 1:2012);
- **month**: mês (1 a 12);
- **hour**: hora do dia (0 a 23);
- **is_holiday**: booleano indicando feriado;
- **weekday**: dia da semana (0: domingo, 1: segunda-feira, ..., 6: sábado);
- **is_workingday**: booleano indicando dia útil;
- **weather_condition**: (1: limpo, 2: nublado, 3: chuva leve, 4: chuva forte);
- **temp**: temperatura escalada entre 0 e 1. Valor original em graus Celsius: -8 a 39;
- **atemp**: sensação térmica escalada entre 0 e 1. Valor original em graus Celsius: -16 a 50;
- **humidity**: humidade relativa (0 a 1);
- **windspeed**: velocidade do vento escalada entre 0 e 1 (máximo original: 67);
- **casual**: número de locações para usuários casuais;
- **registered**: número de locações para usuários registrados;
- **total_count**: contador total de alugueis (casual+registered).

Esta atividade tem como objetivo analisar os dados coletados de compartilhamento de bicicletas em uma cidade, coletados pela Universidade do Porto.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Acessar o ambiente [Google Colaboratory](#) (recomendado) ou qualquer ambiente de desenvolvimento em Python.
2. Importar as bibliotecas do numpy e pandas conforme:

```
# importando as bibliotecas
import numpy as np
import pandas as pd
```

3. Carregar o dataset <https://pycourse.s3.amazonaws.com/bike-sharing.csv> para análise utilizando o pandas. Exemplo:

```
df = pd.read_csv("https://pycourse.s3.amazonaws.com/bike-sharing.csv")
```

4. Responda às seguintes perguntas:
- Qual o tamanho desse dataset?
 - Qual a média da coluna **windspeed**?
 - Qual a média da coluna **temp**?
 - Quantos registros existem para o ano de 2011?
 - Quantos registros existem para o ano de 2012?
 - Quantas locações de bicicletas foram efetuadas em 2011?
 - Quantas locações de bicicletas foram efetuadas em 2012?
 - Qual estação do ano contém a maior média de locações de bicicletas?
 - Qual estação do ano contém a menor média de locações de bicicletas?
 - Qual horário do dia contém a maior média de locações de bicicletas?
 - Qual horário do dia contém a menor média de locações de bicicletas?
 - Que dia da semana contém a maior média de locações de bicicletas?
 - Que dia da semana contém a menor média de locações de bicicletas?
 - Às quartas-feiras (`weekday = 3`), qual o horário do dia contém a maior média de locações de bicicletas?
 - Aos sábados (`weekday = 6`), qual o horário do dia contém a maior média de locações de bicicletas?

Respostas Finais

Os alunos deverão desenvolver a prática e, depois, responder às seguintes questões objetivas: