




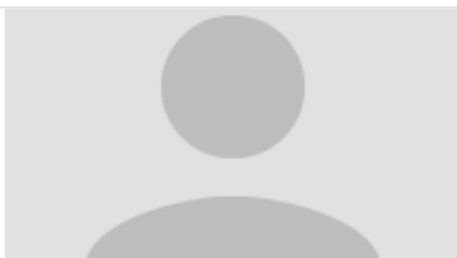
Desafio do Primeiro Módulo

Local no google drive

Meet Google Drive - One place for all your files

Google Drive is a free way to keep your files backed up and easy to reach from any phone, tablet, or computer. Start with 15GB of Google storage - free.

 <https://drive.google.com/drive/folders/1bNP!guE2kAxFG4JLlnMsXrHEjTYyfKti?ths=true>



Questões

Pergunta 1 2,66 pts

No dataset utilizado para o desafio, quantas instâncias e atributos existem, respectivamente?

Grupo de escolhas da pergunta (17379,17).(17,17379).(15,1345).(1345,15).

Pergunta 2 2,66 pts

Quantos tipos diferentes de dados existem no dataset do desafio?

Grupo de escolhas da pergunta 17.12.3.2.

Pergunta 3

Qual é a proporção (em %) de valores nulos existente na coluna "temp" (temperatura ambiente normalizada)?

15%.

0%.

5%.

10%.

Pergunta 4 2,66 pts

Após retirar as linhas que contém valores nulos para a coluna "dteday", passamos a contar com quantas instancias e atributos, respectivamente?

* (17,17379).

* (17379,17).

*** (15641,17).**

* (17,15641).

Pergunta 5 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Qual é o valor médio para os dados da coluna "temp" (temperatura ambiente normalizada)?

Grupo de escolhas da pergunta * 0,156. * 1,192. *** 0,496.** * 0,192.

Pergunta 6 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Qual é o desvio padrão para os dados da coluna "windspeed" (velocidade do vento normalizada)?

Grupo de escolhas da pergunta 0,189. 1,122. 0,192. **0,122.**

Pergunta 7 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Transforme a coluna "season" em valores categóricos.

Quantas categorias diferentes existem?

Grupo de escolhas da pergunta **4**. 2. 1. 3.

Pergunta 8 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Transforme a coluna "dteday" no tipo "datetime". Qual é a última data presente no dataset (YYYY-MM-DD)?

Grupo de escolhas da pergunta 2014-12-31. **2012-12-31**. 2012-01-31.
2012-01-31.

Pergunta 9 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Considerando o boxplot da variável "windspeed" (velocidade do vento), é CORRETO afirmar:

- XX Não existem possíveis outliers, pois existem marcações (pontos) foras dos limites do boxplot.
- **Existem possíveis outliers, pois existem marcações (pontos) foras dos limites do boxplot.**
- Não existem possíveis outliers, pois não existem marcações (pontos) foras dos limites do boxplot.
- XX Existem possíveis outliers, pois não existem marcações (pontos) foras dos limites do boxplot.

Pergunta 10 2,66 pts

Considere o dataset após a retirada das linhas que continham valores nulos para a coluna "dteday". Selecione as colunas "season", "temp", "atemp", "hum",

"windspeed". Plot a matriz de correlação. Sobre as variáveis "hum" e "cnt", é CORRETO afirmar:

Grupo de escolhas da pergunta

Possuem baixa correlação linear negativa.

XX Possuem alta correlação linear positiva.

Possuem alta correlação linear negativa.

XX Possuem baixa correlação linear positiva.

Pergunta 11 2,66 pts

Preencha os valores nulos das colunas "hum","cnt" e "casual" com os valores médios. Utilize as variáveis "hum" e "casual" como independentes e a "cnt" como dependente. Aplique uma regressão linear. Qual o valor de R²? Utilize as entradas como teste.

Grupo de escolhas da pergunta

-0.80.

-0.40.

+0.40.

+0.80.

Pergunta 12 2,66 pts

Utilize os mesmos dados da questão anterior ("hum" e "casual" como variáveis independentes e "cnt" como variável dependente). Aplique Árvore de Decisão como regressão. Qual é o valor aproximado de R²? Utilize as entradas como teste e valores "default".

0,30.

1,00.

0,40.

0,70.

Pergunta 13 **2,66 pts**

Comparando os valores de R^2 encontrado com a **regressão linear** e com a **Árvore de Decisão**, é CORRETO afirmar:

- O valor obtido pela Árvore de Decisão como regressor apresenta maior R^2 .
- O valor encontrado pela regressão linear é superior ao encontrado pela Árvore de Decisão.
- O valor encontrado pela Árvore de Decisão foi negativo.
- Os valores encontrados pela Árvore de Decisão como regressor e a regressão linear são próximos.

Pergunta 14 **2,66 pts**

Comparando o **SVM** com a **Árvore de Decisão** é CORRETO afirmar:

- Árvore de Decisão encontra os hiperplanos de separação do mesmo modo que o SVM.
- **SVM encontra o hiperplano que gera a maior separação entre os dados.**
- SVM e Árvore de Decisão só podem ser utilizados para modelos de classificação.
- Através da Árvore de Decisão não é possível encontrar os limiares de separação entre as classes.

Pergunta 15 **2,76 pts**

Com base na **Árvore de Decisão** é CORRETO afirmar:

- Não é possível encontrar os limites de separação entre os conjuntos de dados a partir de uma classificação realizada através da Árvore de Decisão.
- Árvore de Decisão é utilizada apenas para a classificação de dados contínuos.

- A análise de regressão com a Árvore de Decisão só pode ser realizada com valores categóricos.
- **Pode ser utilizada para classificação e regressão.**