

Desafio Final

Entrega 8 out em 23:59	Pontos 100	Perguntas 15
Disponível até 8 out em 23:59	Limite de tempo Nenhum	

Instruções

O Desafio Final está disponível!

1. Instruções para realizar o desafio

Consulte a data de entrega no teste e em seu calendário.

Reserve um tempo para realizar a atividade, leia as orientações e enunciados com atenção. Em caso de dúvidas utilize o "Fórum de dúvidas do Desafio Final".

Para iniciá-lo clique em "Fazer teste". Você tem somente uma tentativa e não há limite de tempo definido para realizá-lo. Caso precise interromper a atividade, apenas deixe a página e, ao retornar, clique em "Retomar teste".

Clique em "Enviar teste" **somente** quando você conclui-lo. Antes de enviar confira todas as questões.

Caso o teste seja iniciado e não enviado até o final do prazo de entrega, a plataforma enviará a tentativa não finalizada automaticamente, independente do progresso no teste. Fique atento ao seu teste e ao prazo final, pois novas tentativas só serão concedidas em casos de questões médicas.

O gabarito será disponibilizado partir de sexta-feira, 08/10/2021 às 23h59.

Bons estudos!

2. O arquivo abaixo contém o enunciado do desafio

[Enunciado do Desafio – Bootcamp Engenheiro\(a\) de Machine Learning.pdf](#)

Detalhes do envio:

Tempo:	284 minutos
Pontuação atual:	100 de 100
Pontuação mantida:	100 de 100

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	Tentativa 1	284 minutos	100 de 100

ⓘ As respostas corretas estarão disponíveis em 8 out em 23:59.

Pontuação deste teste: 100 de 100

Enviado 7 out em 23:20

Esta tentativa levou 284 minutos.

Pergunta 16,7 / 6,7 pts

Após a utilização da biblioteca pandas para a leitura dos dados sobre os valores lidos, é CORRETO afirmar que:

- ☐ Existem apenas duas características do tipo "string".
- ☐ Os valores nulos encontrados estão presentes apenas no variável "cubicinches".
- ☒ Não foram encontrados valores nulos após a leitura dos dados.
- ☐ Os valores nulos estão presentes apenas para a saída dos dados.

Pergunta 26,7 / 6,7 pts

Realize a transformação das colunas "cubicinches" e "weightlbs" do tipo "string" para o tipo numérico utilizando o pd.to_numeric(), utilizando o parâmetro errors='coerce'. Após essa transformação, é CORRETO afirmar:

- ☒ Essa transformação adiciona valores nulos ao nosso dataset.
- ☐ Após a transformação não ocorre mudanças no tipo de variáveis.
- ☐ Após essa transformação não ocorreu a introdução de valores nulos.
- ☐ Essa transformação deve ser realizada apenas para valores nulos.

Pergunta 36,7 / 6,7 pts

Indique quais eram os índices dos valores presentes no dataset que "forçaram" o pandas a compreender a variável "cubicinches" como string.

- ☐ 38 e 120.
- ☐ 114 e 180.
- ☒ 40 e 180.
- ☐ 33 e 172.

Pergunta 46,7 / 6,7 pts

Após a transformação das variáveis "string" para os valores numéricos, quantos valores nulos (células no dataframe) passaram a existir no dataset?

- ☐ 3.
- ☐ 2.
- ☐ 6.
- ☒ 5.

Pergunta 56,7 / 6,7 pts

Substitua os valores nulos introduzidos no dataset após a transformação pelo valor médio das colunas. Qual é o novo valor médio da coluna "weightlbs"?

- ☒ 3009,83.
- ☐ 1367,13.
- ☐ 3212,48.
- ☐ 1370,82.

Pergunta 66,7 / 6,7 pts

Após substituir os valores nulos pela média das colunas, selecione as colunas ['mpg', 'cylinders', 'cubicinches', 'hp', 'weightlbs', 'time-to-60', 'year']. Qual é o valor da mediana para a característica 'mpg'?

- ☒ 22,0.
- ☐ 31,4.
- ☐ 26,3.
- ☐ 14,8.

Pergunta 76,7 / 6,7 pts

Qual é a afirmação CORRETA sobre o valor de 14,00 para a variável "time-to-60"?

- ☐ Apenas por esse valor é possível afirmar que existe uma menor dispersão entre valores inferiores a 14,00.
- ☒ 75% dos dados são maiores que o valor de 14,00.
- ☐ Como esse valor representa a mediana para essa variável, é possível dizer que 50% dos valores são superiores a 14,00.
- ☐ 25% dos valores de "time-to-60" são superiores ou iguais a 14,00.

Pergunta 86,7 / 6,7 pts

Sobre o coeficiente de correlação de Pearson entre as variáveis "cylinders" e "mpg", é correto afirmar, EXCETO:

- ☐ Caso fosse calculado o coeficiente de determinação entre essas duas variáveis, o valor seria, aproximadamente, 0,6.
- ☐ Mesmo não sendo igual a 1, é possível dizer que à medida que a variável "cylinders" aumenta, a variável "mpg" reduz em uma direção oposta.
- ☐ Quando um coeficiente de correlação de Pearson é igual a 1, o coeficiente de determinação também será igual a 1.
- ☒ Mesmo não sendo igual a 1, é possível dizer que à medida em que a variável "cylinders" aumenta, a variável "mpg" também aumenta na mesma direção.

Pergunta 96,7 / 6,7 pts

Sobre o boxplot da variável "hp", é correto afirmar, EXCETO:

- ☒ Existe uma maior dispersão no segundo quartil quando comparamos com o terceiro.
- ☐ Cada um dos quartis possui a mesma quantidade de valores para a variável "hp".
- ☐ Não foi identificada a presença de possíveis outliers nos dados.
- ☐ Através do boxplot, é possível perceber que a mediana encontra-se entre os valores de 80 e 100.

Pergunta 106,7 / 6,7 pts

Após normalizado, utilizando a função StandardScaler(), qual é o maior valor para a variável "hp"?

- ☐ 1,23
- ☐ 1,18
- ☐ 9,01
- ☒ 3,06

Pergunta 116,7 / 6,7 pts

Aplicando o PCA, conforme a definição acima, qual é o valor da variância explicada com pela primeira componente principal?

- ☐ 13%
- ☒ 72%
- ☐ 34%
- ☐ 92%

Pergunta 126,7 / 6,7 pts

Utilize os três primeiros componentes principais para construir o K-means com um número de 3 clusters. Sobre os clusters, é INCORRETO afirmar que:

- ☐ Cada um dos clusters possui características próprias.
- ☐ Os centroides, utilizando apenas as 3 componentes principais, possuem 3 dimensões.
- ☒ Todos os clusters possuem a mesma quantidade de elementos.
- ☐ Existem 3 centroides após a aplicação da clusterização.

Pergunta 136,7 / 6,7 pts

Após todo o processamento realizado nos itens anteriores, crie uma coluna que contenha a variável de eficiência do veículo. Veículos que percorrem mais de 25 milhas com um galão ("mpg">25) devem ser considerados eficientes. Utilize as colunas ['cylinders', 'cubicinches', 'hp', 'weightlbs', 'time-to-60'] como entradas e como saída a coluna de eficiência criada.

Utilizando a árvore de decisão como mostrado, qual é a acurácia do modelo?

- ☐ Aproximadamente, 0,70.
- ☒ Aproximadamente, 0,87.
- ☐ Aproximadamente, 0,99.
- ☐ Aproximadamente, 0,63.

Pergunta 146,7 / 6,7 pts

Sobre a matriz de confusão obtida após a aplicação da árvore de decisão, como mostrado anteriormente, é INCORRETO afirmar:

- ☐ Os falso-positivos correspondem a instâncias em que o algoritmo considerou a previsão como verdadeira e, na realidade, ela era falsa.
- ☐ A diagonal principal da matriz mostra as instâncias em que as previsões foram corretas.
- ☐ A matriz de confusão se constitui em uma estratégia ainda mais importante quando um dataset não está balanceado.
- ☒ Existem duas vezes mais veículos considerados não eficientes que instâncias de veículos eficientes.

Pergunta 156,2 / 6,2 pts

Utilizando a mesma divisão de dados entre treinamento e teste empregada para a análise anterior, aplique o modelo de regressão logística como mostrado na descrição do trabalho.

Comparando os resultados obtidos com o modelo de árvore de decisão, é INCORRETO afirmar que:

- ☐ Como os dois modelos obtiveram um resultado superior a 80% de acurácia, a escolha sobre qual utilizar deve e pode ser feita a partir de outros critérios, como a complexidade do modelo.
- ☐ A acurácia de ambos os modelos foi superior a 80%.
- ☒ A regressão logística não deveria ser aplicada ao problema, pois ela trabalha apenas com dados categóricos.
- ☐ A árvore de decisão e a regressão logística podem ser utilizadas para previsão em regressões.

Pontuação do teste: 100 de 100

◀ Anterior

Próximo ▶

