DESF5 - Desafio Final

Entrega 23 jun em 23:59 **Pontos** 100 **Perguntas** 15
Disponível até 23 jun em 23:59 **Limite de tempo** Nenhum

Instruções

desafio do módulo..png

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 concluí-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 a partir de sexta-feira, **23/06/2023** às 23h59.

- **O arquivo abaixo contém o enunciado do Desafio. Confira agora:**

Enunciado do Desafio Final – Bootcamp Cientista de Dados.pdf (https://online.igti.com.br/users/144651/files/410053?wrap=1&verifier=N4xt1d20AucCwoLY0Tuto9TMr8PInpDGmczz7cf1) (https://online.igti.com.br/users/144651/files/410053/download?verifier=N4xt1d20AucCwoLY0Tuto9TMr8PInpDGmczz7cf1&download_frd=1)

Bons estudos!

Atenciosamente,

Equipe XP Educação

Histórico de tentativas

	Tentativa	Tempo	Pontuação
MAIS RECENTE	**Tentativa 1**	344 minutos	100 de 100

ⓘ As respostas corretas estarão disponíveis em 23 jun em 23:59.

Pontuação deste teste: **100** de 100
Enviado 16 jun em 3:48
Esta tentativa levou 344 minutos.

Pergunta 1

6,67 / 6,67 pts

Quantas instâncias e características existem, respectivamente, no dataset?

- (5110, 12)
- (12, 5110)
- (7, 5000)
- (5000, 7)

Pergunta 2

6,67 / 6,67 pts

Quantas variáveis do tipo “string” estão presentes no dataset?

- (●) 6
- ( ) 3
- ( ) 2
- ( ) 4

Pergunta 3

6,67 / 6,67 pts

Qual é a idade (age) média dos entrevistados?

- ( ) 45, 28 anos.
- ( ) 55, 12 anos.
- (●) 43, 22 anos.
- ( ) 22, 61 anos.

Pergunta 4

6,67 / 6,67 pts

Sobre a distribuição de AVC em relação ao sexo (gender) dos entrevistados, é CORRETO afirmar:

- ( ) Existe, no dataset, uma maior quantidade de homens que sofreram AVC.
- ( ) Existe, no dataset, apenas dois tipos de gêneros: homens e mulheres.

- (●) Apesar da pouca diferença, existe uma maior quantidade de mulheres que sofreram AVC.
- ( ) Não podem ser identificadas diferenças entre os gêneros, pois o dataset está equilibrado (mulheres=homens).

Pergunta 5

6,67 / 6,67 pts

É correto afirmar sobre o dataset, EXCETO:

- (●) O dataset está balanceado. Existem quantidades similares de instâncias de indivíduos que sofreram AVC e que não sofreram dessa enfermidade.
- ( ) Existem dois tipos diferentes de classes de residências (“Residence_type”) presentes nesse dataset.
- ( ) Existem dados categóricos e numéricos presentes neste dataset. Um exemplo de dados categóricos é o “Residence_type”.
- ( ) A variável bmi possui valores não numéricos.

Pergunta 6

6,67 / 6,67 pts

Qual é o valor da mediana para a variável do nível médio de glicose do entrevistado (“avg_glucose_level”)?

- ( ) 78.

271,74.

120.

95.

Pergunta 7

6,67 / 6,67 pts

Analisando o padrão de dispersão da variável do nível médio de glicose do entrevistado (“avg_glucose_level”), é correto afirmar, EXCETO:

Os possíveis outliers estão localizados após o limite superior do boxplot.

A dispersão dos dados no terceiro quartil é maior que no segundo, pois existe uma maior quantidade de valores diferentes no terceiro quartil.

Existem valores que correspondem a possíveis outliers. Esses valores certamente devem ser eliminados do dataset, pois sempre causam problemas.

O terceiro quartil corresponde ao valor de 120. Desse modo, podemos dizer que 75% dos dados estão abaixo desse valor.

Pergunta 8

6,67 / 6,67 pts

Analisando a dispersão dos dados para a variável idade (“age”), é correto afirmar, EXCETO:

- Pelo Boxplot não é possível identificar possíveis outliers.
- A mediana para essa variável corresponde ao valor de 68 anos.
- O maior existente para a idade dos entrevistados corresponde a 82 anos.
- O primeiro quartil indica que 25% dos dados estão abaixo de 30 anos.

Pergunta 9

6,67 / 6,67 pts

Quantas classes diferentes para a variável “work_type” existem no dataset?

- 2
- 6
- 5
- 4

Pergunta 10

6,67 / 6,67 pts

Dentre as classes de tipos de trabalhos existentes (*work_type*), qual é aquela que possui uma maior quantidade de instâncias?

- Self-employed

- (●) Private
- ( ) Never_worked
- ( ) Govt_job

Pergunta 11

6,67 / 6,67 pts

Qual foi, respectivamente, o percentual de dados utilizados para o treinamento e teste do modelo?

- ( ) (70%, 30%)
- ( ) (30%, 70%)
- ( ) (20%, 80%)
- (●) (80%, 20%)

Pergunta 12

6,67 / 6,67 pts

Analisando as variáveis “bmi” e “smoking_status”, é CORRETO afirmar:

- ( ) Ambas são variáveis numéricas
- ( ) Existem oito classes distintas de “smoking_status”
- ( ) A variável “bmi” possui apenas valores numéricos
- (●) Ambas possuem instâncias com valores desconhecidos

Pergunta 13

6,67 / 6,67 pts

Após o agrupamento dos dados de “smoking_status” e ”stroke”, é CORRETO afirmar que:

- Existem seis classes diferentes de “smoking_status”
- Não é possível realizar o agrupamento, pois os dados possuem dimensões diferentes
- **Dentre os entrevistados que sofreram AVC, existem uma maior quantidade de indivíduos da classe que nunca fumaram (never smoked).**
- Neste dataset, existe uma maior quantidade de indivíduos que sofreram AVC.

Pergunta 14

6,67 / 6,67 pts

Sobre a relação entre a hipertensão (*hypertension*) e o AVC (*stroke*) presente neste dataset, é CORRETO afirmar:

- **A proporção de incidência de AVC é maior nos indivíduos que sofrem de hipertensão.**
- A proporção entre indivíduos hipertensos e não hipertensos no dataset é a mesma.

Existe uma maior quantidade de dados de indivíduos hipertensos.

Os dados mostram que este dataset está balanceado.

Pergunta 15

6,62 / 6,62 pts

Sobre o algoritmo de regressão logística aplicado para a previsão da ocorrência de AVC, é correto afirmar, EXCETO:

A regressão logística não deveria ser aplicada ao problema, pois ela trabalha apenas com dados categóricos.

Como o dataset está desbalanceado, a acurácia (accuracy) resultante pode estar enviesada.

A árvore de decisão também poderia ser aplicada para esse modelo de classificação.

A acurácia do modelo é superior a 90%

Pontuação do teste: **100** de 100