

Pontuação deste teste: **100** de 100

Enviado 12 nov em 21:26

Esta tentativa levou 481 minutos.

Pergunta 1

6,67 / 6,67 pts

Quantas instâncias e características existem, respectivamente, no dataset?

☐ (12, 5110)

☒ (5110, 12)

☐ (5000, 7)

☐ (7, 5000)

Pergunta 2

6,67 / 6,67 pts

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

☐ 3

☐ 4

☐ 2

☒ 6

Pergunta 3

6,67 / 6,67 pts

Qual é a idade (age) média dos entrevistados?

☐ 22,61 anos

☒ 43,22 anos

☐ 55,12 anos

☐ 45,28 anos

Pergunta 4

6,67 / 6,67 pts

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

☐ Não podem ser identificadas diferenças entre os gêneros, pois o dataset está equilibrado (mulheres=homens)

☐ Existe no dataset apenas dois tipos de gêneros, homens e mulheres

☒ Apesar da pouca diferença, existe uma maior quantidade de mulheres que sofreram AVC

☐ Existe no dataset uma maior quantidade de homens que sofreram AVC

Pergunta 5

6,67 / 6,67 pts

Sobre o dataset é correto afirmar, EXCETO:

☐ Existem dois tipos diferentes de classes de residências ("Residence_type") presentes nesse dataset.

☒ O dataset está balanceado. Existem quantidades similares de instâncias de indivíduos que sofreram AVC e que não sofreram dessa enfermidade.

☐ 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”)?

☒ 95

☐ 120

☐ 271,74

☐ 78

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:

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

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

☒ 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

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



A mediana para essa variável corresponde ao valor de 68 anos



O primeiro quartil indica que 25% dos dados estão abaixo de 30 anos



O maior existente para a idade dos entrevistados corresponde a 82 anos



Pelo Boxplot não é possível identificar possíveis outliers

Pergunta 9

6,67 / 6,67 pts

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



6



2



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?

- ☐ Govt_job
- ☐ Never_worked
- ☒ Private
- ☐ Self-employed

Pergunta 11

6,67 / 6,67 pts

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

- ☐ (20%, 80%)
- ☒ (80%, 20%)
- ☐ (70%, 30%)
- ☐ (30%, 70%)

Pergunta 12

6,67 / 6,67 pts

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

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

Pergunta 13

6,67 / 6,67 pts

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

☐

Neste dataset existe uma maior quantidade de indivíduos que sofreram AVC

☐

Existem seis classes diferentes de "smoking_status"

☒

Dentre os entrevistados que sofreram AVC, existem uma maior quantidade de indivíduos da classe que nunca fumaram (never smoked)

☐

Não é possível realizar o agrupamento, pois os dados possuem dimensões diferentes

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:

☐

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

☒

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

☐

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