

# AV1

**Entrega** 20 mar em 23:59      **Pontos** 40      **Perguntas** 16

**Disponível** 14 mar em 0:00 - 20 mar em 23:59 7 dias

**Límite de tempo** 120 Minutos      **Tentativas permitidas** 2

## Instruções

Leia com atenção as orientações abaixo antes de iniciar esta prova:

- Esta prova contem 16 perguntas.
- Serão permitidas duas tentativas para realizar esta avaliação, prevalecendo a maior nota.
- Programe-se para realizar suas avaliações com tranquilidade, pois você terá 120 minutos cronometrados para conclusão e envio das respostas.
- Ao iniciar a avaliação o cronômetro não para, independentemente da plataforma estar aberta ou não.
- Durante a realização da prova:
  - Será exibido uma questão por vez, podendo “Avançar” ou “Voltar” quando necessário dentro do período da tentativa.
  - A tentativa somente será contabilizada após clicar no botão “Enviar”.



## Histórico de tentativas

	<b>Tentativa</b>	<b>Tempo</b>	<b>Pontuação</b>
<b>MANTIDO</b>	<a href="#"><u>Tentativa 2</u></a>	44 minutos	34 de 40
<b>MAIS RECENTE</b>	<a href="#"><u>Tentativa 2</u></a>	44 minutos	34 de 40
	<a href="#"><u>Tentativa 1</u></a>	106 minutos	20 de 40

Pontuação desta tentativa: **34** de 40

Enviado 20 mar em 19:52

Esta tentativa levou 44 minutos.

<b>Pergunta 1</b>	2 / 2 pts
<p>Os descritores estatísticos como frequência, medidas de tendência central e medidas de variabilidades são importantes para a análise descritiva de dados em Data Science.</p> <p>Avalie a seguintes afirmações e a relação entre elas</p>	

I – Um conjunto de dados que apresenta alguns valores ausentes podem ser tratados e preenchidos com medidas de tendência central como média, mediana ou moda

### PORQUE

II – As medidas de tendência central como média, mediana ou moda são ótimas estratégias para se fazer o tratamento de dados ausentes sem que seja preciso o descarte destes dados.

A respeito dessas asserções, assinale a opção correta.

As asserções I e II são proposições falsas.

Correto!



As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.

A asserção I é uma proposição falsa e a II é uma proposição verdadeira.

As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.

A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.

### Pergunta 2

2 / 2 pts

Data Science não pode ser considerada uma ferramenta completamente técnica, pois emprega técnicas e teorias de campos como matemática e estatística, juntamente com ciência da computação e ciência da informação. Uma diferença entre aplicações de Data Science e aplicações de BI tradicionais é que data Science usa dados

**Correto!**

- Estruturados
- Não estruturados.
- Relacionais
- Estatísticos
- De negócios

**Pergunta 3**

2 / 2 pts



Sobre as medidas estatísticas avalie as seguintes afirmações:

- I. Quando a dispersão dos dados for muito grande então o desvio padrão será grande.
- II. O segundo quartil é igual a mediana.
- III. Quando a média é igual ao desvio padrão os dados não estarão dispersos.

Estão corretas:

- Todas.
- Somente uma delas.
- I e II somente.
- I e III somente.
- II e III somente.

**Correto!****Pergunta 4**

2 / 2 pts

Um analista de IA gerou uma matriz de confusão para um algoritmo de aprendizado de máquina. Ele não vai conseguir calcular:

O F-Score do modelo preditivo.

A fórmula do modelo preditivo.

A acurácia do modelo preditivo.

A precisão do modelo preditivo.

O recall do modelo preditivo.

**Correto!**



## Pergunta 5

2 / 2 pts

Sobre aprendizado de máquina é correto afirmar que:

No aprendizado não supervisionado é fundamental rotular os dados.

Tanto no aprendizado supervisionado como no não supervisionado o modelo gerado não apresenta previsões incorretas.

No aprendizado não supervisionado, existem somente as variáveis dependentes que descrevem os dados.

No aprendizado por reforço os algoritmos não precisam de dados.

**Correto!**

No aprendizado supervisionado o conjunto de dados devem ser divididos em base de testes e base de treino.

**Pergunta 6****2 / 2 pts**

A linguagem Python ganhou uma enorme popularidade na área de análise de dados e IA. Avalia as afirmações abaixo sobre Python e a relação entre elas:

I – A linguagem Python é muito utilizada em análise de dados devido a sintaxe simples, uso de código enxuto, possuindo importantes bibliotecas como NumPy.

**PORQUE**

II – A linguagem Python incorpora diferentes bibliotecas como NumPy utilizadas para executar operações matemáticas e funções integradas para álgebra linear, porém, não incorpora bibliotecas para tratamento de medidas estatísticas como desvio padrão.

A respeito dessas asserções, assinale a opção correta.



As asserções I e II são proposições verdadeiras, mas a II não é uma justificativa correta da I.



A asserção I é uma proposição falsa e a II é uma proposição verdadeira.



As asserções I e II são proposições falsas.



As asserções I e II são proposições verdadeiras, e a II é uma justificativa correta da I.

**Correto!**

A asserção I é uma proposição verdadeira, e a II é uma proposição falsa.

**Pergunta 7****2 / 2 pts**

Existem diferentes estratégias de fazer análise exploratória dos dados.  
Em qual categoria enquadra-se o processo de Machine Learning?

Análise Diagnóstica.

Análise Qualitativa

**Correto!**

Análise Preditiva.

Análise Inteligente.

Análise Descritiva.



## Pergunta 8

2 / 2 pts

O algoritmo K-Means tem como objetivo:

Encontrar classificadores rotulados entre os dados e juntá-los conforme o valor do argumento k.

Encontrar agrupamentos rotulados entre os dados e separá-los conforme o número de agrupamentos definidos pelo argumento k.

Predizer as diferenças entre os dados e alocá-las conforme o número de cluster passado pela variável k.

**Correto!**

Encontrar similaridades entre os dados e agrupá-los conforme o número de cluster passado pelo argumento k.

Predizer as classes de agrupamentos mais próximos do hiperparâmetro K.

## Pergunta 9

3 / 3 pts

Desde os tempos mais remotos o ser humano produz dados. Os homens das cavernas escreviam nas paredes. Seus hieróglifos tornaram-se dados que permitem hoje compreendermos o modo como viviam. Tempos depois, os pergaminhos foram escritos e os dados contidos contam a história de povos antigos. Atualmente, milhões de dados são produzidos por hora; logo, armazená-los e tratá-los torna-se um desafio para todos. Há, porém, um desafio maior ainda, que é interpretar esses dados e tirar deles algum valor, assim como fizemos com os hieróglifos e pergaminhos antigos.



De acordo com o seu conhecimento e os conceitos de ETL, assinale a alternativa correta em relação aos tipos e exemplos de visualização de dados.

A fase de transformação (Transform) do ETL é uma técnica de mineração de dados que pode ser usada para descobrir padrões nos dados.

A fase de extração é considerada a mais importante, pois é nela que os dados são enviados para o Data Warehouse.

A extração de dados consiste em limpar os dados, resolvendo problemas como inconsistências nesses dados.

**Correto!**

As ferramentas de ETL automatizam o processo de extração de dados dos diferentes bancos de dados e o entrega para um Data Warehouse.

Ao realizar a carga (load) para um Data Warehouse, os dados são organizados de acordo com o tipo, por exemplo, tipos de texto e tipos de números.

O processo de ETL desempenha um papel fundamental nas estratégias de integração de dados. O ETL permite que as empresas coletem dados de várias fontes e os consolidem em um único local centralizado. O ETL também possibilita que diferentes tipos de dados trabalhem juntos. Você pode ler mais na unidade 1, tópico 4.



## Pergunta 10

3 / 3 pts

A mineração de dados é o processo de análise de grandes volumes de dados para descobrir inteligências que ajudem as empresas a resolver problemas, mitigar riscos e aproveitar novas oportunidades. Esse ramo da ciência de dados tem a nomenclatura derivada das semelhanças entre procurar informações valiosas em um grande banco de dados e extrair uma montanha de minério.

Considerando as informações sobre mineração de dados, analise as asserções a seguir a respeito dos principais conceitos de mineração de dados.

- I. Usa aprendizado de máquina.
- II. Usa dados diretamente dos sistemas de processamento de transações.
- III. Ajuda os gerentes na tomada de decisões de rotina.
- IV. Limpeza e preparação de dados.

Está correto o que se afirma em:

**Correto!** I e III. I e IV. III e IV. II e IV. II e III.

Essa alternativa está correta porque limpeza e preparação de dados é uma etapa na qual os dados são transformados em um formulário adequado para análises e processamentos adicionais, como a identificação e a remoção de erros e a falta de dados (dados incompletos). Do mesmo modo, o aprendizado de máquina é um conceito de mineração de dados no qual um programa, usando probabilidades estatísticas, permite aos computadores ter a capacidade de "aprender" sem que sejam explicitamente programados. Você pode ler mais na unidade 1, tópico 3.



## Pergunta 11

0 / 3 pts

Considere o código abaixo, digitado em Python.

```
a, b = 0, 1  
while b < 10:  
    print b  
    a, b = b, a+b
```

No final da execução do código, o último valor armazenado nas variáveis *a* e *b* serão, respectivamente,

 5 e 10. 13 e 20. 8 e 10.**Resposta correta**

Você respondeu

13 e 21

8 e 13

## Pergunta 12

3 / 3 pts

Considere-se o código Python abaixo.

```
def checa_lista( list ):
    elem = list[ 0 ]
    for a in list:
        if a > elem:
            elem = a
    return elem

print(checa_lista([4, 4, 8, -3]))
```



O que será impresso?

-3

13

2

8

16

Correto!

## Pergunta 13

3 / 3 pts

O proprietário de um restaurante, vendo a baixa nas vendas diárias, decidiu realizar uma pesquisa com o objetivo de saber a opinião dos

clientes sobre os seus pratos. Durante 7 dias ele pediu aos clientes para avaliarem o serviço com notas que iam de zero (muito insatisfeito) até dez (muito satisfeito). No final dos 7 dias, o proprietário estava com os seguintes dados: 20 clientes atribuíram a nota 0 ao restaurante; 180 dos entrevistados atribuíram a nota 10; e 200 clientes deram a nota 5 ao restaurante, totalizando 400 clientes. Como forma de incentivo, o proprietário ofereceu R\$200,00 aos empregados se acaso a média, a moda ou a mediana estivessem acima de 7,5 pontos. Por outro lado, caso os clientes tenham avaliado muito mal e a média for menor ou igual a 5, o proprietário dará desconto de 10% aos clientes.

A partir dessas informações e do conteúdo estudado no livro da disciplina, podemos considerar que o que acontecerá no restaurante será:



A média obtida com o comando `median()` foi 4, a mediana foi 6 e a moda foi 5. Dessa forma, o proprietário dará desconto de 10% aos clientes e não mudará nada em relação aos funcionários



A média foi 4, a mediana foi 4 e a moda que foi obtida pelo método `mode()` resultou em 8. Dessa forma, os funcionários receberão R\$200,00 e os clientes terão desconto de 10%.



A média obtida com o comando `median()` foi 8, a mediana foi 5 e a moda obtida pelo comando `mode()` foi 5, logo não mudará nada no restaurante.



Com o comando `mean()` obteve-se média igual a 7,5, a mediana foi 10 e a moda foi 5. Dessa forma, o proprietário dará R\$200,00 aos funcionários e não mudará nada em relação aos clientes

Correto!



Com o comando `mean()` foi obtida a média 7, com o comando `median()` foi obtida a mediana 5 e a moda obtida pelo comando `mode()` foi 5. Dessa forma, não mudará nada no restaurante

A figura abaixo ilustra o processo de análise para obtermos a moda e a mediana. Note que o tamanho da amostra é igual a 400. Vamos ordenar os dados do menor para o maior (ou seja, de zero a dez). A primeira seta da segunda linha indica que, do intervalo de 0 a 20, essas pessoas deram nota zero. A segunda seta, ainda na mesma linha, indica que, no intervalo de 20 a 220, os entrevistados deram 5 pontos. Por fim, a última seta indica que, do intervalo entre 220 e 400, os entrevistados deram 10 pontos.



Note que, se colocarmos uma marcação bem no centro da segunda linha, essa marcação cairá exatamente sobre o intervalo de 20 a 220, ou seja, intervalo de 5 pontos. Logo, a mediana é 5. A moda é o valor que mais aparece que, nesse caso, também é o 5. Por outro lado, a média é a soma de todos os valores dividida pelo tamanho da amostra, que é  $400 - 2800/400 = 7$ , logo a média é 7. Assim temos: média 7, mediana e moda 5. Os respectivos métodos para encontrarmos esses valores são: para média é `mean()`, para mediana é `median()` e para moda é o `mode()`. Dessa forma, de acordo com os dados do enunciado, não haverá mudança no restaurante. Veja mais no tópico 2 da unidade 3.

### Pergunta 14

3 / 3 pts

Ao lidar com dados estatísticos, é importante distinguir entre "população" dos conjuntos de dados e "amostra" dos conjuntos de dados. Em um volume de dados, a população contém todos os membros de um grupo específico. Por outro lado, a amostra de um grupo de dados contém uma parte, ou um subconjunto, de uma

população. O tamanho de uma amostra é sempre menor do que o tamanho da população da qual é retirada. Considerando essas informações e os conteúdos estudados na disciplina, analise as asserções a seguir:

- I. A população pode ser todas as pessoas que vivem no Brasil.
- II. A amostra pode ser todas as pessoas que vivem no Nordeste.
- III. A amostra pode ser todas as pessoas entre 40 e 50 anos que vivem no Brasil.

Está correto apenas o que se afirma em:

- I apenas.
- III apenas.
- II apenas.
- II e III.
- I e III.

Correto!



A asserção I está correta, pois quando falamos de população estamos tratando o todo, ou seja, todas as pessoas que vivem no Brasil, sem exceção. Já a asserção II está incorreta, pois o conceito de amostra está relacionado a pegarmos uma parte da população, ou seja, uma parte do todo. A asserção diz “todas as pessoas que vivem no Nordeste”, logo isso caracteriza uma população. Se fosse afirmado: “todas as pessoas do Brasil que vivem no Nordeste”, teríamos o Brasil como população e as pessoas que vivem no Nordeste como amostra. Quanto a asserção III está correta, pois pega-se uma parte do todo. O todo (população) são as pessoas que vivem no Brasil. A amostra são as pessoas que vivem no Brasil e que possuem entre 40 e 50 anos. Veja mais sobre amostra e população no tópico 1 da unidade 3.

**Pergunta 15**

3 / 3 pts

O aprendizado de máquina não supervisionado é uma classe de problemas de aprendizado na qual se procura estabelecer agrupamentos de informações que compartilham das mesmas características.

Considere as asserções abaixo:

- I - Determinar grupos de pacientes com base nos resultados dos exames.
- II - Determinar as características de compras dos consumidores de um supermercado, agregando-os.
- III - Determinar a direção de um carro autônomo em uma rodovia com base nas respostas dos sensores.
- IV - Determinar o risco de intrusão em uma rede de computadores com base nos ataques anteriores dos hackers.

Em relação às asserções anteriores, quais dizem respeito ao aprendizado de máquina não supervisionado?

Correto!

I e II.

II e IV.

III e IV.

II e III.

I e III.

#### Feedback:

No aprendizado não supervisionado, o algoritmo de aprendizado de máquina deve procurar por grupos que tenham as mesmas características com base nos valores das variáveis contidas no conjunto de dados, como apresentado na unidade 4, tópico 1.1.

**Pergunta 16**

0 / 3 pts

Considere o conjunto de dados a seguir como sendo o conjunto de dados original a ser utilizado por um algoritmo de aprendizado de máquina supervisionado:

Dia	Tempo	Temperatura	Vento	Pedalar
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não
D10	Chuva	Amena	Fraco	Sim

Considerando o que já foi estudado e a prática usual sobre os conjuntos de dados de treinamento e de teste, uma representação deles pode ser encontrada em:



### Conjunto de treinamento

<b>Dia</b>	<b>Tempo</b>	<b>Temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não
D10	Chuva	Amena	Fraco	Sim

- Conjunto de teste



<b>Dia</b>	<b>Tempo</b>	<b>Temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não
D10	Chuva	Amena	Fraco	Sim

### Conjunto de treinamento

<b>Dia</b>	<b>Tempo</b>	<b>Temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D1	Sol	Calor	Fraco	Sim
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D7	Nublado	Frio	Forte	Sim
D10	Chuva	Amena	Fraco	Sim

- Conjunto de teste

<b>Dia</b>	<b>Tempo</b>	<b>Temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D2	Sol	Calor	Forte	Não
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não

**resposta correta****Conjunto de treinamento**

<b>dia</b>	<b>tempo</b>	<b>temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não
D10	Chuva	Amena	Fraco	Sim

 Conjunto de teste

<b>dia</b>	<b>tempo</b>	<b>temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não

**você respondeu****Conjunto de treinamento**

<b>dia</b>	<b>tempo</b>	<b>temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não

 Conjunto de teste

<b>dia</b>	<b>tempo</b>	<b>temperatura</b>	<b>vento</b>	<b>Pedalar</b>
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não
D9	Nublado	Frio	Forte	Não
D10	Chuva	Amena	Fraco	Sim



### Conjunto de treinamento

Dia	Tempo	Temperatura	vento	Pedalar
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não
D3	Nublado	Calor	Fraco	Sim
D4	Chuva	Amena	Fraco	Sim
D5	Chuva	Frio	Fraco	Não
D6	Chuva	Frio	Forte	Não
D7	Nublado	Frio	Forte	Sim
D8	Sol	Amena	Fraco	Não

### Conjunto de teste

Dia	Tempo	Temperatura	vento	Pedalar
D1	Sol	Calor	Fraco	Sim
D2	Sol	Calor	Forte	Não



### **Feedback:**

A divisão do conjunto original nos conjuntos de treinamento e de teste usualmente é feita na proporção de 80% e 20%, respectivamente, sem repetição dos dados entre um conjunto e outro, de forma que, ao final, o desempenho do classificador possa ser calculado com base em dados previamente não vistos por ele, como apresentado na unidade 4, tópico 2.1.

Pontuação do teste: **34** de 40