

RESIDÊNCIA EM TECNOLOGIA DA INFORMAÇÃO
APLICADA À ÁREA JURÍDICA OU ÓRGÃOS DE CONTROLE
(PÓS-GRADUAÇÃO LATO SENSU)

EDITAL 001/2018 – PROVA DE CONHECIMENTOS ESPECÍFICOS
ÁREA DE CONCENTRAÇÃO: ANALISTA BUSINESS INTELLIGENCE
19 / 03 / 2018

Identificação do Candidato	
Nome completo:	
CPF:	Telefone:

QUESTÕES

1. São exemplos de abordagens de mineração de dados mais apropriados para se criar um modelo de classificação:

- a) Algoritmo memético e algoritmo guloso
- b) Regressão linear simples e regressão multivariada
- c) Redes neurais e árvores rubro-negras
- d) Árvore de decisão e redes bayesianas

2. São exemplos de abordagens para treinamento e validação de modelos em um processo de mineração de dados:

- I. *RapidMiner Tool*
- II. *K-fold cross-validation*
- III. *Leave-one-out*
- IV. RUP

Estão corretas as afirmativas:

- a) I e III
- b) I e IV
- c) II e IV
- d) II e III

3. Quais das técnicas abaixo não é utilizada para realizar redução de dimensionalidade nos dados?

- a) *Principal Component Analysis*
- b) Análise de variância de atributos
- c) *Bootstrap*
- d) Análise de correlação entre atributos

4. Sobre as árvores de decisão, analise as seguintes afirmativas:

- I. Cada nó interno representa um teste para se escolher um caminho que leva a uma determinada previsão.
- II. Cada nó representa a probabilidade de algo acontecer e levar a um caminho de uma determinada previsão.
- III. A poda é uma técnica utilizada para reduzir a complexidade da árvore.
- IV. A árvore de decisão é uma árvore balanceada, tendo as folhas a mesma profundidade.

Estão corretas as afirmativas:

- a) I e III
- b) I e IV
- c) II e III
- d) II e IV

5. Acerca dos tipos de gráficos existentes, assinale a alternativa correta:

- a) Gráficos de barras ilustram tendências ao longo do tempo.
- b) Gráficos de pizza são muito bons para representar correlações entre os dados.
- c) Gráficos de caixa (*boxplots*) representam bem a dispersão dos dados.
- d) Gráficos de linhas são úteis para se ilustrar dados percentuais.

6. Os histogramas são representações gráficas de:

- a) Distribuição dos dados
- b) Correlação entre variáveis
- c) Acurácia de classificação
- d) Predição de classes

7. Uma empresa de auditoria foi contratada para analisar o banco de dados com o cadastro dos advogados registrados na Ordem dos Advogados do Rio Grande do Norte (OAB/RN). A primeira tarefa é encontrar os pares de advogados cadastrados que possuem o mesmo nome e números diferentes de inscrição na OAB para verificar possíveis fraudes. Considere que a tabela que armazena o cadastro dos advogados possui o número de inscrição como chave primária e as colunas referentes a nome, endereço, número de telefone e data de ingresso na Ordem.

Assinale a alternativa que apresenta a consulta em SQL que busca corretamente os resultados, ordenados pelo nome dos advogados:

- a)

```
SELECT a1.nome, a1.inscricao, a2.inscricao
FROM advogados AS a1, advogados a2
WHERE a1.nome = a2.nome AND inscricao1 > inscricao2
GROUP BY nome
```
- b)

```
SELECT a1.nome, a1.inscricao, a2.inscricao FROM advogados a1
JOIN advogados a2 ON a1.nome = a2.nome
WHERE a1.inscricao > a2.inscricao
ORDER BY a1.nome
```

-
- c) `SELECT a1.nome, a1.inscricao, a2.inscricao FROM advogados a1
WHERE nome IN (SELECT nome FROM advogados a2 WHERE a1.nome = nome AND
a1.inscricao > inscricao)
ORDER BY nome`
- d) `SELECT * FROM advogados as a1, advogados AS a2
WHERE a1.nome LIKE a2.nome AND inscricao1 > inscricao2
GROUP BY a1.nome`

8. Antes da construção de um modelo de classificação, é comum realizar uma etapa de pré-processamento. Quais das etapas a seguir não pode ser considerada como um pré-processamento:

- a) Balanceamento dos dados
- b) Amostragem
- c) Transformação de tipos
- d) Treinamento

9. Sobre o processo de amostragem dos dados, analise as seguintes afirmativas:

- I. Para ser considerada representativa, uma amostra deve possuir aproximadamente as mesmas propriedades dos dados originais.
- II. O uso de uma amostra para construir um modelo de classificação tende a produzir um menor custo computacional.
- III. Uma amostragem com reposição tem sempre o mesmo efeito que uma amostragem sem reposição.
- IV. Para conseguir uma amostra representativa, seu tamanho deve ser sempre superior a 20% do total dos dados.

Estão corretas as afirmativas:

- a) I e II
- b) III e IV
- c) II e III
- d) II e IV

10. Ao utilizar uma base de dados, é comum que muitos dos dados não estejam presentes. A falta de alguns valores nas bases de dados se deve, muitas vezes, por falta de preenchimento nos cadastros, falha em sensores, entre outros motivos. Para que técnicas de aprendizado de máquina possam utilizar corretamente os dados, em geral, é necessária a correção desse tipo de problema. Dentre as abordagens utilizadas para tratar esse problema, analise as seguintes afirmativas:

- I. É possível eliminar os exemplos que possuem valores faltando em seus atributos.
- II. Pode-se substituir o campo faltoso por valores retirados a partir dos próprios dados, como a média ou a moda.
- III. É possível utilizar um modelo de classificação para preencher automaticamente os dados.

Sobre as afirmativas anteriores, é correto dizer que:

- a) Somente a afirmativa III está correta.

- b) Todas as afirmativas estão corretas.
- c) Nenhuma afirmativa está correta.
- d) Somente a afirmativa I está correta.

11. Sobre a análise de agrupamentos, é possível afirmar que:

- a) O objetivo principal é encontrar um conjunto de associações entre os itens de uma base de dados.
- b) Tem como meta a sumarização dos conjunto de dados em valores mais simples.
- c) Identificar elementos semelhantes e separá-los para uma análise exploratória mais profunda.
- d) Seu uso requer que os dados estejam devidamente rotulados com as respectivas classes.

12. Sobre os componentes de um *data warehouse*, considere as seguintes características:

- I. informações identificadas por assuntos ou departamentos específicos.
- II. capacidade de encontrar padrões nos dados.
- III. capacidade de analisar informações em múltiplas perspectivas.
- IV. processo de extração, tratamento e limpeza dos dados.

As características de I a IV são, respectivamente:

- a) *Staging Area*, *Data Mining*, OLAP, ETL
- b) *Data Mart*, *Data Mining*, OLAP, ETL.
- c) *Drill Through*, OLTP, *Drill Across*, *Staging Area*.
- d) Cubo de dados, OLTP, *Data Mining*, operações *Drill*.

13. Analise as seguintes afirmativas sobre *Business Intelligence*:

- I. Uma das etapas da abordagem OLAP é a normalização dos dados na 3ª Forma Normal.
- II. A predição de eventos é uma das análises da mineração de dados.
- III. O *data warehouse* é o banco de dados que contém dados sumarizados, sendo otimizado para operações de inserção e atualização.

Está(ão) correto(s):

- a) III, apenas.
- b) I e III, apenas.
- c) II e III, apenas.
- d) II, apenas.

14. Em um *data warehouse* relacionado a vendas, considere a existência de uma dimensão *Localização* com a hierarquia de: *País* > *Estado* > *Cidade* > *Bairro*. Detalhar as informações de venda, deixando de visualizar o país para ver a cidade da venda é uma operação de:

- a) *Slice*
- b) *Drill-Across*
- c) *Drill-Up*
- d) *Drill-Down*

15. Qual ferramenta tem como objetivo realizar o processo de extração, transformação, carga de dados de dados?

- a) PDI
- b) Mondrian
- c) Saiku
- d) Kibana

Considerando uma modelagem multidimensional através do esquema estrela sobre os julgamentos em um tribunal, onde deseja-se analisar quem julgou, o que se julgou, em que data, o valor financeiro especificamente solicitado e o valor liberado após julgamento (em termos de faixas de valores), analise as questões 16 a 19:

16. A realização do julgamento é representada como:

- a) tabela fato
- b) tabela dimensão
- c) atributo de tabela fato
- d) atributo de tabela dimensão

17. O mês do julgamento é representado como:

- a) tabela fato
- b) tabela dimensão
- c) atributo de tabela fato
- d) atributo de tabela dimensão

18. O juiz que julgou é representado como:

- a) tabela fato
- b) tabela dimensão
- c) atributo de tabela fato
- d) atributo de tabela dimensão

19. O valor financeiro solicitado no processo é representado como:

- a) Tabela fato
- b) Tabela dimensão
- c) Atributo de tabela fato
- d) Atributo de tabela dimensão

20. Sobre a chave primária de uma tabela fato em uma modelagem esquema estrela (*Relational OLAP*), assinale a afirmativa correta:

- a) É composta por uma chave autoincrementável.
- b) É composta por uma ou mais chaves estrangeiras.
- c) É composta por um valor de negócio definido pela aplicação cliente.
- d) Tabelas fato não possuem chave primária.

Para as questões a seguir, considere a modelagem apresentada na Figura 1, referente ao banco de dados de um sistema de gerenciamento de eventos classificados por categoria, cujos ingressos podem ser comprados individualmente e nominalmente apenas por usuários cadastrados no sistema, restringindo-se a um único ingresso por usuário para cada evento cadastrado.

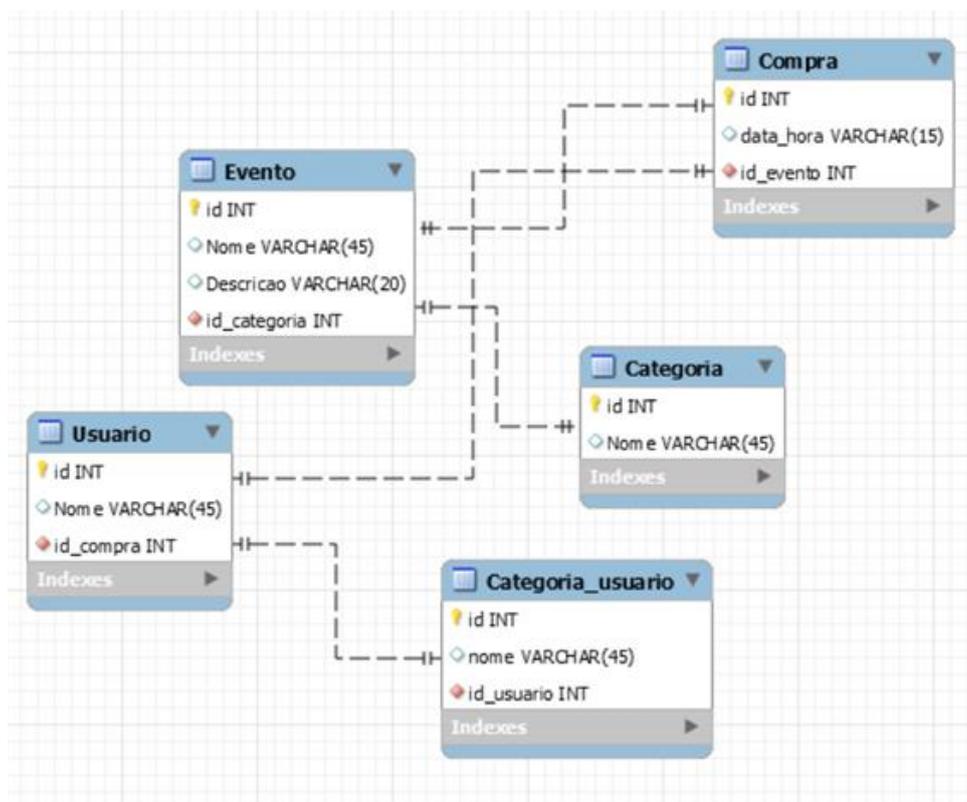


Figura 1. Modelagem do banco de dados.

21. Considerando a modelagem apresentada na Figura 1, analise as seguintes afirmativas sobre problemas existentes ou falta de boas práticas aplicadas nessa modelagem:

- I. O tipo utilizado na coluna `data_hora` da tabela `Compra` não é o mais apropriado.
- II. A tabela `Usuario` não deveria ter a coluna `id_compra`.
- III. A coluna `Nome` da tabela `Categoria` deveria ser uma coluna da tabela `Evento`.
- IV. As linhas que representam relações entre tabelas poderiam não estar se cruzando.

Estão corretas as seguintes afirmativas:

- a) I, II, III
- b) I, III, IV
- c) II, III, IV
- d) I, II, IV

22. Considerando o modelo da Figura 1, analise as seguintes afirmativas:

- I. Para se visualizar para quais eventos cada usuário comprou ingresso, usamos o seguinte comando SQL: `SELECT * FROM evento, categoria, compra, usuario`
- II. Para se visualizar os nomes dos usuários que compraram ingresso para o evento de ID igual a 2, usamos o seguinte comando SQL: `SELECT usuario.nome FROM usuario WHERE usuario.id_compra=compra.id AND compra.id_evento=2`
- III. Para se visualizar os eventos que não foram relacionados a nenhuma categoria, usamos o seguinte comando SQL: `SELECT evento.nome FROM evento WHERE evento.id_categoria=0`
- IV. Para se visualizar a quantidade de eventos cadastrados no sistema, usamos o seguinte comando SQL: `SELECT SUM(id) FROM evento`

Quanto às alternativas anteriores, são falsas:

- a) apenas uma
- b) apenas duas
- c) apenas três
- d) todas

23. Com relação aos tipos de bancos NoSQL, analise as afirmativas a seguir:

- I. Bancos de dados de documentos armazenam dados como documentos (JSON, XML, etc.). Um exemplo de banco deste tipo é o MongoDB.
- II. O banco de dados Neo4J é de um tipo de banco que possuem vértices e arestas representando as relações entre esses vértices.
- III. Banco de dados colunares guardam colunas juntas, ao invés de linhas, sendo o tipo de banco do Neo4J.

Estão corretas as afirmativas:

- a) I, III
- b) I, II
- c) II, III
- d) I, II, III

24. Considerando os comandos disponíveis no *shell* do MongoDB e uma coleção de dados alunos não vazia, com atributos nome, mediaGeral e avaliacao, analise as seguintes afirmativas:

- I. O seguinte comando retorna uma lista vazia, uma vez que os critérios de busca não foram definidos:
`db.alunos.find({})`
- II. No comando abaixo, o programador colocou erroneamente o \$in no lugar do \$or para encontrar alunos baseado nos valores de sua nota média geral:
`db.alunos.find({ mediaGeral: { $in: [7, 10] } })`

III. Para atualizar na coleção `alunos` o atributo `avaliacao` de todos os alunos com média maior que 9, podemos executar o seguinte comando:

```
db.alunos.updateMany(  
  { mediaGeral: { $gt: 9 } },  
  { $set: { avaliacao: "Ótimo desempenho!" } }  
)
```

IV. Para inserir um novo registro na coleção `"alunos"`, podemos executar o seguinte comando:

```
db.alunos.insertOne(  
  {  
    nome: "João",  
    mediaGeral: 7,  
    avaliacao: "Na média"  
  }  
)
```

Estão corretas as afirmativas:

- a) I e II
- b) II e III
- c) III e IV
- d) I e IV

25. Considerando a modelagem de um banco de dados utilizando o esquema estrela, avalie as seguintes afirmativas acerca das tabelas fatos:

- I. Geralmente são as que possuem maior quantidade de dados (linhas).
- II. Geralmente são as que possuem menor quantidade de dados (linhas).
- III. Precisam estar ligadas a todas as tabelas dimensões do *data warehouse*.
- IV. Não precisam estar ligadas a todas as tabelas dimensões do *data warehouse*.

Estão corretas as afirmativas:

- a) I e III
- b) I e IV
- c) II e III
- d) II e IV

26. Acerca da linguagem SQL, analise os seguintes comandos:

- I. `INSERT INTO Clientes (Nome, Idade) VALUES ('Jonas', 18);`
- II. `DELETE FROM Clientes VALUES ('Jonas', 18);`
- III. `UPDATE Clientes (Nome, Idade) VALUES ('Jonas', 18);`
- IV. `INSERT INTO Clientes VALUES ('Jonas', 18);`

Estão sintaticamente corretos os seguintes comandos:

- a) I e III
- b) I e IV

- c) II e III
d) II e IV

27. Analise o seguinte comando SQL:

```
SELECT COUNT(id), Nome FROM Clientes  
GROUP BY Nome  
HAVING COUNT(Nome) > 2  
ORDER BY COUNT(Nome) DESC;
```

Sobre o comando apresentado, analise as seguintes afirmativas:

- I. O comando está sintaticamente incorreto.
- II. A consulta pode identificar nomes iguais de clientes diferentes.
- III. A cláusula `HAVING` não pode ser substituída por `WHERE`.
- IV. O resultado será ordenado de forma decrescente pelo nome do cliente.

Estão corretas as seguintes afirmativas:

- a) II e IV
b) I e III
c) I e IV
d) II e III

Analise o seguinte código para responder às questões 28 e 29.

Código 1. Consulta a um *data warehouse* modelado através de um esquema estrela e implementado em um banco de dados relacional.

```
SELECT Loja.Nome, Data.DataCompleta, Produto.Nome, SUM(Vendas.Quantidade) AS  
Total  
FROM Vendas, Data, Produto, Loja  
WHERE Vendas.idProduto = Produto.idProduto AND Vendas.idLoja = Loja.idLoja  
GROUP BY Loja.Nome, Data.DataCompleta, Produto.Nome  
ORDER BY Data.DataCompleta, Loja.Nome, Produto.Nome
```

28. Baseado no Código 1, analise as seguintes afirmativas:

- I. Produto é uma tabela fato no *data warehouse*.
- II. Vendas é uma tabela fato no *data warehouse*.
- III. A cláusula `WHERE` está adequada, considerando que as chaves primárias das tabelas são armazenadas em colunas de nome `Id` + nome da tabela.
- IV. A consulta retorna uma tabela com quatro colunas, sendo a última de nome `Total`.

Estão corretas as afirmativas:

- a) I e III
b) I e IV
c) II e III
d) II e IV

29. Sobre a consulta apresentada no Código 1, analise as seguintes afirmativas:

- I. A cláusula ORDER BY não é obrigatória, mesmo com o uso da função SUM no SELECT.
- II. Na cláusula SELECT, a indicação do nome da tabela antes do nome da coluna é dispensável.
- III. O uso da cláusula GROUP BY é obrigatória nesse caso, já que a função SUM está sendo utilizada.
- IV. No caso dessa consulta, o uso de “+=Vendas.Quantidade” equivale ao uso da função SUM.

Estão corretas as afirmativas:

- a) I e III
- b) I e IV
- c) II e III
- d) II e IV

30. Em uma modelagem dimensional utilizando o esquema estrela, algumas perguntas sobre operações do negócio são utilizadas para se encontrar tabelas dimensões. Assinale qual das alternativas usualmente não leva a informações de uma tabela dimensão, mas sim de uma tabela fato.

- a) Onde?
- b) Quem?
- c) Quanto?
- d) O que?