

Disciplina: Banco de Dados
AD2 Primeiro Semestre de 2024
Nome: Gaya Marinho de Oliveira

Considere o esquema relacional abaixo de um gerenciador de *Ensembles* de Aprendizado de Máquina para responder às questões de 1 a 7. Um mesmo usuário pode criar *Ensembles* que são submetidos a uma aplicação gerenciadora. A aplicação então treina diversos modelos de Aprendizado de Máquina (por exemplo, *Support Vector Machines*) de acordo com as definições do *Ensemble* do usuário. Assim, diversos modelos treinados estão associados a um mesmo *Ensemble*. Cada modelo é treinado usando um determinado algoritmo de Aprendizado de Máquina. A ideia do *Ensemble* é aumentar a precisão e a resiliência nas previsões, fazendo a combinação de previsões de vários modelos. As chaves primárias se encontram sublinhadas no esquema a seguir.

USUARIO(<u>uid</u> , unome, ulogin, uprovedor, uIP, país)	% Identificador, nome, login, provedor, país e IP do usuário
ALGORITMO(<u>aid</u> , titulo, categoria)	% Identificador, título e categoria de algoritmos de Aprendizado de Máquina
ENSEMBLE(<u>eid</u> , uid, data_submissao, hora_submissao, num_epocas, tamanho_batch, taxa_aprendizado)	% Identificador do <i>ensemble</i> , data e hora da submissão e parâmetros
uid referencia USUARIO(uid)	
MODELO(<u>eid</u> , <u>aid</u> , acuracia, precisao)	
eid referencia ENSEMBLE(eid)	
aid referencia ALGORITMO(aid)	

1. Apresente o comando SQL que retorne o nome dos usuários que nunca criaram um *Ensemble*. (1,0 pontos)

Solução:

SELECT unome AS 'Usuários que nunca criaram um ensemble'

FROM usuario

WHERE

uid NOT IN (SELECT ensemble.uid FROM ensemble);

2. Apresente o comando SQL que retorne o título dos algoritmos que foram usados para gerar algum modelo em algum *Ensemble*. (1,0 pontos)

Solução:

SELECT titulo AS 'Títulos dos algoritmos usados para treinar um ensemble'

FROM algoritmo

WHERE

aid IN (SELECT modelo.aid FROM modelo);

3. Considere que a tabela USUARIO possui o campo unome que inclui o nome e sobrenome da pessoa. Apresente o comando SQL que lista os nomes dos usuários que possuem os sobrenomes OLIVEIRA em conjunto com CARDOSO ou MORAES, sendo que OLIVEIRA deve ser o primeiro sobrenome. (1,0 pontos)

Solução:

```
SELECT unome AS 'Nomes'
FROM usuario
WHERE (unome LIKE '%oliveira%moraes%' OR unome LIKE '%oliveira%cardoso%')
      AND unome NOT LIKE '%moraes%oliveira%'
      AND unome NOT LIKE '%cardoso%oliveira%';
```

4. Apresente o comando SQL que exclui a tabela ENSEMBLE. Observação: o objetivo é excluir a tabela, não somente suas tuplas. (1,0 pontos)

Solução:

```
DROP TABLE ENSEMBLE;
```

5. Apresente um comando SQL que insere a tupla (1, 'Support Vector Machines', 'Não Supervisionado') na relação ALGORITMO. (1,0 pontos)

Solução:

```
INSERT INTO ALGORITMO
VALUES
      (1, "Support Vector Machines", "Não Supervisionado");
```

6. Apresente um comando em SQL que exclua os algoritmos que nunca foram usados para treinar um modelo. (1,0 pontos)

Solução:

```
DELETE FROM algoritmo
WHERE aid IN (SELECT modelo.aid FROM modelo);
```

7. Apresente um comando em SQL que atualize as tuplas da tabela ENSEMBLE dobrando o valor do atributo num_epocas e aumentando em 20% o valor do atributo tamanho_batch. (1,0 pontos)

Solução:

```
UPDATE ENSEMBLE
```

SET num_epocas = num_epocas * 2, tamanho_batch = tamanho_batch * 1.2;

8. Considere a relação R(ID, Nome, UF, Salario). A chave primária se encontra sublinhada. Elabore uma consulta em SQL que apresente a média dos salários por estado (UF), mas somente quando a média for superior a R\$ 5.000,00 e para pessoas que possuem SANTANA no sobrenome (considere que o campo NOME é um VARCHAR que engloba tanto o nome quanto o sobrenome da pessoa. (1,5 pontos)

Solução:

SELECT UF, avg(Salario) as "Média dos salários"

FROM R

WHERE Nome LIKE '%SANTANA%'

GROUP BY UF

HAVING avg(Salario) > 5000;

9. Analise as relações apresentadas a seguir e informe em qual forma normal elas se encontram (NN, 1FN, 2FN ou 3FN). Justifique suas respostas. As chaves primárias de cada relação se encontram sublinhadas. Assuma que todos os professores em um mesmo nível da carreira ganham o mesmo salário. (1,5 pontos)

(a) PROFESSOR(CodDepto, NomeDepto, CodProfessor, NomeProfessor)

(b) PROFESSOR(CodProfessor, NomeProfessor, Salario)

(c) DEPARTAMENTO(CodDepto, NomeDepto, (CodProfessor, NomeProfessor))

(d) PROFESSOR(CodProfessor, NomeProfessor, NivelCarreira, Salario)

Solução:

- (a) **A relação está na 1ªFN.** Não há tabelas aninhadas, mas ela apresenta duas situações de dependência parcial ($\text{CodDepto} \rightarrow \text{NomeDepto}$ e $\text{CodProfessor} \rightarrow \text{NomeProfessor}$), sendo que a chave primária da tabela é (CodDepto, CodProfessor).
- (b) **A relação está na 3ª FN,** pois não se observam nela tabelas aninhadas, dependências parciais ou dependências transitivas.
- (c) **A relação está na forma Não Normalizada (NN),** pois ela contém uma tabela aninhada: (CodProfessor, NomeProfessor).
- (d) **A relação está na 2ª FN.** Isso porque não existem tabelas aninhadas nem há dependências parciais. Porém, há dependência transitiva de Salario em relação a NivelCarreira (de acordo com o enunciado, o nível da carreira determina o salário) e Salario não faz parte da chave primária da relação. Portanto, a relação não chega a atender a um dos pré-requisitos para estar na 3ª FN.