

# Fatec Carapicuíba - 4ADS-Noite

Nome: **Evelyn Juliane Da Hora Sousa**

## **Programação para banco de dados**

### **1. Segundo o texto qual o significado da Sigla MAD? Qual a principal vantagem de fazer análises de dados "In DATABASE"**

MADlib é uma biblioteca de métodos analíticos que podem ser instalados e executados dentro de um mecanismo de banco de dados relacional que suporta SQL extensível. Um instantâneo do conteúdo atual do MADlib, incluindo métodos e portas é fornecida na Tabela 1. Este conjunto de métodos e portas se destina a crescer ao longo do tempo.

Os métodos no MADlib são projetados para execução dentro ou fora do escopo e para o paralelismo de compartilhamento de nada e de expansão. oferecidos pelos modernos mecanismos de banco de dados paralelos, garantindo que o cálculo seja feito próximo aos dados.

### **2. Quais os principais algoritmos disponíveis para análises estatísticas no MADlib?**

Aprendizado supervisionado - que usa os métodos:

- Regressão Linear
- Regressão Logística
- Classificador
- Naive Ba
- yes
- Árvores de Decisão
- Support Vector Machines

Aprendizado não supervisionado – que usa os métodos:

- k-Means Clustering
- Matriz de Fatoração SVD
- Alocação de Dirichlet latente
- Regras de Associação

Estatística Descritiva – que usa os métodos:

- Esboço Count-Min
- Esboço Flajolet-Martin
- Perfil de Dados
- Quantis

Módulo de Apoio – que usa os métodos:

- Vetores Esparsos
- Operações de Matriz
- Conjugate Gradient Optimization

**3. Sobre qual linguagem de programação o MADlib foi desenvolvido? E como o PostgreSQL consegue utilizar essas bibliotecas?**

Ele foi desenvolvido em Python.

E o PostgreSQL consegue utilizar as ferramentas justamente pela implementação do SQL.

**4. Nós fizemos um pequeno exemplo usando o algoritmo APRIORI. Envie a sintaxe do SELECT da função que utilizamos para cálculo das combinações de produtos mais importantes do nosso supermercado. (disponível na documentação do MADlib)**

```
SELECT (linregr(y, x)).* FROM data;
```

```
-[ RECORD 1 ]+-----  
coef | {1.7307,2.2428}  
r2 | 0.9475 std_err | {0.3258,0.0533}  
t_stats | {5.3127,42.0640}  
p_values | {6.7681e-07,4.4409e-16}
```