

## **Lista de Exercício 01**

### **1. O que podemos caracterizar como Big Data e por que esse termo é tão falado?**

O termo Big Data pode ser descrito como a capacidade de lidar com grandes volumes de dados, além do volume, envolve a utilização de técnicas e tecnologias específicas para armazenar, processar e analisar os tipos de dados de maneira eficiente e escalável.

Os 5Vs do Big Data são: Volume, Variedade, Velocidade, Veracidade e Valor

### **2. O que é HDFS e MapReduce? Foram desenvolvidos para solucionar qual necessidade?**

HDFS são arquivos distribuídos e MapReduce são as ordens distribuídas onde são processadas de formas paralelas.

Foram desenvolvidos para solucionar problemas relacionados ao armazenamento e processamento de grandes volumes de dados

### **3. Descreva com suas palavras o que foi entendido dos conceitos de Inteligência Artificial e Machine Learning.**

Inteligência Artificial: máquinas que buscam criar sistemas e programas que simulam a inteligência humana para realizar processos e tarefas

Machine Learning: é a capacidade das máquinas aprenderem usando algoritmos a partir de dados

### **4. Com suas palavras descreva o que é um algoritmo de Machine Learning?**

Um algoritmo de machine learning é um conjunto de regras e cálculos que ensina uma máquina a reconhecer padrões e fazer previsões com base em dados. Ex: Árvore de decisão

### **5. Com suas palavras, descreva o processo de Ciência de Dados, detalhando e explicando cada etapa do processo.**

Definição do problema de negócio: Identifica qual é o desafio a ser resolvido

Aquisição dos dados: Coleta os dados que será utilizado para a resolução do problema

Preparação dos dados: Limpa e organiza os dados coletados, que serão usados na análise

Análise exploratória dos Dados: Aqui o foco é analítico, onde os dados são examinados

Modelagem: Cria-se a representação visual dos dados e seus relacionamentos

Visualização e Comunicação: Criação de gráficos e relatórios para a apresentação dos resultados da análise de forma clara e compreensível.

Deploy e Manutenção: Implementa o modelo em um ambiente de produção e monitora seu desempenho ao longo do tempo.

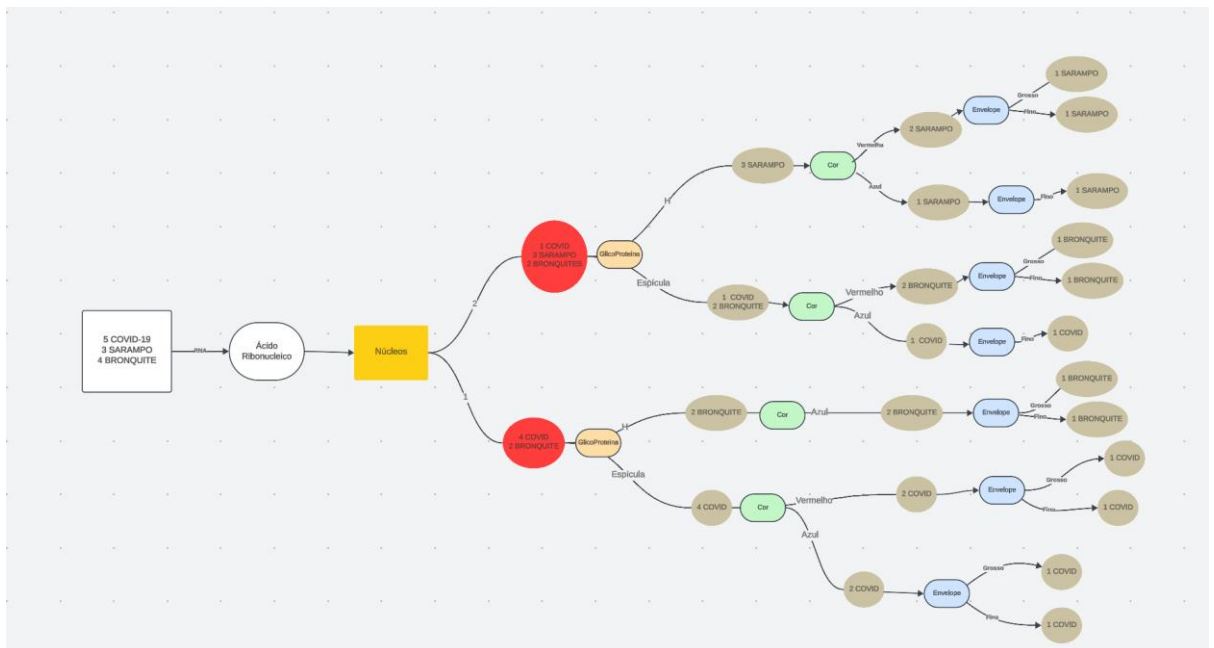
**6. O que é ABT? Descreva e se possível conte um exemplo.**

*Analytical Base Table* , Análise de Negócios em Tempo Real, integra dados e análises em tempo real para ajudar na tomada de decisão.

Um exemplo simples é uma rede de supermercados, onde as análises precisam ser em tempo real para a gestão de estoque.

**7. Qual a diferença entre a ABT que usamos para criar o modelo e a ABT que vai entrar em produção?**

Talvez seja o objetivo, o objetivo para criar o modelo é a resolução do problema, o objetivo em produção é a implementação.



**Descreva todas as regras encontradas na árvore de decisão feita acima.**

Quantidade de núcleos; Tipo GlicoProteína; Cor; Tipo Envelope

Não acho relevante o RNA, pois é estático

**Responda utilizando as regras detalhadas na questão anterior:**

**Se um vírus tem: envelope FINO, glicoproteína H, 2 núcleos, cor AZUL com RNA ácido ribonucleico... que vírus é esse?**

Sarampo

**Se um vírus tem: envelope GROSSO, glicoproteína H, 1 núcleo, cor AZUL com RNA ácido ribonucleico... que vírus é esse?**

Bronquite

**Se um vírus tem: envelope GROSSO, glicoproteína ESPÍCULA, 1 núcleo, cor VERMELHA com RNA ácido ribonucleico, que vírus é esse?**

COVID-19