**1 – Defina com suas palavras o que é Data science?**

R: Data science é uma área multidisciplinar que parte da colet de dados de diversas fontes para analiase e subsidiar a tomada de decisões, de forma preditiva, em grandes quantidades e gerando insights. Nesta área são utilizados conceitos de negócio, como business intelligence, programação, como python e R para limpeza e organização de dados, e também matemática e estatistica, juntamente com a programação para a criação de modelos de predição através da tecnica de machine learning e deep learning.

**2- Descreva resumidamente as principais etapas de um projeto de Data Science.**

R: 1- Entender o problema do negócio, 2- Aquisição dos dados (web scrapping), 3- Preparação dos dados, usualmente o dado de uma base vem faltando informações ou o formato vem diferente. 4- Exploração dos dados (avaliar se há outliers no meio dos dados). 5- Modelagem dos dados (onde se cria o modelo a partir dos dados limpos). 6- Visualização e comunição (onde você apresenta os dados e resultados para a área de negócios). 7-Colocar em produção o modelo criado

**3- As habilidades essenciais de um cientista de dados são:**

1. Comunicação
2. Programação
3. Gestão de Dados Estruturados
4. Matemática
5. Gestão de Projetos
6. Data Mining e Visualização
7. Design de Experimentos
8. Gestão de Dados
9. Design e Desenvolvimento de Produtos
10. Modelagem estatística
11. Desenvolvimento de negócios

Por ter bastante anos de trabalho em industria tenho experiencia em comunição, gestão de projetos, por ser formado em química tenho uma base forte em matemática e bom conhecimento em estatistica. Tenho conhecimento básico em analise de dados usando python.

**4- Quais são as principais ferramentas e linguagens de Data Science?**

R: As principais utilizadas em Data Science são python, uma linguagem mais simples de ser aprendida em relação a outras, também pelo fato de possuir excelentes bibliotecas para análise de dados como pandas, numpy e matplotlib, além de biblioteca para machine learning como o sickitlearn. Também é muito utilizada a linguagem R na área de Data Science.

É muito utilizada a IDE do jupyter notebook para a analise de dados, alem do anaconda, onde já possui pré instaladas diversas bibliotecas uteis para analise de dados. É muito utilizado para a visualização de dados, ferramentas como power BI, tableau e matlab.

O github é uma ferramenta excelente como forma de exibir seu portifólio além de poder contribuir para codigos de terceiros.

**5- Quais as diferenças de Big Data, Business Intelligence e Data Science?**

**R:** O objetivo do Business Intelligence é converterdados brutos em insights de negócios para lideres tomarem decisões. Esta área sempre utiliza dados do passado para gerar insights

Já a ciência de dados emprega o método científico para a exploração de dados, formação de hipóteses e teste de hipótese por meio de simulação e modelagem estatística. Dentro da ciência de dados ainda se utiliza o machine learning como ferramenta para automatizar a transformação de dado em informação

A principal diferença é que o BI trata de dados no passado enquanto o data science vai tratar do futuro, a partir de análise preditiva.

Já Big Data é o termo que trata sobre grandes conjuntos de dados, que podem ser estruturados ou não estruturados. que precisam ser processados e armazenados.

**6- Quais as diferenças entre inteligencia artificial, Machine Learning e Deep Learning?**

**R:** Machine learning é a pratica de usar algoritmos para coletar dados, aprender com eles, e então fazer uma determinação ou predição sobre alguma coisa do mundo.

Já o deep learning são algoritmos complexos construídos a partir de um empilhamento de diversas camadas similares a neurônios, alimentados por grandes quantidades de dados. Então, eles são capazes de reconhecer imagens e fala e aprender a realizar tarefas extremamente avançadas sem interferência humana. Mas, a principal aplicação dos algoritmos de Deep Learning são as tarefas de classificação , em especial, reconhecimento de imagens.

Tanto machine learning quanto Deep Learning são conceitos que podemos dizer que fazem parte do conceito de inteligencia artificial. De grosso modo, podemos definir inteligência artificial como a capacidade das máquinas de aprender, decidir e perceber quais caminhos seguir, de forma racional como os seres humanos. O sistema absorve, analisa e organiza dados de forma a entender e identificar o que são objetos, pessoas, padrões e reações de todos os tipos. Portanto, esse processo de ensinar os computadores a aprender (inteligência artificial) passa por diversas subcategorias da computação, como o machine learning e deep learning.