



ATIVIDADE 1 – PROJETOS DATA SCIENCE

Profa. Franciane Rodrigues

Pergunta 1: Quais são as etapas do ciclo de vida de um projeto de Data Science?

Pergunta 2: O que acontece na etapa de definição do problema em um projeto de Data Science?

Pergunta 3: O que envolve a etapa de coleta e preparação dos dados em um projeto de Data Science?

Pergunta 4: Qual é o objetivo da análise exploratória de dados em um projeto de Data Science?

Pergunta 5: O que é modelagem em um projeto de Data Science?

Pergunta 6: Qual é a importância da etapa de avaliação em um projeto de Data Science?

Pergunta 7: O que significa interpretar os resultados em um projeto de Data Science?

Pergunta 8: Existe alguma etapa adicional que pode ocorrer após a interpretação dos resultados em um projeto de Data Science?

Pergunta 9: Quais estratégias para execução de um projeto Data Science e quando usar cada uma delas?



Aluno: Gabriel Costa Andrade

01 de junho de 2023

Pergunta 1: Quais são as etapas do ciclo de vida de um projeto de Data Science?

Resposta:

- 1 - Entendendo do problema
- 2 - Coleta de dados
- 3 - Processamento de dados
- 4 - Exploração de Dados
- 5 - Comunicação de resultados
- 6 - Feedback

Pergunta 2: O que acontece na etapa de definição do problema em um projeto de Data Science?

Resposta:

É necessário o entendimento do problema a ser tratado. Compreender bem, é a metade da solução, por isso é importante dedicar tempo e energia para um maior detalhamento. Esta etapa é o requisito básico de um projeto de qualidade, uma vez que ela direciona a atuação da equipe.

Pergunta 3: O que envolve a etapa de coleta e preparação dos dados em um projeto de Data Science?

Resposta:

Nesta etapa são identificados todos os dados, a serem coletados e armazenados. Um dos principais objetivos da preparação de dados é garantir que as informações resultantes serão precisas e consistentes. Permitindo assim, uma análise válida e realista. Há diferentes tipos de dados:

- Dados Internos (presentes em bancos de dados)
- Dados Externos (bases de dados públicas ou pagas, etc.)
- Dados estruturados (tabelas)
- Dados semi-estruturados (json, web)
- Dados não-estruturados (conteúdos de redes sociais, de sites externos, etc.)

É válido se atentar que não devemos nos limitar aos dados fornecidos pelo cliente. É preciso analisar se os dados fornecidos são suficientes para a resolução do nosso problema e se é necessário a coleta a partir de novas fontes de dados.

Pergunta 4: Qual é o objetivo da análise exploratória de dados em um projeto de Data Science?

Resposta:

Esta etapa exige muito tempo e estudo para a validação das hipóteses geradas nas etapas anteriores, fazendo análises descritivas dos dados, procurando insights, aplicando estudos estatísticos a fim de entender os dados que foram fornecidos e se estão aptos a resolver o problema proposto.

Pergunta 5: O que é modelagem em um projeto de Data Science?

Resposta:

A modelagem de dados analisa e define os diferentes dados existentes, e gerencia as relações entre eles. Etapa em que se cria uma representação visual de todos os dados coletados e produzidos, e os coloca de forma clara para entender onde e para que serão usados, bem como os separa com requisitos. Etapa de grande importância para facilitar e qualificar o trabalho de análise.

Pergunta 6: Qual é a importância da etapa de avaliação em um projeto de Data Science?

Resposta:

Esta etapa desempenha um importante papel na medição do desempenho do modelo, na garantia da generalização, na comparação de modelos, na tomada de decisões e na validação dos resultados. É uma etapa crítica que deve ser realizada de forma cautelosa e rigorosa para garantir que o modelo seja confiável e útil para o problema em questão.

Pergunta 7: O que significa interpretar os resultados em um projeto de Data Science?

Resposta:

Significa extrair valor dos dados e do modelo. A interpretação nos permite obter insights, fazer inferências e tomar decisões informadas com base nas análises realizadas.

Pergunta 8: Existe alguma etapa adicional que pode ocorrer após a interpretação dos resultados em um projeto de Data Science?

Resposta:

Existe mais de uma etapa:

- Comunicação dos resultados
- Implementação de ações recomendadas
- Validação externa
- Refinamento do modelo ou análise
- Monitoramento contínuo

Pergunta 9: Quais estratégias para execução de um projeto Data Science e quando usar cada uma delas?

Resposta:

- SEMMA - Quando é preciso explorar, modificar, modelar e avaliar dados para atingir seus objetivos analíticos. É frequentemente utilizada para a mineração de dados e análise de dados, apesar de não ser uma metodologia de mineração de dados, mas sim uma organização lógica do conjunto de ferramentas centrais de mineração.
- KDD - Estratégico e tático para a tomada de decisões. Visa a descoberta de conhecimento envolvendo a etapa de modelagem de dados.
- CRISP-DM - Tem o mesmo tipo de processamento que o KDD, porém é utilizado para monitoramento em tempo real, dinâmico. É uma das mais utilizadas em mineração de dados pois não depende de ferramentas para ser implementado ou utilizado e pode ser enquadrado em qualquer tipo de negócio.