



FAQ – Academia BB – IA & Ciência de Dados

Sumário

• O que é Ciência de Dados?	2
• O que é Inteligência Artificial (IA)?	2
• O que significa CRISP-DM?	2
• Por que fazer análise exploratória de dados?	2
• O que é <i>storytelling</i> ?	2
• O que é uma variável <i>dummy</i> ?	2
• Por que normalizar variáveis?.....	3
• O que é aprendizado de máquina?	3
• O que é aprendizado supervisionado?	3
• O que é aprendizado não supervisionado?	3
• O que é aprendizado por reforço?	3
• O que é NLP?.....	4
• O que é série temporal?	4
• O que é visão computacional?	4
• O que é IA Generativa?.....	4
• O que é deep learning?	4
• O que é LLM?.....	4
• O que preciso estudar para me tornar um cientista de dados?	4
• Quais são as habilidades essenciais para um profissional de ciência de dados?	5
• Quais são as principais ferramentas utilizadas na área de ciência de dados?	5
• Ciência de dados vs. análise de dados	5
• Ciência de dados versus engenharia de dados.....	5
• O que é cultura data driven?	5
• O que é Mineração de Dados?.....	6



FAQ – Academia BB – IA & Ciência de Dados

- O que é Ciência de Dados?

Ciência de dados é uma área interdisciplinar que combina técnicas de matemática, estatística, programação e inteligência artificial para analisar grandes volumes de dados e extrair insights significativos. O objetivo é resolver problemas complexos e ajudar na tomada de decisões.

- O que é Inteligência Artificial (IA)?

Inteligência Artificial (IA) é um campo da ciência da computação que se dedica ao desenvolvimento de sistemas e programas capazes de realizar tarefas que normalmente requerem inteligência humana.

- O que significa CRISP-DM?

CRISP-DM (Cross Industry Standard Process for Data Mining) é uma metodologia padrão utilizada para orientar projetos de mineração de dados e ciência de dados.

- Por que fazer análise exploratória de dados?

A análise exploratória de dados (EDA, do inglês Exploratory Data Analysis) é uma etapa crucial em qualquer projeto de ciência de dados. Ela ajuda no entendimento dos dados, permite detectar problemas de qualidade dos dados, permite criar hipóteses, ajuda a identificar as variáveis mais importantes, fornece ideias que ajudam a escolher o algoritmo de modelagem e ajuda a orientar os próximos passos do projeto.

- O que é *storytelling*?

Storytelling é a arte de contar histórias de maneira envolvente e cativante, utilizando elementos como personagens, ambiente, conflito e uma mensagem clara. Essa técnica é usada para transmitir informações e emoções de forma que ressoe com o público, criando uma conexão emocional e tornando a mensagem mais memorável.

- O que é uma variável *dummy*?

Uma variável *dummy*, também conhecida como variável indicadora, é uma técnica utilizada em análise de dados e modelagem estatística para representar variáveis categóricas de forma numérica. Em vez de usar categorias como “sim” e “não”, ou “vermelho”, “verde” e “azul”, uma variável *dummy* transforma essas categorias em valores binários, geralmente 0 e 1. Essa transformação é importante porque muitos modelos estatísticos e algoritmos de aprendizado de máquina exigem que as variáveis sejam numéricas para processá-las adequadamente.



FAQ – Academia BB – IA & Ciência de Dados

- Por que normalizar variáveis?

1-Porque você pode ter variáveis com escalas muito diferentes. Sem normalização, a variável com a escala maior pode dominar a análise e influenciar desproporcionalmente os resultados.

2-Porque muitos algoritmos de aprendizado de máquina, como regressão linear e redes neurais, funcionam melhor quando as variáveis estão em uma escala similar. A normalização pode ajudar a acelerar a convergência do algoritmo e melhorar a precisão do modelo.

3-Normalizar os dados pode reduzir a influência de outliers, tornando o modelo mais robusto e menos sensível a valores extremos.

4-Porque facilita a interpretação: dados normalizados são mais fáceis de interpretar e comparar, especialmente quando se trabalha com múltiplas variáveis que têm unidades de medida diferentes.

- O que é aprendizado de máquina?

Aprendizado de máquina, ou machine learning, é um subcampo da inteligência artificial (IA) que se concentra no desenvolvimento de algoritmos e modelos que permitem aos computadores aprender a partir de dados e fazer previsões ou tomar decisões sem serem explicitamente programados para cada tarefa.

- O que é aprendizado supervisionado?

O modelo é treinado com um conjunto de dados rotulados, onde a resposta correta é conhecida. O objetivo é que o modelo aprenda a mapear entradas para saídas corretas.

- O que é aprendizado não supervisionado?

O modelo é treinado com dados não rotulados e deve encontrar padrões ou estruturas ocultas nos dados.

- O que é aprendizado por reforço?

O modelo aprende a tomar decisões através de tentativa e erro, recebendo recompensas ou penalidades com base nas ações que realiza.



FAQ – Academia BB – IA & Ciência de Dados

- O que é NLP?

NLP, ou Processamento de Linguagem Natural (do inglês, Natural Language Processing), é um ramo da inteligência artificial que se concentra em permitir que os computadores compreendam, interpretem e gerem a linguagem humana de maneira natural.

- O que é série temporal?

Uma série temporal é uma sequência de dados coletados ou registrados em intervalos de tempo sequenciais. Esses dados são organizados cronologicamente, o que permite analisar como uma variável muda ao longo do tempo.

- O que é visão computacional?

Visão computacional é uma subárea da inteligência artificial (IA) que permite às máquinas interpretar e compreender o mundo visual de maneira semelhante aos seres humanos. Isso envolve a análise e processamento de imagens e vídeos para extrair informações relevantes e tomar decisões com base nesses dados.

- O que é IA Generativa?

IA generativa, ou inteligência artificial generativa, é um ramo da inteligência artificial que se concentra na criação de novos conteúdos, como texto, imagens, vídeos, música ou código, a partir de dados existentes.

- O que é deep learning?

Deep learning, ou aprendizado profundo, é um subcampo do aprendizado de máquina que utiliza redes neurais artificiais com múltiplas camadas para simular o processo de raciocínio do cérebro humano.

- O que é LLM?

Large Language Models (LLMs): São modelos avançados de inteligência artificial projetados para entender e gerar linguagem humana. Esses modelos são treinados em grandes quantidades de dados textuais e podem realizar tarefas como tradução automática, geração de texto, resposta a perguntas e muito mais. Exemplos de LLMs incluem o GPT-4.

- O que preciso estudar para me tornar um cientista de dados?

É importante ter uma base sólida em matemática, estatística, programação (especialmente em linguagens como Python e R) e conhecimento em áreas específicas como aprendizado de máquina e visualização de dados.



FAQ – Academia BB – IA & Ciência de Dados

- Quais são as habilidades essenciais para um profissional de ciência de dados?

Habilidades em análise estatística, programação, manipulação de dados, aprendizado de máquina, visualização de dados e comunicação são fundamentais.

- Quais são as principais ferramentas utilizadas na área de ciência de dados?

Algumas das ferramentas mais comuns incluem Python, R, SQL, Hadoop, Spark, Tableau e Power BI.

- Ciência de dados vs. análise de dados

Tanto a ciência de dados quanto a análise de dados desempenham funções cruciais na extração de valor dos dados, mas seus focos são diferentes. A ciência de dados é um campo abrangente que usa métodos, incluindo aprendizado de máquina e análise preditiva, para extrair insights dos dados. Por outro lado, a análise de dados se concentra no processamento e na realização de análises estatísticas em conjuntos de dados existentes para responder a perguntas específicas.

<https://www.datacamp.com/pt/blog/what-is-data-science-the-definitive-guide>

- Ciência de dados versus engenharia de dados

A engenharia de dados concentra-se na criação e manutenção da infraestrutura para coleta, armazenamento e processamento de dados, garantindo que os dados sejam limpos e acessíveis. A ciência de dados, por outro lado, analisa esses dados, usando modelos estatísticos e de aprendizado de máquina para extrair insights valiosos que influenciam as decisões de negócios. Em essência, os engenheiros de dados criam as "estradas" de dados, enquanto os cientistas de dados "dirigem" por elas para obter insights significativos. Ambas as funções são vitais em uma organização orientada por dados.

<https://www.datacamp.com/pt/blog/what-is-data-science-the-definitive-guide>

- O que é cultura data driven?

A cultura data-driven (ou orientada por dados) é uma abordagem em que as decisões e estratégias de uma organização são baseadas na análise de dados concretos, em vez de intuições ou suposições. Implementar uma cultura data-driven pode ajudar as empresas a serem mais ágeis, inovadoras e capazes de atingir seus objetivos com maior precisão.



FAQ – Academia BB – IA & Ciência de Dados

- O que é Mineração de Dados?

Ciência de Dados é um campo mais amplo que abrange várias técnicas para analisar e interpretar dados, enquanto Mineração de Dados se concentra especificamente na extração de insights de dados estruturados usando algoritmos estatísticos e de aprendizado de máquina.