

Relatório do bootcamp - 6

Thassiana C. A. Muller

▶ O que é ciência de dados | Nerdologia Tech

O vídeo inicia exemplificando um problema real de guerra no qual precisa-se decidir quais partes de um avião devem ser blindadas contra projéteis, assim, observa-se, nos aviões que retornaram, que áreas como motores e turbinas eram muito menos atingidas quando comparadas ao restante da fuselagem. Uma análise rasa desses dados poderia levar a uma conclusão equivocada de que os disparos não acertaram nas regiões dos motores, porém o que acontecia é que os aviões atingidos em partes críticas, como motores, não retornavam e, conseqüentemente, não alimentavam os dados. Assim, o cientista de dados é aquele que interpreta além do bruto um conjunto de informações.

Com o estudo de ciência de dados pode-se obter previsões, baseadas em matemática e estatística, de eventos futuros com base em observações passadas, esses estudos quando unidos ao aprendizado de máquina e reconhecimento de padrões levam a excelentes ferramentas para interação e compreensão da máquina sobre o mundo. Dessa forma, o cientista de dados precisa saber modelar, programar e usar ferramentas para garantir que as informações utilizadas estarão em formatos padronizados e “limpos” para que possam ser bem aproveitados por outras técnicas.

▶ What REALLY is Data Science? Told by a Data Scientist

O vídeo inicia esclarecendo que *data science* trata, principalmente, de utilizar dados para criar novos *insights* e padrões valiosos para uma empresa através de ferramentas para esse fim. Dessa maneira, o papel fundamental de um cientista de dados é utilizar os dados para resolver problemas reais e melhorar serviços. O precursor da ciência de dados era a mineração de dados que era usada para retirar alguns conhecimentos úteis de base de dados, porém, após unir esse conceito com ciência de computação e estatística, foi possível expandir a capacidade de processar informações a um nível nunca antes visto.

Com o avanço da globalização pela internet, uma quantidade gigantesca de dados começou a ser gerada, inicia-se então a era da informação e *big-data* e a necessidade de um tratamento especializado de dados para seu aproveitamento em técnicas mais robustas como o aprendizado de máquina, *business intelligence*, análise exploratória, dentre muitas outras.

As principais necessidades da ciência de dados são:

1. Coleta de dados;
2. Armazenamento adequado de dados;
3. Transformação e limpeza dos dados;
4. Classificação e análise dos dados;
5. Extrapolação para outras técnicas como IA e Deep Learning

Um bom cientista de dados deve saber priorizar e extrair de cada etapa a fim de satisfazer as necessidades do seu cliente.

Analista de dados x Cientista de dados x Engenheiro de dados

A diferença entre analista, cientista e engenheiro de dados pode ser entendida com base em suas responsabilidades, habilidades e o foco de seu trabalho. Dependendo da empresa a atribuição das atividades podem ser diferentes mas em teoria os analistas de dados focam na interpretação e visualização de dados para suporte à decisão, os cientistas de dados aplicam técnicas avançadas para criar previsões e insights acionáveis, e os engenheiros de dados garantem que a infraestrutura de dados esteja funcionando de maneira eficiente e eficaz.