

Relatório 6 - Vídeo: O Que É Ciência De Dados (I)

Igor Carvalho Marchi

Descrição da atividade

O card 6 apresenta dois vídeos, onde ambos apresentam o assunto do que é ciência de dados, onde é mostrado que a ciência de dados vai muito além de apenas dados coletados, nos vídeos demonstram de várias maneiras que ciência de dados é entender os dados e interpretá-los de maneira correta, para assim você conseguir incluir esses dados no contexto que você irá usar para que cause um impacto positivo.

O vídeo do canal do Nerdologia é citado o exemplo sobre Abraham Wald na segunda guerra mundial, onde cujo o objetivo era colocar blindagem nos aviões para proteção de tiros para evitar quedas e destruições significativas, a princípio ele estava usando dados apenas de aviões que voltavam da guerra e coletava onde haviam os principais tiros neles. Contudo depois percebeu que não estava totalmente correto, pois os aviões que deveriam ser usado como dados seria os abatidos, então logo começou a verificar os abatidos e percebeu uma diferença onde a maioria dos aviões que foram abatidos foram por tiros no motor e nas asas, esse é um bom exemplo de como devemos interpretar os dados e o contexto que devem ser usados, nesse caso não foi necessário os dados dos aviões que voltaram, pois o intuito era proteger os aviões que fossem destruídos, por isso deveriam pegar os dados dos aviões que foram derrubados.

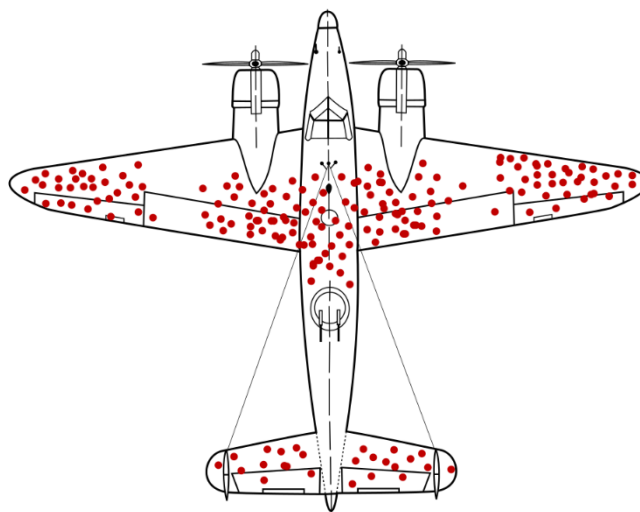


Figura 1 – Principais pontos de perfurações de tiros em aviões.

Agir como Abraham Wald significa adotar a mentalidade de um cientista de dados: buscar, nos dados, informações que vão além do que está visível. Na era do Big Data e dos algoritmos, nosso desafio é transformar grandes volumes de informação em conhecimento útil para antecipar o que pode acontecer. Hoje, já somos capazes não apenas de compreender e explicar eventos passados, mas também de prever, com base em padrões e análises, os acontecimentos futuros.

No segundo vídeo apresentado pelo canal Joma Tech traz uma visão de atuação mais aprofundada do funcionamento do campo de ciência de dados, Joma explica que os dados devem ser usados para gerar resultados significativos, tais como produtos ou até mesmo estratégias.

Joma apresenta também a origem do conceito de “Data Science”, que no início era chamado de “Data Mining” processo de extrair informações úteis a partir de dados. Em 2001, William Cleveland propôs o termo e o conceito de “Data Science”, unindo o poder estatístico da ciência da computação às técnicas do Data Mining, ampliando seu alcance e aplicabilidade.

Com a chegada da Web 2.0 e das redes sociais como MySpace, Facebook e YouTube, passou-se a gerar um volume enorme de dados, caracterizando o fenômeno do Big Data. Como os métodos tradicionais não eram suficientes para tratá-los, surgiram novas tecnologias, como computação paralela, MapReduce, Hadoop e Spark. Esse avanço impulsionou a demanda por cientistas de dados capazes de treinar modelos baseados em dados, tornando reais conceitos como redes neurais, support vector machines e deep learning.

Apesar da visibilidade de Machine Learning e Inteligência Artificial, o trabalho de um cientista de dados vai muito além disso, abrangendo diversas tarefas que impactam diretamente o desempenho e os resultados das empresas.

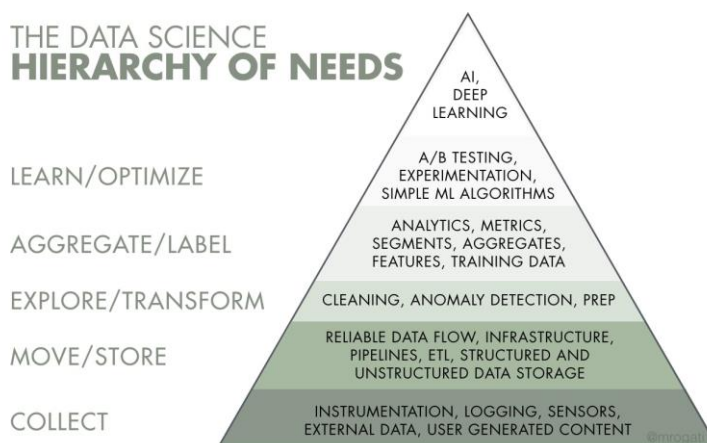


Figura 2 – Pirâmide da hierarquia das necessidades.

A pirâmide de necessidades da Ciência de Dados mostra que o uso de IA e deep learning só é possível após uma base sólida: começa-se pela coleta de dados (sensores, registros, fontes externas), seguida do armazenamento e fluxo confiável, limpeza e detecção de anomalias, agregação e rotulagem para criar features, aplicação de análises e algoritmos de machine learning, até chegar aos modelos mais avançados.

A pirâmide de necessidades da Ciência de Dados mostra o caminho para chegar ao uso de tecnologias avançadas como Inteligência Artificial e deep learning. Tudo começa pela coleta de dados, que pode vir de sensores, registros, fontes externas ou conteúdo criado por usuários. Depois, é preciso organizar e guardar essas informações de forma segura e estruturada.

O passo seguinte é preparar os dados, limpando erros, corrigindo problemas e deixando tudo pronto para análise. Na sequência, eles são agrupados e organizados

em informações úteis, que servirão de base para criar modelos. Com isso pronto, é possível aprender e melhorar usando testes, experimentos e algoritmos mais simples. Só então se chega ao topo da pirâmide, onde entram as soluções mais sofisticadas, como IA e deep learning, que dependem de todo esse trabalho anterior para funcionarem bem.

Conclusões

Os vídeos analisados demonstram que a Ciência de Dados vai muito além de simplesmente coletar e armazenar informações, exigindo interpretação contextual para soluções eficazes, como mostrou Abraham Wald. Com a evolução das tecnologias e o Big Data, surgiram métodos inovadores de análise, como destaca Joma Tech. A pirâmide de necessidades da Ciência de Dados revela que avanços como IA e deep learning dependem de uma base sólida de organização e preparação dos dados. O cientista de dados desempenha um papel estratégico, ajudando a tomar decisões mais precisas e gerar valor para empresas e sociedade.

Referências

Link do vídeo 1 - <https://www.youtube.com/watch?v=ykSILAQQu6o>

Link do vídeo 2 - <https://www.youtube.com/watch?v=xC-c7E5PK0Y>

Link da imagem 1- https://en.wikipedia.org/wiki/Abraham_Wald#

Link da imagem 2 - <https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007>