

# Relatório 9 - Big Data is a Better Data (II)

Yuri Vacelh Zamulhak Zdebski

## Descrição da atividade

Assistir a palestra de Kenneth Cukier do canal TED sobre Big Data e discutir os conhecimentos adquiridos.

## Vídeo

O vídeo aborda o conceito de "big data" e sua relevância para a sociedade. Kenneth inicia com uma analogia sobre a torta mais popular dos EUA para demonstrar como o aumento na quantidade de dados pode alterar nossa compreensão dos fatos. Quando tortas maiores são compradas, a de maçã é a escolha mais comum, mas ao se comprar tortas menores, uma maior diversidade de sabores é selecionada. Isso ocorre porque, ao escolher uma torta grande, a decisão deve agradar a todos (geralmente tortas maiores são consumidas por mais de uma pessoa), e a maçã acaba sendo, nas palavras do próprio Kenneth, o segundo sabor favorito de todos. Esse exemplo evidencia que mais dados não apenas ampliam nossa visão, mas nos possibilita realizar uma análise mais detalhada e precisa, revelando padrões invisíveis em conjuntos menores. Embora possa parecer contraintuitivo, ter mais informações (uma visão mais ampla) gera insights mais profundos e precisos, como demonstrado no caso das tortas.

A maneira como coletamos e armazenamos dados evoluiu significativamente ao longo dos séculos. Antigamente, os dados eram armazenados de forma manual e física, como em placas de argila, livros ou pinturas. Hoje, com a digitalização, os dados se tornaram fluidos e dinâmicos, permitindo que informações sejam acessadas e processadas remotamente. Além disso, a coleta de dados também avançou: localização, expressões faciais e informações médicas podem ser capturadas para análise. Um exemplo recente são os smartwatches, que monitoram em tempo real aspectos como localização, frequência cardíaca e número de passos durante um treino.

Os avanços nas pesquisas de aprendizado de máquina estão diretamente relacionados ao aumento da quantidade de dados disponíveis. Modelos como o ChatGPT só são possíveis graças ao processamento de um grande volume de informações. Atualmente, é possível treinar modelos para quase qualquer área, desde que se tenha os dados necessários. Exemplos incluem carros autônomos, reconhecimento de imagens, entre outros. Essas tecnologias têm o potencial de gerar grandes impactos na sociedade, o que torna necessário considerar a ética envolvida. Podemos realmente confiar na identificação de um crime por inteligência artificial? No diagnóstico de uma doença? Os dados são verdadeiramente neutros, sem viés político ou ideológico? Pessoas perderão seus empregos devido às automações? Esses são apenas alguns dos questionamentos que essas tecnologias podem trazer.

Além dos conteúdos abordados no vídeo, no contexto de big data, é necessário averiguar a qualidade dos dados, afinal, milhares de gigabytes de 0's são simplesmente inúteis, independente do seu tamanho, por isso existem alguns pilares a serem considerados.

- Volume: ainda que não seja o único parâmetro, ainda é necessário que exista uma grande quantidade de informação
- Variedade: a variedade possibilita justamente que o modelo identifique padrões, sejam eles visíveis ou não

- Velocidade: diz respeito ao quão rápido os dados são gerados
- Veracidade: dados equivocados vão inevitavelmente levar o modelo a cometer enganos
- Valor: o quão útil são os dados, treinar modelos hoje em dia ainda é caro e leva tempo

## Conclusão

A era do big data trouxe uma nova perspectiva para a compreensão de fenômenos complexos, revelando informações que antes passavam despercebidas. Com o aumento exponencial dos dados, surgem novas oportunidades, mas também novos desafios éticos. A confiança nas decisões automatizadas e o impacto no mercado de trabalho são questões que exigem uma abordagem cautelosa. À medida que exploramos o potencial do big data, é crucial lembrar que a qualidade, a diversidade e o valor dos dados são tão importantes quanto o volume.