

Relatório 06 - O Que É Ciência De Dados

Guilherme Loan Schneider

Descrição da atividade

O vídeo “O que é Ciência de Dados?” do Nerdologia Tech, cita a utilização da grande produção de dados voltada para a predição futura, ou seja, para onde os dados tendem a se comportar no futuro, isso claro, utilizando cálculos estatísticos.

Atrelado a isso, é exemplificado com a blindagem em aviões na Segunda Guerra Mundial, onde era necessário estudar os pontos onde havia maior incidência de projéteis, e assim definir os melhores pontos onde esse reforço deveria ser colocado, certo? Sim, e não.

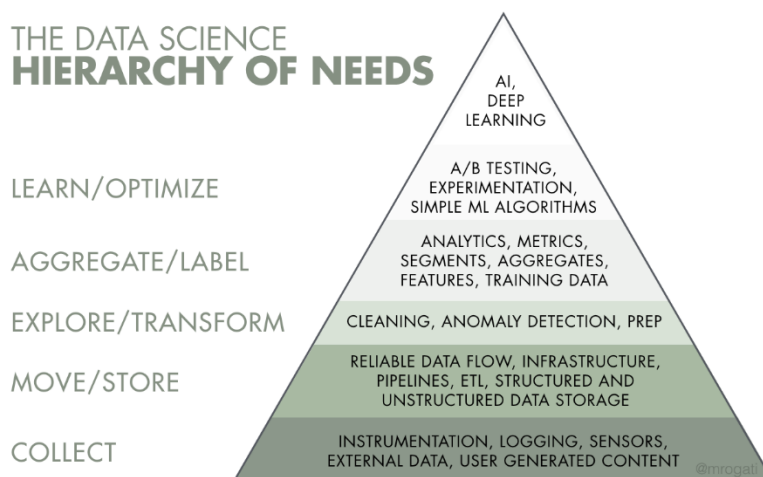
A ideia de analisar os pontos de maior incidência de tiros não é incorreta, mas analisar apenas os aviões que retornaram a base é que não fazia tanto sentido. Assim, o matemático Abraham Wald analisou além do que apenas os dados mostram e concluiu que os motores dos aviões seriam o maior ponto fraco (levando com base o total de aviões que existiam na frota – total aviões que retornaram), visto que eles não conseguiram retornar para a base para ser feita a análise.

Outro importante tópico citado é a garantia da limpeza dos dados a serem utilizados (que muitas vezes não tem um tipo definido) e a qualidade deles, visto que utilizar uma base “incorreta” trará resultados incorretos ou inconsistentes.

O segundo vídeo, “What REALLY is Data Science? Told by a Data Scientist”, inicia comentando que a ciência de dados pode existir de várias maneiras, seja ela por insights, predição de dados, dentre outros. Além disso, comenta um pouco da história da Ciência de Dados, que era chamada, em 1996, de Data-Mining. Com a junção do poder computacional e o Data-Mining, surge a Ciência de Dados.

Isso criou uma era de possibilidades, dando início ao surgimento de inúmeras plataformas digitais, como o Youtube, Facebook, Skype, que em meados de 2005 utilizavam a Ciência de Dados para analisar os seus dados.

É mister citar também que como Cientistas de Dados, nosso objetivo não é apenas analisar dados, mas sim tomar decisões para problemas de alguma empresa, guiando-a para o rumo correto.



Conforme a imagem acima, a hierarquia das necessidades em Ciência de Dados começa na coleta de dados para serem utilizados, esses por sua vez podem vir de qualquer dispositivo

capaz de gerar uma informação. Em seguida é feita a armazenagem desses dados, seja em um Data Warehouse, um Banco de Dados convencional, um arquivo .csv, dentre outros, e a aplicação do principal processo de tratamento desses dados, o processo ETL. Após isso, tem-se a análise desses dados, ou seja, qual ação que deve ser tomada pela empresa, além das métricas do produto, se vale a pena ou não continuar vendendo, dentre outros. E claro, por fim tem-se a melhoria desse modelo, seja fazendo testes A e B, experimentação, adição de mais dados para serem analisados.

Conclusões


Foi possível compreender conceitos centrais da Ciência de Dados, começando pela importância de análises preditivas e estatísticas aplicadas, exemplificado pelo estudo de blindagem de aviões na Segunda Guerra Mundial, onde Abraham Wald demonstrou a importância de considerar dados ausentes (aviões que não retornaram).

Destaca-se também a relevância da limpeza e qualidade dos dados para garantir resultados consistentes. A história da Ciência de Dados é explorada, desde suas raízes no Data Mining nos anos 1990 até sua evolução com o poder computacional, permitindo o surgimento de plataformas digitais como YouTube e Facebook.

A ideia principal é entender que o objetivo do cientista de dados é mais do que análise, é a tomada de decisões orientadas a problemas empresariais.

Referencias

 [O que é ciência de dados | Nerdologia Tech](#)

 [What REALLY is Data Science? Told by a Data Scientist](#)