Ciência de dados, o que é e qual a

importância da visualização dos dados?

Muitos ainda se confundem e vinculam a Ciência de dados apenas à programação de algoritmos, à computação ou até mesmo apenas à matemática aplicada. Ciência de dados é a mistura das ciências com a sua base aplicada da informática, computação, matemática e afins. Mais fácil ainda é confundir BI com Ciência de dados.

Bi ou Business Intelligence trabalha com a análise dos dados procurando saber o que houve e tenta explicar através de dados exatos de eventos que já ocorreram, trazendo um diagnóstico do que já aconteceu. Enquanto a Ciência de dados procura descobrir o que pode acontecer, com base em dados não estruturados para prever o futuro, ajudando então na tomada de decisão de um cliente, por exemplo.

Só que os caminhos ainda se parecem, no início. Ambos necessitam analisar a base de dados, procurando realizar a limpeza das informações, o tratamento de possíveis erros ou formatos e a identificação dos pontos de interesses. O que na base lhe é mais importante negocialmente e que poderá ser trabalhado para responder às nossas perguntas.

Em sequência, enquanto o BI trabalha com as perguntas de: O que aconteceu? e Porque isso aconteceu?, a Ciência de dados terá além dessas informações: O que acontecerá? e O que deve ser feito?

Como? Lembra da base que comentei acima? Com o apoio da ciência da computação, matemática e estatística, especialização científica, que seria o apoio dado por outra área, ou outra ciência para complementar o trabalho que será desenvolvido e analisado. E nisso surgem outras agregações como por exemplo o aprendizado de máquina, a análise exploratória de dados, testes de hipótese ou teste de significância e até mesmo a inteligência artificial.

Alimentados pelo crescente aumento de dados gerados ao redor do mundo, temos ainda mais informações sendo geradas desses dados, tornando ainda mais fácil a assertividade nos projetos. Quanto mais dados, mais dados podem ser explorados e mais informações são geradas intuitivamente sobre os dados. E para isso, para que possamos criar um senso intuitivo e transformá-los em código ou matemática, a visualização dos dados, mesmo que em números, tabelas ou imagens nos ajuda a continuar nossas intuições ou mudá-las para melhor seguir procedendo.

Para melhor visualizar os dados existem inúmeras técnicas, mas sem o devido conhecimento, acaba sendo impossível a correta aplicação e extração de informações. Impedindo que as pessoas consigam ver as coisas óbvias que antes, mesmo com o grande volume de dados, não lhes era tão claro. Os padrões acabam sendo vistos de forma rápida, facilitando a amostragem para o público que seja menos técnico.

Após realizar essas etapas conseguimos através da ciência de dados pensar mais na frente e pensar em “O que acontecerá?”. Através da gama de informação que temos em mãos e seus padrões, conseguimos automatizar o aprendizado das máquinas, tornando-as mais inteligentes. Machine learning é uma parte da Inteligência artificial e o que difere elas é exatamente o fator humano. Enquanto IA tenta reproduzir comportamentos humanos, a machine learning busca modificar o seu comportamento com base na sua própria experiência. Possuindo regras lógicas que visam melhorar o seu desempenho ou de uma tarefa. Regras essas que são geradas com base no conhecimento de padrões que já foram analisados.

Tudo isso e muito mais fazem da Ciência de Dados a nova queridinha do momento, criando novos empregos, cargos, funções, áreas e empresas para trabalhar com ela. Não é de se espantar e hoje não torna-se óbvio graças ao crescimento de dados gerados na internet e fora dela. Se o homem um dia já desejou com todas as suas forças conquistar a lua, hoje em dia os dados são o pote de ouro a ser conquistado.