



***Big Data Real-Time Analytics Com Python e Spark 3.0***

# **Big Data Real-Time Analytics Com Python e Spark Versão 3.0**

## **O Que é Feature Extraction?**

A extração de recursos (feature extraction) é um processo de redução de dimensionalidade realizado dentro da Engenharia de Atributos, no qual um conjunto inicial de dados brutos é dividido e reduzido a grupos mais gerenciáveis. Então, quando você quiser processar, será mais fácil.

A característica mais importante de grandes conjuntos de dados é que eles possuem um grande número de variáveis. Essas variáveis exigem muitos recursos de computação para serem processadas. Portanto, a extração de recursos ajuda a obter o melhor recurso desses conjuntos selecionando e combinando variáveis em recursos, reduzindo efetivamente a quantidade de dados e permitindo compreender os dados por diferentes perspectivas. Esses recursos são fáceis de processar, e ainda são capazes de descrever o conjunto de dados real com precisão e originalidade.

Existem diversas maneiras de realizar feature extraction e no Estudo de Caso usaremos uma abordagem baseada na compreensão do problema de negócio.