

# Introdução a Ciência de Dados - 2025

## Sistemas de Informação

Vahid Nikoofard

Universidade do Estado do Rio de Janeiro - Faculdade de Tecnologia  
Laboratório Interdisciplinar Virtual de Inteligência Artificial (LIVIA)

Março 2025



**FAT** Faculdade de  
Tecnologia  
UERJ-Resende

# Objetivo do Curso

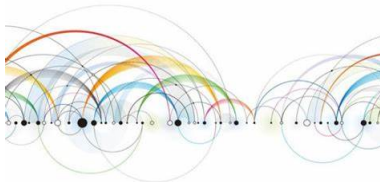
- Enxergar os problemas de negocio via uma perspectiva baseada em dados (Data-Analytic Thinking)
- Entender os princípios e técnicas para extrair **conhecimento** usando dados.

# A Referencia Teórica

*"Um recurso de leitura obrigatória para qualquer pessoa decidida a adotar a oportunidade do Big Data."*  
— Craig Vaughan, Vice-presidente global, SAP

## Data Science *para* Negócios

O que Você Precisa Saber Sobre  
Mineração de Dados e Pensamento  
Analítico de Dados



Foster Provost & Tom Fawcett



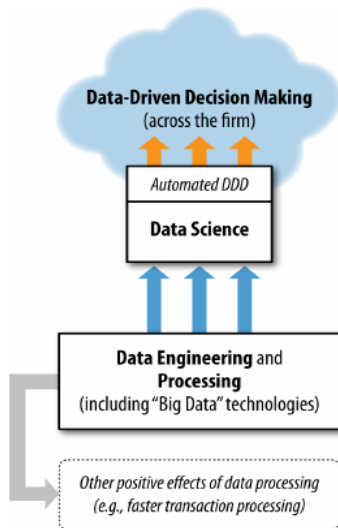
## Exemplo: O Furacão Frances

Em 2004, em Florida nos EUA, foi avisado que o furacão está chegando. A rede WalMart acionou o time de cientistas de dados para estudar os **dados históricos** para construir um modelo preditivo para prever os itens que serão mais demandados nos próximos dias.

## Exemplo: Prevendo a Rotatividade de Cliente (Churn)

O mercado de telefonia celular está muito saturado e a concorrência é alta. As empresas tem uma batalha entre se para atrair os clientes das concorrências. Aí, uma empresa tem que investir em campanhas e propagandas para atrair um cliente de outra empresa. Agora o desafio do time de cientista de dados é, usando os **dados históricos** da empresa, selecionar um grupo de pessoas e montar uma oferta interessante para atrair eles a assinarem um contrato. Este exemplo pode ser aplicado sobre os clientes de bancos também.

# Data Science, Engenharia e Tomada de Decisão



# Tomada de Decisão Orientada em Dados (DDD)

- Basear-se em dados em vez de só a intuição ou experiencia
- Uma empresa é mais produtiva na medida que se basea mais em dados (pesquisa)
- Tipos de decisões:
  - ① usando descobertas via dados
  - ② melhorando a acurácia, especialmente na escala
- O caso da rede Target (concorrente da Walmart) como descoberta
- O Churn como melhorando acurácia (publicidade online, avaliação de credito, detecção de fraude, sistemas de recomendação,...)
- DDD automatizado (industria de credito, sistemas de recomendação)
- Data Science **vs** Data Engineering

# De Big Data 1.0 para Big Data 2.0

- Web 1.0: Páginas estáticas e comércio online
- Web 2.0: Interação com cliente e redes sociais
- Big Data 1.0: Capacidade de armazenamento e processamento de grande quantidade de dados
- Big Data 2.0 (em breve): ????



# Capacidade de Dados como um Ativo Estratégico

- investir para adquirir novos dados
- O caso de Signet Bank in década 90.
- Dados quantitativos no setor bancário e financeiro são difíceis de achar

# Data Analytic Thinking

- Como a data pode resolver um problema de negocio
- Mesmo não trabalhar diretamente com dados é crucial ter essa visão.
- Muitas empresas tem valor porque coletam dados importantes: Meta e Twitter

- Extrair conhecimento útil a partir de dados para resolver problemas de negócios pode ser tratado de forma sistemática, seguindo um processo com etapas razoavelmente bem-definidas. **O Processo Padrão inter-Industrial para Mineração de Dados (CRISP-DM)**
- A partir de uma grande massa de dados, a tecnologia de informação pode ser usada para encontrar atributos descritivos informativos de entidades de interesse.
- Se olhar muito para um conjunto de dados, encontrará alguma coisa! Mas, isto pode não ser generalizada além dos dados para os quais está olhando (overfitting)
- Formular soluções de mineração de dados e avaliar os resultados envolve pensar cuidadosamente sobre o contexto em que serão utilizados.