

Data science aplicada aos negócios

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

O que é e como fazer ciência?

A ciência está baseada na formulação de hipóteses que podem ser testadas para assim fornecerem evidências que corroboram ou desprovam essa hipótese, levando a uma nova hipótese.

Data Science VS Business Intelligence

A diferença está nos modelos que são construídos e testados em Data Science.

- Business Intelligence está interessada em responder perguntas sobre os dados.
- Data Science está interessada em responder perguntas sobre modelos.

Entendendo o problema

Nosso trabalho é responder perguntas mas para fazer as perguntas certas precisamos entender o nosso problema.

- Não adianta resolver o problema errado

Business Intelligence (BI)

A inteligência de negócios (BI) é uma metodologia de coleta, análise e interpretação de dados relacionados ao problema de negócio de um empreendimento.

Geralmente se utiliza muito de visualização de dados, especialmente utilizando “dashboards”

Key Performance Indicators (KPIs)

Métricas e pontos de dados principais para monitorar a integridade de um negócio, departamento ou processos específicos.

- Conectado em bases de dados
- Organizado na forma de *Dashboards*
- Monitoramento em tempo real

Key Performance Indicators (KPIs)

Nesse momento, o *storytelling* é uma ferramenta poderosa na tomada de decisão.

- Observe de maneira crítica o conteúdo de seu relatório para narrativas que podem ter enviesado sua análise.
- Procure determinar o contexto, potenciais conflitos e incorpore-os em sua análise.

Dados

Categorizados

- Escala Nominal
- Escala Ordinal
- Exemplos:
 - Lista de nomes
 - Lista de menções

Quantitativos

- Escala Intervalar
- Exemplos:
 - Lista de salários
 - Preços diários
 - Número de usuários

Dados

Seção transversal

- Todas as observações são realizadas ao mesmo tempo

Série temporal

- Observações realizadas ao longo do tempo

Análise descritiva dos dados

Uma descrição óbvia de um conjunto de dados está nos próprios dados.

- Número de observações
- Valores mínimos e máximos
- Ordenação dos valores

Análise descritiva dos dados

Quando temos muitos dados para analisar cada observação, precisamos de novas ferramentas.

- Tendências centrais
- Dispersão
- Correlação

Análise descritiva dos dados

A análise descritiva utiliza a estatística descritiva para construir conhecimento.

Tendências centrais

- Média
- Mediana

Análise descritiva dos dados

Dispersão

- Amplitude
- Variância
- Desvio-padrão

Análise descritiva dos dados

Correlação

- Covariância
- Correlação

A correlação varia entre -1 (anticorrelação perfeita) e 1 (correlação perfeita)

Bancos de dados e APIs

Bancos de dados e APIs

Dados em geral estão salvos em bancos de dados. Quer você acesse o banco de dados diretamente ou através de uma aplicação, é necessário obter apenas os dados necessários para a análise.

Big Data e Data Lakes

Em geral, projetos de dados envolvem uma quantidade muito grande de dados que nem necessariamente podem ser salvos localmente.

- Consultas a bancos de dados
- Implementação na nuvem

Bancos de dados e ORMs

Bancos de dados em geral são acessados utilizando linguagens de consulta (query) como SQL.

- ORMs são interfaces para bancos de dados
- Funcionam como funções Python

APIs e obtendo dados pela Web

APIs são aplicações Web que permitem o acesso de dados através de pedidos (requests) e respostas (responses).

- Obtém-se o dado geralmente como JSON ou XML

Tratamento e limpeza de dados

Qualidade de dados

Dimensões da qualidade de dados:

- Completude ou integridade
- Conformidade
- Validade
- Acurácia e precisão

Outliers

Outliers são valores encontrados fora de um intervalo razoável com relação às outras observações. É um ponto diferente dos demais, não necessariamente um erro. Sua identificação é importante e pode possuir informações interessantes sobre o sistema.

Vieses

Vieses são desvios sistemáticos na forma de pensar, coletar ou analisar dados, dificultando ou até impossibilitando a compreensão do fenômeno em análise.

- Podem ocorrer em qualquer etapa da análise
 - ~ Coleta/ Tratamento dos dados
 - ~ Análise e visualização



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA