



Aprendizagem Automática em Sistemas Empresariais

PEDRO PEREIRA

AULA 2

Agenda

Enquadramento Teórico

CRISP-DM: Compreensão do Negócio

- Tarefas e outputs esperados
- Exemplo prático

Arranque do projeto

- Registo dos grupos de trabalho
- Instalação das ferramentas



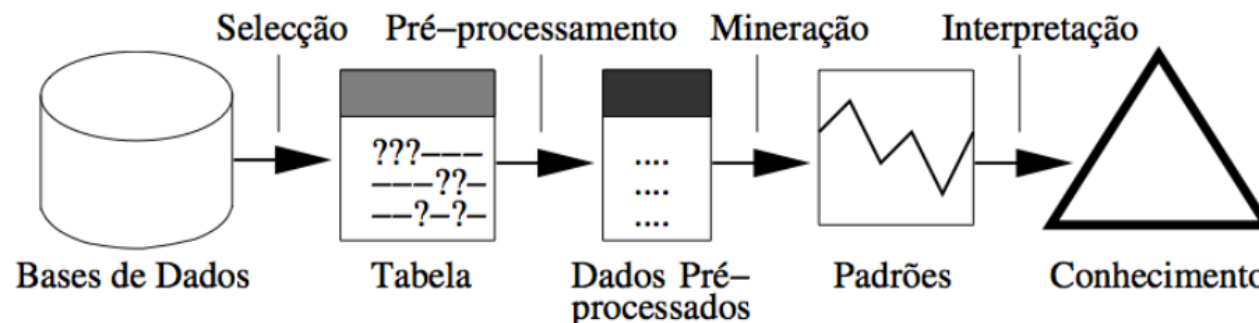
Enquadramento Teórico

Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery from Databases - KDD):

- Na definição original, incluía o termo “Data Mining” como uma das suas etapas.
- Atualmente considerado como um sinónimo de Data Mining.

Data Mining:

- Todo o processo de extração de conhecimento útil a partir de dados em bruto.
- Maior foco no impacto no domínio aplicativo do que em resultados de Machine Learning.
- Assume um método interativo (com analista/decisor) e iterativo (várias fases).
- Grande parte do esforço é gasto na coleta, tratamento e gestão de dados.





Enquadramento Teórico

Sistemas de Apoio à Decisão:

- Área da disciplina de Sistemas de Informação (SI) que se foca em suportar e melhorar a tomada de decisão.

Business Intelligence:

- Termo genérico que inclui arquiteturas, ferramentas, métodos e bases de dados para analisar conjuntos de dados em bruto com o intuito de apoiar decisões.

Data Analytics:

- É definido como a aplicação de sistemas computacionais a grandes conjuntos de dados para apoiar a tomada de decisão.

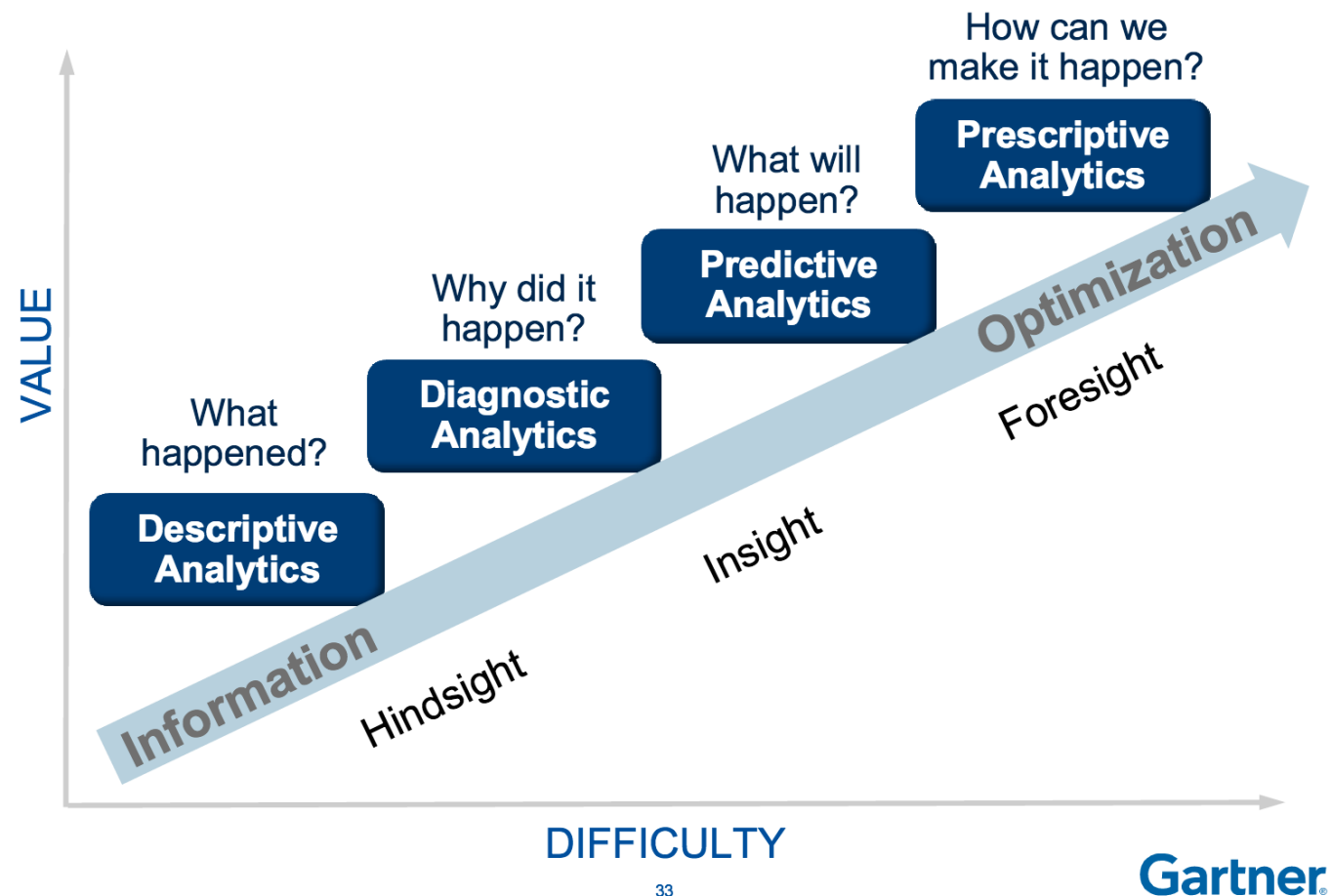
Data Science:

- Campo interdisciplinar relacionado com métodos e processos científicos, e sistemas de extração de conhecimento a partir de dados em várias formas (estruturados e não-estruturados).



Enquadramento Teórico

Gartner Analytic Ascendancy Model

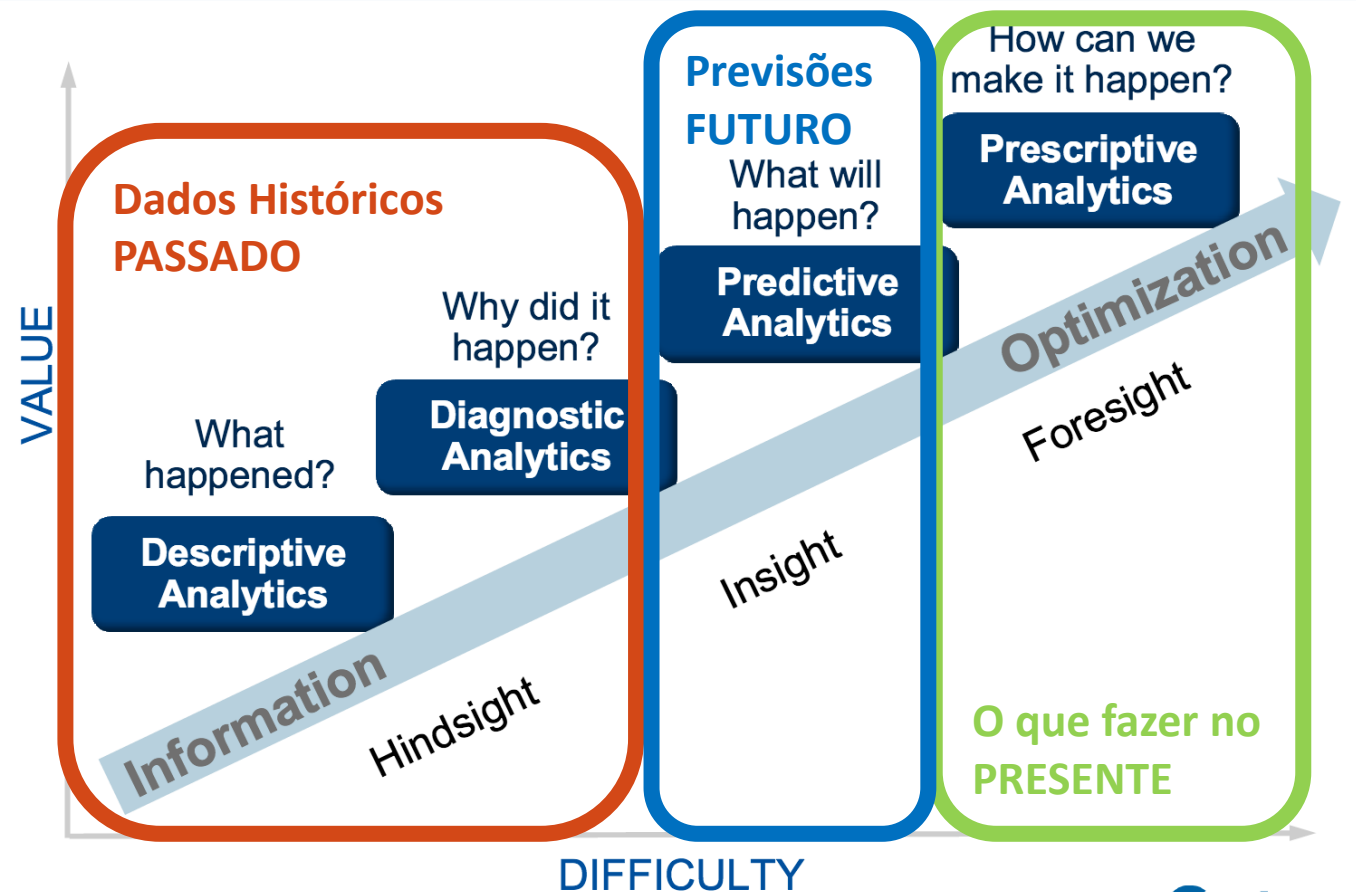


33

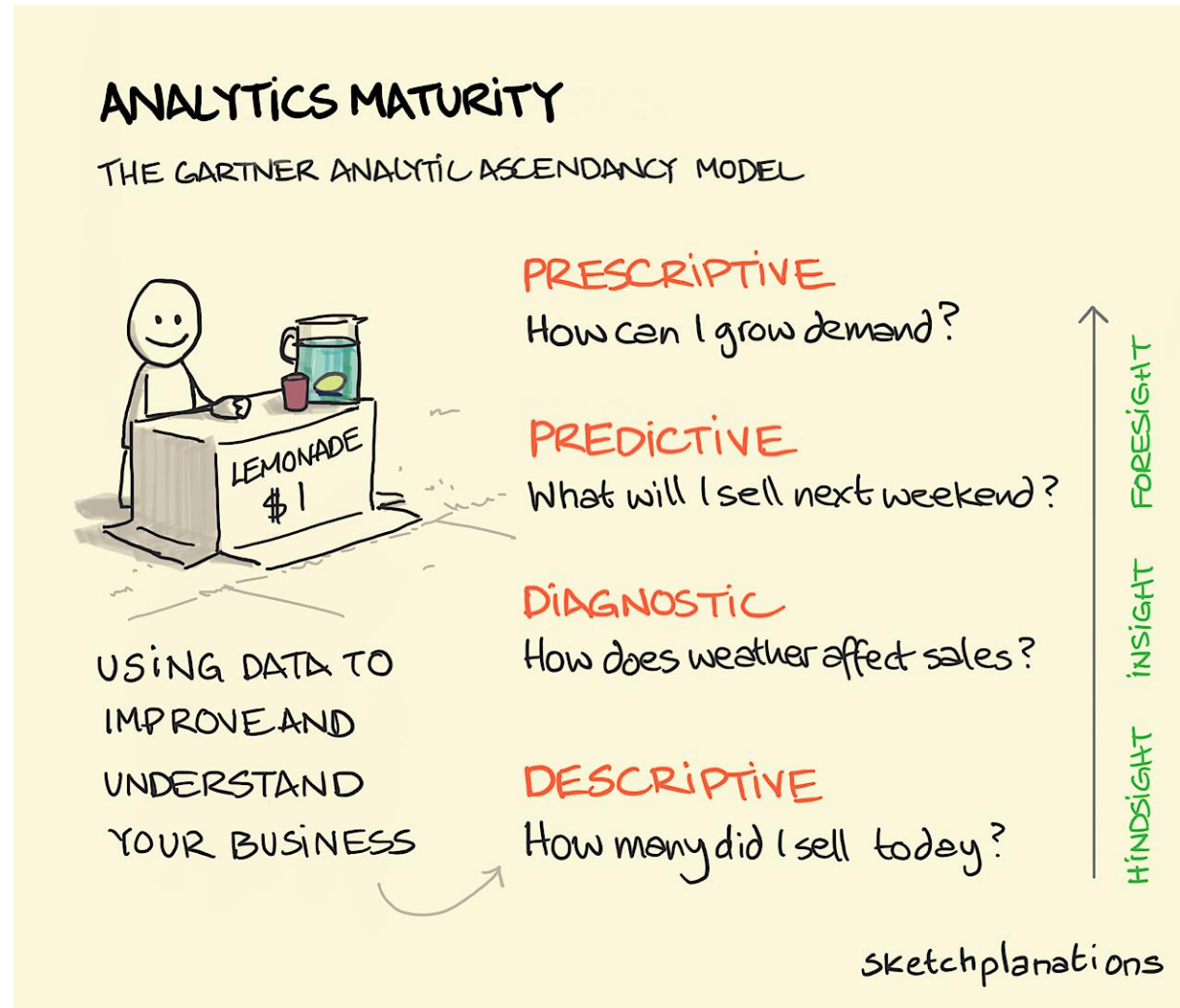


Enquadramento Teórico

Gartner Analytic Ascendancy Model



Enquadramento Teórico





Enquadramento Teórico

Machine Learning:

- Foca-se em conseguir que os computadores/máquinas tomem uma determinada ação sem serem explicitamente programados para tal (ex.: reconhecimento de voz, carros autónomos, pesquisas web,...).

Tipos Machine Learning

Supervised Learning

(Orientado a tarefas; prever o próximo valor)

Classificação (classes, ex.: “aprovado”, “reprovado”)

Regressão (valores numéricos, ex.: preço de produto)

Séries Temporais (sequência de medições, ex.: vendas mensais)

Unsupervised Learning

(Orientado aos dados; identificar semelhanças)

Clustering (agrupar, ex.: perfil de clientes)

Regras de Associação (eventos comuns, ex.: quem compra cereais, também compra leite)

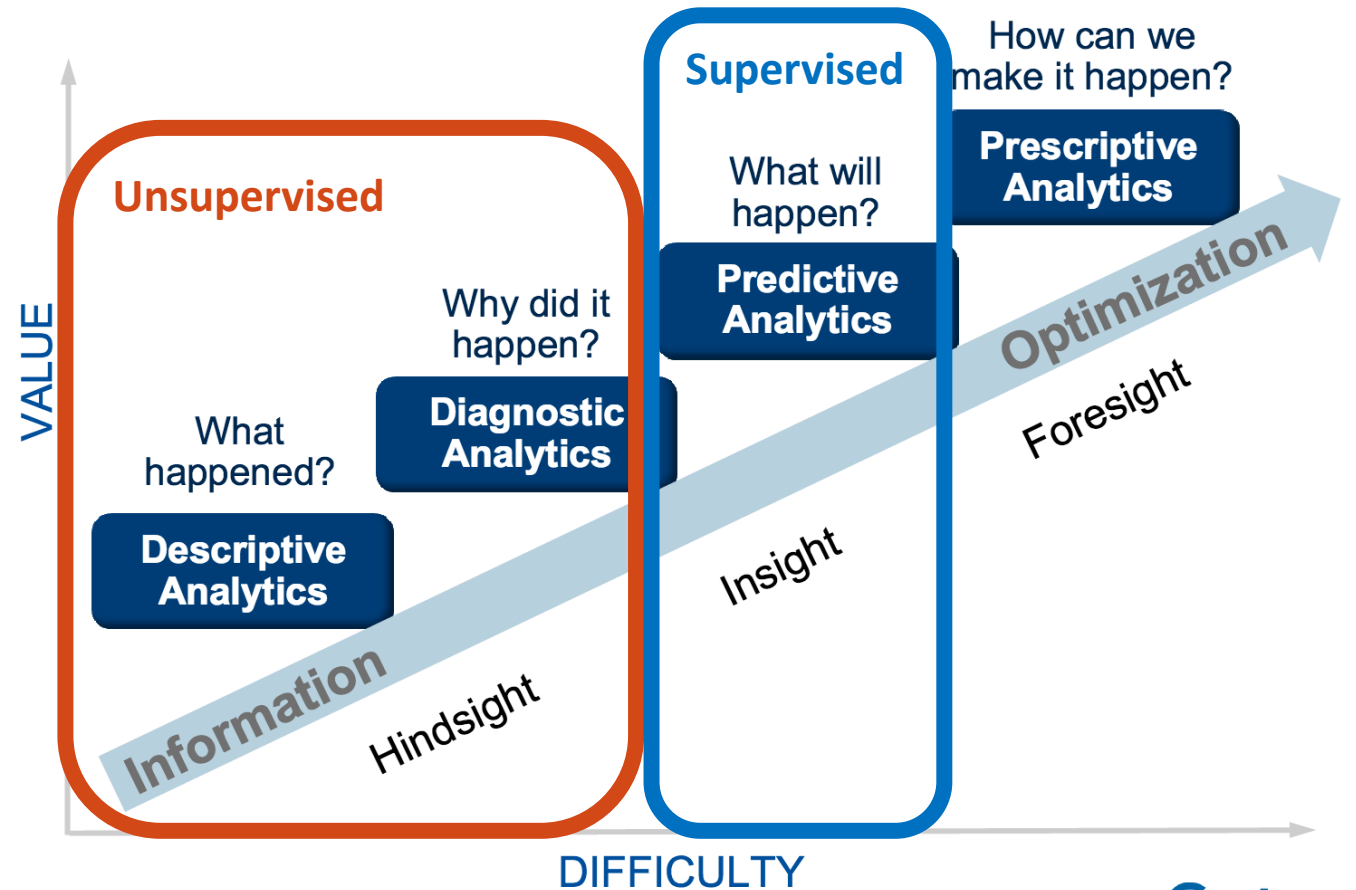
Reinforcement Learning

(Aprender com os erros)

(Ex.: bots em jogos, carros autónomos)

Enquadramento Teórico

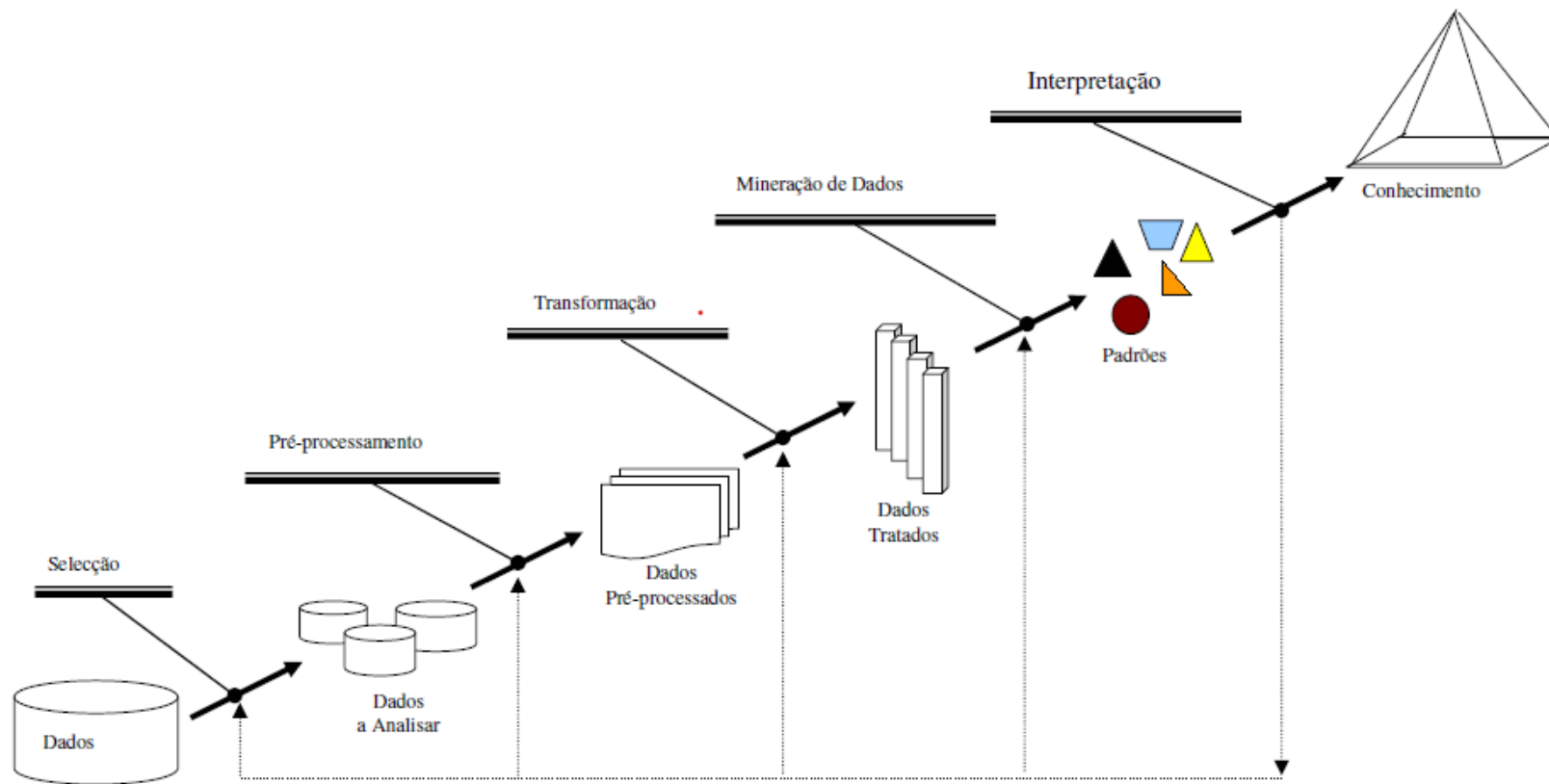
Gartner Analytic Ascendancy Model



33

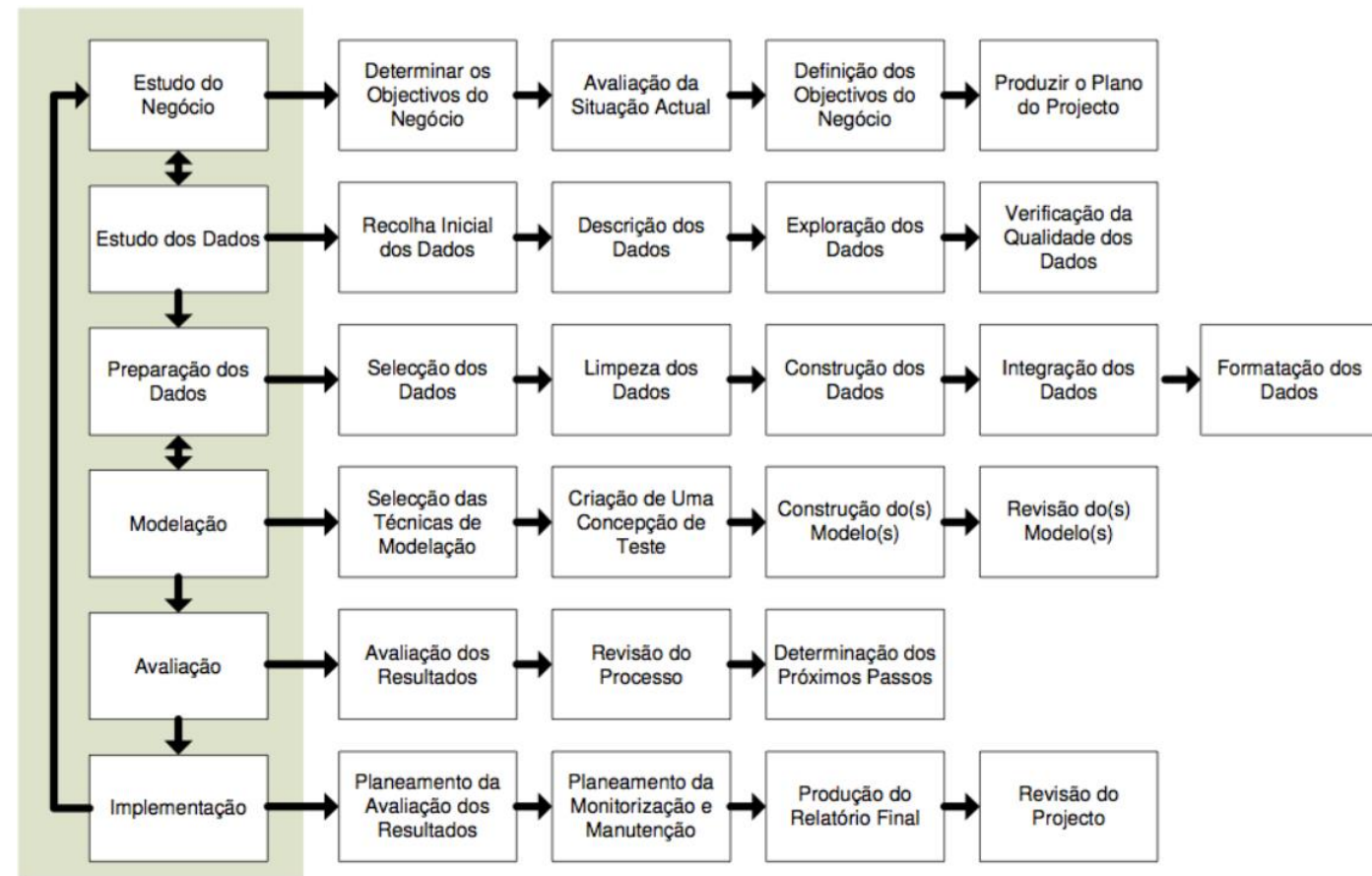
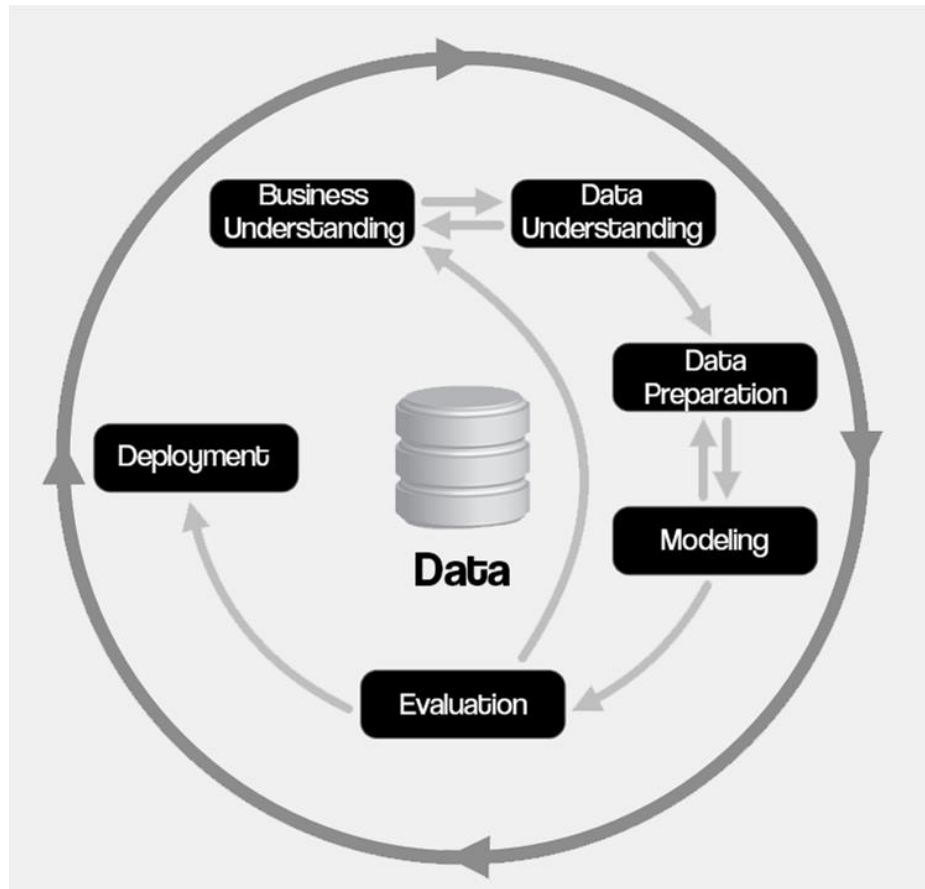
Gartner®

Enquadramento Teórico

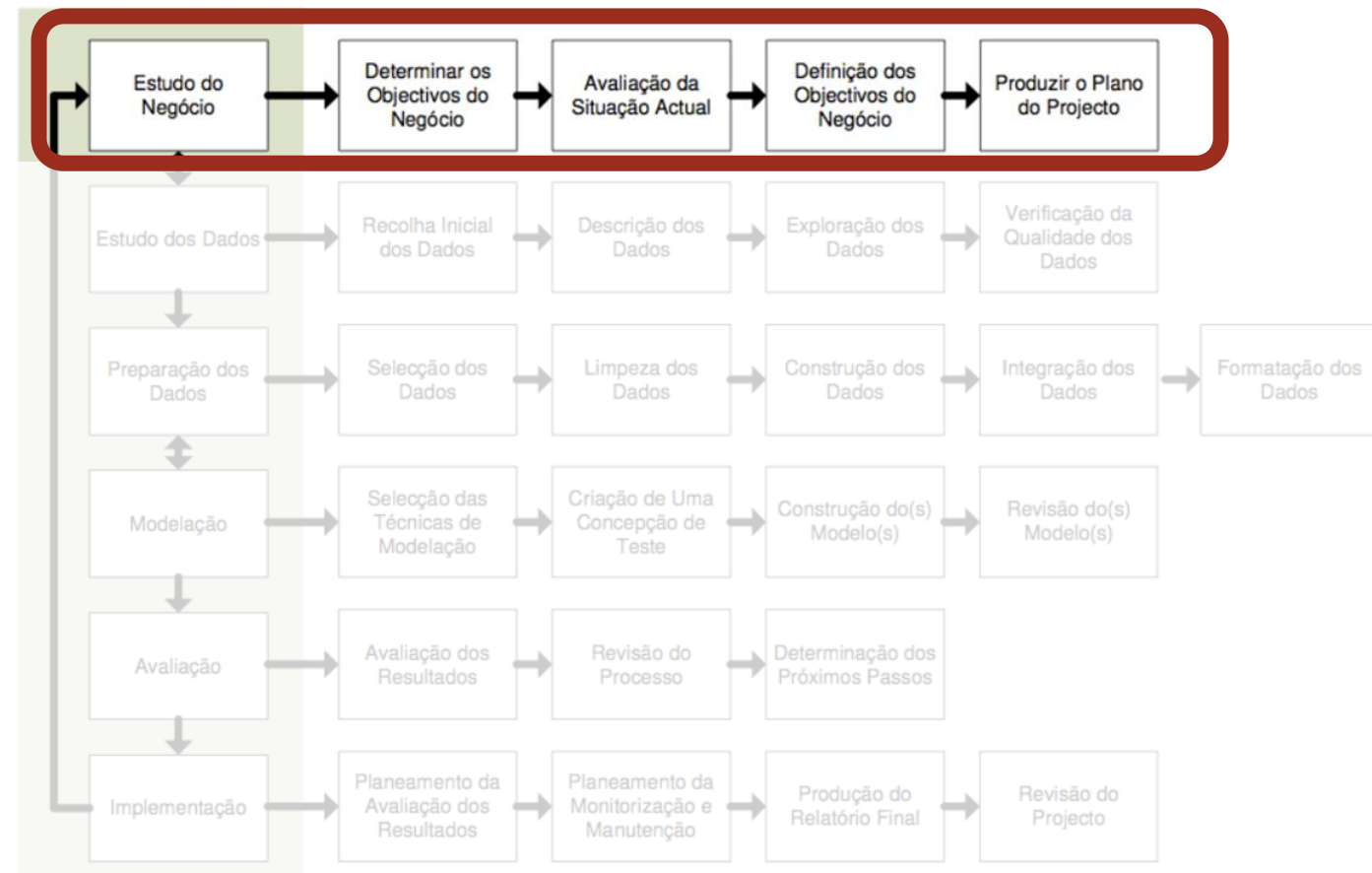
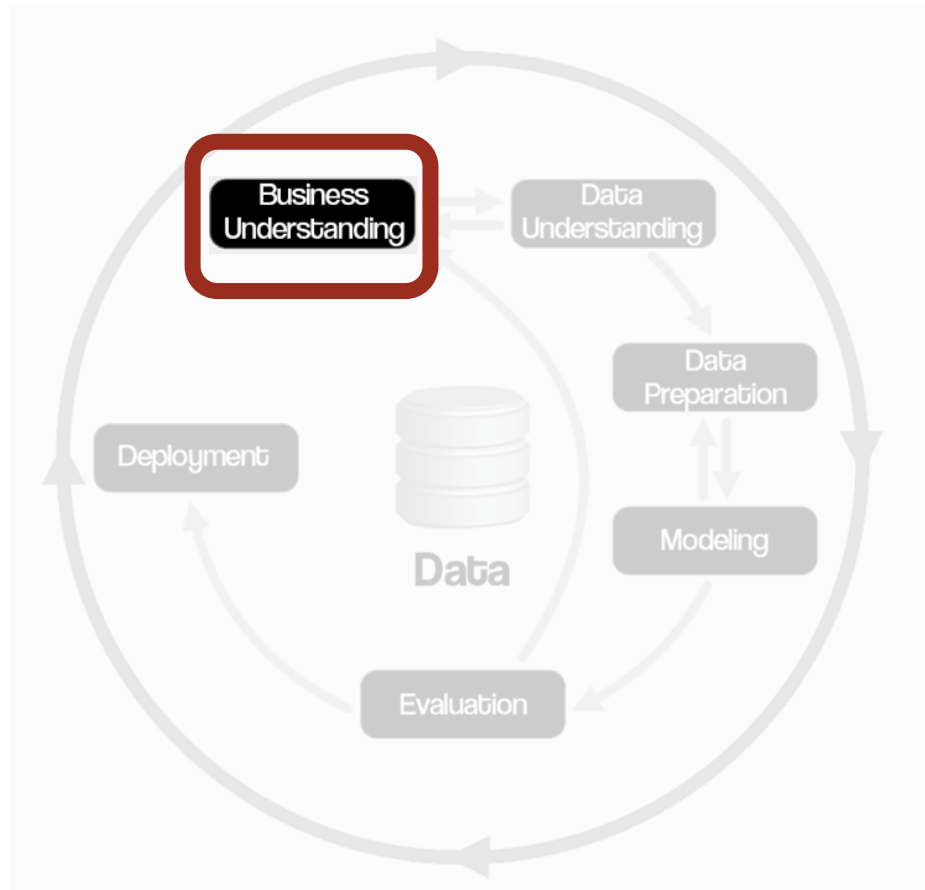


Processo de Extração de Conhecimento

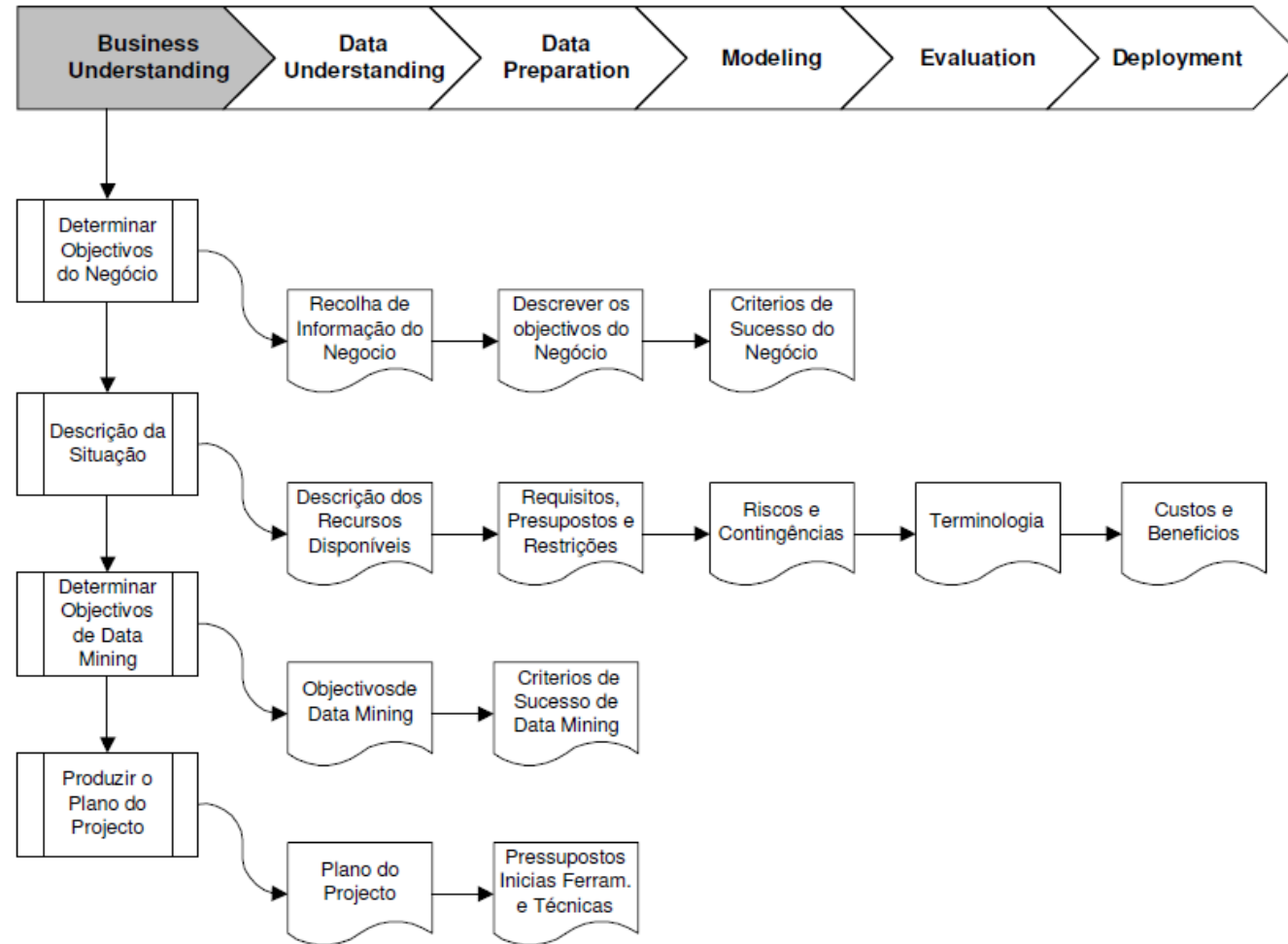
Cross Industry Process for Data Mining (CRISP-DM)



CRISP-DM – Compreensão do Negócio



CRISP-DM – Atividades da Compreensão do Negócio





Determinar objetivos de negócio

Output: **Recolha de Informação do Negócio** (Background)

- Detalhes acerca da situação do negócio no início do projeto para melhor compreender os objetivos de negócio a atingir.

Output: **Objetivos do Negócio**

- Objetivos primários do cliente, tendo em conta uma perspetiva de negócio (ex.: manter os atuais clientes e prever quando é que eles irão mudar para um concorrente).
- Objetivos secundários do negócio (ex.: determinar se uma redução nos preços irá afetar a intenção dos clientes mudarem para um concorrente).

Output: **CrITÉrios de Sucesso de Negócio**

- Descrever os critérios que levarão o projeto a ter sucesso de um ponto de vista do negócio.
- Podem ser específicos (ex.: reduzir a saída de clientes para 5%) ou gerais e subjetivos (ex.: encontrar sinais úteis sobre satisfação dos clientes).



Avaliação da Situação Atual

Output: **Inventário dos recursos**

- Pessoal (técnicos de negócio, de dados, de suporte, de *Data Mining*, ...).
- Dados (extrações pontuais, acesso a um *Data Warehouse*, ...).
- Computacionais (*hardware* e *software*).

Output: **Requisitos, pressupostos e restrições**

- Timings, compreensibilidade e qualidade de resultados, segurança e aspetos legais.
- Descrição de pressupostos e restrições acerca dos dados e do negócio.

Output: **Riscos e contingências**

- Riscos ou eventos que podem comprometer o projeto e respetivos planos de contingência.

Output: **Terminologia** (glossário)

- Termos específicos (do negócio e do *data mining*).

Output: **Custos e benefícios**

- Análise dos custos do projeto vs. benefícios para o negócio.



Definição dos Objetivos do Data Mining

Output: **Objetivos do *Data Mining***

- Descrever os outputs esperados do projeto que levarão a atingir determinados objetivos de negócio.
- Exemplo:
 - Objetivo do negócio: Aumentar as vendas por catálogo para os clientes existentes.
 - Objetivo de *data mining*: Prever que quantidade de um determinado produto vai ser comprada por um determinado cliente, tendo em conta as suas informações (idade, salário, ...) e as compras nos últimos três anos.

Output: **CrITÉrios de sucesso do *Data Mining***

- Determinar critérios para um projeto bem sucedido, em termos técnicos (ex.: determinado nível de precisão na previsão ou um erro médio).



Produzir o Plano do Projeto

Output: **Plano do Projeto**

- Lista de fases do projeto, duração, recursos requeridos, entradas, saídas e dependências.
- Se possível, relacionar com os riscos do projeto.
- O planeamento deve ser dinâmico (revisões intermédias de progresso e conquistas alcançadas).

Output: **Pressuposto Inicial das Ferramentas e Técnicas**

- Selecionar as técnicas e ferramentas de Data Mining que serão utilizadas nas diferentes fases do projeto.



Exemplo: organização de telecomunicações

Background:

- Elevada taxa de rotatividade de clientes (saída para a concorrência): 10%.

Objetivos de Negócio e Critérios de Sucesso:

- Reter clientes pela diminuição da taxa de rotatividade (clientes a saírem para a concorrência $< 5\%$).

Inventário de Recursos:

- 1 especialista do negócio, 3 *Data Scientists*, Servidores de BD e DM, Despesas, ...

Requisitos, Assunções e Restrições:

- O modelo de DM deve ser legível e não utilizar informações sensíveis (ex.: morada de cliente).

Riscos e Contingências:

- Apoio financeiro depende dos resultados demonstrados em fases intermédias.

Terminologia:

- Glossário sobre DM, glossário sobre o negócio e operações de telecomunicações.



Exemplo: organização de telecomunicações

Custos e Benefícios:

- Custo do projeto: 50 K €. Receita de 100 € por cliente que não saiu => reter 500 clientes.

Objetivos de Data Mining:

- Tarefa de classificação binária.
- Modelo de classificação para determinar se um cliente fica ou sai para a concorrência.

Critérios de Sucesso de Data Mining:

- Acerto nas previsões superior a 90%.

Plano de Projeto:

- Fases, entregas, diagrama de Gantt, ...

Pressuposto Inicial das Ferramentas e Técnicas:

- Tableau e Python 3.8 disponível e instalado nos computadores e servidores utilizados.



Aprendizagem Automática em Sistemas Empresariais

PEDRO PEREIRA

AULA 2