

Metodologias de Aplicação de IA

Prof. Dr. Igor da Penha Natal

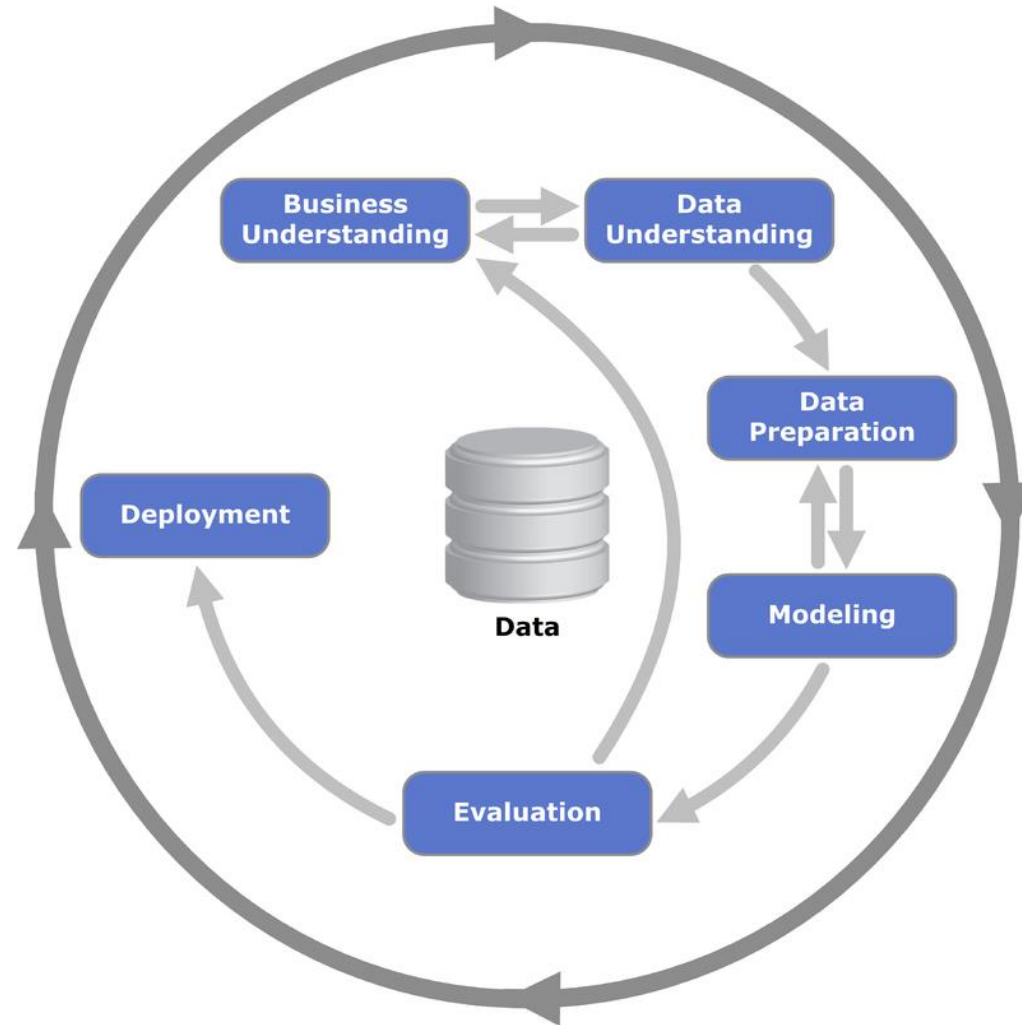
Disciplina: Inteligência Artificial

Metodologia de IA

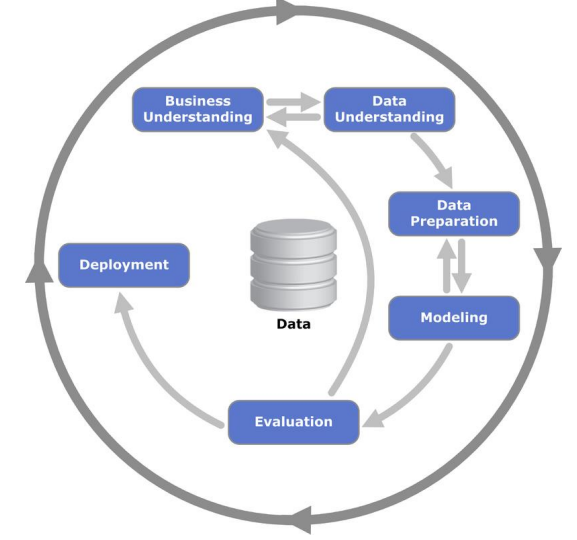
As metodologias de IA são fundamentais para a estruturação e execução eficaz de projetos que envolvem análise de dados e aprendizado de máquina. Elas fornecem um conjunto de diretrizes e melhores práticas que auxiliam os profissionais a transformar dados brutos em conhecimento útil e acionável.

- Três exemplos que estudaremos:
 - CRISP-DM (Cross-Industry Standard Process for Data Mining);
 - KDD (Knowledge Discovery in Databases);
 - SEMMA (Sample, Explore, Modify, Model, Assess).

CRISP-DM



CRISP-DM



1. Business Understanding:

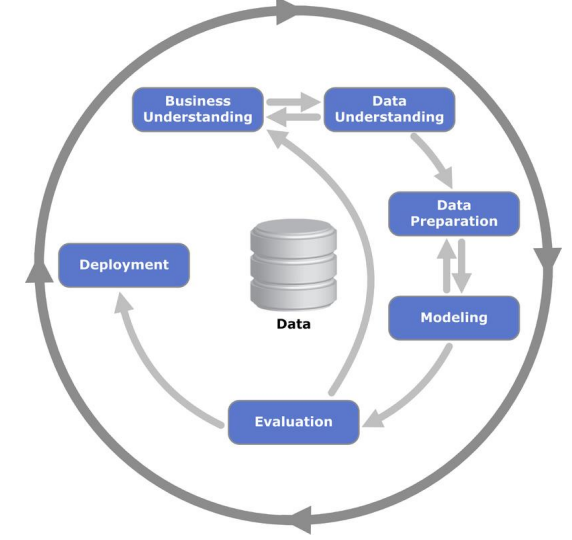
- Identificar o problema.
- Background: Explicar a situação da empresa e como o projeto vai ser direcionado para solucionar o problema.
- Objetivo do projeto: Informar qual o objetivo maior do projeto.
- Critério de sucesso: Deixar bem claro qual será a métrica que ditará se o projeto atingiu o sucesso ou não.



CRISP-DM

2. Data Understanding:

- Coletar dados.
- Descrever os dados.
- Explorar os dados.
- Verifica a qualidade dos dados.

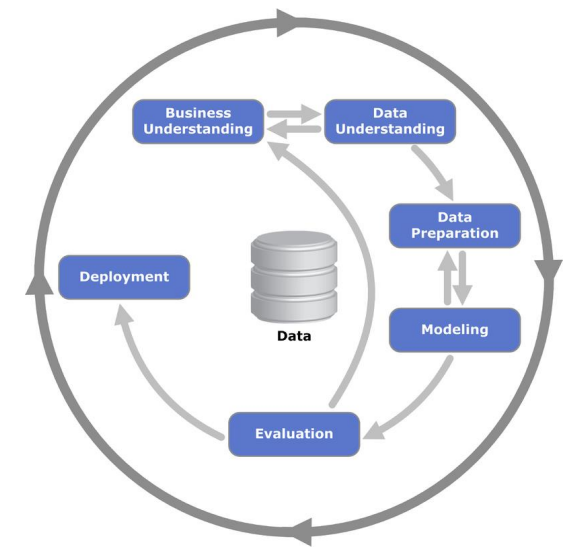


DATA

CRISP-DM

3. Data Preparation

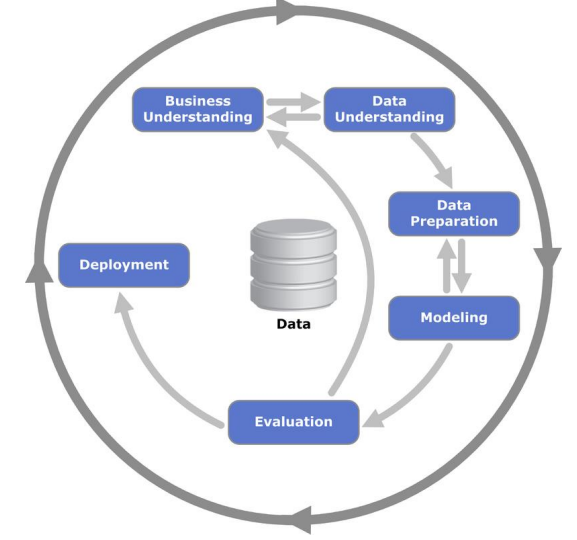
- Data Selection: Selecionar os dados para ser utilizado no modelo.
- Data Cleaning: Tratar dados em formato incorreto, remover colunas da tabela, tratar outliers, etc.
- Construct Data: Enriquecimento de dados, criação de dados derivados, etc.
- Integrating Data: Necessária para juntar diversas fontes de dados diferentes.



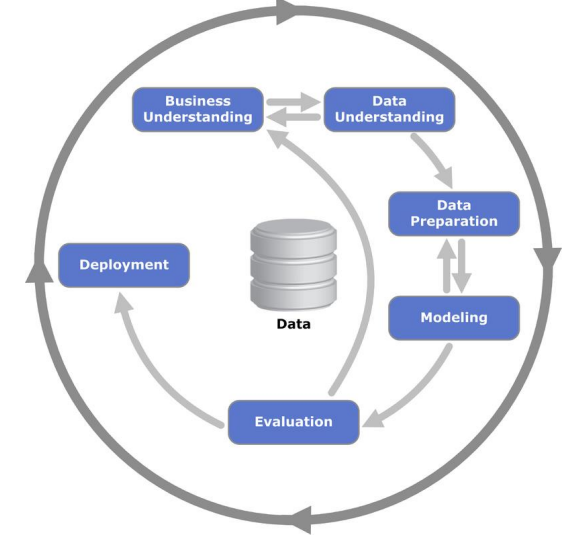
CRISP-DM

4. Modeling

- Realizar a construção do modelo.
- Escolha do algoritmo.
- Configuração dos parâmetros.
- O ideal é testar diversos algoritmos com diversos parâmetros.



CRISP-DM



5. Evaluation

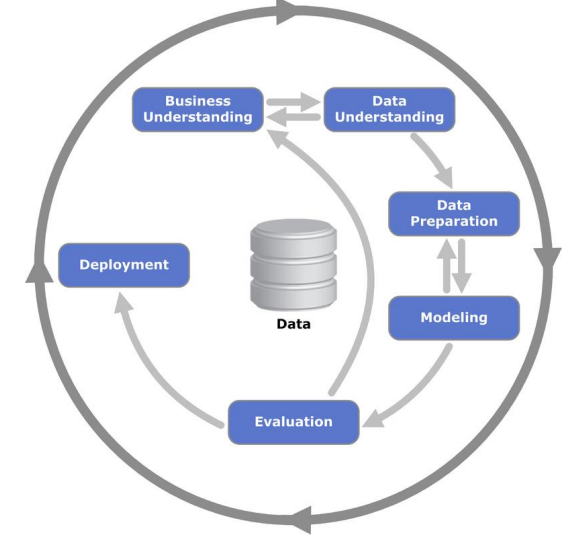
- Avaliar os resultados.
- Utiliza como base os critérios de sucesso definidos na etapa 1.
- Se não ficar dentro do esperado, o recomendado é voltar e analisar o que pode ser melhorado em cada etapa, caso não possa, rever os objetivos e critérios de sucesso e adaptá-los.
- Caso esteja dentro do esperado, o projeto pode seguir para o passo 6.



CRISP-DM

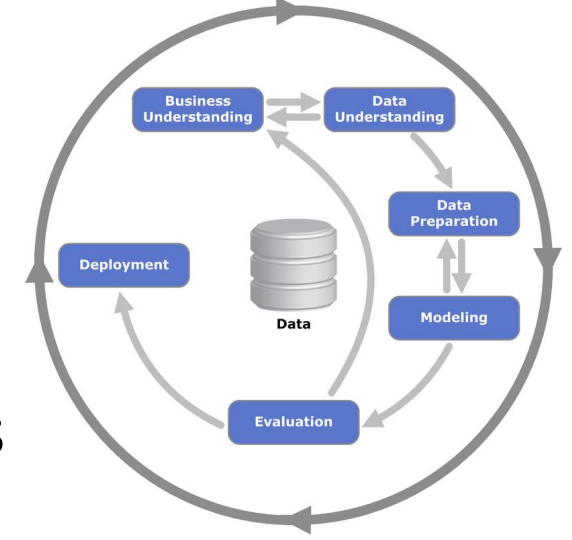
6. Deployment

- Colocar o modelo em ambiente de produção.
- O modelo sendo utilizado em um ambiente real.
- Dependendo da situação, o modelo deve ser avaliado constantemente para que adaptações possam ser feitas.

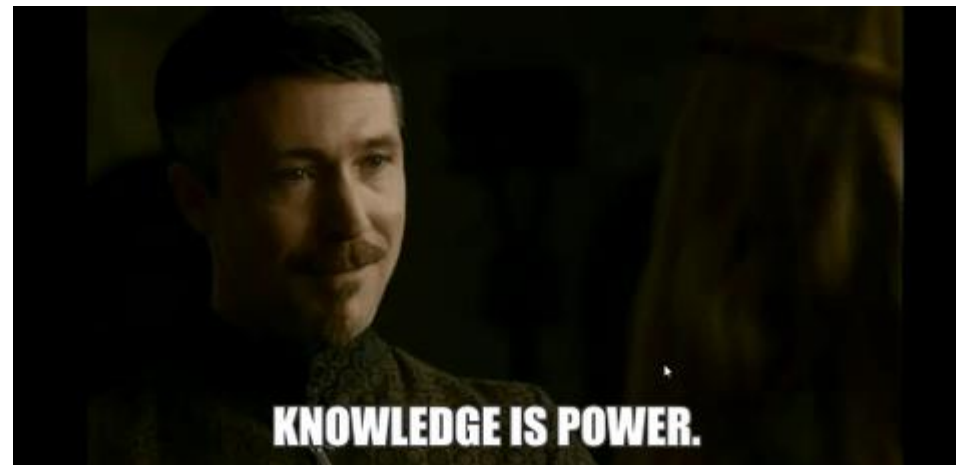
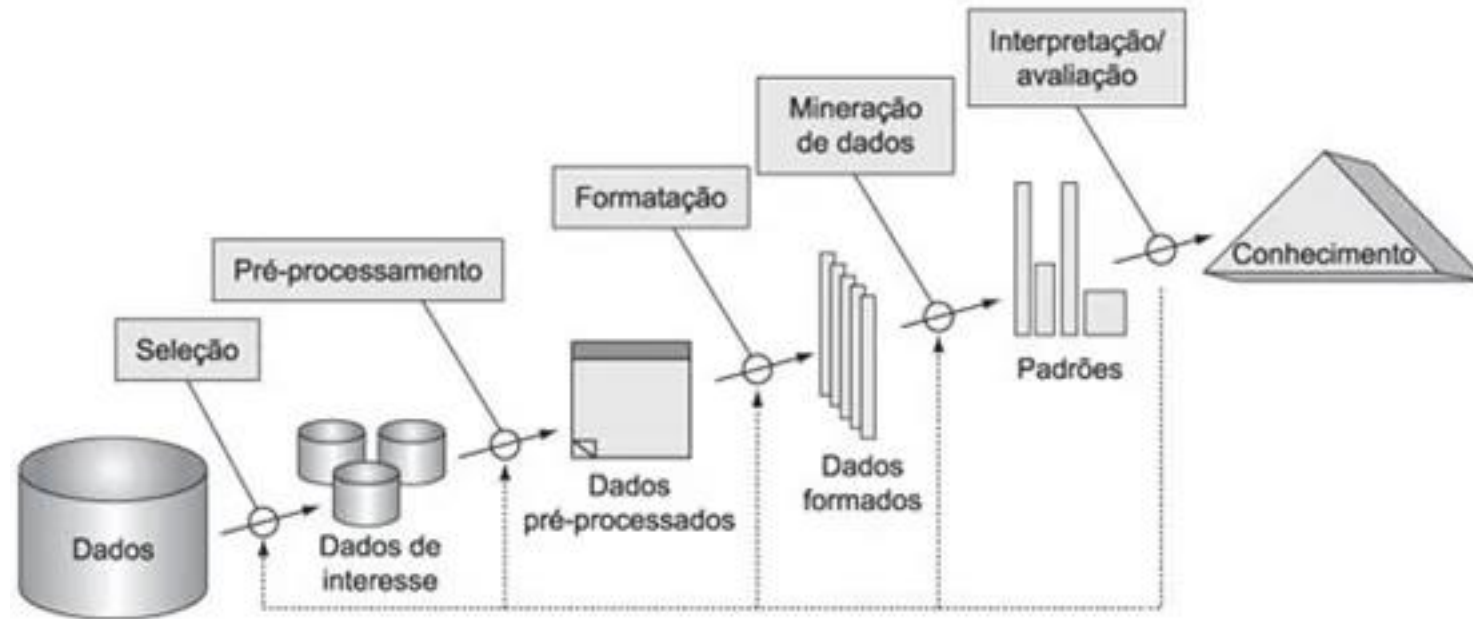


CRISP-DM

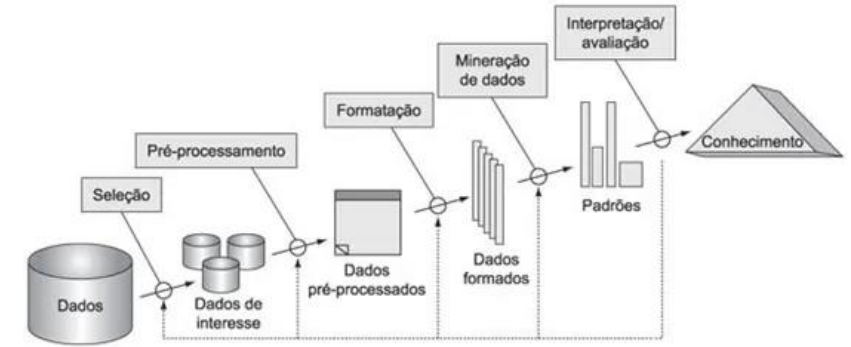
O CRISP-DM é uma metodologia padrão para projetos de mineração de dados, composta por seis fases: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implementação. Ela é amplamente utilizada por sua flexibilidade e aplicabilidade em diferentes setores, proporcionando um processo estruturado para transformar dados em insights acionáveis.



KDD



KDD



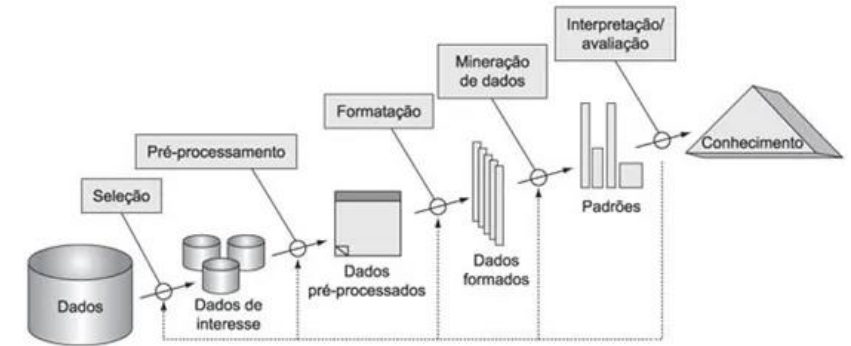
1. Seleção:

- Selecionar um conjunto ou subconjunto de dados que farão parte da análise.
- As fontes de dados podem ser variadas: planilhas, relatórios, etc.
- Dados com formatos diferentes: estruturados, semiestruturados e não estruturados.

2. Pré-processamento:

- Verificação da qualidade dos dados armazenados.
- Limpar, corrigir ou remover dados inconsistentes.
- Verificar dados ausentes ou incompletos.
- Identificar anomalias (outliers).

KDD



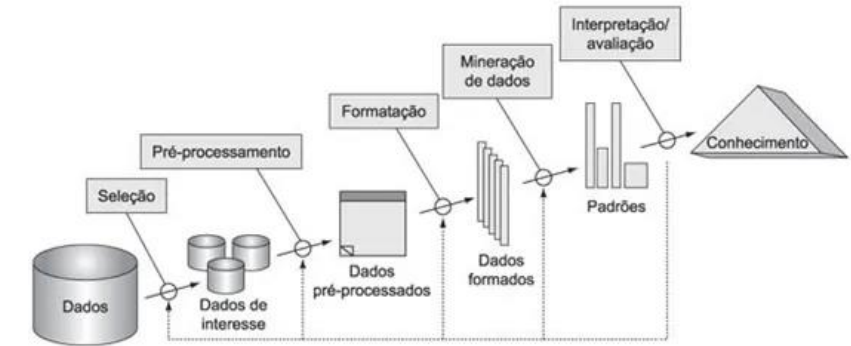
3. Formatação:

- Aplicar técnicas de transformação de dados.
- Normalização, agregação, criação de novos atributos, redução e sintetização dos dados.
- Os dados ficam disponíveis agrupados e formatados em uma única base de dados.

4. Mineração de Dados:

- Construir modelos ou aplicar técnicas de mineração de dados.
- Objetivos: Verificar uma hipótese e/ou descobrir novos padrões de forma autônoma.
- Descoberta pode ser preditiva ou descritiva.
- Essa etapa pode ser repetida inúmeras vezes.

KDD

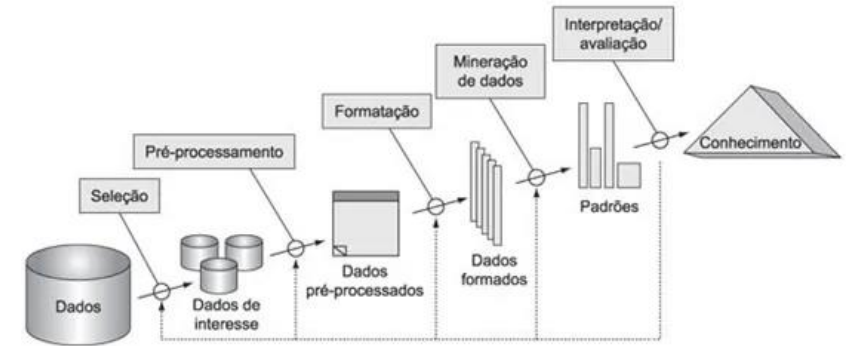


5. Interpretação / Avaliação:

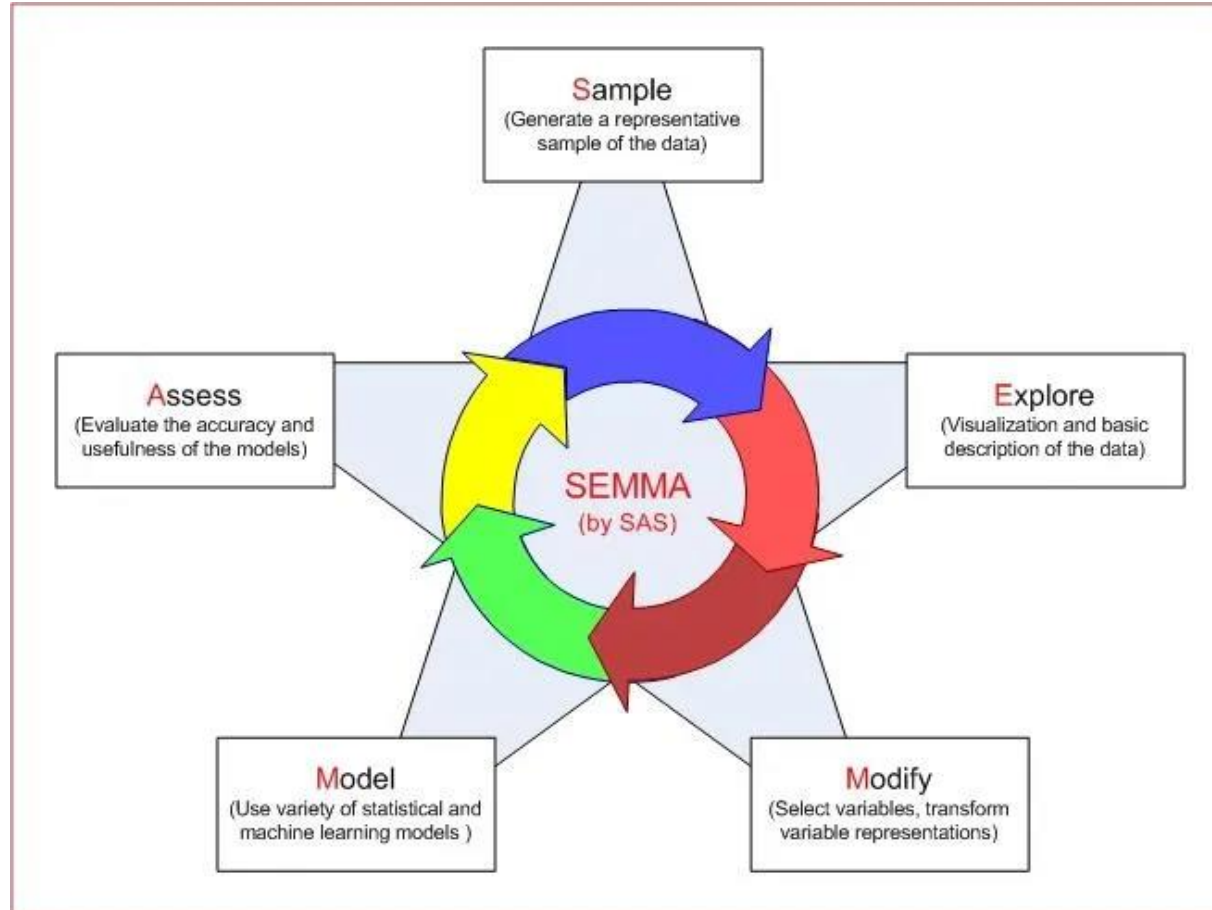
- Avaliar o desempenho do modelo:
 - Utilizar medidas estatísticas.
 - Avaliação dos profissionais de negócio.
 - Aplicação em cima de dados de teste.

KDD

O KDD (Knowledge Discovery in Databases) é um processo de descoberta de conhecimento em grandes volumes de dados, composto por várias etapas: seleção, pré-processamento, transformação, mineração de dados e interpretação/avaliação. Seu objetivo é extrair padrões úteis e conhecimentos significativos a partir de dados, permitindo uma melhor tomada de decisão e a geração de novos insights em diversos contextos.



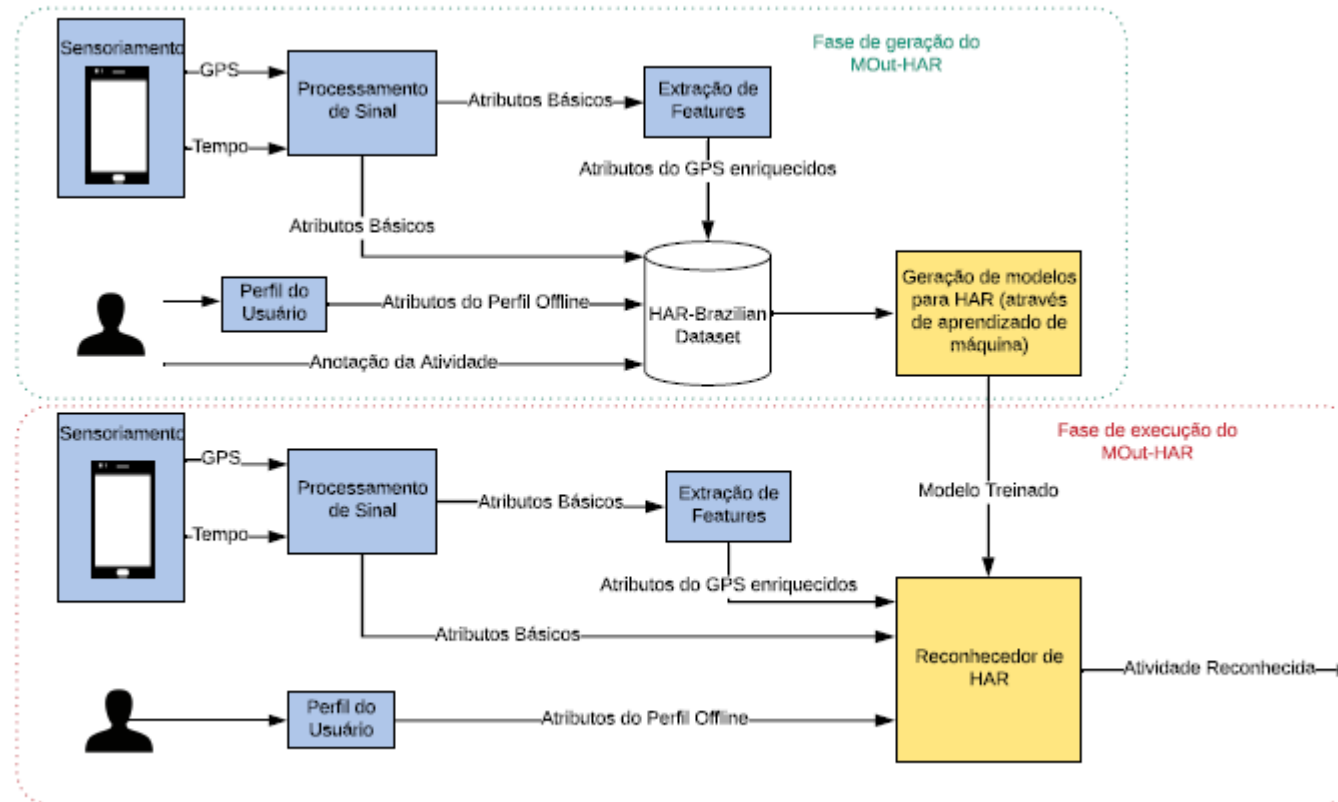
SEMMA



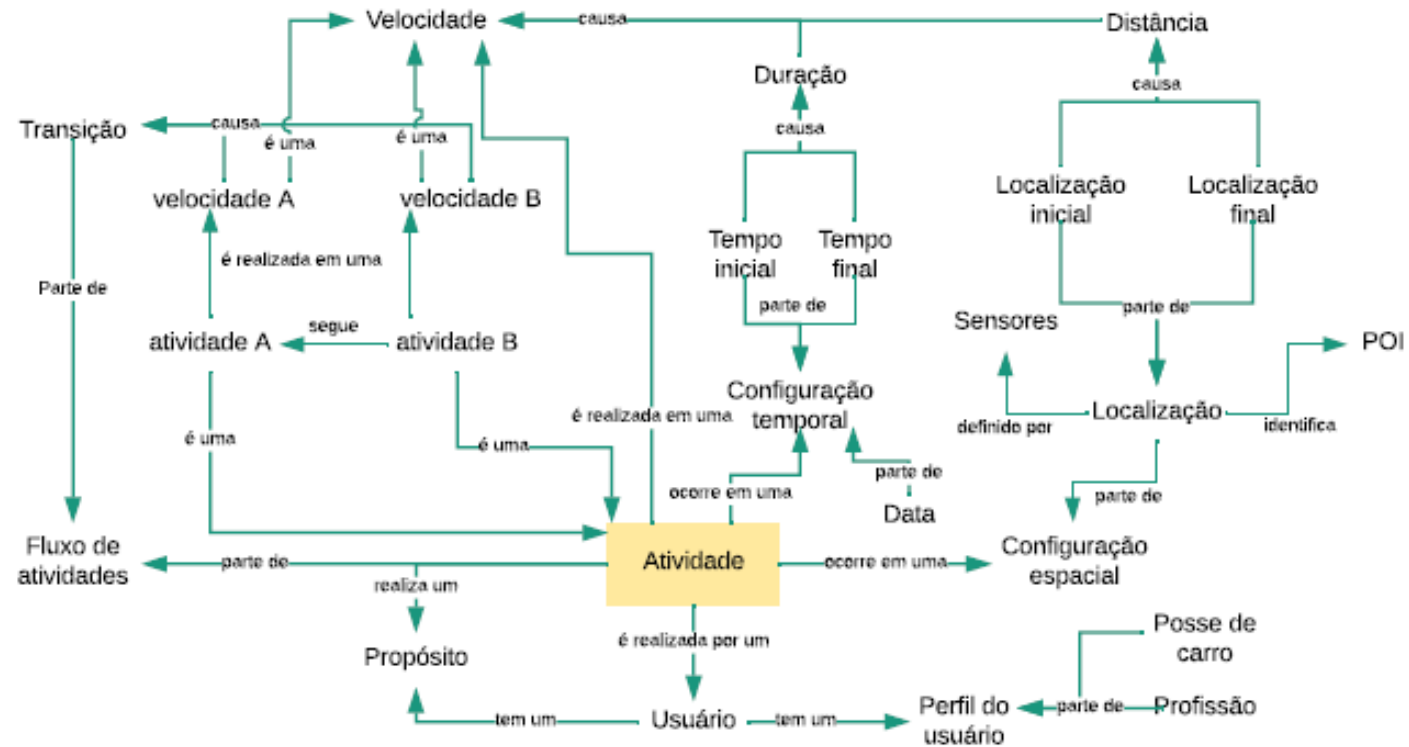
Diferenças entre as metodologias

- CRISP-DM: Flexibilidade e na ênfase no alinhamento com os objetivos de negócios, sendo aplicável em diversos setores.
- KDD: Descoberta de conhecimento em grandes volumes de dados. Ênfase na extração de padrões e insights a partir de dados complexos.
- SEMMA: Preparação detalhada dos dados antes da modelagem, garantindo a qualidade e a precisão dos modelos gerados.

Exemplo na prática



Exemplo na prática



Atividade

- No github da disciplina foram disponibilizados 3 artigos, é fortemente recomendável a leitura deles.
- Cada grupo vai escolher um artigo e fará um breve resumo sobre ele.
- O resumo terá um caráter pessoal, deve apresentar o que você entendeu e o que destaca como importante.