

# Machine Learning 1: Conceitos Primários

## GUIA DA AULA 4



# Metodologia CRISP – DM



Acompanhe aqui  
os temas que  
serão tratados  
na videoaula

● O que é uma metodologia?

● Metodologia CRISP-DM



## A Importância da Metodologia

*Imagine:*

- *a quantidade de empresas aplicando projetos de ciência de dados nesse momento;*
- *a quantidade de modelos estão sendo desenvolvidos;*

*Agora, imagine que cada cientista de dados tem seguido uma ordem de projeto. Muitos podem até ter se esquecido de etapas importantes durante esse processo.*



# A METODOLOGIA

*Afinal o que é metodologia e porque ela nos ajuda na reflexão anterior?*

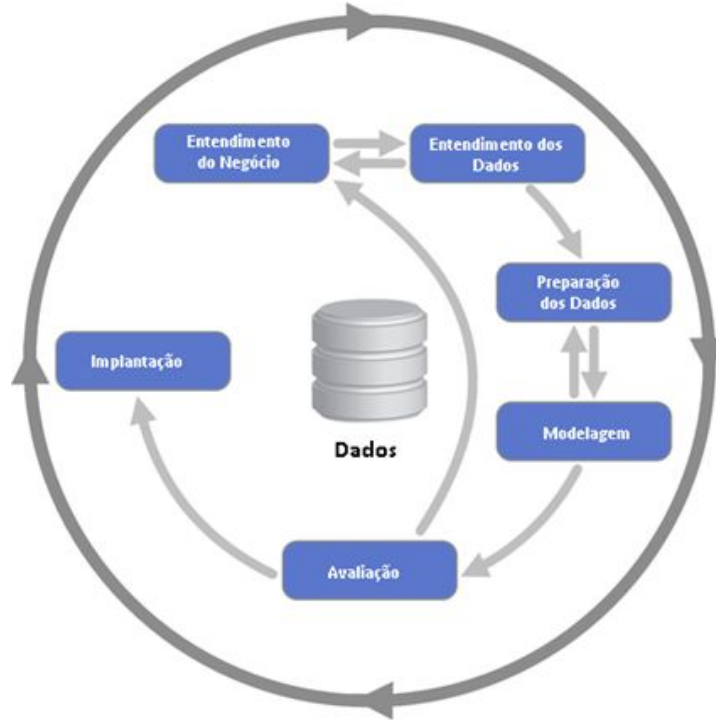
- **O que é?** Um método utilizado com um conjunto de procedimentos padronizados.
- **Organização:** A metodologia nos apresenta um plano estruturado para conduzir um projeto.
- **Reprodutibilidade:** Permite que outros possam reproduzir o mesmo projeto.
- **Controle de Qualidade:** Garante que as etapas não serão esquecidas garantindo a qualidade do projeto.



# CRISP-DM

*“Processo padrão de Data Mining”*

**O que é?** Um método muito utilizado no mercado de trabalho de dados, baseado no método científico.



# ETAPA 1 – Entendimento e Definição de Objetivos



## Entendimento do Objetivo do negócio

Nosso primeiro passo consiste em determinarmos com clarezas o objetivo principal do projeto, mapeado pela necessidade do negócio.



## Avaliação dos Objetivos

Em seguida, devemos entender se os objetivos são tangíveis e entender se os resultados esperados podem ser alcançados através de um projeto de Ciência de Dados.



## Planejar o Projeto

O último passo da etapa 1 é construirmos o planejamento do projeto e entendermos qual tipo de modelagem ou análise será necessária, assim como as ferramentas.

## ETAPA 2 – Preparação dos Dados



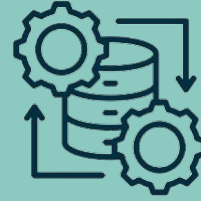
### Buscar fonte dos Dados

Entender se você já tem disponível todos os dados que necessita ou se necessitará de dados externos.



### Exploração dos Dados

- Explorar seus dados, conhecer eles a fundo, utilizar gráficos, filtros;
- Garantir que os dados que você tem são os necessários e serão suficientes.



### Tratamento dos Dados

Verificar se temos campos faltantes, campos com mau preenchimento ou qualquer fator que possa prejudicar a execução do projeto.

## ETAPA 3 – Modelagem



### **Escolha do(s) Modelo(s)**

Escolhemos quais modelos serão aplicados, se aplicamos apenas um ou mais a fim de compararmos resultados.



### **Aplicação da Modelagem**

Nesse passo, realizamos toda construção técnica do modelo.



## ETAPA 4 – Avaliação



### Avaliação de Métricas

- Cada modelo tem suas técnicas de avaliação para conseguirmos identificar se seus resultados são satisfatórios ou não.
- Nessa etapa, aplicamos diversos métodos a fim de validar o modelo.

## COMO APLICAR CRISP-DM AO CURSO?



- Ao longo do curso você notará que, em cada projeto de modelagem apresentado nas aulas, a ordem do CRISP-DM será sempre respeitada;
- Vocês serão encorajados a realizar seus projetos seguindo essas ordens, explorando dados, tratando as bases, modelando e avaliando os modelos;
- Ao final do curso, esse fluxo já estará enraizado de tal forma que será completamente natural seguir essas etapas para a construção de qualquer projeto.



# Bons estudos!

