

Processo Padrão Inter-Indústrias para Mineração de Dados

# CRISP-DM

Cross Industry Standard  
Process for Data Mining

Guia para consulta  
rápida

 /jonasaacampos



# 01

## Entendimento de Negócios

1. Delimitação do escopo
  - a. Background do projeto (mapa geral do projeto)
  - b. Objetivos de negócio
  - c. Critérios de sucesso
2. Avaliação da Situação
  - a. Inventário de recursos
  - b. Requisitos, premissas, restrições
  - c. Riscos e contingências
  - d. Terminologias
  - e. Custos e benefícios
3. Determinar objetivos da mineração de dados
  - a. Definir meta de mineração (critério técnicos do objetivo de negócio)
  - b. Critérios de sucesso de mineração (use um benchmark de mercado)
4. Produzir um plano de projeto
  - a. Plano de projeto
  - b. Avaliação inicial de ferramentas e técnicas



# Etapas mais importantes

- mais demoradas
- mais tensas
- mais técnicas
- mais onerosas



02

Entendimento dos dados

03

Preparação dos dados



## 02

# Entendimento dos dados

1. Coleta inicial dos dados
  - a. Relatório inicial dos dados (de onde os dados vieram, como foram coletados, quais as dificuldades da coleta)
2. Descrição dos dados
  - a. Relatório descritivo dos dados
3. Exploração dos dados
  - a. Relatório exploratório dos dados
4. Verificação da Qualidade dos dados
  - a. Relatório de qualidade (problemas e possíveis soluções)

## Preparação dos dados



## 03

1. Dataset e Data Description
  - a. Selecionar dados (quais atributos, features (colunas) e registros(linhas) fazem sentido). Escrever um relatório que justifique a inclusão/exclusão dos dados.
  - b. Limpar dados (fazer relatório de limpeza)
  - c. Construir data (técnicas para criação de novas features)
  - d. Integrar dados
  - e. Formatar dados (não muda o significado, mas facilitm o entedimento e a modelagem)

1. Definição do modelo
2. Documentar técnicas utilizadas
3. Definir desenhos de teste
4. Ajuste dos Hiperparâmetros
5. Avaliação do Modelo (sumarização do modelo)



1. Avaliar se os objetivos do negócio forma atingidos
  - a. Interpretar os resultados
  - b. Verificar se existem novos objetivos de negócio
  - c. Listar recomendações para projetos futuros de mineração de dados
2. Analisar o fluxo de mineração de dados (Results = Modelos + Descobertas Pessoais)
  - a. foi necessário? Está Otimizado? Poderia ser melhor?
  - b. Identificar falhas e etapas erradas
  - c. Identificar caminhos inesperados
  - d. Identificar tarefas perdidas e/ou que deveriam ser refeitas
3. Avaliar o projeto e visão de negócios
  - a. Analisar potencial de implementação
  - b. Estimar possíveis melhorias do processo
  - c. Recomendar continuações e alternativas



# Deployment

1. Usar o github
2. Testes Unitários
3. Testes de integração
4. Periodicidade de execução
5. Orquestração (útil para cenários complexos, mas concentre-se no seu MVP.)
6. DAG (Grafo Direcionado Acíclico)
7. Reflita sobre a experiência do usuário
8. Definir ambiente com base
9. Cenário
10. Complexidade
11. Orçamento
12. Métricas de Ouro
13. Latência
14. Tráfego
15. Erros
16. Saturação (CPU, Memória, etc.)

07

# Manutenção

1. Programada
2. Não Programada





GOSTOU?  
CURTA,  
COMENTE,  
COMPARTILHE!



**GitHub/jonasaacamos**



**/jonasaacamos**