

A Mineração de Dados é uma das tecnologias mais promissoras da atualidade. Um dos fatores deste sucesso é o fato de dezenas, e muitas vezes centenas de milhões de reais serem gastos pelas companhias na coleta dos dados e, no entanto, nenhuma informação útil é identificada. Além da iniciativa privada, o setor público e o terceiro setor (ONGt's) também podem se beneficiar com a Mineração de Dados.

A seguir são apresentadas algumas das áreas nas quais a Mineração de Dados é aplicada de forma satisfatória:

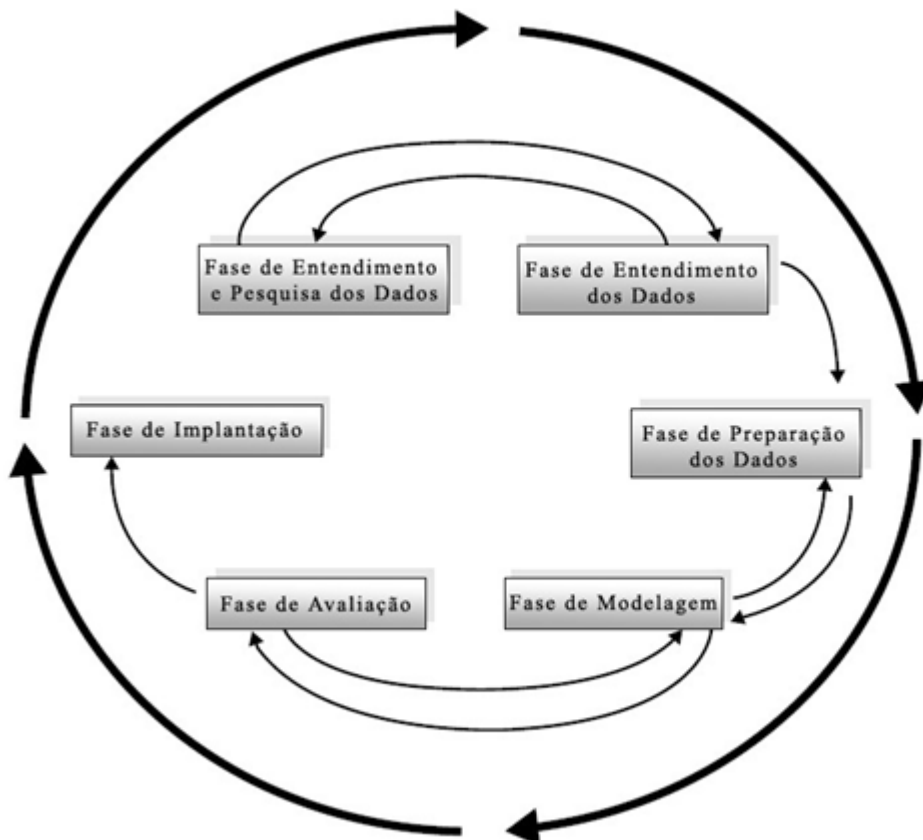
- I. Retenção de clientes: identificação de perfis para determinados produtos, venda cruzada;
- II. Bancos: identificar padrões para auxiliar no gerenciamento de relacionamento com o cliente;
- III. Cartão de Crédito: identificar segmentos de mercado, identificar padrões de rotatividade;
- IV. Matemática: aproximação de equações não-lineares;
- V. Tomada de Decisão: filtrar as informações relevantes, fornece indicadores de probabilidade.

Quais das afirmações acima estão corretas?

A I, II, III e V.

Atualmente diversos processos definem e padronizam as fases e atividades da Mineração de Dados. Apesar das particularidades, todos em geral contém a mesma estrutura. O CRISP-DM (Cross-Industry Standard Process of Data Mining) é um modelo com vasta literatura disponível e pode atualmente ser considerado o padrão de maior aceitação.

O processo CRISP-DM consiste de seis fases organizadas de maneira cíclica, conforme mostra a figura abaixo. Além disso, apesar de ser composto por fases, o fluxo não é unidirecional, podendo ir e voltar entre as fases.



As fases do processo CRISP-DM são:

- I. **Entendimento dos Negócios:** Nessa etapa, o foco é entender qual o objetivo que se deseja atingir com a mineração de dados. O entendimento do negócio irá ajudar nas próximas etapas.

II. **Entendimento dos Dados:** As fontes fornecedoras dos dados podem vir de diversos locais e possuírem diversos formatos. Após definir os objetivos, é necessário conhecer os dados.

III. **Preparação dos Dados:** Devido às diversas origens possíveis, é comum que os dados não estejam preparados para que os métodos de Mineração de Dados sejam aplicados diretamente. Dependendo da qualidade desses dados, algumas ações podem ser necessárias. Este processo de limpeza dos dados geralmente envolve filtrar, combinar e preencher valores vazios.

IV. **Modelagem:** É nesta fase que as técnicas (algoritmos) de mineração serão aplicadas. A escolha da(s) técnica(s) depende dos objetivos desejados.

V. **Avaliação:** Considerada uma fase crítica do processo de mineração, nesta etapa é necessária a participação de especialistas nos dados, conhecedores do negócio e tomadores de decisão. Diversas ferramentas gráficas são utilizadas para a visualização e análise dos resultados (modelos).

VI. **Distribuição:** Após executado o modelo com os dados reais e completos não é necessário que os envolvidos conheçam os resultados.

Quais das afirmações acima estão corretas?

B I, II, III, IV e V.

A Mineração de Dados é comumente classificada pela sua capacidade em realizar determinadas tarefas. As tarefas mais comuns são:

I. Descrição

II. Classificação

III. Estimação ou Regressão

IV. Predição

V. Agrupamento

VI. Distribuição.

Quais das tarefas acima estão corretas?

B I, II, III, IV e V.

Apesar da grande potencialidade oferecida pela Mineração de Dados, alguns fatores devem ser analisados. A seguir discute-se como alguns desses fatores podem prejudicar as técnicas de mineração:

I. a técnica de programação utilizada deve ser bem definida.

II. As relações entre os atributos precisam ser muito bem definidas, caso contrário os resultados podem ser mal interpretados;

III. Permitir que o processo de treinamento execute por muito tempo, até que se consiga obter indícios que possam levar a conclusões factíveis;

IV. Gerar subsídios para uma conclusão errada tornando-a mais plausível. Porém, uma interpretação falha pode disfarçar as falhas nos dados;

V. Usar um grande número de variáveis.

Quais das afirmações acima estão corretas?

B II, III, IV e V.