

PDF - Data Mining, Resumo - 2.

KDD

Knowledge Discovery in Databases, ou Descobrimento de conhecimento em banco de dados é o processo de descoberta de conhecimento útil a partir de uma coleção de dados.

Padrão de Dados

Utilizando uma preparação de dados adequada torna possível extrair informações, auxiliando nas tomadas de decisões e solução de problemas.

- Preparar melhor os dados
- Entender os limites de dados

Principais atividades:

1. Procura de Dados
2. Caracterização de Dados
3. Montagem do Conjunto de Dados

Dataset

Coleção de objetos e seus atributos.

Atributo é uma propriedade ou característica de um objeto.

Coleção de atributos = objeto

Melhoramento de Dados

Enriquecimento - Processo de inserir novas fontes de dados ao conjunto dados.

Melhoramento - Realçar características de dados sem adição de fontes externas.

Caracterização dos Dados

Granularidade - Dados detalhados é preferível a dados agregados

Consistência e Inconsistência - Diferentes coisas representadas pelo mesmo nome/Mesma coisa representada por diferentes nomes.

Poluição - Espaços em branco. Campos incompletos, inexatos, inconsistentes ou não identificáveis.

Relações - Observar a consistência das instâncias dos objetos da estrutura do problema.

Integridade - Observar a integridade das relações avaliadas.

Duplicações ou Redundâncias - Ocorre quando instâncias dependem de diferentes fluxos de dados. Exigem maior esforço computacional.

Lidando com Valores Faltantes

São 6 métodos:

1. Ignorar
2. Preencher manualmente
3. Usar média
4. Usar moda
5. Usar média na mesma classe
6. Valor provável por inferência

Ruído

Dados incoerentes

Base de Dados Desbalanceada

Objetos (instâncias) pertencentes a uma classe, em menor número em relação a outra.



[Material_Apoio_MIN_Parte2.pdf](#)
479.8 KB