

# Exercicio III - Data Mining

Matheus Barbosa Souza, 4º Período

March 2019

## 1 Qual o conceito de data mining ou mineração de dados?

Processo de explorar grandes quantidades de dados à procura de padrões consistentes. [2]

## 2 Cite e descreva duas aplicações práticas de data mining.

### Educação

Inspiradas pelo programa mundial OLPC (One Laptop per Child) proposto pelo Laboratório de Mídias do MIT, as iniciativas UCA (Projeto e Programa "Um Computador por Aluno") vêm promovendo a introdução de laptops de baixo custo em centenas de escolas brasileiras. Em ambas as iniciativas, gestores e docentes brasileiros carecem de informações sobre a efetiva contribuição do uso de computadores nas escolas. Diante deste cenário, participamos do desenvolvimento de um sistema de informação chamado MEMORE para tentar suprir tal carência. Este sistema permite a coleta e a mineração de dados sistemática (tarefa de Descoberta de Associações) de informações sobre como os laptops UCA têm sido pedagogicamente utilizados e os efeitos decorrentes dessa utilização no aprendizado dos alunos. Participante das iniciativas UCA desde o início de sua aplicação, a Secretaria Municipal de Educação de Pirai, no estado do Rio de Janeiro, vem utilizando, com sucesso, o MEMORE em caráter experimental em suas escolas.

### Energia

Um projeto importante na área energética (sobretudo em épocas de poucas chuvas) teve como objetivo aplicar técnicas de Data Mining para geração de modelos que façam a previsão de demanda de consumo de energia elétrica por regiões (tarefa de Regressão/Previsão de Séries Temporais). Para tanto, foram utilizados registros de consumo de energia elétrica ao longo de períodos anteriores. [3]

### 3 Qual a diferença entre dado, informação e conhecimento?

#### **Data**

Dados são quaisquer fatos, números ou textos que podem ser processados por um computador. Hoje, as organizações estão acumulando vastas e quantidades crescentes de dados em diferentes formatos e diferentes bases de dados. Estes incluem:

- dados operacionais ou transacionais, como vendas, custo, estoque, folha de pagamento e contabilidade
- dados não operacionais, como vendas do setor, previsão dados e dados macroeconômicos
- metadados - dados sobre os dados em si, como a lógica design de banco de dados ou definições de dicionário de dados

#### **Informação**

Os padrões, associações ou relacionamentos entre todos esses dados pode fornecer informações. Por exemplo, a análise do ponto de venda Os dados de transação de venda podem gerar informações sobre quais produtos são vendendo e quando.

#### **Conhecimento**

A informação pode ser convertida em conhecimento sobre padrões e tendências futuras. Por exemplo, informações resumidas nas vendas de supermercados de varejo pode ser analisada à luz do esforços promocionais para fornecer conhecimento da compra do consumidor comportamento. Assim, um fabricante ou varejista poderia determinar quais itens são mais suscetíveis a esforços promocionais. [4]

### 4 O que é KDD - Descoberta de Conhecimento em Bancos de Dados?

Descoberta de conhecimento em bancos de dados (KDD) é o processo de descoberta de conhecimento útil a partir de uma coleção de dados. Essa técnica de mineração de dados amplamente usada é um processo que inclui preparação e seleção de dados, limpeza de dados, incorporação de conhecimento prévio em conjuntos de dados e interpretação de soluções precisas a partir dos resultados observados. [8]

## 5 Explique e exemplifique os dois tipos de conhecimento estudados (explícito e tácito)

Conhecimento explícito é o conhecimento que possuímos e de que temos consciência. É o conhecimento que somos capazes de documentar, é o conhecimento que as organizações conseguem armazenar.

O conhecimento tácito, ao contrário, é conhecimento que temos mas do qual não nos apercebemos. É conhecimento que adquirimos através da prática, da experiência, dos erros e dos sucessos. É conhecimento que não somos capazes de descrever nem documentar. É conhecimento que as organizações não podem utilizar depois fora das horas de trabalho dos seus colaboradores. [6]

## 6 Qual a importância da fase de preparação dos dados? Quais os principais problemas envolvendo coleta de dados do mundo real?

Uma preparação de dados adequada torna possível extrair informações, auxiliando nas tomadas de decisões e nas soluções de problemas; Qualidade das informações disponíveis no mundo real. [1]

## 7 Cite 1 exemplo de uma coleção de atributos e 3 objetos dessa coleção.

Dados Nominais: dados nominais estão em formato alfabético e não em um inteiro.[7]

|                   |   |
|-------------------|---|
| Dados Categóricos | Docente, Professor Assistente, Professor          |
| Estado            | novo, pendente, trabalhando, completo, acabamento |
| Cores             | preto, marrom, branco, vermelho                   |

Table 1: Dados Nominais

## 8 Conceitue o processo de enriquecimento dos dados e cite um exemplo diferente dos que constam no material.

É o processo de inserir dados de fontes externas ao conjunto de dados.

Imagine pode ter em sua base, dados demográficos, sócio-econômicos, sociais e comportamentais de todos seus clientes. Saber o poder de consumo, seus

gostos, hábitos e afinidades nas redes sociais, seu comportamento e interação no seu site e qual a afinidade de cada cliente com seu produto. Isso é enriquecimento e inteligência de dados. [5]

## **9 Conceitue o processo de melhoramento dos dados e cite um exemplo diferente dos que constam no material**

É o processo de realçar características dos dados sem adição de fontes externas  
Conexão de um roteador wi-fi apenas pelo modelo não é suficiente para saber se é boa ou suficiente. Sendo assim, verifico a distância e quantas pessoas conseguem usar sem afetar a qualidade da rede. [5]

## **10 Explique cada uma das características dos dados abaixo:**

### **10.1 Granularidade**

nível (detalhes/agregação)

### **10.2 Consistência e Inconsistência**

Diferentes “coisas” representados pelo mesmo nome em diferentes sistemas. A mesma “coisa” representada por diferentes nomes em diferentes sistemas.

### **10.3 Poluição**

Os campos podem conter espaços em branco, estar incompletos, inexatos, inconsistentes ou não identificáveis.

### **10.4 Relacionamento**

É importante observar e analisar a consistência das instâncias dos objetos da estrutura problema.

### **10.5 Integridade**

Deve ser observada a integridade das relações avaliadas.

### **10.6 Duplicação ou Redundância**

Ocorre principalmente quando as instâncias dependem de diferentes fluxos de dados. As variáveis Duplicadas ou Redundantes exigem maior esforço computacional e dependendo do caso podem ser reduzidas. [5]

## References

- [1] Josemar Alves C. Por que estudar preparação de dados?, 2018.
- [2] Cetax. Data mining: o que é, conceito e definição, 2017.
- [3] Ronaldo G. and Eduardo B. Exemplos de aplicações de data mining no mercado brasileiro, 2016.
- [4] Surender K. and Kanwaldip K. Review of data mining (knowledge discovery) in the future. 2016.
- [5] Livia M. Mineração de dados, 2019.
- [6] Ana N. Conhecimento tácito, explícito e implícito, 2003.
- [7] Fzal Shamil R. Attributes types in data mining, 2016.
- [8] Techopedia. Descoberta de conhecimento em bancos de dados (kdd), 2019.