

Análise Inteligente de Dados em Sistemas Big Data

Mestrado em Engenharia Informática – Engenharia do Conhecimento

2023/24

Trabalho Prático

Análise de Dados com PowerBI

Objetivo:

Utilizar o Microsoft *PowerBI* para desenvolver um *dashboard* interativo que permita a análise detalhada da informação constante de um *dataset*. O conjunto de dados escolhido em formato CSV contém informações, possivelmente referentes a vários anos, incluindo dados sobre objetos, quantidades, preços ou datas.

Requisitos:

1. Preparação do dataset

Preparar um *dataset* a partir de um sistema *Big Data* envolve várias etapas essenciais para transformar grandes volumes de dados brutos num conjunto de dados útil para análises estatísticas ou *Machine Learning*. É um processo iterativo e ajustado para maximizar a eficácia das análises estatísticas e dos modelos de *Machine Learning*.

Definição de Objetivos

Defina claramente o que pretende analisar para a seleção dos dados relevantes.

Acesso aos Dados

Diga como pode aceder aos dados no sistema *Big Data*, usando ferramentas adequadas para conectar-se a bases de dados como Hadoop ou Cassandra, ou usando APIs para extrair os dados necessários.

Limpeza de Dados

Diga como limpar os dados para remover erros ou valores ausentes, corrigir formatos inconsistentes e eliminar dados irrelevantes. Esta etapa é crucial para garantir a qualidade dos dados.

Transformação de Dados

Diga como transformar os dados para adequá-los aos requisitos das análises. Isso pode incluir normalização, criação de novas variáveis e agregação de dados.

Redução e Integração de Dados

Diga como reduzir a dimensionalidade dos dados e integrar dados de diferentes fontes, resolvendo conflitos e alinhando esquemas.

Segmentação de Dados

Se for para *Machine Learning*, os dados devem ser divididos em conjuntos de treino, validação e teste para avaliar adequadamente o modelo. Faça referência a essa etapa.

Anonimização de Dados

Se os dados incluírem informações sensíveis, aplique técnicas de anonimização para proteger a privacidade.

Armazenamento de Dados

Diga como guardar o *dataset* num formato adequado, como CSV ou HDF5, pronto para análise estatística ou para criação de modelos. Mostre como o processo de geração dos dados pode ser dinâmico.

2. Análise dos dados

Preâmbulo

Importe o arquivo CSV para o PowerBI e realize a limpeza e transformação dos dados necessárias. Pode incluir a remoção de valores ausentes, correção e formatos de dados e criação de novas colunas calculadas.

Análise e Visualização

Desenvolva visualizações que incluam:

- Valores totais ao longo do tempo (linha do tempo).
- Valores por categorias (gráfico de barras).
- Distribuição de valores por região geográfica (mapa).
- Principais objetos em termos de quantidade e valor (gráfico de barras e gráfico de ranking).
- Análise de tendências.

Utilize filtros e segmentações para permitir a interação do utilizador com o *dashboard*.

Insights e Conclusões

Desenvolva ao menos três insights significativos sobre o desempenho do sistema representado com base nos dados e visualizações criadas. Faça recomendações estratégicas baseadas na análise dos dados.

3. Relatório e Apresentação

Elabore um relatório documentando todo o processo de preparação do dataset e de análise de dados, incluindo a limpeza de dados, as escolhas de visualização e os insights obtidos. Prepare uma apresentação do *dashboard*, demonstrando como o *PowerBI* pode ser utilizado para suportar decisões estratégicas de negócios. O artigo científico a elaborar será baseado em grande escala neste relatório mais detalhado.

4. Entrega

- *Dashboard* finalizado no *PowerBI*.
- Relatório em formato PDF.
- Artigo científico.
- Apresentação em PowerPoint ou formato similar.

5. Critérios de Avaliação:

- Correção e eficácia na limpeza e preparação de dados.
- Clareza, relevância e estética das visualizações criadas.
- Profundidade dos insights e relevância das recomendações estratégicas.
- Qualidade da comunicação no relatório e na apresentação.