

Para um trabalho na disciplina de Introdução à Mineração de Dados, é essencial demonstrar conhecimento prático das técnicas de pré-processamento de dados, fundamentais para a preparação de dados antes da aplicação de métodos de machine learning. Aqui está um esboço detalhado de como você poderia estruturar seu trabalho, utilizando uma das bases de dados selecionadas na atividade anterior. Este esboço serve como um guia para estruturar seu trabalho de forma lógica e acadêmica, destacando a importância e a implementação de técnicas de pré-processamento em mineração de dados.

## **Título do Trabalho**

"Pré-processamento de Dados para Mineração: Uma Abordagem Prática com a Base de Dados [Nome da Base de Dados]"

## **Introdução**

- Contextualização: Breve introdução à importância do pré-processamento de dados na mineração de dados.
- Objetivo: Descrever o objetivo do trabalho, que é aplicar técnicas de pré-processamento na base de dados selecionada para prepará-la para análises futuras.
- Descrição da Base de Dados: Apresentar a base de dados escolhida, incluindo fonte, natureza dos dados e relevância.

## **Trabalhos Relacionados**

Apontar outros trabalhos encontrados na literatura que também utilizaram a base de dados que você escolheu, como eles pré-processaram e que resultados foram encontrados.

## **Metodologia**

- Seleção da Base de Dados: Justificar a escolha da base de dados.
- Ferramentas Utilizadas: Listar as ferramentas, bibliotecas e linguagens de programação (como Python, R, SQL) utilizadas no pré-processamento dos dados.
- Técnicas de Pré-processamento Aplicadas: Descrever cada técnica de pré-processamento que será aplicada, como:
  - Limpeza de Dados: Remoção de dados faltantes, correção de erros de digitação e formatação.
  - Integração de Dados: Caso seja relevante, descrever como diferentes conjuntos de dados foram combinados.
  - Transformação de Dados: Normalização, padronização, e conversão de tipos de dados.
  - Redução de Dados: Técnicas para reduzir o volume de dados sem perder informações importantes, como seleção de características e redução de dimensionalidade.

## **Desenvolvimento**

- Implementação das Técnicas: Descrever e documentar o processo de implementação de cada técnica de pré-processamento, com exemplos de código e discussões sobre as escolhas feitas.
- Desafios Encontrados: Discutir quaisquer desafios ou problemas encontrados durante o pré-processamento e como foram resolvidos.

## **Resultados e Discussão**

- Análise dos Dados Pós-Pré-processamento: Apresentar os dados após o pré-processamento, possivelmente incluindo estatísticas descritivas, visualizações, e discussões sobre como os dados foram transformados e otimizados para análise posterior.
- Impacto das Técnicas Utilizadas: Avaliar o impacto das técnicas de pré-processamento na qualidade dos dados e na eficiência subsequente da mineração de dados.

## **Conclusão**

- Sumário dos Aprendizados: Recapitular os principais aprendizados e a importância das técnicas de pré-processamento aplicadas.
- Possíveis Melhorias: Sugerir melhorias ou outras técnicas que poderiam ser exploradas em trabalhos futuros.
- Aplicações Futuras: Discutir como os dados preparados podem ser utilizados em projetos futuros de mineração de dados.

## **Referências**

- Incluir todas as fontes acadêmicas, documentações e recursos online utilizados para o desenvolvimento do trabalho.