

MINISTÉRIO DA EDUCAÇÃO
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas
IFSULDEMINAS - Campus Poços de Caldas
Avenida Dirce Pereira Rosa, 300. Poços de Caldas/MG. CEP 37713-100
Fone: (35) 3713-5120

Inteligência Artificial - Trabalho Prático 2

Pré-Processamento e Análise de Dados

Prof. Douglas Castilho

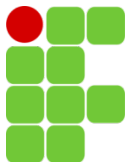
Disponível Desde: 04 de julho de 2024

Data de Entrega: 01 de agosto de 2024

Valor: 1.5 pontos

Neste trabalho, você deverá realizar análise e pré-processamento da base de dados que foi atribuída para você. Cada aluno é responsável por uma base de dados, cujo link para acesso está disponível na página da disciplina. Durante essa etapa de pré-processamento e análise de dados, você deverá seguir o seguinte roteiro como guia para suas atividades. Mesmo que a base de dados não apresente as características analisadas, você deverá realizar a análise e mostrar que elas não se aplicam:

1. Identificação do atributo alvo (saída);
2. Identificação dos tipos de dados dos atributos de entrada (quantitativo, qualitativo);
3. Identificação da escala de dados dos atributos de entrada (nominal, ordinal, intervalar, racional);
4. Exploração dos dados através de medidas de localidade;
5. Exploração dos dados através de medidas de espalhamento;
6. Exploração dos dados através de medidas de distribuição;
7. Identificação e separação do conjunto de teste, que será utilizado para testar o desempenho dos modelos – o conjunto de testes deve ser representativo e ter as características da população completa. Caso sua base de dados já tenha o conjunto de teste definido, analisar se este segue as características do conjunto de treinamento;



MINISTÉRIO DA EDUCAÇÃO
Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas
IFSULDEMINAS - Campus Poços de Caldas
Avenida Dirce Pereira Rosa, 300. Poços de Caldas/MG. CEP 37713-100
Fone: (35) 3713-5120

8. Identificação e eliminação de atributos não necessários;
9. Identificação e eliminação de exemplos não necessários;
10. Análise e aplicação de técnicas de amostragem de dados (caso não seja necessário, analisar o porquê);
11. Identificação e aplicação de técnicas para minimizar problemas de desbalanceamento (caso não seja necessário, analisar o porquê);
12. Limpeza de dados:
 - a. Identificação e eliminação de ruídos ou outliers;
 - b. Identificação e eliminação de dados inconsistentes;
 - c. Identificação e eliminação de dados redundantes;
 - d. Identificação e resolução de dados incompletos (ausentes) – utilização de alguma técnica de preenchimento e justificar;
13. Identificação e conversão dos tipos de dados (caso não seja necessário, analisar o porquê).

Os tipos de conversão que podem ser utilizados são:

 - a. Conversão de tipos (simbólico para numérico, ordinal para numérico, nominal para binário, numérico para ordinal);
 - b. Normalização dos dados (re-escala ou padronização);
14. Análise e aplicação de alguma técnica para redução de dimensionalidade – pesquisar alguma técnica utilizada na literatura e aplicar;