

Atividade 1 – Mineração de Dados

Profa Elaine Ribeiro de Faria

Aluno: Gustavo Borges – 11921BSI240

Aluno: Paulo Diego – 11921BSI237

Aluno: Celso Emiliano – 11921BSI230

1. a. Mineração de Dados é o processo de se aplicar técnicas a fim de extrair informações úteis de bases de dados de maneira automatizada.

1. b. Como exemplos de aplicações no mundo real, podemos citar o consumo de informações de vendas de comércio eletrônico a fim de encontrar padrões de compras de clientes e identificar novas oportunidades de negócios.

Outro exemplo é a previsão meteorológica, onde pode-se minerar os dados de imagens obtidas de satélites e submetê-los a um modelo para prever as condições meteorológicas futuras.

1. c. De maneira geral, eu acredito que é importante estudar mineração de dados, pois o seu conjunto de técnicas e conhecimentos é extremamente útil para a sociedade, pessoalmente, eu gostaria de estudá-la para buscar uma oportunidade no mercado de trabalho e para um futuro projeto de TCC ou pós-graduação.

1. d. Como algoritmos de classificação, podemos citar o algoritmo de k-vizinhos mais próximos e o algoritmo de naive bayes.

Como algoritmos de agrupamento, podemos citar o algoritmo k-means e o algoritmo de clustering.

2. Neste arquivo, coloquei o pseudocódigo feito durante a aula no laboratório. Em anexo à tarefa do Microsoft Teams, coloquei o arquivo resultado. Para utilizá-lo, é necessário passar o nome do arquivo e a quantidade de amostras na linha de comando, conforme exemplo abaixo:

```
> python3 extrai_amostra.py iris.data 15
```

```
"""
```

LÊ A QUANTIDADE DE AMOSTRAS DESEJADA.

ABRA O ARQUIVO DE ENTRADA

CONTADOR = 0

ENQUANTO O ARQUIVO NÃO CHEGAR AO FIM:

LÊ UMA LINHA

CONTADOR = CONTADOR + 1

FECHA O ARQUIVO DE ENTRADA

SE A QUANTIDADE DE AMOSTRAS FOR MAIOR OU IGUAL A
QUANTIDADE DE LINHAS:

COPIA O CONTEUDO DO ARQUIVO DE ENTRADA PARA O ARQUIVO
RESULTADO

TERMINA O PROGRAMA

LISTA = []

ENQUANTO A LISTA NÃO TIVER O TAMANHO IGUAL A QUANTIDADE DE
AMOSTRAS:

NUMERO = NUMERO ALEATÓRIO ENTRE 0 E TAMANHO DO ARQUIVO

SE NUMERO NÃO ESTIVER NA LISTA:

INSERE O NÚMERO NA LISTA

ORDENA A LISTA.

ABRE O ARQUIVO DE ENTRADA

ABRE O ARQUIVO DE SAIDA

$I = 0$

ENQUANTO $I < \text{TAMANHO DO ARQUIVO}$:

LÊ A LINHA

SE I ESTIVER NA LISTA:

ESCREVE LINHA NO ARQUIVO DE SAIDA

$I = I + 1$

FECHA O ARQUIVO DE ENTRADA

FECHA ARQUIVO DE SAIDA

""