



Universidade Federal do Ceará  
Centro de Ciências  
Departamento de Computação

**CKP9011 – Introdução à Ciência de Dados**  
**CK0223 - Mineração de Dados**  
**2025.1**

## **Lista 1**

Exercício: Tratamento de Dados

Objetivos: Exercitar os conceitos referente à manipulação, tratamento e limpeza de dados.

Data da Entrega: 05/05/2025

OBS 1: Exercício Individual.

OBS 2: A entrega da lista deverá ser executada utilizando-se o SIGAA.

### **1. Tarefa**

Crie um arquivo Jupyter Notebook e realize as seguintes operações:

- a) Ler o dataset fakeTelegram.BR\_2022.csv, o qual está disponível no link a seguir:  
[https://drive.google.com/file/d/1c\\_hLzk85pYw-huHSnFYZM\\_gn-dUsYRDm/view?usp=drive\\_link](https://drive.google.com/file/d/1c_hLzk85pYw-huHSnFYZM_gn-dUsYRDm/view?usp=drive_link)
- b) Identificar e listar as posições (células) contendo valores faltantes.
- c) Contar quantas linhas possuem valores faltantes.
- d) Para cada coluna (feature), contar quantas linhas possuem valores faltantes.
- e) Identificar e listar as linhas repetidas (duplicadas).
- f) Identificar e listar as posições (células) contendo valores que não pertencem ao domínio (tipo de dados) esperado.
- g) Crie uma coluna chamada “caracteres” contendo a quantidade de caracteres da coluna “text”.
- h) Crie uma coluna chamada “words” contendo a quantidade de palavras da coluna “text”.
- i) Crie uma coluna chamada “viral” contendo o valor 0 se o texto da mensagem (valor do campo “text”) não for encontrado em outras linhas e 1, caso contrário.
- j) Crie uma coluna chamada “sharings” contendo a quantidade de vezes que o texto armazenado no atributo “text” aparece no dataset.
- k) Crie uma coluna chamada “sentiment” contendo os valores: -1 para textos negativos, 0 para textos neutros e 1 para textos positivos.
- l) Eliminar as linhas cujo valor da coluna “text” contenham “trava-zaps”.
- m) Identificar inconsistências entre os atributos (features).

## **2. Avaliação**

Espera-se com a realização deste trabalho que cada estudante elabore e entregue (de forma digital) os seguintes documentos:

- Jupyter Notebook contendo o código utilizado na implementação das tarefas.
- Vídeo (disponibilizado no Youtube) apresentando e descrevendo as atividades desenvolvidas.

A avaliação deste trabalho se dará em duas etapas:

1ª. Vídeo de Apresentação do Dataset: Cada estudante irá disponibilizar um vídeo (no Youtube) apresentando o código desenvolvido para implementação das tarefas. O estudante pode utilizar slides e notebooks.

2ª. Avaliação do Notebook: O professor da disciplina irá avaliar a qualidade do notebook gerado pelo estudante, bem como dos códigos implementados e análises realizadas.

A avaliação do trabalho irá envolver os seguintes quesitos:

- Abrangência e Organização do Notebook
- Qualidade dos Códigos Utilizados
- Clareza do Texto Utilizado para Descrever as Atividades Realizadas e os Resultados Obtidos
- Domínio do Tema

## **3. Data da Entrega: 05/05/2025**

- PS. Não serão aceitos trabalhos que não forem apresentados (por meio de vídeo disponibilizado no Youtube).
- PS. Cada estudante será responsável pela disponibilização do ambiente (software e hardware) necessário para a gravação da apresentação do seu trabalho.
- Os Notebooks deverão ser disponibilizados, em formato .ZIP, no SIGAA ou em um repositório público (GitHub ou GitLab).

“A Educação, qualquer que seja ela, é sempre uma teoria do conhecimento posta em prática”.

**Paulo Freire**