

# **Mineração de dados**

## **Tia Lu Statics Camacari**

**Autores:** <César Filipe Gomes da Silva>  
<Rodrigo Souza Guimarães>  
<Valnei Sousa Conceição Filho>

# Agenda

Objetivo da apresentação

**1. Introdução:** Importância da estatística na mineração de dados e o objetivo do projeto.

**2. Fundamentação Teórica:** Descrição breve das principais funções presentes no código

**3. Metodologia:** Ferramentas utilizadas na implementações das funções da classe Statics

**4. Resultado e discussões:** Resultado das implementações e desafios enfrentados

**5. Considerações Finais:** Aprendizados adquiridos e próximas etapas.

# Introdução



- Mineração de dados é um processo automatizado que analisa grandes volumes de dados, para transformar esses dados em insights é necessário compreender conceitos estatísticos.
- O objetivo principal do projeto é criar uma biblioteca para métricas estatísticas em Python , que permite reforçar o entendimento de estatística e programação.
- Interessante destacar que foram utilizados apenas recursos nativos da linguagem, sem depender de bibliotecas externas.

# Fundamentação Teórica



- Média -  $\mu = \frac{1}{N} \sum_{i=1}^N x_i$
- Mediana - A mediana é o valor central de um conjunto de dados ordenado.
- Moda - A moda é o valor que aparece com mais frequência no conjunto de dados.
- Variância -  $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$
- Desvio padrão -  $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$
- Frequências - Frequência absoluta, Frequência relativa, Frequência acumulada;
- Pureza - Mede a homogeneidade de uma partição de dados em Árvores de Decisão.
- Ganho de Informação - Redução da incerteza ao dividir dados por um atributo.

# Metodologia



- Validação do Dataset
- Estrutura da Classe Statistics
- Restrições

# Resultado e Discussões



- Comparando com bibliotecas externas os resultados foram consistentes, consolidando as implementações.
- Todas as restrições foram respeitadas.

# Considerações Finais



- Desafios: Validação do dataset, cálculo manual de variância e desvio padrão, lidar com listas vazias ou colunas com valores repetidos.
- Anotações: Integração entre estatística e programação, reforço do entendimento de métricas sem depender de bibliotecas externas.
- Melhorias futuras: Implementação de mais métricas (quartis, coeficiente de variação), otimização dos métodos e criação de uma interface para manipulação direta do dataset.

# Considerações Finais



- Desafios: Validação do dataset, cálculo manual de variância e desvio padrão, lidar com listas vazias ou colunas com valores repetidos.
- Anotações: Integração entre estatística e programação, reforço do entendimento de métricas sem depender de bibliotecas externas.
- Melhorias futuras: Implementação de mais métricas (quartis, coeficiente de variação), otimização dos métodos e criação de uma interface para manipulação direta do dataset.



# Referências



- REUND, J.; PERLIS, D. Estatística e Probabilidade. Rio de Janeiro: LTC, 2018. MANN, P. Estatística Básica. 5. ed. São Paulo: Pearson, 2020. PYTHON SOFTWARE FOUNDATION. Python Documentation. Disponível em: <https://docs.python.org/3/>. Acesso em: 23 ago. 2025