

# Aprendizado Estatístico em Dados Longitudinais

Cícero Hitzschky

cicero.hitzschky@alu.ufc.br

**Departamento de Estatística e Matemática Aplicada  
Universidade Federal do Ceará**

27 de fevereiro de 2025

# Sumário

## 1 Aplicação



# Contextualização

## Osteoporose

é uma doença em que a degradação estrutural e a diminuição da densidade mineral dos ossos (DMO) aumentam o risco de fraturas ósseas.

## Motivações para o estudo

- A prevalência da osteoporose aumentou drasticamente nos últimos anos;
- Representa um problema de saúde pública com alto índice de morbidade (muito comum).



# Contextualização

- Em 2010, na União Europeia, aproximadamente 5,5 milhões de homens e 22 milhões de mulheres foram afetados pela osteoporose;
- 80% das mulheres afetadas não estavam cientes dos seus fatores de risco até o diagnóstico.

# Objetivos

- 1 Comparar a precisão preditiva dos métodos aprendizado de máquina com a regressão linear múltipla tradicional;
- 2 Classificar a importância de vários fatores de risco, incluindo dados demográficos, estilo de vida e bioquímica, na previsão das mudanças futuras no  $\delta$ -T score.

# Metodologia



# Fonte dos Dados

- Coorte MJ de Taiwan
  - Coorte prospectiva em andamento
  - Exames conduzidos pelos Centros de Triagem de Saúde MJ



# Informações Coletadas

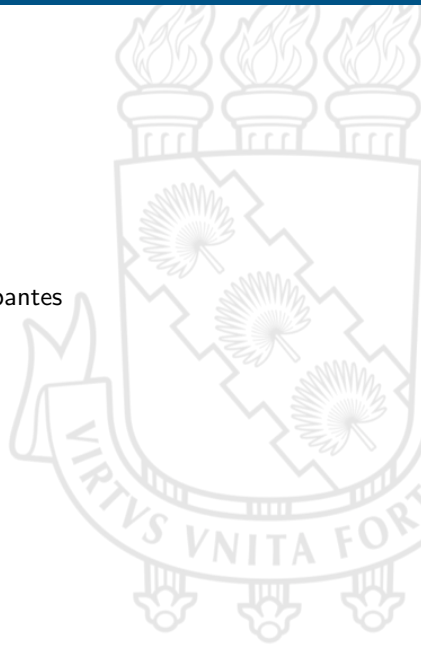
- Mais de 100 indicadores biológicos essenciais
- Questionário abrangendo:
  - Histórico médico pessoal e familiar
  - Estado de saúde atual
  - Estilo de vida e exercício físico
  - Hábitos de sono e alimentares





# Considerações Éticas

- Consentimento informado dos participantes
- Aprovação pelo Comitê de Ética



# Modelos Utilizados

- Floresta Aleatória (RF)
  - É baseado em árvores de decisão que combina as técnicas de bagging e boosting.
  - Minimiza a função de perda e resolve o sobreajuste das árvores de decisão tradicionais.

# Modelos Utilizados

- Gradient Boosting Estocástico (SGB)
  - Classifica objetos com base em características e variáveis específicas.
  - Utiliza o teorema de Bayes para calcular a probabilidade das hipóteses sobre grupos presumidos.

# Modelos Utilizados

- Naive Bayes (NB)
  - Classifica objetos com base em características e variáveis específicas.
  - Utiliza o teorema de Bayes para calcular a probabilidade das hipóteses sobre grupos presumidos.

# Modelos Utilizados

- Extreme Gradient Boosting (XGBoost)
  - Tecnologia de gradient boosting baseada na extensão otimizada do SGB.
  - Treina vários modelos “fracos” e faz ensemble com o Gradiente Boosting.

# Referências I



# Obrigado!!!

**Contato:**

cicero.hitzschky@alu.ufc.br