



## **Pesquisa e Experimentação em Sistemas de Informação - Eixo 7**

### **Obesity Risk**

#### **Integrantes:**

- Douglas Fernandes de Carvalho Jardim
- Bernardo Campos Rodrigues
- Amanda Lima Carvalho Dal Col
- Pedro Henrique Monti
- Thomas Henrique Lousada

#### **Orientadora:**

- Luciana de Nardin

# Introdução

- Mais de 1 bilhão de adultos estão acima do peso, 650 milhões são obesos (OMS, 2023)
- Soluções tecnológicas podem mitigar os efeitos deletérios da obesidade
- Objetivo do projeto: Desenvolver modelos de aprendizado de máquina para classificar, identificar fatores de risco, prever mudanças nos níveis de obesidade e segmentar dados

## Problema

- 2. Problema
- Obesidade no Brasil
- Atinge mais de 20% da população adulta
- 40 milhões de pessoas acima do peso ideal
- Fatores de risco: Má alimentação, falta de atividade física e sedentarismo
- Dataset: Dados de obesidade de México, Peru e Colômbia (Palechor & De la Hoz Manotas, 2023)

# Questão de Pesquisa

## Pergunta Central

Com que precisão é possível prever o nível de obesidade de um indivíduo levando em consideração diversos fatores como hábitos alimentares, tabagismo e condição física?

## Objetivos Preliminares

### 1. Classificar Níveis de Obesidade

- Utilizar os dados coletados para desenvolver um modelo de aprendizado de máquina capaz de classificar os níveis de obesidade com base nos hábitos alimentares, condição física e dados demográficos dos indivíduos

### 2. Identificar Fatores de Risco

- Analisar os atributos relacionados aos hábitos alimentares e à condição física para identificar os principais fatores de risco associados aos diferentes níveis de obesidade

### 3. Prever Mudanças nos Níveis de Obesidade

- Desenvolver um modelo preditivo que possa estimar mudanças nos níveis de obesidade com base em alterações nos hábitos alimentares e de atividade física

### 4. Segmentação de Dados

- Usar técnicas de segmentação para identificar grupos específicos dentro da população estudada que possam se beneficiar mais de intervenções direcionadas, baseadas em seus hábitos alimentares e níveis de atividade física






# Justificativa

- Crescente prevalência da obesidade globalmente (OMS, 2023)
- Associação com doenças crônicas: diabetes, hipertensão, doenças cardiovasculares (CDC, 2023)
- Custos elevados da obesidade para sistemas de saúde (OCDE, 2010)
- Necessidade de entender fatores de risco para intervenções eficazes (NCD-RisC, 2016)



## Público Alvo

- Profissionais da Saúde: Médicos, nutricionistas, psicólogos
  - Pesquisadores e Acadêmicos: Saúde, nutrição, epidemiologia, ciência de dados
  - Organizações Governamentais e ONGs: Políticas de saúde e bem-estar
  - Indivíduos e Famílias: Prevenção e tratamento da obesidade
- 
- 
- 

# Estado da Arte

A prevalência da obesidade entre os jovens é um desafio crescente para a saúde pública global, que causa impactos físicos, emocionais e sociais

## 1. Fatores de Risco para Obesidade Adolescente

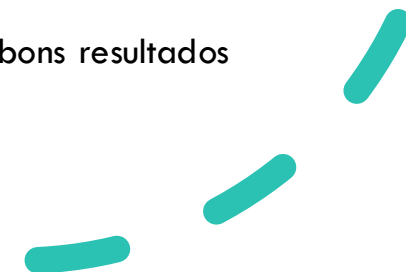
- Antecedentes: Revisão da literatura sobre aspectos biológicos, sociais, psicológicos e nutricionais.
- Algoritmos Utilizados: Regressão logística, Árvores de Decisão, Redes Neurais, SVM.
- Métricas de Avaliação: Acurácia, sensibilidade, especificidade, AUC, F1.
- Resultados: Bons resultados com Regressão Logística e Árvores de Decisão (taxa média de sucesso ~80%).

## 2. Tabagismo

- Antecedentes: Revisão da relação entre tabagismo e obesidade.
- Algoritmos Utilizados: Regressão logística, Análise discriminante linear, Random Forest.
- Métricas de Avaliação: Precisão, sensibilidade, especificidade, AUC.
- Resultados: Random Forest teve melhor desempenho (precisão média 85%, AUC 0,90).

## 3. Obesidade Abdominal e Doenças Cardiovasculares

- Contexto: Relação entre obesidade abdominal e risco cardiovascular.
- Algoritmos Utilizados: Regressão logística, Redes Neurais, SVM.
- Métricas de Avaliação: Precisão, sensibilidade, especificidade, AUC.
- Resultados: Regressão logística foi mais eficaz (precisão média 82%, AUC 0,88). Redes Neurais também apresentaram bons resultados (precisão média 80%, AUC 0,85).



# Descrição do Dataset

## Dataset Utilizado

- Estimativas de níveis de obesidade em indivíduos do México, Peru e Colômbia
- Composto por 2111 registros com 17 atributos
- Inclui hábitos alimentares, condição física e dados demográficos
- Variável de interesse: Nível de Obesidade

## Atributos do Dataset

### Detalhes dos Atributos

- Gênero, idade, altura, peso
- Hábitos alimentares: Consumo de alimentos calóricos, vegetais, refeições
- Comportamentos: Tabagismo, consumo de água, atividade física
- Outros: Uso de dispositivos tecnológicos, consumo de álcool, tipo de transporte, etc...

# Atributos do Dataset

Variável	Tipo	Faixa de Valores
Gênero (Gender)	Categórica nominal	Masculino (Male), Feminino (Female)
Idade (Age)	Quantitativa discreta	17 a 61
Altura (Height)	Quantitativa contínua	1.45 a 1.98
Peso (Weight)	Quantitativa contínua	39 a 173
Histórico familiar de sobrepeso (family_history_with_overweight)	Categórica nominal	Sim (Yes), Não (No)
Consome alimentos ricos em calorias com frequência (FAVC)	Categórica nominal	Sim (Yes), Não (No)
Frequência de consumo de vegetais (FCVC)	Quantitativa discreta	1 a 3
Número de Refeições Principais (NCP)	Quantitativa discreta	1 a 4
Consumo de alimentos entre as refeições (CAEC)	Categórica ordinal	Não (No), Às vezes (Sometimes), Frequentemente (Frequently), Sempre (Always)
Tabagismo (Smoke)	Categórica nominal	Sim (Yes), No (Não)
Consumo de água diário (CH2O)	Quantitativa discreta	1 a 3
Monitoramento do consumo de calorias (SCC)	Categórica nominal	Sim (Yes), No (Não)
Frequência de atividade física (FAF)	Quantitativa discreta	0 a 3
Tempo usando dispositivos tecnológicos (TUE)	Quantitativa discreta	0 a 2
Consumo de álcool (CALC)	Categórica ordinal	Não (No), Às vezes (Sometimes), Frequentemente (Frequently), Sempre (Always)
Meio de transporte utilizado (MTRANS)	Categórica nominal	Transporte público (Public_Transportation), Carro (Automobile), Andando (Walking), Moto (Motorbike), Bicicleta (Bike)
Nível de obesidade (NObeyesdad))	Categórica ordinal	Peso insuficiente (Insufficient_Weight), Normal (Normal_Weight), Sobrepeso grau I (Overweight_Level_I), Sobrepeso grau II (Overweight_Level_II), Obesidade grau I (Obesity_Type_I), Obesidade grau II (Obesity_Type_II), Obesidade grau III (Obesity_Type_III)

# Canvas Analítico

1. QUESTÕES	2. FONTE DE DADOS	3. HEURÍSTICA	4. VALIDAÇÃO
Como identificar padrões comportamentais ligados a obesidade.	Fonte advinda do Kaggle onde foram coletadas dados de pessoas do México, Peru e Colombia.	Podemos supor que as pessoas que não praticam exercícios físicos, não possuem dieta alimentar equilibrada e possuem hábitos sedentários estão diretamente ligada a risco de obsidade	Serão avaliados dados como idade, hábitos diários, alimentação, consumo de água e álcool bem como tempo de uso em aparelhos tecnológicos e meios de transporte.
5. IMPLEMENTAÇÃO		6. RESULTADOS	7. PRÓXIMOS PASSOS
Passos para implementação do processo de informação. 1. Definição do contexto e levantamento de dados. 2. Conhecimento dos dados. 3. Preparação dos dados, construção e avaliação dos modelos propostos. 4. Implementação e avaliação da solução.		Identificação de insights e padrões das informações geradas .	Plano de ação após a obtenção dos insight e dos padrões gerados.



# Avaliação dos Modelos Criados

## Árvore de Decisão

**Acurácia:** 89,91% nos dados de validação e 92,11% nos dados de teste

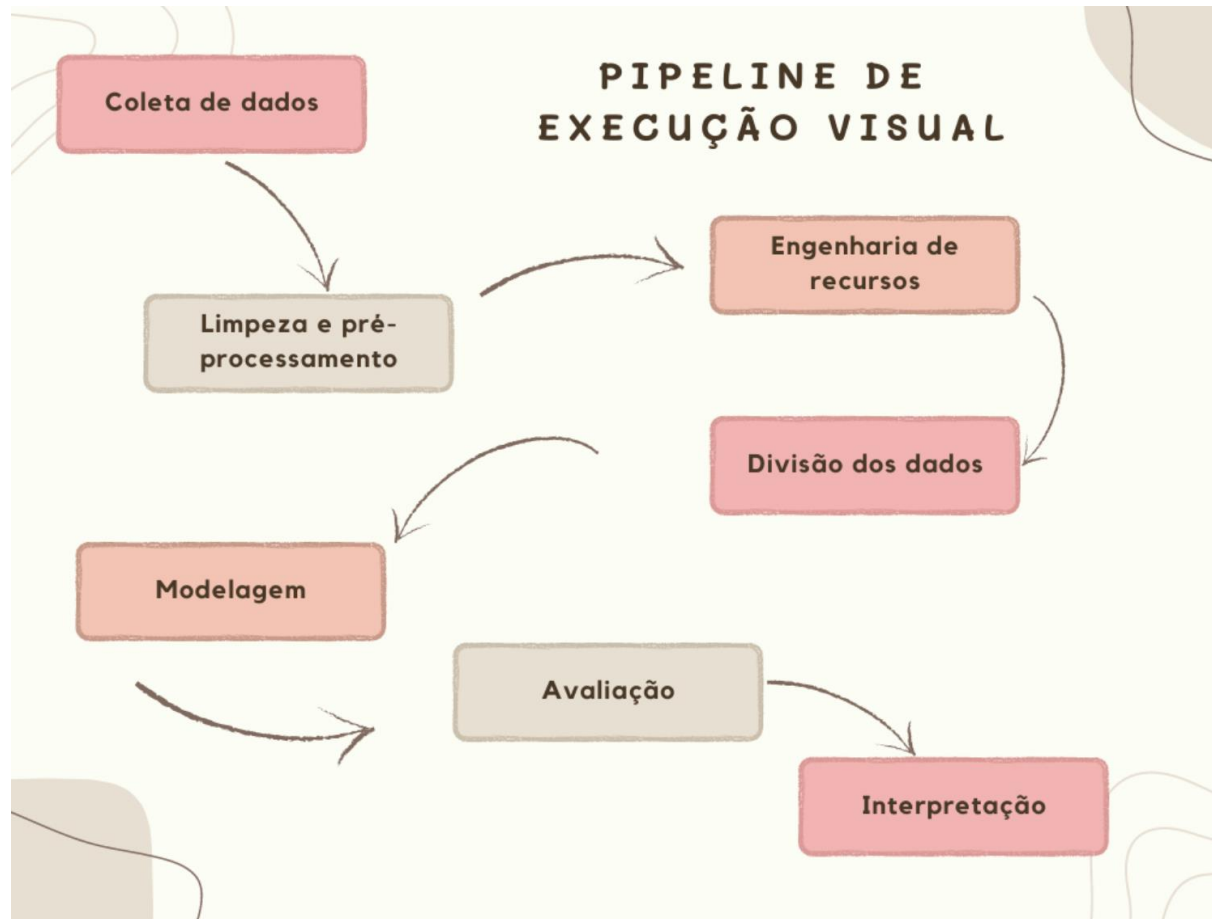
## Regrssão Logística

**Acurácia:** 85%

## Random Forest

**Acurácia:** 95%

# Pipeline de Execução Visual



# Obrigado!

