

Projeto: Previsão de Obesidade

Autora: Carolina Rocha

Data: 15/02/2024

1. Introdução:

- a. A obesidade é um problema de saúde pública global que afeta pessoas de todas as idades, gêneros e origens. Este projeto utiliza uma base de dados abrangente para analisar o índice de obesidade na população mundial fornecendo insights valiosos sobre a prevalência e os fatores de risco, estilo de vida e indicadores de saúde dos indivíduos.

2. Alto volume de obesidade na população mundial

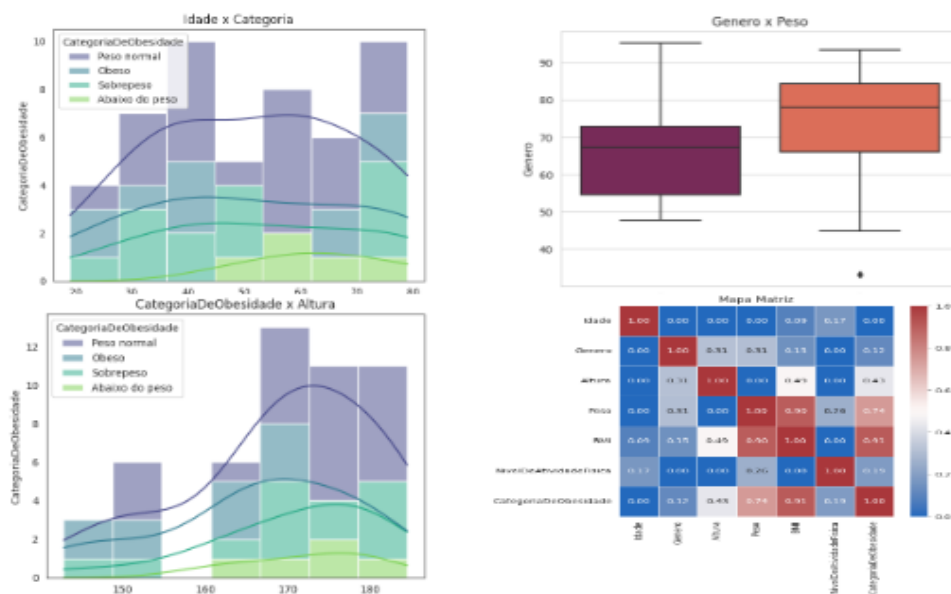
- a. A base de dados, tem a finalidade de apresentar com métricas, dashboards e afins, o índice de obesidade que vem em uma crescente atingindo a população mundial. Os padrões e campos definidos para trabalhar foram (Idade, Gênero, Peso, Nível de atividade física e Índice de gordura corporal) com essas informações, será possível a visualização detalhada de indivíduos jovens, de idade média e idade avançada. Com isso conseguiremos observar onde está o maior índice de obesidade, se a pessoa x pratica algum tipo de atividade ou se é totalmente sedentária.

3. Metodologia:

- a. O estudo utilizou um conjunto de dados de 50 indivíduos com informações sobre as variáveis mencionadas. As seguintes análises foram realizadas:
 - i. **Distribuição das idades por categoria de obesidade:** um gráfico de barras foi utilizado para visualizar a distribuição das idades dentro de cada categoria de obesidade.
 - ii. **Distribuição do IMC por gênero e categoria de obesidade:** um gráfico de linhas foi utilizado para comparar a distribuição do IMC entre homens e mulheres dentro de cada categoria de obesidade.
 - iii. **Mapa de Matriz:** um mapa de matriz foi criado para visualizar as correlações entre as variáveis.
 - iv. **Criação de um modelo de Regressão Linear:** um modelo de regressão linear foi criado para estimar a relação entre a idade, gênero, altura, peso e nível de atividade física com o IMC.
 - v. **Análise de Sensibilidade:** a análise de sensibilidade foi realizada para avaliar como pequenas mudanças nos valores das variáveis afetam o valor do IMC.

4. Exploração e coleta de dados:

- a. Exploração de dados de previsão de obesidade, os dados estão em formato csv, em especial estamos interessados em explicar e analisar a qual a relação das colunas idade, gênero, peso com a categoria de obesidade.
 - i. Utilizei na análise exploratória a função `describe()` para fornecer estatísticas descritivas, como média, desvio padrão, mínimo, máximo.
 - ii. método `info()` fornece informações sobre o DataFrame, como o tipo de dados de cada coluna, números e valores.
 - iii. Cálculo dos valores missing, função `isna()` retorna um valor booleano indicando se o valor é missing (True) ou não (False) enquanto a função `sum()` soma os valores booleanos retornando o número total de valores missing em cada uma.
 - iv. Gráficos:



5. Modelagem

- a. KNN (K-Nearest Neighbors)
 - i. É utilizado para realizar tarefas de classificação e regressão em problemas de aprendizado supervisionado. O principal objetivo do KNN é fazer previsões com base na proximidade dos dados no espaço de características.
- b. Random Forest Classifier
 - i. Avaliar o desempenho do modelo de Random Forest Classifier na classificação da categoria de obesidade com base nas características fornecidas no conjunto de dados. O Random Forest Classifier apresenta um desempenho geral superior ao KNN, identificando corretamente 80% dos indivíduos nas diferentes classes de peso.

- c. Avaliação do modelo de regressão de árvore de decisão: A árvore de decisão apresenta um bom desempenho na previsão da variável dependente, com um erro médio baixo e um R^2 alto.

6. Resultados

- a. Distribuição das idades por categoria de obesidade: a obesidade foi mais prevalente em indivíduos com idade superior a 50 anos.
- b. Distribuição do IMC por gênero e categoria de obesidade: o IMC foi maior em homens do que em mulheres em todas as categorias de obesidade.
- c. Mapa de Matriz: o mapa de matriz mostrou que o IMC estava positivamente correlacionado com a idade, peso e nível de atividade física, e negativamente correlacionado com a altura.
- d. Modelo de Regressão Linear: o modelo de regressão linear apresentou um bom ajuste aos dados, com um R^2 de 0,90.
- e. Análise de Sensibilidade: a análise de sensibilidade mostrou que o IMC era mais sensível a mudanças no peso e nível de atividade física do que a mudanças na idade, gênero e altura.
- f. Os resultados deste estudo sugerem que a idade, gênero, altura, peso e nível de atividade física são fatores importantes que influenciam o IMC. A obesidade é mais prevalente em indivíduos com idade superior a 50 anos e prevalece em homens. O IMC é mais sensível a mudanças no peso e nível de atividade física do que a mudanças na idade, gênero e altura.

7. Recomendações:

- a. Implementar programas de educação alimentar e promoção da atividade física em diferentes faixas etárias.
- b. Criar ambientes obesogênicos que facilitem a escolha de alimentos saudáveis e a prática de atividade física.
- c. Investir em pesquisas para compreender melhor as causas e os impactos da obesidade.
- d. Com o conhecimento adquirido a partir desta análise, podemos trabalhar juntos para construir um futuro mais saudável para a população mundial.