Classificação de Pacientes com Risco de Obesidade e DCV Utilizando Técnicas de Clusterização Não Supervisionada

Vinícius Caputo Resende de Oliveira
Instituto Federal de Minas Gerais (IFMG) - Campus Ibirité
R. Mato Grosso, 02 - Bairro - Vista Alegre, Ibirité - MG - Brasil
viniciuscaputoresende@outlook.com.br

Thalita Vieira Sales
Instituto Federal de Minas Gerais (IFMG) - Campus Ibirité
R. Mato Grosso, 02 - Bairro - Vista Alegre, Ibirité - MG - Brasil
thalitav.sales25@gmail.com

Maria Luísa Clemente dos Santos Tanini Vidal
Instituto Federal de Minas Gerais (IFMG) - Campus Ibirité
R. Mato Grosso, 02 - Bairro - Vista Alegre, Ibirité - MG - Brasil
luisatanini@gmail.com

Thiago Henrique Barbosa de Carvalho Tavares
Instituto Federal de Minas Gerais (IFMG) - Campus Ibirité
R. Mato Grosso, 02 - Bairro - Vista Alegre, Ibirité - MG - Brasil
thiago.tavares@ifmg.edu.br

ABSTRACT

Obesity is a chronic disease that has reached epidemic proportions in Brazil and around the world, being associated with factors such as a high-calorie diet, urbanization and a sedentary lifestyle. This work aims to implement and compare Clustering methods, such as K-means, using the Python programming language. The aim is to predict relevant parameters to identify behavioral patterns and family histories related to the development of obesity. Based on the predictions generated, the study seeks to provide support for the formulation of effective public treatment and guidance strategies, adapted to the specific needs of each risk group. With the results obtained, it is expected to contribute to reducing the rates of type III obesity and its comorbidities in Brazil.

Keywords
Obesity, Statistical Models, Machine Learning, Clustering Models, Machine Learning

RESUMO

A obesidade é uma doença crônica que atingiu proporções epidêmicas no Brasil e no mundo, sendo associada a fatores como alimentação rica em calorias, urbanização e estilo de vida sedentário. Este trabalho tem como objetivo implementar e comparar os métodos de Clusterização, como *K-means,* utilizando a linguagem de programação *Python*. O intuito é prever parâmetros relevantes para identificar padrões comportamentais e históricos familiares relacionados ao desenvolvimento da obesidade. A partir das previsões geradas, o estudo busca fornecer subsídios para a formulação de estratégias públicas eficazes de tratamento e orientação, adaptadas às necessidades específicas de cada grupo de risco. Com os resultados obtidos, espera-se contribuir para a redução das taxas de obesidade tipo III e suas comorbidades no Brasil.

Palavras-chaves
Obesidade, Modelos Estatísticos, Machine Learning, Modelos Clusterização, Machine Learning

1. INTRODUÇÃO

A obesidade é uma doença crônica multifatorial que tem crescido em proporções epidêmicas ao redor do mundo, nas últimas décadas, segundo pesquisas da Organização Mundial de Saúde (2020). Os impactos vão além do acúmulo excessivo de gordura corporal, envolvendo uma série de desregulações metabólicas e hormonais que aumentam significativamente o risco de comorbidades, como diabetes tipo 2, hipertensão arterial e doenças cardiovasculares (Vasques et al. 2007). De acordo com a Organização Mundial da Saúde - OMS (2021), a obesidade é responsável por mais de 2,8 milhões de mortes anuais em todo o mundo, configurando-se como um dos maiores desafios da saúde pública global.

No Brasil, a situação não é diferente: o país tem apresentado um crescimento significativo nas taxas de obesidade, com impactos tanto na saúde individual quanto no sistema de saúde pública. Estudos apontam que cerca de 60% da população adulta brasileira está com sobrepeso, sendo que 25% dessa parcela é clinicamente obesa (Ministério da Saúde, 2022). Este aumento das taxas de obesidade está diretamente relacionado a fatores sociais e econômicos, como a urbanização acelerada, a globalização e as

tendências dos hábitos alimentares e ao estilo de vida sedentário (Pereira et al., 2003).

A Figura 1 ilustra os dados sobre o excesso de peso e a obesidade no Brasil, com base na Pesquisa Nacional de Saúde (PNS) de 2020. O gráfico à esquerda mostra que 1 em cada 4 brasileiros maiores de 18 anos, o que corresponde a cerca de 41,2 milhões de pessoas, sofrem de obesidade. Já o gráfico à direita demonstra que 60,3% da população adulta, ou aproximadamente 96 milhões de brasileiros, apresentam excesso de peso. Esses dados destacam a alta prevalência de problemas relacionados ao peso no Brasil

Figura 1 - Excesso de peso e obesidade no Brasil

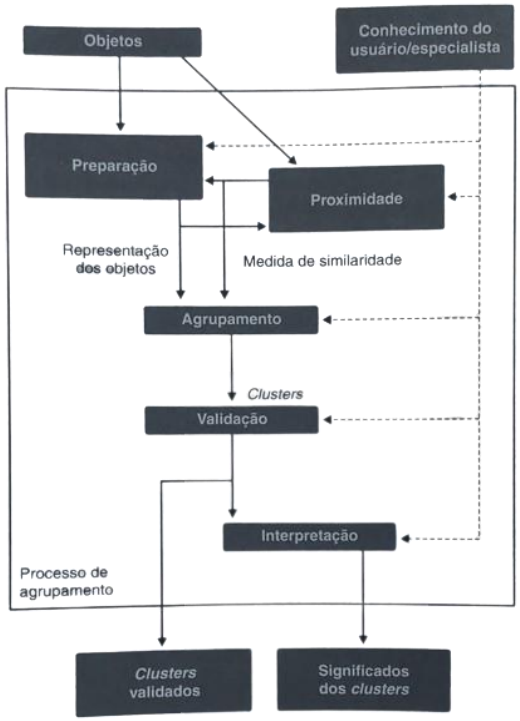
Fonte: Pesquisa Nacional de Saúde, 2020.

No contexto da saúde pública, a obesidade é um fenômeno social, que exige a coordenação de esforços multidisciplinares para sua prevenção e tratamento, assim como as outras doenças. O Sistema Único de Saúde (SUS) no Brasil tem implementado políticas focadas no manejo da obesidade, como a Portaria nº 2.055 de 2016, que estabelece diretrizes para o tratamento de pessoas com obesidade através de uma abordagem que envolve desde o aconselhamento nutricional até a cirurgia bariátrica para casos mais graves. Ainda assim, as iniciativas têm encontrado dificuldades para alcançar a população em larga escala, em parte devido à falta de segmentação dos grupos de risco e ao tratamento uniformizado de pacientes que possuem diferentes perfis metabólicos e comportamentais (Kusters, 2016). Esse cenário demanda uma abordagem mais precisa e personalizada para o combate à obesidade, o que pode ser alcançado por meio de técnicas avançadas de análise de dados (Ministério da Saúde, 2018).

Nesse sentido, o uso de métodos de análise de dados utilizando *Machine Learn*, como a clusterização, tem ganhado destaque como uma técnica capaz para lidar com a complexidade dos fatores associados à obesidade. A clusterização, um método de aprendizado de máquina não supervisionado, permite agrupar indivíduos com base em suas características comuns, como ingestão calórica, padrão de atividade física e histórico familiar de doenças (Manago and Auriol, 1996). Isso permite a identificação de padrões ocultos no comportamento dos indivíduos, bem como, a criação de estratégias de intervenção mais específicas e eficazes (Fayyad, Piatetsky-shapiro, Smyth, 1996). Em vez de tratar todos os pacientes obesos da mesma maneira, as políticas públicas de saúde poderiam ser ajustadas para atender às necessidades específicas de cada grupo de risco (Faceli et al., 2022).

As etapas apresentadas na Figura 2 são essenciais para assegurar que os resultados sejam significativos e úteis, garantindo que qualquer algoritmo de agrupamento identifique uma estrutura subjacente nos dados, de acordo com os critérios estabelecidos (Faceli et al., 2022). Embora o agrupamento seja uma tarefa não supervisionada, o conhecimento do especializado no domínio dos dados pode ser incorporado em várias etapas do processo.

Figura 2 - Etapas do processo de agrupamento.



Fonte: Faceli et al., 2022

Tendo em vista este cenário, o presente trabalho tem como objetivo explorar o uso de métodos de clusterização para segmentar grupos de indivíduos com diferentes perfis de risco para obesidade. A partir dessa segmentação, espera-se que seja possível desenvolver propostas com o intuito de promover estratégias de intervenção eficazes, que levem em consideração as características específicas de cada grupo de risco, contribuindo para a redução das taxas de obesidade.

Este trabalho está dividido em quatro seções principais: Fundamentação Teórica, Metodologia, Resultados e Conclusão. Na Fundamentação Teórica, são apresentados os conceitos e técnicas utilizados para a análise de dados, como os algoritmos de identificação de dados incompletos ou nulos, além de uma abordagem sobre métodos, com destaque para a técnica de Agrupamento (*Clustering*), que visa identificar grupos com características semelhantes nos dados. Na Metodologia, detalha-se o uso de um banco de dados populacionais americanos da plataforma *Kaggle*[1], focado em padrões alimentares, hábitos de transporte e atributos ligados à obesidade. Na sequência, os resultados obtidos com a aplicação das técnicas analíticas são apresentados, como gráficos categóricos e a matriz de correlação individuais. Por fim, na Conclusão, é discutido como a mesma

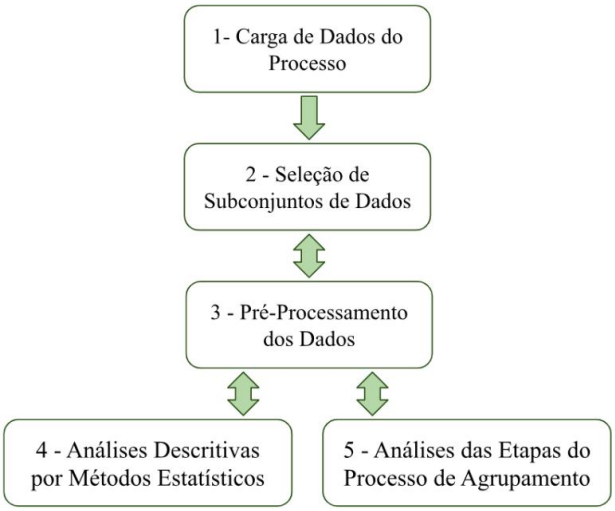
[1]https://www.kaggle.com/competitions/playground-series-s4e2/data?select=test.csv

abordagem pode ser aplicada a dados brasileiros para análise de saúde pública.

2. METODOLOGIA

A metodologia deste trabalho é composta por cinco etapas, conforme ilustrado pelo fluxograma na Figura 3. Além disso, vale destacar que o desenvolvimento do trabalho foi realizado no ambiente virtual do Google Colab utilizando a linguagem de programação Python, com o apoio das bibliotecas *numpy, pandas, sklearn* e *matplotlib.*

Figura 3 - Etapas metodológicas do trabalho



Fonte: Adaptado de Pais, 2020.

Foi utilizado dados disponíveis na plataforma de ciência de dados *Kaggle*, que permite a exploração e o compartilhamento de conjuntos de dados e modelos de aprendizado de máquina. O banco de dados escolhido, intitulado "*Multi-Class Prediction of Obesity*", contém informações populacionais de cidadãos americanos, sendo o foco principal a identificação de padrões de comportamento alimentar, hábitos de transporte e fatores genéticos que possam estar relacionados ao aumento da obesidade. Tal escolha se deu, uma vez que não não há um banco de dados sobre risco de obesidade no Brasil no *Kaggle* e, embora os dados sejam de origem americana, essa mesma abordagem pode ser aplicada a dados populacionais brasileiros.

O banco de dados é dividido em dois arquivos CSV, um contendo a *feature* de classificação "NObeyesdad" e o outro sem. O *dataset* de treino contém 20.757 registros distribuídos em 18 colunas, enquanto o de teste tem 13.840 registros distribuídos em 17 colunas. A coluna omitida nos dados de teste é a da variável-alvo.

As variáveis deste conjunto de dados, conhecidas como *features,* exibem informações detalhadas sobre os entrevistados, como hábitos alimentares e características pessoais. Essas *features* são apresentadas a seguir, bem como suas descrições.

- **ID do Usuário**: Identificador único para cada pessoa, mas irrelevante para os objetivos deste estudo.
- **Gênero**: Indica o gênero do indivíduo.
- **Idade**: Refere-se à idade, variando entre 14 e 61 anos.
- **Altura**: Medida em metros, variando de 1,45m a 1,98m.
- **Peso**: Peso corporal, com variações entre 39kg e 165kg.
- **Histórico Familiar com Sobrepeso**: Indicador binário (sim ou não) para histórico de sobrepeso na família.
- **FAVC**: Consumo frequente de alimentos com alta caloria (sim ou não).
- **FCVC**: Frequência de consumo de vegetais (sim ou não).
- **NCP**: Número de refeições principais por dia, variando entre 1 e 4.
- **CAEC**: Consumo de alimentos entre refeições, com quatro opções: Às vezes, Frequentemente, Nunca e Sempre.
- **SMOKE**: Indicador de hábito de fumar (sim ou não).
- **CH20**: Consumo diário de água, variando de 1 a 3 litros.
- **SCC**: Monitoramento do consumo de calorias (sim ou não).
- **FAF**: Frequência de atividade física, variando de 0 (nenhuma) a 3 (intensa).
- **TUE**: Tempo de uso de dispositivos tecnológicos, variando de 0 a 2.
- **CALC**: Consumo de álcool, com três opções: Às vezes, Não, e Frequentemente.
- **MTRANS**: Meio de transporte utilizado, com cinco opções: Transporte público, Automóvel, Caminhada, Motocicleta e Bicicleta.
- **NObeyesdad**: Representa sete grupos: Peso Insuficiente, Peso Normal, Sobrepeso Nível I, Sobrepeso Nível II, Obesidade Tipo I, Obesidade Tipo II e Obesidade Tipo III.

Foi realizado o pré-processamento dos dados a fim de tratá-los antes de aplicar o método de clusterização. Por exemplo, foi verificada a presença de valores nulos ou ausentes na base de dados. Além disso, foi analisada a distribuição da variável-alvo com as demais *features* do banco de dados através de gráficos de barra, de dispersão e de uma matriz de correlação, para observar as interações entre os *features*.

Para representar o banco de dados de maneira bidimensional e poder visualizar os grupos da variável alvo, foi feita a análise dos Componentes Principais (PCA). Trata-se de uma técnica estatística utilizada para reduzir a dimensionalidade de um conjunto de dados enquanto preserva o máximo de variância possível. A ideia principal do PCA é transformar um conjunto de variáveis possivelmente correlacionadas em um novo conjunto de variáveis não correlacionadas, chamadas de componentes principais.

Primeiro, o número de clusters (K) é definido. Esse número precisa ser escolhido pelo utilizador do modelo e representa quantos grupos o algoritmo deve formar. O algoritmo começa selecionando K pontos aleatórios do conjunto de dados como os centroides iniciais. Os centroides são os "centros" de cada cluster e não precisam ser necessariamente pontos existentes nos dados. Cada ponto de dado é atribuído ao cluster cujo centroide está mais próximo, geralmente calculado usando a distância euclidiana. Isso significa que os pontos são atribuídos ao cluster baseado em qual centroide está mais perto. Após a atribuição de todos os pontos, os centroides de cada cluster são recalculados. O novo centroide de um cluster é a média de todos os pontos que pertencem a esse cluster. O processo de atribuição de pontos aos clusters e a atualização dos centroides é repetido até que os centroides não mudem mais significativamente, ou até que um número pré-definido de iterações seja alcançado. Este ponto de convergência significa que os clusters se estabilizaram.

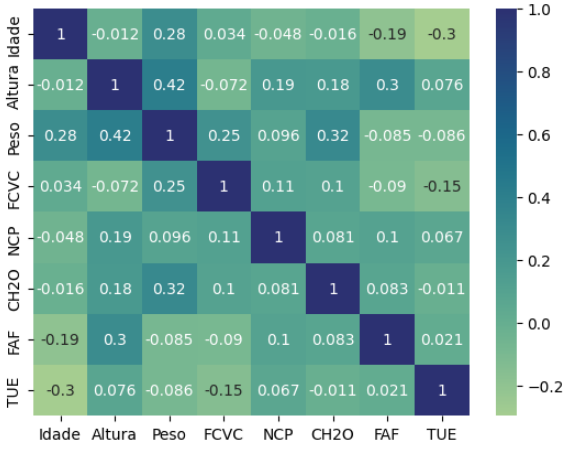
3. RESULTADOS

O Anexo 1 apresenta gráficos individuais, relacionando a variável alvo a cada um dos *features*. Os dados são relativos ao *dataset* de treino. O Anexo 1 (a) é relacionado ao Sexo, onde é possível ver a distribuição da variável-alvo por gênero, entre masculino e feminino. Percebe-se que pessoas classificadas com Obesidade Tipo II são exclusivamente homens, enquanto as classificadas com Obesidade Tipo III são exclusivamente mulheres. Isso é um indicativo de que Sexo é um *feature* relevante. O Anexo 1 (b), é relacionado ao Histórico Familiar com Sobrepeso. Percebe-se que a maior parte dos registros de obesidade possuem familiares com o mesmo problema de saúde, enquanto a maioria das pessoas fora da faixa de obesidade, não possuem registros dessa condição na família. O Anexo 1 (c), é relacionado ao Consumo Frequente de Alimentos com Alto teor Calórico. É nítido que a maior parte das pessoas registradas no dataset possuem uma dieta baseada em alimentos de alta caloria. Posteriormente, no Anexo 1 (d), relacionado ao Consumo de Alimentos Entre Refeições, percebe-se que esse hábito está presente no cotidiano da maior parte das pessoas na faixa de obesidade.

Posteriormente, o Anexo 1 (e), (f), (g) e (h) evidenciam que o hábito de fumar não está presente no cotidiano da maior parte das pessoas na faixa de obesidade. Além disso, é notável que a maior parte delas não possui o hábito de monitorar o consumo de calorias com um nutricionista, realiza consumo de álcool às vezes e utilizam, majoritariamente, transporte público.

A matriz de correlação apresentada na Figura 4 explora a relação entre *features* que não foram explorados nos gráficos anteriores, e demonstra que a maior correlação do *dataset* é entre altura e peso dos indivíduos, com um valor de 0,42. Além disso, a correlação entre idade e peso é de 0,28, sugerindo que, conforme a idade aumenta, o peso tende a aumentar de maneira moderada. Destaca-se ainda a correlação positiva de 0,32 entre o consumo de água (CH2O) e o peso dos indivíduos. Além das correlações negativas entre idade e frequência de atividade física (FAF), com -0,19, e entre idade e tempo de uso de eletrônicos (TUE), com -0,3.

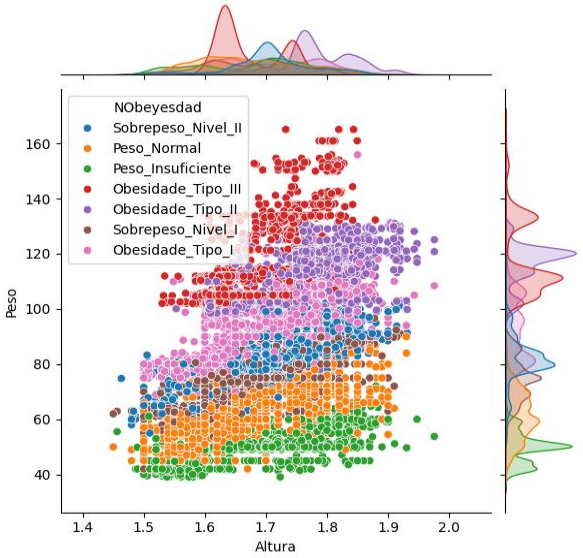
Figura 4 - Gráficos Categóricos Individuais



Fonte: Elaborada pelos autores.

O gráfico de dispersão com distribuições marginais da Figura 5 mostra a relação entre altura e peso, evidenciando a correlação de 0,42 discutida anteriormente. As categorias da variável-alvo (peso insuficiente, normal, sobrepeso e obesidade, com seus diferentes níveis) estão bem segmentadas, com pessoas de peso insuficiente e normal concentradas na faixa de altura entre 1,5 e 1,9 metros e com peso variando de 40 a 80 kg. Já aqueles classificados com sobrepeso ou obesidade estão distribuídos principalmente acima de 80 kg, com alturas que variam de 1,6 a 2,0 metros. A maior concentração de indivíduos obesos ocorre entre 100 e 160 kg, destacando como o peso aumenta de forma desproporcional em relação à altura na classe de Obesidade Tipo III.

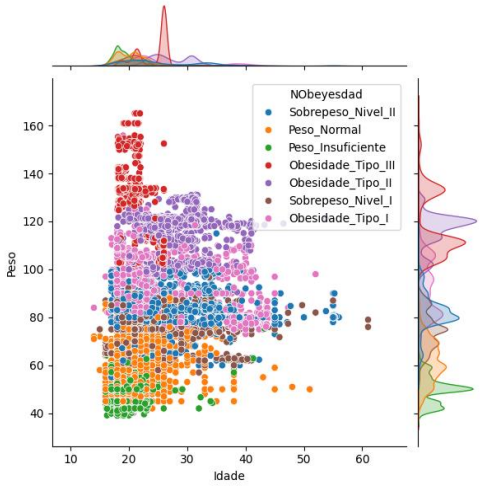
Figura 5- Gráfico de Dispersão (Altura x Peso)



Fonte: Elaborada pelos autores.

O gráfico de dispersão com distribuições marginais da Figura 6 mostra a relação entre idade e peso, evidenciando a correlação de 0,28 discutida anteriormente. Enquanto indivíduos mais jovens estão concentrados nas categorias de peso normal e insuficiente, à medida que a idade aumenta, há uma prevalência maior de sobrepeso e obesidade, com indivíduos atingindo pesos acima de 120 kg nas idades mais avançadas. Vale destacar que indivíduos com Obesidade Tipo III vão contra esse padrão, sendo mais frequente em pessoas na faixa dos 25 anos.

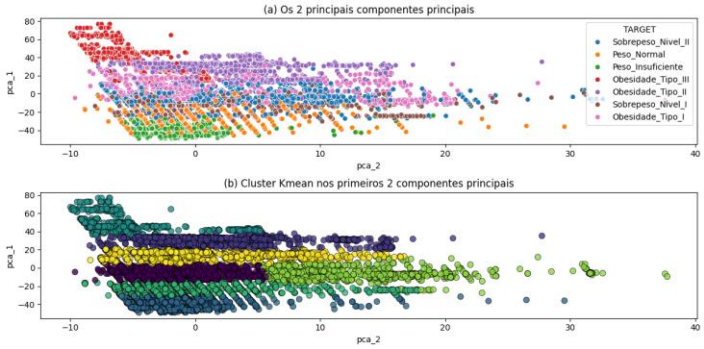
Figura 6 - Gráfico de Dispersão (Idade x Altura)



Fonte: Elaborada pelos autores.

A Figura 7 (a) foi feita a partir da Análise de Componentes Principais (PCA), que reduziu a dimensionalidade dos dados para facilitar a visualização, mantendo o máximo de variância possível. Aqui, os dois primeiros componentes principais (pca_1 e pca_2) são representados nos eixos. Esse gráfico foi feito com a base de dados que possui a variável alvo, então é possível ver perfeitamente os grupos de obesidade. Já na Figura 7 (b), foi aplicado o algoritmo de clusterização K-means nos dois primeiros componentes principais. Aqui, os pontos são coloridos de acordo com o cluster ao qual o algoritmo K-means atribuiu cada ponto. Foi utilizado k = 7 para tentar se aproximar o máximo possível das classes da variável alvo.

Figura 7 - Análise de Componentes Principais e K-Means



Fonte: Elaborada pelos autores.

Percebe-se que, embora o K-means tenha formado grupos visuais distintos, eles não correspondem exatamente às classes reais de peso da Figura 7 (a). Isso sugere que, embora os clusters encontrados pelo K-means tentem agrupar os indivíduos com base em semelhanças nos componentes principais, os grupos obtidos não coincidem perfeitamente com as classes de peso pré-definidas. Isso pode indicar que os dados têm uma estrutura complexa, que o K-means não capturou perfeitamente.

4. CONCLUSÃO

Este estudo explorou um banco de dados de obesidade em uma população norte-americana, retirado da plataforma *Kaggle*. A análise incluiu a geração de gráficos e uma matriz de correlação, que permitiram identificar relações importantes entre variáveis como sexo, histórico familiar com sobrepeso, consumo de alimentos com alto teor calórico, e uso de transporte público.

A maior correlação encontrada no dataset foi entre altura e peso, com valor de 0,42, seguida pela correlação entre idade e peso, com 0,28. Observou-se também uma correlação moderada entre o consumo de água e o peso dos indivíduos, com valor de 0,32. Além disso, ocorrem correlações negativas entre idade e frequência de atividade física, com -0,19, e entre idade e tempo de uso de eletrônicos, com -0,3.

Quanto à clusterização, foi possível perceber que o K-means não obteve um bom desempenho, pois criou *clusters* visualmente diferentes dos grupos originais. Para trabalhos futuros, a aplicação de técnicas avançadas de machine learning, como redes neurais profundas, pode oferecer uma visão mais detalhada e precisa sobre os fatores que influenciam a obesidade. Essa abordagem, além de proporcionar uma compreensão mais aprofundada dos dados populacionais, poderá ser adaptada para o contexto brasileiro, auxiliando no desenvolvimento de políticas públicas de saúde mais eficazes.

5. REFERÊNCIAS

[1] BRASIL. Ministério da Saúde. Saúde Brasil 2018: uma análise da situação de saúde e das doenças e agravos crônicos: desafios e perspectivas. Brasília: Ministério da Saúde, 2018.

[2] FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. 3. ed. Rio de Janeiro: LTC, 2022.

[3] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37-54, 1996.

[4] HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. 3. ed. San Francisco: Morgan Kaufmann, 2011.

[5] MINISTÉRIO DA SAÚDE. *Portaria nº 1.970, de 20 de setembro de 2018*. Diário Oficial da União, Brasília, DF, 20 set. 2018.

[6] MINISTÉRIO DA SAÚDE. *Pesquisa Nacional de Saúde, 2020*. Brasília: Ministério da Saúde, 2022.

[7] SOCERJ. *Obesidade e Doença Cardiovascular: Manual SOCERJ Completo*. Rio de Janeiro: SOCERJ, 2024.

[8] BREIMAN, Leo. Random forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.

[9] CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: ACM, 2016. p. 785-794.

[10] KE, Guolin et al. LightGBM: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*. Long Beach, CA: Neural Information Processing Systems Foundation, 2017. p. 30.

[11] LIAW, S.; WIENER, J. Classifiers for medical applications. In: PEARL, J.; WONG, A. (Eds.). *Proceedings of the Seventh Conference on Artificial Intelligence in Medicine*. Berlin: Springer-Verlag, 2002. p. 25-32.

[12] PROKHORENKOVA, L. et al. CatBoost: Gradient boosting with categorical features support. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal: Neural Information Processing Systems Foundation, 2018. p. 6639-6649.

[13] PAIS, Luis Fernando Reis Tavares. Análise descritiva do deslocamento de pacientes em tratamento de câncer de mama no SUS / Luís Fernando Reis Tavares Pais. - 2020.

[14] PESQUISA NACIONAL DE SAÚDE, 2020. Ministério da Saúde. Pesquisa Nacional de Saúde. Brasília, 2020.

[15] GOEBEL, M.; GRUENWALD, L. A survey of data mining and knowledge discovery software tools. ACM SIGKDD explorations newsletter, v. 1, n. 1, p. 20-33, 1999.

[16] SANTOS, R. S. et al. A data mining system for providing analytical information on brain tumors to public health decision makers. Computer methods and programs in biomedicine, v. 109, n. 3, p. 269-282, 2013

[17] KUSTERS, Daniel. Potencializando os efeitos do *Goal Priming: u*m estudo experimental sobre a influência de *diet reminders e activity equivalent labels* na redução do consumo calórico, 2016.

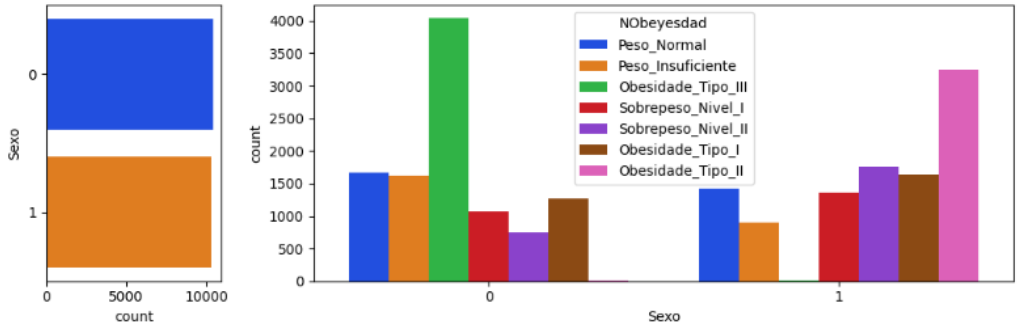
[18] VARQUES, A. C. J et al.. Influência do Excesso de Peso Corporal e da Adiposidade Central na Glicemia e no Perfil Lipídico de Pacientes Portadores de Diabetes Mellitus Tipo 2. Arq Bras Endocrinol Metab 2007.

[19] PEREIRA, L. O. et al. Obesidade: Hábitos Nutricionais, Sedentarismo e Resistência à Insulina. Arq Bras Endocrinol Metab vol 47 nº 2 Abril 2003.

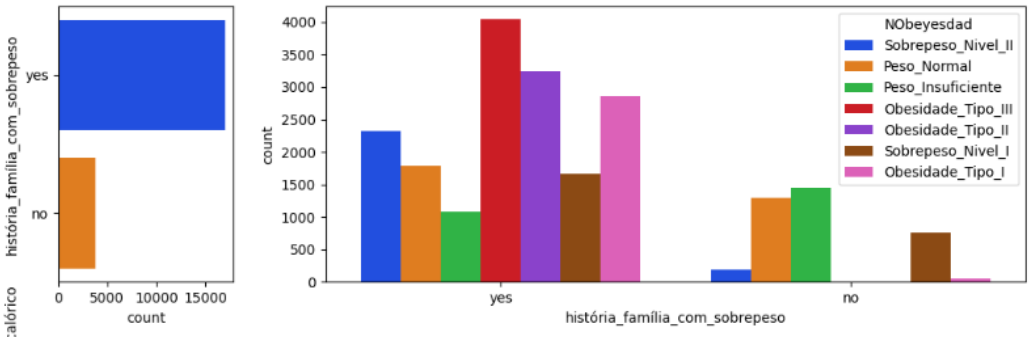
[20] MANAGO, M., and AURIOLl, M. 1996. Mining for OR. ORMS Today (Special Issue on Data Mining), February, 28–32.

ANEXO 1

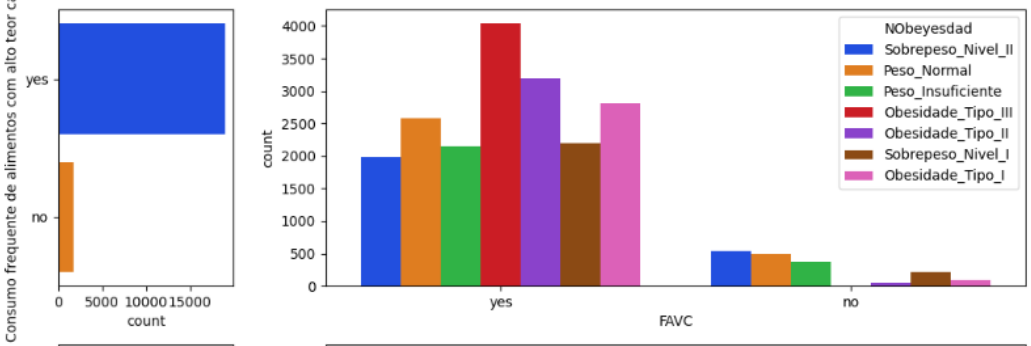
(a)


Sexo
0
1
0
5000
10000
count
NObeyesdad
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Sobrepeso_Nivel_I
Sobrepeso_Nivel_II
Obesidade_Tipo_I
Obesidade_Tipo_II
count
Sexo

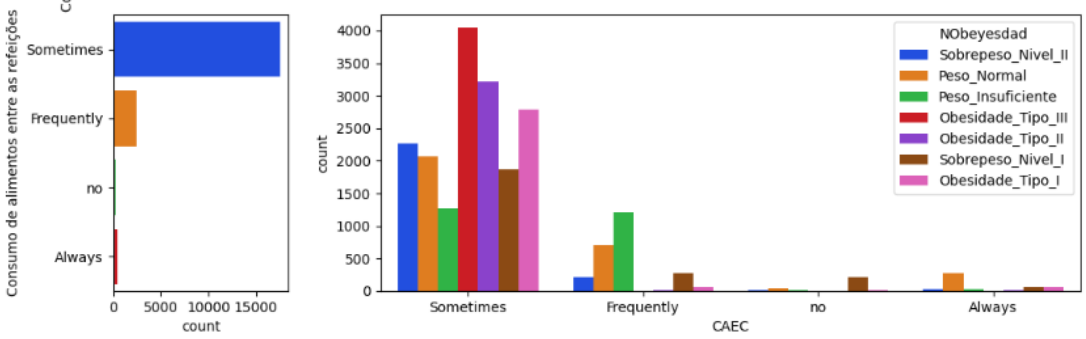
(b)


história_família_com_sobrepeso
yes
no
0
5000
10000
15000
count
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I
count
yes
no
história_família_com_sobrepeso

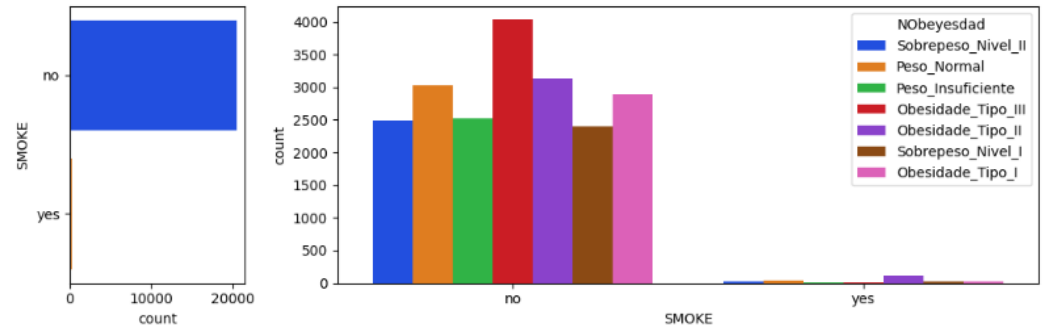
(c)


Consumo frequente de alimentos com alto teor calórico
yes
no
0
5000
10000
15000
count
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I
count
yes
no
FAVC

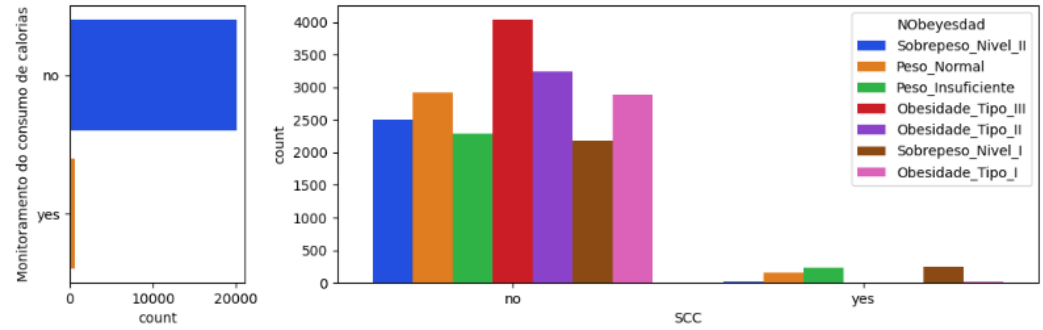
(d)


Consumo de alimentos entre as refeições
Sometimes
Frequently
no
Always
0
5000
10000
15000
count
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I
count
Sometimes
Frequently
no
Always
CAEC

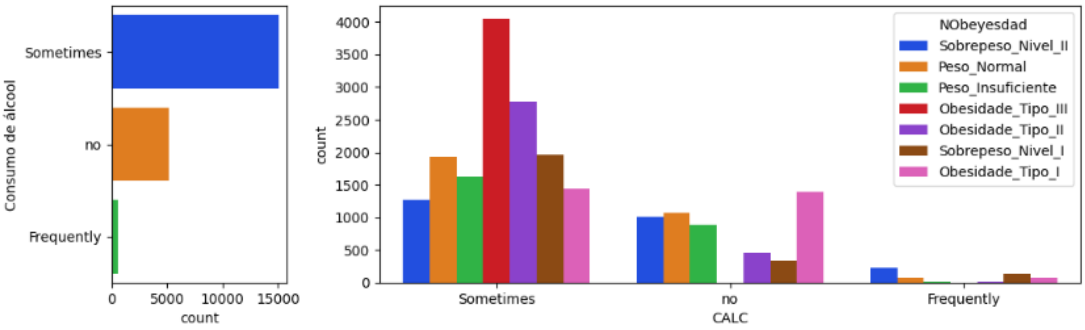
(e)


SMOKE
no
yes
0
10000
20000
count
count
0
500
1000
1500
2000
2500
3000
3500
4000
no
yes
SMOKE
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I

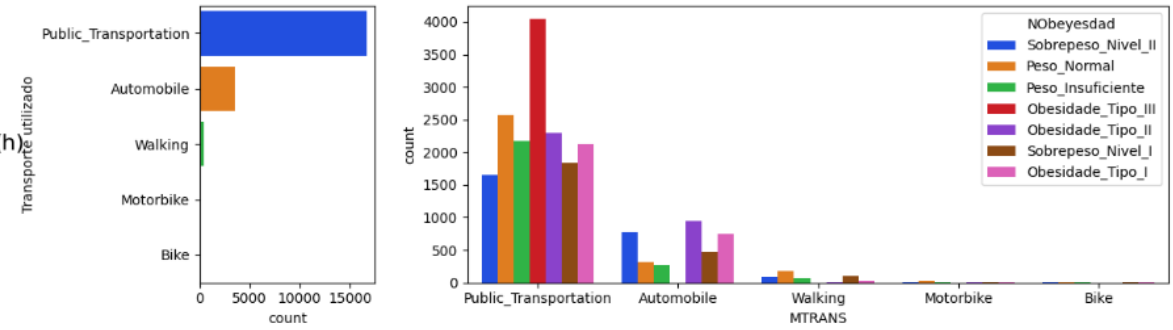
(f)


Monitoramento do consumo de calorias
no
yes
0
10000
20000
count
count
0
500
1000
1500
2000
2500
3000
3500
4000
no
yes
SCC
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I

(g)


Consumo de álcool
Sometimes
no
Frequently
0
5000
10000
15000
count
count
0
500
1000
1500
2000
2500
3000
3500
4000
Sometimes
no
Frequently
CALC
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I

(h)


Transporte utilizado
Public_Transportation
Automobile
Walking
Motorbike
Bike
0
5000
10000
15000
count
count
0
500
1000
1500
2000
2500
3000
3500
4000
Public_Transportation
Automobile
Walking
Motorbike
Bike
MTRANS
NObeyesdad
Sobrepeso_Nivel_II
Peso_Normal
Peso_Insuficiente
Obesidade_Tipo_III
Obesidade_Tipo_II
Sobrepeso_Nivel_I
Obesidade_Tipo_I