

Uso de Inteligência Artificial para Classificação do Nível de Obesidade

Amarildo Junior*

Kauê Otto[†]

Universidade Católica de Santa Catarina, Engenharia de Software

*amarildozej@gmail.com

[†]kaueottoclash@gmail.com

Resumo

O presente artigo mostrará em detalhes o uso de inteligência artificial para classificar o nível de obesidade que um indivíduo se encontra. Tal análise será a junção de informações de dados familiares e dados pessoais de tal indivíduo. A ferramenta tem o objetivo de fornecer informações precisas ao nível de ser usada para orientar o indivíduo sobre sua situação de saúde atual em relação ao peso.

Palavras-chave: Inteligência Artificial, Classificação de Obesidade, Saúde, Dados Familiares, Dados Pessoais, Prevenção de Doenças, Análise de Peso, Saúde Preventiva, Diagnóstico de Obesidade

Abstract

This article will show in detail the use of artificial intelligence to classify the level of obesity an individual is. Such analysis will be the combination of information from family data and personal data of that individual. The tool aims to provide accurate information to the level that can be used to guide the individual about their current health situation in relation to weight.

Keywords: Artificial Intelligence, Obesity Classification, Health, Family Data, Personal Data, Disease Prevention, Weight Analysis, Preventive Health, Obesity Diagnosis

1 Introdução

A conscientização sobre o consumo excessivo de alimentos em prol de proteger a saúde humana e aumentar a qualidade de vida é um assunto importante para todo ser humano, assegurando que somente a quantidade de nutrientes necessários sejam consumidos pelo organismo. O consumo adequado é uma condição básica para que o indivíduo tenha uma vida normal, prezando pelo bem-estar. Embora o planeta esteja se encaminhando com vários projetos para mitigar a desnutrição, temos também problemas com supernutrição. A obesidade é o fato da criação excessiva de tecidos adiposos que é a consequência do consumo excessivo de alimentos que geram energia e que não são gastos o suficiente comparado ao que foi ingerido. De acordo com a POF 2003/0312, cerca de 40% de pessoas adultas no Brasil estão com excesso de peso e destes 40%, dentre elas cerca de 8,9% dos homens e 13,1% das mulheres já são considerados obesos, em relação ao lado financeiro dos participantes da pesquisa, a tendência foi localizada em mulheres de baixa renda e para o lado masculino renda de média a alta. Levando em consideração os dados demográficos, o maior número de pessoas com excesso de peso são nas regiões mais desenvolvidas(Sul, Sudeste e Centro-Oeste) do país para o lado das rendas elevadas, mas também observa-se um aumento em casos nas regiões Norte e Nordeste para o lado das rendas mais baixas. Além do consumo excessivo de alimentos temos uma lista de doenças que podem causar a obesidade sendo elas: Hipotireoidismo, Síndrome de Cushing, Distúrbios do sono, Depressão e ansiedade. A obesidade traz

inúmeros problemas de saúde sendo elas: doenças cardiovasculares, diabetes, problemas ortopédicos, doenças respiratórias, impacto psicológicos, câncer e diminuição da expectativa de vida.

2 Base de Dados sobre Obesidade

Em relação ao levantamento de dados, foi realizada uma pesquisa online com cerca de 1.610 indivíduos para a criação do conjunto de dados. Ressaltamos que como a pesquisa foi feita on-line, não foram capturados dados demográficos. Em relação a lista de variáveis, temos:

Categoria	Opções
Sexo	1 - Masculino 2 - Feminino
Idade	Número
Altura	Número em CM
Familiar com obesidade	1 - Sim 2 - Não
Consumo de fast-food	1 - Sim 2 - Não
Frequência em consumo de vegetais	1 - Raramente 2 - Às vezes 3 - Frequentemente 4 - Sempre
Número de refeições por dia	1 - 1-2 2 - 3 3 - Mais de 3
Ingestão de alimentos fora de horário das refeições	1 - Raramente 2 - Às vezes 3 - Frequentemente 4 - Sempre
Fumante	1 - Sim 2 - Não

Ingestão de líquidos diariamente	1 - Menos de 1 litro 2 - Entre 1 litro e 2 litros 3 - Acima de 2 litros
Calcula a quantidade de calorias ingeridas	1 - Sim 2 - Não
Pratica exercícios físicos	1 - Não pratica exercício 2 - No raio de 1-2 dias 3 - No raio de 3-4 dias 4 - No raio de 5-6 dias 5 - Mais de 6 dias
Quantidade de tempo dedicado a tecnologia	1 - Entre 0 e 2 horas 2 - Entre 3 a 5 horas 3 - Excedente a 5 horas
Meio de transporte utilizado	1 - Automóvel 2 - Motocicleta 3 - Bicicleta 4 - Transporte público 5 - Andando

3 Metodologia e Algoritmos Utilizados

Para a classificação do nível de obesidade, utilizamos o algoritmo de aprendizado de máquina **Random Forest** com uma configuração de 100 árvores de decisão. Isso significa que o modelo gera 100 árvores de forma randômica durante o treinamento. A

decisão final é feita por votação entre essas árvores, onde cada uma contribui com uma “opinião” sobre a classificação através de estatística. Desta forma reduzindo o *overfitting* já que o algoritmo depende de um conjunto de arvores com diferentes amostras de dados.

Como técnica de redução de dimensionalidade, aplicamos o **SelectKBest**, que seleciona as 10 features mais importantes com base em critérios estatísticos. Essa técnica permite reduzir o número de variáveis de entrada, focando apenas nas mais relevantes.

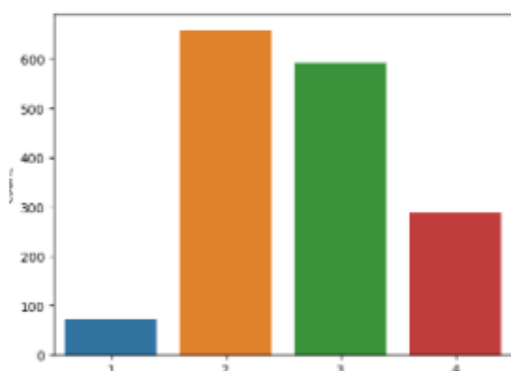
4 Processo de Treinamento e Validação

Inicialmente, realizamos o pré-processamento dos dados, verificando a presença de valores duplicados e nulos, além de avaliar a correlação entre as *features*. Após essa etapa, aplicamos o método *SelectKBest* para identificar as *features* mais relevantes para a classificação, selecionando aquelas com maior peso. Para o treinamento do modelo, utilizamos o protocolo experimental de validação cruzada com 5 dobras, o que possibilita uma avaliação mais precisa. Os hiperparâmetros ajustados no modelo *Random Forest* incluíram `class_weight='balanced'` que serve para tratar o desbalanceamento de classes e `random_state=42` para garantir a reprodutibilidade dos resultados.

5 Métricas de Avaliação

Para avaliação do modelo utilizamos F1-Score que é uma métrica usada para avaliar a performance de um modelo de classificação, especialmente em problemas desbalanceados, onde as classes não têm a mesma quantidade de amostras. No dataset de obesidade contém classes desbalanceadas sendo elas representadas por números.

- 1 - Abaixo do peso
- 2 - Normal
- 3 - Sobrepeso
- 4 - Obeso



que é expressada pela fórmula:

$$F1-Score = 2 \times \frac{Precisão \times Recall}{Precisão + Recall} \quad (1)$$

Onde a precisão mede a proporção de previsões corretas de uma classe positiva em relação ao total de previsões feitas para essa classe. Expressada pela fórmula:

VP = Verdadeiro Positivo

FP = Falso Positivo

FN = Falso Negativo

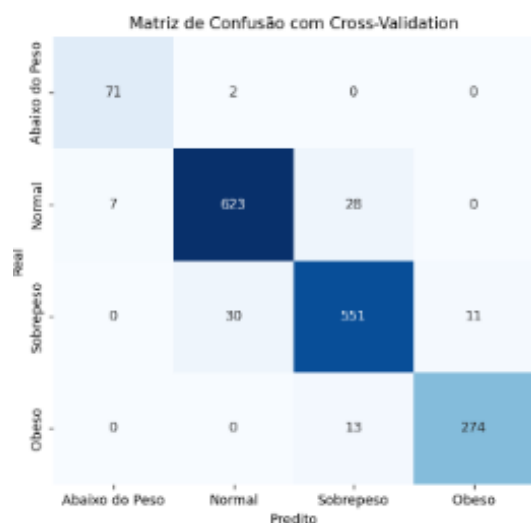
$$Precisão = \frac{VP}{VP + FP} \quad (2)$$

Já o recall mede a proporção de previsões corretas de uma classe positiva em relação ao total de casos que pertencem à classe positiva. Expressada pela fórmula:

$$Recall = \frac{VP}{VP + FN} \quad (3)$$

6 Discussão dos Resultados

Desenvolvemos 2 modelos utilizando RandomForest usando o protocolo experimental cross-validation de 5 com os mesmos hiperparâmetros, sendo assim a única diferença entre os modelos foi a quantidade de features selecionadas, um dos modelos foi feito com todas as features do dataset que o resultado medido pelo f1-score foi de 86%, já utilizando o SelectKBest buscando as 10 features mais relevantes que foram: Sexo, Idade, Familiar com obesidade, Consumo de fast-food, Frequência em consumo de vegetais, Número de refeições principais por dia, Fumante, Calcula a quantidade de calorias ingeridas, Pratica exercícios físico e meio de transporte utilizado. Através dessas features obtivemos o resultado medido pelo f1-score de 94%. Além do percentual vemos o resultado por classe com Matriz de confusão.



Referências

- [1] Organização Mundial da Saúde. "Título do Artigo/Estudo". Disponível em: https://www.scielo.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/csc/v15n1/a24v15n1.pdf

Onde temos no eixo X o valor predito pelo modelo e no eixo Y os valores reais do dataset.

7 Conclusão

Podemos concluir que a obesidade é um fator decisivo para a qualidade de vida quando se trata de facilidade em se locomover no dia a dia contendo grande chance de contrair doenças ósseas. Com base na gama de dados fornecidos que contém cerca de 23 variáveis informativas retiradas de uma pesquisa feita com 1600 pessoas, apresentamos o uso de técnicas de inteligência artificial para classificar o nível de obesidade das pessoas. A combinação do algoritmo Random Forest com a técnica de seleção de Features SelectKBest demonstrou ser eficaz para a melhoria dos resultados, melhorando o F1-Score de 86% para 94% ao reduzir as variáveis para as 10 mais relevantes.

