



**FACULDADE DE INFORMÁTICA DE PRESIDENTE
PRUDENTE
SISTEMAS DE INFORMAÇÃO**

GUILHERME SILVA HORTA

LÍVIA GOMES DE SOUZA

TIAGO PANTAROTO MENA

**TRABALHO DE CIÊNCIA DE DADOS:
APLICAÇÃO COM ÁRVORE DE DECISÃO - SISTEMA ESPECIALISTA**

**Presidente Prudente - SP
2023**

BASE DE DADOS UTILIZADA

A base de dados utilizada foi “Estimation of obesity levels based on eating habits and physical condition” (em tradução livre: Estimativa dos níveis de obesidade baseado nos hábitos alimentares e na condição física).

Link para a base de dados:
<https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition>

SOBRE A BASE DE DADOS

A base de dados contém perguntas relacionadas aos hábitos alimentares e físicos do entrevistado, além de sua condição física atual (idade, peso e altura).

A partir da condição física atual, é calculado o IMC (Índice de Massa Corpórea) e compara-se o resultado com os dados fornecidos pela OMS (Organização Mundial de Saúde).

Com base nos resultados, pode-se treinar o modelo para determinar o nível de obesidade sem fornecer o cálculo do IMC.

$$IMC = \frac{\textit{peso (kg)}}{\textit{altura (m)} \cdot \textit{altura (m)}}$$

A base de dados possui 17 atributos, 2111 registros e 1 classe de variáveis.

ATRIBUTOS

1. Gênero
 - a. Masculino
 - b. Feminino
2. Idade
3. Altura (em metros)
4. Peso (em kg)
5. Algum membro da família sofre ou já sofreu de sobrepeso?
 - a. Sim
 - b. Não
6. Você come comida calórica com frequência? (FAVC)
 - a. Sim
 - b. Não
7. Você costuma comer vegetais nas refeições? (FCVC)
 - a. Nunca (1)
 - b. Às vezes (2)
 - c. Sempre (3)

8. Quantas refeições principais você faz por dia? (*NCP*)
- a. Entre 1 ou 2 (1)
 - b. 3 (2)
 - c. Mais que 3 (3)
9. Você come no intervalo das refeições principais? (*CAEC*)
- a. Não
 - b. Às vezes
 - c. Frequentemente
 - d. Sempre
10. Você fuma?
- a. Sim
 - b. Não
11. Quanto você bebe de água por dia? (*CH2O*; *está como número decimal*)
- a. Menos de 1 litro (1)
 - b. Entre 1 e 2 litros (2)
 - c. Mais que 2 litros (3)
12. Você monitora diariamente as calorias ingeridas? (*SCC*)
- a. Sim
 - b. Não
13. Com que frequência por semana você pratica atividades físicas? (*FAF*; *está como número decimal*)
- a. Nunca (0)
 - b. 1 a 2 dias (1)
 - c. 2 a 4 dias (2)
 - d. 4 a 5 dias (3)
14. Quantas horas por dia você utiliza equipamentos eletrônicos (celular, videogame, televisão, computador, entre outros)? (*TUE*; *está como número decimal*)
- a. Entre 0 a 2 horas (0)
 - b. 3 a 5 horas (1)
 - c. Mais que 5 horas (2)
15. Com que frequência você bebe álcool? (*CALC*)
- a. Nunca
 - b. Às vezes
 - c. Frequentemente
 - d. Sempre
16. Qual meio de transporte você mais utiliza? (*MTRANS*)
- a. Carro
 - b. Moto
 - c. Bicicleta
 - d. Transporte público
 - e. A pé
17. Nível de obesidade (classe de variáveis)
- a. Abaixo do peso

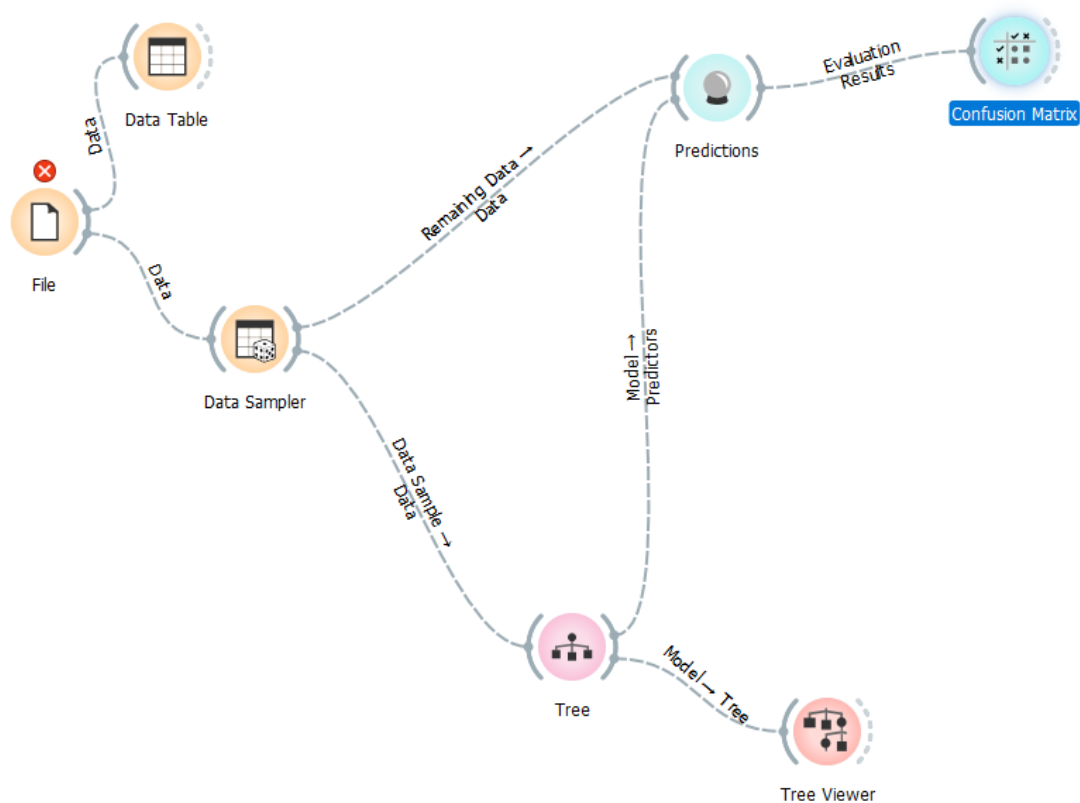
- b. Normal
- c. Sobrepeso Nível I
- d. Sobrepeso Nível II
- e. Obesidade Tipo I
- f. Obesidade Tipo II
- g. Obesidade Tipo III

REGISTROS

23% dos dados foram coletados diretamente dos usuários por meio de uma plataforma web e 77% foram gerados sinteticamente usando a ferramenta Weka e o filtro SMOTE.

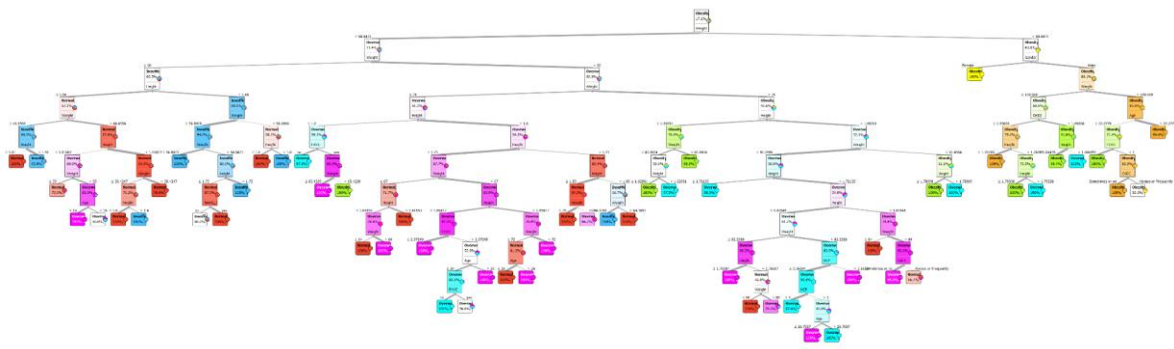
ÁRVORE DE DECISÃO E MATRIZ DE CONFUSÃO

Para gerarmos a Árvore de Decisão e a Matriz de Confusão, montamos a seguinte estrutura no programa *Orange*:



Como indicado nas instruções do trabalho, utilizamos 70% do conjunto de dados para treino e 30% para teste.

ÁRVORE DE DECISÃO



Disponibilizamos em formato .png a imagem da árvore para que possa ser vista com melhor qualidade.

PREDIÇÕES

Os resultados das predições foram os seguintes:

Model	AUC	CA	F1	Precision	Recall
Tree	0.976	0.953	0.953	0.954	0.953

ANÁLISE DOS RESULTADOS

AUC (Area Under The Curve): indica o quão exato o modelo é em sua classificação. Ou seja, o quão preciso é ao indicar a classe de um registro.

Quanto mais próximo de 1, melhor: nosso modelo alcançou uma pontuação de 0,976, indicando que é bastante preciso ao indicar o nível de obesidade de um registro.

CA (Classification Accuracy): a acurácia do modelo é dada pela fórmula $CA = \frac{\text{número de predições corretas}}{\text{número total de predições}}$.

Nosso modelo apresentou um CA de 0,953, o que é excelente, pois o valor máximo e melhor é 1.

Precision: semelhante à CA, a *Precision* mede quantas vezes o modelo acertou em relação ao total de vezes em que ele tenta acertar.

Recall: indica a quantidade de vezes em que o modelo acertou a classe em relação à quantidade de vezes em que deveria ter acertado.

F1: indica o quão *Precision* e *Recall* estão condizentes entre si. Ou seja, quanto maior for *F1*, melhor, pois *Precision* e *Recall* estão em níveis parecidos (não necessariamente alto ou baixo, mas, sim, próximos).

Para as três métricas acima (*Precision*, *Recall* e *F1*) o valor ideal é 1. Nas três, nosso modelo apresentou um valor muito próximo de 1, indicando que acertou a maioria das predições e, levando em consideração o que ele tenta acertar (nível de obesidade), podemos dizer que os resultados foram excelentes.

MATRIZ DE CONFUSÃO

		Predicted							Σ
		Insuffic...	Normal...	Obesity...	Obesity...	Obesity...	Overw...	Overw...	
Actual	Insuffic...	82	2	0	0	0	0	0	84
	Normal...	1	89	0	0	0	4	0	94
	Obesity...	0	0	86	2	0	3	2	93
	Obesity...	0	0	3	77	0	0	0	80
	Obesity...	0	0	0	0	92	0	0	92
	Overw...	0	6	0	0	0	88	1	95
	Overw...	0	0	1	0	0	5	89	95
Σ		83	97	90	79	92	100	92	633

Analisando a Matriz de Confusão, podemos observar que, de 633 registros testados, nosso modelo adivinhou corretamente 603 (soma dos valores na diagonal principal) e errou 30 (soma dos valores que não estão na diagonal principal).

PROGRAMA DESENVOLVIDO COM BASE NA BASE DE DADOS

Criamos um programa em *Python* utilizando o framework *Streamlit*. O programa roda no navegador através de um servidor local.

O objetivo do programa é determinar o nível de obesidade de uma pessoa com base numa série de perguntas.

As perguntas são as mesmas dos atributos da base de dados:

Peso (em kg):	Frequentemente consome alimentos de grandes calorias?
60,00 - +	Não ▾
Seu gênero:	Com que frequência consome vegetais?
Feminino ▾	Sempre ▾
Idade:	Quantas refeições principais por dia?
20 - +	Entre 1 e 2 ▾
Altura (em metros):	Você consome qualquer comida entre as refeições principais?
1,56 - +	Não ▾
Existe casos de obesidade em sua família?	Você fuma?
Não ▾	Não ▾

Como está seu consumo de água por dia?	Qual meio de transporte você geralmente usa?
Menos de 1L ▾	Carro ▾
Tem costume de monitorar o consumo diário de calorias?	
Sim ▾	
Com que frequência faz atividades físicas na semana?	
Não faço ▾	
Por quanto tempo você usa dispositivos digitais diariamente?	
De 0 a 2 horas ▾	
Sobre o consumo de bebidas alcoólicas:	
Não bebo ▾	

Tela inicial do programa:

Peso

Peso Normal

↑ 0 (nível de criticidade)

Parabéns 🎉

Você atingiu um nível saudável considerando suas características e hábitos. Lembre-se de que isso é apenas uma dedução com 93,4% de assertividade. Vale a pena sempre estar atento à sua saúde!

Está feliz com seu resultado? Vamos comemorar juntos!

Comemorar

Peso

Sobrepeso Nível I

↓ - nível de criticidade: 1

Atenção 📢 É necessário dar mais atenção à sua saúde.

Agora que você está ciente da sua situação de peso, é recomendável que você incorpore hábitos saudáveis para promover o bem-estar e cuidar da sua saúde de forma consciente.

Explore as opções disponíveis no menu lateral para criar combinações que levem ao peso normal. Isso permitirá que você identifique as mudanças necessárias e saiba como alcançar o peso recomendado. 😊

Peso

Sobrepeso Nível II

↓ - nível de criticidade: 2

Atenção 📢 É necessário dar mais atenção à sua saúde.

Agora que você está ciente da sua situação de peso, é recomendável que você incorpore hábitos saudáveis para promover o bem-estar e cuidar da sua saúde de forma consciente.

Explore as opções disponíveis no menu lateral para criar combinações que levem ao peso normal. Isso permitirá que você identifique as mudanças necessárias e saiba como alcançar o peso recomendado. 😊

Peso

Obesidade Tipo I

↓ - nível de criticidade: 3

Atenção 📢 É necessário dar mais atenção à sua saúde.

Agora que você está ciente da sua situação de peso, é recomendável que você incorpore hábitos saudáveis para promover o bem-estar e cuidar da sua saúde de forma consciente.

Explore as opções disponíveis no menu lateral para criar combinações que levem ao peso normal. Isso permitirá que você identifique as mudanças necessárias e saiba como alcançar o peso recomendado. 😊

Peso

Obesidade Tipo II

↓ - nível de criticidade: 4

Atenção 📢 É necessário dar mais atenção à sua saúde.

Agora que você está ciente da sua situação de peso, é recomendável que você incorpore hábitos saudáveis para promover o bem-estar e cuidar da sua saúde de forma consciente.

Explore as opções disponíveis no menu lateral para criar combinações que levem ao peso normal. Isso permitirá que você identifique as mudanças necessárias e saiba como alcançar o peso recomendado. 😊

Peso

Obesidade Tipo III

↓ - nível de criticidade: 5

Atenção 📢 As suas informações apresentam um quadro crítico para a sua saúde.

Agora que você está ciente da sua situação de peso, é recomendável que você incorpore hábitos saudáveis para promover o bem-estar e cuidar da sua saúde de forma consciente.

Explore as opções disponíveis no menu lateral para criar combinações que levem ao peso normal. Isso permitirá que você identifique as mudanças necessárias e saiba como alcançar o peso recomendado. 😊

O código foi desenvolvido a partir da Árvore de Decisão. Com base nela, fizemos vários *if-else* para determinar o nível de obesidade do usuário:

```
if i[3] <= 98.44:
    if i[3] > 60.00:
        if i[3] > 76:
            if i[2] > 1.69:
                if i[3] > 91.05:
                    if i[2] <= 1.78:
                        obesidade = "Overweight_Level_I"
                    else:
                        obesidade = "Obesity_Type_I"
                else:
                    if i[2] <= 1.72:
                        obesidade = "Overweight_Level_I"
                    else:
                        if i[2] <= 1.80:
                            if i[3] <= 82.22:
                                if i[2] <= 1.76:
                                    obesidade = "Overweight_Level_I"
                                else:
                                    if i[3] <= 80:
                                        obesidade = "Normal_Weight"
                                    else:
                                        obesidade = "Overweight_Level_I"
                            else:
                                if i[7] == '3 refeições':
                                    if i[7] == 'Mais que 3':
                                        obesidade = "Overweight_Level_II"
                                    else:
                                        if i[1] <= 26:
                                            obesidade = "Overweight_Level_I"
                                        else:
                                            obesidade = "Overweight_Level_II"
                        else:
                            obesidade = "Overweight_Level_II"
```

Printscreen de uma parte do código