

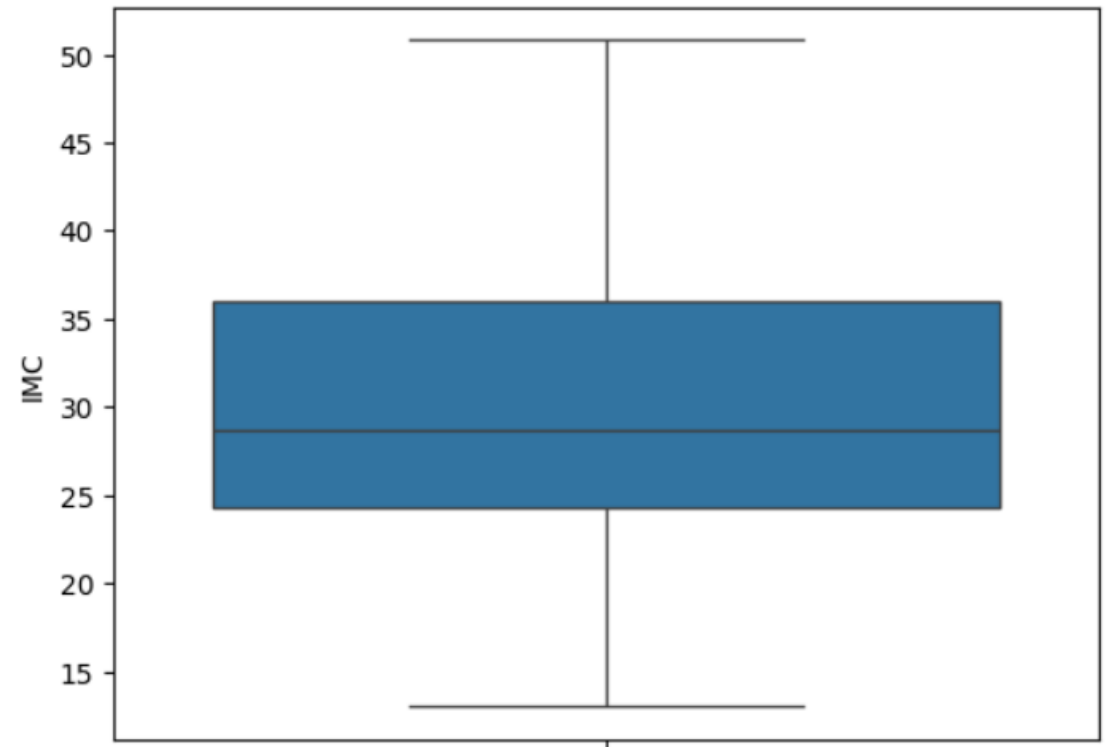
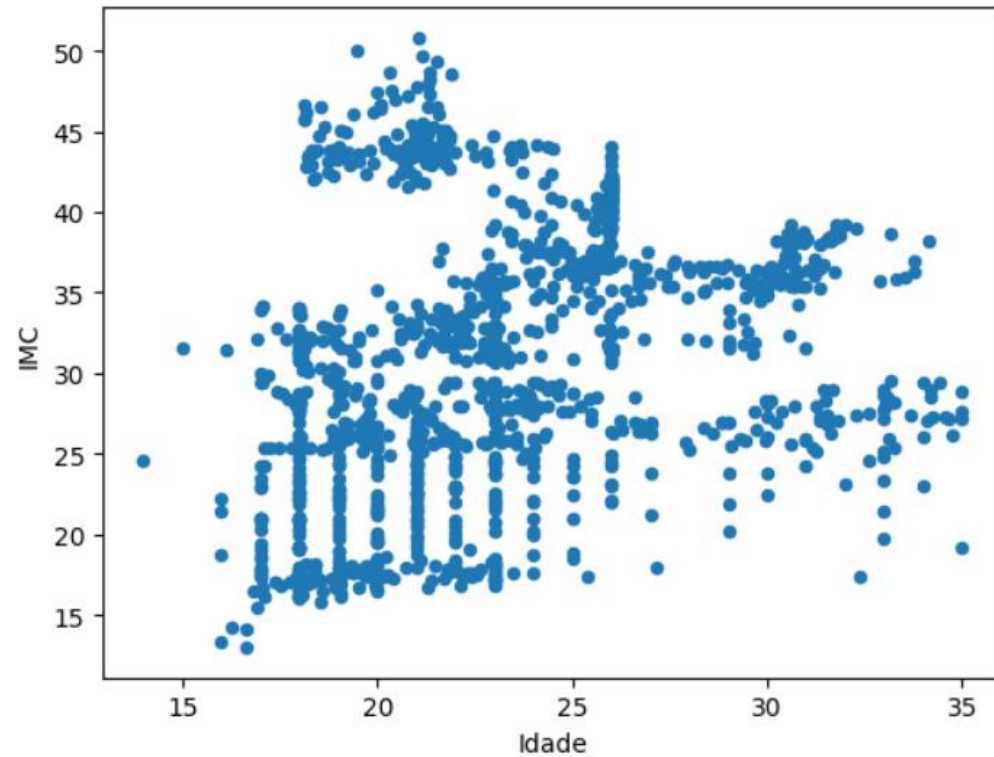
# Trabalho Prático 2

**Análise de Dados em Informática (ANADI)**

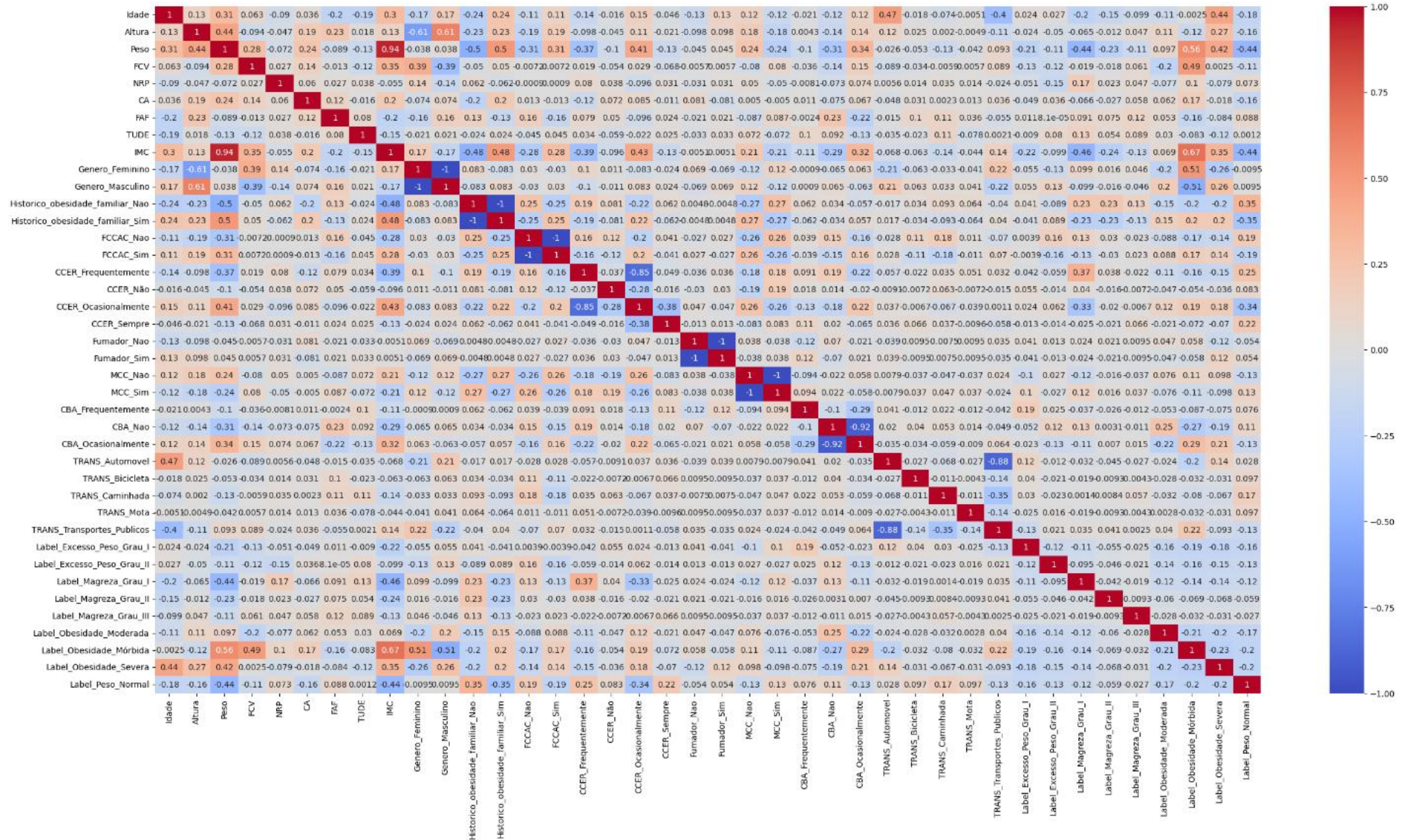
Aprendizagem Automática	Regressão	Classificação
Aprendizagem Supervisionada	Regressão Linear Simples e Múltipla	Árvores de Decisão
Aprendizagem Não Supervisionada	Árvore de Regressão	SVM
Aprendizagem Por Reforço	Redes Neuronal	Redes Neuronal
		K-Vizinhos-Mais-Próximos

# Regressão

## 2 Análise e Tratamento dos Dados



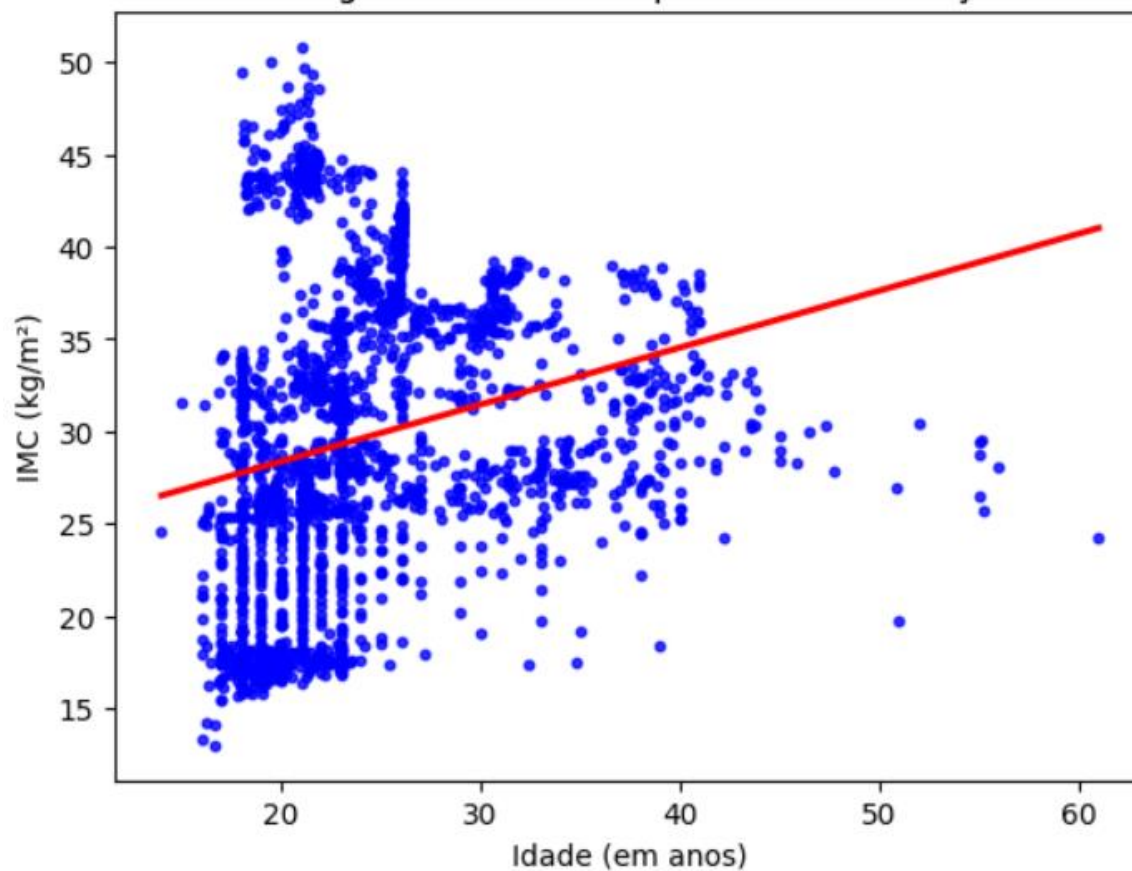
# Diagrama de Correlação





*Equação da reta:  $y = 0.3329 + 0.3320x$*

Diagrama de Dispersão com a reta correspondente ao modelo de regressão linear simples: IMC em relação à Idade



**MAE**

0.1670

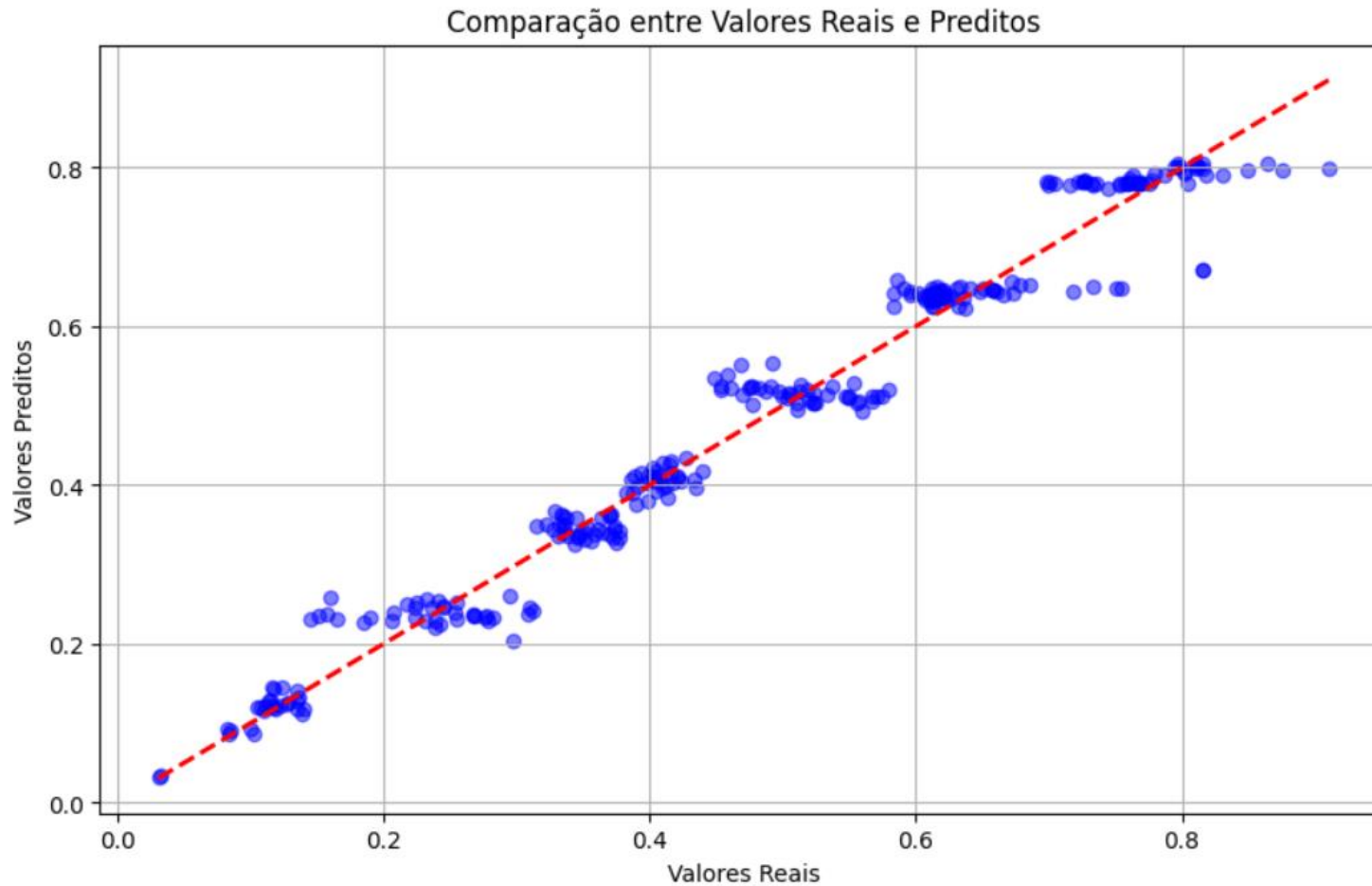
**RMSE**

0.2006

**Melhores variáveis que a Idade**

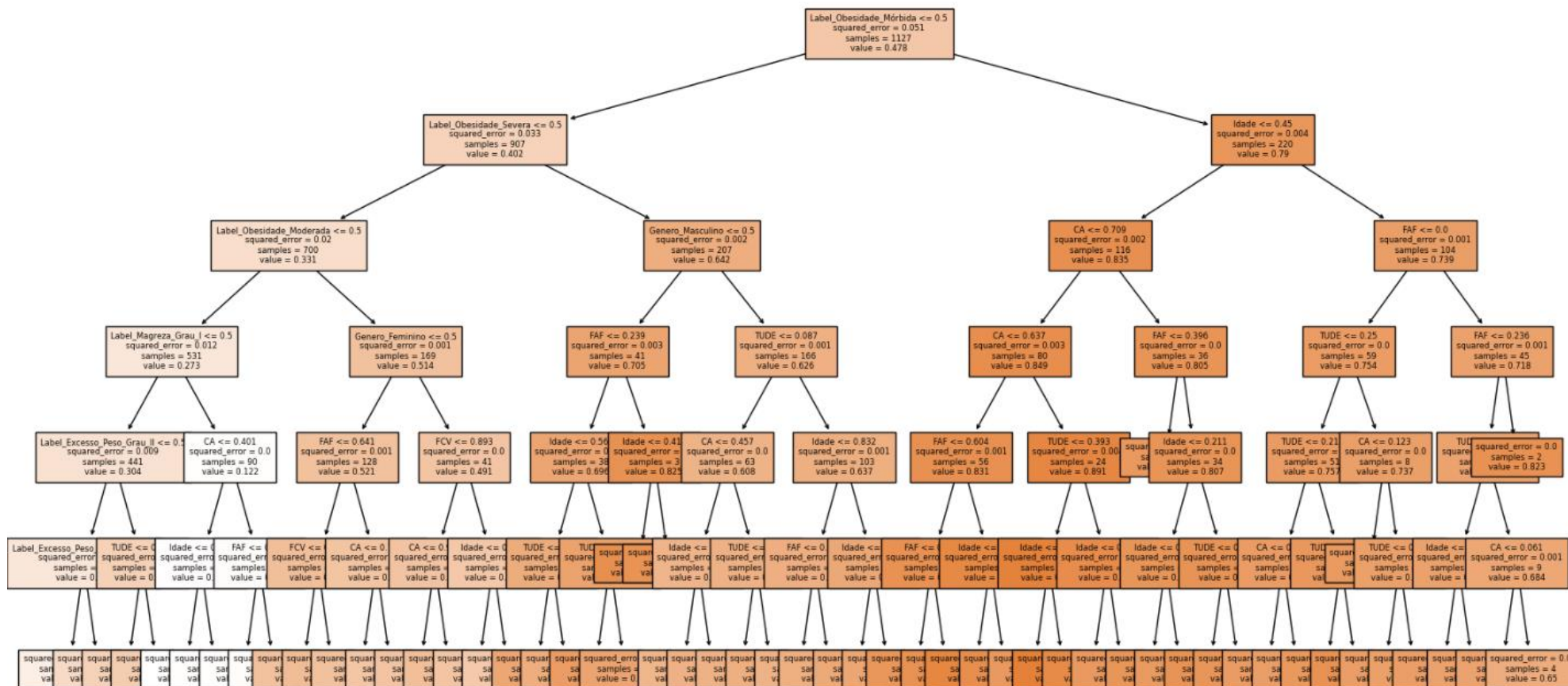
Peso

Frequência Cardíaca de Repouso (FCV)



**MAE**  
0.0280

**RMSE**  
0.0377



**MAE**  
0.0229

**RMSE**  
0.0351



**MAE**  
0.0296

**RMSE**  
0.0426

## 8 Escolha dos 2 Melhores Modelos

2	1	3
Regressão Linear Múltipla	Árvore de Regressão	Rede Neural
MAE 0.0280	MAE 0.0229	MAE 0.0296
RMSE 0.0377	RMSE 0.0351	RMSE 0.0426

Os dois melhores modelos são estatisticamente  
significativos?

Não, pelo Teste T-Student de Amostras Emparelhadas  
(p: 0.3245)

# Classificação

Árvore de Decisão	SVM	Rede Neuronal	K-Vizinhos-Mais- Próximos
Profundidade Máxima 8	Parâmetro de Regularização 10	Número de camadas 2	Métrica de Distância manhattan
Número Mínimo de Amostras por Folha 1	Coeficiente de Kernel 1	Nós na primeira e segunda camada 480 e 80	Número de Vizinhos 5
Número Mínimo de Amostras para Divisão 6	Tipo de Kernel rbf	Ativadores para a primeira, segunda e última camada Softmax, relu e softmax	Pesos dos Vizinhos distance

1

SVM

3

Rede  
Neuronal

2

K-Vizinhos-Mais-  
PróximosMédia  
0.5954Média  
0.7118Média  
0.6913Média  
0.6969Desvio-Padrão  
0.0789Desvio-Padrão  
0.1395Desvio-Padrão  
0.0221Desvio-Padrão  
0.1050

Existe diferença significativa no desempenho dos dois  
melhores modelos?

Não, pelo Teste T-Student de Amostras Emparelhadas  
(p: 0.4954)



Árvore de Decisão	SVM	Rede Neuronal	K-Vizinhos-Mais- Próximos
<b>Accuracy</b> 0.6454	<b>Accuracy</b> 0.7270	<b>Accuracy</b> 0.6135	<b>Accuracy</b> 0.7128
<b>Sensiblity</b> 0.6454	<b>Sensiblity</b> 0.7270	<b>Sensiblity</b> 0.5842	<b>Sensiblity</b> 0.7128
<b>Specificity</b> 0.9565	<b>Specificity</b> 0.9545	<b>Specificity</b> 0.9167	<b>Specificity</b> 0.8636
<b>F1</b> 0.6382	<b>F1</b> 0.7248	<b>F1</b> 0.5517	<b>F1</b> 0.7112

Atributo	Árvore de Decisão	SVM	Rede Neuronal	K-vizinhos-mais-prox
Idade	X	X	X	X
FCV	X	X	X	X
NRP	X	X	X	X
CA	X	X	X	X
FAF	X	X	X	X
TUDE	X	X	X	X
Género	X	X	X	X
Histórico obesidade familiar	X	X	X	X
FCCAC	X	X	X	X
CCER	X	X	X	X
Fumador	X	X	X	X
MCC	X	X	X	X
CBA	X	X	X	
Transporte	X	X	X	

- Densidade Calórica (DC);
- Tempo Ajustado de Atividade Física (TAAF);
- Hábito Alimentar Saudável (HAS);
- Índice do Hábito de Consumo Calórico (IHCC);
- Hábitos Diários Fit (HDF).

**SVM****Média**  
0.7054**Desvio-Padrão**  
0.0237**K-Vizinhos-Mais-  
Próximos****Média**  
0.6850**Desvio-Padrão**  
0.0269

Para um nível de significância de 5%, podemos afirmar que os novos atributos não melhoraram significativamente nenhum dos modelos.

1

Rede Neuronal

Média  
0.9241Desvio-Padrão  
0.0299

2

SVM (C=1000.0,  
Gamma=0.01, Kernel=rbf)Média  
0.8603Desvio-Padrão  
0.1146

3

SVM (C=100.0,  
Gamma=0.1, Kernel=rbf)Média  
0.8539Desvio-Padrão  
0.1227SVM (C=10.0, Gamma=0.1,  
Kernel=rbf)Média  
0.8475Desvio-Padrão  
0.1263

Existe diferença significativa no desempenho dos dois  
melhores modelos?

Não, pelo Teste T-Student de Amostras Emparelhadas  
(p: 0.3204)

Rede Neuronal	SVM (C=1000.0, Gamma=0.01, Kernel=rbf)	SVM (C=100.0, Gamma=0.1, Kernel=rbf)	SVM (C=10.0, Gamma=0.1, Kernel=rbf)
Accuracy 0.9184	Accuracy 0.9220	Accuracy 0.9255	Accuracy 0.9291
Sensiblity 0.9184	Sensiblity 0.9220	Sensiblity 0.9255	Sensiblity 0.9291
Specificity 0.9023	Specificity 0.9098	Specificity 0.9173	Specificity 0.9098
F1 0.9184	F1 0.9220	F1 0.9255	F1 0.9290



# Obrigado pela Atenção!

Gabriel Gonçalves, Tiago Leite e Francisco Bogalho