

Tema do Trabalho Grupo 4

Nosso trabalho se trata de um sistema de predição da probabilidade de desenvolvimento de doenças cardiovasculares baseada em histórico familiar, alimentação e estilo de vida do paciente, assim algumas colunas terão maior relevância na predição, outras não terão relevância alguma, dividiremos da seguinte forma:

Colunas Relevantes:

Idade (<20, 20-40, >40); [valores numéricos]
Tabagismo (fuma, já fumou, nunca fumou); [valores não numéricos]
Consumo de bebida em dias na semana (0, <2, 2-4, >4); [valores numéricos]
Índice de massa corpóreo; [valores numéricos] //pode ser null
Doenças auto-referidas; [valores não numéricos]
Número de refeições diárias (<3, 3-4, >4); [valores numéricos]
Dificuldades para atividades básicas (sim, não)
Existe histórico de doença cardíaca na família: (sim, não); [valores booleanos]

Colunas Irrelevantes:

RP (Registro do Paciente) [valor numérico];
Sexo; (F, M);
Renda;
Escolaridade;(nenhum, 1-5, 6-9, >9);

Para o cálculo da coluna de predição para gerar os casos de treinamento a ideia seria algo como:

1. Percorrer cada tupla gerada aleatoriamente e estabelecer pesos para cada coluna relevante, da seguinte forma:

Idade: peso 2 (para idade >40)
Tabagismo: peso 3 (para fuma); peso 2 (para já fumou)
Consumo de bebida em dias na semana: peso 3 (para >4); peso 2 (para 2-4)
IMC: peso 3 (< 20 ou > 24.9)
Doenças auto-referidas: peso 2 (se houver)
Número de refeições diárias: peso 2 (se <3 ou >4)
Dificuldades para atividades básicas: peso 3 (se sim)
Existe histórico de doenças cardíacas na família: peso 4 (se sim)

2. Soma-se os pesos de cada tupla e se esse valor for maior que 10 então o paciente será predito como alta probabilidade de desenvolver algum tipo de doença cardiovascular.