Amarildo J. F. de Lucena

[amarildolucena@gmail.com](mailto:amarildolucena@gmail.com)

[ajfl@cesar.school](mailto:ajfl@cesar.school)

[Diabetics prediction using logistic regression](https://www.kaggle.com/datasets/kandij/diabetes-dataset)

A opção pela base dados relacionado a predição de Diabéticos buscava analisar a relação das variáveis com o resultado positivo ou negativo para Diabetes.

Fizemos inicialmente uma verificação para existência de dados ausentes, que nos mostrou a inexistência desses dados. Optamos por descartar duas variáveis (pregnancies, skinThickness). Foi necessário realizar a conversão da coluna outcome para string, para ser utilizada como base para a análise final. O cenário no qual utilizamos somente a conversão para string, o resultado da previsão, apresentou uma acurácia acima 0.7 o que é considerado um bom resultado, no entanto, quando realizamos a conversão da coluna outcome para categórica, substituindo o valor 0 por 1, e o valor 1 por 2, e revertendo para string. O Knime gerou um alerta para normalizar os dados, que foi realizado em um novo processo. Esses tratamentos melhoraram o resultado da previsão, chegando a uma acurácia igual 1, melhor resultado possível.

Mapa com linhas pretas em fundo branco

Descrição gerada automaticamente

Mapa com linhas coloridas

Descrição gerada automaticamente

Interface gráfica do usuário

Descrição gerada automaticamente com confiança baixa