

Pima Indians Diabetes Prediction

O dataset *Pima Indians Diabetes Prediction* trata de dados preditivos sobre a ocorrência de diabetes sobre uma população, a de mulheres do grupo nativo americano de Pima e tem como objetivo estabelecer parâmetros que preveem a chance de um indivíduo ter diabetes.

Iniciei a análise desse conjunto de dados tentando determinar quais são os parâmetros que melhor representam uma pessoa que pode vir a ter diabetes, com isso pude determinar que esses eram:

- Glucose: foi escolhido por compor a glicemia, que é a taxa de açúcar no sangue, cuja taxa é maior em pessoas diabéticas;
- Insulina: cuja principal função é controlar a glicemia, sendo portanto um composto crucial em pessoas com diabetes;
- Histórico de diabetes: determina o histórico de pessoas que possuem diabetes na família do indivíduo
- Idade do indivíduo

Com esses atributos determinados foi montado um script que seleciona apenas esses parâmetros e monta um algoritmo de árvore de decisão, que conseguiu obter resultados muito próximos do esperado (coluna outcome), sendo esses:

- Acurácia: 0,942
- Precisão: 0.942
- F1 score: 0.941
- Revogação: 0.941