

# CIÊNCIA DE DADOS PRÁTICA: PREDIZER O SURGIMENTO DE DIABETES COM BASE EM MEDIDAS DE DIAGNÓSTICOS

Encontros Universitários 2018 - Quixadá

Francisco Thomás M. de Oliveira, Ana Victória Araújo Maia, Régis Pires Magalhães

Este artigo descreve a resolução de um problema onde o conjunto de dados (dataset) chama-se Pima Indians Diabetes e está disponível na plataforma Kaggle. O problema é apresentado e resolvido no canal do YouTube do projeto de extensão Ciência de Dados Prática, e disponibilizados documentos detalhando esta resolução no Github. O problema é solucionado usando um algoritmo de aprendizagem supervisionada conhecido como k-vizinhos mais próximos (k-nearest neighbors - k-NN) para classificar se uma determinada amostra do nosso conjunto de dados é de uma pessoa diabética ou não diabética. Ao final são aplicadas métricas de avaliação de modelos para medir o desempenho final do nosso modelo de predição.