



Data Science Academy

www.datascienceacademy.com.br

Big Data Analytics com R e Microsoft Azure Machine Learning

Projeto 2

Prevendo a Ocorrência de Câncer

O objetivo deste projeto é praticar o conhecimento adquirido ao longo dos capítulos anteriores.

A rotina de exame da mama permite que a doença seja diagnosticada e tratada antes de causar sintomas perceptíveis. O processo de detecção precoce envolve a análise de nódulos ou massas anormais do tecido da mama. Se um nódulo for encontrado, uma biópsia de aspiração com agulha fina é realizada, que utiliza uma agulha oca para extrair uma pequena amostra de células. Um clínico, em seguida, examina as células sob um microscópio para determinar se a massa é provável que seja maligna ou benigna. Se a aprendizagem de máquina puder automatizar a identificação de células cancerígenas, isso seria uma conquista importante para o sistema de saúde. Os processos automatizados permitem melhorar a eficiência do processo de detecção, permitindo que os médicos possam diagnosticar o problema, avaliar o gasto necessário para resolver e ainda definir o tempo necessário para o tratamento da doença. Um sistema de rastreamento automatizado também pode proporcionar uma maior precisão de detecção através da remoção do componente humano, inerentemente subjetivo, do processo. Iremos investigar a utilidade da aprendizagem de máquina para detectar o câncer, aplicando o algoritmo de classificação k-NN para medições de células provenientes de biópsias de mulheres com nódulos mamários anormais.

Usaremos dados reais, fornecidos pelo repositório de Machine Learning da UCI, neste link:

<http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

A solução será apresentada com todos os arquivos do projeto.

Equipe Data Science Academy