



Data Science Academy

www.datascienceacademy.com.br

Formação Cientista de Dados

Projeto com Feedback 7

Medicina Personalizada - Redefinindo o
Tratamento de Câncer

Muito tem sido dito durante os últimos anos sobre como a medicina de precisão e, mais concretamente, como o teste genético, vai provocar disrupção no tratamento de doenças como o câncer.

Mas isso ainda está acontecendo apenas parcialmente devido à enorme quantidade de trabalho manual ainda necessário. Neste projeto, tentaremos levar a medicina personalizada ao seu potencial máximo.

Uma vez sequenciado, um tumor cancerígeno pode ter milhares de mutações genéticas. O desafio é distinguir as mutações que contribuem para o crescimento do tumor das mutações.

Atualmente, esta interpretação de mutações genéticas está sendo feita manualmente. Esta é uma tarefa muito demorada, onde um patologista clínico tem que revisar manualmente e classificar cada mutação genética com base em evidências da literatura clínica baseada em texto.

Para este projeto, o MSKCC (Memorial Sloan Kettering Cancer Center) está disponibilizando uma base de conhecimento anotada por especialistas, onde pesquisadores e oncologistas de nível mundial anotaram manualmente milhares de mutações.

Neste projeto, você vai desenvolver um algoritmo de Aprendizado de Máquina que, usando essa base de conhecimento como uma linha de base, classifica automaticamente as variações genéticas.

O dataset completo pode ser encontrado em:

<https://www.kaggle.com/c/msk-redefining-cancer-treatment/data>

Mas também anexamos os arquivos para você. Recomendamos o uso de Python para criação de um modelo de Deep Learning com Keras.

Quando concluir o projeto, envie os scripts e datasets para projeto@dsacademy.com.br. Caso os datasets usados sejam muito grandes, armazene em um diretório virtual (existem vários na internet, como Google Drive ou Dropbox) e envie o link para que nossa equipe possa baixar os datasets. Se os arquivos foram pequenos (uma amostra do dataset original), envie no anexo junto com o script. Documente seu script tanto quanto possível.

Caso prefira, disponibilize seu projeto no Github e envie o link do seu repositório para nossa equipe no e-mail projeto@dsacademy.com.br. Nesse caso, o Readme do repositório deve constar que este trata-se de um projeto da Formação Cientista de Dados da Data Science Academy.

Em até 24 horas, daremos o feedback!

Bom trabalho!