PROJECT: PREDICTING NEURODEGENERATIVE DISEASES

Neurodegenerative diseases are a heterogeneous group of disorders that are characterized by the progressive degeneration of the structure and function of the nervous system. They are incurable and debilitating conditions that cause problems with mental functioning also called dementias.

Neurodegenerative diseases affect millions of people worldwide. Alzheimer's disease and Parkinson's disease are the most common neurodegenerative diseases. In 2016, an estimated 5.4 million Americans were living with Alzheimer's disease. An estimated 930,000 people in the United States could be living with Parkinson's disease by 2020.

The goal of this project is to build a model to accurately predict the presence of a neurodegenerative disease in an individual as early detection of a neurodegenerative disease could be useful for the identification of people who can participate in trials of neuroprotective agents, or ultimately to try and halt disease progression once effective disease-modifying interventions have been identified.

Fortunately, in the last couple of months, you have been acquiring the needed skill to build machine learning models in python. Now it is your time to put your skill to test!

For this task, you should do the following:

1. Obtain any neurodegenerative disease dataset of your choice from a reliable source. For example, UCI ML Parkinson's dataset. Perform exploratory data analysis on this dataset.

2. Highlight at least 3 important questions that you would like to address, which will provide information to the public and help to create awareness about neurodegenerative diseases.

3. Carry out a detailed data analysis to answer your questions above. Your analysis should be done in Jupiter notebook with sections, headings and comments that aid readability.

4. Build a model that accurately predicts the presence of a neurodegenerative disease in an individual using you dataset in (1) above. The python libraries; scikit-learn, numpy, pandas, and xgboost might be useful. Remember to evaluate the accuracy of your model.

5. Submit your complete notebook to your GitHub repository and provide the link here. Submission is due on or before August 7, 2020.

6. Write a blog on neurodegenerative diseases, your dataset, the analysis you performed and the model that you built, your findings and recommendations. Give your blog a unique title. The blog should be written here (https://www.datainsightonline.com/blog) with code from your Jupyter notebook embedded similar to Python: The Language for Data Science. The due date for completing the blog is August 14, 2020.