C964 QuickStart Guide

Alex Brown

11-2-2023

Project Installation:

1. Install either PyCharm IDE with professional/developer subscription, or alternatively you can download just the Jupyter notebook file and run it using the instructions found here: <https://docs.jupyter.org/en/latest/running.html>
2. If you chose to use the PyCharm IDE, you will need to download the file containing the project files from <https://github.com/serkeno/C964_Project> , which includes the cleaned excel file and Jupyter notebook interface.
   1. Once in the PyCharm IDE you will need to select File -> New Project and create a virtual environment with the default settings.
   2. Add the files from the Github link by copying the files and pasting them by right clicking on the main.py file, this will ensure that they are added to the proper folder.
   3. If you have not installed the libraries in Pycharm, you will need to install the following libraries:
      1. Pandas
      2. Sklearn
      3. Matplotlib
      4. Seaborn

Using the interface:

1. Open the Interface.ipynb file by double clicking it under the project folder on the left. Then you have three options for interacting with the interface.
   1. By changing the x\_input in the second cell to any of the named columns, you will be able to see how being part of that category affects the probability of a heart attack in the patient’s past, which will be displayed both as raw numbers and as a chart. Here is a picture of the variable you will need to change and an example of a graph it might produce: NOTE: Column names are capital letter sensitive and will only work if spelled out exactly as written, the sixth cell provides a short list of the excel file with all new columns created for reference. A screen shot of a computer

      Description automatically generatedA blue screen with white text

      Description automatically generatedA graph of a graph

      Description automatically generated with medium confidence
   2. The second option allows the inputting of a new record by setting each variable as a 0 or 1, and will then output whether the model believes the patient has had a heart attack in the past. In the case that it returns a 0, it is unlikely that they have or would have had a heart attack given their attributes, whereas a 1 indicates that they may have. If they have not had a known heart attack in the past, but still receive a 1, this indicates that others with similar attributes have had heart attacks before, and the patient may be at risk. The following is a picture of the inputs and outputA screenshot of a computer program

      Description automatically generatedA screenshot of a computer keyboard

      Description automatically generated Note: 0 indicating low risk of heart attack based on current features.

A screenshot of a computer

Description automatically generated

In either case, you can simply run all of the cells to initialize the program, and then you can rerun all of the cells any time you want to update, or the individual cell for the interface option you want to use.

* 1. The third type of interface interaction allows the user to input two dependent features and produce a bar chart that represents the percentage of each classification that belongs to the other group. For example, you could run “Sex”, and “Diabetes” to produce a chart showing the percent of males and females with diabetes.