ECS 171 One Page Writeup

Arielle Soomi Yoo, Samuel Robert Perelgut, Catherina Castillo, Huy Minh Tran, Mia Ma, William T Wu

April 10, 2021

Group 16 Goals:

The overall purpose of this project is to create a website that can assist medical professionals in diagnosing whether a patient is at risk for a heart disease. This problem is important because these kinds of judgments can save es by allowing us to predict what sort of life threatening issues may affect patients. This problem is interesting because it would give doctors a lot of extra information and help when making decisions on how to treat their patients. Further, a few past studies on the matter have shown that in some cases artificial intelligence programs have superior judgement when making these decisions then medical professionals do [5]. That is not to say that our program is superior to or a replacement for a doctor, but rather it is a reliable extra tool available to them that will help them ensure the best outcome for their patients. While this project focuses only on heart disease, future iterations of this project conducted by us or others will more than likely expand on this idea to encompass a wide variety of possible ailments. Indeed this the first step in reimagining how we, as a society, practice medicine.

Dataset:

The data for this project originally comes from The UCI Machine Learning Repository and contains data donated from four different sources:

1. The Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.,
2. The University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.,
3. The University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D., and
4. The V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:Robert Detrano, M.D., Ph.D. [1, 2].

In general, the 14 variables we will focus on include general background information on the patient, various cardiac health measurements taken from the patient, and the presence or absence of heart disease [2, 4]. The dataset includes information on 303 consecutive patients that were referred to the Cleveland Clinic for coronary angiography between May 1981 and September 1984 [2, 4]. From this data, we should be able to predict whether a patient is at risk of heart disease or not. In general, we will clean the data, removing rows with missing values, and split it into a test set and training set. We will download the dataset directly from the kaggle website, which has some more information on this UCI dataset [1].

Deliverables:

We're planning to create a web application for purposes of allowing a user to interact with our created model. The idea is that the web application will act as an interface where a medical professional can enter in prompted information such as patient's age, sex, cholestorol and more, and our website will output a prediction as to whether or not the patient is at risk for heart disease.

We also need an outline that describes our data, problem, and solution in more detail than is covered here. This can and should be worked on concurrently with the development of the webapp and model, but ultimately will probably be the last thing finished, so we can add the conclusion derived from the finished webapp and model.

Action Plan:

1. Assign roles and responsibilities – 4/12
    a. Roles may be fluid/subject to change
2. Clean/format the data (EDA) – 4/16
3. Frontend – 4/23
    a. Look for web development tools that work with Python
        i. Ex: Flask, nodejs
    b. How to merge frontend with backend
4. Devise algorithm – 4/23
    a. Compare different ML strategies
    b. Data visualization
    c. Evaluation
5. Make an API together and start coding – 4/23
    a. Agree on inputs and outputs of functions
    b. Adjust roles if needed

6. Testing/Test Cases – 5/17
    a. Backend people can probably test things on their own
    b. Frontend people can test button presses/etc
    c. Connecting: might have to test together
7. Final Outline – 5/26
    a. Can be worked on concurrently with some of project, however should be last completed after webapp is functional to our liking
8. Make project public on Github, and a video demo – 5/26

Citations for Data

1. 1. Main dataset: https://www.kaggle.com/ronitf/heart-disease-uci
2. 2. http://archive.ics.uci.edu/ml/datasets/Heart+Disease
3. 3. https://www.kaggle.com/ronitf/heart-disease-uci/discussion/101018
4. 4. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S.,Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64,304–310. https://www.sciencedirect.com/science/article/pii/0002914989905249?via
5. 5. Lysaght, T., Lim, H.Y., Xafis, V. et al. AI-Assisted Decision-making in Healthcare. ABR 11, 299–314 (2019). https://doi.org/10.1007/s41649-019-00096-0