
EE / CPRE / SE 491 - Weekly Report 06

12/01/2019 – 12/08/2019

Group number: sdmay20-41
Project title: Machine Learning for Understanding Aging
Client &/Advisor: Dr. Julie Dickerson
Team Members/Role:

- Ian Simon / Chief Engineer
- Jacob Laing / Chief Engineer
- Nathan Carter / Test Engineer
- Samantha Williams / Meeting Scribe
- Scott Rose / Meeting Facilitator
- Aria Sheets / Report Manager

Weekly Summary:

- This week we focused on extracting the data we believe will be beneficial to our project and inputting this data into our data tables. We also reflected on and updated our design document based on the advice we received from our TA, and then finished the design document for this semester.

Past Week Accomplishments:

- Ian Simon: Our advisor gave us three data sets that surveyed several sets of people for us to use as test data going forward with our project. We needed to start by finding the people who could be found in all three data sets. I was able to extract a little over a thousand unique people who existed in each report. I used Excel to accomplish this.
- Jacob Laing: Finished setting up the database so that the rest of the team can use MySQL workbench to connect and set up the database. I also spent time finishing up both the design doc and worked on the final presentation.
- Nathan Carter: I helped the team research and attempt to find a solution to our SQL columns issue. I also spent time reviewing and learning more about regression and plotly since those things are going to become relevant in the upcoming weeks.
- Samantha Williams: I extracted the columns of interest from our files. Once Ian was able to find which IDs appeared in all 3 files, I was able to further narrow down our data by extracting from Ian's file only the columns that we agreed to focus on. I also edited and updated the design document with some of the changes the TA suggested in our review meeting.
- Scott Rose: My main project for this week was completing the testing plan for our project. I researched testing frameworks and planned a testing cycle. On top of that, I worked on getting the MySQL database working. We needed to insert tables to hold some of the data, so I helped create the table.
- Aria Sheets: I worked on trying to find a solution to the column restriction that MySQL has. I then also worked with the rest of the team to try finding IDs that appear in all three

of the files (respective to three different projects we are looking at). I also spent time working on the final version of the design document, fixing formatting mistakes that were pointed out during our TA design document revision meeting.

Pending Issues:

- Ian Simon: We have a lot to do on the remaining data as far as summing specific columns or eliminating others. I hope to use what I know in excel to help the team accomplish this task.
- Jacob Laing: Right now I just have to continue helping finish the presentation. Other than that I can continue working on how the data will be stored in the database.
- Nathan Carter: I need to figure out how to use regression with our dataset. This should become clear with more discussion with our advisor/client and with more research.
- Samantha Williams: I need to go back and edit the data because we did not have a common decision on what columns to extract from the parent file. This file contains the information about the patient that we want to use as the input for our program.
- Scott Rose: I have to put the testing plan from the design document into the presentation that we are giving on our project..
- Aria Sheets: I need to find a way to deal with data and find a way to deal with both missing data and data that corresponds to a user choosing a “Not Applicable” or “N/A” option in a survey because this appears differently than normal data. I also need to set up MySQL workbench with our database.

Individual Contributions:

<u>Name</u>	<u>Individual Contributions</u>	<u>Hours this week</u>	<u>Hours Cumulative</u>
Ian Simon	Separating IDs.	2	29
Jacob Laing	Finished setting up the database so that the rest of the team can use MySQL workbench to connect and set up the database. I also spent time finishing up both the design doc and worked on the final presentation.	4	32
Nathan Carter	Helped with SQL column issue. Researched into plotly and regression with more of a focus on how to approach our datasets with those topics.	4	31+
Samantha Williams	Extracting columns of interest from the IDs that were present among all three files. Pushed these new files to git. Updated design document.	5	32
Scott Rose	Improved SQL tables, created a test plan for the project, and helped contribute to our final presentation	5	35

Aria Sheets	Find an alternative to the column limit issue. Helped find IDs that appeared only in all three projects. Updated and finished the design document.	5	40
-------------	--	---	----

Comments and Extended Discussion:

- We spent a lot of time finishing the design document and also creating our design document presentation that we will present on Thursday, December 12th.

Plans for the Upcoming Week:

- Ian Simon: I plan to work on the data sets in preparing them for us to start running tests on them. Also I will be contributing on the design document before our deadline on Sunday.
- Jacob Laing: Help rest of team with whatever needs done.
- Nathan Carter: I will begin to focus my research on plotly and regression down so that I can learn more on how to use regression with our dataset and visualize it with plotly. I will also be doing work on our design documents and presentation.
- Samantha Williams: This week I will extract the additional columns of interest from the parent file. I am also going to continue updating the design document and prepare for our presentation this week.
- Scott Rose: I will continue working on getting the data into the database, and help finish our presentation
- Aria Sheets: I will look at the data and find a way to deal with both missing data and data that corresponds to a user choosing a “Not Applicable” or “N/A” option in a survey. This usually appears as a “8” or “9” on a data that is supposed to be on a scale of 1-5 (“8” or “9” allows it to be differentiated from regular data). I will also look at the data to ensure it was inserted correctly into our database and make fixes/changes if I find issues.