E E / CPR E / S E 491 - Weekly Report 02
01/27/2020 – 02/10/2020

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:

- As a group, we worked individually on the various methods we needed for reading files from the user, parsing these files, and extracting the information of interest. We've separated our work thus far into four different directories: Utilities, Enums, DataStructure, and DataExtraction. This work almost completes the input module of the project. Testing is required to complete this module.

Past Week Accomplishments:

- Ian Simon: Worked on the processExtractedColumns method. This method takes a 2D array and converts it to a numpy array.
- Jacob Laing: Worked on creating a method that would take the 2d arrays and join them together into a single 2d array. The method would look for similar ids and combine that information together if they were present in every array. Began work on creating tests as well as going through some code reviews.
- Nathan Carter: Worked on several methods, ExractWantedColumns, calcColAvg, countMissingValues, Handle Missing data. Most of the code was dealing with missing data.
- Samantha Williams: Worked on getColumnProcessingParameters method. It creates ColumnParameter objects using the information in the parameters file provided by the user. These ColumnParameter objects are saved in an array. Also created the class ColumnParameters for this purpose; it contains information such as column name, minimum value, and maximum value. Continued to discuss project architecture and create cards on the trello board for project management purposes.
- Scott Rose: Created the testing framework for the project and began making unit tests for some of the components. Helped create the starting architecture for the projects and created skeleton code for other people to fill in. Created a way for us to parse files that had rows that needed to be merged together. Conducted code reviews for merge requests to our repository. Fixed bugs in the program.

- Aria Sheets: Worked on taking the codebooks and converting them into a parsable format, then traversing through them with a parser to create a file that fit the template we have for handling min/max values for specific columns both, then cleanup by manually comparing the codebook against the created TSV. Worked on code reviews for multiple merge requests. Created Slack channel and Trello board, organized team discussion.

Pending Issues:
- Ian Simon: Going to work on creating unit tests for the processExtractedColumns method.
- Jacob Laing: Continue testing more methods and code reviews.
- Nathan Carter: A few merge conflicts, should work itself out with minor reformatting of the new handleData code.
- Samantha Williams: Continue testing methods and altering as needed.
- Scott Rose: Create more unit tests for some of the components I made.
- Aria Sheets: Finish up some final testing and code cleanup before moving onto creating the machine learning module.

Individual Contributions:

Name	Individual Contributions	Hours this week	Hours Cumulative
Ian Simon	processExtractedColumns.	12	13
Jacob Laing	combineArraysOnM2id, some unit tests for DataExtractor, and code reviews.	14	15
Nathan Carter	ExractWantedColumns, calcColAvg, countMissingValues, Handle Missing data.	12	13
Samantha Williams	Created the ColumnParameters class and the method getColumnProcessingParameters. Began discussing testing techniques.	12	14
Scott Rose	Created Testing framework. Set up architecture so that the individual components made by other people worked together. Created a method for parsing files that had rows that needed to be merged together.	12	14
Aria Sheets	Worked on taking the codebooks and converting them into our template format. Worked on code reviews for multiple merge requests. Handle project organization via Slack and Trello. Create tests for FileUtilities.py.	13	17

Comments and Extended Discussion (Optional):

- We are finishing up some testing and will be working on the actual machine learning module. We feel confident enough in our ability to create this based on the information we have learned last semester.

Plans for the Upcoming Week:

- Ian Simon: Unit tests for processExtractedColumns
- Jacob Laing: Unit tests for DataExtractor and the method that combines arrays
- Nathan Carter: Finishing the data handler methods and creating unit tests for them.
- Samantha Williams: Works on the test cases for extractWantedColumns and creating the method to print out the metadata we want the user to see. Hopefully begin the machine learning module.
- Scott Rose: Help prepare the final parts of the data parsing components so that we can move on to the machine learning components.
- Aria Sheets: Works on testing the FileUtilities file to ensure quality.