E E / CPR E / S E 491 - Biweekly Report 03
02/10/2020 – 02/24/2020

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

Biweekly Summary:

- The past two weeks, we divided and conquered various elements in our program. We focused the first week on implementing tests to cover our various methods that we have developed. The following week, we worked on changing aspects of our project architecture to allow users to use plugins for processing columns in their files. We met as a group to discuss the changes, and the changes were then broken up into tasks for each member to complete on their own.

Past Two Weeks Accomplishments:

- Ian Simon: Worked on implementing tests for process extracted columns. Ensuring the numbers are read and normalized correctly. Second week I worked on reconfiguring parseTSV to use our new flags.
- Jacob Laing: Finished up writing tests for DataExtractor to ensure that the function would work as proposed. Also added a new parameter to allow the user to enter in the location of the column parameters file.
- Nathan Carter: Worked on bug fixing and changing the missingValHandle, calcColumnAvg, calcNumMissingValues, handleMissingData functions. Several of the functions needed to be heavily changed to allow more user input. I also began creating tests for these functions that are not yet complete.
- Samantha Williams: Worked on implementing test functions for extractWantedColumns method that takes in column parameters and returns only the data corresponding to the specified column names. Implemented MetaData singleton class so that each class can edit and append to the one MetaData object in the program. This meta data will serve as an output to the user and includes important information such as input parameters, files used, specified column names, what is done to the data, tensorflow version, and the reduced down data. Worked on testing for the MetaData class.
- Scott Rose: Worked on switching the architecture of our program so that the methods that process the data operate using a plugin structure. This allows users to switch

between how the data is processed. This also allows for other users to create their own processing methods. Created a processing plugin that processed data in a way that recreated the way data was processed in a study we are trying to recreate. Created tests for the components created during the two weeks. Conducted code review on code submitted to the code repository.

- Aria Sheets: Implementing a pipeline that will run on GitLab for the latest commit of a merge request or branch. This pipeline will run our test runner and make sure that new changes don't break old tests, allowing us to do continuous integration. The pipeline brought attention to some broken tests so there was work on fixing those tests up. Modified branch protection for the "master" branch so that developers couldn't push to master, they can only make new changes by making merge requests.

Pending Issues:

- Ian Simon: Continue work on parse tsv implementing the flags instead of the boolean values.
- Jacob Laing: Going through and making sure all test cases are covered when writing my tests while also cleaning them up so they're easier to read.
- Nathan Carter: There are a few pending issues this week with broken tests. We have only started looking at fixing these broken tests, so this issue should be resolved within a few days.
- Samantha Williams: Testing for the MetaDeta class is proving more complicated than anticipated. Looking into doing mock testing.
- Scott Rose: Code is waiting to be merged. It will be merged once a branch that is fixing failed tests is merged.
- Aria Sheets: Work on fixing up broken tests. It may take awhile based on the time it takes to understand why the tests are broken.

Individual Contributions:

Name	Individual Contributions	Past Two Weeks Hours	Hours Cumulative
Ian Simon	Test process extracted columns Configure parseTSV for flags use instead of booleans.	12	25
Jacob Laing	Implemented tests for DataExtractor and implemented a new parameter for the user to insert where their column parameter file is located	15	30
Nathan Carter	Heavily changed the calcColumnAvg, calcNumMissingValues, handleMissingData	14	27

	functions. Began creating tests for these functions and working on fixing broken tests.		
Samantha Williams	Implemented tests for extractWantedColumns. Implemented MetaData class and began MetaData tests.	13	27
Scott Rose	Converted data processing components to use a plugin architecture. Created a data processing plugin Created tests for components Conducted code reviews	15	29
Aria Sheets	Implementing a pipeline that will run on GitLab for continuous integration. Found and worked on fixing broken tests. Modify branch protection for GitLab.	13	30

Comments and Extended Discussion (Optional):

- While breaking tests is taking time to fix, the added pipeline will reduce this from happening again as it will tell us when commits cause breaking tests. Therefore, we won't merge breaking changes like this onto our "master" branch.

Plans for the Upcoming Two Weeks:

- Ian Simon: Works on ParseTSV flags.
- Jacob Laing: Continue cleaning up code within the tests as well as write any new tests that may come up. Start looking into how we can take the data that we have extracted and start plugging it into Tensorflow
- Nathan Carter: I will work on finishing the tests for the functions calcColumnAvg, calcNumMissingValues, handleMissingData. I will also work on fixing broken tests and starting to look into how we will integrate Tensorflow into our project.
- Samantha Williams: Works on completing the MetaData tests. Implementing users entering the file locations of their data rather than hard coding these into the program. Explore tensorflow more.
- Scott Rose: Plan on finishing the implementation of the tensor flow component of the product.
- Aria Sheets: Works on fixing broken tests and starting work on the machine learning module