Note: The assignment instructions and data on the NCCP website did not completely align, so the following assumptions were made.

1. The RESULTANT AVERAGE wind speed and wind direction was used. Not the average. There is no average data for wind direction.
2. Only 6 of the 7 listed training sites were used. There was a missing site.

**Data Setup:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| DATE (DOUBLE) | WIND SPEED (INT) | #ConstituientValues | WIND DIRECTION | #ConstituientValues | Site Number |
| Date 1 site 1 | X1 | 6 | Y1 | 6 | 1 |
| Date 2 site 1 | X2 | 6 | Y1 | 6 | 1 |
| … | … | … | … | … | … |
| Date N site 7 | x1 | 6 | y1 | 6 | 7 |

**Date:**

Matlab is not capable of classifying text data. Therefore, the date time string was converted to a double value representing the date. This was done with a built in matlab function, similar to changing a date time to UTC time in C++.

Since the date time was converted into a double. The relation between years would not have been kept. Therefore, the year on the date was dropped. Since weather patterns generally repeats on a yearly basis, dropping the year allows the time double to repeat as well.

**Site Number:**

All 7 sites were combined into one data group. The siteNumber filed represents the site indexing from west to east.

**Wind Speed:**

Matlab is not capable of predicting a double value. Therefore, the wind speed was converted to an integer value. The decimal points were dropped.

**Classification Model:**

With the above date input, various classification models were tested each site. On average, a Medium Tree classification approach produced the most accurate result as seen in Fig. 1 below. Note, the highest classification accuracy was only 2.3 percent. However, this data is misleading. The accuracy of 2.3 percent means that the classifier was able to predict the *exact* wind speed 2.3 percent of the time (not taking into account the rounding error of moving from a double to an int). This is not the MAE or RMSE error.

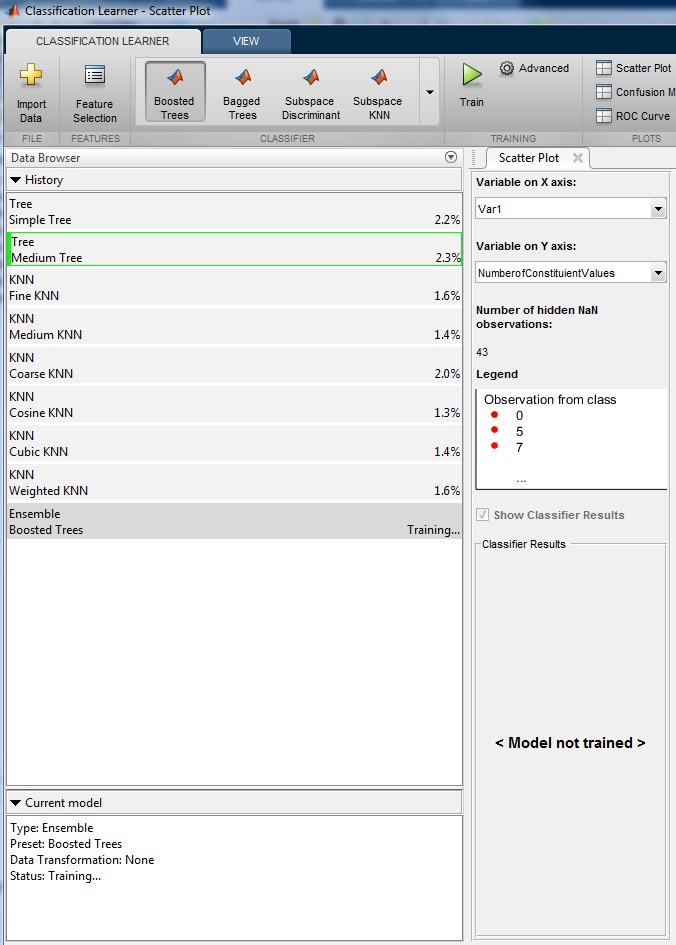


Figure 1: Various classification models applied to the test site data.

The best performing classification task was then exported into the matlab workspace so it could be used to classify the test data for the 3 testing sites. The medium classification tree can be seen in Fig. 2 below.

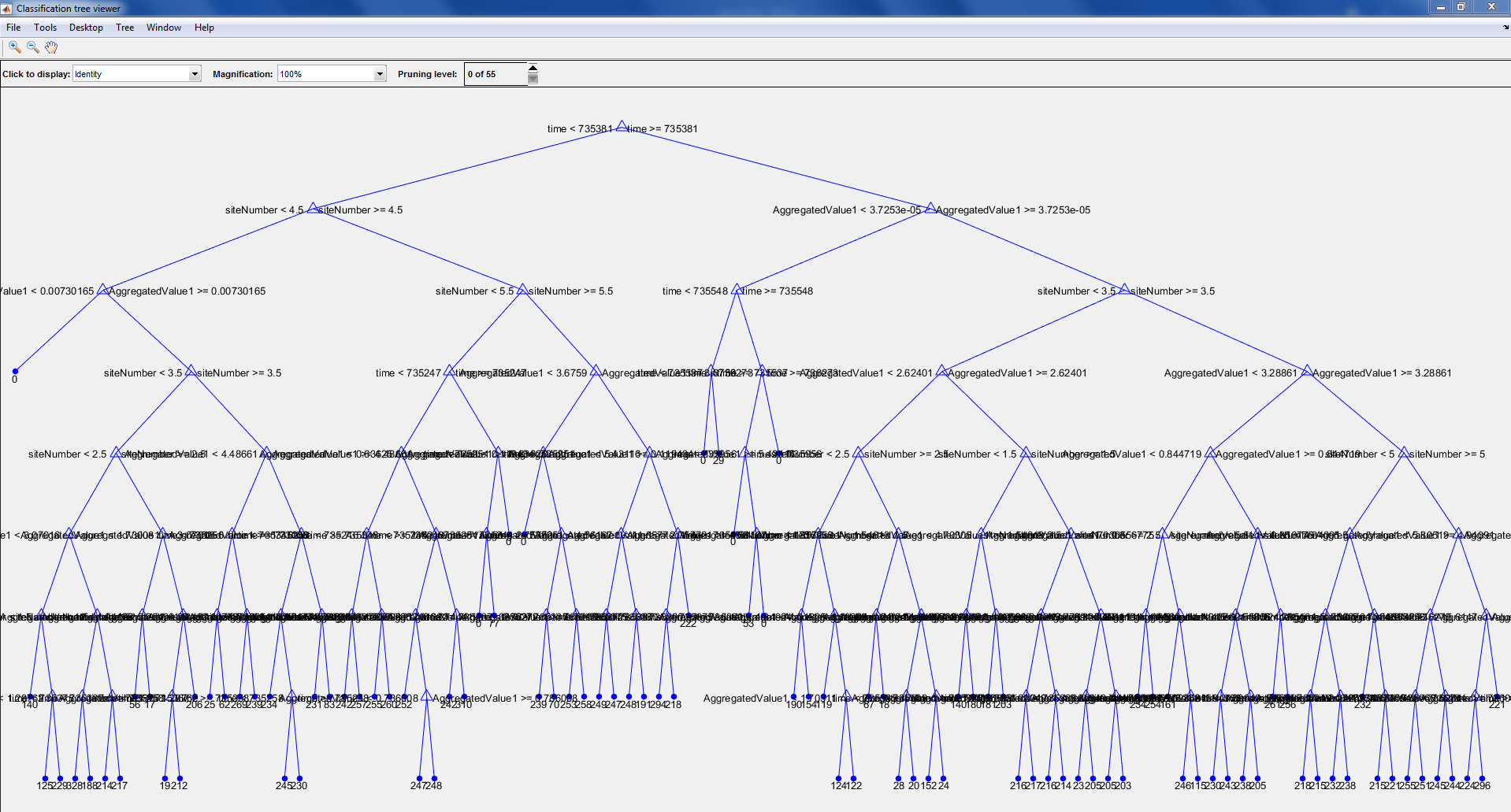


Figure 2: Medium classification tree used to classify wind speed data from all sites.

**Results:**

Applying this classification model to the Snake Range West Pinon Jupiner site produced the following results.

MAE was 59 and RMSE value was 3895.9. Output shown in Fig. 3 below.

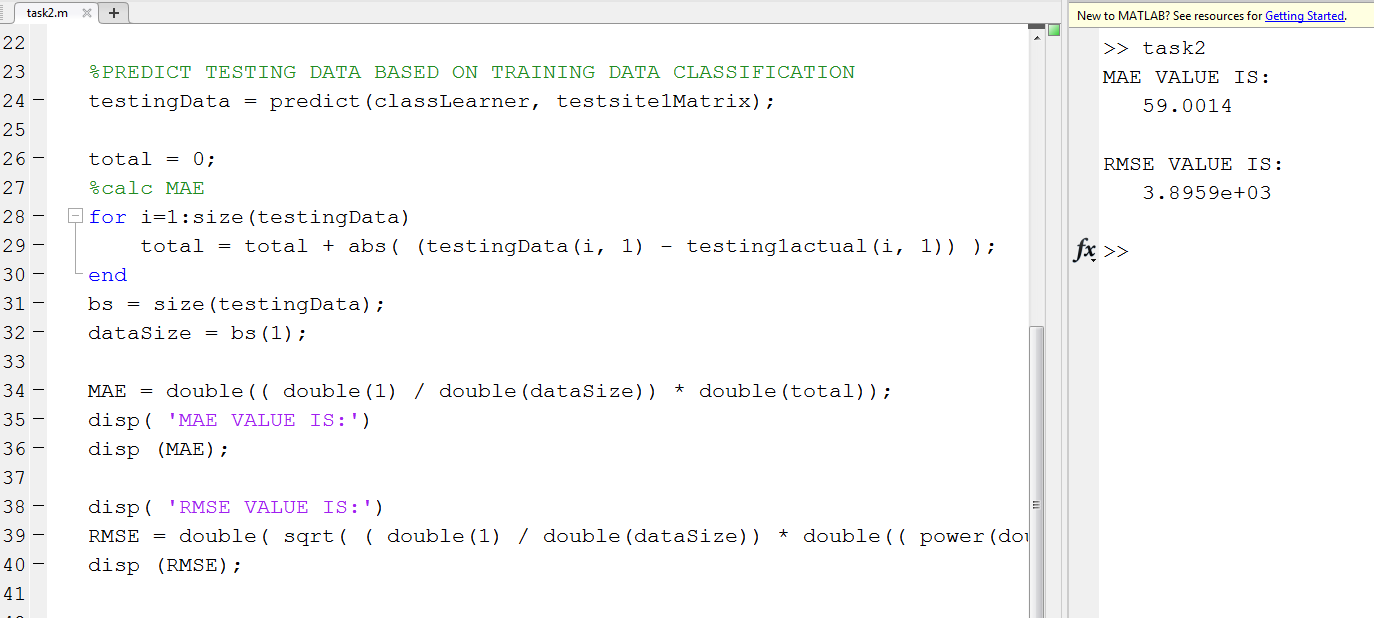


Figure 3: MAE and RMSE values for testing site 1.

For the second testing site the MAE error was 95 and the RMSE was 630.7.

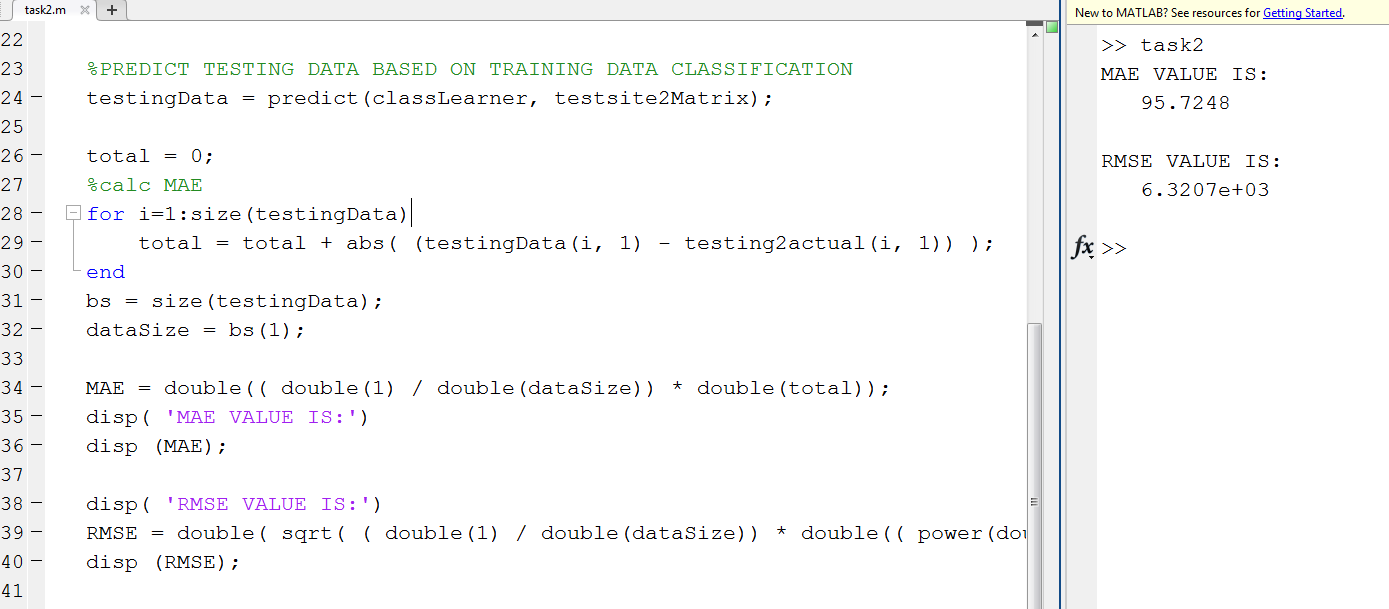


Figure 4: MAE and RMSE values for testing site 2.

For the third site the testing MAE error was 55 and the RMSE was 3683.

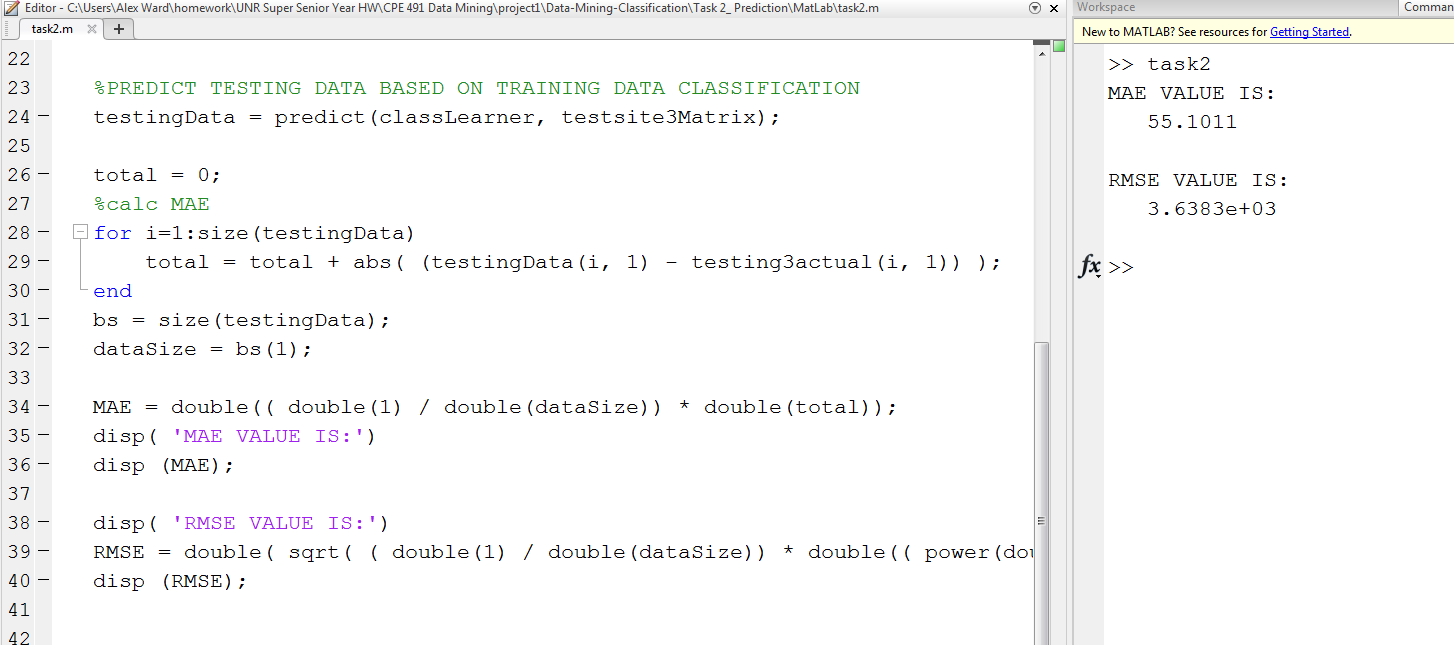


Figure 5: MAE and RMSE values for testing site 3.

**Code:**

Compiling and running the code.

The classification model and data has already been created and included in this project. To test the code simply open the matlab source code file ‘task2.m’ in matlab. Once opened import the classification tree data, training data, and testing data. Click the import data button at the top left of the screen and choose the ‘completed\_workspace’ file as shown in the figure below. Once imported run the program to see the MAE and RMSE error. To change between testing sites, simply change the testsiteXMatrix and testingXactual variables, where X represents the site number you wish to test (1-3).

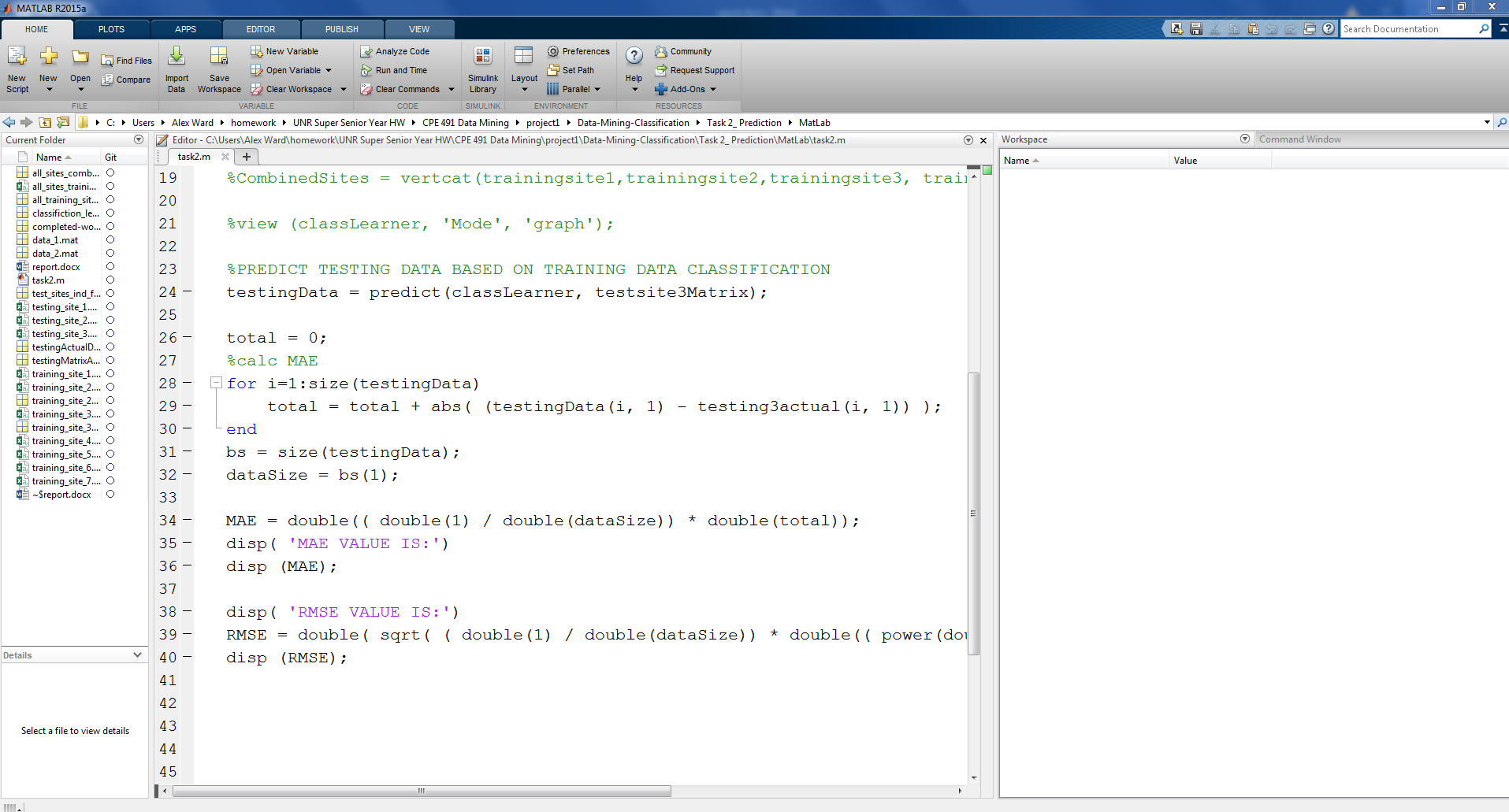


Figure 5: Importing the classification data into the workspace before running the code.

