**1) Reference code:** https://www.statology.org/pandas-train-test/

2) **Difference between reference code and project code:**

\* The code uses both train\_size and test\_size meaning they define the amount of data which is given to train and test. The accuracy for this

code increases when we increase the size of training data but for our code this logic works in reverse way which we concluded after testing the data with

different sizes of training data.

\* We use only train size for code meaning, if we declare the size of training data as 50% then it automatically takes the

size of testing data as the other half 50%.

3) **Authentication:**

The code takes the all the datasets (20 persons datasets) into a array and assigns each person a ID same as the number in their name of file. Meaning a person 0 will be assigned ID 0. And when we run the authentication no matter what name the test file has, the authentication will identify it correctly.

4) **Reference code for GUI:**

https://github.com/zainsiddiqui/Predicting-Survival-on-Titanic-with-Machine-Learning-and-Graphical-User-Interface