**Step 1: Verify the raw data features file**

- Locate the rawdata.features file you created in deliverable 2.  If deliverable 2 completed successfully, the file is located in "C:\it691\BAS\features".
- Using a text editor with word wrap turned off, open the rawdata.features file.   This file is expected to have 182 rows.  180 rows of data plus two rows of header information.

**Step 2: Split the rawdata.features file into a testing and training data files**

- If your editor supports line numbering, enable it and select rows 1 – 92.  This includes the two rows of header information and half of the data.  Since this data file has 180 rows of data, excluding the two row header, 90 rows go into the testing file and 90 rows go into the training file.
- Place rows 1-92 into a text file named **test.txt** and save it in your features directory: "C:\it691\BAS\features"

```
 rawdata.features
  1  Keystroke biometric data, Unknown/Unknown, created 1/26/2010
  2  180
  3  ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7266667491162182, 0.7:
  4  ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7630007359389196, 0.7'
  5  ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7684780204670636, 0.7(
  6  ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7833716752003784, 0.6'
  7  ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.787878446293536, 0.80:
  8  ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 1, 230, 0.7359144654574524, 0.559:
  9  ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6509206147889198, 0.403'
 10  ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6149248356827676, 0.345(
 11  ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6334634891241636, 0.378:
 12  ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6604637938678197, 0.430'
 13  CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.4344496818402506, 0.·
 14  CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.4774499897393189, 0.·
 15  CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.44284629715078305, 0
 16  CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.46092621573770737, 0
 17  CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.5277289710063767, 0.·
 18  DAVID FRONCKOWIAK/Male/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.050704773945909845, (
 19  DAVID FRONCKOWIAK/Male/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.04447343014404843, 0
 20  DAVID FRONCKOWIAK/Male/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.0316682745513522325, (
 21  DAVID FRONCKOWIAK/Male/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.0026977231609936259, (
 22  DAVID FRONCKOWIAK/Male/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.03241572490476756, 0
…
180  YONESY NUNEZ/Male/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.3561416088
181  YONESY NUNEZ/Male/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.3747565250
182  YONESY NUNEZ/Male/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.3770080958
183
```

- Place rows 93-182 into a text file named **train.txt** and save it in your features directory: "C:\it691\BAS\features"

*\* In future experiments you may not have an even number of rows to split between testing and training files.  For this experiment, do not split a subjects data samples between testing and training files to make an even number split.  In other words, split the file on the next or previous person's set of samples, but not in between.  If "Jane Doe" has five rows, do not split Jan Does data rows between testing and training files.*

**Step 3: Fix the file headers**

- **test.txt** has two header rows already, but the value must be changed from 180 to a 90 to reflect the reduced number of data records in the file.  Using your text editor, change the 180 to a 90 and save the **test.txt** file.

```
test.txt
Keystroke biometric data, Unknown/Unknown, created 1/26/2010
90
ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7266667491162182, 0.7
ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7630007359389196, 0.7
ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7684780204670636, 0.7
ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.7833716752003784, 0.6
ANCEY VARGHESE/Female/20 - 29, Right, LAPTOP, FABLE 1, 230, 0.787878446293536, 0.80
ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 1, 230, 0.7359144654574524, 0.559
ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6509206147889198, 0.403
ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6149248356827676, 0.345
ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6334634891241636, 0.378
ANNIE GRENCI/Female/30 - 39, Right, LAPTOP, FABLE 2, 230, 0.6604637938678197, 0.430
CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.4344496818402506, 0.
CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.4774499897393189, 0.
CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.44284629715078305, 0
CAROLINE BOYLAN/Female/Under 20, Left, LAPTOP, FABLE 1, 230, 0.46092621573770737, 0
```

- **train.txt** does not yet have any header information.  Using your text editor, cut and paste a copy of the header you updated in test.txt and place it on the first two rows of **train.txt**.

```
train.txt
Keystroke biometric data, Unknown/Unknown, created 1/26/2010
90
KATHLEEN VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.6974441
KATHLEEN VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.7168998
KATHLEEN VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.7820381
KATHLEEN VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.8826132
KATHLEEN VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.7613447
LAUREN BOYLAN/Female/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.9777190526
LAUREN BOYLAN/Female/40 - 49, Right, LAPTOP, FABLE 1, 230, 0.9840013082
LAUREN BOYLAN/Female/40 - 49, Right, LAPTOP, FABLE 3, 230, 0.8422891358
LAUREN BOYLAN/Female/40 - 49, Right, LAPTOP, FABLE 3, 230, 0.8474558351
LAUREN BOYLAN/Female/40 - 49, Right, LAPTOP, FABLE 3, 230, 0.8565212513
LAURIE VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.191902447
LAURIE VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.357286348
LAURIE VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.187623080
LAURIE VILLANI/Female/50 - 59, Right, LAPTOP, FABLE 1, 230, 0.356070668
```

*\* In future experiments, when creating testing and training data files from feature files, you must always have two header rows in the testing and training data files.   The value on line 2 of each file must reflect the actual count of data rows, excluding the two header rows.  This can be an odd number and the testing and training data files may have different row count numbers on row 2.*

**Step 4: Run a BAS Experiment**

- Navigate to your BAS installation directory "C:\it691\BAS and run the Refactored-BAS.bat script from Windows Explorer or the command prompt.  The following screen should display if you launched BAS successfully:
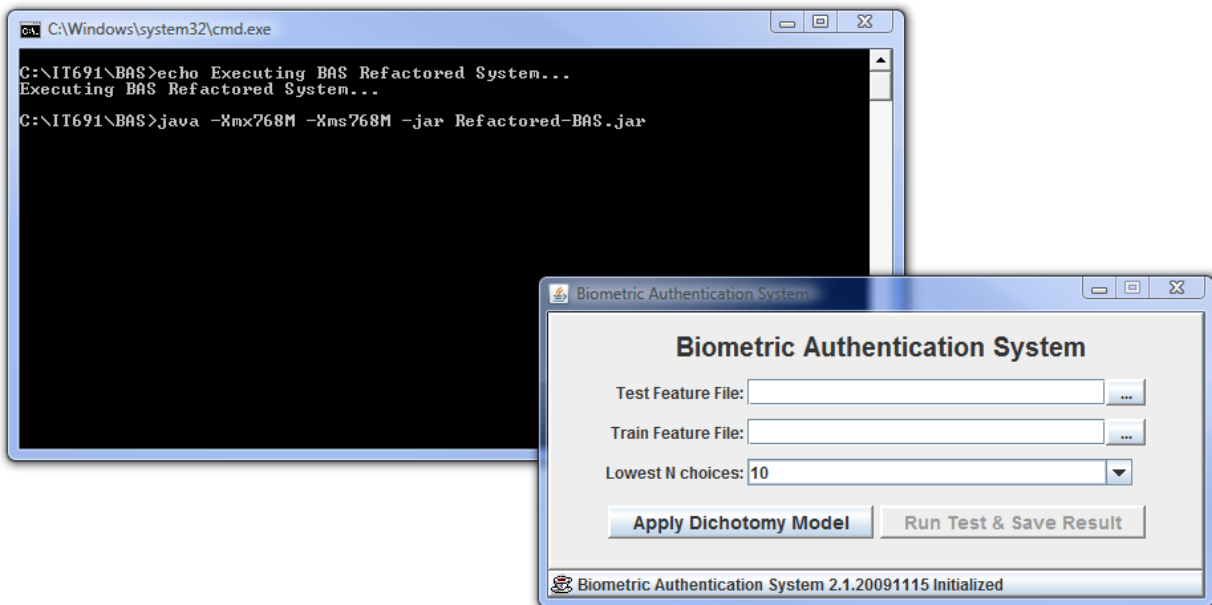


The command window can be minimized if desired, but must not be closed.

**Step 3: Setup the Experiment**

- In the Test Feature File field, enter the testing data file name you created in deliverable 3: "C:\it691\BAS\features\test.txt".

-  In the Train Feature File field, enter the training data file name you created in deliverable 3: "C:\it691\BAS\features\train.txt".

- Leave the Lowest N choices at the default of 10.

  *If the header information is incorrect, you will get an error when attempting to dichotomize the data:  "Error reading feature file(s)":*
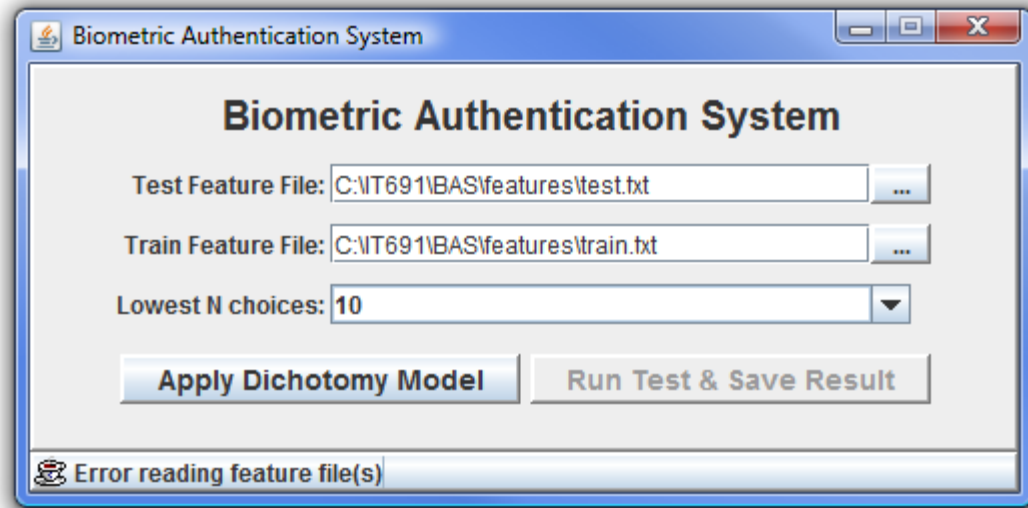
*If the headers are correct, the system will run as expected:*
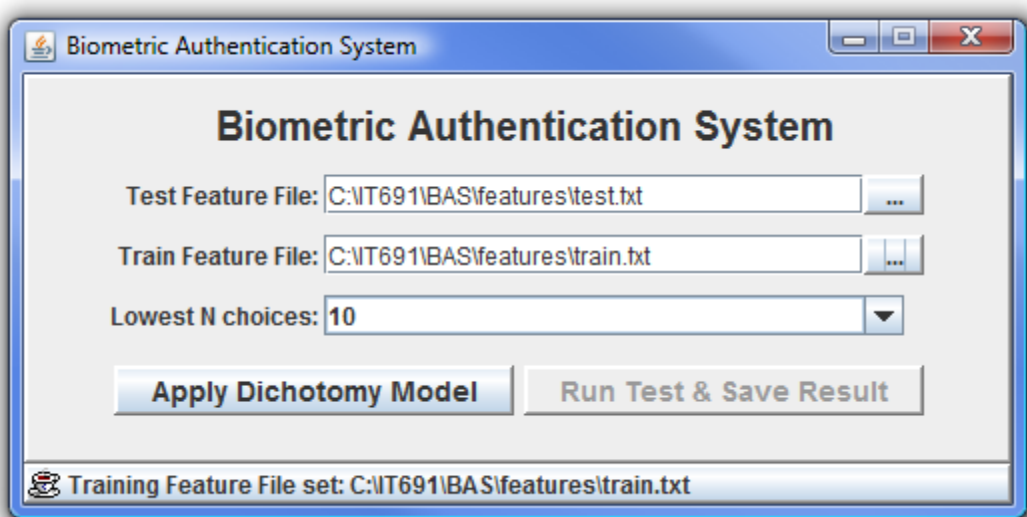


**Step 4: Dichotomize Data**

- Select the 'Apply Dichotomy Model' button.   A dialogue box "Input max data" will open.  Since we will not be restricting the amount of dichotomy data, leave this field at the default of 0 and click "OK".  This will use all data.
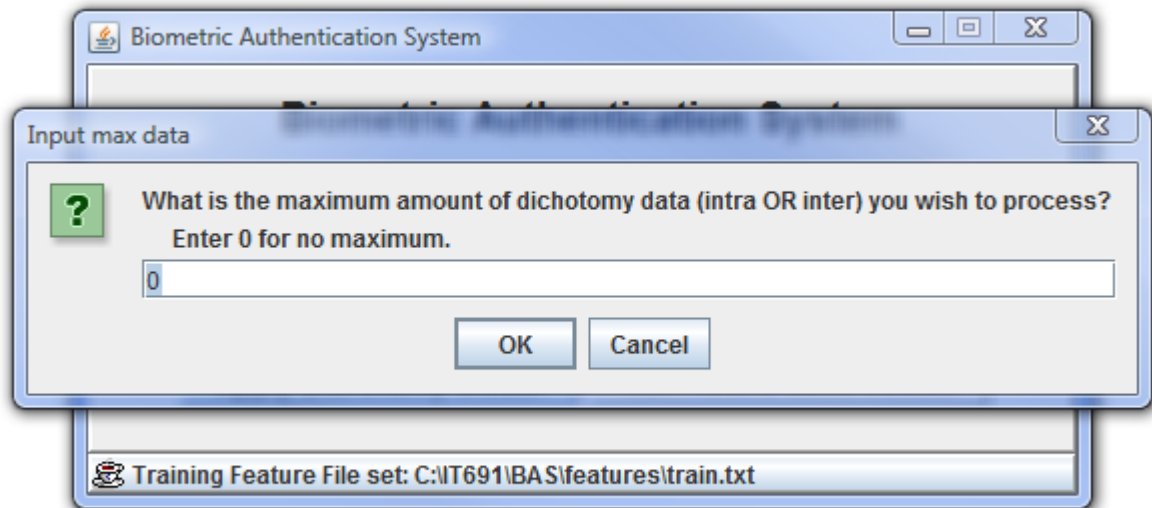
- After pressing "OK" to the maximum dichotomy data size, the system will begin to process the data. It will process in a few seconds. Notice the "Processing Feature" count in the status bar at the bottom of the dialogue box.

- After processing completes, a dialogue box similar to the following will appear. Observe the "Run Test & Save Result" button is now clickable. Before dichotomizing the data, this button was grayed out.
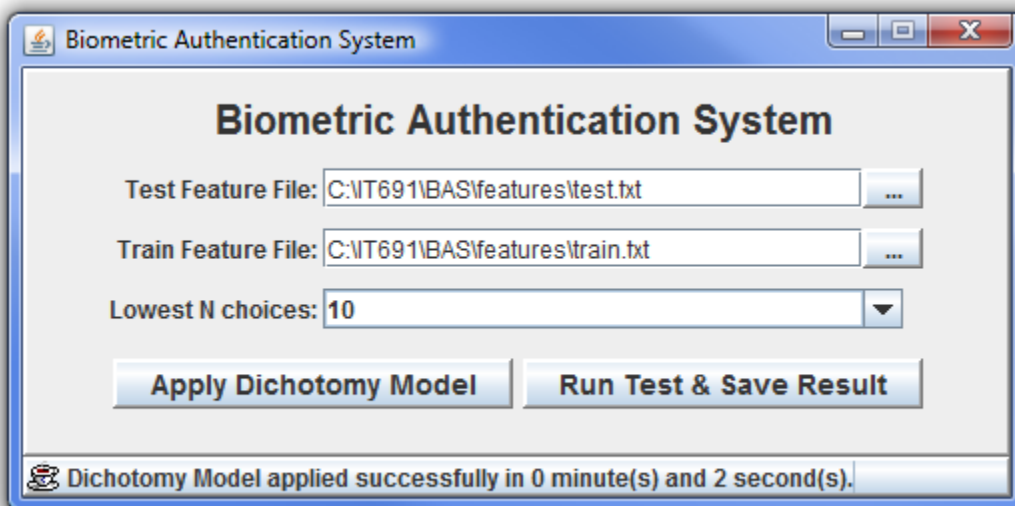
**Step 5: Run Experiment**

- Click the "Run Test & Save Result" button.  A dialogue box will open:
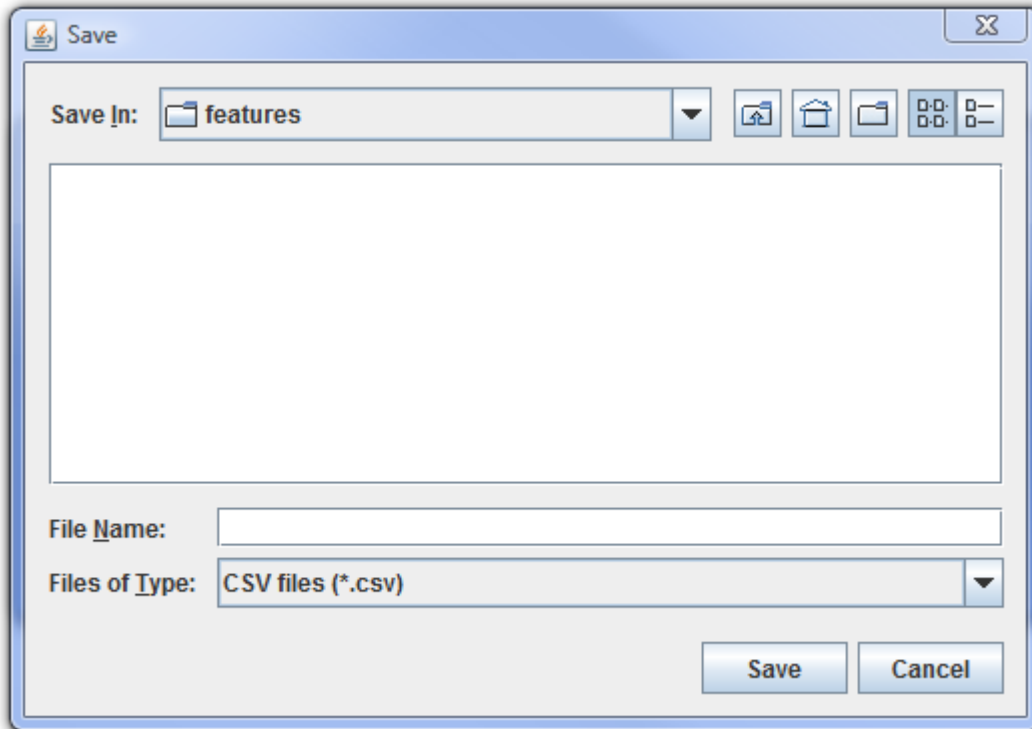


- Enter a name in the File Name field.  For this exercise, use "deliverable3.csv".
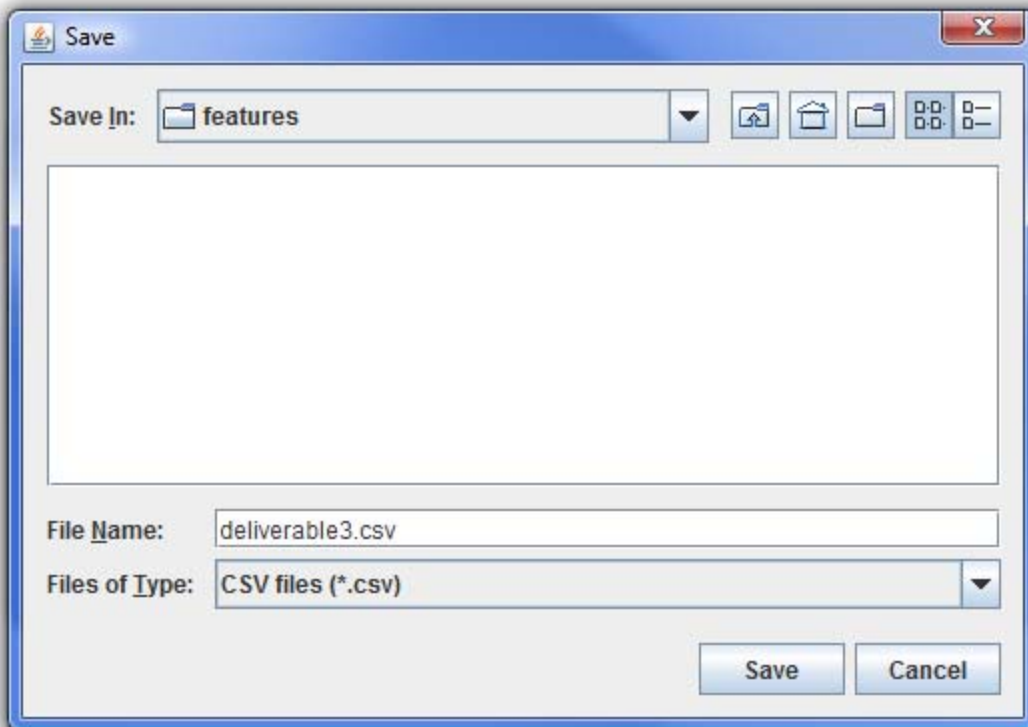
- After pressing the "Save" button, the experiment will run.  Observe the "Test samples" in the status bar at the bottom of the dialogue box.  It will continue to increment as the testing and training samples are processed.  Depending on your computer, this experiment should take approximately 3-10 minutes to run.
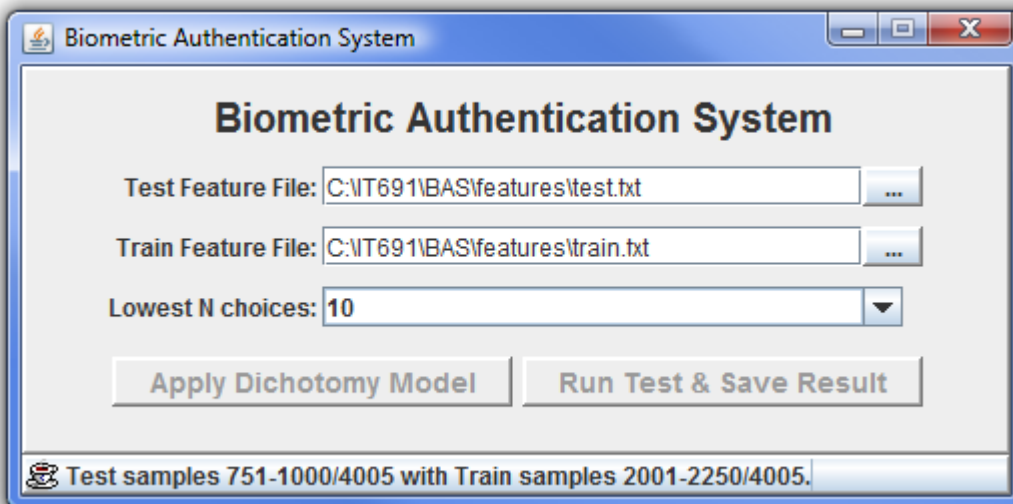


When the processing is complete, the status bar will indicate a successful test run as will the command window.
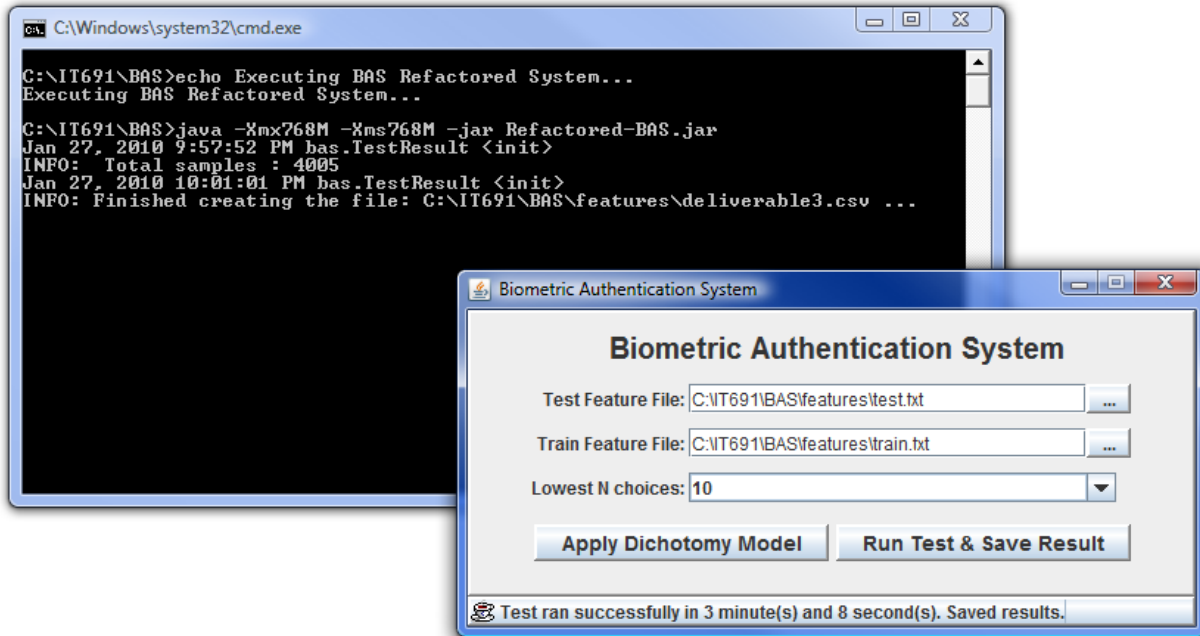
**Step 6: Check the output data file**

The file "deliverable3.csv" should exist in your "C:\IT691\BAS\features" directory.  Other files containing Meta and dichotomy data are also present in the directory and were created during the previous steps. Using a spreadsheet or text editor, you can view this file.  It will be used in the next steps.

- You can now close the BAS application by clicking on the "X" in the top right of the dialogue box.

Step 7: Process the output data file and check the results

- Using Windows Explorer or a command prompt, go to the c:\IT691\BAS directory and run the Calculator.bat program.  The following windows should display:
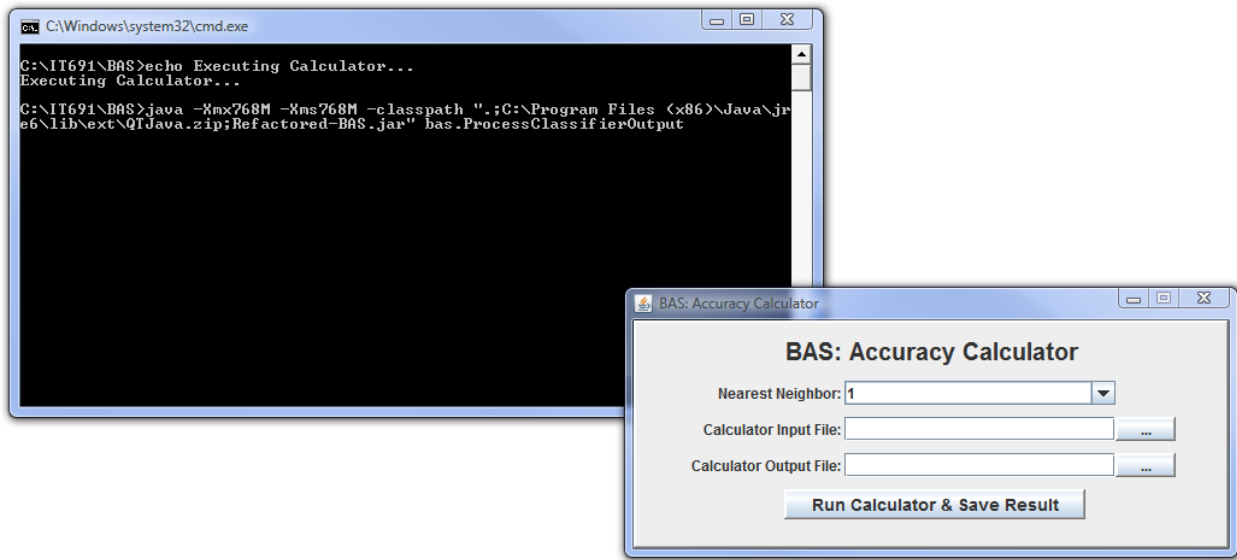
The command window can be minimized if desired, but must not be closed.

- Select the "deliverable3.csv" file you created in the previous step for the File Name and press the "Open" button. **Be sure you are selecting files in the features directory.**
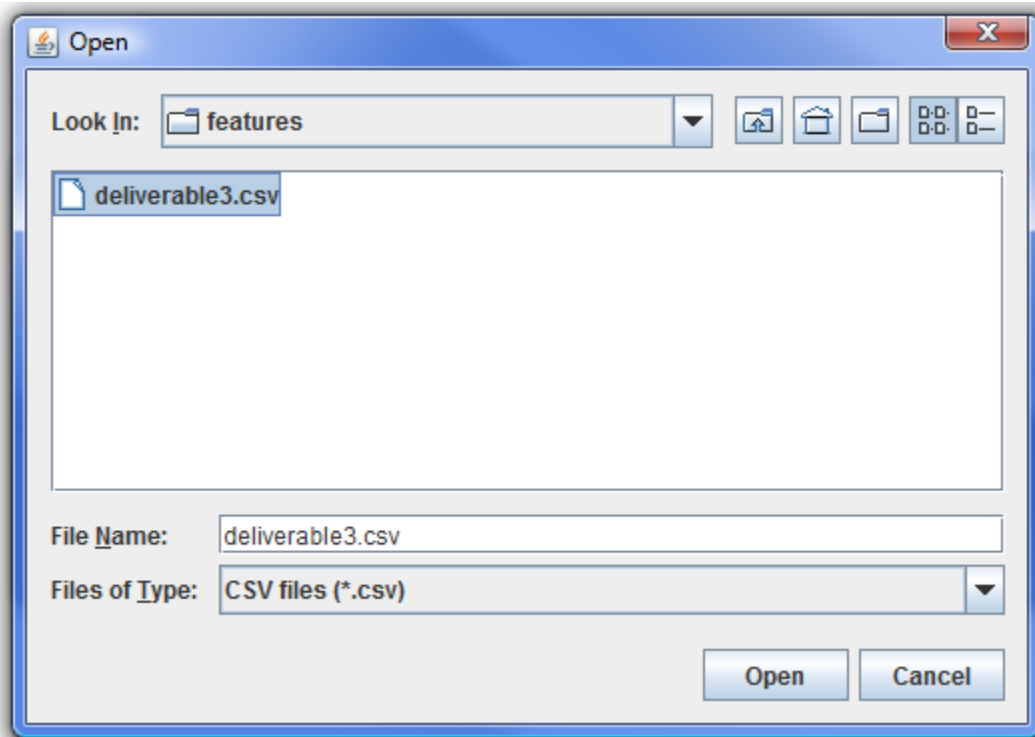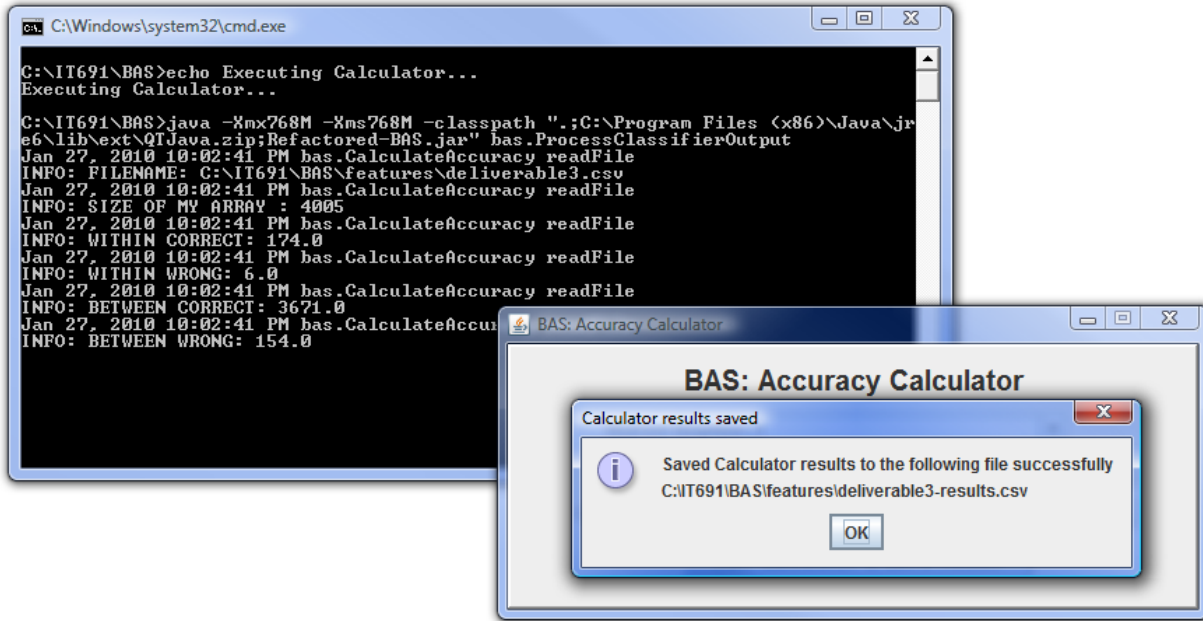
- Enter an Output File Name in the "Calculator Output File" field and press the save button.  For this exercise, enter "c:\it691\deliverable3-results.csv".  The file should be saved in the features directory, "C:\it691\BAS\features".
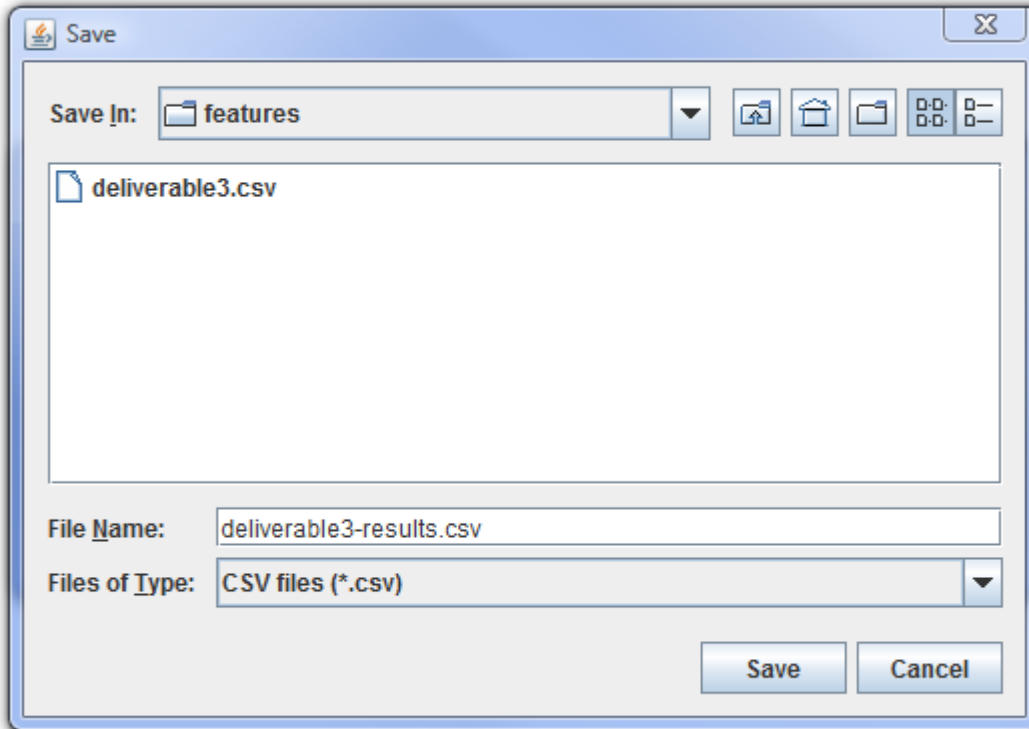
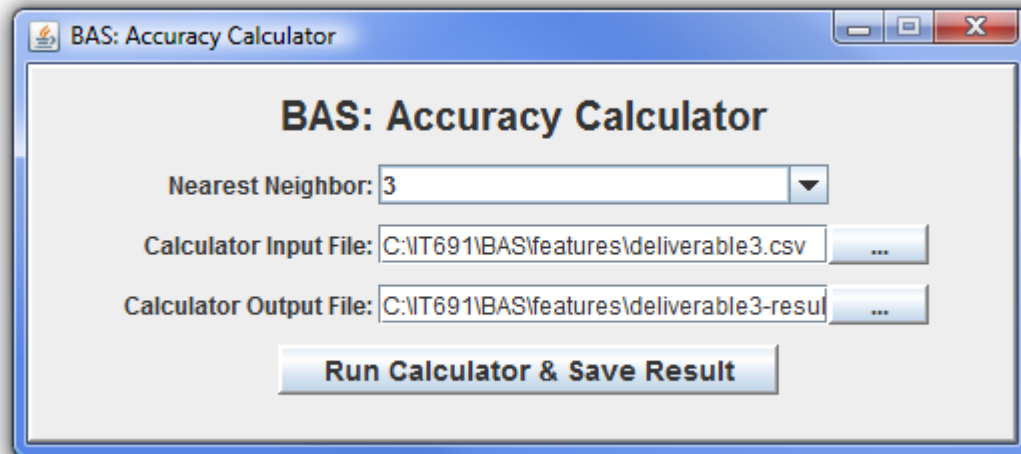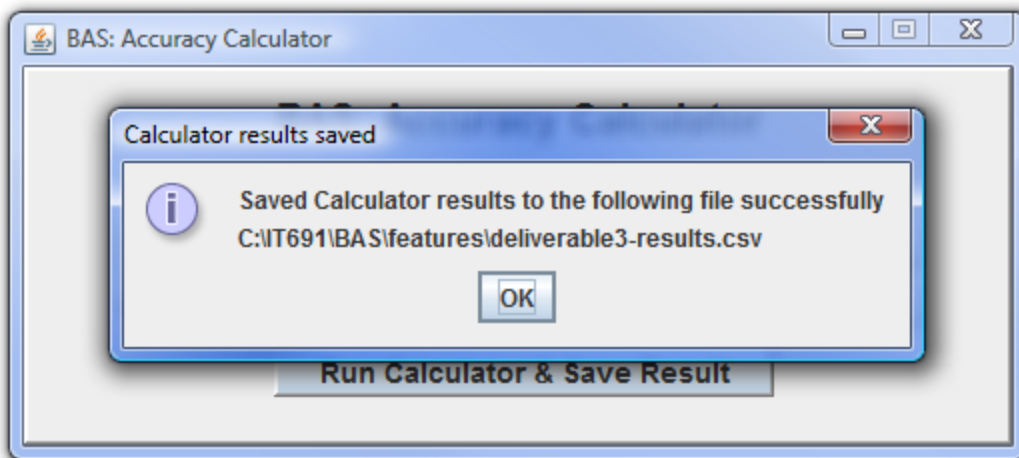- Press the "Run Calculator & Save Result" button.
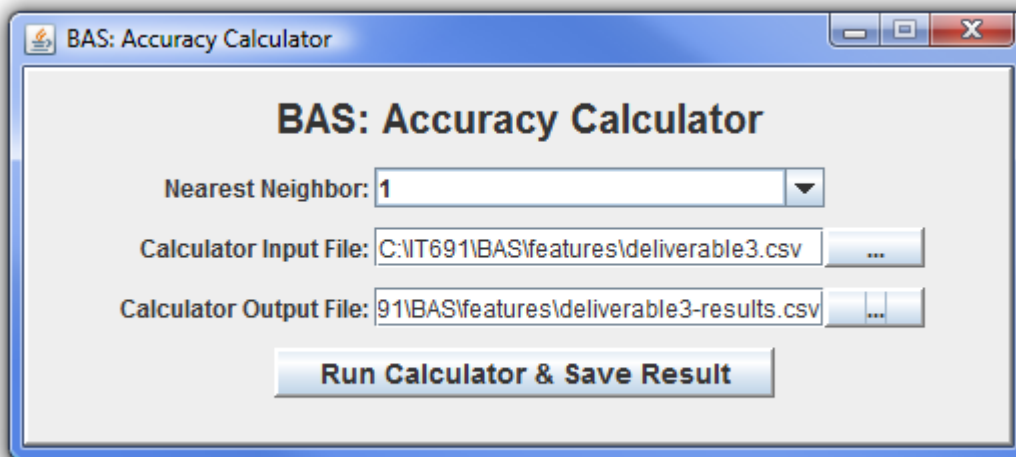
- Press "OK" to close the dialogue box.
- A BAS Result Window will appear.  Press the "Close & Accumulate" button on the bottom left of the dialogue box.

- The BAS: Accuracy Calculator window will appear again.  Change the Nearest Neighbor to 3.  The Calculator Input File and Output File should default to their previous settings.  These should remain the same.
- Press the Run Calculator & Save Result button.   The Calculator results saved dialog box will open.  Press "OK" to clear.  The accumulated results should appear.



- Press OK to continue.

- Press Close & Accumulate and repeat for 5, 7, and 9 Nearest Neighbors.  After 1, 3, 5, 7, and 9 nearest neighbors are processed, the Result window should look exactly like:



- Press Save Result (middle button on bottom of dialogue box).  A dialogue box will open. Enter "deliverable1-processed-results" in the File Name field.  The extension will be .html and the file will be stored in your c:\IT691\BAS directory.  You can open this in many browsers or programs that understand HTML.
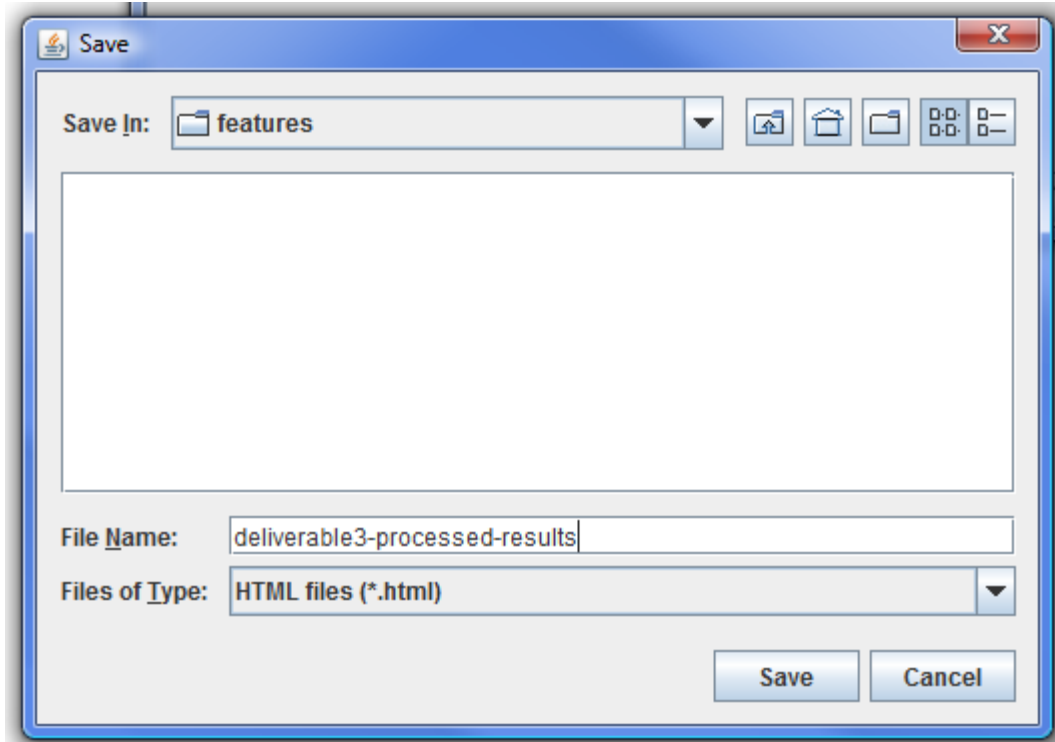
- Press Close & Clear to reset the results and click on the "X" in the top right of the dialogue box to close the program.  Remember, your results are saved.



**Biometric Authentication System Results**

| Biometric | Test | Test Sizes | Train | Train Sizes | FRR | FAR | Performance | Test Subject | AVG(Sample) | Train Subject | AVG(Sample) | kNN |
|---|---|---|---|---|---|---|---|---|---|---|
| Keystroke | LAPTOP/FABLE 1 | 180-3825 | LAPTOP/FABLE 1 | 180-3825 | 3.33% (6/180) | 4.03% (154/3825) | 96.00% (3845/4005) | 18 \| 5.00 | 18 \| 5.00 | 1 |
| Keystroke | LAPTOP/FABLE 1 | 180-3825 | LAPTOP/FABLE 1 | 180-3825 | 4.44% (8/180) | 2.75% (105/3825) | 97.18% (3892/4005) | 18 \| 5.00 | 18 \| 5.00 | 3 |
| Keystroke | LAPTOP/FABLE 1 | 180-3825 | LAPTOP/FABLE 1 | 180-3825 | 5.56% (10/180) | 2.75% (105/3825) | 97.13% (3890/4005) | 18 \| 5.00 | 18 \| 5.00 | 5 |
| Keystroke | LAPTOP/FABLE 1 | 180-3825 | LAPTOP/FABLE 1 | 180-3825 | 5.56% (10/180) | 2.64% (101/3825) | 97.23% (3894/4005) | 18 \| 5.00 | 18 \| 5.00 | 7 |
| Keystroke | LAPTOP/FABLE 1 | 180-3825 | LAPTOP/FABLE 1 | 180-3825 | 5.56% (10/180) | 2.85% (109/3825) | 97.03% (3886/4005) | 18 \| 5.00 | 18 \| 5.00 | 9 |

**Legend**
**Biometric**: Type of Biometric
**Test**: The Biometric test that was ran
**Test Sizes**: The sample size of Intra-class and Inter-class of testing set
**Train**: The Biometric train that was ran
**Train Sizes**: The sample size of Intra-class and Inter-class of training set
**FRR**: False Rejection Rate
**FAR**: False Acceptance Rate
**Performance**: Overall Performance of the test
**Test Subject | AVG(Sample)**: Testing Feature File # of Subject | Average # of Sample per Subject
**Train Subject | AVG(Sample)**: Training Feature File # of Subject | Average # of Sample per Subject
**kNN**: k Nearest Neighbor

**Step 7: Verify your results**

Compare your results to the published "Improved Biometric Authentication" results for the **LapCopy** condition using 1 nearest neighbor
(http://support.csis.pace.edu/CSISWeb/docs/techReports/techReport268.pdf, page 6).

Your results should be identical.

| LapCopy | 180-3825 | 180-3825 | 3.3% | 4.0% | 96.0% |
|---------|----------|----------|------|------|-------|