

NEW YORK INSTITUTE OF TECHNOLOGY
CSCI 860: Biometrics and Its Applications (Spring 2019)
Computer Science

Project Report: Implementation of Manhattan verifier and reporting false accept (impostor pass) and false reject rates on a publicly available keystroke biometric dataset.

Name: Shariq Khan

ID: 1225726

Under Professor
Kiran Balagani

Index

1. Project Objective	03
2. Data Preprocessing	03
3. Development Tools	04
4. Program Flowchart	05
5. Deliverables	06
6. Conclusion	08

1. Project Objective

You are required to implement Manhattan verifier and report false accept (impostor pass) and false reject rates on a publicly available keystroke biometric dataset. You may use any programming language.

2. Data Preprocessing

2.1 Data Details

Sr. No	Name	Format	Source
1.	DSL-StrongPasswordData	Excel	http://www.cs.cmu.edu/~keystroke/#sec1

Table 2.1 Data Description

Total number of data instances	20400
	34
Subjects (Typists)	51
Subjects Values	's002', 's003', 's004', 's005', 's007', 's008', 's010', 's011', 's012', 's013', 's015', 's016', 's017', 's018', 's019', 's020', 's021', 's022', 's024', 's025', 's026', 's027', 's028', 's029', 's030', 's031', 's032', 's033', 's034', 's035', 's036', 's037', 's038', 's039', 's040', 's041', 's042', 's043', 's044', 's046', 's047', 's048', 's049', 's050', 's051', 's052', 's053', 's054', 's055', 's056', 's057'
Typing Repetition	400
Missing Values	00

Table 2.2 Data Details

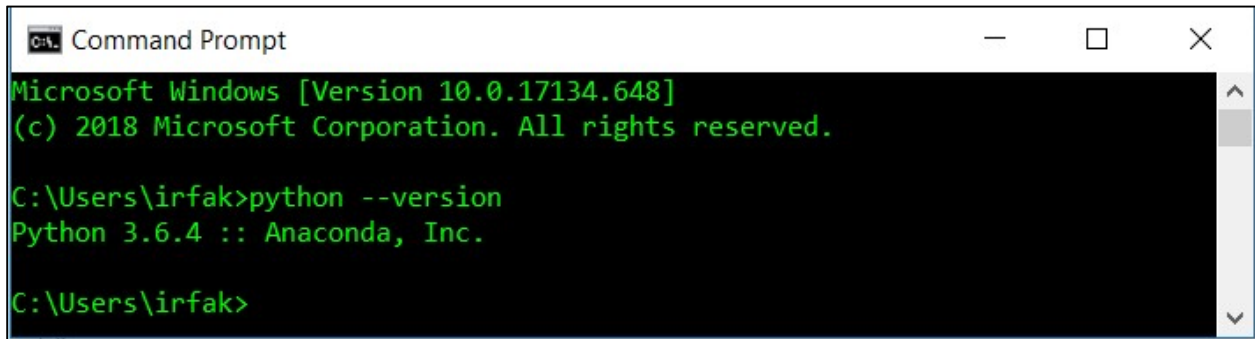
2.2 Data Sampling

Each User	Users	TOTAL INSTANCES	INSTANCES	%
Training Data	51	400	200	50
Training Data	51	400	100	25
Testing Data	51	400	200	50
Testing Data	51	400	300	75

Table 2.3 Data Sampling for training and testing

3. Development Tools

1. Python 3.6.4 - Anaconda Distribution



```
Microsoft Windows [Version 10.0.17134.648]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\irfak>python --version
Python 3.6.4 :: Anaconda, Inc.

C:\Users\irfak>
```

2. Jupyter Notebook



```
C:\Users\irfak>jupyter --version
4.4.0

C:\Users\irfak>
```

3. IDE – PyCharm Professional



4. Packages and Modules

4.1. SciPy

4.2 pandas

4.3 NumPy

```
C:\Users\irfak>python
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import scipy
>>> print(scipy.__version__)
1.0.0
>>> import pandas
>>> print(pandas.__version__)
0.22.0
>>> import numpy
>>> print(numpy.__version__)
1.14.0
```

4. Program Flow

The program has been developed in Anaconda distribution with python version 3.6.4 and the data has been analyzed in Jupyter Notebook version 4.4.0. The Notebook is for testing and analysis on the dataset provided.

Step 1: The file is downloaded as mentioned from in Table 2.1.

Step 2: The downloaded file is in excel format and the details of data is mentioned in Table 2.2.

Step 3: The file is then read in pandas data frame which has a table structure.

Step 4: The samples are created as mentioned in Table 2.3.

Step 5: The mean vector template is calculated using Manhattan distance.

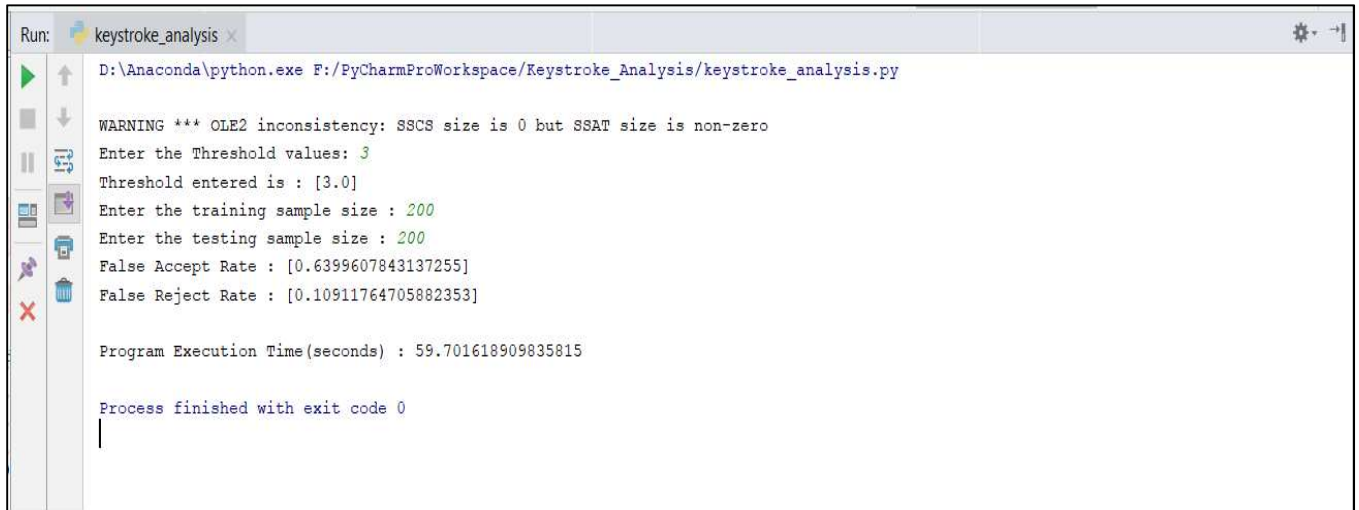
Step 6: The genuine and imposter scores are computed and based on the user thresholds the false accept rate and false reject is computed and same is displayed in console.

5. Deliverables

The deliverables are outputs from the program

1. Well documented, compliable software codes and executables performing template calculation; genuine and impostor score computation with Manhattan distance; and calculation of false accept and false reject rates at a given threshold T . [40 points]

The threshold is taken as 3 for training sample of 200 and testing sample of 200 for each user.



```
Run: keystroke_analysis x
D:\Anaconda\python.exe F:/PyCharmProWorkspace/Keystroke_Analysis/keystroke_analysis.py

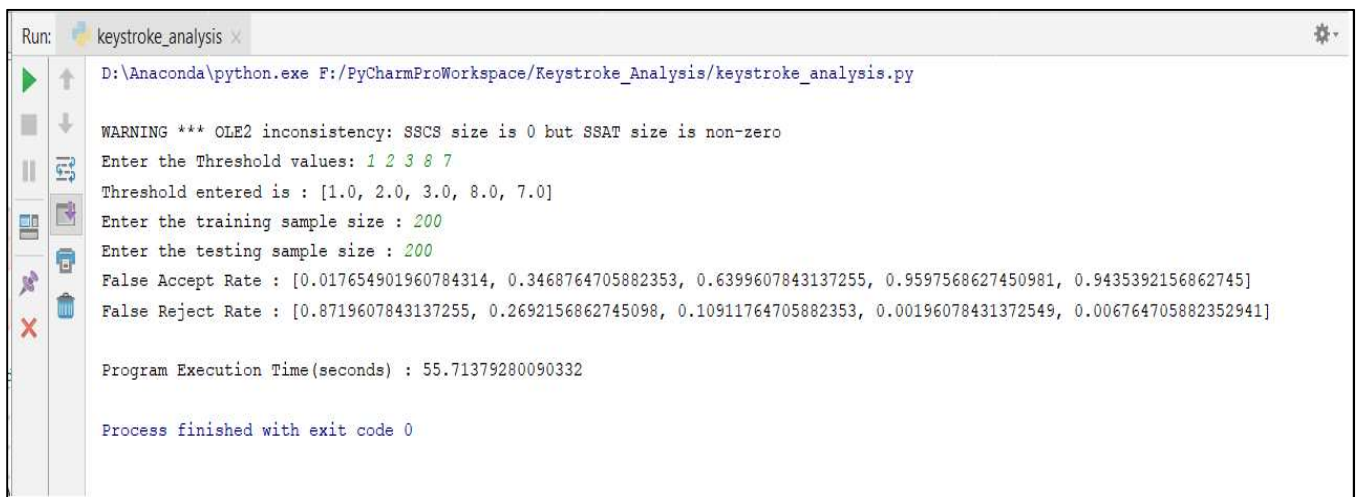
WARNING *** OLE2 inconsistency: SSCS size is 0 but SSAT size is non-zero
Enter the Threshold values: 3
Threshold entered is : [3.0]
Enter the training sample size : 200
Enter the testing sample size : 200
False Accept Rate : [0.6399607843137255]
False Reject Rate : [0.10911764705882353]

Program Execution Time(seconds) : 59.701618909835815

Process finished with exit code 0
```

2. A well-written report containing false accept and false reject rates for $N = 200$ and various threshold values (choose five threshold values that give you the best tradeoff between the false accept and false reject rates). [40 points]

Thresholds are taken as 1,2,3,8,7 for training and testing sample of 200 respectively for each user.



```
Run: keystroke_analysis x
D:\Anaconda\python.exe F:/PyCharmProWorkspace/Keystroke_Analysis/keystroke_analysis.py

WARNING *** OLE2 inconsistency: SSCS size is 0 but SSAT size is non-zero
Enter the Threshold values: 1 2 3 8 7
Threshold entered is : [1.0, 2.0, 3.0, 8.0, 7.0]
Enter the training sample size : 200
Enter the testing sample size : 200
False Accept Rate : [0.017654901960784314, 0.3468764705882353, 0.6399607843137255, 0.9597568627450981, 0.9435392156862745]
False Reject Rate : [0.8719607843137255, 0.2692156862745098, 0.10911764705882353, 0.00196078431372549, 0.006764705882352941]

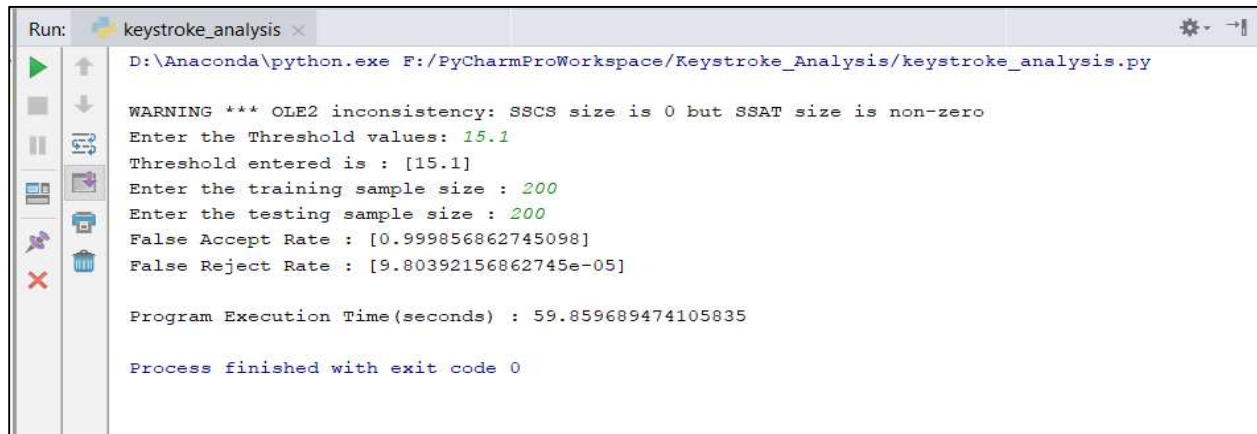
Program Execution Time(seconds) : 55.71379280090332

Process finished with exit code 0
```

3. Report the false accept rate at 0 false reject rate, when $N = 100, 200$, and 300 [20 points] :

The requirement is to calculate false accept rate at 0 false reject rate which has different thresholds value for change in training samples.

For sample of 200; Threshold is 15.1.



The screenshot shows the PyCharm Run console for the file `keystroke_analysis.py`. The output includes a warning about OLE2 inconsistency, followed by user input for a threshold of 15.1, training sample size of 200, and testing sample size of 200. The resulting False Accept Rate is approximately 0.999856862745098 and the False Reject Rate is approximately 9.80392156862745e-05. The program execution time is 59.859689474105835 seconds.

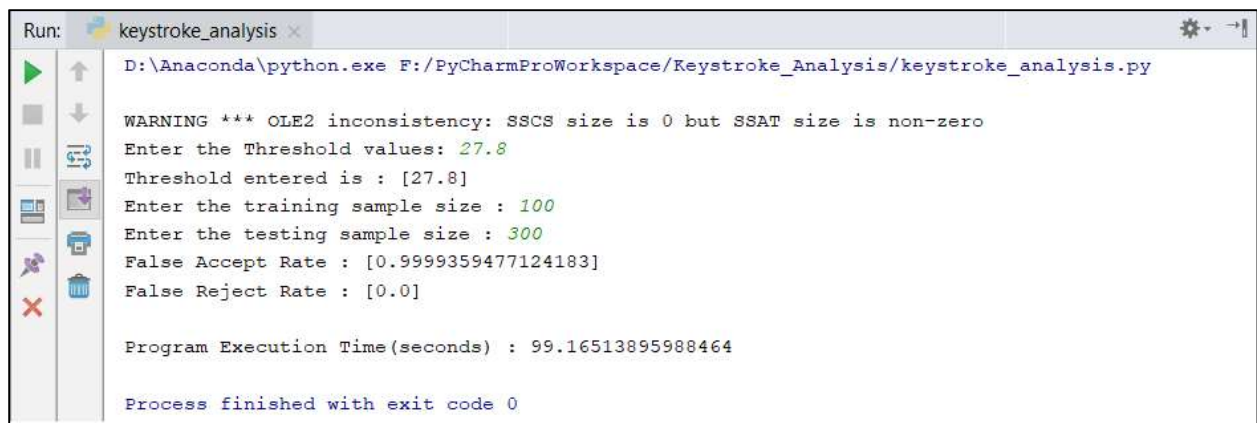
```
Run: keystroke_analysis x
D:\Anaconda\python.exe F:/PyCharmProWorkspace/Keystroke_Analysis/keystroke_analysis.py

WARNING *** OLE2 inconsistency: SSCS size is 0 but SSAT size is non-zero
Enter the Threshold values: 15.1
Threshold entered is : [15.1]
Enter the training sample size : 200
Enter the testing sample size : 200
False Accept Rate : [0.999856862745098]
False Reject Rate : [9.80392156862745e-05]

Program Execution Time(seconds) : 59.859689474105835

Process finished with exit code 0
```

For sample of 100; Threshold is 27.8.



The screenshot shows the PyCharm Run console for the file `keystroke_analysis.py`. The output includes a warning about OLE2 inconsistency, followed by user input for a threshold of 27.8, training sample size of 100, and testing sample size of 300. The resulting False Accept Rate is approximately 0.9999359477124183 and the False Reject Rate is 0.0. The program execution time is 99.16513895988464 seconds.

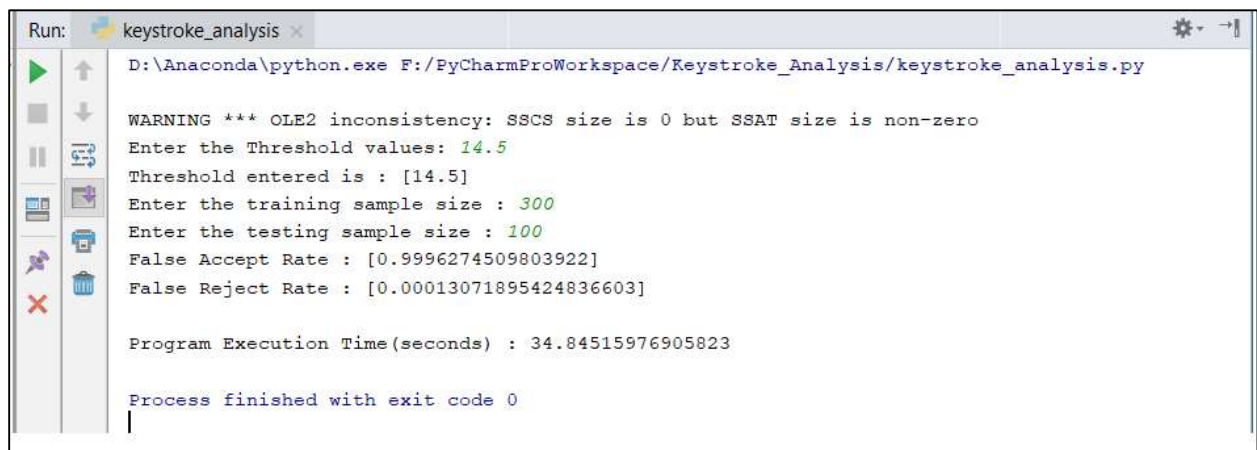
```
Run: keystroke_analysis x
D:\Anaconda\python.exe F:/PyCharmProWorkspace/Keystroke_Analysis/keystroke_analysis.py

WARNING *** OLE2 inconsistency: SSCS size is 0 but SSAT size is non-zero
Enter the Threshold values: 27.8
Threshold entered is : [27.8]
Enter the training sample size : 100
Enter the testing sample size : 300
False Accept Rate : [0.9999359477124183]
False Reject Rate : [0.0]

Program Execution Time(seconds) : 99.16513895988464

Process finished with exit code 0
```

For sample of 300; Threshold is 14.5.



The screenshot shows the PyCharm Run console for the file `keystroke_analysis.py`. The output includes a warning about OLE2 inconsistency, followed by user input for a threshold of 14.5, training sample size of 300, and testing sample size of 100. The resulting False Accept Rate is approximately 0.9996274509803922 and the False Reject Rate is approximately 0.00013071895424836603. The program execution time is 34.84515976905823 seconds.

```
Run: keystroke_analysis x
D:\Anaconda\python.exe F:/PyCharmProWorkspace/Keystroke_Analysis/keystroke_analysis.py

WARNING *** OLE2 inconsistency: SSCS size is 0 but SSAT size is non-zero
Enter the Threshold values: 14.5
Threshold entered is : [14.5]
Enter the training sample size : 300
Enter the testing sample size : 100
False Accept Rate : [0.9996274509803922]
False Reject Rate : [0.00013071895424836603]

Program Execution Time(seconds) : 34.84515976905823

Process finished with exit code 0
```

6. Conclusion

On comparative study taking in account different threshold values for different training and testing sample for each user, false accept rate and false reject rate can be observed in order closely for patterns of keystroke timings of different users each typing passwords.

There is more scope of improvements after analyzing the program which are discussed as below.

1. The design pattern for programming can be improved as the current program executes between 30 and 60 seconds. The program can be more efficient.
2. The threshold estimated for false accept rate at 0 false reject rate can be trained with a better model and techniques.