Eddie Lam **Cuny ID:** 23285373 CIS 4900 Spring 2020
Contact Information
Email: eddielam27@gmail.com
Address: 1641 68 Street Brooklyn NY 11204
Cell Phone: (646)384-5942

Project Provider: Tzipora Halevi
Contact Information
Email: halevi@sci.brooklyn.cuny.edu **Office:** Ingersol Hall, room 2156A

Final Report

The project is about trying to produce similar or even a close result from the research group Touchalytics. The research group was trying to create a machine learning program that is able to determine which user is using the phone by collecting the way they are moving their finger on the screen. After collecting the data, they will compare it with other users. From their program they were able to produce an 87% accuracy from a single stroke and even an almost perfect accuracy when they increased the number of strokes.

So now I am tasked to create a program that will try to reach a similar result. I will be using python and the tensorflow as the library for this project. The program will be using the KNN algorithm to get the results and all of this will be written in Jupyter Notebook. Since I will be using the data that the research group collected, I had to modify it so my program will be able to run it. After a few trials of modifying that dataset, I have a usable dataset because I had to replace some data that contains 'NaN' and 'inf' with the average of the column. When that was done, I needed to write a script that will combine the 2 rows so that I can see if my program result will increase if I increase the number of strokes.

The goal of this project was to create a program that will use data from the research group and try to produce a result similar to theirs. However, the result that I got after a long time of searching and testing wasn't even close to theirs. The research group got about 87% for a single stroke while I was only able to reach 65%. They were also able to lower the error when they increased the number of strokes. When using 11 to 12 strokes, they were able to bring the error to 2% to 3% while I am only able to increase the success rate 1% when I try using 2 strokes. From how it looks, I don't think I will be able to get close to their results.

Programs and other data used:

featMat.csv: original dataset

featMatVersion2_10.csv: modified dataset

DataSet_V4.csv: modified dataset that contains rows that were combined

Project Code 2 Part 1.ipynb: program that run 1 row dataset featMatVersion2_10

Project Code 2 Part 2.ipynb: program that runs the combined row dataset Dataset_V4

EditingDataset_V3.txt: script that combined the previous row with the current row

ConfusionMatric_1R.csv: the result of 1 row shown in a matrix

ClassificaionReport_1R.txt: the result of 1 row that shows the percentage

ConfusionMatric_2R.csv: the result of 2 row shown in a matrix

ClassificaionReport_2R.txt: the result of 2 row that shows the percentage

Project Log:

Week 1:	**2/19/20 - 2/25/20**	
Week 2:	**2/26/20 - 3/3/20**	**(7hr 36 min)**
Week 3:	**3/4/20 - 3/10/20**	**(6hr)**
Week 4:	**3/11/20 - 3/17/20**	**(2hr 38 min)**
Week 5:	**3/18/20 - 3/31/20**	**(8hr 30 min)**
Week 6:	**4/1/20 - 4/5/20**	**(2hr)**
Week 7:	**4/6/20 - 4/14/20**	**(9hr 30 min)**
Week 8:	**4/15/20 - 4/21/20**	**(6hr 7 min)**
Week 9:	**4/22/20 - 4/28/20**	**(8hr 20 min)**
Week 10:	**4/29/20 - 5/5/20**	**(8hr 10 min)**
Week 11:	**5/6/20 - 5/12/20**	**(8hr 14 min)**
Week 12:	**5/13/20 - 5/19/20**	**(8hr 56 min)**
Week 13:	**5/20/20**	**(59 min)**

From how the project has gone, I believe I did an acceptable job by being able to get some kind of result. Even though it wasn't close to the people from Touchalytics, I am still proud of what I did. Yes, I could have done much more or work on the project much longer, but this was all new to me. I had to learn a different programming language, a different algorithm, and use a different program before I can even start writing code. Even after learning the basics for them, I had to find a usable code that worked with the project which took quite a long time. When I thought I was making some progress, I had to switch to a different program than I previously used to run the code. The project may have taken a while to finish, but overall, I believe I did an okay job on it.