# MACHINE INTELLIGENCE AND EXPERT SYSTEMS TERM PROJECT – GROUP 14

# Analyzing effect of mood of a person on mouse dynamics for imposter detection

SK IZAJUR RAHAMAN – 17EC35025

AADI SWADIPTO MONDAL – 17EC10065

KANISHKA HALDER – 17EC35012

RISHAV SHARMA – 17EC10046

NISCHAY RAJ – 17EC10034

## Introduction:

Cyber security very often managed by passwords. History has shown – along with recent news headlines – that passwords, ID numbers or PINs are not safe, and definitely not sufficient as a standalone authentication factor. No matter how complex we make our passwords, there is still a need for additional layers of security. That is what makes biometrics so popular.

Every human being possesses certain unique features in terms of both physical and behavioral characteristics that are different from everybody else on the planet. The first and most common thing that comes to mind when speaking of unique features is the fingerprint, which is a physical characteristic. But there are other characteristics that are more of behavioral in nature, like the way we speak, the way we type on a keyboard, the way we write our signature, and several others.

However, using physical biometrics has its downsides related to the nature of physical traits. Once some of the physical features are revealed it can be reused in the online world multiple times by the fraudsters. For example, voice recordings can fairly easily be used to circumnavigate authentication challenges leveraging speech recognition software. Fingerprints can be printed. Photos or videos might be used to spoof the real user in in some cases.

Behavioral biometrics, on the other hand, checks for patterns of behavior that are characteristic and variable over time. That's the motivation behind using behavioral biometrics.

## Problem Statement:

Mouse Dynamics is another behavioral biometric which is used to authenticate a user. A mouse stroke is defined to be the set of points traversed from one click to the next and a set of one or more strokes are used in order to verify a user. The mood of a person affects these two behavioral biometrics. So, analysis of mood of a person may prove to be very crucial while authentication using keystroke or mouse dynamics.

This project work can be divided in 3 parts: -

1. Data acquisition for Mouse Dynamics.
2. Feature Extraction of acquired data.
3. Training a K Nearest Neighbors Classifier for user prediction.

All the three above steps are described below thoroughly.

- ## Data Acquisition:

We have taken continuous data of Mouse Log, Keylogger, in additional we have taken data of both mouse log and keylogger for emotion classification using GUI as instructed in the problem statement. For each person in our group we have around 15 data points on an average.

The mouse log file for each session contains data in the following format –

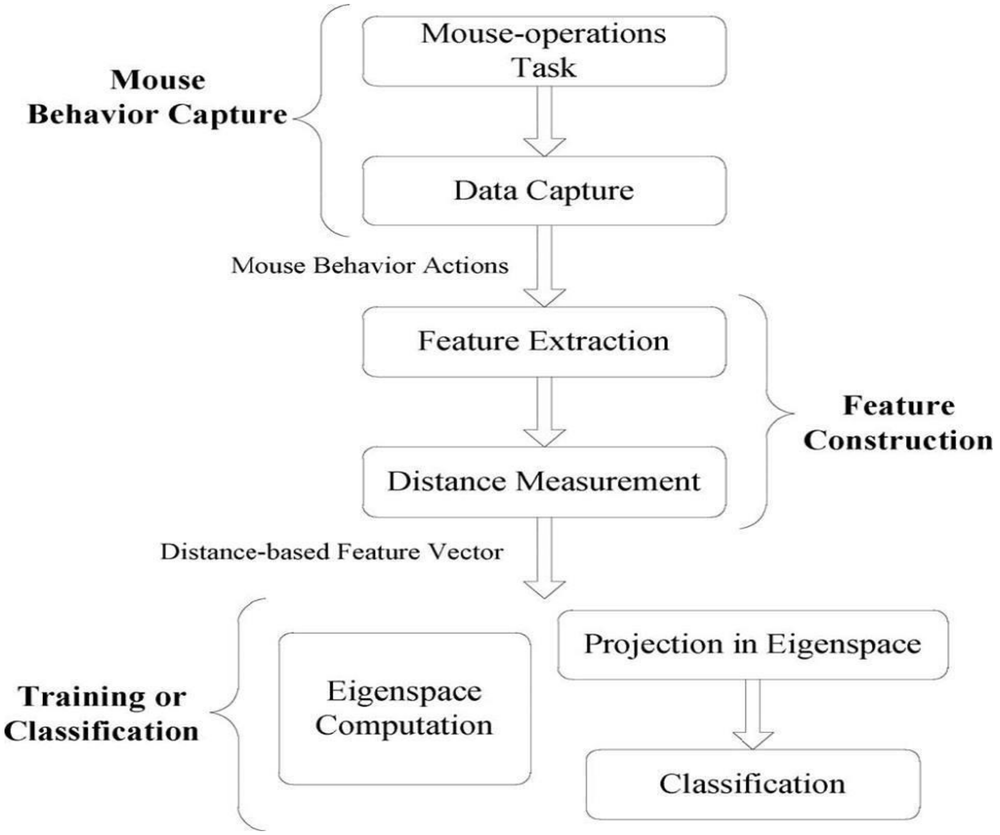
X and Y coordinates of the mouse position and time taken to perform the following tasks:

| | |
|---|---|
| MM: | Mouse Movement |
| MD: | Mouse Drag |
| MP: | Mouse Press |
| MC: | Mouse Click |

- ## Feature Extraction:

The following features are taken into consideration for training the classifier –

Whether it is a left click or a right click.
Position of mouse click or mouse press while scrolling.
Whether it is a single click or a double click.
Time elapsed in each click or press.
Average speed of the drag while scrolling in x and y directions.
Average acceleration of the drag while scrolling in x and y directions.



- ## Training Machine Learning Model:

In our problem statement, we are assigned to train a K Nearest Neighbor Classifier (KNN) to perform the use authentication task.

K Nearest Neighbor Classifier:

1. KNN is a simple machine learning algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).
2. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.
3. Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3-10. That produces much better results than 1NN.

Implementation Details:

For Implementing the above approach, we have used Python3.

The mouse movement, latency and hold time have been extracted from the text files generated.

Each data point is labeled according to their respective user and then fitted in a KNN classifier.

We have used sklearn library to implement our KNN.

- ## Results and Discussions:

1. The accuracy we got on training the data using 5-fold validation is 43.5%.
2. The accuracy can be increased further if we train the model by more and more data.
3. Since this is a high dimension data KNN classification might not be the best approach to classify such data.