National College of Ireland

Keystroke Dynamics for User Authentication using SCM

MSc Research Project
MSc in Cyber Security

Nikita Phadol
Student ID: X21116741

School of Computing
National College of Ireland

Supervisor: Dr. Imran Khan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Nikita Baban Phadol

Student ID: X21116741

Programme: MSc in Cyber Security **Year:** 2021-2022

Module: Academic Internship

Supervisor: Dr. Imran Khan

Submission Due Date: 19/09/2022

Project Title: Keystroke Dynamics for User Authentication using SCM

Word Count: 7960 **Page Count:** 24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Nikita Baban Phadol

Date: 19th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Keystroke Dynamics for User Authentication using SCM

Nikita Phadol
X21116741

Abstract

Today, computer security is an important issue because it is used everywhere to store and process sensitive information. Especially for users of e-banking, e-commerce, virtual offices, e-learning, distribution, computers, and various online services. By using Keystroke Dynamics authentication technology, it is possible to protect your password against many attacks. This technology mimics human behaviour to access their passwords. The analysis in this case is intended to include human behaviour into the device design. Since the device does not require hardware, it does not require additional equipment. Password protection requires only software technology. The results indicate a positive increase in demand for web-based applications. While using the terminal, keystroke dynamics, a sort of behavioural biometrics, may guarantee a consistent user experience. System employs one-time passwords as small files for continuous authentication in this research, extending the usage of support vector machine (SVM) technology for lengthy files. By creating an SVM room for each user that accesses data from 34 user environments, it can reject real users with an astounding mistake rate (nearly 0%) and reject fraudsters. Our findings demonstrate that one-class SVM may be utilized as a tool for continuous user analysis and validation of button dynamics by formulating plans and choosing the proper kernel parameter.

Keywords: Keystroke Dynamics, Authentication, Security, Biometrics, Verification

1 Introduction

The authentication process is basically using login credentials so that users can easily access the system. However, while the conversation is still ongoing, the system cannot check whether the user was first acknowledged on the terminal. (Ho and Kang, 2018) Certification of Quality is based on behavioural biometrics can be used to resolve the trade-off between security and performance. The results of the verification study demonstrate the requirement for the system to recognize the user frequently and get familiar with them. Keyboard Dynamics is a useful and low-cost biometric characteristic that may be used to identify a computer user in the background while the user is working at the terminal. Because each computer user types with a different time, it discovered that there are various differences in the properties of devices used to identify them. Among the machine learning techniques used, vector machine support (SVM) is only used for short scripts such as passwords. (Ho and Kang, 2018) However, authentication mechanisms that use short texts in practice, especially those that require all users to have regular access to the same password during training, may not be sufficient to open the feature. In order to replace authentication without asking the user to input a screen password, the concept of long-term data processing should be used instead.

In our test, users write in a controlled environment and answer free research questions. System suggests a new way to read long texts and delete functions for long-term editing. This paper includes SVM processing of long documents and a new key term deduction method. In addition, it was discovered that SVM can accurately and efficiently identify users.

The most protected and convenient testing tool for testing is biometric one. He cannot live; Stealing, forgetting, and lying are not easy. Physical feature of human behaviour is also known as the biometric analysis. The security system uses three different types of tests:

- Knowledge of passwords, ID numbers, or personal information.
- Items containing key numbers, smart cards, or symbols
- What you are biometric.

Physical biometrics often include fingerprints; hand or palm geometry; and retina, IRIS, or facial features. Practical skills include signature, sound, rhythm, and a speaking class on biometrics, advanced signature design, and sound technology. Security is also provided by the mouse switch, and it is now easier to recognize the key. A biometric process called subject change is based on the idea that every writer writes in the same style.

Neurophysiologic factors make the signature unique to anyone. (Ho and Kang, 2018) Additionally, for each person participating in these activities, a special set of tools must be made. In the nineteenth century, when telegraph operators could identify one another using their own distinctive systems to translate messages on telegraph lines, this idea of key navigation feature is occurred. Biometric typing, key analysis, key strokes are the types of key change.

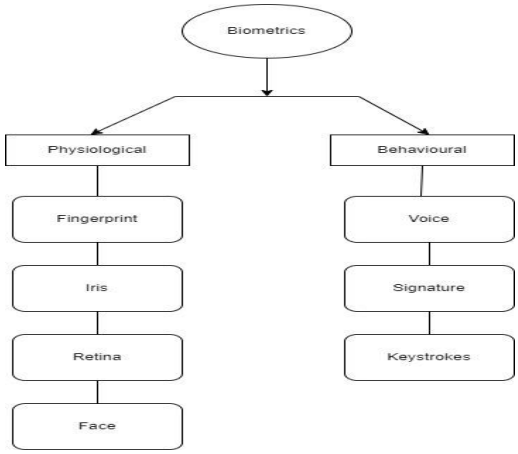


Fig. 1. Biometric Characteristics

2 Related Work

Over the years, many different methods and classifications have been developed in the field of key dynamics to distinguish one person from another. In security, the user model is a behavioural biometric that can be authenticated. In this section, discussion is going on the basics of key dynamics and research using support vector technology (SVM). Additionally, by looking at the experiment's SVM technique. (Ho and Kang, 2018) Consider the authentication problem of legitimate users while rejecting disavowals as an external search problem that can be solved using a one-class SVM. The hypersphere can be used in this way to collect the data content of authorized users and stop the distribution of other stuff. Also, the features that produce the greatest outcomes from random testing are selected using genetic algorithms. If the allocation of the FAR experiment was perfect, the experiment provided the average load rejection (FRR) of 21 users. (FAR reached 0% power). These tests are conducted using a method that is akin to a speedier selection method, producing a greater diversity. Using the live time (when pressing and releasing the same key) and the time-of-flight (when the key is pushed) of the digraphs, it runs a double SVM with dashed lines. (Ho and Kang, 2018) Each user receives a unique one-to-one SVM that treats only valid data as good and all other data as negative. Short passwords were used to sign up users for the trial, but none of these strategies worked well for massive, long-term courses.

During the test, utilize the keypad and let the user choose which keypad to enter the user key change. Additionally, to managing the same type of data with a single SVM class, they also do it by classifying all non-invalid data as negative. One SVM class outperformed two SVM classes, according to the outcomes, and a ROC system was present close to the failure peak. The trials also support the possibility that the user device will select one-to-one (single-class) SVM over one-to-one SVM. Examine how the system of the device evolves over time to look for alterations to the key PIN code of the keypad. These characteristics are properly filtered out and sent to the SVM using radial root function (RBF). Additionally, train the keys to recognize user using the RBF kernel. They use the GREYC caption for the database content. The results are described using a validation process that divides the data into sections and tests one configuration for the other. The tests performed in the above test only use short

characters (e.g., name, password) for confirmation. (Ho and Kang, 2018) They are used as a secondary/secondary identification feature as well as regular user monitoring as well as clear diagnostic procedures. When a user repeatedly inputs their password while training, these features are added. The retrieved elements are compared and examined to determine whether they match the following time the user tries to log in. To buy the system, however, all consumers in our trial had access to long-term data (an average of 13,461 crucial data). Only experiments with long files with GUI-based programs and mouse movements for 3 users, hold time, backup mode, etc. proved to work. The findings revealed that up to 96 percent of the search results originated from various screening techniques. In contrast to subtraction design and estimation, our work simply needs a few digraphs to be properly analysed without the need of any additional tools. Additionally, make use of larger resources to further highlight its advantages throughout this book.

Filtering is a process that allows a machine to personally identify a user. Due to the nature of the problem, analysis is a classification process. (Toosi and Akhaee, 2021) The major changes are one of many options for discussion. This switch's lack of a need for specialized hardware is one of its key advantages. Measurements for a particular sort of activity include memory and recall. Short height makes up a little portion of the good traits that are considered to be favourable in people, just as memory makes up a tiny portion of the positive traits that are considered to be positive. Table 1 lists some examples of information usage, including the level of originality and memory that should be present in a productive system. Depending on the situation, any number of failures (such as a small short circuit) can render the IDS ineffective, but the IDS may record the interruptions (memory) or may not function properly. On the other hand, in questions about the self, one must be sure that the self is accurate (high accuracy), even if in most cases, one does not identify the details (forget), it is noted that the passes are getting shorter and shorter. The kind of system usability with high memory and performance may rely on the length of the keyboard and the length of the keyboard system under test. (2021, Toosi and Akhaee) When the keyboard shortcut system is finished, it may be placed in various places. Consider employing a user password-protected account for this kind of use. The replacement of the keyboard adds a second layer of security throughout the scanning process. During the double discovery process, the attacker must also track the user's device type.

Text	Scenario	Precision	Recall	Input length
Fixed	One-time Authentication	high	high	short
Free	Intrusive Detection	low	high	long
Either	Identification	high	low	either

Table 1: use case for keystroke dynamics-based systems

It should be noted that the second "point" (that is, keystroke dynamics) is transparent from the user's point of view - biometric keystroke data is continuously recorded and does not require further actions by entering the user's password. Dynamic keystroke systems can be helpful even if the keystroke must be long enough to fulfil the criterion. Here, provided a different channel-based method that includes the display of fresh element definitions. To test

method, the system compares the presenting impacts of our component determination strategy to existing element selection techniques that employ Manhattan distance-based appropriation. The results of the exploratory analysis reveal that the classifier that benefits from the highlights that were chosen generally offers the best classifier.

Our examination is completed in the structure of the principal change examination. Collecting keystroke data from the PIN identifier in a phone context and using reliable keystroke-based classifiers for verification. A validation method is known as "key-based validation" makes use of the uncommon example of hits that occur when a client enters a particular string of characters using a device like a scanner. Keystroke data in the flexible environment includes press/discharge timings, hand estimates, hand facilitations, and movement data from sound sensors. Since this study is only getting started, needs to concentrate on the use of essential transformation traits. Additionally, unique key verification calls for a high level of customer data security insurance. As a result of our method for restricting information extraction from other customers' information, key control is an appropriate tool for our element determination procedure.

2.1 Proposed Method

The keystroke Dynamics usually involve to recognizable characteristics such as keystroke timing or set of key and the time to finish keystroke, i.e. keystroke interval which is time to release previous key in enter next key. (Toosi and Akhaee, 2021) All the keystroke dynamics generally includes following points 1. Searching for a program to record data on device and make it run, 2. Capture that data during keystroke, 3. Important feature of removing the training and testing classes, 4. Next training, classifier uses some data from device, 5. Last, by using another remain part of devices data classifier gets tested. Number of issues are fixed by scientists at every level. In this study report, describe some of the many findings achieved by scientists in each of the five assessment steps stated above. There will be a short description of various typing-related features before moving on to the scientific analysis of keystrokes. So, when anyone types on computer it captures the time when key get pressed, then the time of pressing how long it has been pressed and last the time of release. E.g., one key to the next key. Here, time is measured by pressing the key and releasing the same key is called Flight time.

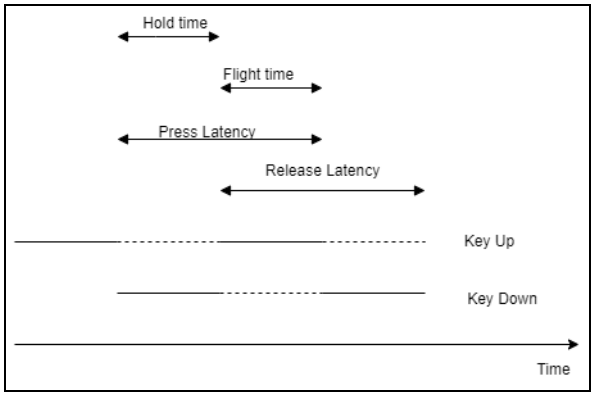


Fig. 2. Flight Time

As a result, three-click, click-to-push (RR), and click-to-click (RP) features may be extracted from the raw data. It is also possible to extract other time records, such as how long it takes to write a single letter, two letters, or a three graph (three letters). (Andrean, Jayabalan, and Thiruchelvam, 2020) The Press-to-Press category this image falls into. Graphics are two in a row, while Trigraphs are three in a row; this is followed by multiple linkages, giving them their look. Using these, the word "Market" requires three ("ma," "rk," and "et") and many ("mar," and "ket"). Key time log data is constructed to obtain a basic system supplied by numbers such as procedures and values, as well as various systems of knowledge of a complicated system to measure. When the user clicks on the file all this data gets saved. Following figure is representing the infinite limit.

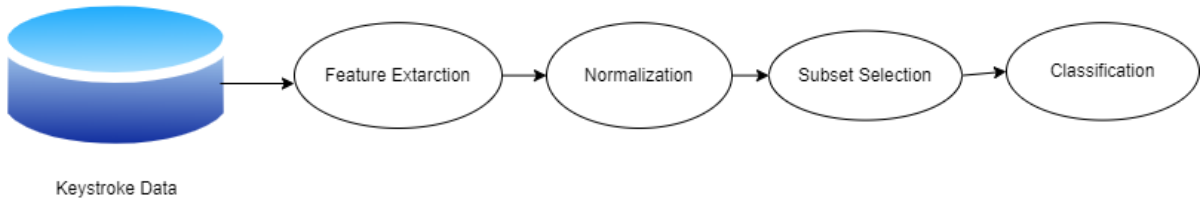


Fig. 3. Keystroke data processing

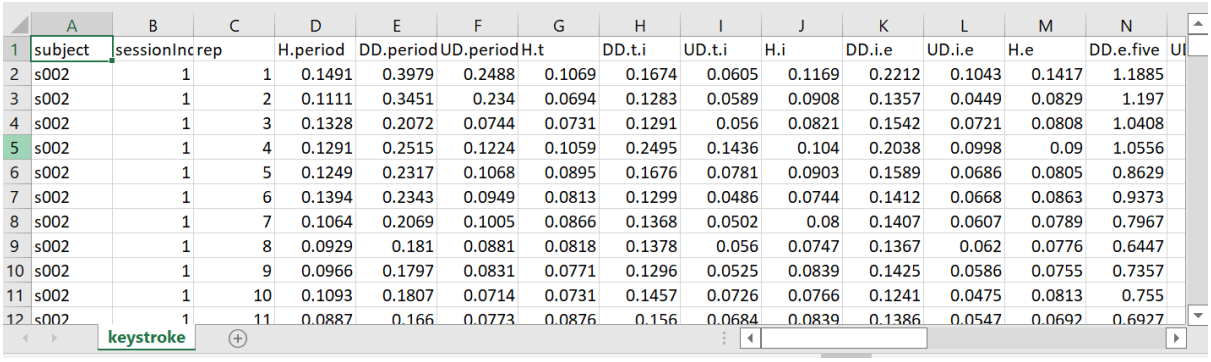
3 Research Methodology

In proposed system deep learning and SCM (spatial distribution, central value distance, and modification algorithm) is used. Also, Buffalo keystroke dataset and Clarkson II keystroke datasets are used.

3.1 datasets

3.1.1 Buffalo keystroke dataset

SUN Buffalo researchers compiled data on Buffalo's free text button from 148 studies. In this case, it is expected that two writing tasks in the lab will be completed with care. Members created a three-part speech by Steve Jobs for Stanford as part of the main assignment. The subsequent errand addresses a few free inquiries. (Alsultan, Warwick, and Wei, 2017) The typical time between two meetings was 28 days. Furthermore, only 75 of the individuals completed two writing tasks using the same console, while the other 73 keyboarders used three different consoles over the course of three meetings. Information about Buffalo is limited. A push button is also included, in addition to a list of deadlines and significant occasions. The average amount of keys on the subject in our section was over 17,000 keys.
Additionally, several participants employed various methods for data collection. The gender of each topic is also included in this information.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	subject	sessionIndex	rep	H.period	DD.period	UD.period	H.t	DD.t.i	UD.t.i	H.i	DD.i.e	UD.i.e	H.e	DD.e.five
2	s002	1	1	0.1491	0.3979	0.2488	0.1069	0.1674	0.0605	0.1169	0.2212	0.1043	0.1417	1.1885
3	s002	1	2	0.1111	0.3451	0.234	0.0694	0.1283	0.0589	0.0908	0.1357	0.0449	0.0829	1.197
4	s002	1	3	0.1328	0.2072	0.0744	0.0731	0.1291	0.056	0.0821	0.1542	0.0721	0.0808	1.0408
5	s002	1	4	0.1291	0.2515	0.1224	0.1059	0.2495	0.1436	0.104	0.2038	0.0998	0.09	1.0556
6	s002	1	5	0.1249	0.2317	0.1068	0.0895	0.1676	0.0781	0.0903	0.1589	0.0686	0.0805	0.8629
7	s002	1	6	0.1394	0.2343	0.0949	0.0813	0.1299	0.0486	0.0744	0.1412	0.0668	0.0863	0.9373
8	s002	1	7	0.1064	0.2069	0.1005	0.0866	0.1368	0.0502	0.08	0.1407	0.0607	0.0789	0.7967
9	s002	1	8	0.0929	0.181	0.0881	0.0818	0.1378	0.056	0.0747	0.1367	0.062	0.0776	0.6447
10	s002	1	9	0.0966	0.1797	0.0831	0.0771	0.1296	0.0525	0.0839	0.1425	0.0586	0.0755	0.7357
11	s002	1	10	0.1093	0.1807	0.0714	0.0731	0.1457	0.0726	0.0766	0.1241	0.0475	0.0813	0.755
12	s002	1	11	0.0887	0.166	0.0773	0.0876	0.156	0.0684	0.0839	0.1386	0.0547	0.0692	0.6927

keystroke

Table 2. Buffalo Dataset

3.1.2 Clarkson II Keystroke dataset

The Clarkson Keynote II is discovered in Clarkson University by researcher and it is free text keynote speed dataset. In this dataset, data has been stored for over 2.5 years and data includes unregulated and natural environments for 101 key time data. (Alsultan, Warwick, and Wei, 2017) Participants share their data among different desktops, browsers, software, keyboards and functions to compare the data with other controllable files. For real life situations, an effective model of data must be appropriate. However, this dataset offers only few special features like timing buttons and closed events. 125,000 is an average number of keys to each research topic. Unfortunately, some users are affected by a small number of keystrokes because the number of them is far from irregular. So, a threshold of 20,000 keys should be set, which results in just 80 entries.

3.2 SCM algorithm

The SCM (spatial distribution, central value distance, and modification algorithm)algorithm is a multi-level decision that involves triple voting. This algorithm is basically including classification of individual parts. Also, it includes priority analysis for two main algorithms one is distance based voting and other is spatial distribution-based voting. The first classification is used by the first sub algorithm, likewise second classification is used by the second algorithm. In both the sub algorithms, prediction class is deciding on the basis of voting. Afterwards, acceleration data is selected to identify the main house which serves as the basis for the third sub-algorithm. With the solution time of the above algorithm, it can be easily identifying the key. By processing these features, the first sub algorithm was developed. E.g., Two points are chosen from the sample training data, and their distribution is shown. It is cleared that the classification is different from different keys but also it is easy to use same key identification. By adding new parameters, new functions were discovered, in this research.

As shown in Figure, Correlation and independent are there are two types of relationships between different key features. Example, the 'E' and 'F' keys are in feature 'A' in Correlation type, there is a link in their feature distribution that is not completely separable. In the other side at independent type, as shown by feature ‘B’, the two characteristics are interrelated and therefore different. Therefore, the ranking algorithm is developed based on the differences

between different features. Finally, when updating the data distribution, have a high and low level for each feature, similar to the time limit. (Sha, Lian, Zhao, Yu, wang and Li, 2020)

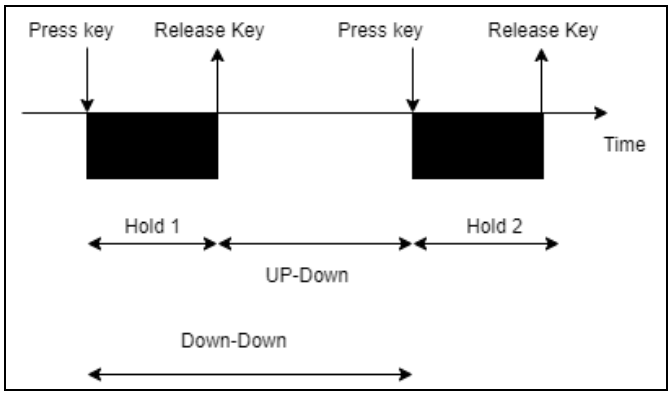


Fig. 4. Distribution of Different features

3.3 Feature selection

With the expansion of mobile devices, current switches are no longer restricted to physical switches, but now encompass the majority of virtual devices that offer access to users. Here timely data and high-quality services are important and it is examined by the pattern of human behaviour. Nevertheless, there is no direct connection between the features of the mind and the study's data. Future datasets produced by mobile devices could contain this information, which might improve the outcomes of authentication and authorisation. (Alsultan, Warwick, and Wei, 2017) Additionally, this research focusing on time-based feature, as it is in dataset of keystrokes. Following figure shows the five-fold estimate. A and B represent the interweaving of the two concepts, the press, and the question indicates the importance and importance of the event. The five are time-based, down-down time (DD time), up-down time (UD time), up-up time (UU time), and down-up time (DU time). The time the user holds the key in the down position and the other four factors are true for the image. Note that you can truncate two important events, say A and B, i.e., long-A, long-B, DD-time, UD-time, UU-time, and DU-time (Abo-alian, Badr and Tolba, 2016).

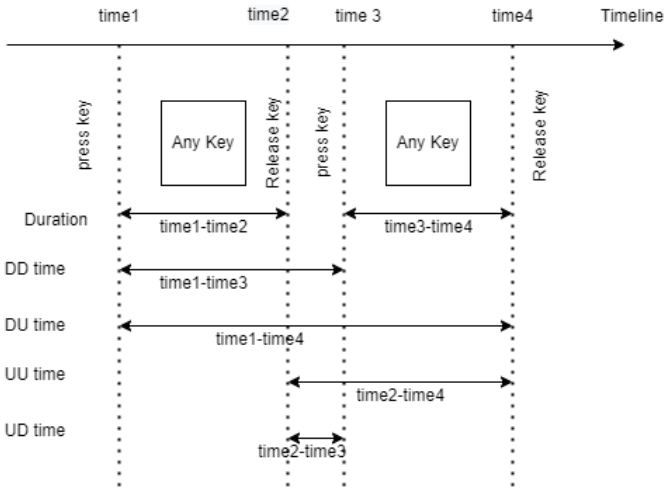


Fig. 5 Key functions

Feature subset selection (FSS) is the subsequent stage of exploration is finished.

Include choice is a cycle that chooses a subset of unique highlights. FSS is very effective at information mining and AI because it reduces the number of components and eliminates information irregularities, duplications, and clamours, hence speeding or enhancing framework insights and outcomes from multiple cycles. 4 steps of FSS- A study procedure called subgroup creation creates products that are chosen for analysis depending on measuring technology. (Akinsowon, 2021) When selecting when to stop, this measuring technique is utilized to compute the substrate. The procedures are what determine if the device is functioning or not. Algorithm FSS is divided in 3 parts and measurement values are dependant on classification , 1) Filter model, 2) Wrapper Model and, 3) Hybrid model.

So, the modification of algorithm is done according to there location of the subgroup features and the accuracy of their functional analysis. A mood vector is used in the Euclidean distance dispersion model as a point in Euclidean space. Each subject's mean vector is established during preparation. A point is classified depending on the separation between each vector record's markings. It also considers displaying raw information with standardized information, where standardization isolates information by the most extreme. A classifier in view of the separation from the assessment vectors of each subject can't distinguish the subjects that are not in the preparation information since it doesn't have the foggiest idea about the mean vector of newbies. What's more, an edge can be set to restrict the scope of every vector estimation, and in the event that the new gadget information is more noteworthy than the scope of every vector estimation of the preparation information estimated by the breaking point, new strategies can be created during testing. Be that as it may, on the off chance that the put down accounts of each subject are in a different area and the new gathering is blended in with different gatherings, the new individual will be characterized erroneously. Separated circulations are not suitable for open testing, therefore. The typical Euclidean division was also used. Classifiers, for example, k closest neighbours and distance techniques, including the Euclidean strategy, can't distinguish new characters that don't show up in the preparation information. So, the accompanying classification of open analysis are tested: Fisher LDA, GPC, SVM, brain organizations, and arbitrary woodland. F1- mean and standard deviation. For SVMs and brain organizations, the F1 fraction of the open test is an irregular gauge.

Lacklustre showing on the open test demonstrates that each subject's example should be rehashed when a renewed individual enters, or the class will misclassify the new individual as a current subject. In the open trial, F1 estimates something similar for all dispersions utilizing crude or unique information, and that implies that the presentation isn't great for standardization. F1 expanded somewhat after standardization.

3.3.1 The Filter Model

The general identification of the data is necessary for the analysis and selection of a strategic plan without a learning curve. In some cases, the optimal part of the identification or filtering system is not properly selected if the identification signal is not used to train the machine. Another drawback of the filter type is that while many filters take each feature's requirements into account separately for each service, the filter technique does not discover them to be an expensive component of the plan. (Akinsowon, 2021) Thus, the efficiency of this kind of learning is reduced.

3.3.2 The Wrapper Model

It needs a learning algorithm and utilizes that algorithm's performance as a standard. When considering the accuracy of prediction, model wrappers do better than others. Unfortunately, because they need more data counts and need specific training techniques, Cover models are less common and involve more costly filter models. (A. and L., 2018) The key advantage of using filters over folders is that filters disappear much faster than folders and hence have a shorter measurement time for files with more services than wrappers. There is no need for additional filters while using several applications. Therefore, filters can offer the same learning advantages as covers.

3.3.3 Hybrid Model

It takes advantage of filters based on the type of folding by using their different tests and different levels of analysis. Decision tree learners, such as ID3, C4.5, and SVM, are some instances of embedded approaches. (A. and L., 2018) Hybrid systems are more costly since, by monitoring the training from the beginning to the completion of the plan for each study, they provide the highest value in terms of equipment and filtration. They are complicated and can only be connected through machine learning. Filter is used for the selection of the input signal GA is detected in 2 parts of the Monk1 database and traffic data is reduced by 50% and 33% respectively.

3.4 Classification

Classification step is next step , in that combination of patterns and stored pattern are entered during the session. Classification algorithms has some categories: ANN (Artificial Neural Networks), Pattern detection, Statistical Algorithm, Learning based algorithms, Heuristic analysis, and Collective algorithms. During implementation, determine the mean and standard deviation of the final plan and model. To compare training data and test data, Distance measuring algorithms are used such as Euclidean distance, Manhattan Distance, weighted Euclidean distance, etc. Because the data collected for analysis and statistical analysis is not needed to be linear, these statistics may not always produce good results.

Because of this, there must be a method that use probabilistic data rather than deterministic data. Decision trees, Bayesian classification (based on posterior probability), and other techniques can also be used for classification. It obtained an average rejection accuracy of 9.62 % and an average acceptance accuracy of 0.88 % by using the Monte Carlo approach for

key exchange. (Kim et al., 2019) The artificial neural network is a different classification technique. Create a lengthy encryption key using the ANN approach and consider the value of key analysis to assess how likely it is that the key will be revealed without the password. You can get good results because lots of negative factors are controlled by this method which is very important. On the basis of algorithms to categorize patterns and other types of unstructured data (such as objects) is the process of pattern recognition. Modelling algorithms include machine learning algorithms, various classification methods like the nearest neighbour rule, Bayesian classifier, and Support Vector Machine, and clustering methods like K-means, etc. They discovered using SVM that retraining boosts recognition performance and learning vector quantization for new detection performs better than other widely used new tools. Adaptive optimization methods like genetic algorithms, ant colony optimization, particle swarm optimization, etc. are typically included in the last approach. These development techniques have the benefit of handling a large amount of data.

3.5 Keystroke Dynamic Classifier

For customers who reject the learning model as a mistake and accept the item as a good client, button-based confirmation provided as an order problem for a class learning model. Although using negative models in preparation can operate on hierarchical structures, it is inappropriate to know the compelling aspects of negative behaviour, not only its instructional content. System utilizes the closest neighbour with the new distance to decide the strength of the key appointed to the client if the closest neighbour distance in the preparation information doesn't meet the edge, or is dismissed as a mistake. Each single key vector of the target customer is used to determine the relationship structure. Embracing new levels during the progress time frame diminishes the adverse consequence of the mouth. (Kim et al., 2019) However, others can in any case mutilate preparing information and twist proof. Use sound-reducing measures while preparing. Analysed the estimates from the preparation data for the I^{th} trademark variable and computed the mean and standard deviation of I, which contains all preparing means except the level and level percentile estimates. Only the preparation vectors are stored, and the preparation information is cleared of lossy modifications and exceptions. After removing the exceptions from the preparation data, the relevant vectors were identified using the closest neighbour classifier and the updated distance measure. In essence, the closest neighbour order approach has resulted in two new scores: one without deduction and one with deduction.

4 Evaluation

The CMU key dynamics benchmark dataset is used to assess the method, as it gives performance statistics for comparison with other current key exchange techniques. The static ".nikita#1234" password string's keystroke variable in the CMU benchmark dataset contains information on the timing of each keystroke as well as the space between two consecutive ones. 51 items are there in this file. Each individual had eight data collecting sessions, with at least one day in between each session. Each of the 400 vectors per subject was produced once for a total of 50 vectors. The figure shows four times the key features written in three elements. Figure also displays another term feature matrix.

The keyword characteristics are correct and have a lot of variability, noise, and discrimination for each subject even if they are accurate and accurate. The equations are used to address situations involving critical processing data. (Kiyani et al., 2020) proposed system use the same policy and evaluation process to ensure that performance is comparable. The first 200 vectors for each individual are used as training data. The remaining 200 feature vectors are used as the positive evaluation data, and the first 50 segments of the remaining 50 individuals are used to generate the 250 negative vectors as individual ratings. When the error rate and false alarm rate are equal, the test accuracy is calculated using the equal error rate (ERR), and the zero false alarm rate (ZMFAR), which is the number of negatives when the value is zero. All topics are the focus of research; the means and standard deviations of error values for 51 subjects were given. The user determines if important dynamics are successful.

5 Design and Implementation

The key dynamic is user-dependent. For instance, the SCM histogram of the highest-rated system in the competition (system 6 from P4) was individually produced for each of the 300 individuals used. The findings revealed significant performance variations across users, with the majority of users reporting SCM variations of up to 20%. (Kiyani et al., 2020) Key dynamics still faces the open task of enhancing the performance of the lowest users. One option is to analyse the algorithms' complementarity. The SCM of the low-performing users using the best system published by P4 and the performance of the same users' systems published by P1 and P2. The machines using P1 and P2 show better overall performance than those using P4. However, the results show the success of the systems as P1 and P2 give better results for this user problem than P4. As a result, we can see the output of five users as well as a table showing the DD, UD, and H (Hold) timings for each user for various key presses.

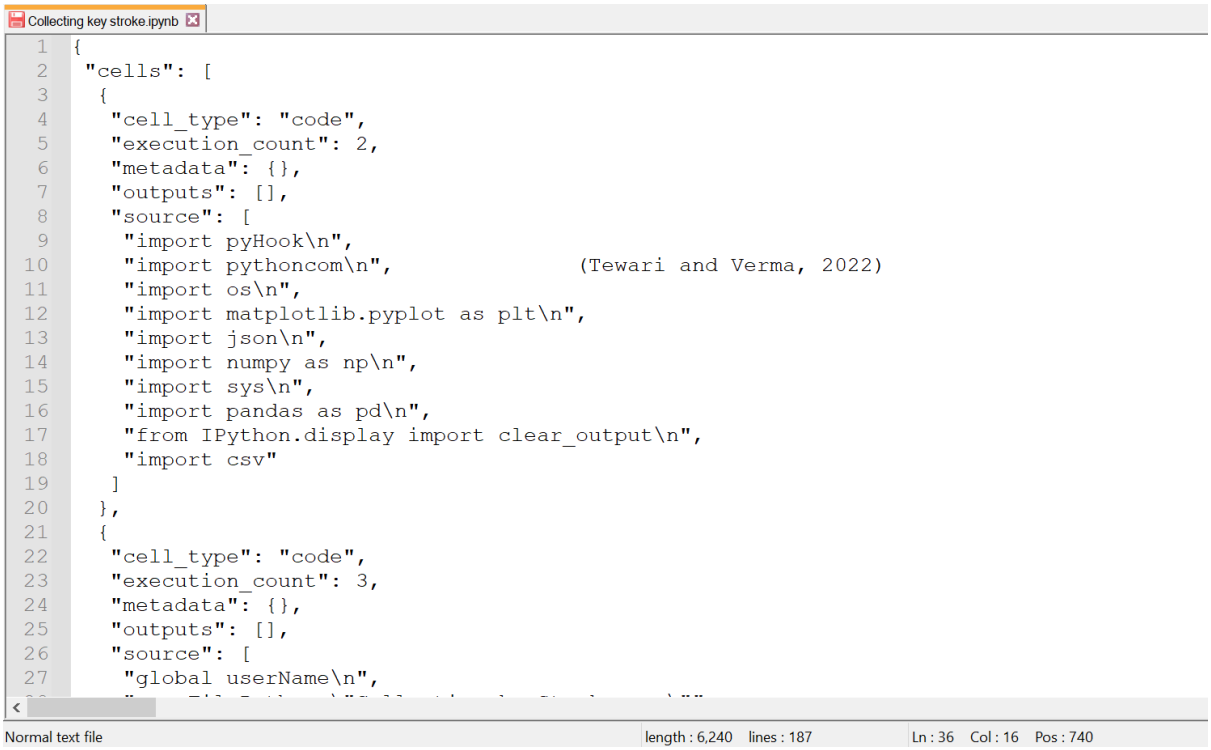
Output shown in the image below:-

subject	sessionIndex	rep	H.period	DD.period.t	UD.period.t	H.t	DD.t.i	UD.t.i	H.i	...	H.a
0	s002	1	1	0.1491	0.3979	0.2488	0.1069	0.1674	0.0605	0.1169	...
1	s002	1	2	0.1111	0.3451	0.234	0.0694	0.1283	0.0589	0.0908	...
2	s002	1	3	0.1328	0.2072	0.0744	0.0731	0.1291	0.056	0.0821	...
3	s002	1	4	0.1291	0.2515	0.1224	0.1059	0.2495	0.1436	0.104	...
4	s002	1	5	0.1249	0.2317	0.1068	0.0895	0.1676	0.0781	0.0903	...

DD.a.n	UD.a.n	H.n	DD.n.I	UD.n.I	H.I	DD.I.Return	UD.I.Return	H.Return	
0.1349	0.1484	0.0135	0.0932	0.3515	0.2583	0.1338	0.3509	0.2171	0.0742
0.1412	0.2558	0.1146	0.1146	0.2642	0.1496	0.0839	0.2756	0.1917	0.0747
0.1621	0.2332	0.0711	0.1172	0.2705	0.1533	0.1085	0.2847	0.1762	0.0945
0.1457	0.1629	0.0172	0.0866	0.2341	0.1475	0.0845	0.3232	0.2387	0.0813
0.1312	0.1582	0.027	0.0884	0.2517	0.1633	0.0903	0.2517	0.1614	0.0818

Table 3. performance of users

To calculate the time, estimate for the first data set, the subkey is identified in the key program for each key ID for each user. For fast keyboard programming, press the key one step lower. So, from the millisecond value stored in the csv file of the keystroke recording, the Hold, Down and Down values are taken into account. Ascii values between 33 and 122 are taken into account (a-z, A-Z, 0-9 and special characters). Values for each keystroke and user combination are extracted from the input data and these values are calculated and stored in the keystrokedistance.csv file.



Sceenshot 1: Collecting key value (Tewari and Verma, 2022)

“When the code will run it will ask for few input

Enter your name- Nikita

Enter your text- 13

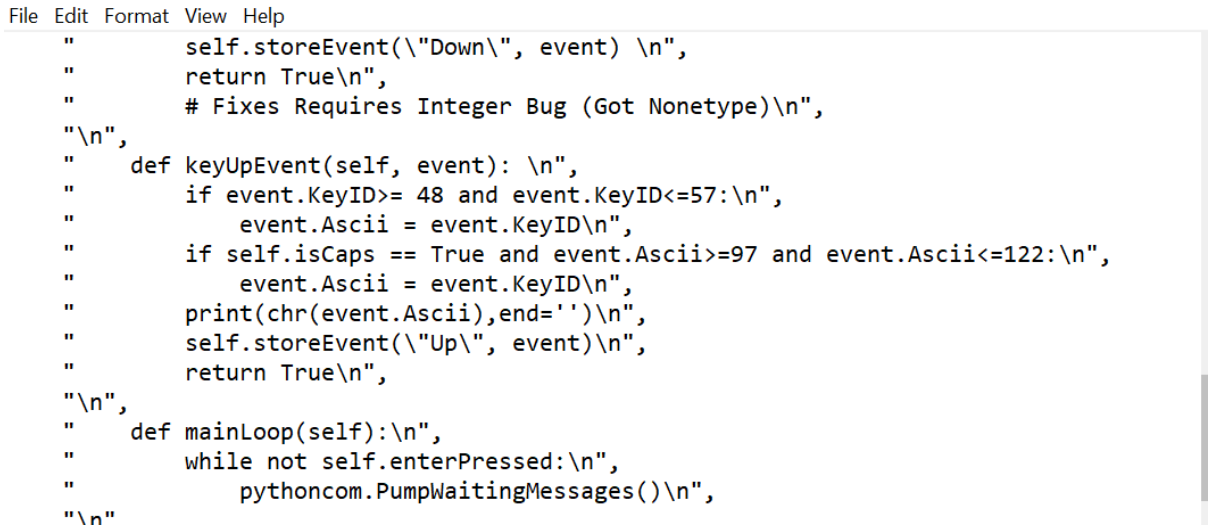
Biometrics-

As soon all thesis will get enter after that the data will get store in the collecting key csv file. The code for these is shown in the screenshot below: -“

```
File  Edit  Format  View  Help
   "execution_count": 6,
   "metadata": {},
   "outputs": [
    {
     "name": "stdout",
     "output_type": "stream",
     "text": [
      "Enter your Name: user1\n",
      "Enter your text: \n",
      "biometrics-\u0000CS-559\n",
      "ouput\n",
      "[('user1', 13, 'Up', 23933390), ('user1', 98, 'Down', 23935312), ('user1', 98, '
3, 'Down', 23948046), ('user1', 53, 'Up', 23948140), ('user1', 53, 'Down', 23948265), (
     ]
    }
   ],
   "source": [
    "class KeyLogger:\n",
    "    def __init__(self):\n",
    "        self.enterPressed = False\n",
    "        self.eventList = []\n",
    "        self.isCaps = False\n",
    "        #self.message = \"\"\n",
    "        \n",
```

Screenshot 2: User Input to calculate Biomatrics

The code will autogenerate the KeyUpEvent and store it in the keystroke.csv file so that at the time of authentication the keystroke will check the event from the csv file and if the user matches the event will be marked as true.



```
File  Edit  Format  View  Help
    "         self.storeEvent(\"Down\", event) \n",
    "         return True\n",
    "         # Fixes Requires Integer Bug (Got Nonetype)\n",
    "\n",
    "    def keyUpEvent(self, event): \n",
    "         if event.KeyID>= 48 and event.KeyID<=57:\n",
    "             event.Ascii = event.KeyID\n",
    "         if self.isCaps == True and event.Ascii>=97 and event.Ascii<=122:\n",
    "             event.Ascii = event.KeyID\n",
    "         print(chr(event.Ascii),end='')\n",
    "         self.storeEvent(\"Up\", event)\n",
    "         return True\n",
    "\n",
    "    def mainLoop(self):\n",
    "         while not self.enterPressed:\n",
    "             pythoncom.PumpWaitingMessages()\n",
    "\n"
```

Screenshot 3: The event on keys (KeyUpEvent, KeyDownEvent, KeyHoldEvent)

The below code is for manging the user list which has been collect by running the code collecting key. The code will access the collecting keystroke csv file and match the biometrics data with the csv file before giving the access to the user.

```
import pandas as pd
import numpy as np
import csv
from scipy.spatial.distance import cityblock
from sklearn.metrics import roc_curve
data = pd.read_csv("Collecting_keyStorke.csv")
userList = data.user.unique()
keyList = data.key.unique()
df = pd.DataFrame(columns=['subject','key','H','UD','DD'])
for i in range(0, len(userList)):
    for j in range(0,len(keyList)):
        queryData = data.query("user=='" +userList[i]+ "' and key==" + str(keyList[j]) + " and key >=33 and key<
        queryLen = len(queryData)
        finalData = {}
        if queryLen > 0:
            if(queryLen > 2):
                for k in range(0,queryLen,2):
                    finalData['subject'] = userList[i]
                    finalData['key'] = chr(keyList[j])
                    finalData['H'] = (int(queryData.iloc[k+1].Time) - int(queryData.iloc[k].Time))/1000
                    keyUpIndex = queryData.iloc[k+1].name
                    if(data.iloc[keyUpIndex + 1].user == userList[i]):
                        finalData['UD'] = (int(data.iloc[keyUpIndex+1].Time) - int(queryData.iloc[k+1].Time))/1(
                        finalData['DD'] = (int(data.iloc[keyUpIndex+1].Time) - int(queryData.iloc[k].Time))/100(
                    else:
                        finalData['UD'] =  finalData['H']
                        finalData['DD'] = finalData['H']
```

Screenshot 4: Key calculating value (Janakiev, N. 2022)

Analysis.ipynb | calculatingValues.ipynb | Collecting key stroke.ipynb | keycalculatingValue.py | collectingValue.py

```
import pyHook
import pythoncom
import os
import matplotlib.pyplot as plt
import json
import numpy as np
import sys
import pandas as pd
from IPython.display import clear_output
import csv
global userName
userFilePath = "Collecting_keyStorke.csv"
class KeyLogger:
    def __init__(self):
        self.enterPressed = False
        self.eventList = []
        self.isCaps = False
        #self.message = ""

    def keyDownEvent(self, event):
        if event.KeyID == 20 and self.isCaps == False:
            self.isCaps = True
        elif event.KeyID == 20 and self.isCaps == True:
            self.isCaps = False
        if event.KeyID>= 48 and event.KeyID<=57:
            event.Ascii = event.KeyID
        if self.isCaps == True and event.Ascii>=97 and event.Ascii<=122:
            event.Ascii = event.KeyID
```

Screenshot 5: Collecting value

The KBOC Baseline testing, which connected 31 computers from four participants, enabled us to get the following new insights regarding the button strength of biometric IDs. To begin with, even under challenging conditions, a small development team of 10 members and a test group of 300 users can compete with SCM at 6%. Second, optimizing the system with various distance and normalization points increases the system's accuracy. (Sheng, Phoha, and Rovnyak 2019) Third, even with a system that isn't the finest, two-month performance is the best. Finally, it's extremely simple to employ the important dynamics performance. Algorithm adaptation to various user behaviours, such as synthetic models, is still an active research topic.

```
In [17]: print("The Datasets for Manhattan distance shows error rates of ", eer1_1, eer1_2)
         print("The Datasets for K-Means distance shows error rates of ", eer2_1, eer2_2)
         print("The Datasets for Euclidean distance shows error rates of ", eer3_1, eer3_2)

The Datasets for Manhattan distance shows error rates of  0.4325842696629214 0.1947450980392157
The Datasets for K-Means distance shows error rates of  0.4245810055865923 0.21787580694749598
The Datasets for Euclidean distance shows error rates of  0.33333333333287907 0.15539215686282523
```

Table 4. Output of distance calculation with error rate

5.1 Application Area

Computer security is very important because computers are used for many transactions. A smart card is required to enter the restricted area, whereas biometric security systems require physical or behavioural access to the computer system. As a result, biometrics offer an extra layer of protection by confirming a person's physical identification. Keystroke Dynamics is one of the biometrics that has been studied in the past but hasn't yet been used to security. Because of the adaptable idea of the centre framework, numerous different applications can profit from such a distinguishing proof framework. One of the security features that KD can provide is digital defence against internet attacks, stopping fraudsters in their tracks, and avoiding reconnaissance attacks. Additionally, the KD framework is used during enrolment.

Keystroke-based confirmation, often known as TOKEN, is used for electronic and portable applications. In such instances, security must define exactly the client's character while discussing data on cell phones. Keypad-put together confirmation has been executed with respect to assembling grade cell phones to decrease the dangers related with new validation-based strategies. (Sheng, Phoha and Rovnyak, 2019) Distributed storage may make use of key trading. All cloud apps must be secure, and client personality counts greatly in security concerns. Associations such as Tera-information, Big-information, and so on are typically beneficial for client assurance, reconciliation, and business understanding. By utilizing console-based verification, their information security system is now safer, and this is the main way they can guarantee their clients that their qualifications are a wise venture.

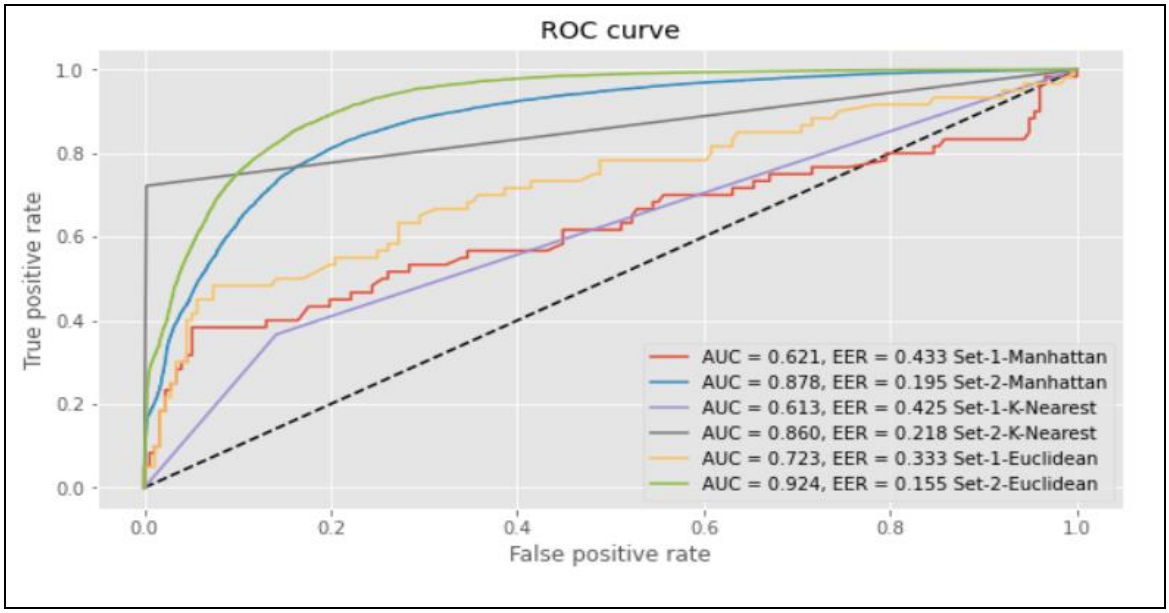
Numerous business arrangements have been created to give client ID. Psylock is a German business that designs security strategies for keystroke components for usage on many platforms, including MS Windows, online access, Citrix, and VPN reconciliation. A Swedish company called BehavioSec develops IT security frameworks based on how keystrokes and mouse movements interact. A Dutch company by the name of ID Control also provides workable solutions, some of which use keystroke components. Scout Analytics also uses social biometrics, such as keystroke patterns, to identify users and prevent them from disclosing their information to third parties.

The study of keystroke dynamics has several uses in the world of computer security. Gaining root-level access to the host server that stores the Kerberos key information is one situation when a static approach to keying dynamics becomes especially helpful. Along with their username and password, every user that accesses the server is required to enter additional

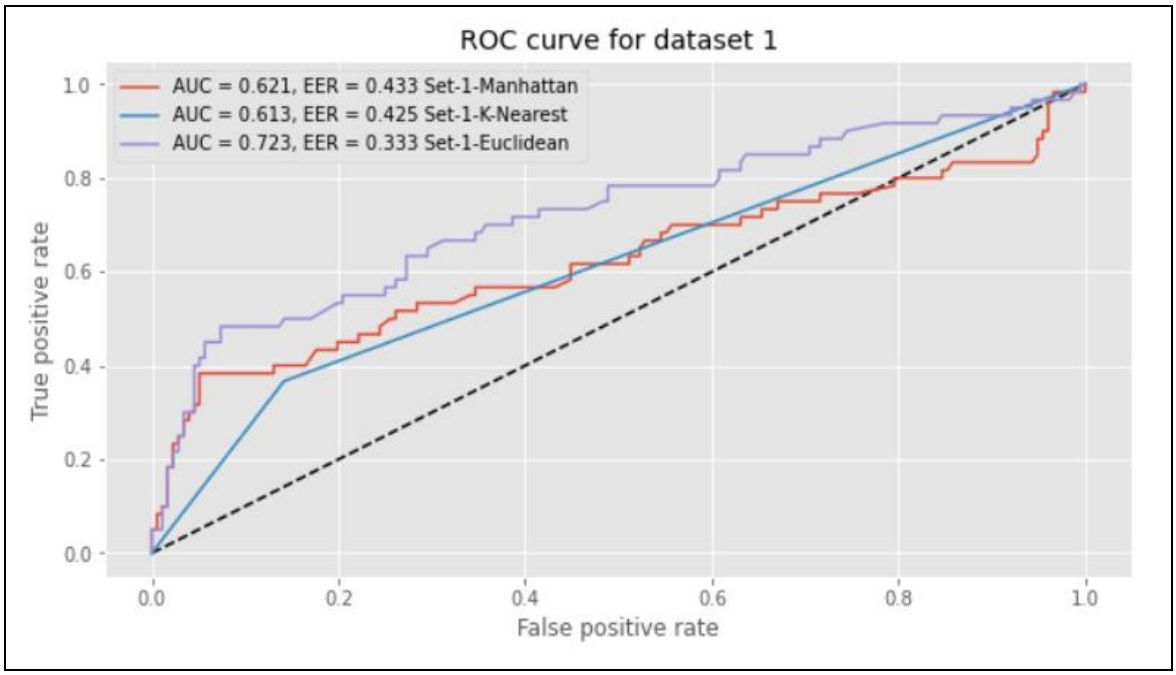
password. Access is provided if your device model is available in the requested band. Because the server typically only has console access and no remote access, this security is helpful. Furthermore, continuous or continuous monitoring of user interaction is best used for keystrokes when accessing restricted information or working in a setting where the user must "alert" (for example, during traffic control). Keystroke dynamics can be utilized to identify the user's odd typing habits (fatigue, tiredness, etc.) and alert others.

6 Result

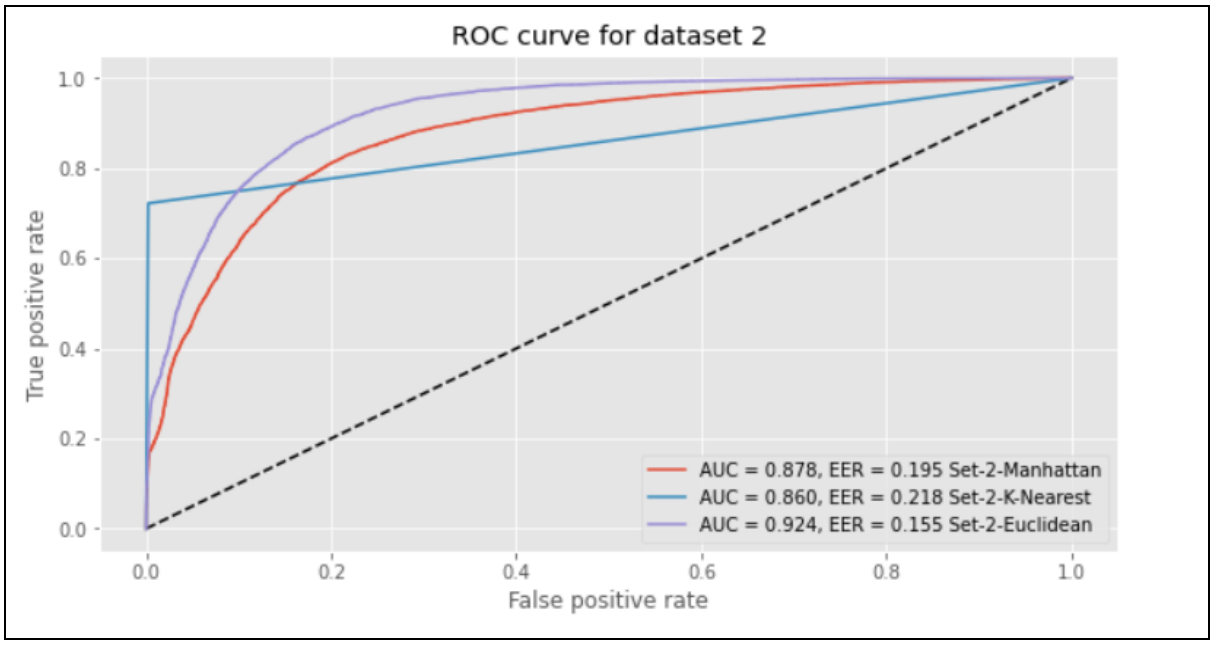
ROC curves are used to compare the findings in this instance. Two datasets are present, one of which is regulated and the other of which is a self-exploring gibberish dataset. The Manhattan distance, the Euclidean distance, and the k-word method are the three models that are utilized. Even if the error rate appears poor in comparison, it produces superior results since it guarantees at least 55–60% accuracy in the worst-case scenario.



Output 1: ROC curve



Output 2: ROC curve for Dataset 1



Output 2: ROC curve for Dataset 2

For the sake of simplicity, the number of sensors that can be drawn on the finger at the time of writing has been limited. (Tewari and Verma, 2022) During testing, it was found that typing with a touch-sensitive finger and a non-touching finger affected the text. Some studies researched the link between the touch and non-touch of the fingers when pressing the keys and obtained some results.

Therefore, it can identify a key with a small sensor. As shown in the image (Output), the "E" and "F" keys are recognized only by the two listed features, which are recognized by the sensor mounted on the middle finger. This result shows that even if an ID cannot be extracted

from a fingerprint, it can be identified from other unauthorized fingerprints. It also makes us realize that with reduced sensors we can make more discoveries.

The speed data is gathered in order to calculate the fingerprint function stated in SCM's third sub-algorithm. A smaller ring means there is less z-axis velocity data from the finger sensor to adjust the algorithm. As a result, the identification of the home button keys will be weak, especially if the number of sensors is lowered from four to two. Another perspective is that SCM algorithms can reveal differences in how different individuals write. This allows us to obtain unique requests for each person.

Although the keys on the keyboard are all unique, there are variations between persons. There is some depth in classifying the properties of the various materials. The form of each individual's keystrokes is unique and varies from person to person as a result of differences in the length and direction of the fingers' movements while pressing a key. In this instance, the output is more than when the identical button must be pressed manually. It may be concluded that this variation may result in a difference in the main results. Furthermore, observational study suggests that variations in typing proficiency may result in variations in subject-to-subject recognition accuracy. (Tewari and Verma, 2022) This trend can be interpreted as a change in people's behaviour. As shown in Figure, one result, for example, is that the keys "E" and "W" can be easily confused with B items due to their similar features, but they are complicated by the meaning of C, which it seems to confuse the keys 'C' and 'X'. With the help of our SCM technique, one can quickly discover various key distributions based on the above study. This allows us to better understand the various user's device models, which allows us to not only solve the algorithmic problem but also identify which feature is optimal for each user. Additionally, will be possible to use various strategies to change how each individual sees the differences between the keys, including the layout of the keyboard and other components.

7 Conclusion and Recommendation

In this report, explained the importance of using key variables such as biometrics to determine access to the workplace. Keyboard Dynamics is a method of analysing how people type by looking at keyboard input and modifying it based on typing patterns. (Tewari and Verma, 2022) In the proposed system review the current state of critical change and current distribution based on comparative models and Bayesian probabilistic models. System contends that even while using behavioural features (such as personal characteristics) as indicators of behaviour has limits when using practices, the challenge of key qualities permits the formation of a more comprehensive research than other approaches focused on culture.

The limitations of using keystrokes as an analysis method are only defined by the nature of the "signature" of the user - identify the user according to the type of sound they make in their mode and use the performance as it is. performance - assembly is a function of the user and the environment. The problem with biometrics is that unlike static biometrics (such as voice) there are no attributes or special variables for carrying separate information

Fortunately, in the past few years, scientists have produced remarkable findings that show that different people perceive characteristics in their voices that may be in identity and these characteristics can be effectively used and used for identification. The effectiveness of data categorization for 63 users ranges from 83.22% to 92.14% depending on the strategy selected. (Toosi and Akhaee, 2021) According to this study, typists produce graphics in a variety of ways that are clearly distinct from one another. In order to avoid this, the research suggests using a digraph difference scaler rather than a low-pass filter. Also suggests use text while performing timing analysis rather than allowing users to input random text (like "blank text"). Although plain text recognition is more difficult, it shown substantial changes in performance due to the fact that concepts are disconnected, the user might be uncooperative, and environment variables are uncontrollable, limiting what can be done with free text.

The paper lists the key characteristics of laptops, PCs, and other devices. Because no additional hardware connections are required, the desktop keyboard's capability is affordable. (Toosi and Akhaee, 2021) The paper lists the key characteristics of laptops, PCs, and other devices. Because no additional hardware connections are required, the desktop keyboard's capability is affordable. The thesis concludes with a comparison of the information saved on the user and the login ID for analysis. Keystroke Dynamics uses two-point biometric security, thus in order to enter the system, you must first know the password and then match the type. Individual keyboard security is required in human behaviour to develop a programme. In other words, hardware biometrics are vital, but our human behaviour may form a security key to secure the password. This key is produced by human behaviour, such as When the user inputs the password, utilise the keyboard shortcut to enter the password. In milliseconds, the key's programming determines the timing difference. The different keyboard is a drawback to this function.

However, if you put a lot of effort into this task and figure it out, you'll do better on this crucial task. The greatest results are obtained when using a similar material across all keyboards. The research paper shows a number of helpful aspects. Future research will thus focus on finding additional characteristics or feature combinations that help to improving the accuracy of biometric systems. The size and form of the password are two factors that have an impact on performance. The difficulty with long passwords is that they are hard to remember and recognised if a short password is used, so the size and type of passwords should be an area that requires more research. Static and continuous analytic techniques were the topics of discussion. (Wang, Guo and Ma, 2017) The type of application determines which approach is better and which method should be used. The instrument samples include noise and several limitations as a result of user conflicts, which might lead to low accuracy. As a result, routing is a critical step in key exchange that should not be performed manually. The primary issue in this area is the lack of a standard method for measuring keys in order to assure accuracy and comparability. (The majority of FAR, FRR, and EER areas aren't included.) Key encryption is still a growing area that faces several obstacles before it can be used as a biometric.

The properties of computer user authentication keystrokes were investigated, and a new distance was presented that separates relevant information, normalizes feature changes, and suppresses content. Since keystroke dynamics data always contain incoherence and correlation, it is not surprising that classifiers using the new distance outperform keystroke dynamics ensembles using the distance measure. (Wang, Guo and Ma, 2017) Although use of the new distance measure for the dynamic button features matching problem, it is a general metric that can be used for all estimators in feature vector regions where the traditional Mahala nobis distance is required, with the advantage that relating to robustness. Who use the static text technique of distance assessment to increase the accuracy of packet dynamics? Future work should focus on improving keystroke characteristics, studying content sub-words and related structures, and applying our new technique to the challenging problem of keystroke biometrics using free text.

References

1. A., R. and L., A. (2018). Keystroke Dynamics for User Authentication and Identification by using Typing Rhythm. *International Journal of Computer Applications*, 144(9), pp.27–33. doi:10.5120/ijca2016910432.
2. Abo-alian, A., Badr, N.L. and Tolba, M.F. (2016). Keystroke dynamics-based user authentication service for cloud computing. *Concurrency and Computation: Practice and Experience*, 28(9), pp.2567–2585. doi:10.1002/cpe.3718.
3. Akinsowon, O. (2021). Keystroke Dynamics for User-Authentication on Mobile Devices using Ensemble Method. *Communications on Applied Electronics*, 7(36), pp.33–38. doi:10.5120/cae2021652885.
4. Alsultan, A., Warwick, K. and Wei, H. (2017). Non-conventional keystroke dynamics for user authentication. *Pattern Recognition Letters*, 89, pp.53–59. doi:10.1016/j.patrec.2017.02.010.
5. Alshanketi, F., Traore, I. and Ahmed, A.A. (2016). Improving Performance and Usability in Mobile Keystroke Dynamic Biometric Authentication. *2016 IEEE Security and Privacy Workshops (SPW)*. doi:10.1109/spw.2016.12.
6. Andrean, A., Jayabalan, M. and Thiruchelvam, V. (2020). Keystroke Dynamics Based User Authentication using Deep Multilayer Perceptron. *International Journal of Machine Learning and Computing*, 10(1), pp.134–139. doi:10.18178/ijmlc.2020.10.1.910.
7. Baynath, P., Soyjaudah, S. and Khan, M. (2018). Feature Selection and Representation of Evolutionary Algorithm on Keystroke Dynamics. *Intelligent Automation and Soft Computing*, p.-1--1. doi:10.31209/2018.100000060.
8. Boakye Osei, M., Opanin Gyamfi, E. and Okoe Alhassan, M. (2020). Keystroke Dynamics Algorithm for Securing Web-based Password Driven Systems. *Asian Journal of Research in Computer Science*, pp.1–26. doi:10.9734/ajrcos/2019/v4i430119.
9. Campisi, P., Maiorana, E., Lo Bosco, M. and Neri, A. (2017). User authentication using keystroke dynamics for cellular phones. *IET Signal Processing*, 3(4), p.333. doi:10.1049/iet-spr.2008.0171.
10. Chandranegara, D.R., Wibowo, H. and Minarno, A.E. (2020). Combined scaled manhattan distance and mean of horner's rules for keystroke dynamic authentication. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(2), p.770. doi:10.12928/telkomnika.v18i2.14815.
11. Chandrasekar, V., Kumar, S.S. and Maheswari, T. (2015). Authentication based on keystroke dynamics using stochastic diffusion algorithm. *Stochastic Analysis and Applications*, 34(1), pp.155–164. doi:10.1080/07362994.2015.1112291.
12. Chandrasekar, V. and Suresh Kumar, S. (2015). A dexterous feature selection artificial immune system algorithm for keystroke dynamics. *Stochastic Analysis and Applications*, 34(1), pp.147–154. doi:10.1080/07362994.2015.1110707.

13. citeseerx.ist.psu.edu. (n.d.). *Download Limit Exceeded*. [online] Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.738.8408&rep=rep1&type=pdf [Accessed 12 Aug. 2022].
14. Crawford, H. (n.d.). *Keystroke Dynamics: Characteristics and Opportunities*. [online] Available at: https://research.fit.edu/media/site-specific/researchfitedu/l3hiai/documents/l3hiai-publications/Keystroke-dynamics_-Characteristics-and-opportunities.pdf [Accessed 12 Aug. 2022].
15. Euler Movement Firefly Algorithm and Fuzzy Kernel Support Vector Machine Classifier for Keystroke Authentication. (2019). *International Journal of Innovative Technology and Exploring Engineering*, 8(11), pp.2267–2274. doi:10.35940/ijitee.k2063.0981119.
16. Fouad, K.M., Hassan, B.M. and Hassan, M.F. (2016). User Authentication based on Dynamic Keystroke Recognition. *International Journal of Ambient Computing and Intelligence*, 7(2), pp.1–32. doi:10.4018/ijaci.2016070101.
17. Garba, N., Rakshit, S., Maa, C.D. and Vajjhala, N.R. (2021). An email content-based insider threat detection model using anomaly detection algorithms. *SSRN Electronic Journal*. doi:10.2139/ssrn.3833744.
18. He, L., Li, Z. and Shen, C. (2018). Performance evaluation of an anomaly-detection algorithm for keystroke-typing based insider detection. *Tsinghua Science and Technology*, 23(5), pp.513–525. doi:10.26599/tst.2018.9010014.
19. Ho, J. and Kang, D.-K. (2018). Mini-batch bagging and attribute ranking for accurate user authentication in keystroke dynamics. *Pattern Recognition*, 70, pp.139–151. doi:10.1016/j.patcog.2017.05.002.
20. Janakiev, N. (2022). *Biometric Prediction on Keystroke Dynamics*. [online] GitHub. Available at: https://github.com/njanakiev/keystroke-biometrics/blob/master/keystroke-biometrics.ipynb [Accessed 12 Aug. 2022].
21. Kim, D.I., Lee, S. and Shin, J.S. (2020). A New Feature Scoring Method in Keystroke Dynamics-Based User Authentications. *IEEE Access*, [online] 8, pp.27901–27914. doi:10.1109/ACCESS.2020.2968918.
22. Kim, Park, Kim, Cho and Kang (2019). Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms. *Applied Sciences*, 9(19), p.4018. doi:10.3390/app9194018.
23. Kiyani, A.T., Lasebae, A., Ali, K., Rehman, M.U. and Haq, B. (2020). Continuous User Authentication Featuring Keystroke Dynamics Based on Robust Recurrent Confidence Model and Ensemble Learning Approach. *IEEE Access*, 8, pp.156177–156189. doi:10.1109/access.2020.3019467.
24. Pashchenko, D.V., Balzannikova, E.A. and Sergina, I.G. (2018). USER IDENTIFICATION METHOD BY MEANS OF BIOMETRIC IMAGE OF KEYSTROKE DYNAMICS WITH DOUBLE-CHAINED REPRESENTATION. *Issues of radio electronics*, (12), pp.83–89. doi:10.21778/2218-5453-2018-12-83-89.
25. Sae-Bae, N. and Memon, N. (2022). Distinguishability of keystroke dynamic template. *PLOS ONE*, 17(1), p.e0261291. doi:10.1371/journal.pone.0261291.

26. Sheng, Y., Phoha, V.V. and Rovnyak, S.M. (2019). A Parallel Decision Tree-Based Method for User Authentication Based on Keystroke Patterns. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 35(4), pp.826–833. doi:10.1109/tsmcb.2005.846648.
27. Tewari, A. and Verma, P. (2022). An Improved User Identification based on Keystroke-Dynamics and Transfer Learning. *Webology*, 19(1), pp.5369–5387. doi:10.14704/web/v19i1/web19360.
28. Toosi, R. and Akhaee, M.A. (2021). Time–frequency analysis of keystroke dynamics for user authentication. *Future Generation Computer Systems*, 115, pp.438–447. doi:10.1016/j.future.2020.09.027.
29. Wang, X., Guo, F. and Ma, J. (2017). User authentication via keystroke dynamics based on difference subspace and slope correlation degree. *Digital Signal Processing*, 22(5), pp.707–712. doi:10.1016/j.dsp.2012.04.012.