A User Authentication Tool Based On Behavioral Biometrics.

Kiprop, Valerie Cheruto

146640

CNS

Supervisor Name

Mr. James Gikera

Submitted in Partial Fulfillment of the Requirements of the Bachelor of Science in

Computer Networks and Cybersecurity at the Strathmore University

School of Computing and Engineering Science

Strathmore University

Nairobi, Kenya

May 2023

# Declaration and Approval

I declare that this work has not been previously submitted and approved for the award of a

degree by this or any other University. To the best of my knowledge and belief, the research

proposal contains no material previously published or written by another person except where

due reference is made in the research proposal itself.

Student Name: Valerie Kiprop

Admission Number: 146640

Student Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The Proposal of Valerie Kiprop has been reviewed and approved by Mr. James Gikera

Supervisor Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# ABSTRACT

With the evolution of mobile services and technologies, mobile threats are becoming increasingly prevalent. One solution to this problem is adding an extra layer of protection for example behavioral biometrics. This project focuses on using keystroke dynamics to develop a tool for user authentication on a payment form. It involves analysing factors such as a person's typing speed, and mouse movement, which are almost impossible for someone to replicate or steal. This paper presents a tool uses a machine learning algorithm, Gaussian Naïve Bayes, to create unique digital identities that can be used during online card transactions to distinguish legitimate users from fraudsters. It receives user input data entries from a graphical user interface, similar to an online payment form, and transforms them into unique digital identities. The tool is implemented in Python. This provides a stronger security system.

# Table of Contents

[Declaration and Approval ii](#_Toc136618063)

[ABSTRACT iii](#_Toc136618064)

[Table of Contents iv](#_Toc136618065)

[List of Figures v](#_Toc136618066)

[List of Tables vi](#_Toc136618067)

[CHAPTER 1: INTRODUCTION 1](#_Toc136618068)

[1.1 Background 1](#_Toc136618069)

[1.2 Problem Statement 2](#_Toc136618070)

[1.3 Objectives 2](#_Toc136618071)

[1.3.1 General Objectives 2](#_Toc136618072)

[1.3.2 Specific Objectives 2](#_Toc136618073)

[1.4 Research Questions 3](#_Toc136618074)

[1.5 Justification 3](#_Toc136618075)

[1.6 Scope and Limitations 3](#_Toc136618076)

[CHAPTER 2: LITERATURE REVIEW 5](#_Toc136618077)

[2.1 Introduction 5](#_Toc136618078)

[2.2 Viability of Keystroke Dynamics in User Authentication 5](#_Toc136618079)

[2.3 Designing and Implementing a Prototype Authentication Model 5](#_Toc136618080)

[2.5 Determining How well the System Works. 5](#_Toc136618081)

[2.6 State of Art in the Current Situation 6](#_Toc136618082)

[2.7 Related Works 6](#_Toc136618083)

[2.8 Conceptual Framework 8](#_Toc136618084)

[CHAPTER 3: METHODOLOGY 9](#_Toc136618085)

[3.1 Introduction 9](#_Toc136618086)

[3.2 Methodology 9](#_Toc136618087)

[3.2.1 Justification 10](#_Toc136618088)

[3.3 Design and Development tools 10](#_Toc136618089)

[3.3.1 User Interface 10](#_Toc136618090)

[3.3.2 Data Processing 10](#_Toc136618091)

[3.3.3 Machine Learning 11](#_Toc136618092)

[3.4 Deliverables 12](#_Toc136618093)

[References 13](#_Toc136618094)

# List of Figures

[Figure 1.1 8](#_Toc136616802)

[Figure 2.1 11](#_Toc136616803)

[Figure 3.1 11](#_Toc136616804)

# List of Tables

[Table 1.1 6](#_Toc136617092)

[Table 2.1 9](https://d.docs.live.net/b3988de70d9dae2f/Documents/project%201/Project%201.docx#_Toc136617093)

# CHAPTER 1: INTRODUCTION

## 1.1 Background

In today's digital landscape, ensuring secure and reliable user authentication is of paramount importance. User authentication methods, such as passwords and PINs, are mostly used for authentication. There’s also physical biometrics which refer to the identification of individuals based on their unique physical characteristics and traits such as fingerprints, facial features, iris and retina patterns. However, due to numerous security-oriented threats, it has previously been demonstrated that authentication with a single element is insufficient to provide effective protection for users’ identities. (Pilar et al., 2012). As a result, there is a growing need for more robust and advanced authentication techniques that provide stronger security while maintaining user convenience. Behavioral biometrics has emerged as a promising solution to address the limitations of traditional authentication methods. By leveraging the unique behavior patterns exhibited by individuals, such as keystroke and mouse use characteristics, commonly known as keystroke, mouse and touchscreen (KMT) dynamics, gait, or voice, behavioral biometrics offers a new dimension of authentication that enhances security and accuracy. (Yampolskiy & Govindaraju, 2008) The concept behind behavioral biometrics lies in the recognition and analysis of individual behavior patterns. This technology capitalizes on the fact that human behavior is inherently difficult to replicate or imitate, providing an additional layer of authentication that is challenging for malicious actors to bypass. This paper seeks to provide a comprehensive understanding of behavioral biometrics and its applications in user authentication. It presents a software that uses keystroke dynamics to provide secure payment authentication, without compromising the user experience. The software captures user keystroke dynamics, whilst users enter card details on a payment form. Then, measurable patterns generated from the data processing unit are processed using a machine learning algorithm to generate a uniquely identifying model for the user. The model can be tested subsequently with KMT dynamics input from the same and/or other user(s) and authentication is either approved or denied. Furthermore, this study will discuss the potential benefits of integrating behavioral biometrics into existing authentication systems. It will explore how behavioral biometrics can enable continuous and passive authentication. Additionally, the paper will delve into the possibilities of combining behavioral biometrics with other authentication factors to create multi-factor authentication systems that offer heightened security measures. While behavioral biometrics offers exciting opportunities for improving user authentication, this paper will also address potential challenges and considerations. Privacy concerns and performance issues are among the factors that need to be carefully examined when implementing behavioral biometrics solutions. (Tran, Turnbull, & Hu, 2021). In conclusion, this paper aims to shed light on the innovative field of behavioral biometrics and its potential in revolutionizing user authentication. The integration of behavioral biometrics into authentication systems has the potential to enhance security, improve user experience, and mitigate the risks associated with traditional authentication methods.

## 1.2 Problem Statement

Maintaining the security of digital and non-digital assets requires the capacity to identify or authenticate a person. Passwords or physical tokens (such as ID cards) are frequently used to offer security. However, due to numerous security-oriented threats, it has previously been demonstrated that authentication with a single element is insufficient to provide effective protection for users’ identities. Passwords, for example, have long been a popular method for user authentication due to their simplicity and familiarity, they are although susceptible to a range of security risks, including weak passwords, password reuse, and the potential for brute-force attacks. Physical biometrics also suffer from similar drawbacks and can be subject to spoofing attacks, where attackers mimic or replicate biometric characteristics to gain unauthorized access. (Syed, 2019) These authentication methods can be difficult to use at times, and behavioural biometrics offers a viable solution.

## **1.3 Objective**s

### 1.3.1 General Objectives

1. To develop a robust and reliable user authentication tool based on behavioral biometrics.

### 1.3.2 Specific Objectives

1. To identify and analyze behavioral biometrics modalities specifically keystroke, mouse and touchscreen dynamics to determine their viability for user authentication.

2. To design and implement a prototype authentication system that integrates the keystroke dynamics to capture behavioral data that is used for authentication.

3. To conduct tests to assess how well the system works

## 1.4 Research Questions

1. How viable are keystroke, mouse and touchscreen dynamics for user authentication?

2. How can a prototype authentication system that integrates keystroke dynamics to capture behavioral data be designed and implemented?

3. What tests can be conducted to assess how well the system works?

## 1.5 Justification

User authentication is a critical aspect of securing digital systems and protecting sensitive information from unauthorized access. Traditional authentication methods, such as passwords and PINs, are susceptible to various security threats, including password cracking, social engineering, and phishing attacks. Consequently, there is a growing need for more robust and reliable authentication solutions. Behavioral biometrics has emerged as a promising field within the realm of user authentication, by analyzing and recognizing these patterns, a behavioral biometrics-based authentication tool can provide a more secure and convenient user authentication experience. Its resistance to attacks, user-friendly nature, continuous authentication capabilities, and compatibility with modern devices make it an attractive alternative to traditional authentication methods. By implementing such a tool, organizations can enhance security, protect sensitive information, and provide a seamless user experience in an increasingly digital and interconnected world. (Alsaadi, 2021).

## 1.6 Scope and Limitations

The scope of this project focuses on the development and implementation of a user authentication tool utilizing keystroke, dynamics captured from an online payment form. The primary objective is to leverage the unique behavioral patterns exhibited by individuals during input activities, such as typing, mouse movements, and touchscreen interactions, to establish a reliable and convenient authentication mechanism. The selected features from the data pre-processing unit which are the mean time of keyboard dwell and the mean flight time, as well as the mean mouse trajectory per keystroke dynamics test session, are fed into the machine learning unit that uses scikit-learn implementation of a simple Gaussian Naïve Bayes classifier to build a classifier that can differentiate between genuine users and potential impostors, enhancing the security of online payment transactions.

However, it is important to acknowledge certain limitations. Firstly, the accuracy of KMT dynamics analysis heavily relies on the quality of data collection and the consistency of user behavior. Factors like variations in device types, network conditions, and user fatigue may introduce noise or affect the reliability of the captured dynamics.

However, this project aims to make it easier to capture the data as it comes from only one source, a hardware keyboard hence mitigating some of these limitations eg device type.

# 

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

The use of keystroke dynamics biometrics for identifying user was first studied in the 1970’s. The typing rhythm, i.e., pressed and released of each key, of each individual was recorded and measured to distinguish their uniqueness. It is still used in many current biometrics systems [6].

## 2.2 Viability of Keystroke Dynamics in User Authentication

Since the late twentieth century, desktop computers have been widely used, and keyboards are the most common input devices. Numerous research in personal identification have focused on computers and hardware keyboard [7]. There are various reasons why keystroke dynamics is a viable method for user authentication. It is unique: It has the ability to measure timing data up to nanoseconds for keystroke events to precisely identify a user. 2. Low Implementation and Deployment Cost: Keystroke dynamics recognition does not require additional hardware equipments and can be completely implemented by software. 3. Transparency: User may not be aware with the fact that an extra authentication layer using keystroke dynamics is implemented in some systems, as it is implemented at the back end, thus helping users even with no technical background. 4. Reliable: With the involvement of Keystroke dynamics in password authentication scheme, its reliability can be increased.[17](Akshat Shah et al., 2015)

## 2.3 Designing and Implementing a Prototype Authentication Model

To implement keystroke dynamics, a machine learning algorithm, Gaussian Naive Bayes (GNB) is used. GNB is a probabilistic classifier that uses Bayes’ theorem to estimate the probability of a user belonging to a particular class based on the features of their typing pattern. It is a simple and effective algorithm that can be used to classify keystroke dynamics data. To get started, keystroke dynamics data from users is to be collected. This will be done in the data pre-processing unit where user dynamics are captured and reduced to a set of features which include, are the mean time of keyboard dwell and the mean flight time, as well as the mean mouse trajectory per keystroke dynamics test session. Once you have collected the data, you can use it to train a GNB model.[8]

## 2.5 Determining How well the System Works.

The software is preloaded with 10 sets of false user keystroke dynamic instances collected from multiple individuals entering the ‘fictitious’ card details displayed on the user interface. Upon initial instantiation, the software requires the true user to enter the same card details 10 times. With a saved model in place, subsequent entries of the same card details displayed on the user interface are considered test data instances that are used to test the model’s authentication accuracy in distinguishing between fraudsters and legitimate users/card owners.

## 2.6 State of Art in the Current Situation

Keystroke dynamics has been considered in various research areas some of them include: Continuous Authentication: Keystroke dynamics can be used for continuous user authentication, where the typing patterns are continuously monitored and compared against the user's baseline profile. This allows for real-time detection of suspicious activities or unauthorized access, (Herath, Dulanga, Tharindu, & Ganegoda, 2022)[9].

Huang et al. [10] A Piezoelectric touch panel was used to extract Keystroke functionality. Machine learning classifiers are used to determine whether a user is legitimate or not. An EER of 0.720 percent was achieved using the Piezoelectric method for data collection and a Machine Learning classifier for authentication. Piezoelectric force touch panels are less expensive because they only need to record 3D touches and do not require any additional components. Since smart devices and systems use pressure-based touch input, this is an efficient process.

Kim et al. [11] Implementation of keystroke dynamics-based authentication with the added factor of PIN-based authentication by collecting data from actual users is tested. The method to select the features is applied and compared to other methods. The proposed feature selection method outperforms the others. This is because it only concentrates on the mean of the data selected. The same feature scoring method can be used in other data sets as well other than keystroke dynamics as it proves to be efficient.

## 2.7 Related Works

**Table 1.1**

|  |  |  |
| --- | --- | --- |
| **OTHER WORKS** | **LIMITATIONS** | **THIS PROJECT** |
| In the paper “Understanding keystroke dynamics for smartphone users authentication and keystroke dynamics on smartphones built-in motion sensors"[12] A multimodal approach is used for biometric authentication by using fixed as well as variable patterns while typing OTP. Two different fusion models are used. | Obtained their data from various sources and it is concluded that more work is to be done on obtaining typing patterns. | The typing patterns in this project are generated from 3 main features the mean time of keyboard dwell and the mean flight time, as well as the mean mouse trajectory per keystroke dynamics test session |
| The paper “Statistical median-based classifier model for keystroke dynamics on mobile devices” [13] discusses how keystroke dynamics-based authentication systems could be built into mobile phones | This project focuses on using a virtual keyboard on touchscreens and thus concluded that due to the inclusion of touch features in the mobile a lower EER is recorded. | This project focuses on keystroke dynamics where the input is taken from a hardware keyboard. The error rate from a hardware keyboard is less than from a virtual keyboard |
| The paper “A survey paper on keystroke dynamics authentication for current applications" [14] focuses on different areas where keystroke dynamics could be used for authentication eg mobile devices, online learning, health etcIt discusses ways keystroke dynamics could be used in these sectors | This paper focuses on all the areas keystroke dynamics could be applied for authentication but fails to highlight how effective it works in those areas and their limitations | This project focuses on the financial sector, mainly a payment form. It shows how the tool created will be able to use keystroke dynamics to authenticate a user as they enter their card details and password to purchase an item. The model built will also show the effectiveness of the tool. |

## 2.8 Conceptual Framework

A picture containing diagram, line, plan, text

Description automatically generated

Figure 1.1

# CHAPTER 3: METHODOLOGY

## 3.1 Introduction

The methodology chapter aims to outline the systematic approach employed to develop a robust user authentication system using the Gaussian Naive Bayes algorithm. User identity verification is a crucial aspect of ensuring secure access to digital systems and protecting sensitive information. By leveraging the Gaussian Naive Bayes algorithm, which is a probabilistic classifier, this research seeks to create a reliable and efficient mechanism for accurately identifying users. This chapter will provide a detailed overview of the data collection process, feature selection, model training, and evaluation techniques utilized to construct the user identity model. Furthermore, it will address the limitations and potential challenges associated with the implementation of this methodology. The ultimate objective is to establish a highly accurate and dependable user authentication system that enhances security measures in various domains.

## 3.2 Methodology

The software algorithm captures and stores the unique characteristics of a user, so that when authentication is needed, new data is captured and compared to the stored record. The identity is confirmed if the data matches and rejected otherwise. A machine learning model is trained and stored for making such predictions on new data. The model requires ten data samples from the legitimate card owner (true data) and ten from fraudsters (false data). Thus, the software presented in this paper is pre-loaded with ten pre-processed false data captured from multiple users entering a fictitious card detail displayed on the graphical user interface. Then, the current user is required to enter the same fictitious card detail ten times on the graphical user interface, which forms the true data that is combined with the pre-loaded false data to train and store a model. Subsequent entries of the same fictitious card details (irrespective of the user) are regarded as test data and used by the model to predict whether it is the user who entered the true data or not. This illustration of the software from a user’s viewpoint is shown in Algorithm 1. The software is implemented in Python and designed to run on a laptop or personal computer with Microsoft Windows operating system installed.

Table 2.1

**Ensure:** trainData ← falseData trainData ← trueData **do** test model

**Require:** trainData ← trueData **do** feature engineer trainData **end if**

newData = input (“enter card details”) **do** train model **end if**

trueData ← newData from 1 to 10 **else**

testData ← newData from 11 to ∞ **if** model ←True **then**

**if** model ← False **then do** feature engineer testData

\*falseData contains pre-existing entries from multiple users

\*trueData must be from a single user

### 3.2.1 Justification

The utilization of a Gaussian Naive Bayes classifier in this research serves as an effective method due to its simplicity, efficiency, and ability to handle continuous data. The selected features, namely the mean time of keyboard dwell, mean flight time, and mean mouse trajectory per keystroke dynamics, have been carefully chosen due to their significance in accurately capturing and distinguishing individual typing patterns [16]. These features generated by the data processing unit ensure a robust foundation for user identification.

Overall, the combination of the Gaussian Naive Bayes classifier and the selected features derived from the data processing unit provides a robust foundation for authenticating user identities based on their typing patterns. The classifier's ability to capture the inherent patterns within the selected features and its computational efficiency make it a suitable choice for our authentication system.

## 3.3 Design and Development tools

### 3.3.1 User Interface

The user interface is designed to mirror a standard online payment form. The Kivy Python library is used to capture user events related to keystroke dynamics from the form’s entry fields. The data is stored in JavaScript Object Notation (JSON) format, which provides convenience in working with Python due to its native JSON support. The data is captured captured from consenting individuals (male or female, aged ≥ 18), whilst they entered fictitious card details onto the software user interface.

### 3.3.2 Data Processing

A wide variety of KMT dynamics is collected per user session from keyboard (e.g., ‘pressed’ and ‘released’) and mouse (e.g., ‘movement’, ‘left press’, ‘left release’, ‘right press’, ‘right release’, ‘scroll up’, and ‘scroll down’). It is also possible to capture keystroke dynamics from other user interface types, such as touch screens, but this is beyond the scope of this paper. The captured keystroke dynamics are transformed into a reduced set of features by calculating various functions, such as minimum, maximum and mean, per test session. From these, the optimal subset of features are determined, through a technique commonly known as feature selection[15]. The selected features were used subsequently to train a machine learning model for prediction. The selected features are the mean time of keyboard dwell and the mean flight time, as well as the mean mouse trajectory per keystroke dynamics test session (fig 2). For example, the mean mouse trajectory in (fig 2) is the total distance (d1+d2+d3) over the number of pauses beyond 500 ms from the beginning to and including the end (i.e., 3 times, in this case). The selected features are stored in a Pandas frame ready for training and/or testing machine learning models.

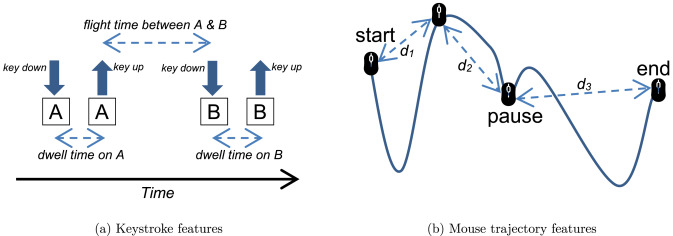


Figure 2.1

.

### 3.3.3 Machine Learning

For the classification task, the scikit-learn implementation of a simple Gaussian Naïve Bayes classifier is used in its default settings. The three selected features, presented in figure 2, collected over ten user data entry sessions, are considered to demonstrate how the software builds a classifier that can be used for user authentication.

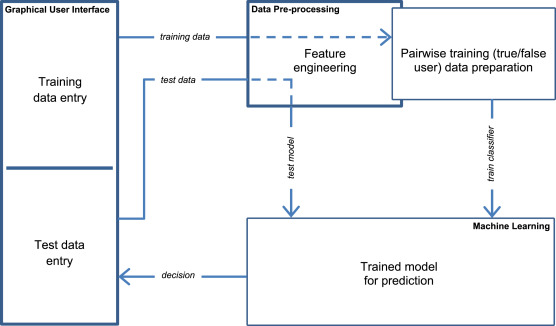


Figure 3.

## 3.4 Deliverables

1. Development of a User Authentication Tool:

- Design and implementation of a user authentication tool based on behavioral biometrics, keystroke dynamics specifically.

- Integration of a machine learning algorithm, Gaussian Naive Bayes, for creating an identity for the true user during the authentication process.

- Utilization of selected features, namely the mean time of keyboard dwell, mean flight time, and mean mouse trajectory, extracted from the data preprocessing unit.

2. Data Preprocessing Unit:

- To obtain the raw data from keystroke dynamics test sessions.The captured keystroke dynamics are transformed into a reduced set of features by calculating various functions, such as minimum, maximum and mean, per test session. From these, we determined the optimal subset of features, through a technique commonly known as feature selection.

3. Documentation and Validation:

- Preparation of comprehensive documentation detailing the design, implementation, and functionality of the developed tool.

- Validation of the tool's effectiveness through rigorous testing and analysis.

# References

1. Pilar, Denise & Jaeger, Antonio & Gomes, Carlos & Stein, Lilian. (2012). Passwords Usage and Human Memory Limitations: A Survey across Age and Educational Background. PloS one. 7. e51067. 10.1371/journal.pone.0051067.
2. Yampolskiy, Roman & Govindaraju, Venu. (2008). Behavioural biometrics: A survey and classification. International Journal of Biometrics. 1. 10.1504/IJBM.2008.018665.
3. Q. N. Tran, B. P. Turnbull and J. Hu, "Biometrics and Privacy-Preservation: How Do They Evolve?," in IEEE Open Journal of the Computer Society, vol. 2, pp. 179-191, 2021, doi: 10.1109/OJCS.2021.3068385
4. Syed, Najamuddin Aftab. (2019). Weaknesses of Biometric Authentication.
5. Alsaadi, Israa. (2021). Study On Most Popular Behavioral Biometrics, Advantages, Disadvantages And Recent Applications : A Review. International Journal of Scientific & Technology Research. 10. 15-21.
6. BrajeshSingh, S. Sonawane, Y. Shah, V. Singh, “Literature Survey on Keystroke Dynamics for User Authentication,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, Issue 5, pp. 280-282, 2017
7. Tappert, C., Cha, S., Villani, M., & Zack, R., “A keystroke biometric system for long-text input,” International Journal of Information Security and Privacy, 4(1), 32–60, 2010.
8. Hoobi, Mays & Hoobi, Mais. (2015). Keystroke Dynamics Authentication based on Naïve Bayes Classifier. 1176-1184.
9. H. M. C. K. B. Herath, K. G. C. Dulanga, N. V. D. Tharindu and G. U. Ganegoda, "Continuous User Authentication using Keystroke Dynamics for Touch Devices," 2022 2nd International Conference on Image Processing and Robotics (ICIPRob), Colombo, Sri Lanka, 2022, pp. 1-6, doi: 10.1109/ICIPRob54042.2022.9798728.
10. A. Huang, S. Gao, J. Chen, L. Xu and A Nathan, "High Security User Authentication Enabled by Piezoelectric Keystroke Dynamics and Machine Learning", IEEE Sensors Journal, vol. 20, no. 21, pp. 13037-13046, Nov. 2020,
11. D. I. Kim, S. Lee and J. S. Shin, "A New Feature Scoring Method in Keystroke Dynamics-Based User Authentications", IEEE Access, vol. 8, pp. 27901-27914, 2020,
12. Lee Hyungu, Hwang Jung Yeon, Kim Dong In, Lee Shincheol, Lee Sung-Hoon and Shin Ji Sun, "Understanding keystroke dynamics for smartphone users authentication and keystroke dynamics on smartphones built-in motion sensors", Sccur. Commun. Netw., 2018,
13. N. M. Al-Obaidi and M. M. Al-Jarrah, "Statistical median-based classifier model for keystroke dynamics on mobile devices", 2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC), pp. 186-191, 2016.
14. Sadikan, Siti Fairuz Nurr & Ramli, Ts. Azizul Azhar & Md Fudzee, Mohd Farhan. (2019). A survey paper on keystroke dynamics authentication for current applications. AIP Conference Proceedings. 2173. 020010. 10.1063/1.5133925.
15. Nnamoko, Nonso & Arshad, Farath & England, David & Vora, Jiten & Norman, James. (2014). Evaluation of Filter and Wrapper Methods for Feature Selection in Supervised Machine Learning.
16. Krishnamoorthy, Sowndarya & Rueda, Luis & Saad, Sherif & Elmiligi, Haytham. (2018). Identification of User Behavioral Biometrics for Authentication Using Keystroke Dynamics and Machine Learning. 50-57. 10.1145/3230820.3230829.
17. Akshat Shah, Parth Shah, Hitarth Shah, Chetashri Bhadane, & Dwarkadas J Sanghvi College of Engineering. (2015). Strengthening user Authentication using Keystroke Dynamics . International Journal of Engineering Research And V4(11), IJERTV4IS110437.