

Cognitive Stress Detection Using Keystroke Dynamics

¹Akanksha K P, ²Anupama A M, ³Deekshitha C, ⁴Maanya K V, ⁵Uma H R

¹Student, ²Student, ³Student, ⁴Student, ⁵Assistant Professor

¹Computer Science Department,

¹B G S Institute of Technology B G Nagar, Mandya(di), Karnataka, India.

Abstract: Stress comes from certain work parameters, daily life, habits, and other biological, internal and external factors that we exist. Stress is a ubiquitous phenomenon that is integrated into our daily lives consciously or unconsciously. This research focuses on stress analysis, cognitive stress detection, and adopts new concepts of keystroke dynamics and pattern changes. Application and stress factors are also triggered. Approximately 13 clearly important and unique characteristics are collected for each user. Modeling is also performed using classification algorithms to predict the final result, which is the stress level of a person at a given point in time.

Index Terms – Keystroke Dynamics, Cognitive Stress, Characteristics.

I. INTRODUCTION

Stress is a common, familiar and pervasive phenomena. While typically viewed as nuisance at worst and good motivator at best, Stress can be used as an indicator for a wide variety of phenomena. However, keystroke dynamics has advantages over these methods because it is a non-invasive method, does not require additional sensors to collect and analyze data, and is inexpensive. Increased automation has increased number of industries. In order to meet requirements of the client or profit point of view all the industries will exercise strict deadlines, commitment towards cost and quality of the product etc., which increases the responsibility of each employee towards his or her work. This overload of work causes employees to undergo stress. Students get stressed during their exams, doctors during surgeries etc. like this everyone will experience stress in either way. Although it has become common but it's an important factor that has an effective impact on one's health. Stress detection and analysis helps to identify an individual's level of stress and to suggest a healthy lifestyle.

Stress is nowadays present in most occupations. IT professionals are one of the groups that is exposed to stress the most. Different ways of identifying stress are measuring heart rate variability, finger temperature, pupil diameter, skin conductance etc. These methods use body worn sensors or controlling equipment attached to the body. But Keystroke dynamics has an advantage over these methods, because it's a non-invasive methodology which doesn't require additional sensors for data collection, analysis and also cost effective. Keystroke dynamics is a non-invasive methodology. Keystrokes are nothing but depression done on keyboard. Keystroke dynamics are also called as keystroke biometrics, typing dynamics. Lately typing biometrics is nothing but detailed timing information which describes when the key was exactly pressed and released.

Keystroke dynamic approach would allow for the uninfluenced determination of stress using technology that is in wide spread use today. Keystrokes are collected and used as dataset for further functioning of the website. Key dynamics are also called key biometrics and typing dynamics. Recently, biometric data sets are nothing more than detailed time information, which can accurately describe the time when a button is pressed and released. The dynamic keystroke method can use today's technology for stress-free recognition. The keystrokes are recorded and used as data for later operations of the website. The idea behind keystroke dynamics is that people have different writing styles, and by analyzing the keystroke time, this person can be identified. This indicator is that the keyboard can be used to measure the number of keystrokes, offsetting the cost of a typical physiological biometric system that requires expensive equipment to measure physical characteristics. One of the main contributions of this document is to use the website to collect data collectively and determine the stress level of specific people using the website to check their stress levels.

The remainder of the paper is structured as follows: Section 2 analyzes the state of the art while Section 3 explains the new idea's methodology. Section 4 reports the performance of the proposed system and finally, Section 5 concludes and discusses future directions for further improvements to the research.

II. STRESS DETECTION RECENT SURVEYS

Machine learning uses algorithms to transform data sets into models. Which algorithm works best depends on the problem. This enables the system to automatically learn from its own experience and make improvements without the need for explicit programming. It aims to enable computers to learn, self-learn, and generalize behavior from databases, and is becoming increasingly important in the healthcare industry.

These recent surveys consists of 8 papers published in various National and international conferences. The author of [1] has mentioned in this paper that, unlike using machine learning algorithms, statistical analysis was used. Here, the keystroke data of a user was collected twice, in order to understand the significance of this collection, the statistical apparatus was used (tests). However, the classification is based on the statistical approach, which is completely dependent on the values being passed. And it is more like hypothesis testing. The author of [2] in the paper explains the advantage of stress identification using keystroke dynamics over body worn sensors with the limitation, data collection is time consuming because there exists difficulty in identifying stressed samples. The author of [3] has mentioned of how data collection is done using a model platform. The author of [4] has told how data classification is done through SVM, neural network and decision tree algorithm. The disadvantage of this method is that data collection is done using a standard keyboard, which is very expensive. The author of [5] has mentioned about the work that involves two different types of activities that the human body perceives, namely cognitive activities and physical activities, and mainly focuses on invasive and non-invasive methods based on these two broad areas of surveillance activities. However, this work is limited to research. Only in this particular area, so there is no comprehensive and holistic attention. The author of [6] has published which are the best limits for which the classification can be applied to obtain better results and accuracy. However, the paper talked about

effects, not correct classification. The author of [7] has used the keystroke data that was collected from the users by showing them a short-animated clip and later asked them to describe the clip in their own words. Two different agents were considered whose accuracy was considered. The disadvantage of this paper is that the graphical structure used in this analysis does not significantly and exactly draw out the conclusions made. The author of [8] this paper talks about the modern practices of ML algorithms are used hand-in-hand with the traditional practices involving generic hardware appliances like the Keyboard and Mouse tools. The outcome of this research led to proving the fact that these external hardware devices are generally influenced by typing and holding patterns, however the hypothesis was not proved.

III. METHODOLOGY

Random forest classification algorithm is one approved Machine learning algorithm related to supervised learning methods. It can be used for classification and regression problems in machine learning. It is based on the concept of ensemble learning, combining multiple classifiers to solve a complex problem. Problem and improve the performance of the model. As the name implies, "Random forest is a classifier that uses multiple decision trees in different subsets of a given data set and averages them to improve the prediction accuracy of the data set." Random forest does not depend on decision making Trees, instead of making predictions from each tree, and predicting the final result based on the majority of the predicted votes. The more trees in the forest, the higher the accuracy, and there is no conversion problem.

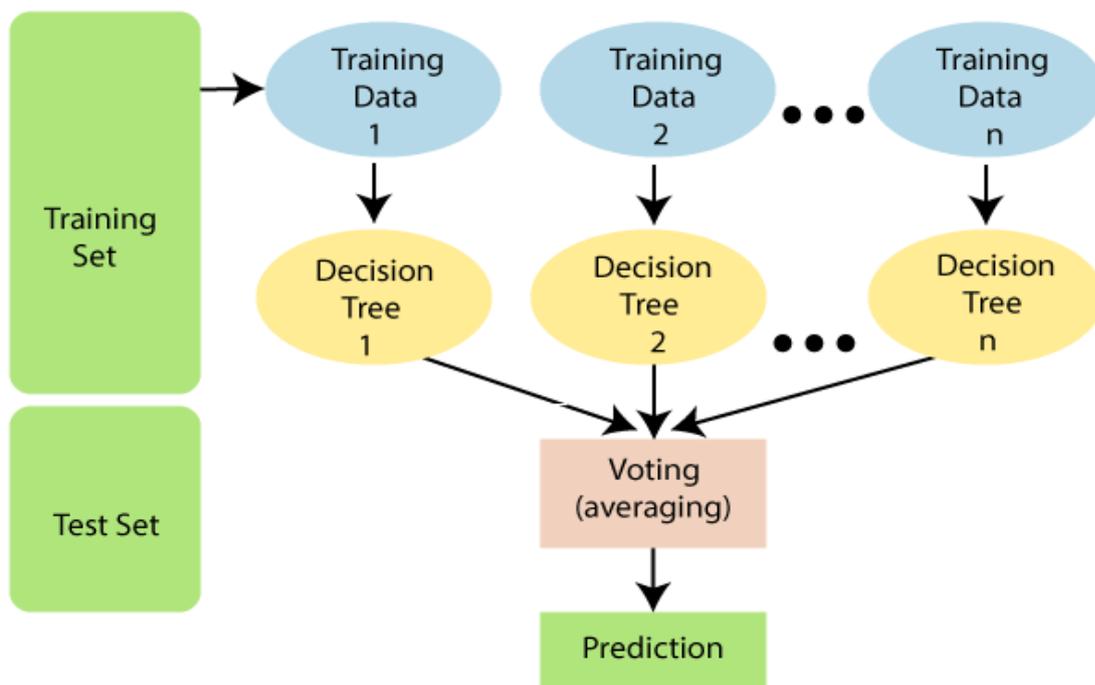


Fig 1: Random Forest Classification

IV. IMPLEMENTATION

In modern life, stress has become a common problem in day--to-day life. It has highly hazardous and consequential side effects on an individual's health therefore, it has to be strictly monitored. This research is aimed at developing a website which is capable of identifying whether the user is stressed or not, along the users stress level (Low, Medium, High). The website has been carefully and strategically developed with the help of random forest classification algorithm. It comprises of several invaluable parameters that are used in this analysis such as Key Press Pressure, Key Press Area, Key Hold Time and so on. This information is collected about each and every key being typed by the user thereby, not losing data of utmost importance. The collected attributes values will be passed as input to the classification model integrated in the website, classify it and display the result along with suggestions for inference. This website offers one key features namely: Typing Session.

The primary page is the one in which a user enters his email-id and logs into the portal for data entry. Here, further information about the user is recorded and kept-ready. The privacy policy is displayed so as to keep the user's privacy in mind. A static message is entered by the user which can be modified accordingly. There is a login for the admin page where-in the admin can send the data/export it. In the admin page, the user-id and the number of passwords can be mentioned as the sample set. The modules that were adopted in the implementation of the project are explained as follows:

Typing Session: The primary page is the one in which a user enters his email-id and logs into the portal for data entry. Here, further information about the user is recorded and kept-ready. The privacy policy is displayed so as to keep the user's privacy in mind. A static message is entered by the user which can be modified accordingly. There is a login for the admin page where-in the admin can send the data/ export it. In the admin page, the user-id and the number of passwords can be mentioned as the sample set for the trials.

4.1 Implementation Steps

- Creation of website.
- Creation of cognitive stress dataset using a above website.
- Preprocessing of created dataset and splitting the dataset into training and testing.
- Model building using the above preprocess dataset.
- Integration of model into the website.
- Identification of user stress level.

4.2 Dataset

- With the help of this project using website entire comprehensive dataset were collected, stored, preprocess and changed.
- It consists of all the collected keystrokes.

userid	gender	handedness	birth	usage	collected	date
100	F	R	1998	3	1	
101	F	R	1998	2	1	
102	F	R	1999	3	1	
103	M	R	1998	3	1	
104	M	R	1999	3	1	
105	M	R	1998	3	1	
106	M	R	1998	3	1	
107	F	R	1998	2	1	
108	F	R	2000	3	1	
109	M	R	1998	2	1	
110	M	R	1999	3	1	
111	F	R	2000	3	1	
112	F	R	1997	3	1	
113	F	R	1998	3	1	
114	F	R	1998	3	1	
115	F	R	2000	3	1	
116	M	R	1998	3	1	
117	F	R	1999	3	1	
118	F	R	2000	3	1	
119	F	R	2000	3	1	
120	F	R	1998	3	1	
121	F	R	1998	3	1	
122	M	R	2000	2	1	
123	M	R	1997	3	1	
124	M	R	1998	3	1	
125	E	R	2003	1	1	

Fig 4.2.1: Dataset

Email	UserID	Gender	Handedness	Birth	Usage	Collected	Date
sahitya@gmail.com	100	F	R	1998	3	1	
payasvini@gmail.com	101	F	R	1998	2	1	
nidhikalmene@gmail.com	102	F	R	1999	3	1	
ramesh@gmail.com	103	M	R	1998	3	1	
raghu1098@gmail.com	104	M	R	1999	3	1	
kiranraj2208@gmail.com	105	M	R	1998	3	1	
kiranrajhk2208@gmail.com	106	M	R	1998	3	1	
snehake@gmail.com	107	F	R	1998	2	1	
ep.703322@gmail.com	108	F	R	2000	3	1	
vikasbgangadhar95@gmail.com	109	M	R	1998	2	1	
hemath.palil@gmail.com	110	M	R	1999	3	1	
iqraameen200@gmail.com	111	F	R	2000	3	1	
mailsoumu15@gmail.com	112	F	R	1997	3	1	
kmanasvibhat@gmail.com	113	F	R	1998	3	1	
mmbht@gmail.com	114	F	R	1998	3	1	
iyoti.gonsvs.20.10@gmail.com1	115	F	R	2000	3	1	
guptaranjan314@gmail.com	116	M	R	1998	3	1	
shrisowmya@msrit.edu	117	F	R	1999	3	1	
imeenakshi1832@gmail.com	118	F	R	2000	3	1	
lbellubbi@gmail.com	119	F	R	2000	3	1	
anikshashetty98@gmail.com	120	F	R	1998	3	1	
anikshashetty47878@gmail.com	121	F	R	1998	3	1	
pdpndey26@gmail.com	122	M	R	2000	2	1	
vkdosaya@gmail.com	123	M	R	1997	3	1	
syamsagar72@gmail.com	124	M	R	1998	3	1	
kulsikulsom@gmail.com	125	E	R	2003	1	1	

Fig 4.2.2: Dataset

V. RESULTS AND DISCUSSION

This method directly determines the level of stress a person has at any given time. This research shows that new techniques using keystroke dynamics and pattern changes can be used for correct analysis. Save money and time for setup. The website is easy to use, accessible to everyone, and helps you make quick decisions with real-time user input. The server is implemented using powerful machine learning algorithms. People of all ages and backgrounds can use this method to promote the perception of upward trends in growth, which in turn helps determine the level of public performance and how it positively affects their daily lives. This research demonstrates that suitable analysis is possible using the novel technique of keystroke dynamics and pattern variations. This is highly advantageous as external sensors are unused, thereby saving on cost and setup time. This website is easy to use, accessible to all and helps in performing a quick judgement using real time values entered by the user. The backend is implemented using Machine Learning algorithms with a high accuracy. Overall, this methodology can be used by individuals across all age groups and multiple environments to facilitate a sense of an upward trajectory of growth. This in turn, helps with the efficiency levels of the general public and how it positively impacts their day-to-day life.

VI. CONCLUSION AND FUTURE SCOPE

The project gradually focuses on using the aforementioned tools, techniques, and operations to define the cognitive stress that a person experiences in daily life, which is very helpful because a person’s performance and the importance of mental health can be directly related. Prevent harmful and unfavorable accidents. Because the stress level is identified, calculated and quantified, this can be applied to any situation in our lives. In short, it can be iteratively extended to almost any field, including people, platforms and operations with facial recognition capabilities, systems Or biometric authentication so that specific users can log in and view their stress at any time of the day. This classification method is accurate and meaningful because it directly indicates a certain degree of cognitive pressure. The software industry or educational institutions can also use it to track their employees or students. Performance level or determining whether stress is the cause of poor performance, etc.

REFERENCES

- [1] Author Lisa, M, V., Author Lina, Z., & Author Andrew, S composed “Automated stress detection using keystroke and linguistic features: An exploratory study” (2009).
- [2] Authors Rodrigues, M., et al. wrote the paper “Keystrokes and clicks: Measuring stress on e-learning students”(2013).
- [3] Author Shukla, P. & Solanki, R., published the paper “Web based keystroke dynamics application for identifying emotional state”(2013).
- [4] Authors Trojahn, M., et al. wrote the paper “Emotion Recognition through Keystroke Dynamics on Touchscreen Keyboards”(2013)[5] Authors Gunawardhana, S.D.W., et al. published the paper “ Non Invasive Human Stress Detection Using Keystroke Dynamics and KeyStroke Pattern Variations”(2014).
- [6] Author, Lim, Y.M., Author Ayesh, A. & Author Stacey, M., wrote the paper “The Effects of Typing Demand on Emotional Stress, Mouse and Keystroke Behaviours” (2014).