BigPsy: A Big Data framework to support Psycho-informatics

by

Rashik Al Muizz
17101158
Md. Sakib Uddin
16201035
Mirza Md. Nazmus Sakib
17101144
S.M.Fahmidul Islam
17101161
Nourin Ahmed
17101022

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2021

Declaration

It is hereby declared that

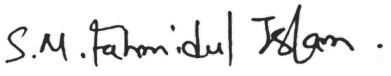
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

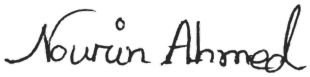
Rashik Al Muizz
17101158

Md. Sakib Uddin
16201035

Mirza Md. Nazmus Sakib
17101144



S.M.Fahmidul Islam
17101161



Nourin Ahmed
17101022

Approval

The thesis/project titled "Developing a Psycho-informatics platform using Machine Learning and Big Data" submitted by

1. Rashik Al Muizz (17101158)
2. Md. Sakib Uddin (16201035)
3. Mirza Md. Nazmus Sakib (17101144)
4. S.M.Fahmidul Islam (17101161)
5. Nourin Ahmed (17101022)

Of Spring , 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May, 2021.

Examining Committee:

Co-supervisor: (Member)

Afrina Khatun

Afrina Khatun
Lecturer
Department of Information and Communication Technology
Bangladesh University of Professionals

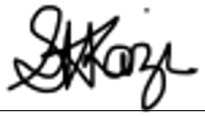
Supervisor: (Member)

Mostafijur Rahman Akhond
Lecturer
Department of Computer Science and Engineering
Jashore University of Science and Technology

Program Coordinator: (Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department: (Chair)



Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Human conduct and state of mind are fundamental pieces of both psychiatry and psychological research. A person's mental state can be determined by analyzing his behavioral patterns which contain subtle signals. This paper outlines the technical vision, sketches the signals that can be detected and illustrates the tremendous benefits over traditional methods of psychometrics. In particular, it suggests tracking user behavior with smartphones, a particularly rich and intimate source of data. The proposed method will help researchers and psychologists to study human minds more efficiently. This study also aims to dissect human conduct in detail with the help of large information by estimating the seriousness of an individual's downturn and web fixation. Because of this kind of excessive behavior, a person experiences depression, anxiety, insomnia and some other deteriorated mental health conditions. It is hard to see how genuine it is as the methods for evaluating mental health are not entirely solid. Despite, everything analysts rely upon psychometric tests, studies and perceptions, which face difficulties to address the issues of each distinctive individual. In the paper, the ordinary exercises can be recorded with the help of cell phones or computers. It will gather information on a finer scale and search for transient action designs. Additionally, the information will be collected in an electronic structure and be stored in Big Data Storage. This information-driven system will turn out to be less tedious and less expensive than conventional strategies for both specialists and patients. The most significant challenge of this research is dealing with streaming data. It will be stored in the big data storage and then, using machine learning with the useful data, a sample feature such as sleep analysis can be created. Besides, the stigma related to mental health is also a matter of concern to know more about their mental health condition. In addition, it will create another sub-zone of psychometrics which can make a new research area in the field of psychology by exclusively examining data. Finally, the proposed paper will show a portion of the moral issues natural to Large Information advancements. Therefore, the proposed approach might incite the methodological move since the arrival of psychiatry or psychological research.

Keywords: Psychology; Mental Health; Depression; Big Data; Big Data Storage; Psycho-informatics; Behavior; Psychological; Sleep Analysis.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our Supervisor Mostafijur Rahman Akhond and Co-Supervisor Afrina Khatun for their kind support and advice in our work. They helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

Chapter 1

Introduction

Psychology is the investigation of the psyche and conduct. It incorporates the natural impacts, prevailing burdens, and ecological variables that influence how individuals think, act, and feel. When psychology is combined with modern technology, it creates Psycho-informatics. Psycho informatics is a field where human behavior and mental health can be analyzed simultaneously. Formally, it can be said that the usage of electronic appliances can provide the necessary information for a person's psychiatric evaluation. Data can be parsed using a combination of big data, data mining and machine learning. Big data stores analyze and extract information out of bulk data sets. On the other hand, Machine learning can learn and improve from experience without being explicitly programmed automatically. Data mining is a cycle utilized by organizations to transform raw information into useful data. It includes investigating and breaking down enormous squares of data to gather significant examples and patterns. By using those three fields together, the mental state of human beings can be analyzed.

In 1995, Lovibond and Lovibond's developed the depression, anxiety, stress scale DASS which is increasingly used in diverse settings [1]. The authors of Lovibond and Lovibond's intended to develop a measure consisting of two scales (i.e. anxiety and depression), a third factor emerged from their analyses of scale structure consisting of items relating to difficulty relaxing, irritability, and agitation. Accordingly, the third scale was the Stress and then combing them it become DASS. DASS provides strong support for the internal consistency of the three scales. Therefore, their construction did not cover all the necessities. Specifically, items about changes in appetite, sleep disturbance, guilt, tiredness, concentration loss, indecision, agitation, loss of libido, the diurnal variation in mood, restlessness, irritability and crying were excluded from the measure. After that, in 2013, Alexander Markowetz his fellow members proposed a method by which the user's behavioral pattern can be recognized by analyzing his way of using electronic appliances [2]. These appliances include smartphones, laptops etc and they can create a resourceful data source if put together. The collection of this data can go on in the background and the collected data can be stored in a secured server. The traditional methods include questionnaires and reports created by the patients themselves which makes the reports slightly biased. These papers did not contain a finding of mental problems. Rather, it recommends helping the clinical specialist with extra data, to screen the

course of sickness and therapy. It might even make the job of the specialists even harder. Finally, in the most recent paper in 2016, Muhammad Bilal Amin showed that gadgets are autonomous and amassed a client's tangible information from multimodal information sources continuously [3]. In this way, user's daily activities will be the real source of the data.

In today's generation, young people are constantly facing mental sickness, which they cannot express to anyone. Many meritorious students fall in clutches of depression, and as a result, their academics are getting hampered. Moreover, they cannot concentrate on their studies properly. Also, many job holders can't concentrate on their works and later on, some of them suffers from anxiety problem and they committed suicide. Some people are facing mental disorder which they even do not know. As younger people are future leaders and job holders are current leaders, and their mental situation needs to protect. Now, the goal is to develop a system that will analyze their daily life activities and show an outcome of whether they are affected by a mental health problem or not. So, the system is going to be designed in such a way by which the data can be collected daily from the patients and the data collected from at least one month, and thus the mental state of the affected ones can be determined.

If the user's communications can be put on record, which means if it can be recorded whom the user speaks with over the phone or texts, it can provide a good overview of his regular contacts which will pave the way to a viable solution. In this way, the amount of social interaction the user has with the help of technology can be determined. By doing this, it can be recorded which of his friends, relatives or accomplices are in touch with him and whether he starts the conversations or they do. But it does not actually matter who is on the other side or what they are talking about. Additionally, as far as the user's movement records are concerned, the user's geographic position will be recorded regularly. If the depression is in a alarming level, it is necessary to check whether he goes out of his house or not.

The data stored in Big Data Storage requires to go through the analysis process. Two popular methods of data analysis are data mining and machine learning. In the process of data mining, the entire data is marked with proper labels and after all the required inspection, its prototype is recognized. After that, the other datasets are categorized with the help of this prototype. And as far as the machine learning algorithm is concerned, first the datasets are thoroughly observed and analyzed, then it gives a report about which way they are indicating.

Chapter 2

Literature Review

In the proposed system, some papers were thoroughly studied where information on internet addiction, Depression Anxiety Stress Scales (DASS), the impacts on mobile phone addiction were included. Moreover, whether a person is satisfied with their life is also determined by the result of studying these papers. Some of the papers are presented below for better clarification:

The Depression Anxiety Stress Scales (DASS): [**1**] In a large non-clinical sample, normative data and latent structure: The Depression Anxiety Stress Scale (DASS) was designed to have two subscales, one for anxiety and the other for depression. Because the author's primary objective was to establish measures and differentiate between depression and anxiety, this method has been used. They can't say that it will look different based on gender and age so this research was limited to gender and age. Mainly, the younger person like 0 to 17 age is not depressed, and they are not suffering in anxiety also 18 to 59 age people are suffering in this problem the most. As they, have to take responsibility for the family, works and other also. They observed this in the survey. DASS, meanwhile, does not include appetite, sleep disturbance, guilt, anxiety, concentration, indecision, agitation, lipid loss, diurnal mood change, restlessness or crying.

Psychometric properties of the depression anxiety stress scales (DASS) in a clinical sample: [**2**] Psychometric properties of the depression anxiety stress scales (DASS) in a clinical sample: This paper provides an idea about the proper use of DASS (Depression, Anxiety and Stress scale) where they respectively show the evolutions. Anxiety and depression have been typically regarded by researchers to be distinct at the conceptual level. Nevertheless, attempts to quantify these constructs using questionnaires and clinical ratings have often demonstrated a high degree of overlap (inter-correlation) between measures of anxiety and depression. Although the authors meant to make a measure using two scales (anxiety and depression), their investigations of scale structure revealed a third factor: problem relaxing, irritation, and restlessness. As a result, the third scale was the Stress that was later likely to produce the DASS. The internal consistency of the three scales is well supported by DASS. As a result, the existing data give strong evidence for the three scales' internal consistency. Even though the existing data fully support the

psychometric and conceptual aspects of the DASS, researchers discovered that most of this research has been performed with non-patients.

Psycho-Informatics: Big Data shaping modern psychometrics: [3] Although the evaluation process is quite systematic and well-organized, both the amount of data and the percentage of usable data can be lower than the required amount. This is why the work process of the conventional methods is not as foolproof as someone would expect. But if modern psychometrics is implemented with the help of big data, the sensors will become more user-friendly and do the job in a better way. The quality of the data, along with its collecting and storing process will be completely in favor. In this way, it can be predicted if the psychological condition of the patient will reach a serious phase and the patient can become cautious long before that happens. If all of these are successfully executed, many psychologists will prefer an evaluation process that is based on electronic data. The collection, storage and analysis of the data will be possible at a reasonable price without causing any problems for the patients. Big Data has the ability to create a massive impact in the field of psychology. For doing that, data mining and machine learning can be put to use. When the usable data will be retrieved, it will provide a more accurate result than the traditional methods.

Relationships among smartphone addiction, Stress, academic performance, and satisfaction with life: [4] The study's main goal was to investigate the connection between mobile phone addiction risk and satisfaction with life. This study had two objectives: one would be to contribute to the existence of information on this subject, and the other was to contribute to the body of knowledge on this subject. One goal was to look into the link between smartphone addiction risk and life satisfaction, as mediated by stress and academic achievement. The other goal was to see if happiness with life is promoted by smartphone addiction once it is controlled by stress and academic performance. The results indicate that smartphone addiction has a strong relation with stress, a negative connection with academic achievement, and a moderated negative relationship with life satisfaction. On the one hand, there was a zero-order connection between smartphone addiction and personal wellbeing and on the other hand, there was a zero-order relationship between felt stress and academic performance. Most of the other findings, specifically the relationship to academic performance, are similar to those of previous research. They noticed no relation between it and a family issue, depression, anxiety.

Recorded Behavior as a Valuable Resource for Diagnostics in Mobile Phone Addiction: [5] Evidence from Psycho informatics: Universally there is a developing discussion regarding whether abuse of cell phones and the Internet can be characterized as a type of conduct enslavement. Abuse plainly cannot be seen essentially by alluding to the time spent on cell phones or the Internet. Therefore, instead of other habit related ideas, for example, distraction, withdrawal, the advancement of resistance or individual enduring in light of the utilization are more significant factors. Given the argument over the idea of cell phone usage, "prob-

lematic mobile phone use" may be a more appropriate phrase than "cell phone addiction" or "smartphone addiction." Cell phone applications can likewise be of an incentive in the treatment of addictions, for example, liquor abuse. Also, the self-announced information and straightforwardly recorded information can be contrasted more than a few cell phones factors to anticipate cell phone compulsion. Secondly, speculation of recorded cell phone conduct contrasted with oneself announced appraisals ought to be all the more firmly connected with results from the cell phone habit survey. Although many opportunities arrive from the use of new technological trends, the misuse of these technological trends is multifarious. Therefore, the mutual connection and understanding between psychologists and patients will provide a good amount of information in case of this data pool. So, it can be said that Psycho informatics does introduce with a good medium for treating the over usage of smartphones and internet addiction.

The Impact of Psycho informatics on Internet Addiction: [**6**] Psycho-informatics can be an invaluable part of the treatment and research for internet addiction. As far as the traditional methods are concerned, they are not capable of creating a complete view of a person's behavior because they do not go that deep to the person's mental state. The traditional methods include experiments, self report questionnaires or direct interviews. Although these techniques prove to be effective to a certain extent, they still have many limitations. In this report, Psycho-informatics can solve these problems by responding to the issues regarding internet addiction. Right now, there are a lot of data to support the proposed hypotheses. If they are put to good use, these methods can prove to be very effective in dealing with internet addiction.

On Curating Multimodal Sensory Data for Health and Wellness Platforms: [**7**] Since the conventional methods have a few restraints, DCF (Data Curation Framework) has been executed which has an algorithm based on cloud storage. The collection of data is totally dependent on the records of the user's daily activities. DCF can acquire data by coping up with different situations or circumstances. Since it is a cloud-based system, there is no doubt about its ability to perform calculations. A big data storage will be required because the acquired data will be of huge volumes. The information of the users' daily activities will be stored in one database but the data which does not need to be updated as much will be stored in a separate database which will not be altered as frequently as the former. The attributes and qualifications of DCF has been determined keeping all of these necessities in consideration.

Toward Psycho Informatics Computer Science Meets Psychology: [**8**] This paper provides insights into an emerging research discipline called 'Psycho informatics'. It is the combination of psychiatry and modern computer science. In this system, the user's mobile usage is the basic source for the data to work with. This data includes how the user interacts with others in social media and what he uses the internet for. Psychology basically works with a person's lifelog data. So if the

data is acquired with the help of technology, the level of accuracy increases. Their actual target was the 'Menthal' project. It was necessary to understand how much a person's life is dependant on his mobile devices. The amount of phone usage for many users was logged by tracing a huge number of phones. Their sleeping schedules were also monitored by the same tactic. Although all the collected information gives an overview, it still was solely based on the user's internet usage. This put the impact of phone usage on the user's psychological condition in question. Some future advancements were being considered, but as of now, their idea is still based on how the user uses his phone.

The Depression, Anxiety and Stress Scale (DASS-21) as a Screener for Depression in Substance Use Disorder in Patients: A Pilot Study: [**9**] Combining results from worldwide epidemiologic and clinical investigations show that Substance Use Disorder is regularly connected with burdensome issues and the other way around. In every one, 27% people with a lifetime SUD are influenced by a co-sullen lifetime significant discouragement, while 41% people with a lifetime significant burdensome scene additionally have a lifetime SUD. Assessments of co happening burdensome issues in SUD treatment-chasing populaces are considerably higher than those in everyone: 44% of patients remembered for SUD treatment offices had a co-happening significant depression. The announced commonness of Depression in SUD people looking for treatment shifts unequivocally due to fluctuation in study strategies, for example, contrasts in populace test, treatment setting, SUD under examination, analytic instruments and challenges in diagnosing. A few examinations are distributed on its unwavering quality and legitimacy around the world, all demonstrating the DASS-21 is an entrenched instrument to gauge side effects of gloom, tension and worry in both clinical and non-clinical examples of grown-ups (11-16 years). In the Netherlands, the DASS-21 is utilized in most SUD treatment offices during admission to evaluate manifestations of gloom, tension and worry in patients of 24 years and more seasoned. In conclusion, Substance use disorder patients report significantly higher levels of symptoms of depression at intake compared to shortly after clinical detoxification. No significant differences were found between DASS-21 scores for patients with and without a depressive disorder at intake. In contrast, there were significant differences after detoxification. Based on the findings in this pilot study, the use of the DASS-21 at intake as a screening instrument for depressive disorders for SUD patients is not advised. The DASS-21 seems to be a suitable instrument to screen for depressive disorder when scored after detoxification.

Using the Depression Anxiety Stress Scale 21 (DASS-21) across cultures: [**10**] The Depression Anxiety Stress Scales was developed to measure the constructs of depression and anxiety and to address the failure of earlier emotional measures in discriminating between anxiety and depression. The original DASS has 42 items measuring three dimensions of negative emotional states, namely depression, anxiety, and stress/tension. Depression refers to low levels of positive affect, e.g., dysphoria, hopelessness, lack of energy, and anhedonia, while anxiety refers to a mixture of general distress such as irritability, agitation, difficulty relaxing, and impatience.

How the differences might appear remains unknown. This indicates that the DASS-18 depression scale did not solely figure out the level of anxiety. Since it was difficult to conduct a comparison among the results, the DASS-4 and DASS-18 should be approached with proper precautions to avoid errors.

New Horizons at the Interface of the Psychological and Computing Sciences: [**11**] According to Yarkoni, Psycho-informatics is a recent system that uses computer and information science tools and methods to enhance the acquisition, organization, and synthesis of psychological data.

Diving into further study, it is noticed that further advancements can be introduced to the traditional methods of treating psychological problems. With the help of these advancements, psychologists will be able to treat their patients in a better and more efficient way. A psychologist can differentiate between a normal person and a patient. Also, the daily activities of a person can also be traced time to time in order to understand their mental conditions. With this advancement, a revolutionary change is about to emerge in the field of psycho-informatics.

Chapter 3

Problem statement

As discussed in the previous chapter, it is a challenging task to understand a person's mental condition completely because self-reports are used quite often which are made from the person's own judgement and the treatment has to be based on assumptions. In addition, the process of collecting the required information of the patient is also time-consuming and managing all those information is also quite difficult. Normally it takes a good amount of time to observe the psychological changes that occur in a patient from time to time. During this period, the psychologists have to go through multiple data of a patient several times. It often becomes difficult to keep track of everything and attend to the specific needs of every single patient.

Since collecting the patient's information and properly managing it are the issues here, use of technology can be the missing piece of the puzzle. To assist the psychologists, a system is needed which can help them to treat patients and common people by measuring their daily activities. If a person's daily activities are tracked with the help of his electronic devices, it will lay the groundwork for collecting the information required to get an idea about his mental state. The report that will come after examining the data, will make the job of the psychologists a lot easier.

Severe depression and inconvenient mental health have become common occurances of the recent era. An example scenario was collected after discussing with a psychologist, where he informed about a patient named Rashi. Rashi is a university student. She has been performing well academically since the beginning of her university life. Therefore, her father passed away suddenly due to which she is facing financial problems. She is facing obstacles to continue her studies, and also her mother is pressurizing her to get married so that her husband can take care of her expenses. Because of all these issues, she can no longer concentrate on her studies, and her mental health has been deteriorating ever since as she is not prepared to get married so soon. She also wants to continue her higher studies. Now she is at a loss of how to decide her future. For all these reasons, she has become less talkative, less interactive with her friends and relatives, loss of appetite, sleep disturbance, tiredness, concentration, indecision, agitation, the diurnal variation in mood and restlessness. Thus, from the proposed idea, the psychologist can measure the anxiety level, what amount of time she has been spending on social media. As a result,

the psychologist finds out that she has become more addicted to smartphone usage and has started interacting with strangers online. Before, she was constantly in touch with her beloved ones. Now she has distanced herself from the people she was close to.

The scenario mentioned above represents one of the many situations that are occurring frequently in society. In the proposed system, an approach has been made where the collection of data will be done by electronic devices, the storing of the data will be conducted with the help of Big Data and the analysis of the data will be done by clustering and machine learning algorithm. Hence, no part of the information will go missing because it will be collected as electronic data. In addition, the data storage and data management system of Big Data technologies are impeccable. So when it comes to keeping track of the massive amount of data, there will be no issues. The data analysis process will efficiently filter out the unimportant data and only let the useful information pass through so that their patterns can be recognized. In this way, there will be no hindrances in the path of psychologists to solve one of their initial motives, which is treating their patients in a more advanced way. Therefore, after a proper data analysis process, the psychologists will have a notion of the patient's improvements and deterioration simultaneously and they will be able to treat their patients accordingly. The system still has room for improvement and it will require additional expenses for further development and modification.

Chapter 4

Proposed System Model

Due to the problems mentioned earlier, a more advanced system is required so that each of the issues the patients face, gets attended to. If a simpler method is created and implemented, it will eradicate most the confusion and assist the psychologists to a greater extent. In this chapter, the model of a system will be introduced which will make use of the perks of modern technology as tools to create an overview of a person's psychological condition. The proposed model will be able to interact with technology and bring about further improvements in the field of psychology. The model will assist psychologists and psychiatrists to communicate with their patients in a better way. In the overall work, a mobile application and a browser extension have been developed for the patients of the psychologists who need a more thorough observation for a time being.

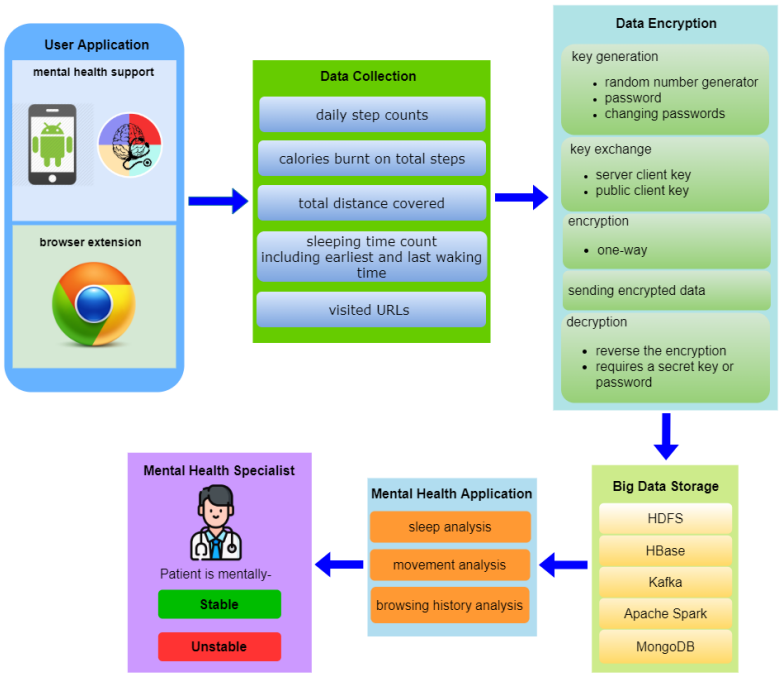


Figure 4.1: Framework of the Proposed System

A mobile application named "Mental Health Support" has been created to track a user's movements. In this application, there is a pedometer function to calculate burnt calories and cover distance with the user's step count. This will go on for each day and the movement data will be shown against the date of that specific day. This application will also measure the user's sleeping time for each day. It will include the time the user went to sleep and the time he woke up. To track the user's online activities, a browser extension has been created which will mainly work for google chrome. It will record the mail address the user is logged into google chrome with and the URLs visited against the date and time.

After the necessary data is acquired through the mobile application and the browser extension, it needs to be encrypted. Data encryption includes the following steps:

1. Key Generation: It will use a random number generator that creates temporary passwords and keeps changing them after a fixed period of time.

2. Key Exchange: Computers will exchange keys via reliable channels. In this process, only public keys will be exchanged. The private keys are not to be exposed.

3. Encryption: The readable data will be encoded and can be read again only after it is decrypted. Since this is a one-way function, the computer cannot discover the initial data.

4. Sending Encrypted Data: The data will be sent in an encoded format so that its original format does not get exposed.

5. Decryption: Decryption is the reverse process of encryption. When the receiver gets the encoded message, it will require a secret key or password to decode. In this case, it can be done only with the user's private key which will never be exchanged with other computers.

After the encryption process, the encrypted data is stored in Big Data Storage. Among the different methods of Big Data Storage, Apache HBase is a NoSQL database written in Java.[15] It runs with the help of HDFS and has several implementations such as in Mozilla Firefox or Facebook. Although HBase is not an applicable option in every situation, its functions will come in handy in specific cases. Firstly, when it comes to data volume, it will be used to store and process a huge amount (petabytes) of data. Secondly, for the application, access to the database will be granted by a key or password and in this system, HBase will work very efficiently. Thirdly, HBase works with the help of HDFS. Since HDFS can work properly with a large number of nodes, HBase will be a suitable choice when the hardware environment is in favor. Lastly, HBase has a very efficient data management system, which will allow the extraction of data at any random time.[16] In short, HBase will be very useful for data queries and a proper solution for NoSQL deployments.

Some other options to consider are Kafka, Apache Spark and MongoDB. Kafka is a software that creates frameworks for data storage and the analysis of streaming data. It is open-source which means it will be free to use and the users and developers will be able to play an important role in its advancements. The developers can bring new features or add updates to the existing ones and the users can provide feedback if they want. Another option is Apache Spark. It is widely used for processing data on large scale. It is also open-source similar to Kafka. Its interface will program entire clusters and have reasonable fault tolerance. Lastly, MongoDB is a NoSQL database system and is a very popular tool in modern applications. Its data management system will help data storing and retrieving go on smoothly and developers will be able to work more efficiently since its structure can manage a large amount of data even in thick traffic.

Mental Health Application

Good mental health is a very essential need in daily life. In this application, a patient's mental conditions can be analyzed by taking the required data. For example, calories burned, sleep analysis, distance covered, browsing history analysis. In sleep analysis, this application will run in the background it will measure the patient is sleeping or not. In addition, when this application will gather all the data, then this data will be stored in big data storage. After that, it will use the machine learning method to filter out the data and will show if the patient's sleeping pattern is normal or abnormal.

The proposed application will keep track of the user's daily movements, which includes the number of steps taken, covered distance in kilometers and number of calories burnt. This information will be recorded every day and added to the database. Now, there is an average amount of daily movement for a person that is natural. Even if a person stays at home the whole day, he normally does not sit idle at one place all time. If he moves around the house, he burns some calories. Any disturbance in this regular amount will be considered as an irregularity. If the distance covered has a huge difference from the normal amount, it will be noted as a sign of abnormality. With the data acquired from the database, any irregularity or any pattern of irregularities can be noticed. If the pattern goes on for a long period of time, it can be determined that the user's daily movement habits are not completely normal.

Browsing history analysis will help the psychologist a lot under the mental conditions of the patient. As the browser extension will be added to the user's browser, after that through the synchronized method whenever the patient is going to browse any kind of website the URL will be automatically stored to the main big data storage. All sorts of URLs will be stored in the storage. Then the data will be filtered out using machine learning methods. Filtered data will be shown as to which types of websites are visited the patient visited more and which are less visited. For example,

if he visits the 18+ websites more than the other websites it will be also figured out. Through this analysis, the psychologist could easily confirm that he is addicted to porn.

Mental Health Specialist

After seeing the sleeping analysis, movement analysis and browsing history analysis, the mental health specialists will decide whether the patient is mentally stable or not. For that, they have to see the average sleeping time and the time the patients are going to sleep and their everyday routine is normal or not. The everyday average movement of the patients, which means how much they walk or how much distance they cover every day will also be checked. With the help of the browser extension, their everyday browsing history will be analyzed. It will be checked which of the social media or any other websites they visit the most and in which ones they spend most of their time. It will also indicate which patients' habits are usual and which are not. If it is found that they are not in their normal routine and also their sleeping time is not normal, in fact they spend most of the time on social media like Facebook, Instagram, and others, then it can be determined that their psychological condition is not perfectly fine. When psychologists can come to such a conclusion, they can start the treatment for those specific patients.

Chapter 5

Data Collection

A mobile application has been developed and a browser extension has been made to order to add in the web browser to collect the desired data for the research. All the data from those two systems are stored in the firebase data storage.

The mobile application "Mental Health Support" will be a revolutionary step in the field of psycho-informatics. This application will be provided to the patients who need to be observed by some of their daily activities for a specific time limit such as step counts, the number of calories burnt during the step counts, and distance. Also, this application will track the sleeping pattern such as how much time the patient has slept at night from the starting time to the last waking time. After installing the application, when a user starts it on the mobile screen, it will require signing in with a google mail the user's phone number to sign in with a one-time password (OTP). Then always the app will be running in the background to collect data from the user. The app is quite user-friendly and one only needs to sign up to start it. After that, it will not ask for any kind of additional data from the user.

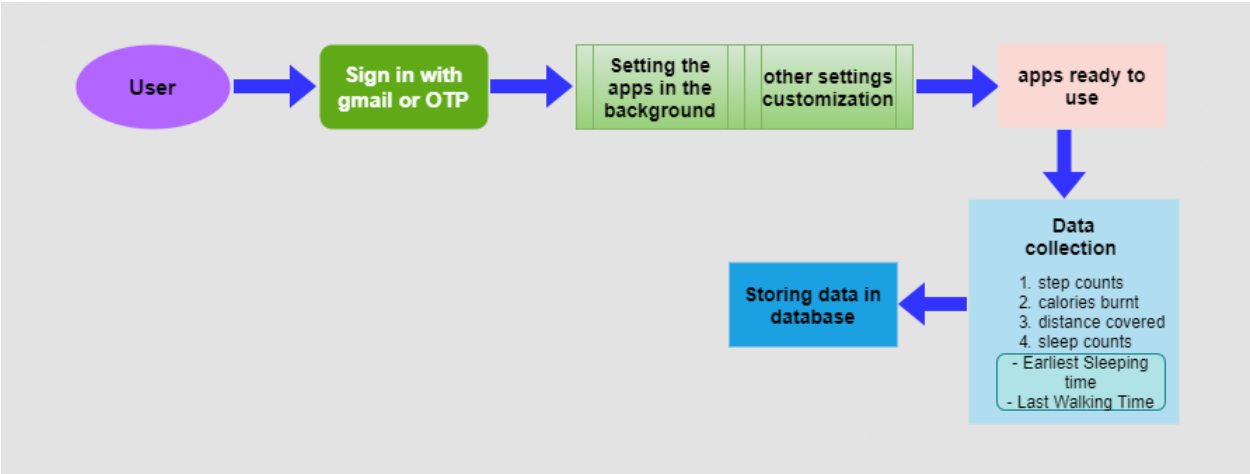


Figure 5.1: Application Framework

Figure 5.2: Logo of the Application

As the app will be running in the background after the sign up it will start collecting the data from that moment onwards. If the user starts walking the application starts counting the steps along with calculating the number of calories burnt and the total distance covered.

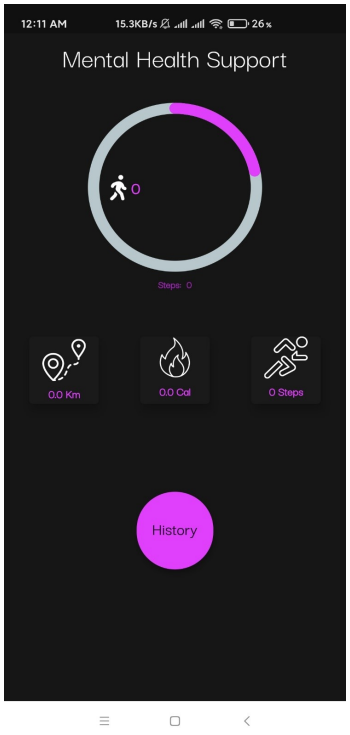


Figure 5.3: Starting Interface of the Application

When the user starts walking, the steps will be updated along with the calories and total distance covered as shown below.

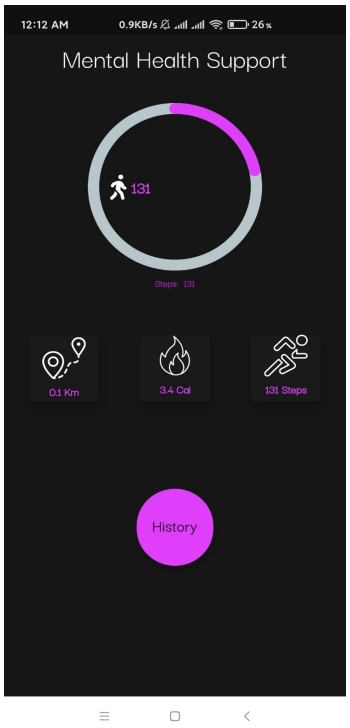


Figure 5.4: Step Count

The calory burnt has been set to approximately 0.03 units per step. The distance has been set to approximately 0.763 meters per step as normal.

After a user's completion of a certain number of steps, the application's interface is showing user has completed 131 footsteps. And as a result of those 131 footsteps, the distance covered by that user is 0.1 kilometer and the total calory burnt is 3.4 Cal. Those three features always get updated in real time data updating method and updated in history page.

The sleeping pattern of the user will not appear on the homepage. Instead, there will be a button labeled "History" which will lead to the history page and the sleeping data will be updated on this page. The sleeping time will record the start time and waking time of the user and the data will be stored in the cloud firebase but it won't be shown on the history page. If the user locks his phone sometime between 10:00 pm and 8:00 am, a timer will run for 15 minutes. If the phone is found locked for 15 minutes straight, a stopwatch will start recording time until the user unlocks his phone. The amount of time for which the phone was locked, will be counted as the user's sleeping time of that day. The starting and stopping time of the stopwatch will also be recorded which will be considered as the user's starting and waking time respectively. The user will receive the overall calculated data that is, the overall distance covered, the number of calories burnt, the number of steps accomplished and the sleeping time of the user only after 24 hours or a day has been completed. The data of each day are automatically updated on the history page after every 24 hours. This will help the psychiatrists and psychologists to determine the current mental condition of the user. The psychologists will read the data from

the application and check if there is any irregularity in the user's daily movements or sleeping schedule.

Figure 5.5: Data Storage's Records

In the same day, measuring the sleep time it has been seen that the user slept at 12:47am and then woke up at 9:59am which has been recorded. The total sleep time is recorded 33260 seconds in total.

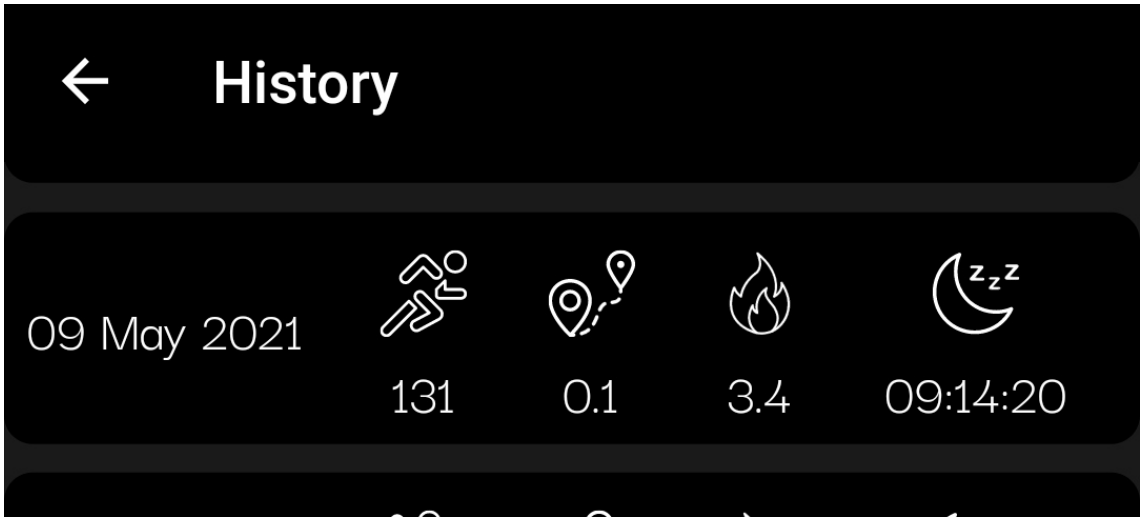


Figure 5.6: The Application's History Page

The data is also updated in the history page of the user application. The data was on 9th May, 2021 where the recorded data were foot steps 131, covered distance

0.1 kilometers, calory burnt 3.4 Cal and the total sleep time 9 hours 14 minutes 20 seconds. This is how the data has been collected.

As mentioned before, a mobile application has been developed along with a browser extension to collect the desired data for the research. Data from both the application and extension are stored in the firebase data storage.

The browser extension can leave a crucial impact in the field of psycho-informatics. This extension will be provided to the patients who need to be observed by some of their daily search in their search engine such as URL saving, date and time saving, user id saving, after adding this extension. When a user will search something in the search bar like www.facebook.com , any other website when they entered in the website in the meantime the URL will be saved in our database and also there will be saved their google mail and also the date and time here. Moreover, if they delete their history from there google history it won't affect because the data will be saved in our database.

In addition, every user id will be containing a unique id. Like 100 patients added this extension in their chrome and 100 ids will be shown there.

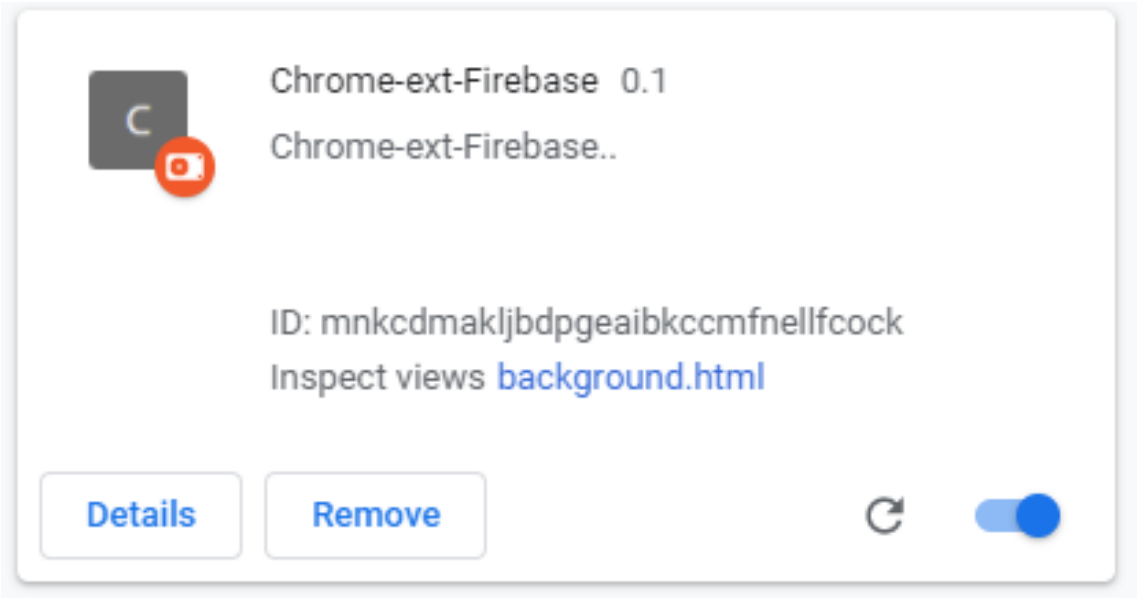


Figure 5.7: The Browser Extension

In the 2 following picture, we can see the user's Gmail id, then the time and date when he entered, and the URL he entered.

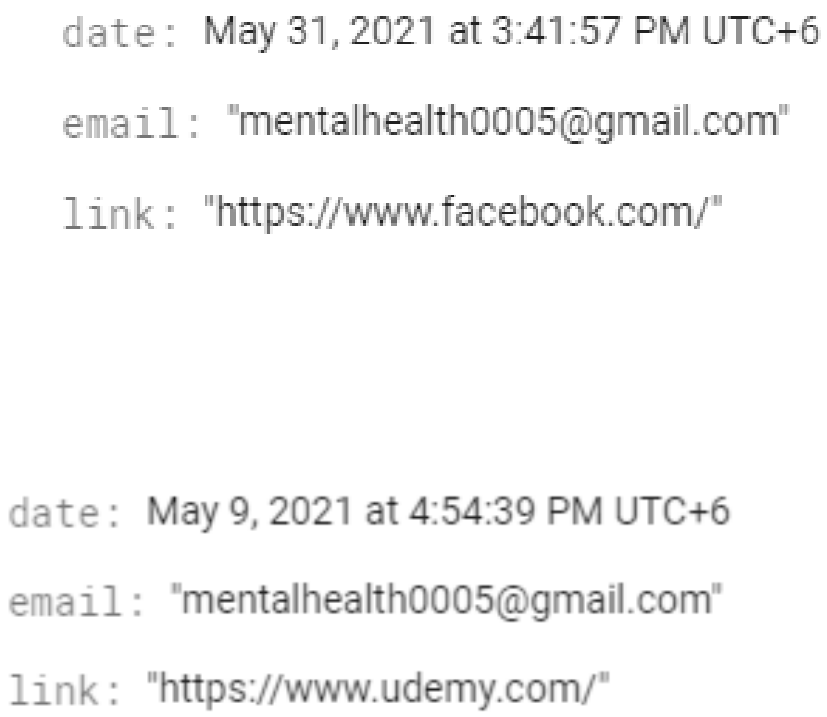


Figure 5.8: The user's Gmail id, date and time in the storage

Chapter 6

Data Analysis

In the developed system, if the collected data is accurately analyzed and if it can be determined without a shadow of a doubt whether the user's behavior pattern falls in the abnormal category or not, it will be a revolutionary step in the field of psycho-informatics. So far, the data for various activities of multiple users has been collected. The history of these data is updated after every 24 hours on the history page. In the data analysis part, the aim is to collect all this data and separate the useful ones from the unimportant ones with the help of clustering. Among the various methods, K-means clustering, Hierarchical clustering and DBSCAN clustering algorithms will be the smartest approaches to gather all the useful data of multiple users.

K-means Clustering

There are not many calculations or computations in the process of K-means clustering, which allows it to be quite swift. When the user data will be collected, not all of it will be useful or even usable. In this case, K-means clustering algorithm will be an efficient method to filter out the unimportant data and keep only the useful parts. In its work process, the value of K will be set beforehand. The clusters will be identified and the datasets will be properly labeled. If this algorithm is executed for multiple values of K, the number of clusters can be calculated by a comparison among the results. Generally, the value of K cannot be determined precisely. But by using some specific techniques, an estimation is possible.

To begin with, some sets of data are selected and center points are randomly picked. It's important to determine how many classes are in use. Here as consequence, different features are discovered in overall data. The center points are vectors of the same length as the data point vectors. The total duration between data points and their cluster center is a common statistic for comparing outcomes across different values of K. Because increasing the number of clusters increases the gap between data points, increasing K decreases this measurement to the point where it remains constant when K equals the number of data points. As a result, this measurement cannot be applied as the primary benchmark. Rather, it will display the mean distance to the centroid as a function of K, and use the "elbow point," where the

rate of decrease unexpectedly shifts, to calculate K. Then the distance between each data point and the group center is calculated and the result of this calculation is used to categorize each data point. The following step is to classify the object as belonging to a particular group with the closest center to it. The mean of all the vectors in the group is determined using this information, and the group area is generated again. This procedure is repeated many times, with the group centers remaining consistent when the process will be repeated.

This specific method of clustering uses iterative refinement to produce a final result. The algorithm inputs are the number of clusters and the data set. The data set is a collection of features for each data point.[17] The algorithms start with initial estimates for the centroids, which can either be randomly generated or randomly selected from the data set.

Ultimately, the best machine learning algorithm to use for any given project depends on the data available, how the results will be used, and the data scientist's domain expertise on the subject. Understanding how they differ is a key step to ensuring that every predictive model your data scientists build and deploy delivers valuable results.

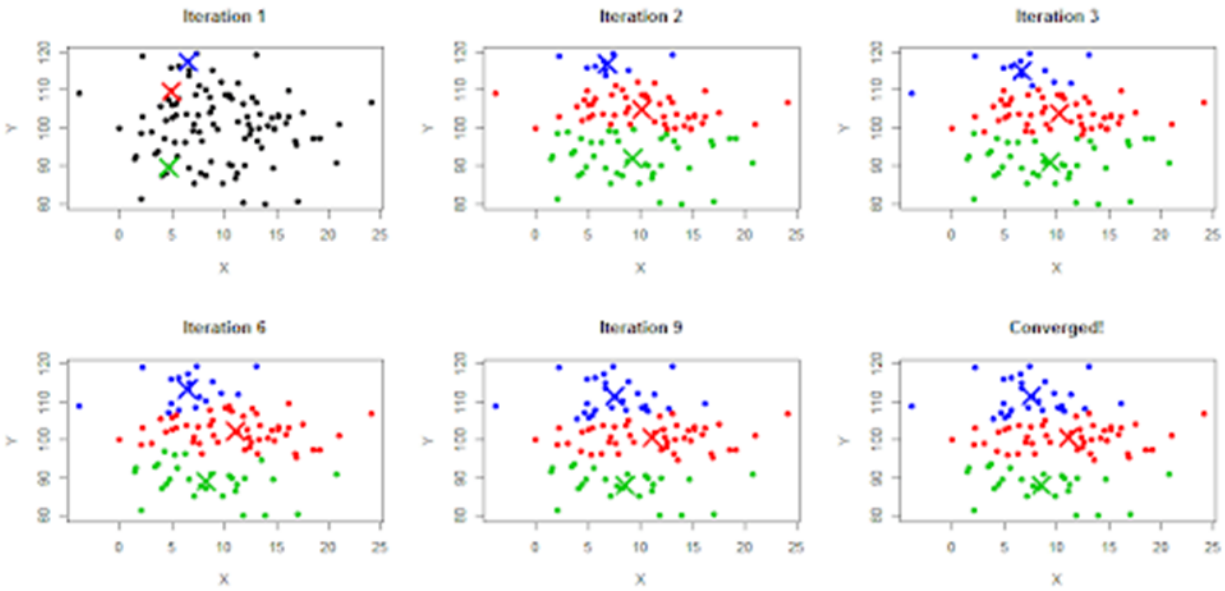


Figure 6.1: K-Means Algorithm

Hierarchical Clustering

Another method of clustering is the Hierarchical clustering algorithm. Hierarchical clustering includes the organization of the data in a type of hierarchy. This is a kind of bottom-up approach, where the data is considered as individual data points. Then these data points are lumped together into clusters little by little until eventually, the entire data set is just one big cluster.

At first, the points of data look like dots. Then they add up and acquire circular shapes. After that, the shapes keep getting bigger and finally form a large enough body which is known as a cluster. This is the basic idea of hierarchical clustering method. This algorithm calculates the distance between two data points and then the points are combined to generate a new one. One advantage is that a relational diagram can be generated using the distance between the points. The process of combining points go on until all the points are combined into one large cluster.

DBSCAN Clustering Algorithm

DBSCAN or density-based spatial clustering of applications with noise, is a clustering algorithm which works on the speculation, to break down clusters that have high density from clusters that have low density. It groups all the densely grouped data into a single cluster. In this clustering method, pattern can be recognized by developing clusters with the help of the changes in density. DBSCAN is the base algorithm which is also used to figure out data structures which will be a trustworthy means of finding patterns.

For DBSCAN clustering, each and every data point has to be explored and the algorithm approaches by randomly picking the data points one by one. When the minimum number of points will be found in a specific radius (epsilon), all of those points will be considered as parts of the same cluster. The clusters are then developed by repeating the neighborhood computation for each surrounding point in a recursive manner.

DBSCAN is an underrated but very useful clustering method. For the proposed system, DBSCAN can be used for clustering as it selects random points and then filters the data until all points have been visited. In this way, the usable data from the application and browser extension can be retrieved which will later generate the sleeping and movement patterns. Among the three methods described, the most applicable one for the proposed system will be used to find out the important datasets from the storage.

There are many different reasons why one would want to run an analysis on a doc-

ument. In this scenario, the documents need to be organized quickly and efficiently. The problem is, there might be a limited amount of time to conduct this organizing. To be able to complete this task, the theme of the text needs to be understood. The next step is to compare it with other documents and classify it. It has to be decided whether K means clustering algorithm or hierarchical clustering algorithm will be used in the application utilizing the neural network in machine learning.

Before the clustering operation, every type of data is mixed up. So, there are many junk data in the storage. It becomes very difficult to find out useful data. After clustering the dataset, groups of similar data have been created automatically.

After the collected data is filtered by the clustering process and only the useful datasets are acquired, they will go through another analysis which will be conducted by machine learning. This process dictates which of the data patterns show normal results and which ones show something unusual.

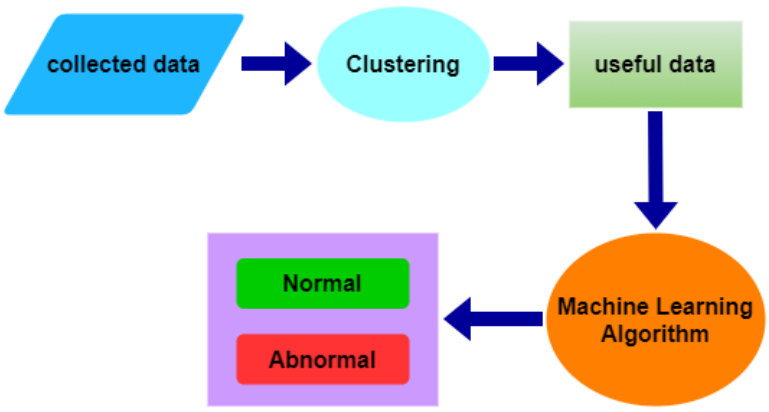


Figure 6.2: Data Analysis Process

Utilizing this useful data will help the psychologist to determine which person is mentally stable or unstable.

Chapter 7

System Privacy And Security

Modern Technology has made tremendous advancements in the last few years, which has opened many doors in every sector. To operate everything with the help of technology, it is crucial to store data. And to keep those data safe and unscathed, security systems need to be established. Data security plays an essential role in helping security systems run smoothly. Generally, when it comes to keeping personal data secured, most people tend to think that only a strong password does the job. But in reality, a lot more security operations run in the background which not everyone can see.

When any information is put on the internet, it becomes open to being stolen for personal gain. Digital information is exchanged all the time throughout the world. Every time people log into their e-mail or any social networking website, they knowingly or not knowingly exchange digital information. System privacy and security are very important to prevent this type of theft. If the stored information is sensitive, it definitely needs protection.

The process of system privacy and security for the proposed system is discussed below:

Application Security Models

The neural network is used to analyze the computer network security evaluation simulation model. At First, the system security detection algorithm is built and the calculation scale selection method is evaluated by using a system security experimental platform. The relationship between incursions and their relationships is explored via data mining. Second, the method's development steps are described. Finally, the algorithm is proven correct, and the outcomes are evaluated in terms of the three factors: entropy value, active entropy algorithm, and cloud platform performance. Furthermore, the algorithm is accurate and adaptable, and it can be used to support system safety practice training. Also, the encryption protocol in the network security protocol has been upgraded, establishing a strong foundation for the network's security development.

Each of these attributes can be briefly defined as below:

User Management: Within an identity system, users, groups, and roles are created, deleted and modified to support an application or a group of applications. Adding or removing attributes, changing or resetting passwords, locking or unlocking user accounts, adding or removing group membership, and other similar acts are characteristics of modifications.

Authentication: The process of justifying the user's identity to the system.

Authorization: Authorization is the process of determining whether or not an authenticated identity is authorized to enter a requested server resource based on a specified authorization policy.

Confidentiality: Only allowing those who need to see the transmitted data or comfort access to it.

Integrity: Ensuring that the message was not modified in transit between two actors.

Non-Repudiation: The certainty that the sender of the message will not be able to deny it sent the information.

Availability: Determining whether or not the system is operating as expected.

Auditability: The Data Security Auditor's capability to reconstruct an outcome of interest after it has occurred and certify that suitable security policies are in place for all components of the system (for application users and administrative users).

Identity Propagation: Identity propagation is a method of safely transmitting an authenticated identity from one system actor to another (think SAML 2.0 Bearer Tokens or JWT tokens).

Data Encryption

Passwords should be kept confidential and preferably encrypted. Passwords should be compared encrypted, never decrypt the data. Because application security does not demand specific database permissions, the threat of database disruptions is minimized. Passwords should be encrypted securely, perhaps using private and public keys. The architecture of the APPLICATION USER table was changed in this case by adding columns to contain public and private keys. This application security method refers to the application verifying users by collecting the end users' encrypted passwords in a database. The system allows the access permissions based on databases instead of roles and positions.

The key derivation function can be used to convert the user's password into an encryption key. Then the derived key encrypts the generated key. In this way, the

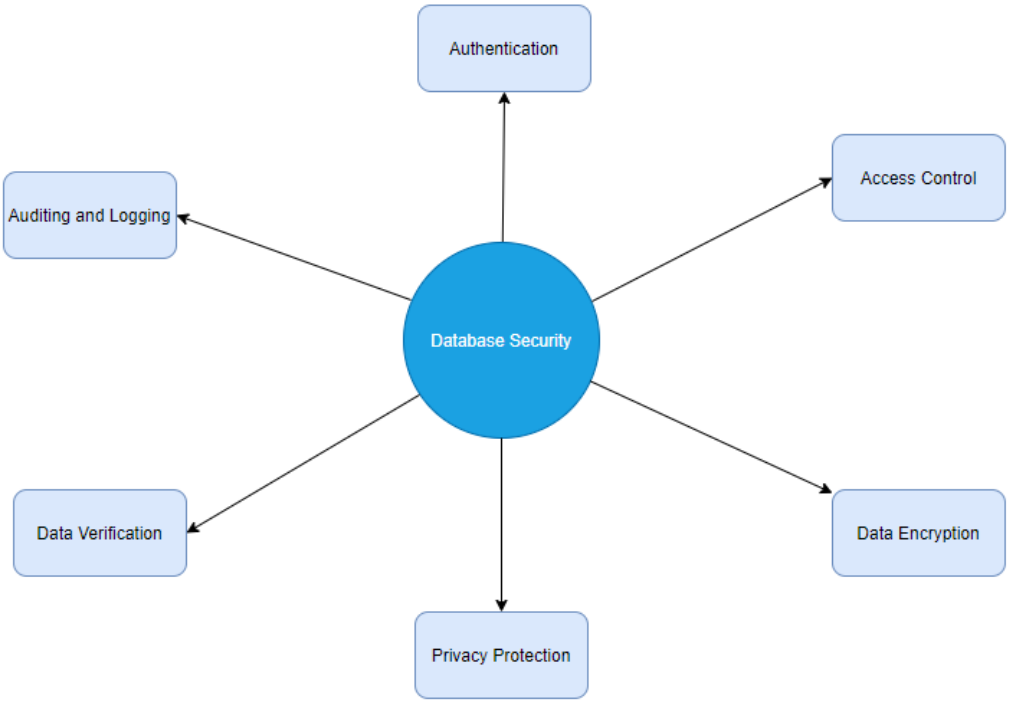


Figure 7.1: Database Security Model

user's password will become the means to decrypt the user's encrypted key. The key which actually encrypts the data is only decrypted long enough to decrypt the data that it encrypted. That key needs to be stored in the session in order to avoid the need to ask the user for password on each decryption occurrence.

Key Encryption

Random Number Generator: Helps create Key#1

This key encrypts the data. It stays constant over time. This key must be generated when the user first registers. A CSPRNG (cryptographically secure pseudo-random number generator) can be used to ensure sufficient randomness and unpredictability.

Password: Converted to Key#2 with PBKDF2

This key, Key#2, is used to encrypt Key#1. Key#2 needs to be persisted in the user's session. The encrypted form of Key#1 will be stored in the user table, in a field called (perhaps) "encrypted key".

Changing Passwords

Whenever the user changes their password, only step#2 has to be executed again, rather than encrypting all of the data all over again. The new password has to be converted into a key (Key#2), then Key#1 has to be re-encrypted and the old value for the encrypted form of Key#1 will be overwritten.

Encrypting/Decrypting Data

When the user has logged in, execute step#2. Once the password is converted into a key, Key#1 has to be decrypted. Now that Key#1 is decrypted, it can be used to encrypt and decrypt the data.

The applicable method to privacy is being used in the earlier models. These models are limited and lack the flexibility to establish access controls based on the data content or the type of activities in a WFMS. WFMSs have achieved success as a key technology for providing activity-intensive Web applications that contain a wide range of automated transactional tasks.

The ethical implications of using Big Data in research and therapy are unavoidable. As a result, data privacy becomes incredibly valuable, and the possibility of misuse cannot be ignored. In a hospital environment, monitoring depression fulfills the highest ethical standards. Since its origin, both medicine and psychology have focused on the idea of a transparent human. Both have had to deal with sensitive data for the same period. As a result, there is a strong record of privacy and security, whose methods can be used to install Big Data technology. Scaling and extending these ideas to an entirely new level is a major task that will require a lot of work from researchers, users, and professional organizations. From a different point of view, privacy concerns are simply an issue. Instead, the medical sciences must keep to another of their great traditions: a type of situation balancing risks and benefits.

Despite the rapid growth of Web-based devices connected to the internet, information security and Web security issues have been mostly ignored.

How to Prevent User Data Theft

Any unprotected data on the internet is privy to theft and can also be used for cyber crimes. If the stolen data contains anything confidential, it can expose a person to a lot of threatening situations. So it is necessary to have some precautions to prevent any of this from happening. Some of the methods of preventing data theft are discussed below:

1. Using a firewall to protect data

Keeping the internal network safe from outsiders. To secure remote access, set up a firewall to prevent access to user data.

2. Having an elaborate password policy

The most common passwords are "123456" and "password," with "123456" and "password" continuing to be the most often used passwords in 2018. It is important to have a detailed and specific password policy in a way to protect unauthorized access. All passwords should be changed daily.

3. Backing up the data regularly to avoid ransomware attacks

The latest trend in cyber crime is to steal access to the information, encode it, and then demand money in exchange for the data. This is known as a ransomware attack, and it can be easily avoided by regularly upgrading backups. These backups should be password-protected.

4. Preventing user account theft

Security issues can be avoided by applying various development methods. It should be checked and requirements should be regularly updated.

5. Not storing plain text passwords

It's not easy to keep track of user passwords. Too many people keep their passwords in plain text or apply poor hashing methods. For any hacker, plain text passwords are a treasure, because they would endanger users' trust.

6. Monitoring network and actions

It should be detectable if someone is taking data. Network traffic should be observed to set up email alerts. User behavior monitoring software is meant to capture every step taken by a user and to detect and examine any unusual conduct instantly.

7. Data storage encryption

Any gathered information should be stored in an encrypted format. Any data user should provide an option to encrypt all data automatically. It needs to be switched on. A Key Management Service is being used to protect all of the data in someone's database. This makes it impossible for any hacker who obtains this data to misuse it without acquiring the master key.

Authentication

In the proposed application, the user gets to sign in with both OTP and email addresses. In the application, direct encryption is not possible. Any user can log in to the system and measure the distance covered, calorie burnt and sleeping pattern. However, the firebase can detect the old and new users who are trying to log in to the system. The firebase will allow only the old users to log in and run the application and the firebase also will update from time to time which OTP and email address users are logging in. In case of new logins, the firebase will try to help to notify the owner or application administrator about who logs in, through email. Thus, encryption is fully assured. Furthermore, inside the application, when a user uses the application for a full day, the data of distance covered, calorie burnt and the sleeping pattern in that particular day only is updated in the firebase. The history of the users' data will not be retrieved instantly. That way the data will be secure in the firebase and won't fall into wrong hands even if someone breaks into the application. The full data is only visible after 24 hours is complete, the data of a user recorded will only show in the history page of the application once a full day or 24 hours is complete. However, further information like the person's web surfing history will not be updated in the application. It will only be visible to the owner or system administrator and thus full encryption of both the user and the application can be ensured finally.

Chapter 8

Big Data Storage

Big data storage is a specific kind of architecture that collects huge volumes of data and stores, computes and analyzes it. Big data and its analysis are highly used for scientific, technological and business purposes. The origin of this data is the user's online activities such as online videos, Emails and other social networking interactions. When such a massive amount of data is stored in databases, there is often a lot of similar data and managing all of them becomes quite troublesome.[18] When the cumulative size of data keeps increasing, analyzing it becomes tougher as well. But still, the goal is to manage this amount of data and also retrieve any part of it whenever necessary. The data management system of big data storage helps to manage all of this data more efficiently. That is why, big data has been implemented in the proposed system where the analysis of daily activities such as sleeping, movement and browsing history of multiple users will be conducted and will be figured out through a pattern.

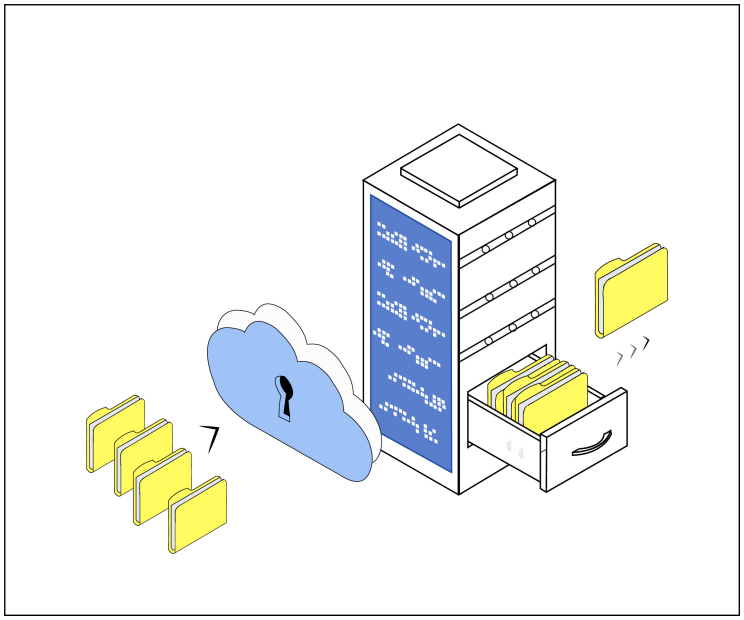


Figure 8.1: Work Process of Big Data Storage

In this application, different data sets of some individuals has been collected. The collected data contains the records of their daily movements, sleeping time and browsing history. Then the data will go through the stages of the encryption process so that anything personal does not get exposed.[19] After that, the collected data will be stored in big data storage. Several categories of data models can be implemented to store big data. The categories that have been used in this system are Apache Spark, HBase, HDFS, Kafka and MongoDB.

Apache Spark is an open-source real-time platform for large amounts of data that is aimed to provide the performance, adaptability, and computing found to be associated with Big Data technology, particularly streaming data, graph data, machine learning, and artificial intelligence (AI).

Among the methods of big data storage mentioned earlier, HBase is a very fault-tolerant one. It also plays the most significant role in data storing for this system. HBase is a non-relational database where a multiple server hardware is preferred to use instead of a single server. It is a powerful method where the contribution of HDFS is noteworthy.[20,21] HBase is also used in several world-renowned projects like Adobe, Facebook, Twitter etc. HBase can successfully conduct its operations in other file systems. There is a work in process in HBase with Java's new I/O, with the help of which the whole system can be amplified. One of the most significant attributes of HBase is its adaptability. It can manage divided or clustered data very well.[22,23] This process is operated through multiple nodes and as a result, the system will not face major disruption if one node goes offline for any reason. This makes the system of HBase more flexible and so it does a better job than the other relational database systems. Due to all of these advantages and qualifications, HBase has been vitally used in the system model.

Hadoop-HBase Data Model

The model contains four tables: User, User Application, Browser Extension and Results. They are shown in Figure 1, 2, 3, and 4 respectively. "User" is designed to maintain the users of the system indexed by emails as the row key. "User" table contains two columns "Personal Info" and "Login Info". "Personal Info" contains "first_name", "last_name" and "phone_no". "Login Info" has the attributes "user_name", "id", "encrypted_password", "last_login", "date_created", "remember_me" and "last_updated".

To clarify how the concept of de-normalization has been introduced to avoid complex joins, some examples relevant to the system are given below:

Two tables have been taken in traditional DBMS for User Application and Browser Extension and another table is there to maintain their relations. "User Application" is in "one-to-one" relationship with "Browser Extension" as app_id and ext_id (user_id) are connected. To get rid of this join table an add-on-fly column has been imposed in both "User Application" and "Browser Extension" tables. In "User Application" table, "ext_id" has been added as the column name for each new id attached with that application. On the other side, "dataset_id" has been added as a column name for every application user who has been attached to the browser extension.

Regarding "one-to-one" relationship, composite row key has been introduced here. So, join operation in relational BDMS has been created. Therefore, using composite row key in "User Application" table all the user's app data can be retrieved in single selection operation.

Row Key	**User**									
	Personal Info			**Login Info**						
Email	first_name	last_name	Phone_no	user_name	user_id	encrypted_password	last_login	date_created	remember_me	last_updated

Figure 8.2: USER HBase Table, to hold user logs

Row Key	User Application							
	General Info	RealTime Data				Browser ClsFtre	Feature ClsFtre	
user_id + app_data	user_id	step_counts	calories_burnt	distance_covered	sleep_time	ext_id	feature_text	class_level

Figure 8.3: App HBase Table, to store all information about app data sources

Row Key	Browser Extension					
	General Info					Dataset ClsFtre
user_id + ext_data	date_created	time_created	user_id	description	link_visited	dataset_id

Figure 8.4: Browser Extension HBase Table, to manage browser extension services

Row Key	Results
	General Info
app_data + ext_data + user_id	result

Figure 8.5: Table to store all possible results generated by a service

HDFS is an invaluable part of the system of HBase. It is also fault-tolerant by design. The collected data is stored in different blocks and several replicas of the data are generated in multiple nodes. If there is any disturbance in any node or even in an entire branch, it will cause a very limited effect on the overall system.

Similarly, Kafka is also an open source software which allows to store, read, and analyze streaming data. MongoDB is a document-oriented database that uses a dynamic structure to store data in JSON. It stores the records without any issue with the data structure, such as the number of fields or the types of fields used to hold values.

Chapter 9

Sample Application: Sleep pattern Identification

In the proposed system, 3 different types of data are being analyzed. There are movement analysis, sleep analysis and browsing history analysis to be done. The process of identifying the sleep pattern is going to be discussed here. In the application, there is a specific time allocated for the start and stop of the sleeping cycle. The sleep count takes up to 15 minutes to start once the mobile screen is locked and then it goes to sleep mode. The data takes up to 24 hours to get updated on the history page. Using the acquired data, a pattern can be developed which will clarify the person's sleeping habits even more.

For the sleeping pattern identification, artificial nodes can be of important usage where all the data of multiple users will be preserved and thus help the psychologist to check if the sleeping time of an individual is regular or not. To identify the sleeping pattern here neural network can prove to be a great platform.[24,25] After the clustering operation, the user data will be classified by the neural network and serve its purpose to recognize the sleeping pattern.

An ANN is a type of neural network which depends on a collection of all the units or nodes known as "artificial neurons". Each collection can pass information from one artificial neuron to another.[26] Artificial neurons are divided into multiple layers. These layers may cause different changes in their inputs. Signals move from the first layer (input layer) to the last layer (output layer), possibly after going through all the layers multiple times.[27,28,29,30]

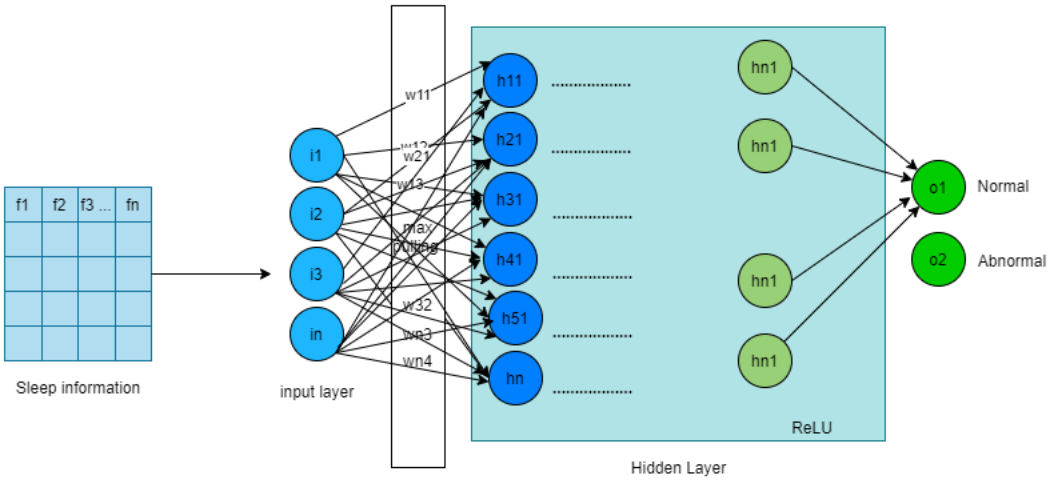


Figure 9.1: Architecture of Sleep Pattern Analysis using ANN

Artificial Neural Networks in Machine Learning is used in determining how the application collects and showing the retrieved data. After gathering the data from the application and browser extension, neural network and machine learning will be used to recognize any pattern and confirm the patient's psychological condition.[31,32] In the light of the proposed system, the artificial neurons will be able to break down all the collected sleeping data.[33,34] These sleeping data will belong to different users of different ages. The neurons will send various signals of data throughout the whole procedure.[35,36] As a result, the regular and irregular sleeping data of the users will be accumulated and thus help the psychologist to differentiate between a normal person and a mental patient.[37]

Chapter 10

Conclusion and Future Work

In today's dynamic world data science and big data are creating a huge impact on computer science. Many researchers are going on to improve people's way of living. In addition, psycho-informatics is a part of it. Based on the previous research papers it has been seen that there have been many endeavors to bring many advancements in the world of psychology. Although the advancements have been noteworthy, there is still room for improvement. Having said that, there are not many software or applications that can monitor all the daily activities of a person. But with the help of the proposed application, a person's step counts, distance covered, calories burnt and sleeping time can be calculated and thus a person can be monitored continuously. There is an average amount of calories burnt, distance walked or sleeping time which is deemed to be normal. If the user's records show something much lower or much higher than the normal amount, it can be considered as a signal of abnormality. The proposed application will keep these records for each day which can help to build a pattern of abnormality if there is any. This application will not provide the judgment about the user's mental condition itself but will present the data as a report to the psychologist. In this way, the psychologist or psychiatrist can determine whether the person is mentally stable or not.

In the future, if further researches are conducted in this area, there will be a huge improvement in the field of Big Data. The psychologists will be benefited as a result. Also, an approach can be made in the future, where Apache HBase database and HDFS (Hadoop Distributed File System) can be executed and more work in machine learning can be implemented to store necessary data such as sleeping analysis and movement tracking utilizing clustering.

To conclude, this research will result in less time-consuming for both patients and psychologists and it will be convenient for both parties in the long run. In comparison to the traditional methods of treatment, this research will help to modify and extend the advantages of psychology and psycho-informatics for the betterment of the mental health of an individual.

References

[1] J. R. Crawford and J. D. Henry, "The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample," Br. J. Clin. Psychol, vol. 42, no. Pt 2, pp. 111–131, 2003.

[2] T. A. Brown, B. F. Chorpita, W. Korotitsch, and D. H. Barlow, "Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples," Behav. Res. Ther, vol. 35, no. 1, pp. 79–89, 1997.

[3] A. Markowetz, K. Błaszkiewicz, C. Montag, C. Switala, and T. E. Schlaepfer, "Psycho-informatics: Big Data shaping modern psychometrics," Med. Hypotheses, vol. 82, no. 4, pp. 405–411, 2014.

[4] M. Samaha and N. S. Hawi, "Relationships among smartphone addiction, stress, academic performance, and satisfaction with life," Comput. Human Behav, vol. 57, pp. 321–325, 2016.

[5] C. Montag, "Recorded behavior as a valuable resource for diagnostics in mobile phone addiction: Evidence from psychoinformatics," Behav. Sci, vol. 5, no. 4, pp. 434–442, 2015.

[6] C. Montag, M. Reuter, and A. Markowetz, "The impact of Psycho-Informatics on internet addiction," Cham: Springer International Publishing, pp. 143–150,2015.

[7] M. Amin, "On curating multimodal sensory data for health and wellness platforms," Sensors (Basel, vol. 16, no. 7, p. 980, 2016.

[8] C. Montag, É. Duke, and A. Markowetz, "Toward Psycho-informatics: Computer science meets psychology," Comput. Math. Methods Med, p. 2983685, 2016.

[9] I. N. Beaufort, G. H. D. W.-V. Oene, V. A. J. Buwalda, J. R. J. Leeuw, and A. E. Goudriaan, "The depression, Anxiety and Stress Scale (DASS-21) as a screener for depression in substance use disorder inpatients: A pilot study," Eur. Addict. Res, vol. 23, no. 5, pp. 260–268, 2017.

[10] T. P. S. Oei, S. Sawang, Y. W. Goh, and F. Mukhtar, "Using the depression anxiety stress scale 21 (DASS-21) across cultures," Int. J. Psychol, vol. 48, no. 6, pp. 1018–1029, 2013.

[11] Yarkoni T. Psycho-informatics: New Horizons at the interface of the psychological and computing sciences. Current directions in psychological science,21(6):391–397, 2012.

[12] R. Hecht-Nielsen, "Theory of the backpropagation neural network**based on 'nonindent' by Robert Hecht-Nielsen, which appeared in proceedings of the international joint conference on neural networks 1, 593–611, June 1989. © 1989 IEEE," in Neural Networks for Perception, Elsevier, pp. 65–93, 1992.

[13] J. K. Basu, D. Bhattacharyya, and T.-H. Kim, "Use of artificial neural network in pattern recognition,"

[14] A. Markowetz, K. Błaszkiewicz, C. Montag, C. Switala, and T. E. Schlaepfer, "Psycho-informatics: Big Data shaping modern psychometrics," Med. Hypotheses, vol. 82, no. 4, pp. 405–411, 2014.

[15] S. Sagiroglu and D. Sinanc, "Big data: A review," in 2013 International Conference on Collaboration Technologies and Systems (CTS), 2013.

[16] V. Bhupathiraju and R. P. Ravuri, "The dawn of Big Data - Hbase," in 2014 Conference on IT in Business, Industry and Government (CSIBIG), 2014.

[17] K.-L. Du, "Clustering: a neural network approach," Neural Network, vol. 23, no. 1, pp. 89–107, 2010.

[18] J. Cohen, et al., "MAD skills: new analysis practices for big data," Proc. VLDB Endow., vol. 2, pp. 1481-1492, 2009.

[19] F. Chang, et al., "Bigtable: A distributed storage system for structured data," ACM Transactions on Computer Systems (TOCS), vol. 26, p. 4, 2008.

[20] M. N. Vora, "Hadoop HBase for large scale data," in Computer science and network technology (ICCSNT), 2011 international conference on, pp. 601-605, 2011.

[21] P. Hunt, et al., "ZooKeeper: Wait free Coordination for Internet scale Systems," in USENIX annual technical conference, p.9, 2010.

[22] A. Thusoo, et al., "Hive: a warehousing solution over a map reduce framework,"

Proceedings of the VLDB Endowment, vol. 2, pp. 1626-1629, 2009.

[23] A. Khurana, "Introduction to HBase schema design," White Paper, Cloudera, 2012.

[24] M. Y. Rafiq, G. Bugmann, and D. J. Easterbrook, "Neural network design for engineering applications," Computer. Struct., vol. 79, no. 17, pp. 1541–1552, 2001.

[25] P.E. Ross, "Flash of Genius", Forbes, pp. 98-104, Nov, 1998.

[26] A.K. Jain, J. Mao, and K.M. Mohiuddin, "Artificial Neural Networks: A Tutorial", Computer, pp. 31-44, Mar, 1996.

[27] T.Kohonen, "Self-Organizing Maps", Springer Series in Information Sciences, Berlin, vol.30,1995.

[28] Syed Ayaz Ali Shah, Azzam ul Asar and S.F. Shaukat, "Neural Network Solution for Secure Interactive Voice Response", World Applied Sciences Journal 6 (9), 1264-1269, 2009.

[29] Xinyu Guo, Xun Liang and Xiang Li, "A Stock Pattern Recognition Algorithm Based on Neural Networks", Third International Conference on Natural Computation, Volume 02,2007.

[30] Wenjin Dai and Ping Wang, "Application of Pattern Recognition and Artificial Neural Network to Load Forecasting in Electric Power System", Third International Conference on Natural Computation, Volume 01,2007.

[31] Shahrin Azuan Nazeer, Nazaruddin Omar, Khairol Faisal Jumari and Marzuki Khalid, "Face detecting using Artificial Neural Networks Approach", First Asia International Conference on Modelling Simulation, 2007.

[32] Lin He, Wensheng Hou, Xiaolin Zhen and Chenglin Peng, "Recognition of ECG Patterns Using Artificial Neural Network", Sixth International Conference on Intelligent Systems Design and Applications, Volume 02, 2006.

[33] Hongwei Yang, Chen He, Wentao Song and Hongwen Zhu, "Using Artificial Neural Network approach to predict rain attenuation on earth-space path", Antennas

and Propagation Society International Symposium, IEEE-Volume 02,2000.

[34] Nallasamy Mani and Bala Srinivasan, "Application of Artificial Neural Network Model for Optical Character Recognition", IEEE international conference, 12-15 Oct, 1997.

[35] Young-Sang Han, Seong-Sik Min, Won-Ho Choi and Kyu-Bock Cho, "A Learning Pattern Recognition System using Neural Network for Diagnosis and Monitoring of Aging of Electrical Motor", International Conference, 9-13 Nov, 1992.

[36] P. M. Grant, "Artificial neural network and conventional approaches to filtering and pattern recognition", Electronics Communications Engineering Journal, 225, 1989.

[37] Z. Liang, M. A. C. Martell, and T. Nishimura, "A personalized approach for detecting unusual sleep from time series sleep-tracking data," in 2016 IEEE International Conference on Healthcare Informatics (ICHI), 2016.