

A PROJECT REPORT
on
MACHINE LEARNING MODEL FOR MENTAL HEALTH TESTING

*Submitted in partial fulfillment of the
requirement for the award of the degree of*

Bachelor of Computer Applications



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

**Under The Supervision
of
Mr. Samson Ebenezar U.
Assistant Professor**

Submitted By

Khushi Singh
19021180012 / 19SCSE1180013
Aditya Prakhar
19021011239 / 19SCSE1010027

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER, 2021**



**SCHOOL OF COMPUTING SCIENCE AND
ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA**

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the project, entitled “**MACHINE LEARNING MODEL FOR MENTAL HEALTH TESTING**” in partial fulfillment of the requirements for the award of the **B.TECH** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Mr. Samson Ebenezar, Assistant Professor, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by us for the award of any other degree of this or any other places.

Khushi Singh 19021180012 / 19SCSE1180013
Aditya Prakhar 19021011239 / 19SCSE1010027

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Mr. Samson Ebenezar
Assistant Professor

CERTIFICATE

The Final Project Viva-Voce examination of Khushi Singh 19SCSE1180013 and Aditya Prakhar 19SCSE1010027 has been held on _____ and his/her work is recommended for the award of B.TECH.

Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Project Coordinator

Signature of Dean

Date: 04, December, 2021
Place: Greater Noida

ACKNOWLEDGEMENT

This is the matter of great privilege for all of us to submit this project entitled “**MACHINE LEARNING MODEL FOR MENTAL HEALTH TESTING**”. We take pleasure in expressing our deep sense of gratitude to **MR. SAMSON EBENEZAR U.** in the department of Computer Science & Engineering at GALGOTIAS UNIVERSITY, Greater Noida for providing us necessary guidance and for his kind and constant encouragement which made it possible for us to complete this project work.

We also extend our gratitude to our external examiner **MR. G. NAGARAJAN** who has given us pragmatic sense to look into the matter and we are also highly obliged for his persistence in making the project complete.

Last but not the least, we express sincere thanks our colleagues, who helped us directly or indirectly to complete the project. We would also like to express our gratitude towards our parents for their kind co-operation and encouragement which helped us in completion of this project.

**KHUSHI SINGH
ADITYA PRAKHAR**

ABSTRACT

In this world of continuous and technological development a major issue is being neglected from people's concern that is their mental health. Mental health of a well-being is equally important to everyone's physical health. This is the reason by which we are inspired to propose our idea so we planned to analyze the mental health of people using Machine Learning Model which will test the mental health status of the people by diagnosing them with the help of a dataset obtained from the survey from a psychiatric hospital over patients.

The proposed model will be made with the help of the machine learning algorithms and NLP and we will be making a survey over different machine learning algorithms and after the selection of the best algorithm we will load our labelled dataset into it to predict the future mental health status of the people.

We will be using the programming Language like python and the Jupyter Notebook platform for the implementation of our proposed model.

Our model will help people to know that whether their mental health status is high, low or medium and hence will help in the timely detection and treatment of the mental illness problems.

In future, we will be preparing a model that will be more efficient and will try to improve the previous model.

Keywords: Mental Health, Machine Learning, NLP, Supervised Learning, Depression, Anxiety, Python, Jupyter Notebook.

CONTENTS

Title	Page No.
Candidates Declaration	I
Acknowledgement	II
Abstract	III
Contents	IV
List of Figures	V
Acronyms	VI
Chapter 1 Introduction	8 - 9
1.1 Introduction	8
1.2 Problem Formulation	9
1.3 Tool and Technology Used	9
Chapter 2 Literature Survey	10 – 26
2.1 Research Paper 1	10 – 12
2.2 Research Paper 2	13 – 15
2.3 Research Paper 3	16 – 19
2.4 Research Paper 4	20 – 23
2.5 Research Paper 5	24 - 26
Chapter 3 Functionality/Working of Project	27 - 32
3.1 Model Selection Survey	27 – 29
3.2 Dataset	30 - 32
Chapter 4 Results and Discussion	33 - 34
4.1 Results	33
4.2 Source Code	34
Chapter 5 Conclusion and Future Scope	35
5.1 Conclusion	35
5.2 Future Scope	35
Reference	36 - 37

LIST OF FIGURES

S.No.	Caption	Page No.
1	Figure 2.1.1	11
2	Figure 2.1.2	12
3	Figure 2.2.1	14
4	Figure 2.2.2	14
5	Figure 2.2.3	15
6	Figure 2.4.1	21
7	Figure 2.4.2	23
8	Figure 2.4.3	23
9	Figure 2.5.1	26
10	Figure 3.1	29
11	Figure 4.1	34
12	Figure 4.2	34
13	Figure 4.3	35
14	Figure 4.4	35

ACRONYMS

&	And
NLP	Natural Language Processing
EPDS	Edinburgh Postnatal Depression Scale
ML	Machine Learning
PHC	Primary Healthcare
EBP	Evidence Based Practice
IEEE Xplore	Institute of Electrical and Electronics Engineers and Institution of Engineering and Technology
ACM	Association for Computing Machinery
PIB	Parent Based Intervention

CHAPTER-1

INTRODUCTION

Introduction:

With the fast moving and continuously developing world full of immense pressure, challenges and illnesses humans are being regularly tested on different bases. While passing all these tests the one thing we just don't put our focus on is our mental health. Due to this carelessness hazardous damage is held to us hence, its importance should be understood and its treatment should also be further promoted and should evolve over the time.

Hence, we had proposed the idea to build a model with the help of the machine learning algorithms which will help us in the detection of the mental health status of the people. We had also made the survey over the mental health of the different age groups, professional groups and also had compared the different ways to identify the mental illness of the humans using the one of the most common tool of today's technological world which is Machine Learning algorithms.

We were able to get the different rates and the trends that had been followed in the number of mental illness among several age groups as well as among different professional groups also we were able to determine the treatment rates and the identification rates that helped us to know that how important are the mental health issues, their identification and their treatment.

Problem Formulation:

The mental health is always being neglected and this is the main reason that it is not being identified at the correct time which lead severe problems in the future. The reason for this is the lack of awareness among the people about the importance of the mental health in one's life, how they can identify if they or someone they know is suffering from the mental illness and how they can get themselves treated. Proper systems or models are required which should be efficient and also accurate to determine people's mental health status. This is the problem we had focused and for its solution we had proposed the idea of the machine learning model which will provide us the efficient solution to this problem.

Tools and Technology Used:

Different technologies that were used by us are python programming language, supervised Learning concepts and algorithms, csv file creation knowledge. Also, we had compared several machine learning models and had choose the best and had implemented it. Hence, we had used a python platform Jupyter Notebook for the implementation of the Machine Learning algorithms through which we had loaded our dataset for training out model and making a system that can predict people Mental Health Status.

CHAPTER 2

LITERATURE SURVEY

RESEARCH PAPER 1

Research Paper Name:

Scaling up care for perinatal depression for improved maternal and infant health (SPECTRA): protocol of a hybrid implementation study of the impact of a cascade training of primary maternal care providers in Nigeria.

Authors Name:

Oye Gureje, Bibilola Oladeji, Olatunde Olayinka Ayinde, Lola Kola, Jibril Abdulmalik ,Waheed Akinola Lanre Abass, Neda Faregh & Phyllis Zelkowitz.

Proposed Model / Implemented Plan:

This paper proposed the hybrid implementation method or study which is primarily focusing on improving the maternal health care facilities and helping women at 6 months or more post-partum to recover from the depression in the Maternal Care Clinics present at the Low- or Middle-Income countries by developing a sustainable method or implementation.

They have divided their study into different phases which are:

1 - Pre-Conditions: this study phase duration is from 1 - 6 months. In this phase they will be conducting engagement/interaction meetings with the people and planning workshops so that they can review their proposed plan. Then they will do the designing and drafting for the training and intervention manuals. In this phase the recruitment of cohort one will be done where women who had registered will

be screened by the EPDS which stands for the Edinburgh Postnatal Depression Scale.

2 - Pre-Implementation: In this phase they are going to analyze the structure of the different clinics and their rate of determining the depression rate analysis of the clinics and also determining the training required. Here the cohort two will be screened and further assessment will be done which can be demonstrated by the following figure 2.1.1.

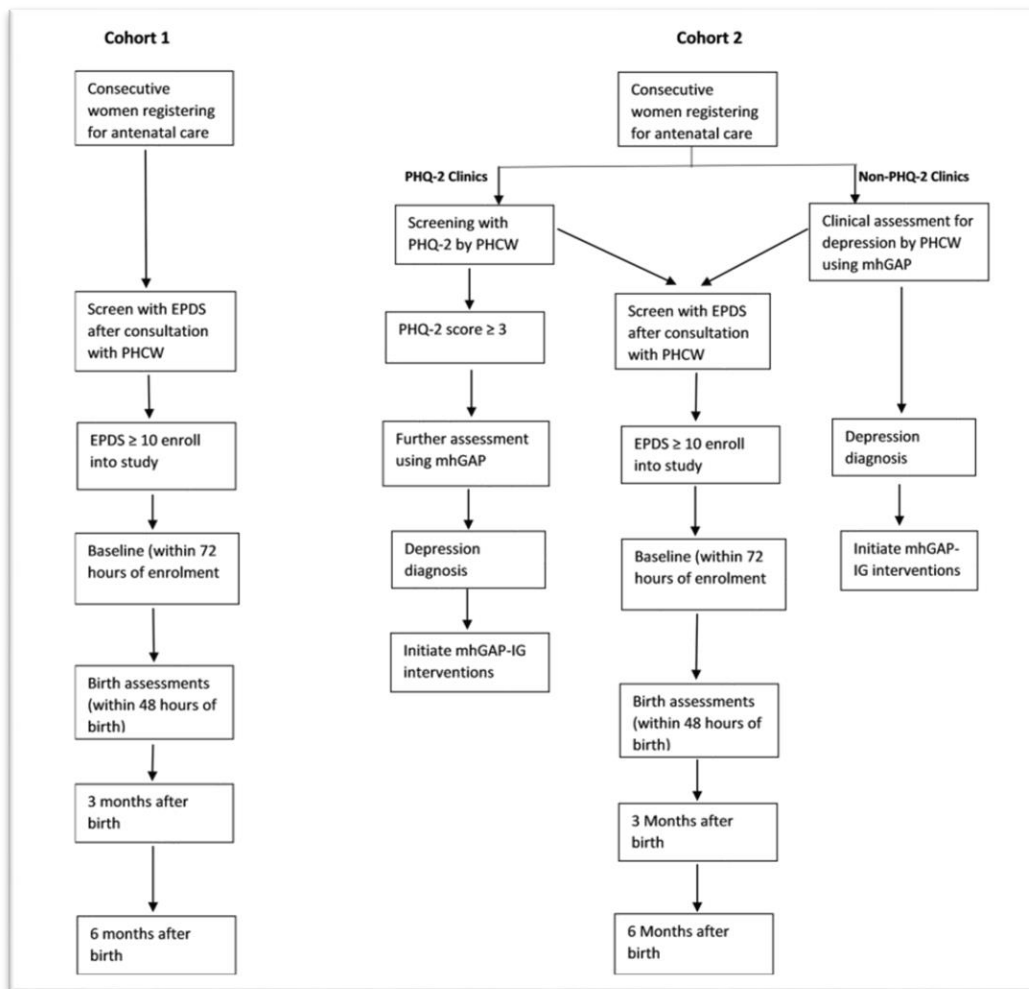


FIGURE 2.1.1

3 - Implementation: In this phase the selection of the trainers will be done and then they will be trained for achieving their proposed plan objective. They will

train all the primary care providers. This activity and assessment will take place according to this flow represented below in figure 2.1.2:

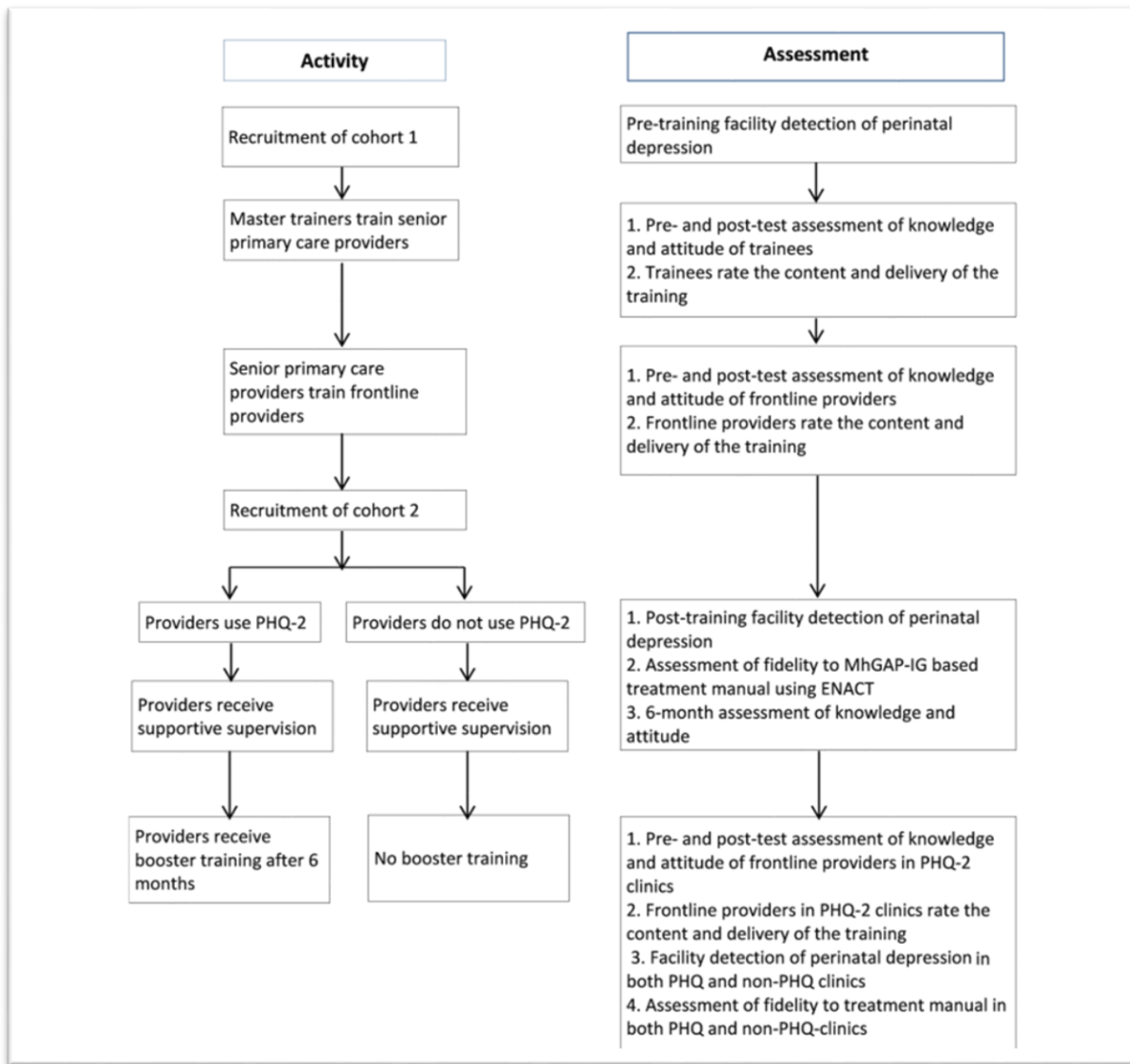


FIGURE 2.1.2

4 - Maintenance and Evolution: In this final phase the process is being evaluated and the evolution will be done on the basis of the feedback. They will collect all the outcomes that they have achieved and then they will be evaluating on the analysis gained and required changes will be made after that.

RESEARCH PAPER 2

Research Paper Name:

The architecture of the primary mental healthcare system for older people in India: what public policies tell us.

Author's Name:

Tom Kafczyk and Kerstin Hämel

Old Proposed Model / Implemented Plan:

This paper has primarily have put their focus on the mental healthcare of the old age people. They had performed the survey and have evaluated all the measure and public national policies that have been put forward for improvising the health of the older citizens. Its primary focus is on the primary healthcare level also referred as the PHC.

Proposed model covers all the Indian national policies for the public from the year 2007 to 2019 framing all the mental healthcare systems proposed for the older people. They had informed it with the help of a triangle names as the Policy Triangle Model which was developed by Gilson and Walt.

They had evaluated that how much part does the different components play for the betterment of the mental health at old age few of these components are:

- > Family
- > Supporting and empowering the health workers
- > Ayurveda
- > Yoga

Several already laid down policies which are represented below in figure 2.2.1:

Policy field	Policy (included documents)	First launched/ published	Publication year of included documents	Policy type	Primary issuing authority	Further information
Mental health	National/District Mental Health Programme (N/DMHP) [68, 115]	1982	2015/2017	Programme	Ministry of Health and Family Welfare	The programme has been subsumed under the NHM
	National Mental Health Policy [7, 116]	2014	2014	Strategic vision	Ministry of Health and Family Welfare	
	Mental Healthcare Act (MHCA) [57, 63, 117]	2017	2017/2018	Legislation	Ministry of Law and Justice	
Old age	Integrated Programme for Older Persons (IPOP) [118]	1992	2016	Programme	Ministry of Social Justice and Empowerment	The programme's latest revision of 2016 was included
	The Maintenance and Welfare of Parents and Senior Citizens Act (MWPSCA) [64]	2007	2007	Legislation	Ministry of Law and Justice	
	National Policy for Senior Citizens (NPSC) [54, 56]	2011	2011/2014	Strategic vision	Ministry of Social Justice and Empowerment	The NPSC 2011 is in a draft stage
	National Programme for the Health Care of the Elderly (NPHCE) [66, 69, 119]	2011	2011/2016	Programme	Ministry of Health and Family Welfare	The NPHCE has been subsumed under the NHM

FIGURE 2.2.1

Policy field	Policy (included documents)	First launched/ published	Publication year of included documents	Policy type	Primary issuing authority	Further information
General health	National Health Mission (NHM) [36, 38, 65, 120–124]	2005	2009/2010/2012/2013/2015	Strategic vision and guidelines	Ministry of Health and Family Welfare	The included documents primarily provide information about the NRHM, the NUHM, and CHWs
	National Programme for Control and Prevention of Cancer, Diabetes, Cardiovascular Diseases and Stroke (NPCDCS) [125–127]	2010	2010/2016/2017	Programme	Ministry of Health and Family Welfare	The NPCDCS has been subsumed under the NHM

FIGURE 2.2.2

Indian Public Health Standards (IPHS) –Guidelines for PHCCs and SHCs [40, 128]	2012	2012	Guidelines	Ministry of Health and Family Welfare	The IPHS are a health system strengthening component of the NHM
National Action Plan and Monitoring Framework for Prevention and Control of Noncommunicable Diseases (NCDs) [129]	2013	2013	Guidelines	Ministry of Health and Family Welfare	The Framework is complimentary to the NPCDCS
National AYUSH Mission (NAM) [70, 130]	2014	2014	Strategic vision	Ministry of AYUSH	The Department of AYUSH in the MoH&FW, which oversaw the NAM, became the Ministry of AYUSH in 2014
Comprehensive Primary Health Care (CPHC) [61, 67, 131]	2015	2015/2018	Strategic vision and guidelines	Ministry of Health and Family Welfare	
Rights of Persons with Disabilities Act (RPDA) [132–135]	2016	2016/2017/2018	Legislation	Ministry of Law and Justice	
National Health Policy (NHP) [62, 136]	2017	2017	Strategic vision	Ministry of Health and Family Welfare	

FIGURE 2.2.3

After evaluating they had proposed both the limitation as well as few policy recommendations which are as follows:

Limitations:

According to their survey the Indian constitute consider the health as the matter of the state the policies as being laid down for the whole nation but their implementation as well as their interpretation is completely dependent on the state only which according to their survey is the biggest limitation.

Policy Recommendations:

- > At the PHC level the intersectoral as well as collaborative policy field should be developed for the heath care of the old age.
- > Strengthening the PHC will be really important in India.
- > Unambiguous implementation rules and congruent policies should be created.
- > Policies that can support the growing of group of older people should be made and promoted among the nation.

RESEARCH PAPER 3

Research Paper Name:

College Students: Mental Health Problems and Treatment Considerations

Authors Name:

Paola Pedrelli, Maren Nyer, Albert Yeung, Courtney Zulauf & Timothy Wilens

Old Proposed Model / Implemented Plan:

This proposed model is a commentary which is focusing the problem of the prevailing mental illness among the college students of both the traditional and the non-traditional colleges. They had outlined the onset of these issues along with that the several issues or we can conclude them as the reasons due to which the psychiatric problems and substance use problems are being emerging in a large amount among the college peers are also highlighted.

According to their survey which was occurred over 274 institutions, it is being analyzed that almost 88% of the directors of the different counseling centers had reported that in comparison to the previous 5 years there is an increase in the number of severe psychiatric problems among the students. They have analyzed that the onset of the mental illness is usually in the young adulthood. It is being observed by the Kessler et al that 75% of those who are suffering from the mental illness experience their first onset by the age of 25. The reason for the triggering of onset varies from the significant disorders associated to attending college among traditional students to the pressure of demands from work and home for the non-traditional students.

Among the percentage of the numerous mental illness the most prevailing one among the students is the anxiety disorder which is upto 11.95%. Another prevalent mental illness among college students is the depression with the rate of 7

to 9%. And the third major mental illness among college students is the suicide, whose rate or percentage according to a large survey are that for among 8155 students the suicide ideation percentage is 6.7% while the percentage of suicide planning is 1.6% and 0.5% is the actual suicide committing percentage rate. They also had discovered that almost 9.5% students of every 2822 suffers from eating disorders such as bulimia, binge eating and anorexia. This problem has a higher percentage in the females in comparison to the males.

They also found that use of different substances is also common among the college students due to their mental health and therefore early detection of the mental health problems among college students is important to reduce the loss that can occur and also to obtain better output. But unfortunately, it has been discovered that it often is persisted for many years. Often it is being examined that students having any mental illness get to know about it after a year or 2 which is severely harmful for the students.

Zivin et al proclaimed in his study that among the college students facing mental illness only less than half of them had got the treatment within the correct time period. It is being concluded by several studies that the treatment rate of college students is really low no matter they are traditional students or non-traditional students. According to NESARC which stands for National Epidemiologic Survey on Alcohol and Related Conditions in 2001 – 2002 survey which was conducted on over 2188 college students the obtained rates were that only 18% students having the last year mental illness had been treated in the past year only. Among them the treated with different illness was as follows:

34% - Received stder treatment.

15% - Received anxiety disorder treatment.

5% - Received treatment for drug or alcohol disorder.

In another large survey which was conducted online by Healthy Minds it was being evaluated that only 36% of the students got the treatment last year who were diagnosed with the depression, anxiety, GAD, self-injury or self-harm, suicide planning or ideation, panic disorder, etc. Also, they got to know that the number of females receiving the treatment is more than the number of boys getting the treatment.

Eisenberg et al claimed that the appropriate care of the depression is only received by approximately half of the students. This claim of him is based on the survey of the Healthy Minds.

In this proposed commentary several daunting pertinacious natures and its implications are also being explained along with the crucial treatments that could be taken into consideration for both the traditional and non-traditional.

Some of the treatment considerations are:

The colleges and universities should do tie-ups or hold bonds and relationships with the several health care's centers for the treatment and checkup of the students facing any type of mental issues.

EBP which stands for Evidence Based Practice its interventions can help in the treatment of these type of illness for which several sites are present like:

(<http://psychiatryonline.org/guidelines.aspx>) which is being provided by the APA which stands for the American Psychiatric Association. It is basically for the online assessment and treatment of the psychiatric disorders.

(<http://www.samhsa.gov/ebpWebguide/>) which is being provided by the SAMSHA which stands for the Substance Abuse and Mental Health Services Administration. It is also an online way for registry that too on a national level. Some other

resources are NGC which stands for National Guideline Clearinghouse. It has provided (<http://www.guideline.gov/index.aspx>).

Another consideration in treatment is that the EBP used for the college students should be integrated with the MI which stands for the Motivational Interviewing as according to the surveys it helps in a large amount for the treatment of the college students.

While treating the students their context of college and development stage both should be given importance. As there can be both common as well as specific considerations for both the traditional and non-traditional students. As in the treatment of the traditional students it's really important to involve their parents in the treatment as here in this specific case the students are dependent of their parents. The parents should be briefed about their children's state of mind, problems, symptoms, treatment, recovery, their role in treatment, etc. Also, if the children don't won't to get involve in treatment or take treatment than also parents can help them by working as a medium between children and health services. Even in this regard a PIB which stands for Parents-Based Intervention is also being developed by the Turrisi et al. Similarly, for the non-traditional students it should be keenly insured that their treatment hours should be flexible as well as extended. They should be given special care management so that their financial issues, family issues, work issues, etc could be solved. And another important consideration is both the groups which is the care continuity during their breaks which plays a very important role.

RESEARCH PAPER 4

Research Paper Name:

The impact of social housing on mental health: longitudinal analyses using marginal structural models and machine learning-generated weights.

Author's Name:

Rebecca Bentley, Emma Baker, Koen Simons, Julie A Simpson, Tony Blakely.

Old Proposed Model / Implemented Plan:

In this paper they had performed the analysis over the survey of the longitudinal panel (HILDA) which had gathered the data over the tenure and the mental health of the people of over 12 years. They had evaluated about the social housing which is also termed as the welfare housing under which the house is being rented and sold at very low prices to those who need it the most and their effect on the mental health of the people.

They had thrown light over the different challenges that comes in the way to estimate the effects of the social housing on the mental health. These challenges are as follows:

- i) That major part of the data that is available is from observational studies.
- ii) For social housing allocation an important criterion is poor health.
- iii) The population keeps on changing.

They had used the marginal structural models along with the machine learning for measuring and evaluating the social housing residency over the effects of the previous health which is time varying.

Hence, the MSM is used to provide us the average effects which are not biased under some assumptions.

They had represented the relationship between the social housing tenure and the health outputs using the following directed acyclic graph.

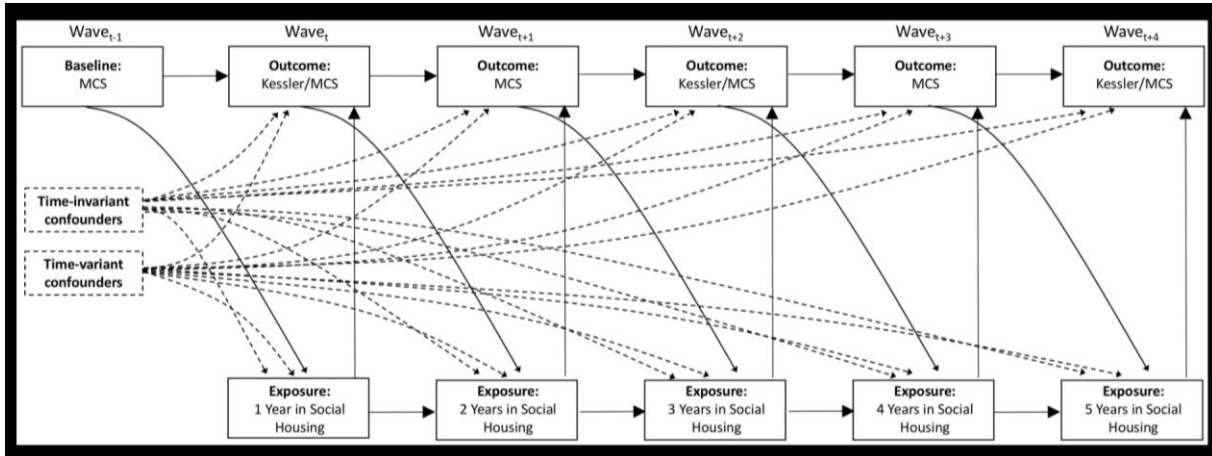


FIGURE 2.4.1

The output or target variable in this study and analysis is the psychological distress which is being measured with the help of the Kessler – 10. They had constructed the total number of the years as both categorical as well as continuous data variables whose values ranges from 0 to 5 in the social housing for the 5 consecutive years time period. Also, they had constructed the number of the total transitions from and in the social housing over the time period of 5 consecutive years, these are both categorical and continuous having the data variable values ranging between 0 and 4.

They had then used and implemented the inverse probability of treatment weights to enhance the prediction of the outcome for our model. And for this they had firstly included some cofounders which can be termed as the variables which varies over the time and are being affected by the earlier exposure. Those 5 cofounders were:

- i) Age
- ii) Gender

- iii) Country of birth
- iv) Torres Strait Islanders status
- v) Educational qualification

In inverse probability of the treatment weights the probability of the observing the ‘A’ treatment within a particular time interval ‘t’ for an individual value ‘i’ at every wave is estimated by the stabilized weights (SW).

Its mathematical representation is :

$$SW_{it} = \frac{p(A_{it}=a_{it}|\bar{A}_{it-1}=\bar{a}_{it-1})}{p(A_{it}=a_{it}|\bar{A}_{it-1}=\bar{a}_{it-1}, \bar{L}_{it}=\bar{l}_{it})}$$

Here, $\bar{a}_{it} = \{a_{it}, a_{it-1}, \dots, a_{i1}\}$ and by the multiplication of the time specific weights the final weight is being obtained.

After all these a double adjustment is made to the model for which the baseline covariates are being included in the model and it is being done to resolve the issue of the residual imbalance that used to be remained.

They tested 3 types of base learners:

- i) Logistic Regression
- ii) Gradient Boosting Machine
- iii) Conditional Inference Forest

They had concluded on their finding that the mental health of the people under the continuous social housing is worse than the people who are under other continuous tenures. And they also had found that if the number of total transitions is increasing in and out of the social housing than the mental health of that person decreases. Hence less no. of transitions is better. The descriptive result of their model is as follows in figure 2.4.2.

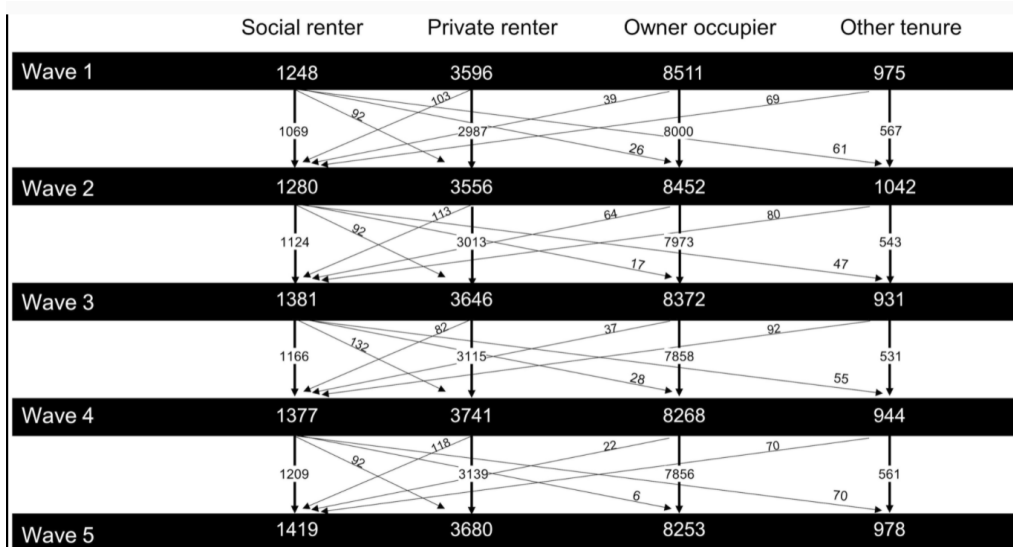


FIGURE 2.4.2

They had represented their output by the following figure 2.4.3.

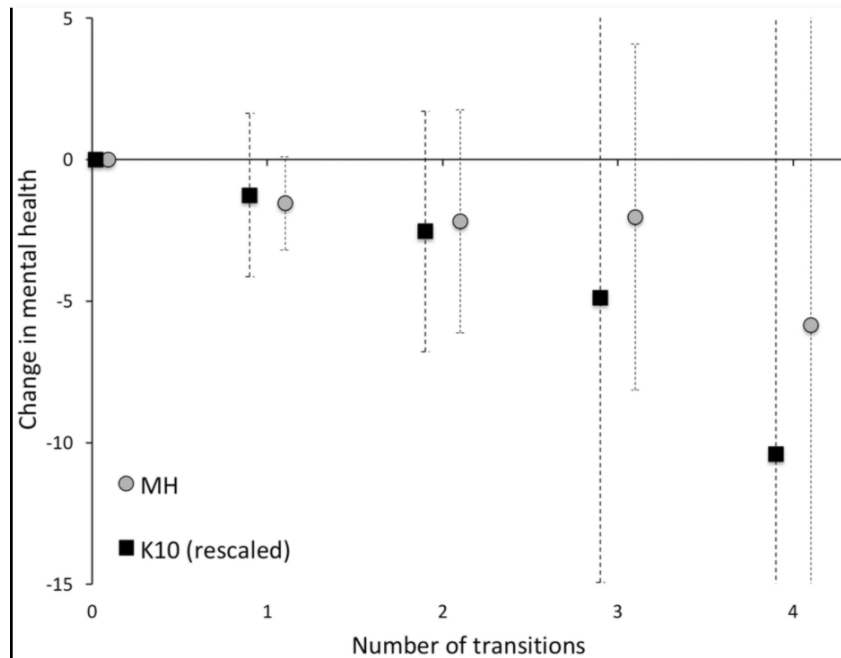


FIGURE 2.4.3

RESEARCH PAPER 5

Research Paper Name:

Machine learning in mental health: a scoping review of methods and applications.

Author's Name:

Adrian B. R. Shatte, Delyse M. Hutchinson, Samantha J. Teague

Old Proposed Model / Implemented Plan:

This paper mainly focused on the scoping review methodology to rapidly map the field of ML in mental health. In this paper 8 information technology research database were studied and applied. In this paper the concept of two reviews are used which are published till date related to this topic. The first review is Luo et al (2016) in which reviewers systematically investigated the use of big data application in the field of biomedical research and health care. In this article Luo discussed the concept that how they detected depression (using social media) by predictive models for classifying psychological condition. In the second article which is written by Bone et al(2017), in this article the reviewer described signal processing and ML for mental health research and clinical application in which they conclude how collaboration of clinicians with data scientists in leading to important scientific breakthrough.

In this paper a scoping review methodology was chosen to achieved our aim, which is to mapping the state of field of ML in mental health.

Scoping review is defined as that it is study that aims “to map rapidly the key concepts underpinning a research area and the main sources and types of evidence available, and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before”.

Search Strategy:

This strategy is adopted from Luo et al's. In this strategy a literature search conducted through health related research database including PsyInfo, the Cochrane library and PubMed, now next step is that we searched IT database IEEE Xplore and the ACM Digital library and at last we searched database that index both field including springer, science direct and Scopus.

ML Application Domains In Mental Health:

There are 4 domains of mental health application they identified through synthesis of the data:

1. **Detection and diagnosis:** This article aimed to identify or diagnose mental health conditions in individuals.
2. **Prognosis, treatment and support:** This articles that aimed to predict the progression of mental health conditions, or explore treatment or support opportunities for such conditions.
3. **public health applications:** This article used large epidemiological or public datasets (e.g. social media data) to monitor mental health conditions and estimate prevalence.
4. **Research and clinical administration:** This article aimed to improve administrative processes in clinical work, mental health research, and health-care organizations.

PRISMA Procedural Flow Chart:

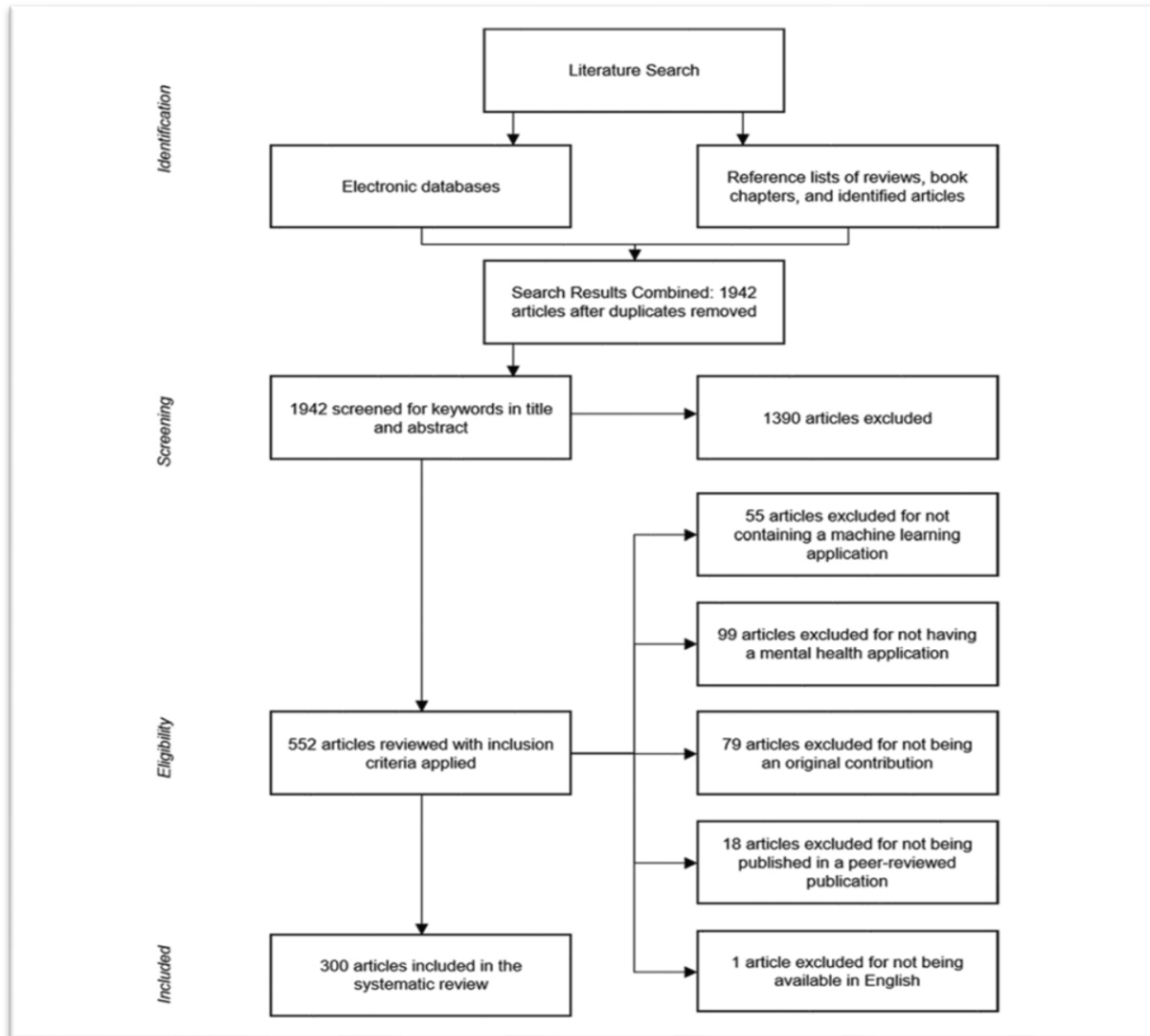


FIGURE 2.5.1

So, we conclude that in this paper we use ML at advance level which help us to significantly improve the detection and diagnosis of mental health condition. This paper also shows how ML is use to improve the area of mental health.

CHAPTER 3

FUNCTIONALITY / WORKING OF THE PROPOSED PLAN

Our proposed plan is to build an efficient and accurate machine learning model which will be able to predict people's mental health status. So, we had firstly made a survey over various machine learning algorithms to choose the best algorithm for us.

MODEL SELECTION SURVEY:

So, the different algorithms we had made the survey over, to find the best algorithm for our model or proposed plan are:

1 – Logistic Regression

It is a supervised learning algorithm which is used for the prediction of the categorical independent data with the help of the dependent variables or data.

The outcome of the logistic regression is either categorical or discrete. In the logistic regression an S shaped function is being fitted to find or predict the two maximum values. The formula on which logistic regression work is

$$\text{Log}[y/1-y] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

The likelihood of the something in the dataset is being displayed with the help of the curve of the logistic regression.

2 – K Nearest Neighbor

It is also a type of supervised learning in which we match the similarities of the new data point that we have to predict with all other and puts those data points together who shares the most similarities. It is being used for both regression as well as classification problems. In the training phase of this algorithm the data is

only stored in this algorithm and when it phases new data than at that time it classifies the data into different groups on the basis of their similarities.

In this algorithm, we find the Euclidean distance among the new data point and the other grouped data points.

Its formula will be

$$\text{Distance between point A \& B} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

3 – SVM

This algorithm works by finding the best separation line or a type of boundary through which we will be able to segregate the whole n – dimensional space that can help us to easily classify the new data points into different classes. Hyperplane is known to be the best decision boundary. This algorithm is named as support vector machine as in this algorithm the boundary points are known to be support vectors and they help in the creation of the hyperplane that will help us in the classification.

4 – Naïve Bayes Algorithm

This is also a supervised learning in which the predictions are made on the basis of their probabilities and hence its whole working is based on the mathematical concept of the Bayes theorem. The main purpose which it is widely used is text classification.

The Bayes theorem formula on which it works is

$$P(A / B) = \frac{P(B/A) * P(A)}{P(B)}$$

5 – Random Forest

It is an ensemble learning in which more than one algorithm are combined to produce a desired output. Particularly, in random forest there is collection of different independent decisions tree algorithms. The output of the random forest algorithm is based on the majority vote of the different decision tree. The increasing number of trees in the random forest increases the accuracy of the algorithm to predict any certain output.

After surveying all these algorithms, we find out some attribute or functionalities through which we can compare their performance and these feature or functions are:

1 – Recall Score

$$\text{Recall} = \text{TruePositives} / (\text{TruePositives} + \text{FalseNegatives})$$

2 – Precision

$$\text{Precision} = \text{TruePositives} / (\text{TruePositives} + \text{FalsePositives})$$

3 – F Score

$$\text{F-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

And on applying few random related datasets and surveying their performance we concluded that the K- Nearest Neighbor is the best algorithm for our purpose as for its average scores of all the random tests results are likely to be equal for all the 3 attributes and it can be represented by the following while comparing it with other algorithms graphs.

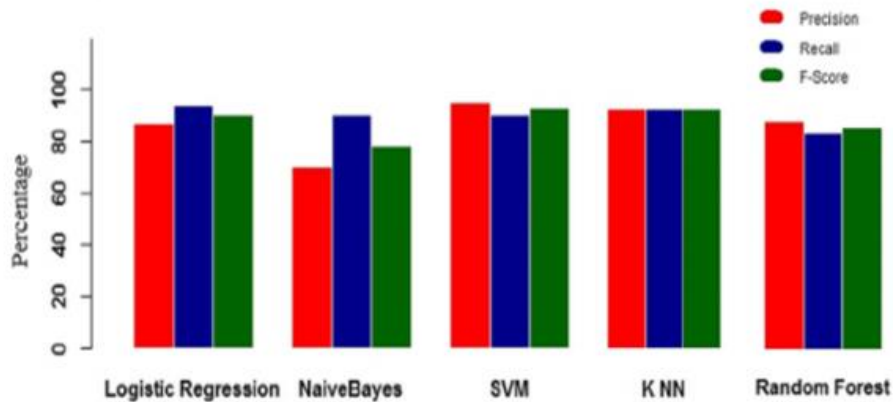


FIGURE 3.1

DATASET

After the selection of the model we had made a dataset from the survey we had made from a psychiatric hospital records which consists of 50 entries which are labelled as we are going to apply a supervised learning KNN over it. The dataset that we are going to use for the training of our model consists of fourteen parameters which are crucial to determine if a person is suffering from the mental illness or disorder or not and hence will help us to get know about one's mental status. Those 14 parameters are:

1. **Emotional Disbalance** – It comprises all the mood changes, feeling anxious, low, lonely, etc. at the various times.
2. **Loss of Interest** – It comprises the action in which one loses interest in things in which once he/she was interested in. It also plays a very important role in determining of the mental status of the person.
3. **Sleeping Habit Changes** – It depends on how the person sleeping habits had changed, do his/her sleeping duration had increased or decreased over the period of time.
4. **Low Energy** – This factor tells us if the energy rate of the person is always high, low or normal. Is there no change in the energy of the person no matter what happens.

- 5. Interaction Difficulty** – This factor is valued over if the person is getting detached to his family, friends and other people. Do he/she are not able to communicate to others and prefers to stay alone.
- 6. Weight Change** – Do the person have any pattern of weight change over the time that have no medical reason for it.
- 7. Appetite Change** - Do the person have any pattern of appetite change over the time that have no medical reason for it.
- 8. Anger Issues** – This factor is judged on the basis, if people are getting very irritated over silly things and having continuous fluctuation in his temper either people tries to harm himself or tries to harm others in their rage of the anger.
- 9. Suicidal Thoughts** – This tells us how much the person have thoughts of committing suicide at different situations.
- 10. Bi-polar Behavior** – Tells us if the person often behaves as a completely different personality and have completely different traits for only a specific time period and that too in any particular pattern common.
- 11. Concentration issues** – This factor tells us if the person finds it difficult to focus very easily and this had occurred since a particular time.
- 12. Decision making issues** – How much problem do a person have in decision making even if it is a small decision.
- 13. Excessive Imagining** – Do person remains in a completely different world of imagination and tries to create things on his own or circumstances of his own. The person prefers to spend time imagining rather than living the present moments.
- 14. Time** – This factor represents the time from which the person is experiencing the changes in these above factors. It is also one of the most important features while predicting the mental health status of a person.

These factors are being observed over a period of time that varies from 3 weeks to 4 weeks or even 2 – 3 months if the symptoms are really mild over time. These all factors are being given values ranging from 0 to 5 on the basis of the severity of the symptoms of the following factors.

So, the total score of all 13 parameters and time is 65. The algorithm on which we classified them in 3 groups are:

45-65	low	if $w \leq 2$	medium	$2 < w \leq 3$	low	$w > 3$	low
30-45	medium	if $w \leq 2$	medium	$2 < w \leq 3$	low	$w > 3$	low
Below30	high	if $w \leq 2$	high	$2 < w \leq 3$	medium	$w > 3$	low

Hence, with the help of this dataset and KNN model we will be able to predict the mental status of the person, if his/her mental health status high which is a good indication, average which most of the people do have or low that is really harmful state and one needs to consider it for treatment.

So, we fit our KNN model over this dataset and find the accuracy of our model to get to know if it will be working fine over the new data or not.

CHAPTER 4

RESULTS AND DISCUSSION

After the completion of our survey over the various supervised learning algorithms for which we compared their precision, F score and Recall score. The result of our survey is that KNN algorithm will be the most suitable for our proposed plan as its precision, F score and Recall score is almost equal and high.

Then after the implementation of the machine learning KNN algorithm over the dataset of the patients records that was collected from the survey of a psychiatric hospital we found that our model is working accurately and is able to predict the people mental health status correctly.

So, with our proposed model we are able to predict the mental health status of any person that too with the help few factors which will solve the problem of lack of knowledge of people regarding their own mental health status.

Also, with our model people will be able to find out at the very early stage of their mental illness or disorder and hence, will get treated earlier at the initial stage where less harm has been caused to them.

The results of our model can be described by the following code of our implementation.

```
In [41]: #KHUSHI SINGH 19SCSE1180013
#ADITYA PRAKHAR 19SSE1010027
import pandas as pd
```

```
In [42]: data = pd.read_csv('dataset.csv')
print(data)
```

```

    Emotional Disbalance  Loss Of Interest  Sleeping Habit Changes  \
0                        5                  3                      3
1                        4                  5                      2
2                        2                  3                      0
3                        3                  4                      1
4                        1                  4                      5
5                        2                  4                      3
6                        5                  3                      2
7                        1                  3                      2
8                        3                  2                      1
9                        3                  1                      4

    Low Energy  Interaction Difficulty  Weight Changes  Appetite Chnages  \
0             3                      3                3                2
1             1                      4                3                5
2             2                      5                1                5
3             3                      5                2                3
4             2                      4                2                1
5             5                      2                4                2
6             5                      2                5                3
7             1                      3                2                1
8             5                      4                2                1
9             2                      1                3                5

    Anger Issues  Suicidal Thoughts  Bi-polar Behavior  Concentration Issues  \
0               5                   4                   4                   4
1               2                   3                   4                   1
2               0                   2                   1                   1
3               1                   4                   5                   2
4               4                   2                   2                   2
5               1                   1                   3                   5
6               1                   5                   3                   5
7               2                   3                   1                   3
8               3                   1                   4                   2
9               2                   5                   2                   2
```

FIGURE 4.1

```

    Decision Making Issues  Excessive Imagining  Time Status
0                          2                   4           1       Low
1                          4                   5           2       Low
2                          1                   2           2       High
3                          4                   2           1       High
4                          4                   1           2       High
5                          2                   4           3       Low
6                          1                   4           4       Low
7                          2                   1           5       Low
8                          1                   4           2       High
9                          1                   3           3       low
```

```
In [43]: print(data.iloc[:,0:14])
```

```

    Emotional Disbalance  Loss Of Interest  Sleeping Habit Changes  \
0                        5                  3                      3
1                        4                  5                      2
2                        2                  3                      0
3                        3                  4                      1
4                        1                  4                      5
5                        2                  4                      3
6                        5                  3                      2
7                        1                  3                      2
8                        3                  2                      1
9                        3                  1                      4

    Low Energy  Interaction Difficulty  Weight Changes  Appetite Chnages  \
0             3                      3                3                2
1             1                      4                3                5
2             2                      5                1                5
3             3                      5                2                3
4             2                      4                2                1
5             5                      2                4                2
6             5                      2                5                3
7             1                      3                2                1
8             5                      4                2                1
9             2                      1                3                5

    Anger Issues  Suicidal Thoughts  Bi-polar Behavior  Concentration Issues  \
0               5                   4                   4                   4
1               2                   3                   4                   1
2               0                   2                   1                   1
3               1                   4                   5                   2
4               4                   2                   2                   2
5               1                   1                   2                   5
```

FIGURE 4.2

```

Decision Making Issues Excessive Imagining Time
0 2 4 1
1 4 5 2
2 1 2 2
3 4 2 1
4 4 1 2
5 2 4 3
6 1 4 4
7 2 1 5
8 1 4 2
9 1 3 3

In [44]: print(data.iloc[:, 14:15])

Status
0 Low
1 Low
2 High
3 High
4 High
5 Low
6 Low
7 Low
8 High
9 low

In [59]: X = data.iloc[:,0:14]
y = data.iloc[:, 14:15]

In [77]: from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.4, random_state=1)

In [78]: print(y_train)

Status
0 Low
3 High
1 Low
7 Low
8 High
5 Low

```

FIGURE 4.3

```

0 Low
3 High
1 Low
7 Low
8 High
5 Low

In [79]: from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=3)

In [80]: knn.fit(X_train,y_train)

<ipython-input-80-889312abc571>:1: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
knn.fit(X_train,y_train)

Out[80]: KNeighborsClassifier(n_neighbors=3)

In [81]: from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report

In [82]: score = accuracy_score(y_test,y_pred)

In [83]: y_pred=knn.predict(X_test)
cm=confusion_matrix(y_test,y_pred)

In [84]: print(cm)

[[0 2 0]
 [0 1 0]
 [0 1 0]]

In [85]: print(" correct prediction", (1-score))

correct prediction 0.75

In [86]: print(" wrong prediction",score)

wrong prediction 0.25

```

FIGURE 4.4

CHAPTER 5

CONCLUSION & FUTURE SCOPE

Conclusion

The problem statement has been solved with our proposed plan in which we had first made a survey over the various supervised learning algorithms to determine which one is the best and we concluded that the KNN algorithm will be the best for the implementation of our proposed idea and then we had used the supervised learning KNN algorithm and had trained it over an labelled dataset of the different patients who had come for testing if they suffer from any mental health disorder or illness and find the accuracy of the model to assure that our model will determine the correct results for the various new inputs or patients values.

Future Scope

1. We are planning to make a survey among the college student of the different universities at a region and then will predict which college peer do have the highest ratio of poor mental health status.
2. Also, we are planning to collaborate this idea or model with a mobile application with some additional functionalities and tie-ups with several psychiatric hospitals.

REFERENCES

1. Funk M, et al. Integrating mental health into primary healthcare. *Mental Health Fam Med*. 2008;5(1):5–8.
2. World Health Organization. *Mental Health Atlas*. Geneva: World Health Organization; 2005.
3. Baron EC, et al. Maternal mental health in primary care in five low- and middle-income countries: a situational analysis. *BMC Health Serv Res*. 2016;16:53.
4. Thara R, Padmavati R, Aynkran JR, John S. Community mental health in India: a rethink. *Int J Ment Heal Syst*. 2008;2(1):11.
5. Das A. The context of formulation of India's mental health program: implications for Global Mental Health. *Asian J Psychiatr*. 2014;7:10–4.
6. Goel DS. Why mental health services in low- and middle-income countries are under-resourced, underperforming: an Indian perspective. *Natl Med J India*. 2011;24(2):94–7.
7. Blanco C et al. Mental health of college students and their noncollege-attending peers: Results from the National Epidemiologic Study on Alcohol and Related Conditions. *Arch Gen Psychiatry*. 2008;65(12):1429–37.
8. Statistics, N.C.f.E., Nontraditional Undergraduates, U.S.D.o.E. Institute of Education Sciences, Editor 2013.
9. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. *Am Psychol*. 2000;55(5):469–80.
10. Miner, L., et al., Practical predictive analytics and decisioning systems for medicine: Informatics accuracy and cost-effectiveness for healthcare

administration and delivery including medical research. Cambridge: Academic Press, 2014.

- 11.**Luxton, D. D., (ed.) *Artificial Intelligence in Behavioral and Mental Health Care*. Amsterdam: Elsevier Inc., 2015.
- 12.**Hahn, T., Nierenberg, A. A., and Whitfield-Gabrieli, S., Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Mol. Psychiatry* 22(1):37–43, 2017.
- 13.**Abou-Warda H, Belal NA, El-Sonbaty Y and Darwish S (2017) A random forest model for mental disorders diagnostic systems. In Hassanien A, Shaalan K, Gaber T, Azar A and Tolba M (eds), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016. AISI 2016. Advances in Intelligent Systems and Computing*, vol 533. Cham: Springer, pp. 670–680
- 14.**Agarwal A, Baechle C, Behara RS and Rao V (2016) Multi-method approach to wellness predictive modeling. *Journal of Big Data* 3, 15.
- 15.**Aguilar-Ruiz JS, Costa R and Divina F (2004) Knowledge discovery from doctor-patient relationship. In *Proceedings of the 2004 ACM Symposium on Applied Computing SAC '04*. New York, NY, USA: ACM, pp. 280–284.
- 16.**Baker E, Bentley R, Lester L, Beer A. Housing affordability and residential mobility as drivers of locational inequality. *Appl Geogr* 2016; 72:65–75.
- 17.**Productivity Commission. Steering Committee for the Review of Government Service Provision, Volume C: Justice Sector Overview. Table C. Parkes, ACT: Government Department of Treasury, 2016.
- 18.**Seelig T, Han JH, O'Flaherty M et al. *Housing Consumption Patterns and Earnings Behavior of Income Support Recipients Over Time*. Melbourne, VIC: Australian Housing and Urban Research Institute, 2005.