IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Detection of Impact of Covid-19 Pandemic on Emotional well being of people using Machine Learning Approach

Deepali Joshi, Harshali Patil
Thakur College of Engineering & Technology Mumbai, Maharashtra

***Abstract*— Coronavirus disease 2019 is a global pandemic caused by the (SARS-CoV-2) Severe Acute Respiratory Syndrome. During the outbreak of an infectious disease, the population's psychological reactions play a crucial role in affecting both the spread of the disease and the occurrence of emotional distress. It is recognized that there are number of techniques and tools which are used to detect the trauma caused by it. To assess whether the individual is affected or not, machine learning techniques are used. The course of action is based on texts. Among different techniques, it is analyzed that none of the technique is using the probability method to detect the impact of pandemic. In probability method, one can create a database, which is a collection of words related to emotions like happiness, sadness and anger etc., The probability would be set for all such words in the database and then it will be calculated by various techniques which shows the level of emotional well being of a person. This paper summarized the findings on the identification of emotional disorders, using emotion analysis methods and techniques. The author focused on research that identifies irregular activity patterns on social networks automatically. The studies selected used the classic off-the-shelf classifiers to evaluate the knowledge available for lexicon use.**

Keywords*— *Machine Learning, Random Forest, Naive Bayes, Covid-19, KNN, Emotional Analysis

I. INTRODUCTION

COVID-19 has caused unimaginable trauma and stress all over the world. On 30th January 2020, after all the regions of China were detected to be prone to novel coronavirus infection cases and the total number of cases had crossed the total case count of SARS (2003), the outbreak was declared as a pandemic, a public health emergency of international concern by the World Health Organization (WHO) [1]. Traumas like pandemics lead to feelings of uncertainty about life that ultimately cause stress, frustration, fear, anxiety, panic, sleep problems [2]. It can be seen that this pandemic has led to direct and indirect psychological and social effects which are pervasive and are affecting mental health now and might continue in the future too [3]. Isolation strategy with quarantine has turned out to be an effective model in controlling transmission and rapid spread of the virus [4]. There is a definite possibility for this because over 10 million total cases all over the world have been registered by the end of June 2020 [5] and have reached almost 10.5 million by July mid. As per a research, in India, under normal circumstances, around 10–12% of people are affected by a mental disorder either due to stress, anxiety, or any other cause [6]. It can be assumed that the level of mental issues could have increased manifolds by now, under the current circumstances.

The outbreak of the novel Coronavirus Infectious Disease 2019 (COVID-19) has caused an unprecedented impact on people's daily lives around the world. People's lives are at risk because the virus can easily spread from person to person either by coming in close contact with the infected person or sometimes may even spread through community transmission , which then becomes extremely challenging to contain. The infection had rapidly spread across the world and there had been more than 10.3 million confirmed cases and more than 505,000 people have died because of the infection until 30 June 2020. Almost every country in the world is battling against COVID-19 to prevent it from spreading as much as possible. As a result, this outbreak has caused immense distress among individuals either through infection or through increased mental health issues, such as depression, stress, worry, fear, disgust, sadness, anxiety (a fear for one's health, and a fear of infecting others), and perceived stigmatisation. These mental health and emotional health issues can even occur in people, not at high risk of getting infected. There could be even several people who are exposed to the virus may be unfamiliar with it as they may not follow the news, or are completely disconnected with the general population.

II. RELATED WORKS

A number of surveys have been conducted in order to detect the prevalence of emotional health problems in many sectors. These surveys clearly show that various factors like layoffs, work pressure, job evaluation and appraisals, competition, et cetera cause depression in employees at varying levels and stages. Countries like Singapore uses the impact of event scale-Revised (IESR), which is a standard instrument used by healthcare workers, for depression detection. Another survey was carried out in 2015, in industrial workers of Bangalore, India where the researchers used a Depression Anxiety Stress Scale (DASS)-21, as the mental health screening

tool that has been discussed below. One of the approaches to determine the depression among firm employees was proposed by Rao et al where they made use of the Depression Anxiety Stress Scale (DASS-21). Logistic Regression was used to test the relationship between the DASS-21 scores and the variables like age, the number of years in a firm, leaves, and sick leaves. It is said that the pandemic has various effects on a individual even if the person is not actually infected. Various emotional issues are seen in people of different age groups[12]. The person may be a child, teenager, middle age person or senior citizen. Educational sector has witnessed a drastic change as everything had shifted to online mode and students are being affected due it since the teaching learning methods have changed[11]. The students who are in school and are on the verge of selecting their streams for their future are drastically affected as there is lot of confusion in their minds[13]. The students are searching for different career options that they can opt[14][15].

.

III. EXISTING SYSTEMS

The most recent work using Twitter to analyze emotional and mental health issues due to COVID-19 is developed. These work focus more on public sentiment analysis. Furthermore, little work such as Li et al classify each tweet into the emotions of anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The two emotions of sadness and fear are more related to severe negative sentiments like depression due to COVID-19.

Little work is done to detect depression dynamics at the state level or even more granular level such as suburb level. Such granular level analysis of depression dynamics not only can help authorities such as governmental departments to take corresponding actions more objectively in specific regions if necessary but also allows users to perceive the dynamics of depression over the time to learn the effectiveness of policies implemented by the government or negative effects of any big event.

IV. METHODOLOGIES

A. Proposed System:

The proposed model consists of various steps such as data collection in which the data that is required is gathered. The data pre-processing steps actually pre-processes the data, for example data to be updated in a particular format or manipulating or deleting of data to enhance the performance. The data is divided into two sets that is training and testing, where the proposed model is trained using training set and the proposed model is tested using test set. Once this operation is done then the algorithms are applied and after the processing the results are predicted.

In this paper a structural model is presented that identifies users' emotions from the dataset as well as the questionnaire. It consists of two data sets. The first one is the training and the second one is training. After collecting the data, pre-processing it the algorithms are applied along with fresh inputs.

Steps:

Step 1: Data Collection

Step 2: Data Pre-Processing

Step 3: Training and Testing

Step 4: Application of Algorithms

Step 5: Prediction of Results

The proposed system architecture is given below:

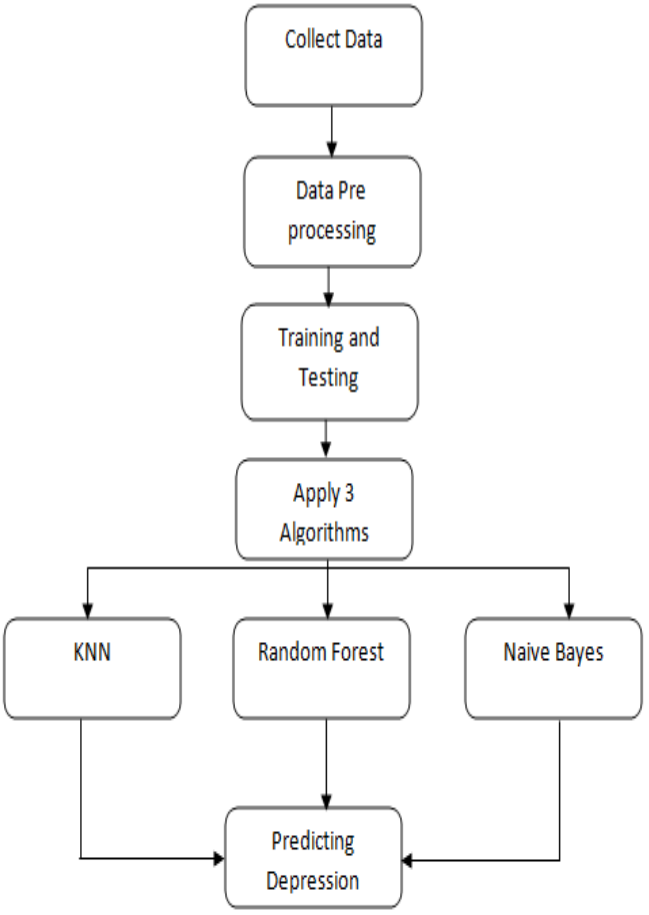

Figure 1: System Flow Chart

B. Datasets

This dataset is from a 2014 survey that measures attitudes towards emotional and mental health and frequency of disorders in the tech workplace.

This dataset contains the following data:

- Timestamp
- Age
- Gender
- Country
- state: If you live in the United States, which state or territory do you live in?
- self_employed: Are you self-employed?
- family_history: Do you have a family history of mental illness?
- treatment: Have you sought treatment for a mental health condition?
- work_interfere: If you have a mental health condition, do you feel that it interferes with your work?

- no_employees: How many employees does your company or organization have?
- remote_work: Do you work remotely (outside of an office) at least 50% of the time?
- tech_company: Is your employer primarily a tech company/organization?
- benefits: Does your employer provide mental health benefits?
- care_options: Do you know the options for mental health care your employer provides?
- wellness_program: Has your employer ever discussed mental health as part of an employee wellness program?
- seek_help: Does your employer provide resources to learn more about mental health issues and how to seek help?

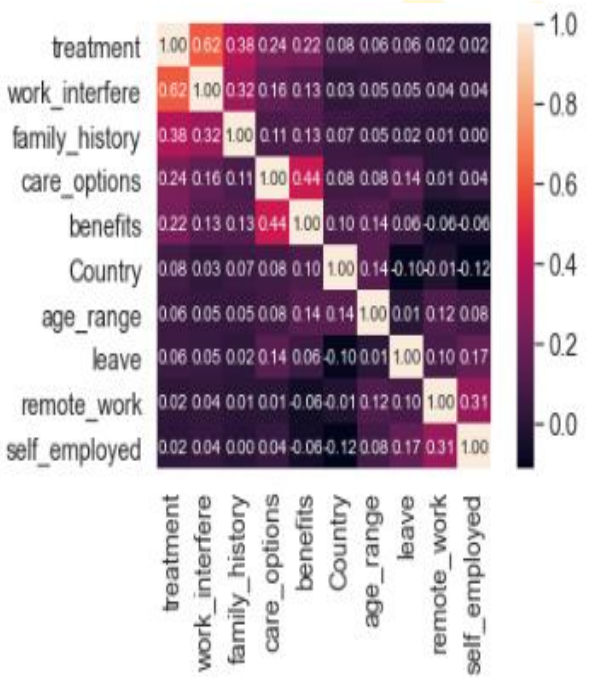

Figure 2: Covariance Matrix

Data Cleaning

Data cleansing is an essential tool for data and NLP research. Text cleanup removes stop words (words that don't affect the model very much). These are usually pronouns, conjunctions, and so on. In addition, code cleanup in NLP removes unique characters such as @, #, and trailing characters. Stemming translates the same term as a word with a specific root, so the model is applied in all cases where similar words are used, and it works better and minimizes problems.

C. Tokenization and Sequence Padding

Tokenization is a greater state-of-the-art shape of statistics control that could fit and change exceptional textual content files. It can be the proper alternative for big projects. The Tokenizer ought to be constructed to suit both uncooked textual content or encoded textual content files in an integer. Next is sequencing and collection padding in which we remodel the tokens into sequences used for a sequential pattern.

These sequences change in duration, so we pad positive sequences to get all of the statements or sequences to a popular length.

D. Naïve Bayes Classifier

Naïve Bayes classifiers are highly used for text based classification. It is a probabilistic learning model that applies Bayes Theorem [7]. Naïve Bayes algorithm is a supervised learning algorithm that works on Bayes theorem and solves classification problems. It is used in text classification which includes a training dataset. Naïve Bayes algorithm is the most simple and effective Classification algorithms which helps in building the machine learning models that can generate predictions quickly.

Bayes' Theorem:

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

The dataset is first subjected to the pre-processing stage. Pre-processing consists of following steps:

1. Tokenization: This process splits the given text into relevant tokens like characters, words, phrases etc., Here word –level tokenization is used [7].
2. Bag-of-words: Bow counts the number of times a particular token appears in the given text. To achieve this a class called Count Victories from scikit-learn is used.[7]

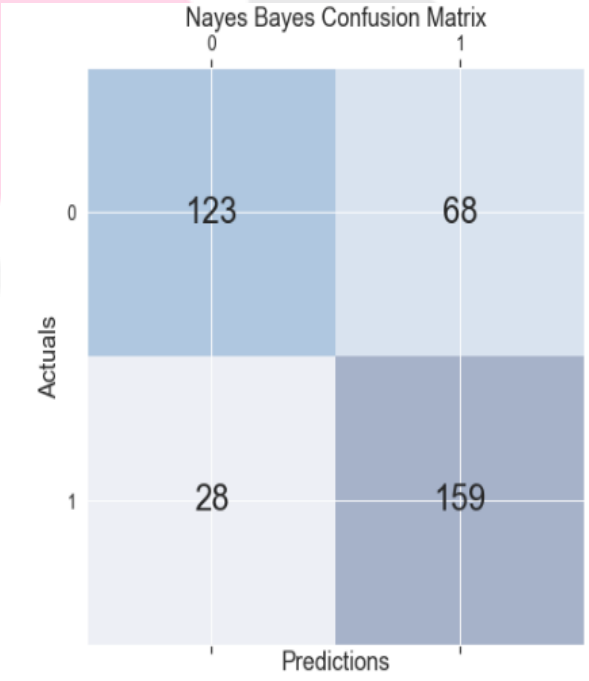

Figure 3: Navie Bayes Confusion Matrix

E. Random Forest

Random Forest algorithm is a supervised learning method. It can perform Classification as well as Regression problems in Machine Learning. It is based on ensemble learning. The Ensemble learning is a process of combination of multiple classifiers to solve a complex problem and will improve the performance of the model.

In Random Forest Algorithm, the dataset is divided into several individual trees and forms a class with similar functionalities. Then from the different classes, the final class is selected by voting the majority. The main advantage of any decision tree-based algorithm like Random Forest is the clear and understandable prediction rules that may be generated from the training dataset.

Features of Random Forests

- It is unexcelled in accuracy among current algorithms.
- It runs efficiently on large data bases.
- It can handle thousands of input variables without variable deletion.
- It gives estimates of what variables are important in the classification

The confusion matrix for Random Forest is given below and the actuals and predictions can be seen in the figure.

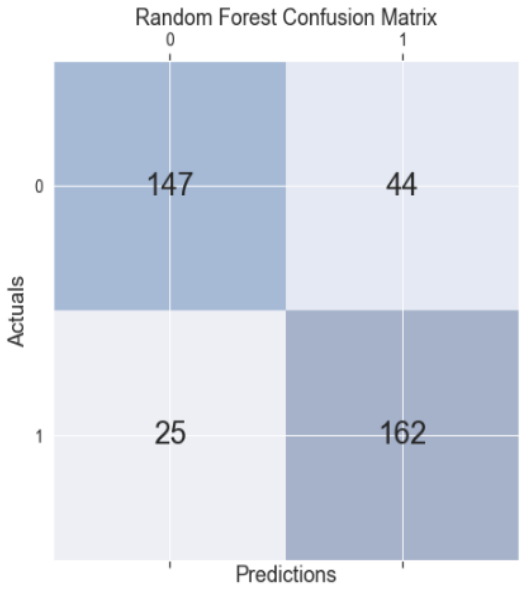

Figure 4: Random Forest Confusion Matrix

F. KNN

The KNN is K-Nearest Neighbour algorithm. It is the most simplest type of Machine Learning algorithm. The KNN is a Supervised Learning algorithm. K-NN identifies the similarity between the new data and the old data that is available and adds the new data into the category that is most similar to the available categories. The K-NN stores all the available data and classifies new data based on the similarity. The K-NN algorithm can be used for Regression and Classification both but generally it is used for the Classification.

To decide which information relates to which group, the KNN examines the distance between features inside a data set. When the distance between data points is close and there are lots of them, a group is formed.When the gap between data points is large, groups arise. K-Nearest Neighbours (KNN) is a simple algorithm for regression and classification problems in Machine Learning. KNN algorithms take data and uses similarity measurements (such as the distance function) to classify fresh data points. A majority vote is taken to classify its neighbours.

Characteristics of kNN

- Between-sample geometric distance.
- Classification decision rule and confusion matrix.
- Feature transformation.
- Performance assessment with cross-validation.

The confusion matrix for KNN is given below and the actuals and predictions can be seen in the figure.

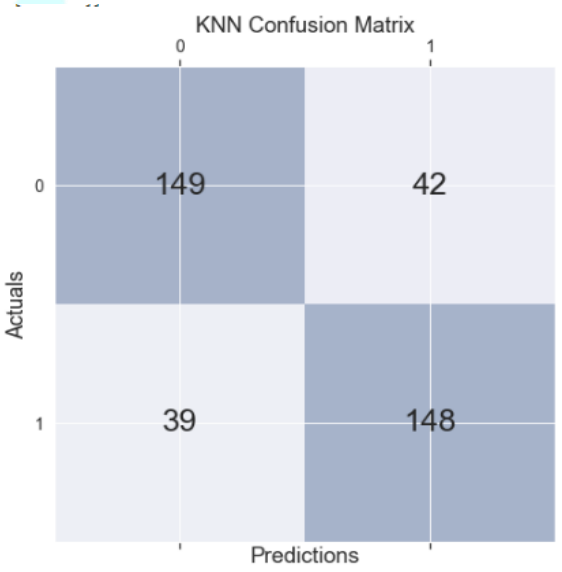

Figure 5: KNN Confusion Matrix

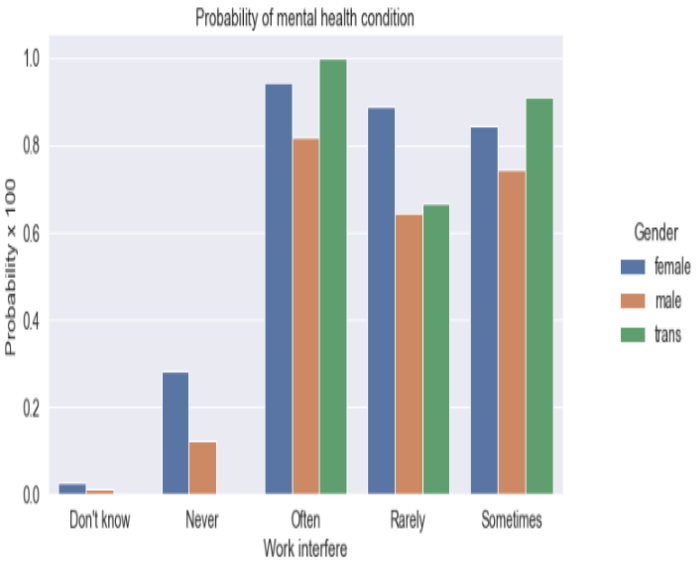

Figure 6: Barplot to show probabilities

The above diagram indicates the barplot of probabilities for mental conditions. The parameters on the x axis are don't know, never, often, rarely and sometimes. The values that are plotted are for gender consisting of male, female, trans.

V. RESULTS

The system provides a result that indicates if the individual is emotionally stable or not. It will be evaluated whether the person's mindset is positive, negative, or neutral. A positive mind-set is associated with a lower risk of emotional instability and trauma, while a negative mind-set is associated with the opposite.

The dataset is evaluated with help of machine learning algorithms and along with that a questionnaire is utilized. The result is evaluated with help of score that is obtained. Three machine learning algorithms are applied on the data that is Naïve Bayes, Random forest and KNN.

The parameters that are considered are f-1 score, recall and accuracy. The f-1 score for Naïve Bayes Classifier is 74%, Random forest is 82% and KNN is 79%. The recall score for Naïve Bayes Classifier is 75%, Random forest is 82% and KNN is 79%. The Random Forest algorithms provides 82% accuracy rate, Naïve Bayes with 76% and KNN with 79% accuracy rate and is mentioned in Table 1 with other parameters. Thus from the table it can be observed that Random forest provides results with good accuracy percentage.

TABLE 1: Model Accuracy on dataset

Model	f-1 score	Recall	Test Accuracy
Naïve Bayes Classifier	74%	75%	76%
Random forest	82%	82%	82%
KNN	79%	79%	79%

VI. CONCLUSION

The covid-19 pandemic has affected every individual physically or mentally and in some situations in both ways. There is a huge amount of difference in a people pre and post pandemic. Physical as well as emotional well being is very essential. It is very essential to find out if a person has changed pre and post pandemic. The implemented system helps to identify and predict if a person is going through emotional turmoil or trauma. With the help of machine learning algorithms and questionnaire the predictions are made. Three algorithms are applied on the dataset and along with it questionnaire is used. The algorithms like Naïve Bayes Classifier, Random Forest and KNN are utilized. The results are obtained and from the results it is observed that Random Forest algorithms works efficiently with 82% accuracy rate, Naïve Bayes Classifier provides accuracy of 76% and KNN provides accuracy of 79%. This from the accuracy rate it can be seen that Random Forest works best on the given dataset.

References

[1] 1. Galea, S.; Merchant, R.M.; Lurie, N. The mental health consequences of COVID-19 and physical distancing: The need for prevention and early intervention. JAMA Intern. Med. 2020, 180, 817–818. [CrossRef]

[2] Balkhair, A.A. COVID-19 Pandemic: A New Chapter in the History of Infectious Diseases. Oman Med. J. 2020, 35, e123. [CrossRef]

[3] Wang, C.; Pan, R.; Wan, X.; Tan, Y.; Xu, L.; Ho, C.S.; Ho, R.C. Immediate 1155 Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int. J. Environ. Res. Public Health 2020, 17, 1729. [CrossRef] [PubMed]

[4] Mazza, C.; Ricci, E.; Biondi, S.; Colasanti, M.; Ferracuti, S.; Napoli, C.; Roma, P. A Nationwide Survey of Psychological Distress among Italian People during the COVID-19 Pandemic: Immediate Psychological Responses and Associated Factors. Int. J. Environ. Res. Public Health 2020, 17, 3165. [CrossRef]

[5] Adamson, M.M.; Phillips, A.; Seenivasan, S.; Martinez, J.; Grewal, H.; Kang, X.; Coetzee, J.; Luttenbacher, I.; Jester, A.; Harris, O.A.; et al. International prevalence and correlates of psychological stress during the global COVID-19 pandemic. Int. J. Environ. Res. Public Health 2020, 17, 9248. [CrossRef] [PubMed]

[6] Hopfenbeck, T. The Role and Value of International Datasets and Comparisons in Education Research. Res. Intell. 2012, 119, 7–8. 7. Polit, D.F.; Beck, C.T. Generalization in quantitative and qualitative research: Myths and strategies. Int. J. Nurs. Stud. 2010, 47, 1451–1458. [CrossRef]

[7] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. Health Information Science and Systems, 6(1). https://doi.org/10.1007/s13755-018-0046-0

[8] Uddin, M. Z., Dysthe, K. K., Følstad, A., & Brandtzaeg, P. B. (2021). Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Computing and Applications. https://doi.org/10.1007/s00521-021-06426-4

[9] Tolba Marwa, Ouadfel Salima, Meshoul Souham, "Deep learning for online harassment detection in tweets", 2018 3rd International Conference on Pattern Analysis and Intelligent Systems(PAIS).

[10] Monirah A. Al-Ajlan , Mourad Ykhlef, "Optimized Twitter Cyberbullying DetectionbasedonDeepLearning",201821stSaudiComputerSociety NationalComputerConference(NCC).

[11] Ms. Harshali Patil, Ms. Jesal Joshi, Ms. Deepali Joshi, Teaching Learning Methods' shift and its impact during pandemic, National Symposium on theme Enhancing quality and driving change in higher & technical education through NEP 2020, Journal of Engineering Education Transformations (JEET) a UGC Care journal, ISSN:2349-2473 and E-ISSN: 2394-1707, DOI: 10.16920/jeet/2021/v35i0/167016, Volume 35, Index in Scopus and EBSCO

[12] Deepali Joshi, "Impact of Covid-19 Pandemic on Emotional well being of people", IJERA, February 2022

[13] Deepali Joshi, "Prediction of Study Track using Decision Tree and Aptitude Test", IJCA 152(8):33-36, October 2016.

[14] Deepali Joshi, "Prediction of Study Track by Aptitude Test using Java", ISSN: 2231-1963, Volume 7, Issue 3, pp.1018-1026, July 2014.

[15] Deepali Joshi, "Prediction of Study Track using Decision Tree", IJERA, ISSN: 2248-9622, Vol. 4, Issue 5(Version 7), May 2014, pp.37-40

[16] Anuradha Khattar; Priti Rai Jain; S. M. K. Quadri, Effects of the Disastrous Pandemic COVID 19 on Learning Styles, Activities and Mental Health of Young Indian Students - A Machine Learning Approach, IEEE Xplore: 19 June 2020, Electronic ISBN:978-1-7281-4876-2, DVD ISBN:978-1-7281-4875-5

[17] Natasha Shaukat, Daniyal Mansoor Ali, Junaid Razzak, Physical and mental health impacts of COVID-19 on healthcare workers: a scoping review, International Journal of Emergency Medicine volume 13, Article number: 40 (2020)

[18] Johannes H. De Kock, Helen Ann Latham, Stephen J. Leslie, Mark Grindle, Sarah-Anne Munoz, Liz Ellis, Rob Polson, Christopher M. O'Malley, A rapid review of the impact of COVID-19 on the mental health of healthcare workers: implications for supporting psychological well-being, , BMC Public Health volume 21, Article number: 104 (2021)

[19] NorahAlballaaIsraAl-Turaikib, Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review, https://doi.org/10.1016/j.imu.2021.100564, Science Direct

[20] Cornelia Herbert, Alia El Bolock , Slim Abdennadher, How do you feel during the COVID-19 pandemic? A survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, and behaviour during the COVID-19 pandemic among university students, BMC Psychology volume 9, Article number: 90 (2021)

[21] Safwan Wshah, Christian Skalka, and Matthew Price, Predicting

Posttraumatic Stress Disorder Risk: A Machine Learning Approach, , JMIR Ment Health. 2019 Jul; 6(7): e13946. Published online 2019 Jul 22. doi: 10.2196/13946

[22] Jillian Ryan; Hamza Sellak; Emily Brindal, The Psychosocial Impacts of COVID-19 on a Sample of Australian Adults: Cross-sectional Survey and Sentiment Analysis, JMIR Publications, Vol 7, No 7 (2021): July