A picture containing text

Description automatically generated

FACULTY OF ENGINEERING & INFORMATICS

SCHOOL OF ENGINEERING

**MSC GROUP PROJECT (GROUP 3)**

**COS7048-B**

***PREDICTING PRELIMINARY SIGNS OF MENTAL DISORDER   
USING MACHINE LEARNING***

*Keywords: Machine Learning, Mental Disorder, Anxiety, Depression*

|  |  |  |
| --- | --- | --- |
| NAME | UoB NUMBER | COURSE |
| ANTHONIA ALIU | 21025728 | MSc IoT |
| ASAD MUNAWAR SHAH | 21025957 | MSc IoT |
| NILOY SIDDIQUEE KHANDOKER | 18015200 | MSc IoT |
| ASIM ZUBAIR | 21025150 | MSc Big Data |
| MAJID AL HUMAIDI | 21012650 | MSc Big Data |
| MICHAEL ABIEGBE | 20033150 | MSc Big Data |
| EDWARD TURKSON | 20023142 | MSc Big Data |

May 10, 2022

**Table of Contents**

[ABSTRACT 1](#_Toc103089248)

[1. INTRODUCTION 1](#_Toc103089249)

[1.1 Background 1](#_Toc103089250)

[1.2 Problem Statement 2](#_Toc103089251)

[2. LITERATURE REVIEW 2](#_Toc103089252)

[3. METHODOLOGY 3](#_Toc103089253)

[3.1 Aims & Objectives 3](#_Toc103089254)

[3.2 Achievements 3](#_Toc103089255)

[3.3 Dataset 4](#_Toc103089256)

[3.4 Prototype 4](#_Toc103089257)

[3.5 Requirements 5](#_Toc103089258)

[3.6 Assumptions and Risks 6](#_Toc103089259)

[3.7 Activity Plan 7](#_Toc103089260)

[3.8 Version Control Management 8](#_Toc103089261)

[4. EXPERIMENTS 9](#_Toc103089262)

[4.1 Web APPLICATION 9](#_Toc103089263)

[4.2 Machine Learning 11](#_Toc103089264)

[4.3 Using Python for Machine Learning Models 12](#_Toc103089265)

[4.5 Results 14](#_Toc103089266)

[4.6 Testing & Validation 15](#_Toc103089267)

[4.7 Computational Challenges 15](#_Toc103089268)

[5. LSEP ISSUES 16](#_Toc103089269)

[6. DISCUSSIONS 17](#_Toc103089270)

[7. CHALLENGES 18](#_Toc103089271)

[8. CONCLUSION & FUTURE OF WORK 18](#_Toc103089272)

[9. PEER EVALUATION 18](#_Toc103089273)

[10. REFERENCES 19](#_Toc103089274)

[11. APPENDICES 21](#_Toc103089275)

# ABSTRACT

Mental illnesses among employees cause productivity loss. It is also a huge burden on health expenditures by Government. Depression, anxiety, post-traumatic stress disorder (PTSD), and bipolar disorder are among the several sorts of mental health issues.

The use of machine learning models and big data applications in determining an initial state of mental health in the context of anxiety and depression at work has been examined and explored in this study. Professional research is consulted to align questions of our dataset with professional survey to improve assessment of anxiety and depression in employees. The research has produced a machine learning solution supported by a web Application for data collection from employees of an organization.

The solution has solved the complexity of privacy rules by introducing online web application. The overall data collected from an organization would be anonymous if not disclosed. However, the data collected will help to ascertain the overall mental health of all employees and an organization’s approach towards providing safe environment for employee's mental health. The paper will suggest improvement in mental health of employees. The paper has also provided a scientific as well as a humanistic reason for better addressing the often-overlooked problem of workplace mental health using machine learning approaches. Moreover, the solution will be helpful for the government to assess company's current state of preparedness towards employee's mental health.

# INTRODUCTION

## 1.1 Background

Machine learning has gained more popularity in every occupation in recent years. Globalization has connected people, and the data coming from this connectivity is in terabytes which is impossible for normal commuters and even advanced software to manage, process, and manipulate (W., 2017). The use of social media, mobile applications, and websites have provided researchers with real-time data to get information regarding user mental health and substance abuse (Smith, 2020).

Without the help of machine learning, it is impossible to manipulate the data and get the desired results. Various programs allow researchers to process this data and develop future trends and patterns among people (Panchal, Mukherjee, & Kumar, 2016). The need of the hour is to identify the ways to predict preliminary signs of mental disorder using machine learning.

Our project is unique as it focused adult employees of an organization. We have linked our research to the overall productivity and the government expenditures on mental health. Our solution is helpful for both the regulatory body as well as a private entity. There are several research in our literature review that highlighted the increased cost of the US government on mental health that is not only a burden on the government expenditures but is the major cause of low productivity. Our solution will help get an overview of the company current state of its responsibility towards employee's mental health.

## 1.2 Problem Statement

Machine Learning applications are more focused on diagnosis, treatment, and support, as well as research and clinical administration. Most of such studies concentrated on detection and diagnosis in children and young people. **However, prediction of the preliminary signs of mental disorders in adults at the workplace has not been actively explored with the use of machine learning**.

# LITERATURE REVIEW

The most phenomenal preliminary sign of mental disorder, according to psychologists, is occupational stress. Occupational stress alone is responsible for million-dollar losses to organizations throughout the world. The loss of employee productivity creates a burden on the organization and leads to employee demotivation and turnover (Chen L, 2015;). Stress affects the family and social life of people in the most harmful ways and may turn into depression if not detected at early stage.

The World Health Organization (WHO) showed in numerous reports the importance of mental health prevention in the workplace. Thus, many companies and scientific resources have started to develop attention to the psychological health of employees (Moreno, 2020) (A., 2020).

(Cai et al., 2018) achieved the accuracy of 79.27% when KNN classifier was used to detect depression. This study used the data collected from EEG reports. However, in (Abou-Warda et al., 2016), the same Random Forest classifier observed 92% accuracy on data from different mental illness patients was collected. Most of the patients were suffering from bipolar disorder, depression, schizophrenia, and obsessive-compulsive disorder. At the same time 97.2% was observed accuracy for depression detection when ANN (artificial neural network) was used in (Sau and Bhakta, 2017). In this study, a dataset of 105 patients was analysed.

In another study, using Facebook data, 73% accuracy was observed using Decision Tree. The study gave use comparatively reliable results for Random Forest model (Cacheda et al., 2019) to obtain textual features along with the log time gap and writing hours. The best accuracy was observed in a study of 348 people using Naïve Bayes of 85%. Again, in another study comprising Bangladeshi students (Rois et al., 2021), the accuracy for stress was found to be 89%. In the study of 268 respondents (Nguyen et al., 2019), satisfactory results were achieved comprising 95% confidence and a 6% margin of error.

# METHODOLOGY

The project implemented **Evolutionary Methodology** due to the complexity of the report. The reason we selected this methodology was because of the complexity foreseen in the beginning of the project. The step-by-step implementation of this model was chosen to allow for changing requirements as well as all work in broken down maintainable work chunks, prioritising them.

The project utilised existing data sets, professional survey, and list of symptoms contributing to the anxiety and depression. Before generalising our models, we have aligned our datasets with the most relevant questions for depression and anxiety. Some questions had been ignored due to their minor impact on depression and anxiety. So, the final dataset allowed us to find the best fit model. As we mentioned, the research focused on detecting depression and anxiety based on some questions asked to target people.

Furthermore, the data is divided into two sets: training data for building the model and test data for evaluating the model. Four machine learning algorithms are applied to create the model in this research, namely, **Logistic Regression**, **Random Forest Tree**, **Support Vector Machine (SVM)**, and **Decision Tree**. The machine learning algorithms were applied in Python programming language. This can identify an employee’s initial state of mental health.

## Aims & Objectives

The aim of this project is to achieve a machine learning approach in precisely predicting depression and anxiety of employees in the workplace by:

* Building an online Web Application for employee mental disorder assessment.
* Predicting, through Machine Learning (ML), the severity level of depression and anxiety of employees of a certain organization to formulate preliminary assessment for the same organization, health centers and Governments.
* Evaluating the shortcomings of the various approaches as well as methods to select the best fit ML model.

## Achievements

The proposed machine learning model with the help of structured questions was proposed to have predicted capacity with at least an **accuracy level of 70%**. But our final model except for the decision tree showed over 70%. Whereas the **highest accuracy of 96.8%** was recorded with logistic regression. Moreover, Web APPLICATION is developed, and most relevant questions are constructed to gather data from employees. Web APPLICATION is helpful in ensuring privacy by allowing features for either hiding identity or disclosing it.

## Dataset

We explored two datasets from Kaggle and UCI ML to produce the most relevant question. The project used the questions finalised with theoretical research for our main dataset. Please refer **“Computational Challenges”** for details. The Final Data set comprised 21 columns with 1000 records using the Depression and Anxiety Scale questionnaire. The data was explored using four machine learning algorithms – namely Logistic Regression, Support Vector Machine, Random Forest, and Decision Tree.

Table

Description automatically generated with low confidence  
*Fig 1: There are 21 attributes in the dataset from 20 questions and 1 output (type of disorder)*

## Prototype

As the project is unique that predicts the initial mental health of an employee in a controlled environment such as workplace. We had to find a structured questionnaire by a professional to help us not only to finalise the questions and modify or main data set. One of the data sets was broken into three unreadable formats. We exported these files into SQL server tables and then joined those tables to make data into readable format. The data was then exported to CSV for further use. Second data set from Kaggle had questions worded improperly which was a little offensive such as “Panic Attack”. So, we align our data sets with the professional questionnaire to produce a question for our Web Applications.

Graphical user interface, text

Description automatically generated  
*Fig 2: Web API*

We developed web application for those questions for the backend for the online survey. For the front-end, we have built a web application. Take the information from users and predict the mental condition by utilizing previous datasets.

Diagram

Description automatically generated  
*Fig 3: flow diagram for our solution*

## Requirements

***Functional***

|  |  |
| --- | --- |
| The online survey Web application | All question is mandatory to answer |
| Applying machine learning on the dataset | Predict the symptoms using machine learning |
| Testing |  |

***Non-Functional***

|  |  |
| --- | --- |
| Scalability | Capacity |
| Availability | Reliability |
| Manageability | Recoverability |

***Resources***

**Skills**

|  |  |
| --- | --- |
| Research | Antonia, Niloy, Asim |
| Data Analysis | Edward, Asim, Michael |
| Machine Learning | Asad, Majid, Michael |
| Application Development | Asad, Majid, Edward |
| Testing | Antonia, Niloy |

**Software**

|  |  |
| --- | --- |
| Python | Machine learning |
| GNIME | Machine learning |
| Visual Studio | Web APPLICATION |
| Python | Data pre-processing |
| MS Project | Project Management & Planning |
| Lime Survey | Survey Content Management System |
| GitHub | Repository Version Management |

**Data**  
Depression Dataset  
Anxiety Dataset

**Facilities**

|  |  |
| --- | --- |
| Microsoft Teams | Document Management System, Online meetings |
| WhatsApp | Communication |

**Time**  
Start Date: February 1, 2022  
End Date: May 10, 2022

## 3.6 Assumptions and Risks

***Risk Assessment Matrix***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Severity** | | | |
| **Likelihood** |  | 1 | 2 | 3 |
| 1 | Group Disagreements | Lack of Strategy and Experience | Unable to Achieve target |
|
| 2 | University Closure (Covid) | Biased data set | Unavailability of Lab Resources |
| 3 | Emergencies and Natural Disasters | Non-Disclosure of Information by subjects | Time loss in case of unavailability of team members |
|

Risks have been discussed in the challenges to this research. The research required more than available time which affected the initial plan and we had to change it in the middle of the project because of the unavailability of team members to gather in the University. We had to face time loss due to the Major religious events observed during our project timeline.

## Activity Plan

Graphical user interface, application

Description automatically generatedChart, line chart

Description automatically generated

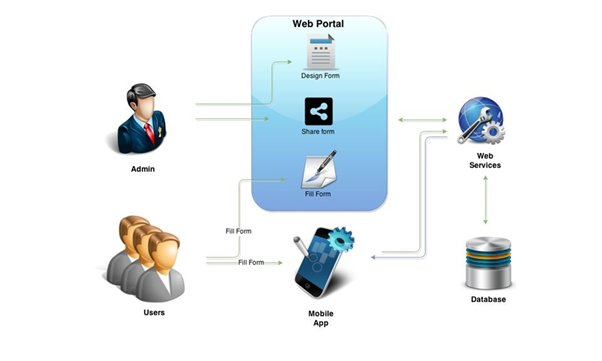
*Fig 4: Breakdown of the activities into five phases namely research work, data analysis, machine learning, application development and testing phases*

Graphical user interface, application, table, Excel

Description automatically generated

*Fig 5: Assigned activities to team members with their dependencies*

***Development Architecture***

  
*Fig 6: Overview of the survey system*

## 3.8 Version Control Management

***Microsoft Teams Document Management System***

Graphical user interface, text, application, email

Description automatically generated

*Fig 7: Use of MS Teams for collaborative work on project report.*

***GitHub Repository***

Graphical user interface, text, application, email

Description automatically generated  
*Fig 8: Use of GitHub repository for versioning control*

# EXPERIMENTS

## 4.1 Web APPLICATION

We decided to use Lime Survey for the questionnaire because it is an open-source statistical survey tool written in PHP based on MySQL database with several features and is supported on both Microsoft Windows and Linux operating systems.

Graphical user interface, text, application

Description automatically generated  
*Fig 9: The questionnaire is broken down into demographic (2 questions) and self-assessment (20 questions)*

Chart, waterfall chart

Description automatically generated  
*Fig 10: Statistical view of the responses from the questionnaire*

Graphical user interface, application, table, Excel

Description automatically generated  
*Fig 11: Preparing to export the survey responses to different formats csv format*

Graphical user interface, text, application

Description automatically generated  
*Fig 12: Exporting the survey responses to csv format for machine learning experiments*

Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated

*Fig 13: The collection of the demographics of the subjects will help analyse the most vulnerable groups likely to have depression or anxiety.*

Graphical user interface, text, application

Description automatically generated  
*Fig 14: Questionnaire welcome page*

Graphical user interface, application, Teams

Description automatically generated  
*Fig 15: Questionnaire first question*

## 4.2 Machine Learning

ML algorithms have been extensively used in mental health illnesses. For most predictive research, different ML algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), k-Nearest Neighbours (KNN), Random Forest (RF), Ensemble Methods (EM) and Instance Based Learning (IBL) were used (Wade et al., 2015, Abou-Warda et al., 2016, Nguyen et al., 2019).

The project selected four Machine Learning (ML) approaches and evaluated their accuracies in detecting mental health concerns using a variety of criteria. The four ML learning approaches are as follows Logistic Regression, SVM, Random Forest, and Decision Tree.

The performances of the four machine learning algorithms are completely different. These techniques have been compared, with the Logistic Regression technique obtaining the most accurate prediction of **96.8%.**

|  |  |
| --- | --- |
| Methods | Accuracy (%) |
| Decision Tree | 67.2 |
| Random Forest | 79.1 |
| Support Vector Machine | 90.8 |
| **Logistic Regression** | **96.8** |

*Comparison of methods*

## 4.3 Using Python for Machine Learning Models

Graphical user interface, text, application

Description automatically generated  
*Fig 16: As from the above figure we have loaded the packages needed for the experiments for machine learning.*

Calendar

Description automatically generated with low confidence  
*Fig 17: After all the packages and libraries, we have read the dataset from excel and read it as a data frame.*

Graphical user interface, text, application

Description automatically generated

*Fig 18: Loading more packages to enable the split of the data and build the models*

A picture containing calendar

Description automatically generated  
*Fig 19: Now we encoded the attributes from yes/no to binary (1/0) and then disorder (normal, depression, and anxiety to (0,1,2) respectively before creating the models.*

Graphical user interface, text, application

Description automatically generated

*Fig 20: Preparing the independent variable (Y) and dependent variables (X) before building models and splitting the data into test and train data*.

## Results

Graphical user interface, text, application, email

Description automatically generated  
*Fig 21: Now we applied Support Vector Machine (SVM) algorithm with train and test accuracies, and we got 90.9% accuracy.*

Graphical user interface, text, application, email

Description automatically generated

*Fig 22: Let us examine the Random Forest Classifier algorithm with train and test accuracies which is almost 79%.*

Graphical user interface, text, application, email

Description automatically generated

*Fig 23: Outcome of the Logistic Regression algorithm with train and test accuracies is about 96.8%*

Graphical user interface, text, application, email

Description automatically generated  
*Fig 24: Outcome of Decision Tree algorithm with train and test accuracies is about 67.23%.*

## Testing & Validation

Since logistic regression algorithm provided the best results, we have chosen it to do some tests on the obtained model using some random answers to the questions. The model results could be **0 for normal**, **1 for depression**, or **2 for anxiety**. We did validation by distributing the dataset into 35% for testing and 65% for training. The following are some examples for newly predicted mental conditions using machine learning.

Text

Description automatically generated  
*Fig 25: Testing and Validation with the model (logistic regression) with highest accuracy.*

## Computational Challenges

Collection of data was the one of the vital computational challenges of our project. Accuracy of prediction depends on the collection of data and its authenticity. Different surveys were consulted to construct questions in a proper manner that gives more accurate information about an employee. Additionally, questions had to be constructed in a way that ensures comfortability of the employee so that he/she is not to annoyed or offended. Therefore, we modified questions of our original data set and aligned them with the professional surveys. Our work in this regard was more theoretical. NHS literature on mental health specifically on anxiety and depression was consulted in detail to come up with the most important questions in our data set.

One of the data was converted to CSV with various codes. The data set was filled with textual answers. We tried to merge those questions into our initial data set, but it was a very time-consuming task that required NLP applications. Moreover, it was observed that it would have a little effect on the overall project. So, only important questions were consulted from that data set instead of using it.

Development of Web APPLICATION was another computational challenge to ensure privacy and ease of collection of data. Web APPLICATION was the requirement for this project as we were providing solutions to company that have some legal obligations to observe.

# LSEP ISSUES

***Professional and Social Issues***

**Data Privacy:** This means that people can choose how and what information is collected about them is conveyed to others. The personal data collected in our research contains only age group and gender.

**Coercion:** The extent to which the employee is coerced to contribute to the development of data analytics services is referred to as coercion. People are often forced or urged to use various networking apps, such as social networking sites. The web application gives private interface for each employee to fill the survey

***Legal Issues***

**Data Ownership and Protection:** Personal health and identifiable information are typically protected and kept private. Having the proper data protection is important for organisations to ensure that their operations are secure and profitable. It can help minimise the risk of data breaches and improve the legal responsibilities of their users. We have deployed our web application on a local server for data protection and safeguarding the information.

**Database Licensing:** Another major legal issue with data is database rights. There may be regulatory safeguards to prevent data misuse when performing data analytics on data related to mental health.

***Ethical Issues***

**Data Brokerage:** This is when a company obtains data from various sources, enhances, filters, examines it, and then distributes it to other companies. Data brokers can also directly acquire or process another company’s data to get better results. The project has not utilised any broker as per the sensitivity of the information collected.

# 

# DISCUSSIONS

Mental and behavioural health issues are common in young population. Most of the disorders related to mood and use of substances take place at the age of 20 to 30 years (Kessler et all, 2005). Workplace is normally overlooked in terms of addressing mental health issues at an initial stage. Tackling health issues in controlled environments such as workplace where people spend most of their time can lessen the burden on public health spendings. (Goetzel, et all, 2018). Goetzel et al have demonstrated that physical health payments by companies far exceed compared to the payment on mental health.

Mental disorders are the most burdensome and expensive illnesses in the United States, costing more than $200 billion per year, more than heart disease, stroke, cancer, and obesity combined (Roehrig, 2016).  Employees demonstrated low productivity with benign or modest depressive cycles (Dewa et al, 2011).  Therefore, it is inevitable to predict the mental health of individuals in a workplace to suggest measures to a company to shift resources to this area.

Numerous studies are confidently describing that machine learning algorithms can predict high degree of accuracy. From other research works, it is observed that previous studies used only a small group of participants. Additionally, the dataset they used contain small number of features while some of the studies used large number of features however, these studies contained small group of instances. We can conclude that those studies randomly selected predicted methods. They did not know, on that specific dataset, how other algorithms performed. The data set we had with 1000 instances is a sufficient for the Machine learning predictions.

The U.S. Surgeon General (1999) and the WHO (2001) pointed put stigma a vital hindrance in proper patient engagement. Stigma is a common predicament, but it repeatedly demonstrates in several forms. There are also varying ways in which it builds up in society, which all have consequences for social work at larger and smaller level. Privacy is another issue that is specially linked with employees and their future in the company. The Web APPLICATION is the unique idea which was not explored before. The interface is a particularly useful tool that provides a high accuracy in terms of comfortability of self-analysis. A private interface is not only useful to extract information in the beginning but also monitoring the progress of the subject post medical intervention.

The project selected four Machine Learning (ML) approaches and evaluated their accuracy in detecting mental health concerns using a variety of criteria. ML algorithms have been extensively used in mental health illnesses. For most predictive research, different ML algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), k-Nearest Neighbours (KNN), Random Forest (RF), Ensemble Methods (EM) and Instance Based Learning (IBL) are used (Wade et al., 2015, Abou-Warda et al., 2016, Nguyen et al., 2019). The research has achieved higher accuracy with Logistic Regression. The data was normalised to increase the performance of the model.

# CHALLENGES

Initially, we planned the project in a slightly different way where we decided to take the employee questions and provide prediction in a real-time basis. But due to the Ramadan, we must drop this idea as most of the group members were fasting. Moreover, the time limit and the scope of the course work. During holidays it was difficult to meet in person in labs that made us to change overall planned project in the middle.

# CONCLUSION & FUTURE OF WORK

Various algorithms were tried to assess the accuracy of the prediction, but the Logistic Regression provided the highest accuracy. Previously KNN provided 97 percent accuracy on a similar work on depression detection. However, our model provided 96% accuracy much higher than the proposed achievement. There is an enormous potential of applying ML techniques to predict mental health of employees.

However, it was observed that prediction of mental health of employees through Machine learning is not explored in length. The research is initial effort to motivate further exploration in this path. Such ML models can also assist physicians, psychologists, and psychiatrist to obtain initial state of mental health as well as provide self-assessment opportunity to patients to ascertain the improvement through medical interventions.

The research provides a food of thought for decision makers in governments to devise laws that can hold organizations responsible for ensuring working environment safe for mental health of employees. Companies can be categorised as safe, unsafe, or partially safe through the implementation of this Model. Further studies can be initiated to improve this model or use the data from this model to devise a 360 solution for health centres. Further studies can combine the facial expressions, voice, and body language with this model to optimise the accuracy of the responses and resultant predictions.

# PEER EVALUATION

Group members met through Teams, WhatsApp and in person. Edward managed the meetings and the construction of the report. Asim contributed towards the background research and conception of idea. Majid and Asad developed the technical part which was also reviewed by Asim and Edward. Anthonia helped in LESP issues and research. Niloy and Michael proofread the final report for errors and omissions and helped in other parts.

# REFERENCES

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):593–602. doi: 10.1001/archpsyc.62.6.593. [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/15939837)] [[CrossRef](https://dx.doi.org/10.1001%2Farchpsyc.62.6.593)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Arch+Gen+Psychiatry&title=Lifetime+prevalence+and+age-of-onset+distributions+of+DSM-IV+disorders+in+the+National+Comorbidity+Survey+Replication&author=RC+Kessler&author=P+Berglund&author=O+Demler&author=R+Jin&author=KR+Merikangas&volume=62&issue=6&publication_year=2005&pages=593-602&pmid=15939837&doi=10.1001/archpsyc.62.6.593&)]

Goetzel RZ, Ozminkowski RJ, Sederer LI, Mark TL. The business case for quality mental health services: why employers should care about the mental health and well-being of their employees. *J Occup Environ Med.* 2002;44(4):320–330. [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/11977418)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=J+Occup+Environ+Med&title=The+business+case+for+quality+mental+health+services:+why+employers+should+care+about+the+mental+health+and+well-being+of+their+employees&author=RZ+Goetzel&author=RJ+Ozminkowski&author=LI+Sederer&author=TL+Mark&volume=44&issue=4&publication_year=2002&pages=320-330&pmid=11977418&)]

Roehrig C. Mental disorders top the list of the most costly conditions in the United States: $201 billion. *Health Aff.* 2016:10–1377. [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/27193027)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Health+Aff&title=Mental+disorders+top+the+list+of+the+most+costly+conditions+in+the+United+States:+$201+billion&author=C+Roehrig&publication_year=2016&pages=10-1377&)]

Dewa CS, Thompson AH, Jacobs P. The association of treatment of depressive episodes and work productivity. *Can J Psychiatry.* 2011;56(12):743–750. [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/22152643)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Can+J+Psychiatry&title=The+association+of+treatment+of+depressive+episodes+and+work+productivity&author=CS+Dewa&author=AH+Thompson&author=P+Jacobs&volume=56&issue=12&publication_year=2011&pages=743-750&pmid=22152643&)]

Aldwin, C. (2020, February 11). Culture, Coping and Resilience to Stress. Retrieved 2020, from <https://www.researchgate.net/publication/241639325_Culture_Coping_and_Resilience_to_Stress> :

Chen L, H. P. (2015; ). Perceived workplace health support is associated with employee productivity. . Am J Heal Promot. , 29(3):139–146.

Chen, J., Li, J., Cao, B., Wang, F., Luo, L., & Xu, J. (2020). . Mediating effects of self-efficacy, coping, burnout, and social support between job stress and mental health among young Chinese nurses. J. Adv. Nurs, 76, 163–173.

Mensah, J., & Amponsah-Tawiah, K. (2016). Mitigating occupational stress: The role of psychological cApplicationtal. Workplace Behav. Health , 4, 189–203.

Oppong, S. (2019, ). Seth Oppong Overcoming obstacles to a truly global psychological theory, research, and praxis in Africa. . J. Psychol. Afr. , 29, 292–300.

Panchal, S., Mukherjee, S., & Kumar, U. (2016). Optimism in relation to wellbeing, resilience, and perceived stress. Int. J. Educ. Psychol, 5, 1–6.

Smith, N. (2020). An Application of the Dual Factor Model of Mental Health in Elementary School Students. Implications for Social Functioning and Psychopathology. .

W., A. D. (2017). Cross-cultural understanding of stress and coping mechanisms. . Int. J. Res. Eng. IT Soc. Sci. , 7, 75–80.

Wagner SL, K. C. (2016). Mental Health Interventions in the Workplace and Work Outcomes: A Best-Evidence Synthesis of Systematic Reviews. Int J Occup Environ , 601-607.

Yunus WMAWM, M. P. (2017 ). Systematic review of universal and targeted workplace interventions for depression. . Occup Env Med. , oemed-2017.

K. W. M. Fulford (2001) *‘What is (mental) disease?’: an open letter to Christopher Boorse.* Journal of Medical Ethics. Pages 80-85. Accessed 20 February 2022.

Moreno, A. (2020) *Occupational Stress and Employees Complete Mental Health: A Cross-Cultural Empirical Study.* NBCI. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7277686/#B1-ijerph-17-03629> Accessed 15 March 2022.

ABOU-WARDA, H., BELAL, N. A., EL-SONBATY, Y. & DARWISH, S. A random forest model for mental disorders diagnostic systems. International Conference on Advanced Intelligent Systems and Informatics, 2016. Springer, 670-680.

CACHEDA, F., FERNANDEZ, D., NOVOA, F. J. & CARNEIRO, V. 2019. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research,* 21**,** e12554.

CAI, H., HAN, J., CHEN, Y., SHA, X., WANG, Z., HU, B., YANG, J., FENG, L., DING, Z. & CHEN, Y. 2018. A pervasive approach to EEG-based depression detection. *Complexity,* 2018.

NGUYEN, M.-H., HO, M.-T., NGUYEN, Q.-Y. T. & VUONG, Q.-H. 2019. A dataset of students’ mental health and help-seeking behaviors in a multicultural environment. *Data,* 4**,** 124.

ROIS, R., RAY, M., RAHMAN, A. & ROY, S. K. 2021. Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. *Journal of Health, Population and Nutrition,* 40**,** 1-12.

SAU, A. & BHAKTA, I. 2017. Artificial neural network (ANN) model to predict depression among geriatric population at a slum in Kolkata, India. *Journal of clinical and diagnostic research: JCDR,* 11**,** VC01.

WADE, B. S., JOSHI, S. H., PIRNIA, T., LEAVER, A. M., WOODS, R. P., THOMPSON, P. M., ESPINOZA, R. & NARR, K. L. Random forest classification of depression status based on subcortical brain morphometry following electroconvulsive therapy. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015. IEEE, 92-96.

Talend (2020) Big Data and Privacy: What Companies Need to Know to Ensure Trusted Data, <https://www.talend.com/resources/big-data-privacy/>

Nielsen M. (2015) Who Owns Big Data? https://www.technologyreview.com/2015/01/05/169719/who-owns-big-data/ PostgresSQL(2020)PostgresSQL Database Limit, <https://link.springer.com/content/pdf/bbm%3A978-1-4302-0018-5%2F1.pdf>

Nathan N (2002) How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population

Someh, I., Davern, M., Breidbach, C. F., and Shanks, G. (2019). Ethical Issues in Big Data Analytics: A Stakeholder Perspective. Communications of the Association for Information Systems, 44, pppp. h

# APPENDICES

1. Teams Meeting (Annex-1, Annex-1(a), Annex-1(b))
2. WhatsApp (Annex-2, Annex-2(a))
3. Minutes of Meetings (Annex -3)
4. Web Application (Annex-4)
5. Machine Learning (Annex-5)
6. Dataset (Annex-6)