Predicting the Utilization of Mental Health Treatment with Various Machine Learning Algorithms

MEERA SHARMA
N/A
St. Lukes' School
New Canaan,
UNITED STATES OF AMERICA
meera.bluesnowflake16@gmail.com

SONOK MAHAPATRA
N/A
Westhill High School
Stamford,
UNITED STATES OF AMERICA
sonokmahapatra@gmail.com

ADEETHYIA SHANKAR
N/A
Brookfield High School
Brookfield,
UNITED STATES OF AMERICA
adeethyia@gmail.com

XIAODI WANG
Department of Mathematics
Western Connecticut State University
Danbury,
UNITED STATES OF AMERICA
wangx@wcsu.edu

Abstract— In 2017, about 792 million people (more than 10% of the global population) lived their lives with a mental disorder [24]– 78 million of which committed suicide because of it. In these unprecedented times of COVID-19, mental health challenges have been even further exacerbated as home environments have been proven to be major sources of the creation and worsening of poor mental health. Additionally, proper diagnosis and treatment for people with mental health disorders remains underdeveloped in modern-day's society due to the widely ever-present public stigma attached to caring about mental health. Recently there have been attempts in the data science world to predict if a person is suicidal (and other diagnostic approaches) yet all face major setbacks. To begin, big data has many ethical issues related to privacy and reusability without permission—especially in regards to using feeds from social media. Additionally, people diagnosed with specific mental health conditions may not actually seek treatment, so data may be incorrect. In this research, we address both of these problems by using anonymous datasets to predict the answer to a different question—whether or not people are seeking mental health treatment. We also use a large variety of machine learning and deep learning classifiers and predictive models to predict with a high accuracy rate through statistical analysis.

Keywords— Mental Health, Mental Health disorders, Big Data, Machine Learning Algorithms

Received: June 27, 2020. Revised: November 14, 2020. Accepted: December 5, 2020. Published: December 31, 2020.

1. Introduction

Mental health (the health of one's mind) is an integral component of the overall state of a person. The WHO formally defines an individual with a healthy mind as someone who "realizes [their] own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to [their] community" [19]. Someone with poor quality mental health is far more susceptible to mental disorders, either occasional or chronic, which negatively impact one's emotions, mood, and overall behavior.

According to a WHO led study, mental health issues cause an estimated $1 Trillion in productivity loss to the worldwide economy [21]. Here in the US, mental health disorders are among the most burdensome. The Center for Disease Control (CDC) states that nearly 20% of all adults in 2017, about 18.3% of the US population, reported the official presence of one or more mental illnesses and 71% reported one or more symptoms of stress. In addition, they claim "many people with mental health disorders also need care for other physical health conditions". The costs of treating people with both mental health disorders and other physical conditions are 2-3 times higher than those without co-occurring illnesses [20]. Mental health problems have startling social and economic costs that place a heavy burden in the workplace, including employee presentism, absenteeism, and disability days [5].

Furthermore, in these unprecedented times of COVID-19 caused lockdowns, mental health challenges have been even further exacerbated. In any emergency, it is common for individuals to be stressed and worried, however, specific to COVID-19, there is additional stress that comes from risks related to transmission, managing work and family, higher demands in work setting, reduced capacity to use social support etc. [12]

Different types of mental disorders are characterized by individuals exhibiting different combinations of one or more of the following: abnormal thoughts, perceptions, emotions, or behavior. According to the WHO [19], mental disorders include: depression, bipolar disorder, schizophrenia and other psychoses, dementia, and developmental disorders including autism. There are many effective ways to prevent and treat mental disorders and alleviate the suffering caused by them, however, critical to providing this type of care are both diagnosis and access to healthcare and social services.

Diagnosis and treatment of people with mental health disorders remains challenged partly due to the public stigma attached to mental health. People with mental illness are challenged by stereotypes and prejudice that result from misconceptions about mental illness while struggling with the symptoms and disabilities resulting from the illness[6]. People are reluctant to seek help especially early in the detection phase, which if avoided, could lead to much better outcomes down the road.

In recent times, artificial intelligence and big data have begun increasingly used in healthcare, especially in mental health. Data scientists can use data to improve public health strategy, medical research and even the medical care that is provided to patients. Data can be sourced from many different sources like providers (pharmacy and patient's history) and non-providers (cell phone and internet searches). The use of big data is only expected to grow in the coming years.

2. Data-Cleaning and Feature Extraction

The purely numerical dataset we used to perform Principal Component Analysis (PCA) and other machine learning models on originated from combining the responses from Open Sourcing Mental Illness' (OSMI's) mental health surveys from 2014 [37] and 2016 [38].

The raw arrays of qualitative data from these two years differed in the number of columns they had: While the response array from 2014 consisted of 1260 rows and 27 columns, the 2016 array was composed of 1467 responses across 63 columns (in which each column represented one question from each survey).

This inconsistency in categorical variables resulted in us taking the inner-joint union of all characteristics of each data set, which combined these two datasets entirely and reduced the supposed number of columns to 21. However, three of these columns were labeled with very similar questions and including them would've decreased the performability rates for every method we use. So we deleted them, resulting in our concatenated data having a final value of 18 categorical columns (which would eventually leave us with 17 features and 1 boolean output vector). This entire process automatically resulted in our final array containing only 2391 complete rows (compared to the initial 2727).

Finally, we converted any/all categorical data into numerical data. The list below depicts every category we converted:

❖ Age: {18,19,20, … 74} kept the same

❖ Gender: {Female, Male, NB} was converted to {1,2,3}.

➢ Everything other than cis male or cis female was categorized as "NB".

❖ Family History: {no, yes, Other} was converted to {1,2,3}

❖ Treatment: {no, yes} was converted to {0,1}.

➢ This would later become our binary output variable.

❖ Work Interference: {NA, Never, Often, Rarely, Sometimes} was converted to {1,2,3,4,5}

❖ Work Benefits: {I don't know, no, yes} was converted to {1,2,3}.

❖ Care Options: {Not sure, yes, no} was converted to {1,2,3}

❖ Wellness program: {Don't know, yes, no} was converted to {1,2,3}

❖ Seek help: {Don't know, yes, no} was converted to {1, 2, 3}

❖ Anonymity: {Yes, no, Don't know} was converted to {1, 2, 3}

❖ Leave: {Somewhat easy, Don't know, Yes, Somewhat difficult, Very easy, Neither easy nor difficult, Very difficult} was converted to {1, 2, 3, 4, 5, 6, 7}

❖ Mental Health Consequence: {Maybe, No, Yes} was converted to {1, 2, 3}

❖ Coworkers: {Yes, No, Some of them} was converted to {1, 2, 3}

❖ Supervisors: {Yes, No, Some of them} was converted to {1, 2, 3}

❖ Mental Health Interview: {Yes, No, Maybe} was converted to {1, 2, 3}

❖ Physical Health Interview: {Yes, No, Maybe} was converted to {1, 2, 3}

❖ Mental vs Physical health awareness: {Yes, No, Maybe} was converted to {1, 2, 3}

❖ Observable Consequence: {Yes, No} was converted to {1,2}

After we completely preprocessed or cleaned our initial data, we had to decide which column of this array would be set aside as our output vector. Our output vector needed to be boolean, so we chose to run our models to predict whether or not a person sought out treatment. As this was represented by the fourth column of our array, we used the MATLAB code below to separate the entire array into an IV and a DV:

load('MentalHealthArr.mat');

MentalHealthArr_Independent = MentalHealthArr(:, 4)

InputMatrix = horzcat(MentalHealthArr(:,1:3), MentalHealthArr(:,5:18));

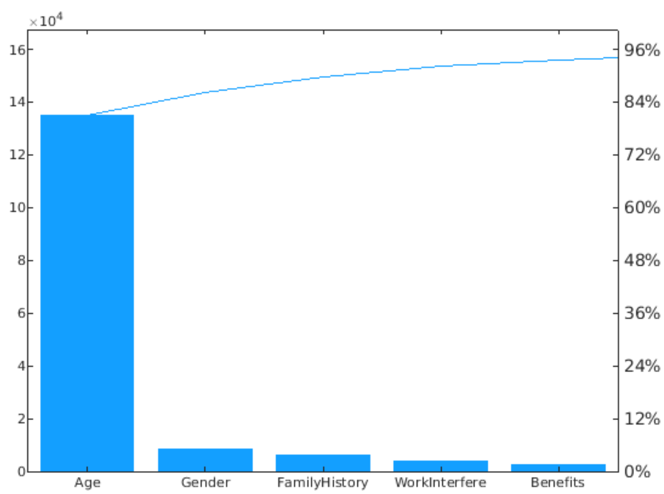
After performing this data cleaning, we used PCA to the cleaned data.

Although we do not use it as a predictive/classification model, PCA is a very efficient dimensionality-reduction algorithm and feature extraction tool, widely utilized across a variety of machine learning environments, which decreases the dataset's number of predictors while simultaneously retaining the highest percentage of the original dataset as it can. The algorithm does this by preserving the most influential variables (with the greatest percentage of the original variance) while removing the least influential ones (with the closest amount percentage of the original variance to 0). It does this by creating new uncorrelated variables that successively maximize variance.

To perform PCA, we calculate the eigenvalues of a Covariance matrix C in order to identify the Principal Components of the original dataset X. We then project the original dataset X onto the much lower dimensional space, called eigenspace, spanned by principle components of X to get a new dataset Y which has much lower dimension than that of X.

PCA helps in increasing interpretability but at the same time minimizing information loss from the dataset. By utilizing PCA on any dataset, the degree of variance specific to each independent variable can be determined. For this research specifically, we were able to use PCA to analyze each variable and determine which ones contribute the most to whether a person is receiving mental health treatment or not (while simultaneously determining which variables had negligible significance).

We found out that of all the independent variables that could contribute whether or not a person receives mental health treatment is correlated primarily by the age of the person.



3. Machine Learning Algorithms and Visual Representations

The following is a list and description of the machine learning algorithms we used to analyze the dataset.

3.1. Support Vector Machine Classification

Support Vector Machine (SVM) is a supervised learning models with associated learning algorithms which analyzes "training" data with f features (each corresponding with one binary output y where y is either 0 or 1) to create a model which constructs a hyperplane to separate different classes so that one can predict the correct output corresponding to each datapoint in a new "test" dataset.

If a dataset is not linearly separable, then we transform the dataset to a nicer and higher dimensional space where the transformed data set would be linearly separable. We applied the nonlinear hyperbolic tangent kernel function to our dataset as our data is mapped across 5 dimensions.

$$k(x, x') = \tanh(\alpha x^T x' + c)$$

3.2. Naive Bayes Classification

Naive Bayes is a simple but yet effective algorithm for imperative analysis that's characterized by the algorithm assumption of independence among predictors even while certain features depend on each other or upon the existence of the features of these properties.

The advantages of using Naive Bayes classification is that it is an easy and fast way to predict a dataset, especially when the assumption of independence holds. You need less training data. It performs well in instances where there are categorical input variables versus numerical variable(s).

The disadvantage of Naive Bayes is that if in test data, a categorical variable has a category that was not observed in training data, then the model will assign a 0 probability, thus making it not feasible to make a prediction. This is known as "Zero Frequency". This can be addressed by using a smoothing technique (e.g. Laplace estimation). Another disadvantage is the assumption of completely independent predictors, which is almost impossible in achieve real life.

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)}$$

3.3. k-Nearest Neighbors

K-nearest neighbors or KNN is used for classification and regression. It denotes class membership. An object is classified by assigning it a class that is most common among its k-nearest neighbors (k is a small positive integer). If k = 1, then the object is assigned to the class of that nearest neighbor.

We have focused on KNN for classification. A technique used by many is to assign weights to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the ones with larger distance. Neighbors are typically taken from a set of objects for which the class is known.

$$R_R(C_n^{wnn}) - R_R(C^{Bayes}) = (B_1 s_n^2 + B_1 t_n^2)\{1 + o(1)\}$$

3.4. Decision Tree

Decision Tree model is an information-mapping method with a tree-like structure commonly used for a variety of purposes in machine learning. There are multiple types of decision trees, but the one we used on our research is a Classification and Regression Tree, commonly abbreviated as CART, to conduct binary classification on our data.

Every CART utilizes an attribute-selection method called the Gini Index during the training process to calculate the probability that an attribute is classified incorrectly at a split (aka the split criterion).

$$Gini = \left(\frac{T}{D}\right) * g(T) + \left(\frac{F}{D}\right) * g(F)$$

3.5. Random Forest

Random Forest algorithm, as implied in its name, shares many similarities with Decision Tree mapping. However, they are not identical. A Random Forest consists of a multitude of Decision Trees built from randomly-selected observations and features and presents the total average of all of the results (from individual Decision Trees) as its final output.

The main benefit of using a Random Forest over a single Decision Tree is that it allows you to increase the accuracy of your results without overfitting your data. However, in a Decision Tree, the accuracy rate is directly proportional to the number of splits — As the number of splits in a tree increases, so does the accuracy. Therefore, this makes overfitting training data easily happening without our realizing, which would result in decreasing the accuracy of output-prediction on the test dataset.

3.6. Gradient Boosting Machines

While random forests build an ensemble of deep independent trees, Gradient Boosting Machines ("GBM") build an ensemble of shallow and weak successive trees with each tree learning and improving on the previous [13]. The main idea of boosting is to sequentially add new models to an ensemble. This way each new weak base learner model is trained with respect to the error of the whole ensemble learnt so far. This converts weak learners into stronger learners.

$$Loss(x) = sign(\sum_{m=1}^{M} \theta_m f_m(x))$$

3.7. Binary Logistic Regression

Binary logistic regression is a subset of logistic regression which analyzes a dataset (notated by the variable *x*) categorized by labels (features) and fits it onto a logistic curve in order to predict the accurate classification of data into two dichotomous outcomes (visually represented as 0 and 1 on the *y*–axis on a cartesian coordinate plane). Usually, these outcomes are opposing answers to the same question.

$$p = \frac{1}{1+e^{-(b_0+b_1 x)}}$$

4. Deep Learning

The brain of any mammal contains basic tissue. This tissue consists of a complex web of many individual cells called neurons (nerve cells). This web, officially named the Nervous system, allows action commands and sensory data to travel throughout the mammal's entire body. Artificial Neural networks (ANN) work in the same way. To summarize, an artificial neural network is a type of machine-learning model that is loosely referred to as an "artificial brain". This is due to the structural similarities that biological brain tissue and an artificial neural network share.

We delved into the vast field of deep learning and created a deep feedforward neural network with 100 hidden layers. While this didn't produce the most accurate results, the results produced by running the network on the test dataset have been very consistent and accuracy was the second-highest out of all of the methods used. The following is the structure of activation function used in ANN deep-learning model.

$$K\left(w_l^{(1)} x_1 + \cdots + w_{il}^{(1)} x_i + b_l^{(1)}\right) = K(\left\langle \vec{w}_l^{(1)}, \vec{x} \right\rangle + b_l)$$

5. Experimental Results

TABLE I.

Model Type	**Train Results**	**Test Results**					
	Accuracy	***Accuracy***	***Precision***	***Recall***	***F1 Score***	***Support***	***AUC***
SVM	*81.5%*	*82.5%*	*84.0%*	*82.0%*	*83.0%*	*718*	*0.83*
Naive Bayes	*79.4%*	*80.4%*	*81.0%*	*80.0%*	*80.0%*	*718*	*0.80*
KNN	*84.6%*	*79.1%*	*80.0%*	*79.0%*	*79.0%*	*718*	*0.79*
Decision Tree	*91.5%*	*76.3%*	*76.0%*	*76.0%*	*76.0%*	*718*	*0.76*
Random Forest	*91.6%*	*77.4%*	*78.0%*	*77.0%*	*78.0%*	*718*	*0.78*
GBM	*90.3%*	*80.4%*	*81.0%*	*80.0%*	*80.0%*	*718*	*0.80*
Logistic Regression	*78.4%*	*79.6%*	*79.0%*	*78.0%*	*79.0%*	*718*	*0.79*
Neural Network	*82.7%*	*80.8%*	*81.0%*	*80.0%*	*80.0%*	*718*	*0.81*

[a] In the table are the training accuracies and the test accuracies, precisions, recalls, F1 scores, supports, and AUCs.

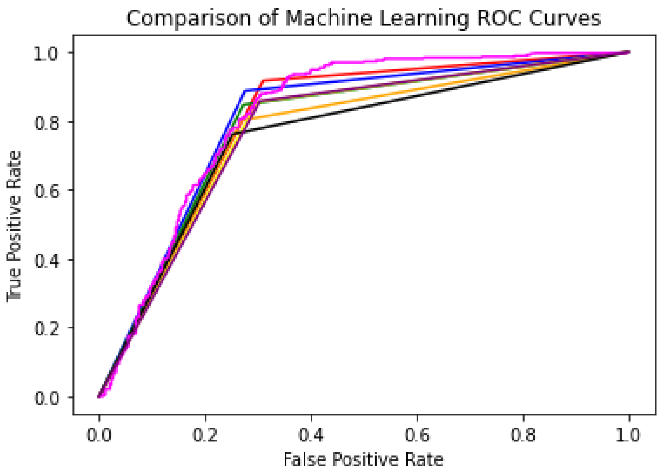


Fig. 1. The ROC curves of the machine learning algorithms used plotted for comparison.

Support Vector Machines model yielded the best overall results post training. It outperformed the other models in all key metrics (accuracy, precision, recall, FI score). Gradient Boosting Machines and Naive Bayes had the next best performances with very similar results across key metrics (differences are only at second decimal).

While decision tree based classifiers, like the CART and Random Forest, showed very promising results in the training set, the results did not generalize well in the test sample, likely due to the relatively smaller number of observations. Since the number of factors had already been reduced to five, in the Principal Component Analysis, we have already taken care to minimize chances of overfitting. With larger sample sizes, we would expect to see a closer performance between Random Forest and SVM/GBM. within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

6. Conclusions and and Further Research

From our research, we were able to conclude that machine learning can be used to predict likelihood of individuals seeking treatment with a high degree of accuracy (76.3% - 82.5%) by utilizing a self-reported questionnaire. Similarly, through a simple questionnaire that asks enough questions relevant to mental health, machine learning should also be able to determine if the person requires treatment. Despite stigma surrounding mental illness, individuals would be able to utilize machine learning to determine the correct course of action for their mental illness. As a result, these individuals would be more productive, reducing social and economic costs at the tech workplace.

We were also able to utilize Principal Component Analysis (PCA) to identify the five most important feature factors that predict the likelihood of a person receiving treatment. By utilizing PCA, the number of features (the columns of the dataset that contribute to whether a person is receiving mental health treatment or not), was reduced from 21 to the five that had the most effect on the prediction. These five features are the most important out of the 21, and should definitely be considered whenever discussing mental health. Also, these features (person's age, gender, family history, whether their mental illness interferes with their work, and work benefits) can be analyzed deeper with further research involving each feature individually, which would demonstrate the extent of the effect of each feature.

Utilizing Machine Learning on self-reported data avoids all the ethical issues related to privacy because employees are choosing to respond and provide permission to use the data. This is a novel approach that is able to avoid the challenges in current applications of Machine Learning in the field. We believe this can be a first step in the overall diagnosis and treatment of mental health issues.

To leverage our research, we propose sending a simple questionnaire to employees with key questions that are the factors we have identified. Since it will be a short questionnaire, it will likely have high response rates. Two of these factors (age, gender) do not need to be asked in a questionnaire from an employer if they can be tagged at the back end with the employee data. The other 3 factors, Family history, work interference, and work benefits are the only 3 questions that need to be answered. As a result, the time burden in answering this much simpler questionnaire can be as little as 2-3 minutes as opposed to 15-30 minutes in a typical questionnaire.

Our approach will be much faster to implement and with low burden on technology, allowing for a wide-spread adoption from employers. Simplicity and speed of administration also allows administration of the questionnaire in mobile format, which typically has significantly higher response rates vs. web-based designs.

We believe using Machine Learning in such a manner will have direct applications for employers, insurance companies and healthcare providers like Hospitals and doctors to allow for a very quick response and deliver diagnoses in a timely fashion as well as allow providers to plan for capacity and insurers and employers to identify and prevent mental health issues early.

Future research should seek to fill some of the shortcomings in the data by leveraging a larger, more recent sample across industries. Our research was based on a previously administered questionnaire and the data on it is both dated (from 2014 & 2016) and relatively limited (with only 21 features). The questionnaire was limited to technology workspace, hence some of the conclusions may not hold true for other industries. There may have been bias in the results due to the self-reporting nature of the survey, since the incidence of people seeking treatment seems very high.

Taking the PCA results further, future research can also be used to develop a nested questionnaire with 5-7 questions that both determine the likelihood of receiving mental health treatment as well as classifying what type of mental health treatment is needed. The questionnaire can dynamically change the next question based on the answer to the previous question. This can be done using the decision tree type of classification.

Mental illness is very common in society, but often, mental illness is not acknowledged as an illness like a physical illness would be. In the workplace, there may even be stigma associated with mental illness, invoking anxiety and loss of productivity in employees. However, machine learning is able to objectively decide whether a person should receive treatment based on numerous factors even without human interaction. Machine learning can play a pivotal role in determining whether certain people should receive treatment to improve their quality of life, which this research demonstrates.

Acknowledgment

We would like to enumerate the contributions of each team member.

Meera kept the team on task, led them through the entire process, and wrote an overwhelming majority of the final draft of the paper. Specifically, she wrote the entire body sections revolving around the following models: Naive Bayes, Decision Tree Classification, Random Forest, Binary Logistic Regression, Gradient Boosting Machines, and k-Nearest Neighbors. Outside the body, she wrote the entirety of the Introduction (Chapter 1) and ‘Conclusions and Future Research’ (Chapter 6). She created the table of numerical results and analysis in Chapter 5, and was also responsible for the completion of subsections 4.0 and 4.1. She worked with co-author Adeethyia to complete the Abstract and 2.1.1 on the ‘Basic Procedure of PCA’.

Sonok worked on coding all of the machine learning and deep learning models. Although unseen, he created the Pareto chart (graphical result of PCA) and recorded all mathematical results (AUC, ROC graphs, & confusion matrices) at the end of each model-specific section of the paper. He also assisted Meera and Adeethyia with the final formatting and edits of the submitted paper.

Adeethyia wrote foundational descriptions embedded with mathematical formulas for SVM and Neural Networks in the paper. He contributed to the primary draft, which can be seen in parts in the final draft of the paper.

We are very grateful for Hieu Nguyen, who taught us the fundamentals of Python coding through Google Colaboratory. Also, we would like to extend our special thanks to our advisor Dr. Xiaodi Wang who, along with the Western Connecticut State University, has provided us with hundreds of hours over Zoom meetings to teach us about Machine Learning and Deep Learning, as well as invaluable assistance throughout the process and many resources. We have learned many concepts in advanced mathematics and data science, as well as the invaluable skill of working with a group, in a very short amount of time.

References

[1] Axén, I., Björk Brämberg, E., Vaez, M. et al. Interventions for common mental disorders in the occupational health service: a systematic review with a narrative synthesis. Int Arch Occup Environ Health 93, 823–838 (2020), from https://doi.org/10.1007/s00420-020-01535-4

[2] Boehmke, Bradley. “Gradient Boosting Machines.” Gradient Boosting Machines · UC Business Analytics R Programming Guide, 14 June 2018, uc-r.github.io/gbm_regression.

[3] Breslow, Norman E., and Richard Holubkov. “Royal Statistical Society Publications.” Royal Statistical Society, 6 Jan. 2002, rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467 - 9868.00078.

[4] Bronshtein, Adi. “A Quick Introduction to K-Nearest Neighbors Algorithm.” Medium, Noteworthy - The Journal Blog, 6 May 2019, blog.usejournal.com / a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7.

[5] Colin G. Walsh, Jessica D. Ribeiro. “Predicting Risk of Suicide Attempts Over Time Through Machine Learning - Colin G. Walsh, Jessica D. Ribeiro, Joseph C. Franklin, 2017.” SAGE Journals, Apr. 2017, journals.sagepub.com/doi/abs/10.1177/2167702617691560.

[6] Corrigan, Patrick W, and Amy C Watson. “Understanding the Impact of Stigma on People with Mental Illness.” World Psychiatry : Official Journal of the World Psychiatric Association (WPA), Masson Italy, Feb. 2002, www.ncbi.nlm.nih.gov/pmc/articles/PMC1489832/.

[7] Dewa, C. S., Lesage, A., Goering, P., & Caveen, M. (2004). Nature and prevalence of mental illness in the workplace. Healthcare Papers, 5(2), 12-25, from https://doi.org/10.12927 /hcpap..16820

[8] El Karoui, Noureddine. “The Bootstrap, Covariance Matrices and PCA in Moderate and High-Dimensions.” Https://Www.stat.berkeley.edu/, 18 June 2016, www.stat.berkeley.edu/~nkaroui/files/BootstrapEigPaper Combined.pdf.

[9] Fisher, Aaron. “Fast, Exact Bootstrap Principal Component Analysis for p > 1 Million.” Taylor & Francis, Journal of the American Statistical Association, 18 Aug. 2016, amstat.tandfonline.com/doi/abs/10.1080/01621459.2015.1062383.

[10] Ha, Park. “An Introduction to Logistic Regression: from Basic Concepts to Interpretation with Particular Attention to Nursing Domain.” Journal of Korean Academy of Nursing, U.S. National Library of Medicine, Apr. 2013, pubmed.ncbi.nlm.nih.gov/23703593/.

[11] Hanisch, S.E., Twomey, C.D., Szeto, A.C.H. et al. The effectiveness of interventions targeting the stigma of mental illness at the workplace: a systematic review. BMC Psychiatry 16, 1 (2016), from https://doi.org/10.1186/s12888-015-0706-4

[12] IASC. IASC Interim Briefing Note ADDRESSING MENTAL HEALTH AND PSYCHOSOCIAL ASPECTS OF COVID-19 OUTBREAK Version 1.5 , Feb. 2020.

[13] Jaadi, Zakaria. “A Step by Step Explanation of Principal Component Analysis.” Built In, builtin.com/data-science/step-step-explanation-principal-component-analysis.

[14] Ji, Shaoxiong, et al. “Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications.” ArXiv.org, Cornell University, 6 Sept. 2020, arxiv.org/abs/1910.12611.

[15] Kamran Ul haq, Ayesha, et al. "Data Analytics in Mental Healthcare." Scientific Programming, Hindawi, 4 July 2020, ww.hindawi.com/journals/sp/2020/2024160/

[16] Lawrie, Stephen M, et al. "Predicting Major Mental Illness: Ethical and Practical Considerations." BJPsych Open, Cambridge University Press, 8 Mar. 2019, www.ncbi.nlm.nih.gov/pmc/articles/PMC6469234/.

[17] Linting, Mariëlle. "Stability of Nonlinear Principal Components Analysis: An Empirical Study Using the Balanced Bootstrap." Leiden University, PubMed, 2007.

[18] Marmot M (2017) Social justice, epidemiology and health inequalities. Eur J Epidemiol 32(7): 537–546, from https://doi.org/10.1007/s00420-020-01535-4

[19] "Mental Disorders." World Health Organization, World Health Organization, 28 Nov. 2019, www.who.int/en/news-room/fact-sheets/detail/mental-disorders.

[20] "Mental Health in the Workplace." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 10 Apr. 2019, www.cdc.gov/workplacehealthpromotion /tools-resources/workplace-health/mental-health/index.html.

[21] "Mental Health in the Workplace." World Health Organization, World Health Organization, 9 Aug. 2019, www.who.int/mental_health/in_the_workplace/en/.

[22] Passos , Ives C. "Identifying a Clinical Signature of Suicidality among Patients with Mood Disorders: A Pilot Study Using a Machine Learning Approach." Journal of Affective Disorders, U.S. National Library of Medicine, 15 Mar. 2016, pubmed.ncbi.nlm.nih.gov/26773901/.

[23] Park, H. (2013). An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. Journal of Korean Academy of Nursing, 43(2), 154. doi:10.4040/jkan.2013.43.2.154

[24] Ritchie, Hannah, and Max Roser. "Mental Health." Our World in Data, 20 Jan. 2018, ourworldindata.org/mental-health.

[25] Ritchie, Hannah, et al. "Suicide." Our World in Data, 15 June 2015, ourworldindata.org/suicide.

[26] Roy, Arunima, et al. "A Machine Learning Approach Predicts Future Risk to Suicidal Ideation from Social Media Data." Nature News, Nature Publishing Group, 26 May 2020, www.nature.com/articles/s41746-020-0287-6.

[27] Saxena, Shekhar, and Etienne Krug. World Health Organization, 2014, Preventing Suicide: a Global Imperative.

[28] Schott RL. Managers and Mental Health: Mental Illness and the Workplace. Public Personnel Management. 1999;28(2):161-183, from https://doi.org/10.1177/009102609902800201

[29] Shlens, Jonathon. "A Tutorial on Principal Component Analysis." Https://Arxiv.org, 10 Dec. 2005, arxiv.org/pdf/1404.1100.pdf.

[30] Singh, Harshdeep. "Understanding Gradient Boosting Machines." Medium, Towards Data Science, 4 Nov. 2018, towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab.

[31] Stone, Charles J. "Consistent Nonparametric Regression." Annals of Statistics, Institute of Mathematical Statistics, 2020, projecteuclid.org/euclid.aos/1176343886.

[32] "Suicide Prevention." National Institute of Mental Health, U.S. Department of Health and Human Services, www.nimh.nih.gov/health/topics/suicide-prevention/index.shtml.

[33] Tch, Andrew. "The Mostly Complete Chart of Neural Networks, Explained." Medium, Towards Data Science, 4 Aug. 2017, towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464.

[34] Theorell T, Hammarstrom A, Aronsson G, Bendz LT, Grape T, Hogstedt C, Marteinsdottir I, Skoog I, Hall C (2015) A systematic review including meta-analysis of work environment and depressive symptoms. BMC Public Health 15(1):738, from https://doi.org/10.1186/s12889-017-4153-7

[35] Torous , John, et al. "Smartphones, Sensors, and Machine Learning to Advance Real-Time Prediction and Interventions for Suicide Prevention: a Review of Current [28] Progress and Next Steps." Current Psychiatry Reports, U.S. National Library of Medicine, 28 June 2018, pubmed.ncbi.nlm.nih.gov/29956120/.

[36] Werth, Mark T., et al. "Analysis of Metabolomic PCA Data Using Tree Diagrams." Analytical Biochemistry, Academic Press, 21 Dec. 2009, www.sciencedirect.com/science/article/abs/pii/S0003269709008690?via=ihub.

[37] Mental Health in Tech Survey. (2014, November 3). Open Sourcing Mental Illness. Retrieved July 22, 2020, from https://www.kaggle.com/osmi/mental-health-in-tech-survey

[38] OSMI Mental Health in Tech Survey 2016. (2016, November 14). Open Sourcing Mental Illness. Retrieved July 22, 2020, from https://www.kaggle.com/osmi/mental-health-in-tech-2016

[39] Choi, Kenneth. Differentially Private M-Band Wavelet-Based Mechanisms in Machine Learning Environments. Aug. 2019.