

Using Data Science to Understand Mental Health in Business Environments

Waylon Abernathy, Brian Hogan, & David Madsen

Class: Applied Data Science, IST707

Professor Ying Lin, Ph.D

Syracuse University

June 13, 2019

Introduction

Mental illness encompasses many facets of society, family, personal and business health. According to the National Alliance on Mental Illness, “approximately one in five adults in the US, or 9.8 million people, experiences mental illness each year” (Clark & Woepel, 2019). This requires different approaches for understanding the array of mental health information and perhaps data science and analytics can help. Analytics may be useful to help policy planners, communities and businesses better understand mental health issues, their perceptions, and perhaps how to approach their existence in the workplace in a different manner.

Decisions about whether an individual will or will not disclose a mental health issue to an existing or potential employer are complex. Known or perceived stigmas as associated with mental health disorders and both employees and employers are uncomfortable approaching this topic. Specifically, supervisors may be in a gateway position to understand individuals with a condition while simultaneously being able to help a business environment be more supportive of mental health while encouraging both personal and business health (Kirsh, Krupa, & Luong).

Employees are going to keep their mental health issues to themselves if they don't believe their anonymity is being protected. Employees may experience a negative consequence in the workplace further guarding against disclosure for fear of losing a job or being categorized as not being able to fulfill a position's requirements successfully.

From a business perspective, the team is interested in providing data science analytics to investigate these perceptions and learn about attitudes towards disclosing mental health issues in the workplace. “Employer interest in understanding and supporting workplace mental health is increasing worldwide...as prevalence rates reveal common mental health and substance abuse disorders affect at least 1 in 5 workers” (Attridge, p.627, 2019). Perhaps data analytics can encourage employers to look at the data differently regarding employee mental health (MH)?

Data Overview

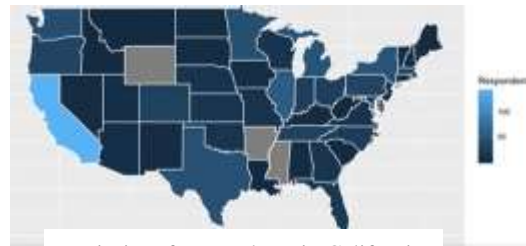
According to Mark Attridge (2019), “employer interest in understanding and supporting workplace mental health is increasing worldwide” (pg. 627). Available on Kaggle is a data set from the “Open Sourcing Mental Illness, LTD” (OSMI) initiated in 2016 and currently has over 1400 responses across 63 data categories and questions. The data set is “interested in gauging how mental health is viewed within the tech/IT workplace and the prevalence of certain mental health disorders within the tech industry” (Esther, 2016).

The OMSI data set provides an opportunity to explore this perception from an employee perspective and whether respondents felt comfortable disclosing mental health conditions to a current or potential employer. It also includes: percentage of time affected by a mental health disorder, whether resources made available by an employer to help with mental health conditions, and other characteristics such as age, gender, country, state, and survey respondent work status (work in person, remotely, or both). The data set is robust with 76% of the respondents answering all questions. The data includes numerous field types data analytics can use to investigate MH attitudes in workplace environments.

Data Preprocessing

Fifty-six fields were brought in and figure A. details the main category types:

Data Type Category	Data Values	Count
# Employees	numeric	1
Country	alph	2
Likert-Scale	nominal/ordinal	12
M/F/Transgender	binary	1
Age	numeric	1
Position	numeric	1
Primary Role	text / alpha	1
RemoteYN	binary	1
State	text / alpha	2
Y.N.Other	binary (yes/no)	34
Grand Total		56



Majority of respondents in California

Figure A.

Figure B.

Significant attention was paid to NA categories across numerous fields to maximize total available observations. This involved extensive programming around “self-employed,” and “text” fields transforming freeform data into discrete buckets. Location data included ‘country’ and ‘state’ (for those living or working in the United States).

Although many respondents were specifically from California (Figure B), feature extraction of ‘continent’ and ‘region’ were still possible to assist in adding significance to other areas. Of the model’s 70,217 available data points, 1,000s of cells were recaptured with the “Abernathy_Hogan_Madsen_Code_3” removing NAs. Other significant preprocessing included:

Field	Description	Result
Gender	Survey respondents included male, female and transgender persons. For all three categories respondents provide a variety of names and types requiring consolidation. Examples: non-binary, transitioned, mail, and “human.”	Male, female, transgender
Age	Age data was mostly clean and was put into bins for associative rule mining.	Twenties, thirties, forties, fifties, >50

Figure C.

Analysis

Random Forest Analysis:

Random Forest was run initially to build an understanding of significant variables to guide analysis work (figure C). Tree modeling revealed the overall significance of “*mental health consequence*” even though it was coded “yes, no and maybe.” This field guided the design for associative rule mining revealing age and gender, particularly “male”, as contributing to not disclosing MH issues. Descriptive statistics (next) indicated “*diagnosed*” was more important for prediction and helped focus analysis work between these key variables.

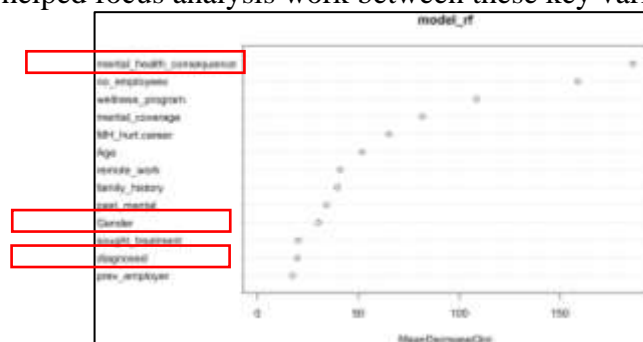


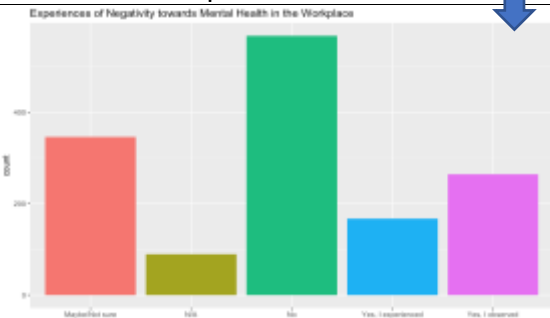
Figure C.

Descriptive Statistic Analysis:

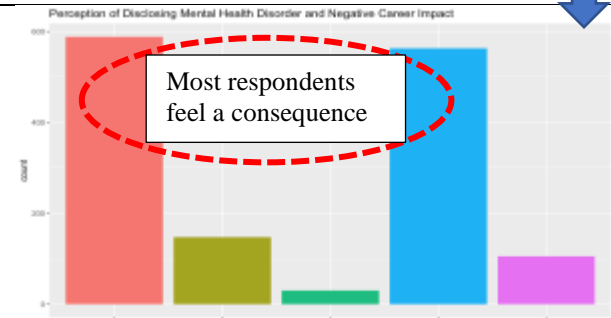
Descriptive statistics were important in assessing “our playing field” as data leans to not ‘disclosing’ MH (*) & most respondents either ‘have or may have’ a MH condition(**)

Experiences of Negativity in Workplace

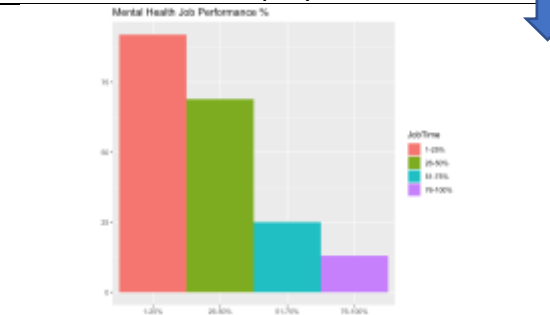
When combining three responses (maybe, yes/experienced, yes/observed vs. No) MH is cast in a negative light in the workplace. (maybe, yes/experienced/, and yes/observed) = 777 of respondents; No = 567 of respondents

**Fear of Discussing Mental Health due to Negative Career Impacts**

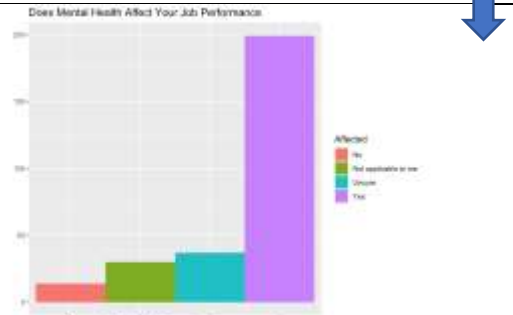
Displays heavy weight of mental health stigma within workplace

**Perceived Amount of Time Mental Health issue effects Job Performance**

Reveals impact of MH effects on performance; stresses the need for businesses to take proactive approach towards state of employee's mental health

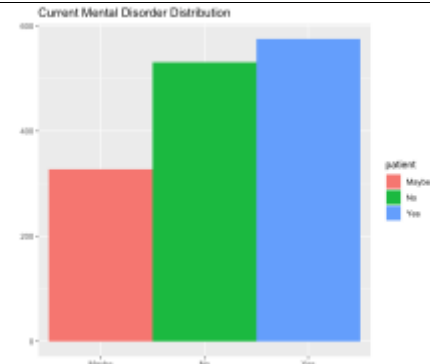
**Distribution of those who believe mental health affects job performance**

Overwhelming “yes”

**Current Mental Disorder Distribution**

Most respondents either ‘have or may have’ a MH condition

Figure D.



Associate Rule (AR) Analysis:

Respondents are going to keep MH issues to themselves if they don't believe their anonymity is protected. AR rules help indicate a business may witness a negative consequence in the workplace. The following shows how workplace negativity towards MH creates an environment that limits productivity

	lhs	rhs	support	confidence	lift	count
[1]	{anonymity=No, mental_health_consequence=Yes, Age=twenties}	=> {supervisor=No}	0.01116539	1.0000000	4.264881	16
[2]	{coworkers=No, anonymity=Yes, mental_health_consequence=Yes}	=> {supervisor=No}	0.01674808	0.9600000	4.094286	24

Figure E.

AR mining indicates not having wellness programs, believing MH disclosure would hurt their career, and keeping MH disclosure from coworkers validates a lack of MH communication will be present with a direct supervisor. This data confirms the supervisor's position and ability to both positively and negatively influence MH issues in a business environment.

```
> inspect(head(sort(SuperRules, by = "support", decreasing = T), 10))
```

	lhs	rhs	support	confidence	lift	count
[1]	{wellness_program=No, coworkers=No, mental_health_interview=No}	=> {supervisor=No}	0.1221214	0.7028112	2.997406	175
[2]	{coworkers=No, MH_hurt.career=Yes, I think it would}	=> {supervisor=No}	0.1172366	0.7058824	3.010504	168

Figure F.

Respondents don't trust MH anonymity in workplace and believe negative consequences result. The following find negative responses towards MH leaving respondents cautious of disclosure.

```
> inspect(SuperRulesMH) #Rule 1 = Anonymity=No, MH Consequence=Yes -> supervisor=No
```

	lhs	rhs	support	confidence	lift	count
[1]	{anonymity=No, mental_health_consequence=Yes}	=> {supervisor=No}	0.02023726	0.8055556	3.435599	29
[2]	{anonymity=No, Exp_neg_response=Yes, I observed}	=> {supervisor=No}	0.01046755	0.7142857	3.046344	15
[3]	{care_options=No, anonymity=No}	=> {supervisor=No}	0.01465457	0.7000000	2.985417	21

Figure G.

AR confirms age group "twenties" are least likely to report MH to a supervisor. This speaks to organizations mistrust which could have been learned from MH in a respondent's family history.

```
> inspect(head(sort(SuperRulesMH, by = "confidence", decreasing = T), 10))
```

	lhs	rhs	support	confidence	lift	count
[1]	{anonymity=No, mental_health_consequence=Yes, Age=twenties}	=> {supervisor=No}	0.01116539	1.0000000	4.264881	16
[2]	{coworkers=No, anonymity=Yes, mental_health_consequence=Yes}	=> {supervisor=No}	0.01674808	0.9600000	4.094286	24
[3]	{coworkers=No, anonymity=No, family_history=Yes}	=> {supervisor=No}	0.01465457	0.9545455	4.071023	21
[4]	{coworkers=No, anonymity=No, Age=twenties}	=> {supervisor=No}	0.01395673	0.9523810	4.061791	20

AR mining confirms males have most support in not wanting to disclose MH issues. (Figure H/I)

[29]	{supervisor=No} => {Gender=male}	0.1688765	0.7202381	0.9895505		
------	----------------------------------	-----------	-----------	-----------	--	--

Decision Tree (DT) Analysis:

MH “stigma” effect is nicely revealed as respondents feel “maybe” there would be negative consequence if disclosed. A general atmosphere created by coworkers, i.e. “*feeling comfortable discussing*,” (* in Figure J) seems to play a big decision for those not self-employed.

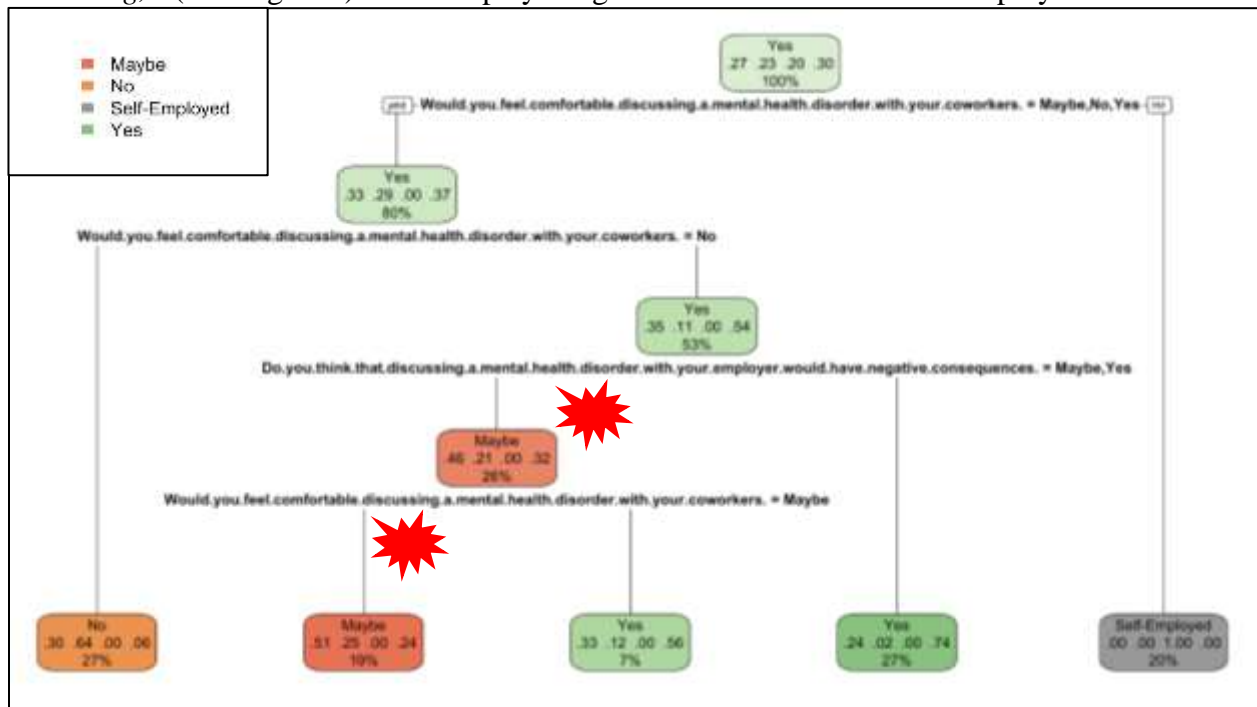


Figure J.

A wider decision tree blurred analysis results, potentially misleading an analyst on conclusions.

- i. Would you be willing to bring up a physical health issue with a potential employer in an interview?
 - a) Do you feel being identified as a person with a MH issue would hurt your career?
 - i) Have you had a mental health disorder in the past?
 - (1) Do you think that team members would view you more negatively if they knew you suffered from a MH illness?
 - ii) Would you be willing to bring up a physical health issue with potential employer in interview?
 - (1) Would you be willing to discuss a mental health issue with your supervisor?
 - (2) Would you feel comfortable discussing a MH disorder with employer?

- 41% of respondents did not think being identified as having a MH issue had affected or would affect their career. Of these only 15% expressed definite willingness to bring up a MH issue in an interview.

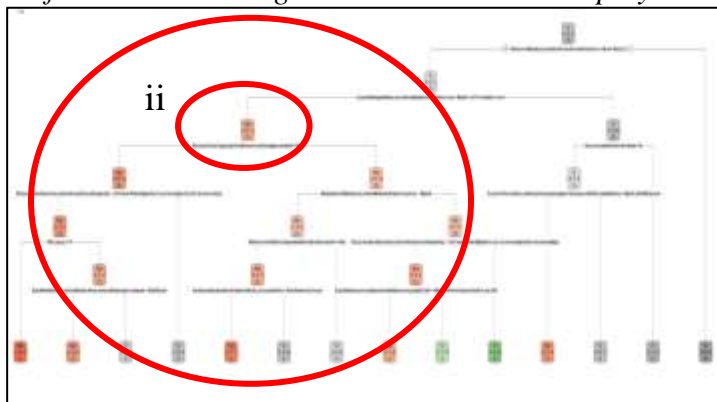


Fig. K.

Naïve Bayes Analysis:

What is the likelihood a respondent would or would not reveal a mental health issue with a potential employer during an interview? The model can calculate the conditional probabilities of the mutually exclusive questions learning the amount of joint probability in the event family. Naïve Bayes was helpful in this prediction resulting in a 70% accuracy score.

This algorithm makes sense as numerous questions centered around disclosure or no disclosure of MH, and the importance of features for this model agrees with those found in other models.

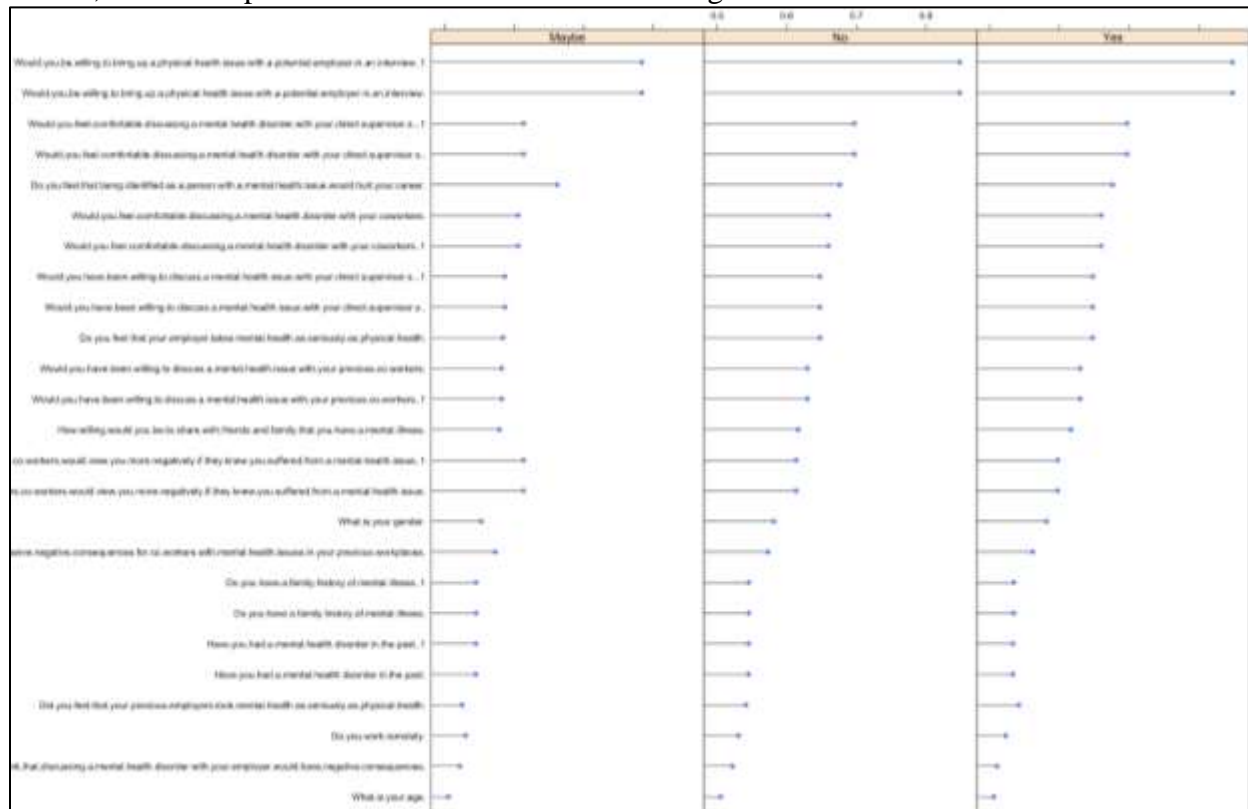


Figure L.

Model Performance

Decision Tree Summary:		
Would you feel comfortable discussing a mental health issue with a supervisor?		
Model	Parameters	Accuracy
Decision Tree (rpart)	Xval = 10, cp = 0.01, minsplit=7	73.76%
Random Forest (caret – rf)	Repeated Cv, repeats=10, number = 3, cp = seq(0, 0.1, 0.01)	69.5%
Boosted Logistic Regression (caret – LogitBoost)	Bootstrapping, number = 25, allowParallel = T, niter = seq(0, 10, 1)	74.26%
Would you bring up a mental health issues with a potential employer during an interview?		
Decision Tree (rpart)	Xval = 10, cp = 0.01, minsplit = 20	73.05%
Naïve Bayes	Cv, number = 3, fL = 1:3, useKernel = c(F,T), adjust = 1:3	70.92%

Association Rule Analysis Summary: (top three models)				
lhs	rhs	Support	Confidence	Lift
anonymity=no, MH.consequence=yes, age=twenties	supervisor=No	0.0112	1.0000	4.2648
coworkers=no, anonymity=yes, MH.consequence=yes	supervisor=No	0.0167	0.9600	4.0942
wellness.program=no, coworker=no, MH.consequence=no	supervisor=No	0.1221	0.7028	2.9974
coworker=no, MH.hurt.career='yes, I think it would'	supervisor=No	0.1172	0.7058	3.0105
anonymity=no, MH.consequence=yes	supervisor=No	0.2023	0.8055	3.4355
anonymity=no , exp.neg.reponse='yes,I observed'	supervisor=No	0.0105	0.7143	3.0463

Conclusion

The available data and data mining approaches suggest caution when considering revealing a MH issue either to peers or a supervisor. It is difficult to conclude companies are not receptive to employees discussing MH disorders, but companies do not seem to foster an environment where employees feel comfortable to speak openly about mental illness.

An untreated MH issue leads to decreased work performance and stigma. “Stigma has been recognized as a significant barrier to the full participation of people with mental illness in the workforce” (Kirsh, Krupa, Luong, p.547, 2018). Stigma has many forms, but any type decreases an individual’s work performance which impacts business performance. This exercise uncovered two forms of stigma in both associative rules (AR) and decision tree (DT). AR revealed significant support and confidence rules surrounding mental health consequences. DT found similar support as “maybe” there would be consequences discussing MH disorders with an employer. Analysis revealed discussing MH with supervisors either did not make sense (AR rules: supervisor=NO) or (DT: “maybe=yes”) in support across all age groups but specifically in males in the twenties age group.

Employers can benefit by stressing importance of getting MH help and endeavoring to remove stigma associated with MH issues. For this to occur a company must: a) make employees feel their career will not be adversely affected by coming forward about mental illness, and b) help to create a culture where employees talk openly with each other about MH without stigmatizing those who suffer. Supervisors are uniquely positioned to assist as an intermediary.

A supervisor is key in establishing the work environment and supervising their employee assets. A supervisor is also a bridge between the employee and the major decision makers within a company. “Workplace supervisors are well positioned to influence the employment pathways and success of workers with mental illness, by virtue of their involvement in establishing workplace culture and communications, allocating job duties, team development and other workplace functions” (Kirsch, et al. p. 548, 2918). Training may also be needed in many cases to help supervisors identify those individuals who may need assistance, to know how to approach these individuals, and to help foster an atmosphere of openness among their work groups without fear of repercussion for an employee.

Programs Written

Abernathy_Hogan_Madsen_Code_1.R	Corrected for gender & age
Abernathy_Hogan_Madsen_Code_2.R	Corrected for NAs in remaining data set
Abernathy_Hogan_Madsen_Code_3.rmd	randomForest, AR, descriptives, Decision tree, nB

References

- Attridge, M., 2019. A global perspective on promoting workplace mental health and the role of employee assistance programs. American Journal of Health Promotion. Retrieved from: https://journals-sagepub-com.libezproxy2.syr.edu/doi/full/10.1177/0890117119838101c?utm_source=summon&utm_medium=discovery-provider
- Clark, L., Woeppel, J. (2019). Importance of Analytics in Mental Health. Journal of AHIMA. Retrieved from: <https://journal.ahima.org/2019/02/01/february-2019/>.
- Kirch, B., Krupa, T., Luong, D. (2018). How do supervisors perceive and manage employee mental health issues in their workplaces? IOS. DOI 10.3233/WOR-182698
- Esther, K. Open Sourcing Mental Illness (OMSI). <https://osmihelp.org>. Retrieved from: https://www.kaggle.com/osmi/mental-health-in-tech-2016/version/1#mental-health-in-tech-2016_20161114.csv