REPORT

Mental Health Disorders

A great deal in the Tech World

Abstract

Several mental health conditions fall under the category of mental health disorders. For example, they may affect your mood, behavior, and thinking. However, many people may suffer from such health concerns, but it turns out to be a mental illness when the symptoms and signs affect their capability to function. Mental illness can be a cause of serious problems in everyday life at work as well. Thereby, it is supposed to be beneficial for people who have a mental illness to recognize it first and seek help. In addition to this, Tech companies are anecdotally notorious for endless and high-stress work environments (Mayoclinic, 2019).

Objectives

To explore and analyze the data provided in the survey for insights into Mental Health.

Attempt to build classifiers in order to predict whether an employee:

- will be willing to raise Mental Health (MH) issues with future employers
- will seek professional diagnosis for Mental Health Disorders (MHD)
- will seek treatment for Mental Health Disorders (MHD)

Description of Dataset

Problem of Interest

In this paper, research was conducted to find out the prevalent attitudes regarding Mental Health (MH) in the tech workplace. Moreover, company features which are most critical in

inclining employees to recognize their Mental Health Disorders (MHD) and seek help were also identified.

Data Source

The OSMI Mental Health in Tech Survey 2016 dataset was taken from Kaggle and it currently contains over 1400 responses, from an ongoing 2016 survey which aims to measure attitudes towards mental health in the tech workplace, and examine the frequency of mental health disorders among tech worker (Mental Health in Tech, 2016**)**

Detailed Explanation of Dataset

Data Description

As the dataset is based on a survey therefore, there were a total of 63 questions which were asked in the survey. The column headings in the dataset represent the questions while all the rows contain the respective response of an individual. Thus, the dataset had 1433 rows and 63 columns.

Data Issues

This particular dataset contained lots of missing values. The dataset was quite noisy for analysis and questions like gender had inconsistent answers as they were left as an open question. Much transformation was needed in the dataset as most of the answers were categorical while the machine learning algorithm takes in numerical data.

The Figure below shows the following questions which were inquired from the respondent in the survey.

SURVEY QUESTIONS
How many employees does your company or organization have?
Is your employer primarily a tech company/organization?
Is your primary role within your company related to tech/IT?
Does your employer provide mental health benefits as part of healthcare coverage?
Do you know the options for mental health care available under your employer-provided coverage?
Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?
Does your employer offer resources to learn more about mental health concerns and options for seeking help?
Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?
If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:
Do you think that discussing a mental health disorder with your employer would have negative consequences?
Do you think that discussing a physical health issue with your employer would have negative consequences?
Would you feel comfortable discussing a mental health disorder with your coworkers?
Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?
Do you feel that your employer takes mental health as seriously as physical health?
Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?
Do you have medical coverage (private insurance or state-provided) which includes treatment of Â mental health issues?
Do you know local or online resources to seek help for a mental health disorder?
If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to clients or business contacts?
If you have revealed a mental health issue to a client or business contact, do you believe this has impacted you negatively?
If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to coworkers or employees?
If you have revealed a mental health issue to a coworker or employee, do you believe this has impacted you negatively?
Do you believe your productivity is ever affected by a mental health issue?
If yes, what percentage of your work time (time performing primary or secondary job functions) is affected by a mental health issue?
Do you have previous employers?
Have your previous employers provided mental health benefits?
Were you aware of the options for mental health care provided by your previous employers?
Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?
Did your previous employers provide resources to learn more about mental health issues and how to seek help?
Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?
Do you think that discussing a mental health disorder with previous employers would have negative consequences?
Do you think that discussing a physical health issue with previous employers would have negative consequences?
Would you have been willing to discuss a mental health issue with your previous co-workers?
Would you have been willing to discuss a mental health issue with your direct supervisor(s)?
Did you feel that your previous employers took mental health as seriously as physical health?
Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?
Would you be willing to bring up a physical health issue with a potential employer in an interview?
Why or why not?
Would you bring up a mental health issue with a potential employer in an interview?
Why or why not?
Do you feel that being identified as a person with a mental health issue would hurt your career?
Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?
How willing would you be to share with friends and family that you have a mental illness?
Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?
Have your observations of how another individual who discussed a mental health disorder made you less likely to reveal a mental health issue yourself in your current workplace?
Do you have a family history of mental illness?
Have you had a mental health disorder in the past?
Do you currently have a mental health disorder?
If yes, what condition(s) have you been diagnosed with?
If maybe, what condition(s) do you believe you have?
Have you been diagnosed with a mental health condition by a medical professional?
If so, what condition(s) were you diagnosed with?
Have you ever sought treatment for a mental health issue from a mental health professional?
If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?
If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?
What is your age?
What is your gender?
What country do you live in?
What US state or territory do you live in?
What country do you work in?
What US state or territory do you work in?
Which of the following best describes your work position?
Do you work remotely?

Basic Statistics of Data

Data Types

The dataset had different data types for different columns. The following table shows the proportion of data types for all columns:

Data Type	Count
Object	56
Int64	4
Float64	3

Statistics of Numerical Data

The following table shows statistics of Numerical Data:

	Are you self-employed?	Is your employer primarily a tech company/organization?	Is your primary role within your company related to tech/IT?	Do you have medical coverage (private insurance or state-provided) which includes treatment of mental health issues?	Do you have previous employers?	Have you ever sought treatment for a mental health issue from a mental health professional?	What is your age?
count	1433.00	1146.00	263.00	287.00	1433.00	1433.00	1433.00
mean	0.20	0.77	0.94	0.64	0.88	0.59	34.29
std	0.40	0.42	0.23	0.48	0.32	0.49	11.29
min	0.00	0.00	0.00	0.00	0.00	0.00	3.00
25%	0.00	1.00	1.00	0.00	1.00	0.00	28.00
50%	0.00	1.00	1.00	1.00	1.00	1.00	33.00
75%	0.00	1.00	1.00	1.00	1.00	1.00	39.00
max	1.00	1.00	1.00	1.00	1.00	1.00	323.00

Data Preprocessing

Due to inconsistent, noisy and missing values, it becomes necessary to clean and preprocess our data before visualizing and gaining insight from it.

Data Cleaning

Handling Missing Values

The missing values are filled with -1.

Handling Noisy Values

The age attribute contains outliers which are replaced by the mode of age which is 32.

Data Reduction

The dataset has been reduced to 1146 rows by removing all those who are self-employed as we are concerned with those employed in some Tech Company.

From the remaining dataset irrelevant and empty columns are further dropped.

All the respondents are dropped who are neither in a tech company nor in a tech role.

Data Transformation

The categorical values are transformed into numerical data and columns are renamed to short meaningful keywords.

After complete Data Preprocessing the dataset is left with 1131 rows and 43 columns. The following table lists all 43 questions along with their respective renaming:

	QUESTION	RENAMED
0	How many employees does your company or organization have?	num_employees
1	Is your employer primarily a tech company/organization?	tech_company_or_role
2	Does your employer provide mental health benefits as part of healthcare coverage?	cep_benefits
3	Do you know the options for mental health care available under your employer-provided coverage?	cep_know_options
4	Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?	cep_discuss
5	Does your employer offer resources to learn more about mental health concerns and options for seeking help?	cep_learn
6	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?	cep_anon
7	If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:	cep_mh_leave
8	Do you think that discussing a mental health disorder with your employer would have negative consequences?	cep_mh_ncsq
9	Do you think that discussing a physical health issue with your employer would have negative consequences?	cep_ph_ncsq
10	Would you feel comfortable discussing a mental health disorder with your coworkers?	cep_comf_cw
11	Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?	cep_comf_sup
12	Do you feel that your employer takes mental health as seriously as physical health?	cep_serious
13	Have you heard of or observed negative consequences for co-workers who have been open about mental health issues in your workplace?	cep_others_ncsq
14	Do you have previous employers?	pep_have
15	Have your previous employers provided mental health benefits?	pep_benefits
16	Were you aware of the options for mental health care provided by your previous employers?	pep_know_options
17	Did your previous employers ever formally discuss mental health (as part of a wellness campaign or other official communication)?	pep_discuss
18	Did your previous employers provide resources to learn more about mental health issues and how to seek help?	pep_learn
19	Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers?	pep_anon
20	Do you think that discussing a mental health disorder with previous employers would have negative consequences?	pep_mh_ncsq
21	Do you think that discussing a physical health issue with previous employers would have negative consequences?	pep_ph_ncsq
22	Would you have been willing to discuss a mental health issue with your previous co-workers?	pep_comf_cw
23	Would you have been willing to discuss a mental health issue with your direct supervisor(s)?	pep_comf_sup
24	Did you feel that your previous employers took mental health as seriously as physical health?	pep_serious
25	Did you hear of or observe negative consequences for co-workers with mental health issues in your previous workplaces?	pep_others_ncsq
26	Would you be willing to bring up a physical health issue with a potential employer in an interview?	fep_ph_willing
27	Would you bring up a mental health issue with a potential employer in an interview?	fep_mh_willing
28	Do you feel that being identified as a person with a mental health issue would hurt your career?	hurt_career
29	Do you think that team members/co-workers would view you more negatively if they knew you suffered from a mental health issue?	cw_view_neg
30	How willing would you be to share with friends and family that you have a mental illness?	comf_ff
31	Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?	neg_response
32	Do you have a family history of mental illness?	mh_fam_hist
33	Have you had a mental health disorder in the past?	mh_hist
34	Do you currently have a mental health disorder?	mh_cur
35	Have you been diagnosed with a mental health condition by a medical professional?	mh_diag_pro
36	Have you ever sought treatment for a mental health issue from a mental health professional?	sought_treat
37	If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?	work_affect_effect
38	If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?	work_affect_ineffect
39	What is your age?	age
40	What is your gender?	gender
41	What country do you work in?	work_country
42	Do you work remotely?	work_remote

Data Visualization

Histogram

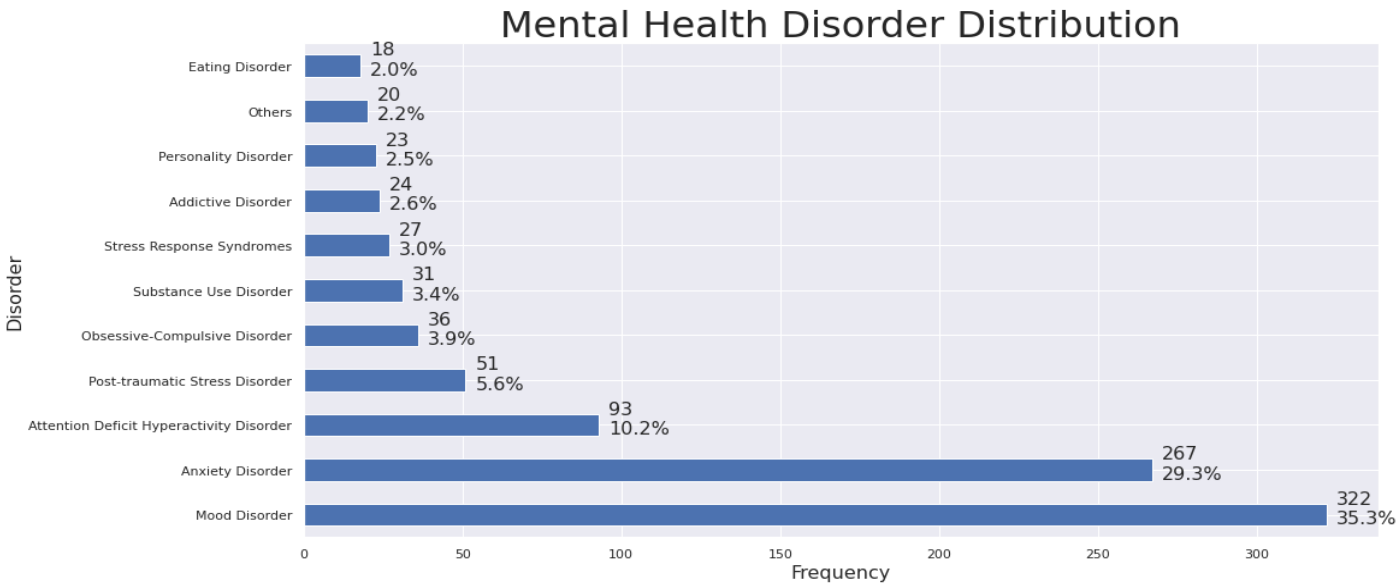
A Mental Health Status across Ages and Genders

- The largest proportion of genders is held by males, followed by females and then others.
- The proportion of males who do not currently have MHD is larger than the proportion of males who do have MHD.
- This is different from that observed among females. The proportion of females who currently have MHD is larger than the proportion of females who do not have MHD.
- For others, most have MHD while none do not currently have MHD.

Some graphs have been plotted for understanding mental health with gender

Bar Graph and Pie Charts

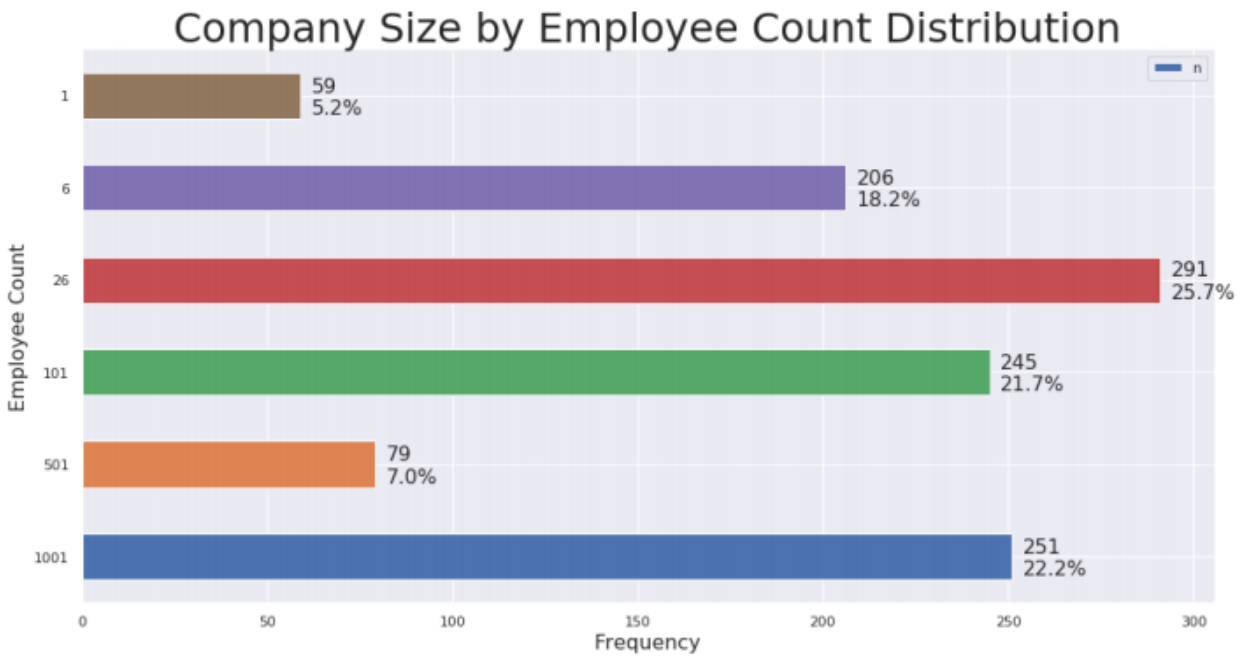
Mental Health Disorder Distribution



- Mood Disorders form the largest proportion of MHDs among the surveyed, at 35.3% or 322 out of 912 instances of MHDs.
- Mood Disorder causes one's general emotional state or mood to be distorted or inconsistent with circumstances, interfering with one's ability to function.
- Anxiety Disorder is 2nd at 29.3%.
- Anxiety Disorder also affects one's mood and often occurs alongside depression, which falls under Mood Disorders.
- This may be why Anxiety Disorder is the 2nd most frequent MHD among the surveyed, slightly less than the frequency of Mood Disorders but much more than the 3rd most frequent disorder.
- ADHD is 3rd at 10.2%.
- A myriad of other disorders follows at around 2.0-5.6% in frequency, or less than 56 persons among 1000.
- Given that the most frequent disorders are Mood and Anxiety, which appear to largely affect the mood, it may not be a surprise that employers find it hard to accept or treat employees with these MHDs kindly.

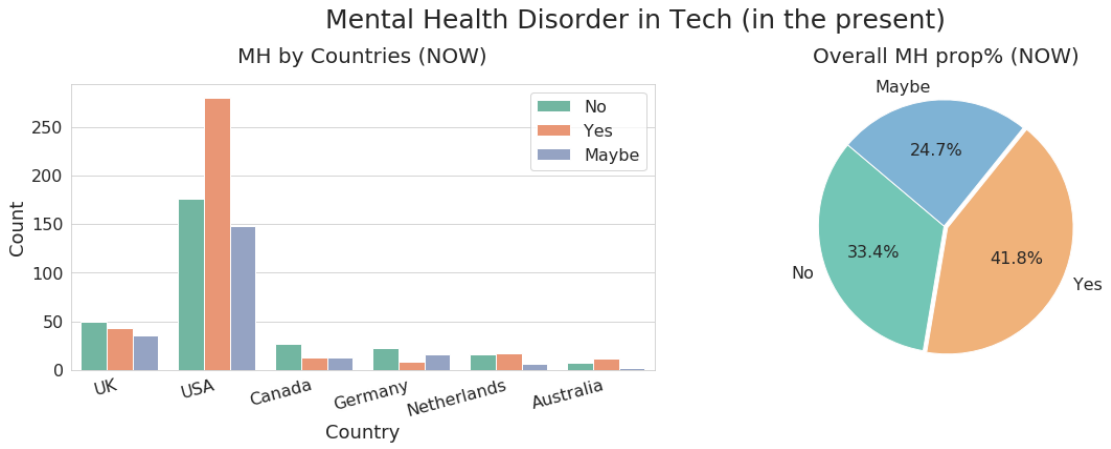
- That is, an employer may perceive an employee to simply be moody than actually suffering from a disorder that requires medical attention.
- Disorders like substance-abuse, PTSD, addictive might be more likely to raise medical concern, but these appear infrequently

Mental Health Status and Company Size



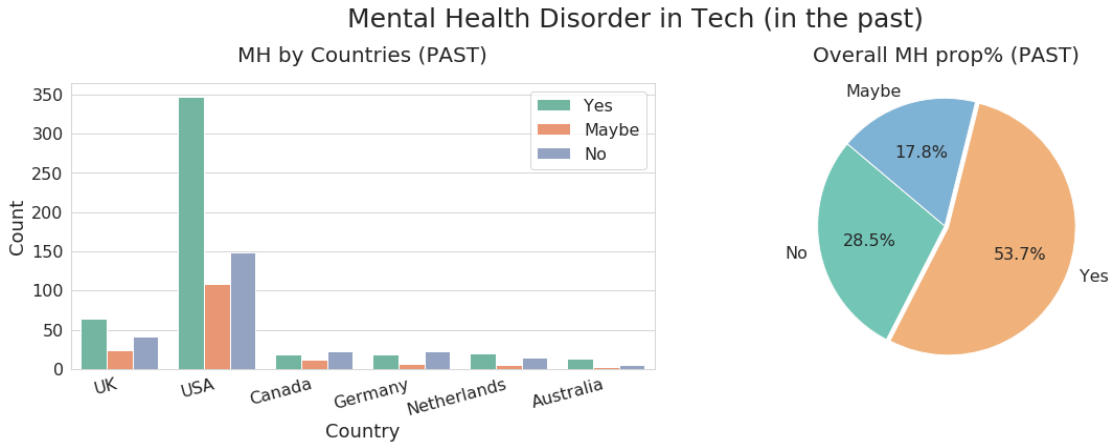
- The varying proportions across company size is largely due to the distribution of company sizes.
- For companies with 26-100 employees, there is a noticeably larger proportion of employees with current MHD than employees without.
- Furthermore, there is relatively less uncertainty (mh_cur = 2) for these companies.
- See how the proportion of people who are unsure is comparable across companies of sizes 6-25, 26-100, and 101-500, although the frequency of these companies differs.
- Similarly for companies with 101-500 employees, although the increment is less obvious.
- There are noticeably more people who do not have MHD in companies of size 501-1000 than people who do.

Mental Heath Disorder in Tech (in the present)



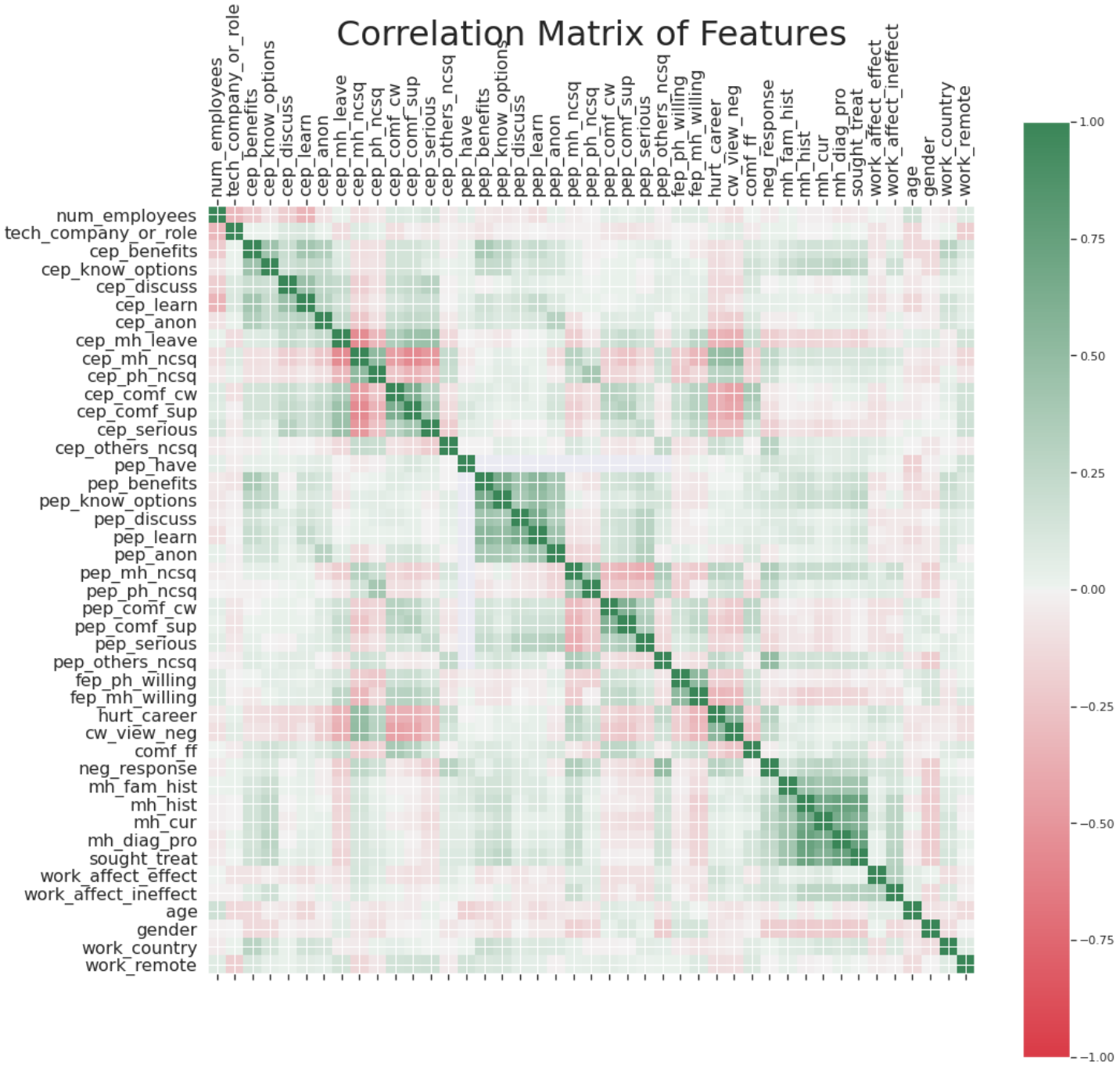
The pie chart shows that 41% of the respondents suffer at the moment from a mental health disorder.

Mental Health Disorder in Tech (in the past)



The pie chart shows that more than half of the respondents had a mental health disorder in the past.

Correlation Using Heat Map

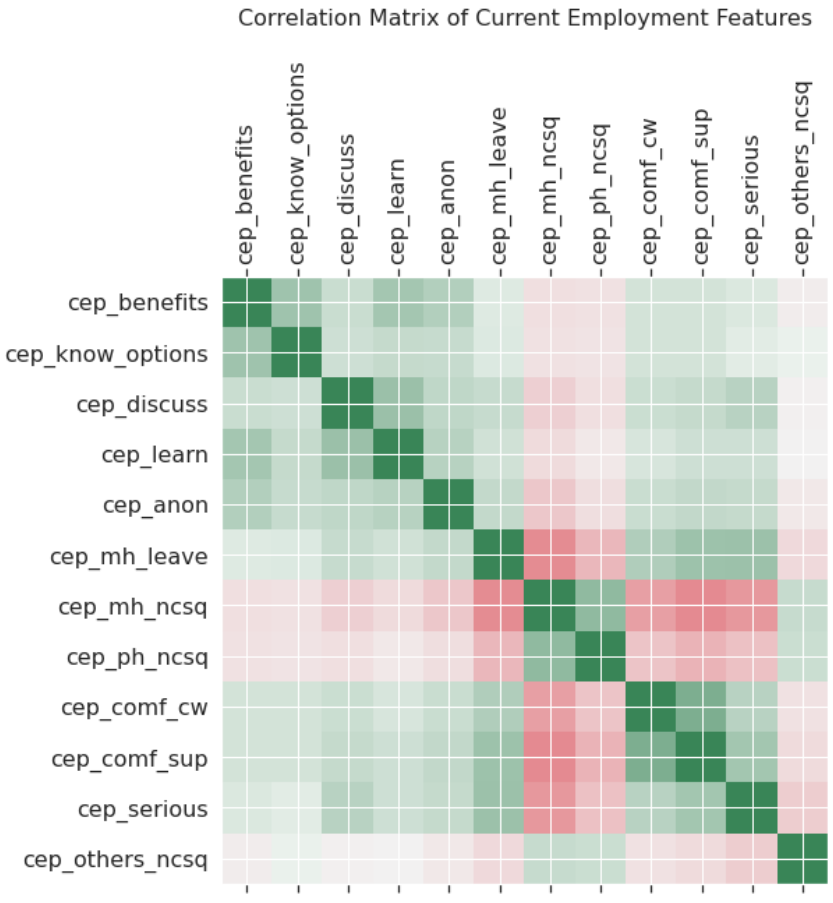


As the above figure shows, there are several zones in the above heat map with strong correlation, namely:

1. **Top Left:** It comprises features regarding current employment.
2. **Middle:** It comprises features regarding past employment.

3. **Middle, Bottom Right:** It comprises features regarding future employment and views of others.
4. **Bottom Right:** It comprises features regarding mental health history, diagnosis, treatment, and effect upon work.
5. **Bottom Left (and Top Right):** It comprises features regarding sentiment towards mental health and comfort in sharing about it.

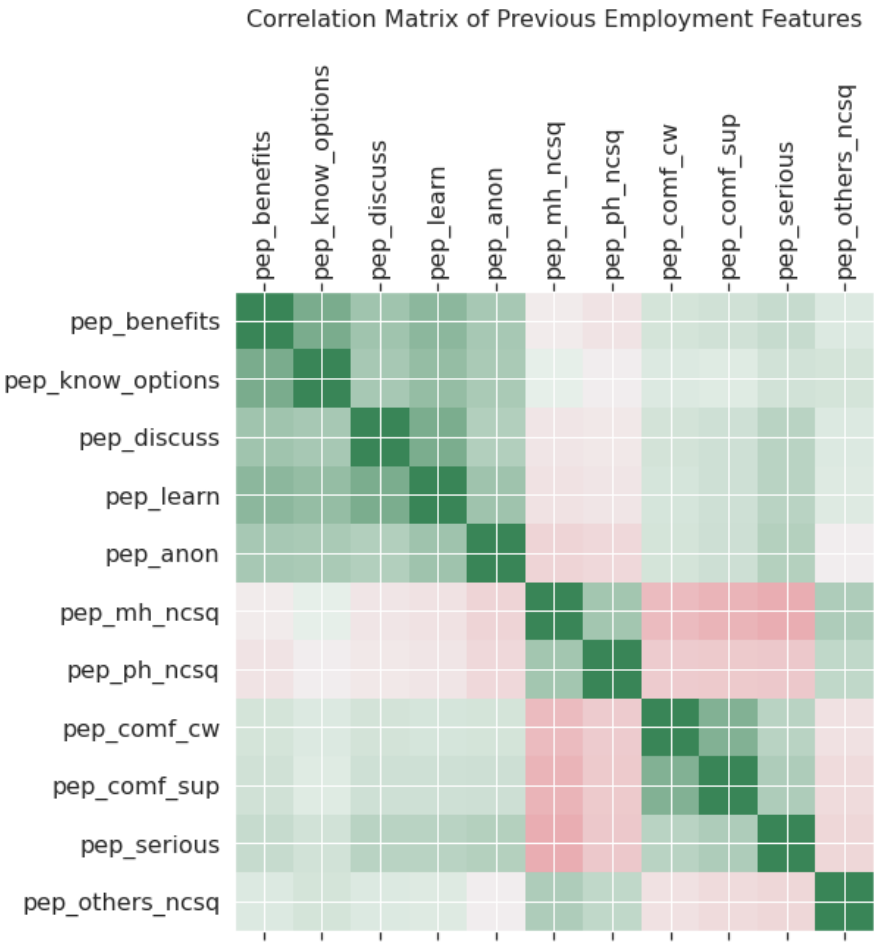
Notable Correlations from Current Employment Zone



It can be observed from the zone mentioned above in the figure that the lesser sense of threat from bringing up Mental Health (MH), the easier it is to obtain leave for Mental Health issues. Alongside, the greater the sense of threat from bringing up Physical Health (PH), the greater the sense of threat from bringing up Mental Health. Similarly, the greater the comfort in sharing about Mental Health with coworkers or supervisors, the easier it is to obtain leave for Mental

Health issues. Consequently, the more employers view Mental Health as seriously as Physical Health, the less the sense of threat from bringing up Mental Health.

Notable Correlations from Previous Employment Zone

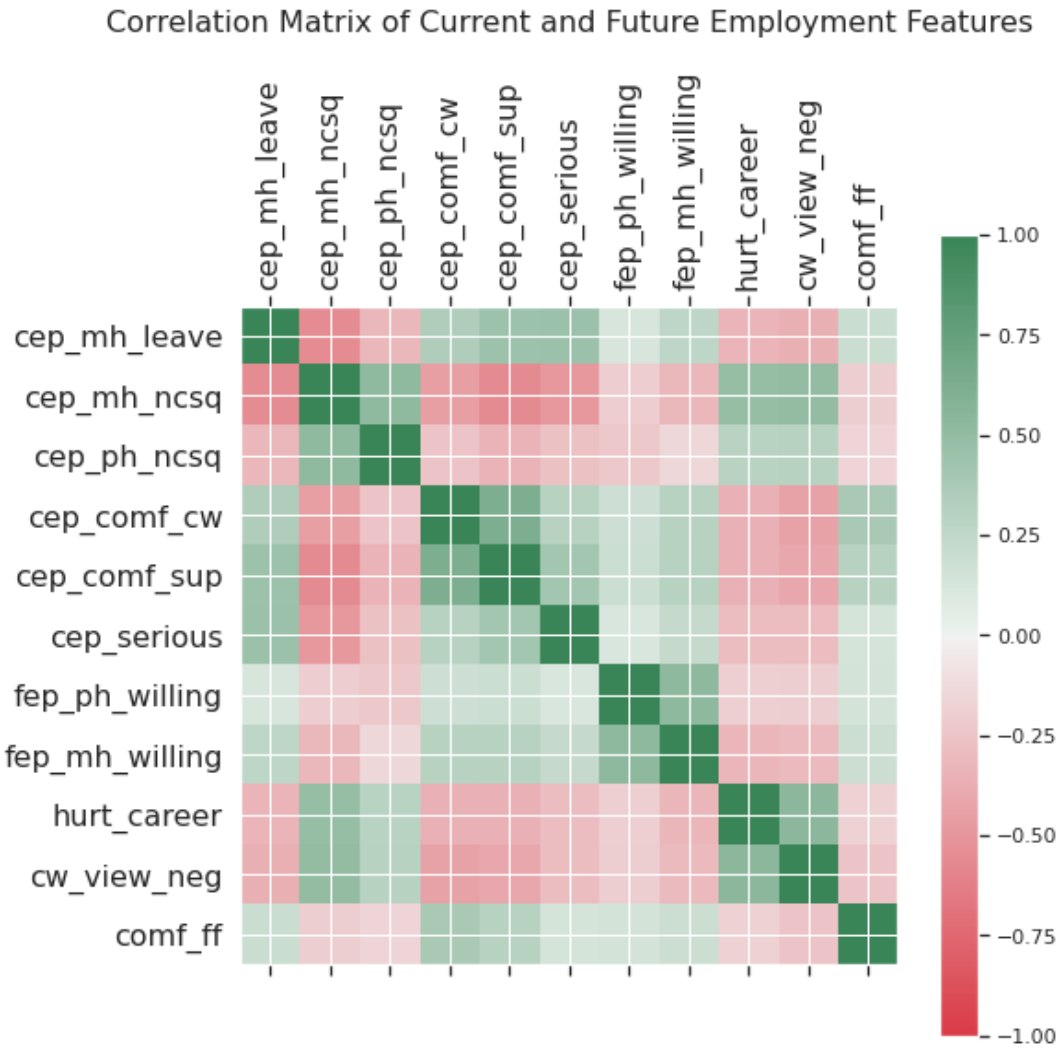


Features of previous employment related to company policy (e.g., availability of Mental Health benefits) are more strongly positively related than compared to features of current employment. However, the negative correlation between the comfort of sharing about Mental Health issues with supervisors & coworkers and the sense of threat of raising Mental Health issues is milder for previous employment than current employment. Nevertheless, the general direction of correlation is similar for both previous and current employment features.

Notable Correlations from Future Employment Zone

The heat map shows that the more willing one is with sharing about Physical Health, the more willing one is with sharing about Mental Health. The more one feels that being identified as having Mental Health will hurt one's career; the less willing one is to share about Mental Health. The more one feels that coworkers will view one more negatively if one has Mental Health, the more one feels that being identified as having Mental Health will hurt one's career.

Notable Correlations from Current and Future Employment Zone



It can be observed that current experiences affect future willingness. The more one senses a threat of negative consequences when raising Mental Health concerns with current employers; the less likely one is willing to raise Mental Health issues with future employers. More likely, one will feel that being identified as having Mental Health will hurt one's career and cause coworkers to view one negatively. However, the more willing one is to share about Physical Health with future employers, the more willing to share about Mental Health. The more one feels that coworkers will view one negatively for having Mental Health, the more one feels that being identified as having Mental Health will hurt one's career.

Machine Learning & Prediction

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention (sas, n.d.). Machine learning can be divided into 3 groups: supervised, unsupervised and reinforcement learning. Our methodology involves the use of supervised learning algorithms and classification techniques

Supervised learning

Supervised learning provides you with a powerful tool to classify and process data using machine language. With supervised learning you use labeled data, which is a data set that has been classified, to infer a learning algorithm. The data set is used as the basis for predicting the classification of other unlabeled data through the use of machine learning algorithms (Ryan, Talabis, Kaye, 2015).

Applying Classification

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data. From a modeling perspective, classification requires a training dataset with many examples of inputs and outputs from which to learn (Jason Brownlee, 2020).

Input Variables

After exploring the dataset, relevant input variables were selected. Those variables concerning current and previous employers along with those variables related to respondent's profile were chosen as input variables.

Target Variables

As described in the objectives section, different classifiers will be built for each of the target variables. There are 3 target variables are:

1. fep_mh_willing
2. mh_diag_pro
3. sought_treat

Model:

The dataset was split, with 25% for testing data and 75% for training data respectively. Different Classification models were applied on the input variables to predict the target variables.

The following classification models were used:

Support Vector Machines (SVMs)

SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes.

K-Nearest Neighbors (KNN)

KNN is a simple algorithm which uses data and classifies new data points based on similarity measures (e.g., distance function).

Random Forest Classifier

Random forest classifier creates a set of decision trees from a randomly selected subset of the training set. It then aggregates the votes from different decision trees to decide the final class of the test object.

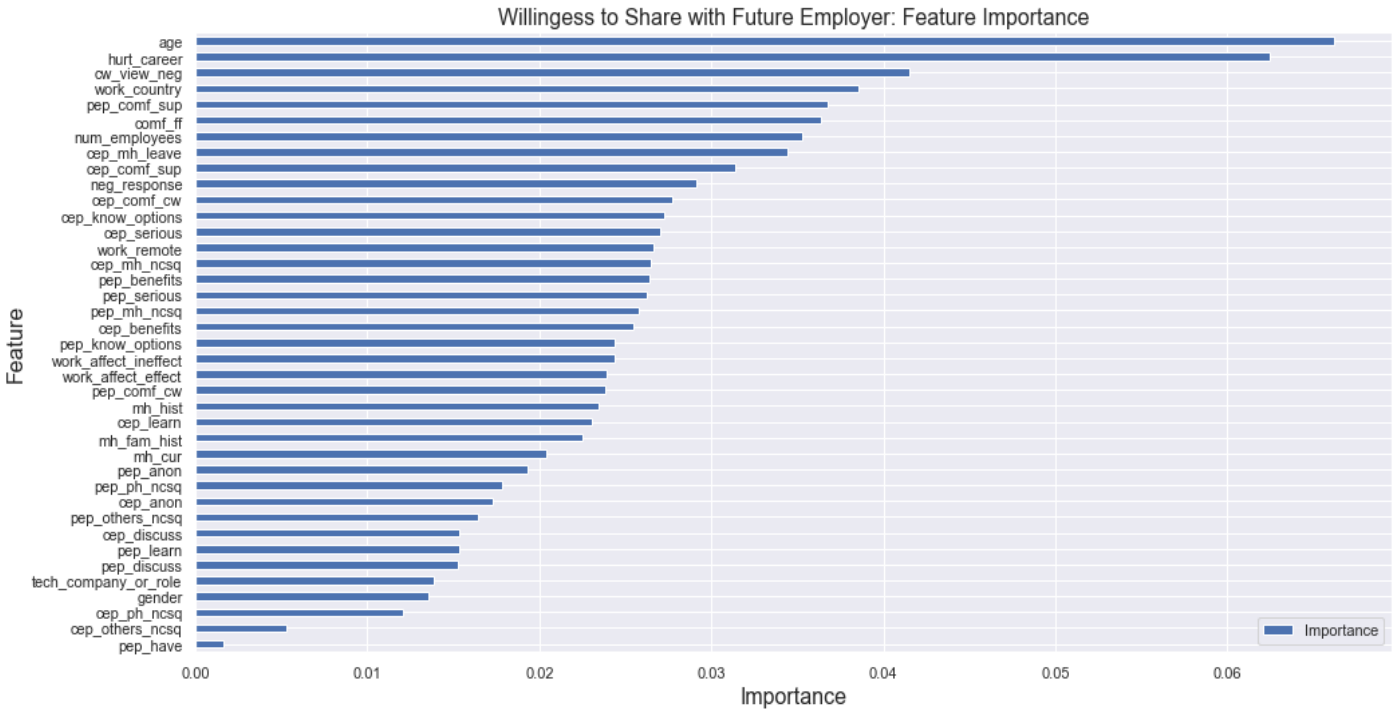
Performance Evaluation & Comparison of Methods

ACCURACY TABLE	Target Variable 'fep_mh_willing'	Target Variable 'mh_diag_pro'	Target Variable 'sought_treat'
Random Forest Classifier	0.7243	0.8939	0.8763
Kneighbors Classifier	0.6431	0.7561	0.7844
SupportVector Machines	0.6996	0.5265	0.5512

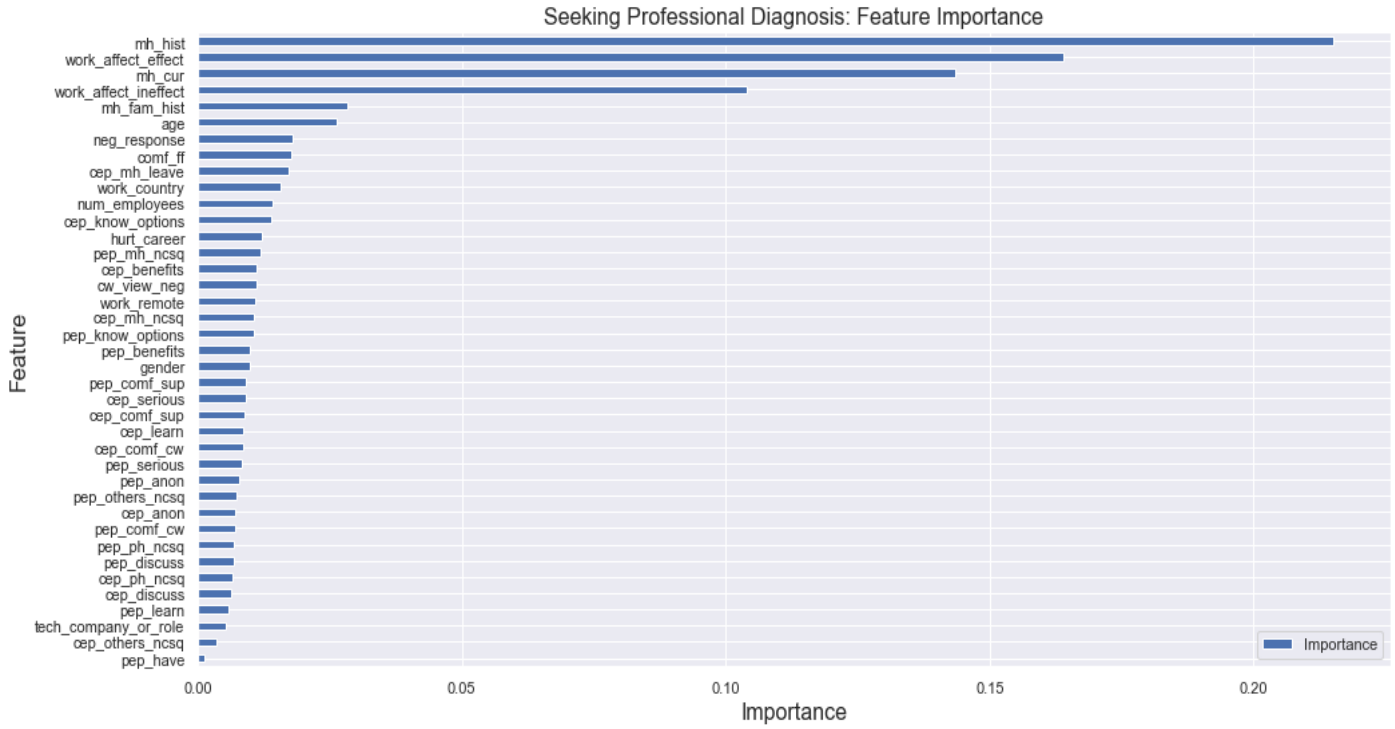
The Table above shows that 3 different classifiers were built for each target variable. All classification models were trained on input dataset and their accuracy was calculated on test dataset. Out of all 3 classification models, RandomForestClassifier was the most accurate on all target variables.

Discussion of Findings

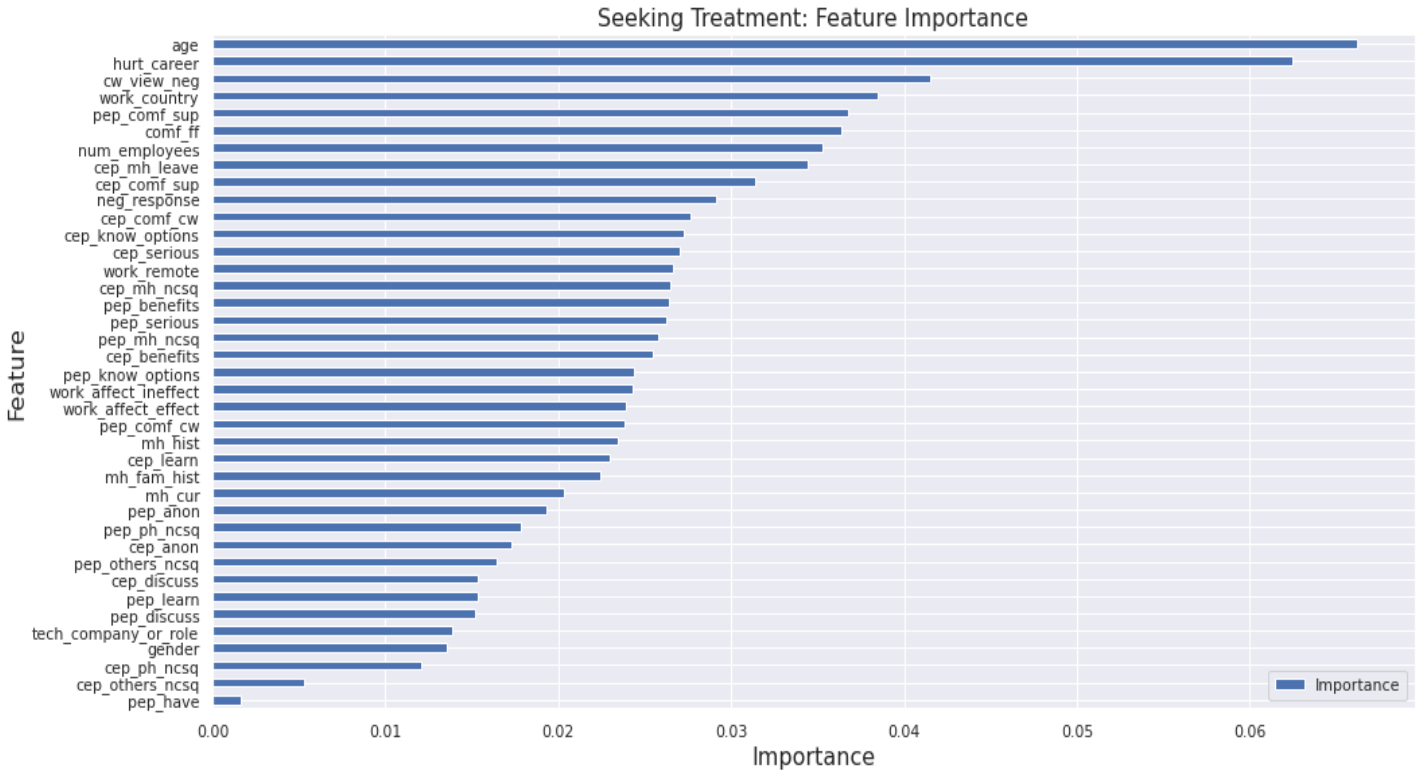
As RandomForestClassifier performed the best therefore feature importance with different targets were found.



- Age is the most important predictor, although there appears to be nonlinear correlation with willingness to share about MH with future employer, as observed in the Correlation Matrix.
- The sentiment that being identified as having MHD as detrimental to one's career is the next most important predictor.
- Intuitively, the more one feels that being identified with MHD will hurt one's career, the less likely one will bring it up with a future employer



- Having a family history of MH issues is the most important predictor of seeking professional diagnosis.
- Age and openness in sharing about MHD with family or friends are also important predictors over the remaining features.
- The older, the more likely to seek professional diagnosis.
- The more comfortable with family or friends, the more likely to seek
- Diagnosis.
- The most important predictors related to current employment are knowingly the MH care options provided by current employment as well as ease of obtaining leave for MH.
- The more one knows one's options with the current employer, the more one may seek diagnosis.
- Curiously, the harder to obtain leave for MH, the more likely one may seek diagnosis.
- The experience of negative responses to MH issues is a similarly important predictor.
- The less negative response experienced or observed, the more one may seek diagnosis



- Feature importance for Seeking Treatment is similar to Willingness to Share with Future Employer

Conclusions and Further Work

Conclusively, the older one is, the more likely one will have sought treatment for Mental Health. Age is found to be the most important predictor in the model. The sentiment that to be identified as having MHD as detrimental to one's career is the next most important predictor. Intuitively, the more one feels that being identified with MHD will hurt one's career; the more likely one would have sought treatment for it. There is still a lot more improvement needed regarding the company's openness and people in general to mental health. Different classifiers can be applied to improve accuracy and different target variables can be selected to gain more insight. The dataset contains a smaller number of rows, more data can be beneficial in seeing general trend of public.

References

Mental illness. (2019, June 8). Mayoclinic. https://www.mayoclinic.org/diseases-conditions/mental-illness/symptoms-causes/syc-20374968

Machine Learning. (n.d.). Sas. https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.

Ryan, Talabis,. . . Kaye. (2015). Analytics Defined. Information Security Analytics. Published. https://www.sciencedirect.com/topics/computer-science/supervised-learning

Jason Brownlee, J. B. (2020). 4 Types of Classification Tasks in Machine Learning. Python Machine Learning. Published. https://machinelearningmastery.com/types-of-classification-in-machine-learning/#:~:text=In%20machine%20learning%2C%20classification%20refers,one%20of%20the%20known%20characters.

OSMI Mental Health in Tech Survey 2016. (2016). Kaggle. https://www.kaggle.com/osmi/mental-health-in-tech-2016