

Mental Health in the Workplace

(COMP3125 Individual Project)

Lili Stearns
School of Computing and Data Science

Abstract—Mental health is a topic people tend to shy away from, especially in the workplace. This project aimed to dive into how workplace factors impact an employee’s mental health and their attitudes towards it. Two datasets were used to find answers: An ongoing Open Sourcing Mental Illness (OSMI) survey starting in 2016 and a Gallup survey taken in 2022. A Python notebook was used to analyze the data, build a machine learning (ML) model, and visualize results. In the end, several conclusions were made regarding mental health in the workplace and around the world.

Keywords—Mental Health, Workplace, Machine Learning, Python

I. INTRODUCTION

The topic of mental health has historically been a topic not to be discussed with others. The main reasoning comes from some believing it leads others to look at or treat you differently. It was better to bottle up these feelings than to seek help and be potentially labeled with an official diagnosis. In this new day and age, however, it has started to become more acceptable to open up about mental health to friends and family without fear of judgement. Despite this, one place where people still feel hesitant to share their mental state is the workplace. The fear of being replaced because peers and supervisors no longer see you as capable of doing your job is unfortunately still a common occurrence.

The view of mental health in the workplace can depend upon which age groups are at a specific company. Older individuals, who have been in the workforce a long time, may have a completely different attitude towards how the workplace affects their mental health compared to younger individuals just starting their careers. Another influencing factor is the company culture surrounding mental health. Whether or not the company covers medical expenses for mental conditions or provides helpful resources for getting help can sway how comfortable employees are with requesting medical leave due to their mental health. Perhaps the biggest influencing factor is the day-to-day interactions with peers and supervisors. Witnessing or experiencing either positive or negative responses to sharing mental health at work may impact how safe others feel to share theirs.

The most important thing to remember is that even though the number of professional diagnoses for mental health can vary per country, mental health is a global societal issue. In fact, the World Health Organization (WHO) reports an equivalent of 1 trillion USD is lost globally every year due to the 12 billion working days taken off due to depression and anxiety [1]. With an estimated 60% of the population working around the world, it is crucial to study mental health in the workplace [1]. As a result, workplace culture can not only foster a positive mental impact on its employees but also create an environment for more quality work to be done.

II. DATASETS

A. Source of datasets

Dataset 1 is from an Open Sourcing mental Illness (OSMI) survey started in 2016. The dataset is hosted on Kaggle and was last updated in 2019. OSMI is a non-profit organization focusing on studying mental health, particularly in the tech community. Even though Kaggle is not the most credible provider (not peer reviewed), OSMI appears to be a credible source for the survey. OSMI’s goal with the survey was to examine mental health in the tech workplace, including how prevalent it is.

Dataset 2 is from a Gallup online survey conducted in 2022. The dataset is hosted on Statista. Both Gallup and Statista are credible providers for surveys and datasets. Only individuals in the United States who were currently either full- or part-time employees participated in the survey. The goal of the survey was to look at job impact on mental health by different age groups (positive or negative).

B. Character of the datasets

Dataset 1 has 1,433 responses and comes as a CSV file. The rows correspond to each respondent. There are 63 columns that each represent a question asked during the survey. For the purpose of this project, only seven columns were used to analyze the data and find conclusions. These columns are:

1	Does your employer provide mental health benefits as part of healthcare coverage?
2	Does your employer offer resources to learn more about mental health concerns and options for seeking help?
3	If a mental health issue prompted you to request a medical leave from work, asking for that leave would be:
4	Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?
5	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?
6	Have you been diagnosed with a mental health condition by a medical professional?
7	What country do you work in?

Columns 1, 2, 3, and 5 contained missing data and column 5 converted one of the answer categories ('N/A') to NaN resulting in missing data. For columns 1, 2, and 3, any missing data was assigned a new category of 'Self-Employed' since only self-employed individuals did not answer these three questions. For questions 4 and 5, any missing data was assigned to the existing category for maybe to keep it neutral.

Dataset 2 has 15,809 responses and comes as an excel file. The first sheet contains detailed information regarding the survey. The second sheet contains the resulting data. The rows are broken down into five different age groups [18-29, 30-39, 40-59, 50-64, 65+]. There are two columns that represent positive job impact on mental health and negative job impact on mental health. The data is measured by the percentage of each age group that reported a positive impact and the percentage that reported a negative impact. The data did not have any missing entries and therefore did not need any data cleaning. For analysis, the data was taken and reformatted into a CSV file with the same rows, columns, and data percentages (Note: the column and row names were renamed for simplicity purposes).

III. METHODOLOGY

Below will be a discussion of the methods/models used to answer each question. The Pandas Python library was used to initialize the datasets for each method/model.

A. Method A

Question to Answer: Can we predict how a workplace's attitude towards mental health (whether or not they cover care for it and provide resources for getting help) impacts an employee's perceived ability (on a scale of very easy to very difficult) to take medical leave due to their mental health?

Answering this question involved creating a machine learning (ML) model. The response variable (or the variable to be predicted) is categorical with more than two classes. Based on the response being a non-binary classification problem, a decision tree model was used. A decision tree can handle multi class categorical problems well and the resulting tree can be visualized to allow anyone to predict the response based on the predicting features used. A decision tree model will not work as well if the boundary between classes is more linear.

The Scikit-learn (Sklearn) Python library was used to create the decision tree model. Specifically, these Sklearn imports were used: the `train_test_split` import created training and testing subsets from the data, the `DecisionTreeClassifier` import built the decision tree, the `tree` import visualized the final tree, and the `metrics` import calculated model accuracy on the test set as well as creating a confusion matrix.

B. Method B

Question to Answer: How do different age groups view the impact of the workplace on their mental health (positive or negative impact)?

Answering this question involved analyzing the data and creating a visualization. Since there was a comparison of the different age groups (categorical) with the percentage of negative and positive impact (numerical), a line plot visualization was used. And, based on how the data was formatted, it was simpler to create two visualizations, one for a negative impact and one for a positive impact.

The Seaborn Python library was used to create the two line plots with negative/positive impact on the y-axis and age range on the x-axis. The Matplotlib Python library was used to plot, add a title, and add x and y labels to the visualization.

C. Method C

Question to Answer: Did more employees feel deterred about sharing their mental health in the workplace after witnessing an unsupportive response towards a co-worker sharing theirs?

Answering this question involved analyzing the data and creating a visualization. Since the specific question only required results from witnessing an unsupportive response, a count plot was used (indexing only the rows answered as 'Yes, I observed'). A line plot was also created to capture the overall trend across all observation types and feelings of being deterred.

The Seaborn Python library was used to create the count plot and line plot with the count on the y-axis and if they felt deterred on the x-axis. For the line plot the legend represented the observation type. The Matplotlib Python library was used to plot, add a title, add x and y labels, and format the legend of the visualization.

D. Method D

Question to Answer: How does the rate of a professional diagnosis of mental health conditions compare across different countries? In other words, does one country have significantly more diagnoses than the others?

Answering this question involved analyzing the data and creating a visualization. Since the main comparison was the number of diagnosis (numerical) per country (categorical), a count plot visualization was used. Considering there was a total of 53 countries listed by respondents, some of which only being listed by 1-2 respondents, only results from the top five most frequently listed countries were visualized.

The Seaborn Python library was used to create the count plot with the count on the y-axis, country on the x-axis, and if they were diagnosed (yes/no) as the hue on the legend. The Matplotlib Python library was used to plot, add a title, add x and y labels, and format the legend of the visualization.

IV. RESULTS

Below will be a discussion of the results and findings for each question.

A. Result A

The decision tree model only resulted in a 38% prediction accuracy. Out of the 286 data entries in the testing set, 110 of them were correctly predicted, while the other 176 were incorrectly predicted.

Figure 1 tells us a lot about how the predictions compare to the actual values. The 'Self-Employed' and 'Somewhat easy' labels were most accurately predicted with 62 and 38 correct predictions, respectively. The 'Somewhat easy' label was the most predicted label with 174 total predictions, and it accounted for 136 of the incorrect predictions. There were no predictions of the 'I don't know' and 'Neither easy nor difficult' labels.

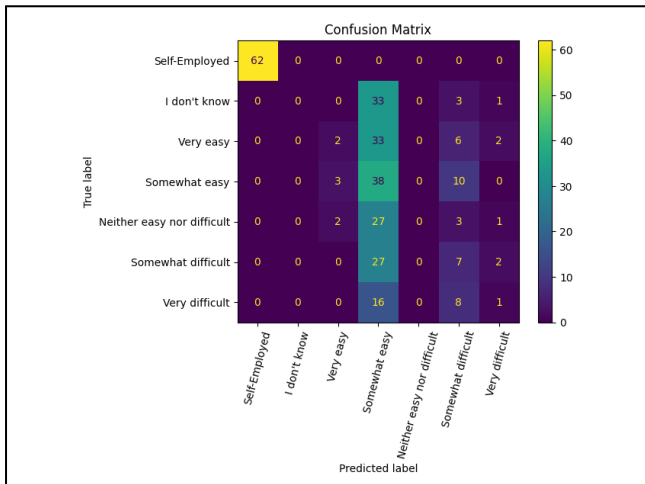


Figure 1. Confusion matrix to show actual v. predicted

B. Results B

There was a noticeable trend in the reporting of work having a negative impact on mental health. Starting with the 18-29 age group, a large percentage of individuals reported a negative impact. As the age groups got older, the percentage shrank. By the 65+ age group, there was a significant drop of more than 30%.

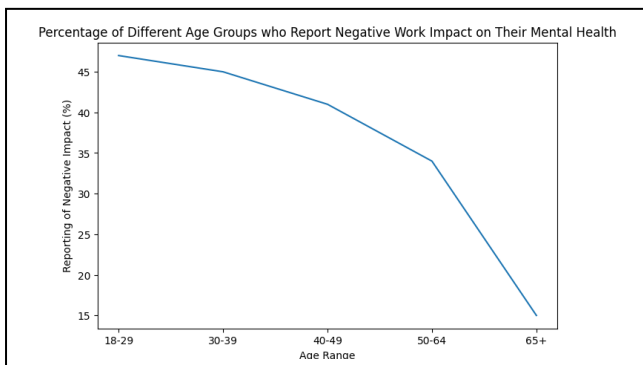


Figure 2. Line plot of the percentage of individuals reporting a negative work impact on their mental health

There was not as a straightforward trend in the reporting of work having a positive impact on mental health. The lowest reporting came from the 50-64 age group, while the highest reporting came from the 65+ age group.

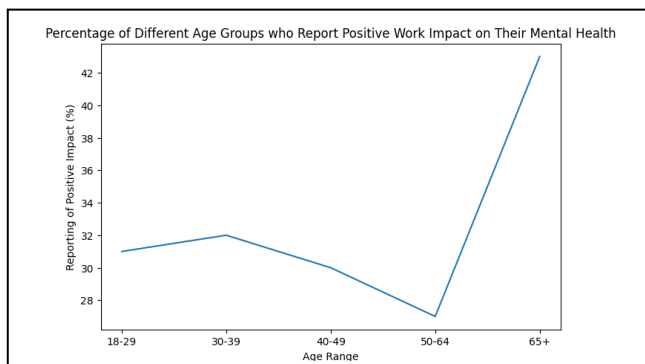


Figure 3. Line plot of the percentage of individuals reporting a positive work impact on their mental health

Overall, a negative impact had a higher reporting percentage than a positive impact. Also, the 65+ age group had the greatest difference in reporting the two impacts with 15% reporting negative and a little over 42% reporting positive.

C. Results C

Out of the 264 individuals who reported witnessing an unsupportive response towards a co-worker sharing their mental health at work, 128 felt deterred to share their mental health. In fact, observing a bad response was the number one factor in someone feeling deterred, with experiencing a bad response being the second greatest factor.

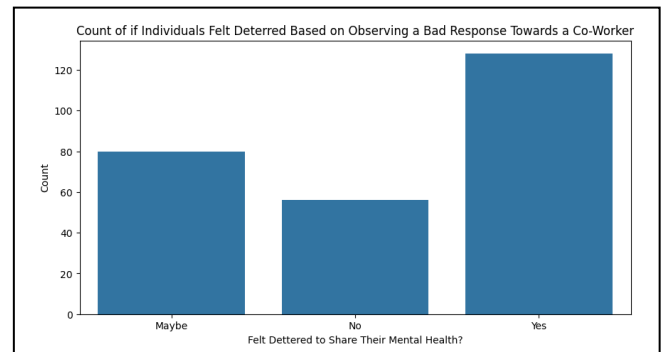


Figure 4. Count plot of if individuals felt deterred after observing an unsupportive response

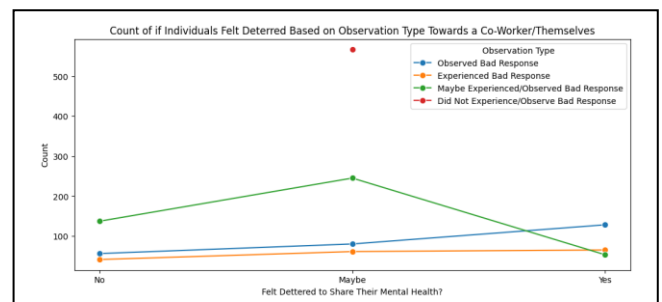


Figure 5. Line plot of if individuals felt deterred based on observation type

D. Results D

The top responded countries individuals work in are: United States, United Kingdom, Canada, Germany, and Netherlands. The United States had the most respondents by more than 500, so naturally it had the highest total diagnoses count. Looking at ratios though, the United States and the Netherlands are the only countries that have more diagnoses than non-diagnoses. Germany has the highest ratio of non-diagnoses at roughly 64%, compared to the United States at about 44%.

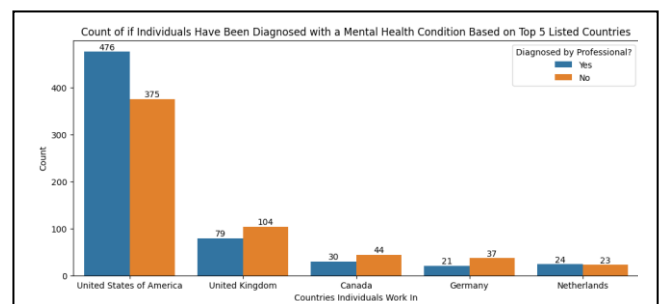


Figure 6. Count plot of mental health diagnoses by country

V. DISCUSSION

The biggest unsatisfactory result was the low accuracy of the decision tree model, especially since it was below 50%. There are a few possible fixes in the future to improve the accuracy. The first step would be to experiment with different decision tree parameters (or otherwise tune the model). With proper tuning, the trained model can find the optimal flexibility without under or overfitting. Additionally, other ML models could be used to compare accuracies. Perhaps the data falls within the linear boundaries and another ML technique, such as neural networks, would have better luck at predicting the response. Finally, experimenting with other predictor variables could be beneficial, as the two used for the current model may have a minimal relation to the response.

VI. CONCLUSION

Discussion around mental health grows in acceptance as the years go on. Yet, in professional places like the workplace it is still a commonly avoided topic. The reason for avoidance can range from not wanting peers to judge you, to believing personal problems should stay separate from work. But, mental health is not just a personal problem. It affects every aspect of someone's life, such as their productivity level and how they interact with others.

The workplace can even have an impact on an individual's mental health. A much higher percentage of younger age groups, who are just entering the workforce or have only been in it a decade or so, experience a negative impact to their mental health compared to older age groups nearing retirement. A large percentage of individuals 65+ report experiencing a positive impact to their mental health due to their work.

A workplace's culture and environment can greatly impact how its employees feel towards mental health. If the culture meets mental health discussion with unsupportive and judgmental responses, employees will be timid to continue sharing their mental health struggles with peers and supervisors. If, however, the sharing of mental health is met with supportive responses, especially from supervisors and management, employees will develop a positive relationship with being open about their mental health.

The recognition of mental health as a serious topic does not just vary company by company, but also country by country. Access to medical professionals who can diagnose and help manage mental health, and each societies' view on mental health acceptance both contribute to the difference. Findings show the United States had the highest ratio of diagnoses to non-diagnoses at roughly 56% and Germany had the lowest at about 36%.

Taking mental health concerns as seriously as physical health concerns will not only benefit the individual, but also the workplace community. Creating a supportive and safe work environment will improve productivity, employee relationships, and company culture as a whole. And, someday, talking about mental health may be just as common as talking about last night's game.

ACKNOWLEDGMENT

A big thank you to Dr. Weijie Pang for guidance throughout this project.

REFERENCES

- [1] World Health Organization (WHO), "Mental Health at Work," *World Health Organization*, Sep. 02, 2024. <https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work>