Predicting Mental Health Issues in the Tech Industry

Hasan Ahamed
Department of Economics
Rice University
Houston, TX, USA
ha63@rice.edu

Matthew Hong
Department of Mathematics
Rice University
Houston, TX, USA
mmh12@rice.edu

Abstract:

Background: Mental Health issues have become increasingly prevalent, especially in the stressful, fast-paced environment of technology companies. This hinders employees' efficiency, earnings, employment, and most importantly happiness, hurting both the individuals and employers. To curb these mental health issues, we aimed to use machine learning methods to predict whether people would seek mental health treatment. This machine will have policy implications as workspace even the government can take policies to control the features that cause mental health issues.

Data and Methods: The dataset contains 748 employees in tech industries, of which 54% sought treatment for mental health issues. We primarily used SVM and Random Forest Models.

Results: Our models can predict with 70% to 83% accuracy if a person will be seeking treatment or not based on the explanatory variables. We were also able to generalize this result for the employees living in the UK.

Conclusion: We believe that using machine learning methods is a viable way to predict who will seek mental health treatment and who will not. Using these methods, companies could set up programs, learning sessions, and other benefits for employees who may be at risk in order to help them seek treatment earlier, therefore reducing the total effect the mental health issues will have on them in the long run.

1. Introduction:

Mental Health issues have been increasing globally for the past few decades [4]. 23% of the US population suffers from anxiety and depression in the US ([5]). Young people in the US are suffering from mental health issues at a significantly higher rate than the overall population. A study showed that anxiety and depression disorders increased by 63% among the young population in the US from 2005 to 2017 [6]. Additionally, COVID-19 increased the prevalence of mental health issues by 25% across the world [7].

Researchers have found severe consequences, in both the personal and professional lives, of people suffering from mental health issues. For instance, A study concluded that anxiety and depression were highly associated with reduced productivity such as absenteeism and presenteeism in the workplace [2]. Similarly, using the National Longitudinal Study of Youth 1979 dataset, another study found that depression causes employment to drop by 10% and earnings by 27% [3]. A systematic review of qualitative research depicts that the quality of life among people with severe mental health disorders often leads to poor quality of life [1].

The primary motivation of this project is to identify the features that have a high influence on the mental health issues of people. As we know, machine learning techniques have been proven superior to traditional statistical methods to analyze complex patterns in the data. Hence, we think using machine learning techniques will improve the understanding of predictors of mental health issues significantly. After identifying the important features that are responsible for poor mental health, we will be able to design policies for the workspace to alleviate those symptoms and ensure a productive and healthy workspace. Simultaneously, individuals can take precautions to avoid those situations since that will guarantee them a higher earning and better quality of life.

In the rest of the report, we will delve into the previous literature that tried to predict mental health issues in tech industries using machine learning methods. Then we will discuss our models and methodologies, and the results we get using those models. Finally, we will conclude with a discussion of what we found and what further research can be conducted to understand this topic in more detail.

2. Literature Review:

In the last few decades, machine learning techniques have become extremely popular in predicting various outcome variables across different industries. The medical industry was not an exception. Consequently, researchers have examined the various machine learning models and their accuracy in predicting the level of anxiety, distress, depression, and other mental health-related disorders. For example, A study used the Convolutional Neural Network (CNN), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) to predict the intensity of anxiety and depression and found that CNN predicts anxiety and depression with 96% accuracy whereas SVM was able to predict with 95% accuracy [1]. Another research tested several machine-learning techniques including decision trees, probabilistic, ensembles and deep learning-based classifiers and found that Extreme Gradient Boosting works best in predicting the risk of mental health crises [4]. A systematic review conducted by another group of researchers of the machine learning methods employed to predict mental health issues found that researchers typically use Convolutional Neural Networks (CNN), Random Forest (RF), Support Vector Machine (SVM), Deep Neural Networks, and Extreme Learning Machine (ELM). They also found that CNN works best among these models but the challenges in explaining the differences in the population, study settings, and background still persist [6].

Even though many machine-learning methods have already been implemented in various settings, predictors of mental health in the tech industry have not been explored much and it is imperative to think that the predictors of the mental health issues in the tech industry might be different than the other industries. Moreover, almost no studies have tried to generalize the findings of the US to some other countries such as the UK. We plan to fill this gap in the literature through our research.

3. Data & Methodology:

3.1 Data:

3.1.1. Data Description:

The dataset we will be using for this project comes from a 2014 survey that measures the attitudes toward mental health among workers in the tech industry [8]. Data has been collected mostly from the US (60%) and UK (15%), but there are some other countries as well including Canada, Brazil, France, Mexico, Ireland and others. The survey asked questions about mental health issues, family history of mental issues among the respondents, the impact of those issues in their workspace, and how they communicate about those issues with their supervisors, colleagues, and others. Even though we wanted our primary outcome variable to be measures of anxiety or depression, owing to lack of data, we will be using the answer to the question “Have you sought treatment for a mental health condition” as our outcome variable. Since most of the observations are coming from the US, we will work primarily with the US data to avoid high variability among different countries. Finally, we will test our models, developed using the US data, on the UK data to see how well those models can be generalized.

3.1.2. Data Processing:

Most of the features we have in the dataset are binary (except for the age variable) in nature since the survey participants had to answer questions with yes or no. However, the respondents could choose not to answer a question (or respond “Don’t Know”) which in some cases is more than 50% of the observations. Hence, removing all the observations with no answers or ‘Don’t Know” would significantly shorten the dataset. We therefore recorded 0 as “No”, 1 as “Don’t Know”, and 2 as “Yes” (assuming that both “Yes” and “No” are stronger responses than “Don’t Know”, hence “Don’t Know” should be in the middle of both). We also dropped all the observations except for the US as we only wanted to work with the US data as described earlier. Our outcome variable is binary as the respondents were allowed to answer only yes or no to that question.

3.2 Feature Selection:

We plan to explore the variables that are most important in predicting the outcome variables. To do that, we divided our explanatory variables into three sets. First of all, we used all the variables in the dataset to predict the outcome variable. Secondly, using the knowledge of medical and psychology literature, we hypothesised that 13 variables (see Table 3a) in the dataset might play an important role in predicting the outcome variable, which we called the Hypothesized Important features. Finally, we used all the variables that have a correlation of greater than 0.1 with the outcome variable to create a final model, which we called High Correlation (High Corr.) features (see Table 3a). Note that we used SVM and Random Forest with all three sets of explanatory variables to see which model works the best to predict mental health issues.

3.3 Model Selection:

The primary goal of this project is to predict if an individual will be seeking treatment or not based on the explanatory variables we have in our dataset. Because of the fact that our explanatory variables are mostly categorical, (since we have “Don’t Know” as a category in most cases), it makes sense to use classification algorithms instead of regression algorithms. Therefore, we plan to employ a Support Vector Machine and a Random Forest Model to predict if someone will be seeking treatment or not. Our goal is to see which predictive algorithm will work better to predict the outcome variable.

3.4 Performance Metrics:

We will be using the train-test-split method to train the data on 80% of the observations and the rest of 20% to test the models. We will primarily use test accuracy and precision to evaluate the models, however, we will be looking at the confusion matrices and f-scores as well.

4. Results:

4.1 Exploratory Analysis:

The dataset comes with a total of 1259 observations but after removing observations from all the countries except for the US, we have 748 observations. We will conduct our primary analysis on these 748 observations. Since we only have categorical variables (except for age), it is reasonable to look at the histogram of these variables. Figure 1a (in the appendix) shows the histogram of the outcome variable. As we can see from Figure 1a, the outcome variable is well-balanced. 340 (45.5%) individuals responded that they had sought treatment for mental health issues whereas 408 (54.5%) persons said they did not, among the people who have been living in the US.

In the dataset, we have 553 males and 176 females, 76 people are under 25 while 356 people are between 25 and 35. Among the states, California has the most participants (18.4%) followed by Washington (9.4%) and New York (7.5%). This shows that people from the states that have highly dense tech firms participated in a higher number compared to the other states. Figure 1 shows the histogram of the variables that are vital in predicting the outcome variable.

Figure 1: Histogram of Selected Features

4.2 Model Evaluation:

In creating and testing our models, many of the parameters were tuned to increase accuracy. For both SVM and RF, our process for tuning the parameters consisted of manually adjusting each parameter and running the model, taking note of the training accuracy, testing accuracy, and precision. When we arrived at a parameter set that yielded the highest accuracies (or so we thought at the time), we then ran a GridSearchCV process over the parameter set, in which for each parameter we made a buffer of around 5 units above the so-called best parameter. GridSearchCV then ran the model over all combinations of these parameter sets to see if one existed that gave better accuracy scores, and in some cases it did.

From Table 1a, we see that the RF model outperformed the SVM model in testing accuracy and precision in all feature set configurations. When all features were used, RF achieved a testing accuracy of 0.777 compared to SVM’s 0.721; similarly, precision was higher for RF at 0.752 versus 0.724 for SVM, indicating that the accuracy of treatment predictions for the RF model was higher. When using the High Corr features, RF yielded a training accuracy of 0.813, testing accuracy of 0.825, and precision of 0.825, compared to SVM’s 0.783, 0.713, and 0.720. Another interesting outcome is that the machines run on the High Corr features, for both SVM and RF, have higher training accuracy, testing accuracy, and precision than their respective machines run on the Hypothesized Important features. This shows that our human intuition of which features are truly important is not as accurate as the correlation formulas.

When running the same machines on the UK data (Table 2a), we see similar results. First, the High Corr. feature set strictly dominated the Hypothesized Important feature set across both models in terms of testing accuracy and precision. Also, the SVM being run on just the eight High

Corr features actually had better training accuracy, testing accuracy, and precision than the SVM running on all features, with scores of 0.809, 0.717, and 0.750 versus 0.735, 0.701, and 0.667, respectively. Interestingly, this is the opposite case for RF, where running on all features produced higher scores for those three metrics. We believe that culturally, the US and UK are similar enough that the models work with almost equal accuracies. However, these models may not extrapolate to countries with different cultures, like Japan, China, or Brazil.

5. Discussion:

The objective of this study was to evaluate the use of ML techniques to predict whether individuals working in the tech industry would seek treatment for mental health issues. The findings from our analysis show the effectiveness of both SVMs and Random Forest (RF) models to achieve accuracy in both the training and testing datasets. Here, we interpret the results, highlight the implications, and discuss the limitations of our project.

Performance of Feature Subsets: The analysis of the feature subsets provides valuable insights:

- All Features: On the US data, using all available features resulted in the best training accuracy for both models but did not generalize as well to the testing dataset, suggesting some overfitting.
- Hypothesized Important Features: Models performed worst when restricted to hypothesized features, indicating that these variables may not capture the full complexity of the problem or are insufficient for prediction alone.
- Highly Corr. Features: This subset achieved the best balance of accuracy and precision across both training and testing datasets, suggesting that feature selection based on correlation is a practical and effective strategy.

Strengths and Limitations:

Strengths:

- The models showed strong performance, particularly RF, with testing accuracy exceeding 80% in some cases. This shows the viability of machine learning in predicting mental health treatment-seeking behaviour.
- By evaluating different subsets of features, we were able to identify the most influential variables, improving both interpretability and generalizability.
- Testing the models on UK data provided an additional layer of validation and highlighted the applicability of the RF model.

Limitations:

- While the outcome variable was relatively balanced in this study, the inclusion of more diverse datasets with a better outcome variable (i.e. “Diagnosed with a mental health issue?”) could better assess the accuracy of these models.
- Despite the performance of UK data, further validation of data from other countries or industries is necessary to confirm the applicability of the models.
- The reliance on categorical variables with a “Don’t Know” category may introduce noise, as this response may not align well with "Yes" or "No."

Conclusion:

In conclusion, our analysis demonstrates the feasibility of using machine learning to predict mental health treatment-seeking behaviour in the tech industry. While RF proved to be the best model, SVM showed competitive performances. These results provide a foundation for further research and practical applications in the workplace to support employee mental health and productivity.

6. References

[1] Janice Connell, John Brazier, Alicia O'Cathain, Myfanwy Lloyd-Jones, and Suzy Paisley. 2012. Quality of life of people with mental health problems: a synthesis of qualitative research. *Health Qual Life Outcomes* 10, 1 (2012), 138. https://doi.org/10.1186/1477-7525-10-138

[2] Claire De Oliveira, Makeila Saka, Lauren Bone, and Rowena Jacobs. 2023. The Role of Mental Health on Workplace Productivity: A Critical Review of the Literature. *Appl Health Econ Health Policy* 21, 2 (March 2023), 167–193. https://doi.org/10.1007/s40258-022-00761-w

[3] Giuseppe Germinario, Vikesh Amin, Carlos A. Flores, and Alfonso Flores-Lagunes. 2022. What can we learn about the effect of mental health on labor market outcomes under weak assumptions? Evidence from the NLSY79. *Labour Economics* 79, (December 2022), 102258. https://doi.org/10.1016/j.labeco.2022.102258

[4] Share of population with mental disorder. *Our World in Data*. Retrieved December 17, 2024 from https://ourworldindata.org/grapher/share-with-mental-health-or-development-disorder

[5] Mental Illness - National Institute of Mental Health (NIMH). Retrieved December 17, 2024 from https://www.nimh.nih.gov/health/statistics/mental-illness

[6] The Rise of Anxiety and Depression among Young Adults in the United States. *Ballard Brief*. Retrieved December 17, 2024 from https://ballardbrief.byu.edu/issue-briefs/the-rise-of-anxiety-and-depression-among-young-adults-in-the-united-states

[7] COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. Retrieved December 17, 2024 from https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

[8] Mental Health in Tech Survey. Retrieved December 17, 2024 from https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey

[9] Ranking the States 2022. Retrieved December 17, 2024 from https://mhanational.org/issues/2022/ranking-states

7. Appendix

Github Link: https://github.com/mmhong12/HongAhamedProject.git

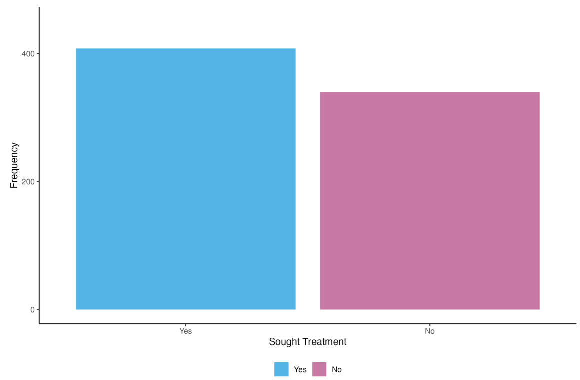

Figure 1a: Histogram of the Outcome Variable

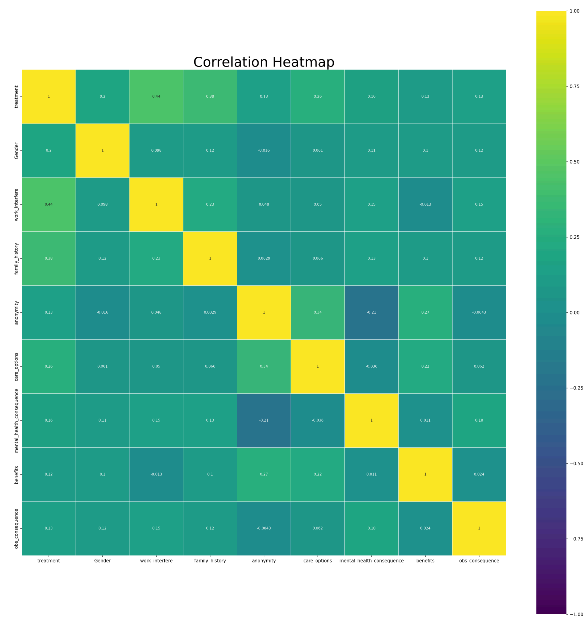

Figure 2a: Correlation Heatmap for the highly correlated variables

Features/Model	***SVM***			***Random Forest***		
	Training Accuracy	Testing Accuracy	Precision	Training Accuracy	Testing Accuracy	Precision
All features	0.845	0.721	0.724	0.878	0.777	0.752
Hypothesized Important Features	0.756	0.705	0.705	0.748	0.717	0.726
***High Corr. Features* (8)**	0.783	0.713	0.720	0.813	0.825	0.825

Table 1a: Model Accuracy (Test on US Data)

Features/Model	***SVM***			***Random Forest***		
	Training Accuracy	Testing Accuracy	Precision	Training Accuracy	Testing Accuracy	Precision
All features	0.735	0.701	0.667	0.866	0.815	0.791
Hypothesized Important Features	0.888	0.663	0.692	0.774	0.663	0.662
***High Corr. Features* (8)**	0.809	0.717	0.750	0.837	0.774	0.740

Table 2a: Model Accuracy (Test on UK data)

Category	Features
Hypothesized Important Features	Age, Gender, self_employed, family_history, remote_work, benefits, care_options, wellness_program, leave, mental_health_consequence, coworkers, supervisor
***High Corr. Features* (8)**	Gender, work_interfere, family_history, anonymity, care_options, mental_health_consequence, benefits, obs_consequence

Table 3a: Feature Categories

Note: We attempted to see if a US state's mental health ranking would have an impact on *treatment*. In a new column, we mapped the best 10 states to a 2, the worst 10 to a 0, and the others to 1 [9]. However, the correlation between this new feature and *treatment* was only 0.01, which is 1/10th the correlation we used for our High Corr feature set. So, we did not continue with running models on this new data frame. We theorized this could be due to the much lower number of data points from "bad" states (103) compared to "good" states (180) skewing the results a certain way.