Using AI to Promote Employee Mental Health

Keaton Chak

Keaton Chak, St-Georeges Senior School, Keaton.chak@gmail.com.

Acknowledgements

Thank you for the guidance of Dashiell Young-Saver from Harvard University in the development of this research paper.

*Abstract* - **The purpose of this paper is to discuss the effect of company policies on employees seeking mental health treatment, specifically in tech. This data will allow tech companies to make choices on policies to promote mental health treatment in the workplace. Using surveyed data on mental health in the workplace, the AI determines the extent that specific company policies predict whether employees seek mental health treatment. The AI uses logistic regression and neural networks to predict if an employee would seek mental health treatment through personal details and company policies. The effect of each variable in the logistic regression can be interpreted to show the answer to the questions of what company policies could affect employees seeking mental health treatment the most.**





**I. Introduction**

Mental health is a huge problem plaguing society today. 1 in 4 adults in America suffer from a mental illness in a given year[1]. This leads to Suicide, the fourth leading cause of death for people ages 15-29[2]. The industry that depression has the most impact on is tech. In 2019, The Bima Tech Inclusion & Diversity Report[3] reported that 52% of tech workers have suffered from anxiety or depression. The goal of my research is to create an AI model that helps tech companies identify policies that promote employees’ likelihood of seeking mental health treatment. If successful, such a model could have highly positive downstream effects, such as creating changes in company policies and lowering the chance of suicide from depression among tech workers. In this paper, I seek to answer the question: What tech company policies are most associated with employees seeking (or failing to seek) mental health treatment?

There has already been work done in this area. Specifically the study “Algorithms and Anxiety: An Investigation of Mental Health in Tech Exploring the Relationship between Mental Health in Tech and the Hardships Faced by Its Employees.[4]” This study used statistical models to find the relationship between a set of independent variables like age, gender, country, and personal and family mental illness history and 2 dependent variables: the employee's willingness to discuss mental health issues with supervisors and their thoughts on facing negative consequences when discussing mental health issues with their superiors.

My study expands upon this prior work by using both inferential statistical models and predictive AI models, rather than just inferential statistical models. In addition, whereas prior work has focused on whether employees feel comfortable reporting their mental health status to their employer, my study focuses on how company policies influence whether employees actually pursue the ultimate step: seeking mental health treatment.

**II. METHODOLOGY**

In this research, I used publicly accessible data from Open Sourcing Mental Health’s 2014 Mental Health in Tech survey [5]. The survey data displays over 1,200 responses from tech workers across the globe. A majority of respondents worked in the United States; however, responses were gathered from 33 countries. The data provides an unprecedented window into company policies around mental illness, mental health benefits packages available to tech employees, and the incidence of mental health disorders among the tech workforce.

The first step was to clean the data. I removed the timestamp column from the dataset which details the time the survey was submitted and removed any country which didn't have more than 10 participants in the survey. I then created a new column called” has\_mh\_disorder,” based on whether the employee otherwise identified in their responses that they had a mental health disorder. The dataset was then filtered for only the employees that reported having a mental health disorder. I also removed the “state” variable, as only America had state information. Next, I graphed the values of the data. Realizing the age variable had many outliers from -1726 to 329, I removed all data that wasn't between the ages of 17 and 100, a realistic working age range. Surveyees who inputted ages outside that range would be considered false data as they likely didn't do the survey seriously. I then categorized all the imputed genders into 4 main categories: Male, female, Non-binary and Trans-female (there were no Trans-male respondents in the dataset).

To prepare the data for model fitting, I one-hot encoded all the categorical variables and standardized all quantitative variables (e.g. divided by the maximum, so that variables like “age” ranged from 0-1). I then split my data into a test (20%) and train (80%) set. Finally, I isolated the target variable: a 0-1 indicator for whether the employee sought mental health treatment.

For the baseline model, I predicted that employees would seek mental health treatment if the company provided medical benefits for mental health. I predicted they wouldn’t otherwise. No other variables were considered. The purpose of this simple baseline model is for performance comparison with the more advanced AI models.

The next model I created was a logistic regression model. I created the model by fitting my data into the sklearn[6] logistic regression model. A logistic regression model is useful, as it can determine the association between multiple predictive variables and our target outcome: seeking mental health treatment. In addition, unlike the baseline model, the logistic regression can consider multiple predictive variables at once. So, I used the model with all my predictor variables, to see which was most predictive of seeking treatment (controlling for the other factors).

The last model I created was a neural network. A neural network is the most powerful of the three. A neural network uses nodes in a structure that replicates human neurons. A neural network is useful as it is usually the most accurate model and the type most AI use today. I created the neural network model by looping through different iterations of neurons and layers of the TensorFlow [7] sequential neural network model(“ tf.keras.models.Sequential”) and using the iteration with the best accuracy.

**III. RESULTS**

**Table I: Accuracy of models**

| **Model** | **Test Set Accuracy** |
| --- | --- |
| Baseline | 0.5393258426966292 |
| Logistic | 0.8314606741573034 |
| Neural Network | 0.8483 |

Both AI models (Logistic Regression & Neural Network) had much higher accuracies on the test set than the baseline model. So, a host of variables beyond a company’s mental health benefits package are important in predicting whether an employee seeks treatment. Although the neural network performed slightly better than the logistic model, it didn’t beat the logistic model’s accuracy by much. In addition, the logistic model is more interpretable than the neural network, as each predictor is assigned a coefficient that indicates its association with the outcome. So, I decided to look further at the strongest predictors from the logistic regression model. See Table II below.

Table II: Coefficient of Logistic variables

| Variable | Coefficient |
| --- | --- |
| benefits\_No | -0.061507 |
| benefits\_Yes | 0.512661 |
| work\_interfere\_Never | -2.276094 |
| work\_interfere\_Often | 1.329171 |
| work\_interfere\_Sometimes | 0.591654 |
| mental\_health\_interview\_Maybe | -0.510137 |
| mental\_health\_interview\_No | -0.024375 |
| mental\_health\_interview\_Yes | 0.534432 |
| anonymity\_Don't know | -0.310535 |
| anonymity\_No | 0.048185 |
| anonymity\_Yes | 0.262270 |
| care\_options\_No | -0.189915 |
| care\_options\_Not sure | -0.305678 |
| care\_options\_Yes | 0.495512 |
| wellness\_program\_Don't know | 0.414579 |
| wellness\_program\_No | -0.020271 |
| wellness\_program\_Yes | -0.394389 |

The positive coefficients indicate variables that are positively associated with employees seeking treatment. The negative coefficients indicate variables that are negatively associated with seeking treatment. Because all variables were standardized to a 0-1 scale, the magnitudes of the coefficients can be easily compared, indicating the relative magnitude of their associations.

**IV. DISCUSSION**

The data shows that the neural network model was only a tiny bit more accurate than the logistic regression model. Although this doesn’t mean that the neural network can’t be optimized to create a greater accuracy. The low accuracy could be due to a lack of data as neural networks require lots of data to create an effective model. Additionally it could mean the combinations of neurons and layers that I tested were not the best for the data. In relation to the neural networks I tested ,about 16 iterations, the logistic model are viable in determining the effect of each dependant variable on the independent variables, as the logistic model performed just as good as the neural networks.

Analyzing the coefficient data we can see that the variable that had the most association with seeking treatment is the “work\_interference” variable. The “work\_interference\_Never” coefficient had the biggesest coefficient with -2.276094. This would imply that employees where their mental health doesn’t affect their work has a much lower chance of seeking mental health. One possible reason for the negative association is that employees might not feel the need to seek help if it doesn’t affect their work. This theory is further supported by looking at “work\_interference\_sometimes” and “work\_interference\_Often” as they have a coefficient of 0.591654 and 1.329171 respectively. This means that it is more likely for employees to seek mental health if it would sometimes interfere with work and would greatly influence(with 1.329171 being the largest positive coefficient) the likelyhood of seeking mental health if it often interfered with work. Since “work\_interference\_Never” and “work\_interference\_Often” are the largest negative and positive coefficients, it can be inferred that the most important aspect influencing employees seeking treatment is if their mental health would interfere with work.

Another large aspect influencing seeking treatment is having benefits and having care options which have a coefficient of 0.512661 and 0.495512 respectively; that said, having no benefits and no care options doesn’t decrease the chance of seeking treatment nearly as much with a coefficient of only -0.061507 and -0.189915 respectively. In fact, even the coefficient for being unsure of care options has a larger negative effect than not having care options with a coefficient of -0.305678.

**V. CONCLUSION**

In conclusion, the most important predictor of seeking help seems to be whether mental health interferes with work. Of course, companies shouldn’t change policies to make work unnecessarily strenuous, in hopes that this will prompt more employees to seek treatment. We must keep in mind that this is self-reported survey data. So, this data may show that employees who are inclined to “tough it out” by not seeking treatment are also the ones who think their work isn’t affected by mental health disorders. This may be an inaccuracy driven by the “tough it out” mindset. Instead, companies should look to policy levers they *can* control, to see if any of those could induce employees towards seeking treatment. Among employer-controlled variables, the most important variables are having health care benefits for mental health and knowing that such benefits exist. These variables were strongly associated with employees being more likely to seek mental health treatment. So, to encourage seeking treatment, companies should consider providing employees with more robust mental health benefits packages and very clear messaging that those benefits exist. Additional research could perfect the use of the AI by trying different models with the data to see if it can get a much higher accuracy than the logistic model.

**Reference**

1. *Mental health disorder statistics*. JHM. (2023, February 1). https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics
2. World Health Organization. (n.d.). *Suicide*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/suicide>
3. Johnson, R. (n.d.). (rep.). *BIMA Tech Inclusion & Diversity Report 2019*. Retrieved July 16, 2023,.
4. Joseph,V.(2020). Algorithms and Anxiety: An Investigation of Mental Health in Tech Exploring the Relationship between Mental Health in Tech and the Hardships Faced by Its Employees. <https://static1.squarespace.com/static/5c196f59b98a788f4d1b9fc5/t/5ecff928cfba127f3c11c526/1590688041301/Algorithms+and+Anxiety.pdf>
5. Publicly accessible via kaggle: https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey
6. Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.
7. Abadi, Mart&#x27;in, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … others. (2016). Tensorflow: A system for large-scale machine learning. In 12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16) (pp. 265–283).