Project 10

Predicting Mental Health in the Workplace

Kristie Kooken

Business Problem

Nearly 75.8 million Americans suffer from mental illness (about 22.8%); this represents 1 in 5 adults (www.nami.org). In a 12-month period, an employee with mental health issues can miss 12 days of work while an employee without mental illness will be absent 2 days. When looking at this finding across the United States work force, the estimated loss of productivity is $47.6 billion annually (Witters, D & Agrawal, S., 2022).

Considering the potential loss of productivity of an employee due to mental illness, it is imperative that employers offer mental health benefits as part of a comprehensive health care package as well as be supportive if any employee needs to seek out mental health services. By exploring the factors that can predict mental illness and certain attitudes about mental health in the workplace, it may be possible to help employers know how to offer their employees a set of health care tools to deal with challenges such as mental health illness.

Background/History

Mental health costs have increased a whopping 202% from 79B USD to 240B USD from 2000 to 2020 (www.statista.com). These sourcing costs suggest that affording mental health treatment can be challenging for anyone suffering from a mental health condition. Though the United States government covers about ¼ of the costs of treating mental health (www.whitehouse.gov), 75% of the costs to pay for mental health treatments come from elsewhere.

In the United States, employers with 50 or more employees must offer affordable/minimum value medical coverage (www.cigna.com) however, mental health benefits are not part of this government mandate. Likewise, though many employers do offer mental

health benefits, they may not be used by their employees. In 2021, less than 50% of U.S. adults with mental illness received treatment and the average delay between onset of mental illness symptoms and treatment was 11 years (www.nami.org).

For an employer, having their employees get mental health services is key to helping them to maintain treatment as well as maintain productivity. Likewise, researchers have found that people who are aware that their co-workers are using employer-provided mental health benefits may be more likely to use them too (www.safetyandhealthmagazine.com). By investigating what factors are the strongest predictors of mental health illness and certain attitudes towards mental health in the workplace, it may shed light on barriers to services for those suffering from mental health issues and help employers ensure their benefit packages are designed for both parties' best interest.

Data Explanation (Data Prep/Data Dictionary)

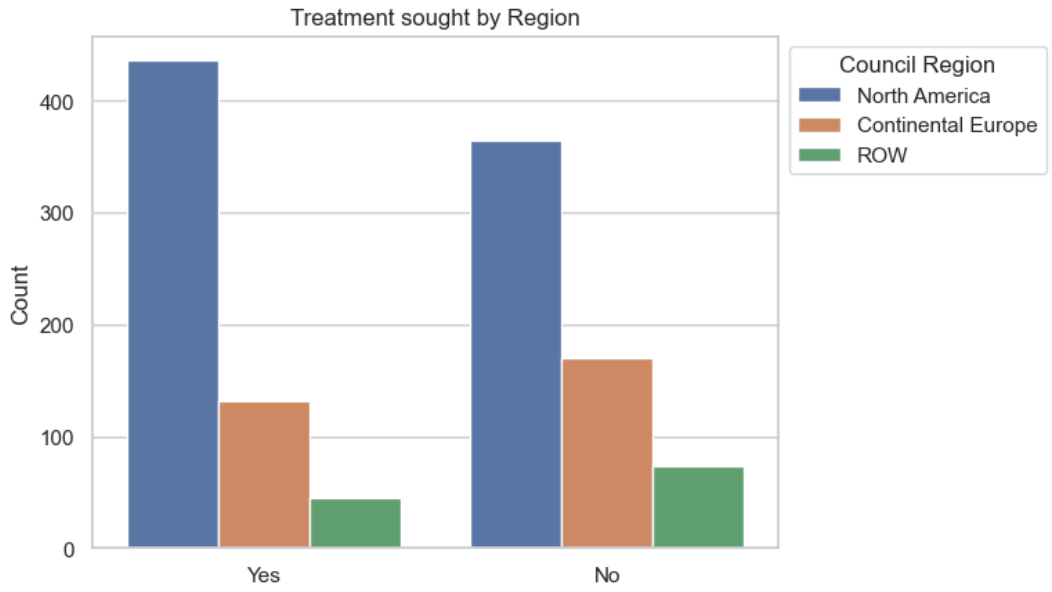
The dataset for this project has 1260 rows of mental health data from tech workers from 2014; it includes 23 questionnaire items plus 4 columns of demographics: Age, Gender, and Country and State (www.kaggle.com). The dependent variable named treatment used for this analysis was "have Sought treatment for mental health conditions" with values of Yes/No, this variable was well balanced in the dataset (Yes, N = 613, No, N = 607). All other variables were independent variables and included both dichotomous response and Likert scale response variables. All data cleaning and analyses were conducted in python using Jupyter Notebook. The data dictionary for these 27 variables is found in Appendix 1.

Methods

Data was cleaned, recoded as needed to single values per category (e.g., different spellings of Male were all collapsed to one spelling of Male). Eighteen records of missing values for the variable "self-employed" were dropped. Likewise, for the variable "work_interfere", counts of

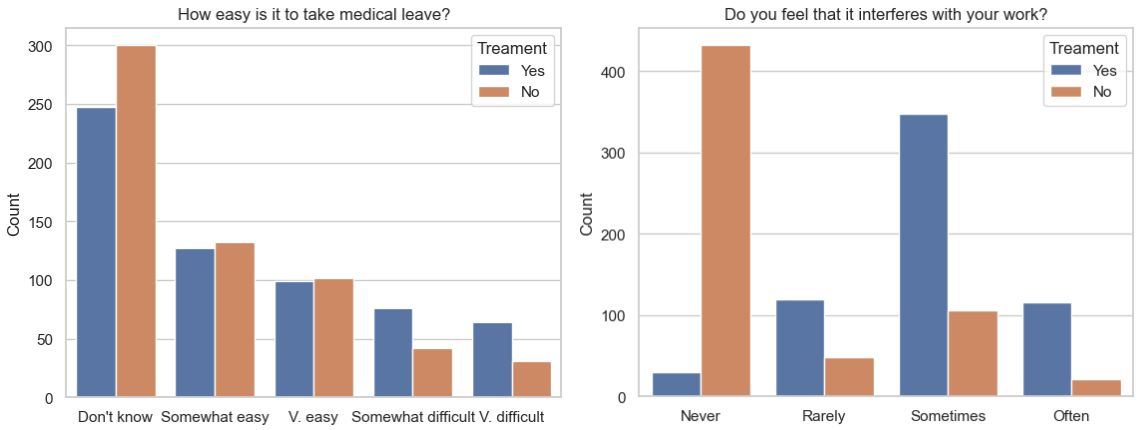
missing were replaced to the highest frequency within treatment. So, those with treatment = “No” were replaced with “Never” and those with treatment = “Yes” were replaced with “Sometimes”. Both Country and State variables were grouped to regions with Country then further aggregated to “North America”, “Europe” and “ROW”. Analyses were run using “Council Region” only. Duplicate values were dropped. The “Age” variable was trimmed to include only aged 18 to 99 years old. The resulting dataset had 1220 records. Exploratory data analysis was conducted to identify variables with any skew or unexpected distributions. All variables were found to be as expected. Figure 1 below shows treatment by Council Region.

Figure 1.



The data was also explored to examine the relationship between treatment and how easy it is to take medical leave as well as if mental health interferes with work (Figure 2).

Figure 2.



Analysis

Categorical variables were dummied for analysis (removing n-1 to avoid collinearity) and the single continuous variable was standardized as these data were skewed with most subjects aged 20 to 40 years old. “Timestamp” was dropped from analysis as well as the original “Country”, “State” variables and only the grouped country variable along with others remained. The data was split to train and test and base modeling was performed on these data using logistic regression which yielded an acceptable model. To explore reducing the number of features, chi-square analyses were conducted on all categorical dummied variables (the variable “Age” was dropped off for this step). Chi-square values were generated for each feature and those meeting the < 0.2 criterion were dropped from analysis. This resulted in 10 out of the 44 dummied categorical features being dropped. Another logistic regression was run on these data which also yielded an acceptable model with modest improvement over the base model. K-nearest neighbor classification model was also run on these data to see if this would yield a better model; it did not

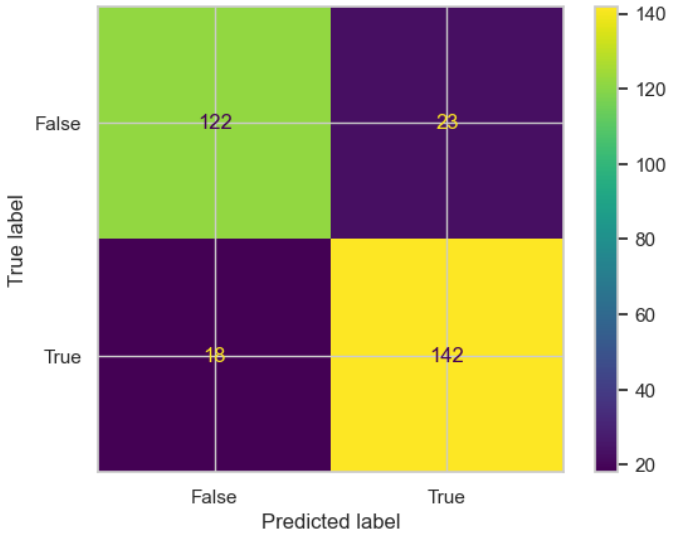
though the model was acceptable. Likewise, random forest classification model was also explored and did yield a better model than knn but not logistic regression. Because of this, final model tuning was performed using logistic regression. Upon inspection, the model results for hyper-tuning did not yield a better model and thus, it was assumed this was due to overfitting. Because of this, the reduced features logistic regression model had the best results and was deemed the final model. The results from these models are presented in Table 1.

Table 1. Model results with reduced features.

Model	Accuracy	Predictor value	Precision	Recall	F1-score	Support
Logistic reduced	0.87	0	0.87	0.84	0.86	145
		1	0.86	0.89	0.87	160
Logistic hyper-tuned	0.86	0	0.86	0.83	0.85	145
		1	0.85	0.88	0.87	160
Random Forest	0.84	0	0.84	0.83	0.83	145
		1	0.85	0.86	0.85	160
K-nearest neighbor	0.70	0	0.63	0.88	0.74	145
		1	0.83	0.53	0.65	160

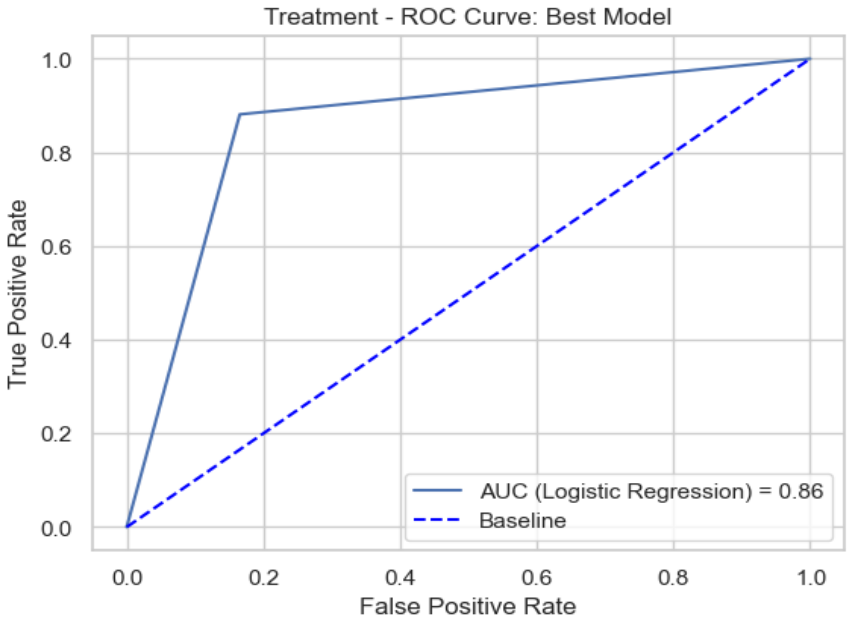
The confusion matrix is a visualization of the true negatives, false positives, false negatives, and true positives which are the correct and incorrect number of predictions for each class. This matrix shows 122 true negatives, 23 false positives, 18 false negatives and 142 true positives.

Figure 3.



The ROC curve is used to show how the logistic regression model discriminates between the positive and negative classes at different classification thresholds. The ROC plot shows a line above the 0.5 threshold, with the AUC score of 0.86 indicates an acceptable ability to discriminate between the positive and negative classes.

Figure 4.



Conclusion

This project investigated factors that can predict mental illness and certain attitudes about mental health in the workplace, the final logistic regression model had a relatively high accuracy score and demonstrated good recall, F1 and precision scores. Important features that predicted treatment for this model include “interfere with work” (all levels of the variables), “coworkers”, “family history”, “care options” and “benefits”. The results from this model are very compelling as they started to tell a story that is often not heard where the employee has concerns if their mental health issues will interfere with work, the willingness to share their situation with a coworker, if they have a family history of mental illness, what care options there are and if the employer provides benefits. Employers need to be diligent in letting their employees know about what mental health benefits are available and having an encouraging and open environment that welcomes employees getting additional or outside support when they believe they need it or are advised to get it.

Assumptions

The data for this project did not list how it was collected or where the scale was developed and if the scale demonstrated sufficient psychometric properties to be used in a formal setting. Another assumption is that all different statistical models available were tested and the final model is the most appropriate model. Likewise, it is assumed that the results from this survey of tech workers are applicable to other groups when other groups may not have as many of the benefits offered to them.

Limitations

Limitations include not knowing enough information about the sample or not having more data. Likewise, about 700 subjects in this dataset were from the United States thus this may be a

sample that is more representative of how the United States views mental health and the services that are common here to support such issues.

Challenges

The challenges are that I have a limited amount of time to investigate this data and there is a potential that I am not looking at this analysis fully or seeing if I could find additional data where I could have cross-validation with a similar scale. Likewise, I am assuming the scale has been conducted to measure something which is not clear based on the information available at the kaggle.com site. This type of modeling work to answer what I am interested in exploring would be a much larger project and would take months and possibly years to fully develop a working, robust model.

Future Uses/Additional Applications

Further work is needed to better identify a full set of covariates to predict mental health. Likewise, it would be beneficial to have similar scales administered to the same population to see if the results correlate, this would strengthen the usage of the scale as well as have a more robust picture of the types of support that are needed by different types of employees and workers and what challenges they perceive prevent them from seeking mental health treatment as well as encourage them to seek mental health treatment.

Recommendations

Recommendations include getting more survey data from different years as well as similar surveys to build a more robust model and to have a deeper investigation of the performance of different independent variables included in this model.

Implementation Plan

Currently, there is no implementation plan as this research effort is preliminary and additional data and modeling is needed to develop a robust set of characteristics for mental health or certain attitudes about mental health in the workplace.

Ethical Assessment

Ethical considerations for this data include that I was not able to verify the source of this data and thus, it is not clear how this data was collected or if this survey was ever assessed for psychometric reliability and validity. Likewise, results found with this analysis would need to be verified using a scale that demonstrated reliability and validity to ensure that results from the scale were measuring what they are intended to measure. Further, this scale is mostly tested on tech workers and would need to expand to include workers in many different types of settings in order to have a representative sample. Mental health is an important topic and the results found here would need to be replicated in order to give employers the best chance to help their employees (which of course, would then help the employer).

References

Employer Mandate (accessed 2024, May 12). Employers must offer health insurance or pay a penalty. Cigna Healthcare. https://www.cigna.com/employers/insights/informed-on-reform/employer-mandate#:~:text=Employer%20mandate%20overview,known%20as%20the%20employer%20mandate.

Mental Health By the Numbers (accessed 2024, May 12). About Mental Illness. https://www.nami.org/about-mental-illness/mental-health-by-the-numbers/#:~:text=22.8%25%20of%20U.S.%20adults%20experienced,represents%201%20in%2020%20adults.

Mental Health in Tech Survey (accessed 2024, May 12). Survey on Mental Health in the Tech Workplace in 2014. https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey/code

Reducing the Economic Burden of Unmet Mental Health Needs (accessed 2024, May 19). The White House. https://www.whitehouse.gov/cea/written-materials/2022/05/31/reducing-the-economic-burden-of-unmet-mental-health-needs/#:~:text=The%20Federal%20Government%20covers%20some,from%20the%20U.S.%20Medicaid%20program.

Total U.S. expenditure for mental health services from 1986 to 2020 (accessed 2024, May 19). State of Health - Statista. https://www.statista.com/statistics/252393/total-us-expenditure-for-mental-health-services/.

Witters, D., & Agrawal, S. (2022, December). The Economic Cost of Poor Employee Mental Health. Gallup. https://www.gallup.com/workplace/404174/economic-cost-poor-employee-mental-health.aspx

Appendix 1.

#	**Variable name**	**Variable description**
1	Timestamp	Timestamp of survey
2	Age	Participant age
3	Gender	Participant gender
4	Country	Participant country
5	State	If you live in the United States, which state or territory do you live in?
6	self_employed	Are you self-employed?
7	family_history	Do you have a family history of mental illness?
8	treatment	Have you sought treatment for a mental health condition?
9	work_interfere	If you have a mental health condition, do you feel that it interferes with your work?
10	no_employees	How many employees does your company or organization have?
11	remote_work	Do you work remotely (outside of an office) at least 50% of the time?
12	tech_company	Is your employer primarily a tech company/organization?
13	benefits	Does your employer provide mental health benefits?
14	care_options	Do you know the options for mental health care your employer provides?
15	wellness_program	Has your employer ever discussed mental health as part of an employee wellness program?
16	seek_help	Does your employer provide resources to learn more about mental health issues and how to seek help?
17	anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
18	leave	How easy is it for you to take medical leave for a mental health condition?

#	**Variable name**	**Variable description**
19	mental_health_consequence	Do you think that discussing a mental health issue with your employer would have negative consequences?
20	phys_health_consequence	Do you think that discussing a physical health issue with your employer would have negative consequences?
21	coworkers	Would you be willing to discuss a mental health issue with your coworkers?
22	supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?
23	mental_health_interview	Would you bring up a mental health issue with a potential employer in an interview?
24	phys_health_interview	Would you bring up a physical health issue with a potential employer in an interview?
25	mental_vs_physical	Do you feel that your employer takes mental health as seriously as physical health?
26	obs_consequence	Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
27	comments	Any additional notes or comments