Project 10

Strongest predictors of mental health illness in the workplace

Kristie Kooken

May 12, 2024

The topic that I will investigate for this project will be exploring what factors are the strongest predictors of mental health illness and certain attitudes towards mental health in the workplace.

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).

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). 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.

Dataset:

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). Potential dependent variables include Self-employed, Family history and have Sought treatment for mental health conditions. Independent variables include both dichotomous response and Likert scale response variables.

Method:

The data will be cleaned, normalized as needed and duplicates dropped. Descriptive statistics and frequencies will be run on all data to ensure there are no outliers and that the distributions are as expected. Exploratory data analysis (EDA) will include different visualizations of these data, looking at the demographic variables in context of each survey question. Statistical modeling will be performed on these data using logistic regression, k-nearest neighbor, support vector machine and potentially decision trees in order to determine the best model for these data. I am hoping to be able to examine at least two dependent variables of Family history and Sought treatment.

Ethical considerations:

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 exercise 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).

Challenges/Issues:

Issues that could arise are that the classification modeling doesn't go as planned and I will have to determine if the data source is viable. Likewise, I could face challenges with interpretability of the data – meaning the sample is not very large and potentially the survey questions were not properly designed and only have face validity. 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.

References:

To validate the results with outside resources, I would need to investigate to see if another institute had created a similar scale or if one of the larger tech companies had done a similar scale in hopes of creating a great workplace by offering great benefits to attract top talent. Whereas different government data could be helpful, the best way would be to find a similar type of effort (workplace questions around mental health) and determine if the results from the two surveys correlated to each other.

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

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