Isabella Delionado
Brown University
Github: https://github.com/IDelionado/data-project.git

Mental Health in the Tech Workspace

I. Introduction

The goal of my project was to explore attitudes towards mental health in the tech workspace and how these attitudes affect an individual's likelihood to receive professional treatment. My dataset was from the Open Science Monitoring Initiative. Its goals are to fight stigma around mental health disorders by speaking openly about experiences and educate the tech community on the economic impact of mental disorders and how it affects worker productivity. I used their Mental Health Tech Survey from 2016 which includes about 1400 responses measuring mental health in the tech workspace. The primary difficulty of the dataset was missing values due to unanswered survey questions.

II. EDA

My project dealt with a classification problem. My target variable was the question "Have you ever sought treatment for a mental health issue from a mental health professional?"

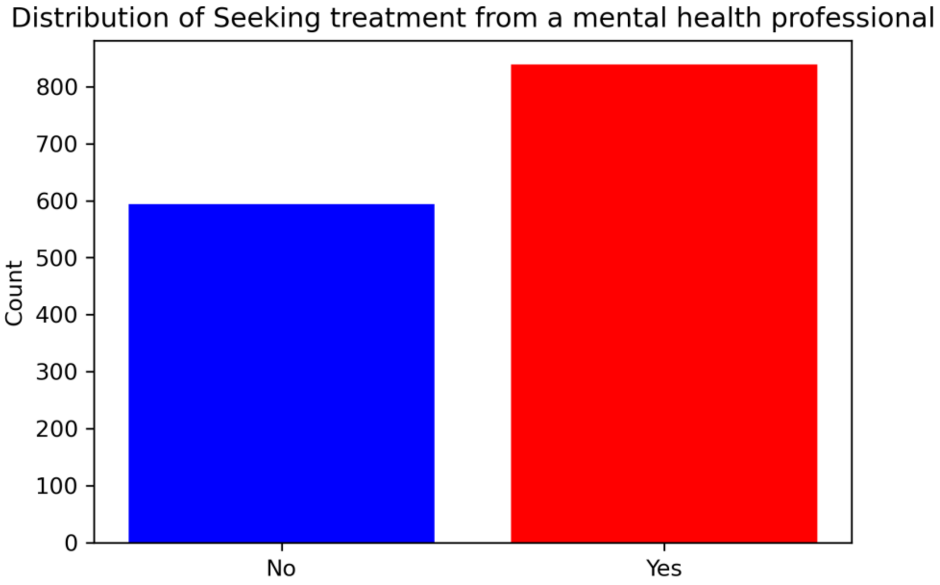


Fig. 1. The plot of responses to the target variable "Have you ever sought treatment for a mental health issue from a mental health professional?"

The data were primarily categorical and only had one continuous feature and one ordinal feature. The free response columns "Why or why not?" were dropped since they were not necessary for the goal of the project. As a result, 61 columns of questions and 1433 rows of responses were used.

As previously stated, the primary difficulty of the dataset was the missing values. There were no missing values in the ordinal and continuous features. However, the categorical features were missing anywhere from 20.03% to 89.95% of responses to a particular question, while not all categorical features had missing values. The frequency of the missing values was random and did not correspond to a certain type of question.

As reflected in Fig. 2., most of the survey questions did not have binary answers. Responses to the target variable (seeking professional treatment) and having a mental health disorder are consistent; the majority of "No" responses received no treatment and the majority of "Yes" responses received treatment. Additionally, the "Maybe" responses are more evenly split, while having sought treatment accounts for slightly more.

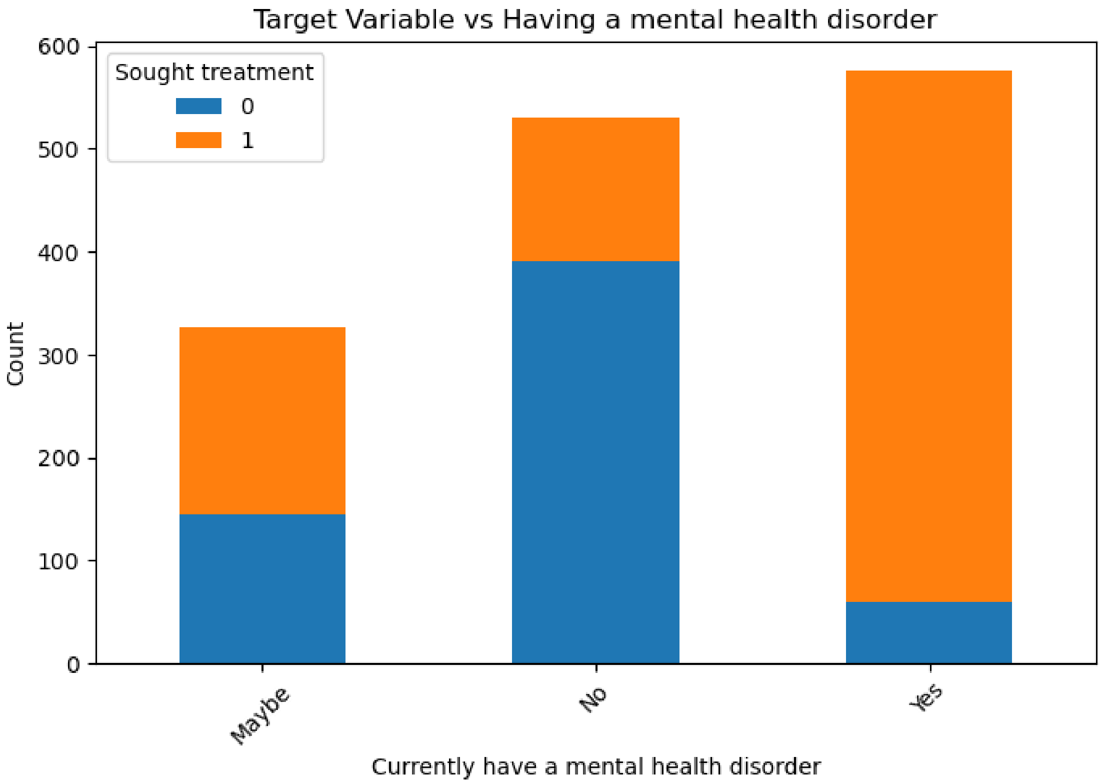


Fig. 2. The plot of responses to "Do you currently have a mental health disorder?" compared against the target variable.

When comparing the continuous feature age against the target variable (Fig. 3), we can see that age does not have a significant impact on seeking treatment as evident by the similar shape of the violin plots.

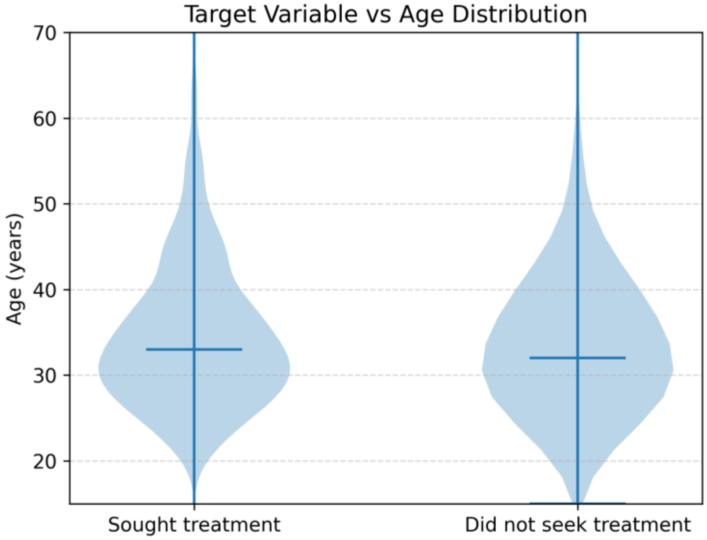


Fig. 3. The plot of responses to “What is your age?” compared against the target variable.

While the responses were generally consistent between questions, there were some noteworthy discrepancies. For example, in Fig. 4 when asked “Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?”, around 450 responded that they would be comfortable while about 400 responded that they might be comfortable. However, when asked “Do you feel that being identified as a person with a mental health issue would hurt your career”, just under 600 respondents said that they think it might hurt their career and a little less than that believe that it would (Fig. 5). It is logically inconsistent that respondents could be comfortable discussing mental health with a supervisor while simultaneously not wanting to be identified as an individual with a mental health disorder.

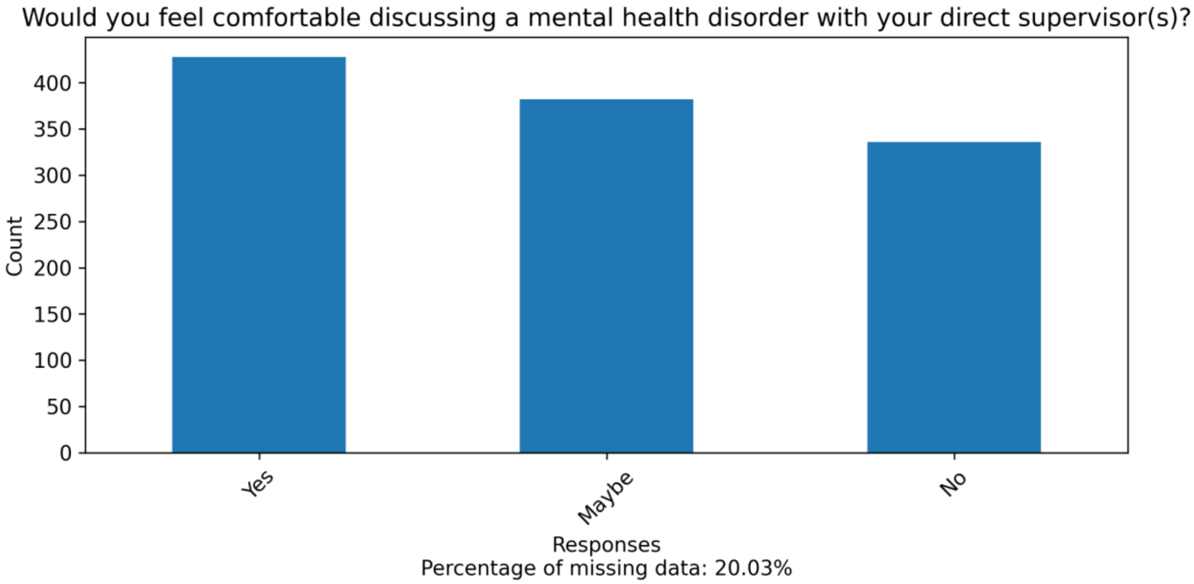


Fig. 4. The plot of responses to "Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)"

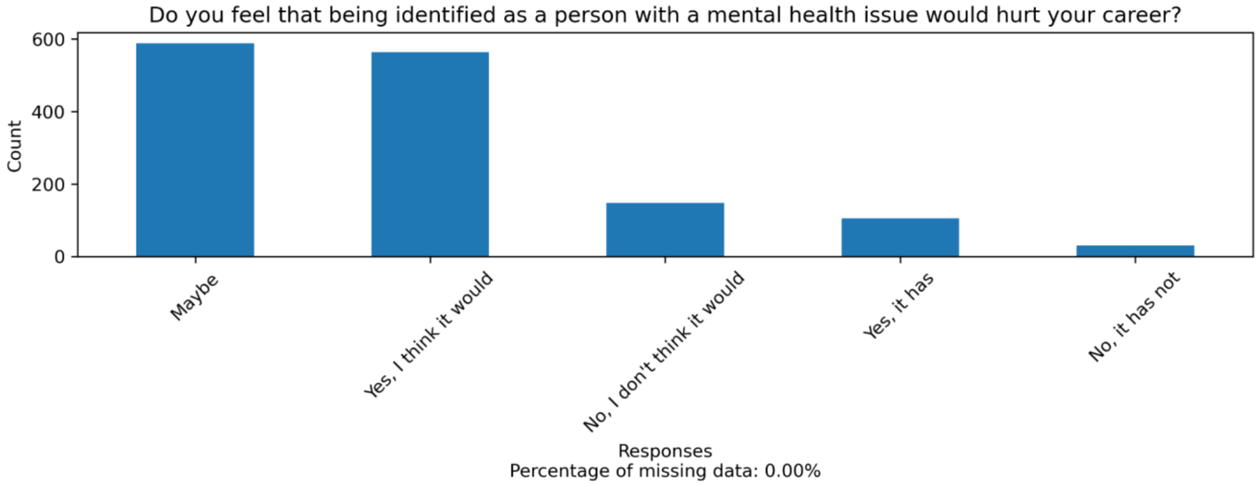


Fig. 5. The plot of responses to "Do you feel that being identified as a person with a mental health issue would hurt your career?"

III. Methods

I performed a K-fold split with five folds since the dataset was smaller (about 1400 entries). I already had missing data and wanted to prevent more data leakage, so it was better for the model performance to be averaged across all k iterations. I used ordinal, one hot, and standard scalar encoders for the different features. Before preprocessing, I had 1433 data points (rows) and 61 features (columns). After preprocessing, my ordinal encoder produced 1433 data points (rows) and 1 feature (column). My one hot encoder produced 1433 data points (rows) and 1093 features (columns). My standard scalar encoder produced 1433 data points (rows) and 1 feature (columns). The total was 1433 data points and 1095 columns.

The missing data were handled by the encoders implicitly. While there were no missing values for the ordinal feature, I still set handle_unknowns to "use_encoded_value" and "unknown_value = -1" to assign a placeholder value of -1 to any missing or unseen categories. For my continuous feature, any potential missing values were handled by Standard Scalar since NaN values are automatically ignored during fitting and scaling. For my categorical features, where I actually had missing values, I first converted missing values to strings then set the One Hot encoder parameter "handle_unknown = "ignore"".

For my Machine Learning algorithms, I used Logistic regression, K-Nearest Neighbors, Decision Tree, Random Forest, and XGBoost, with early stopping, and tuned the following hyperparameters (Fig. 6):

ML Algorithm	Hyperparameters
Logistic Regression	model__penalty: 'l1', 'l2' model__C: np.logspace(-3,3,7) model__solver: 'saga' model__max_iter: 10000
K-Nearest Neighbors	model__n_neighbors: 3, 5, 7, 9 model__weights: uniform, distance model__p: 1, 2
Decision Tree	model__max_depth: None, 5, 10, 20 model__min_samples_split: 2, 10 Model__min_samples_leaf: 1,4
Random Forest	model__max_depth: None, 15 model__min_samples_split: 2, 5, 10 model__n_estimators: 100, 200, 300
XGBoost	model__learning_rate: 0.03 model__max_depth: 3, 5, 10 model__colsample_bytree: 0.5, 0.7, 0.9 model__subsample: 0.5, 0.66, 1.0

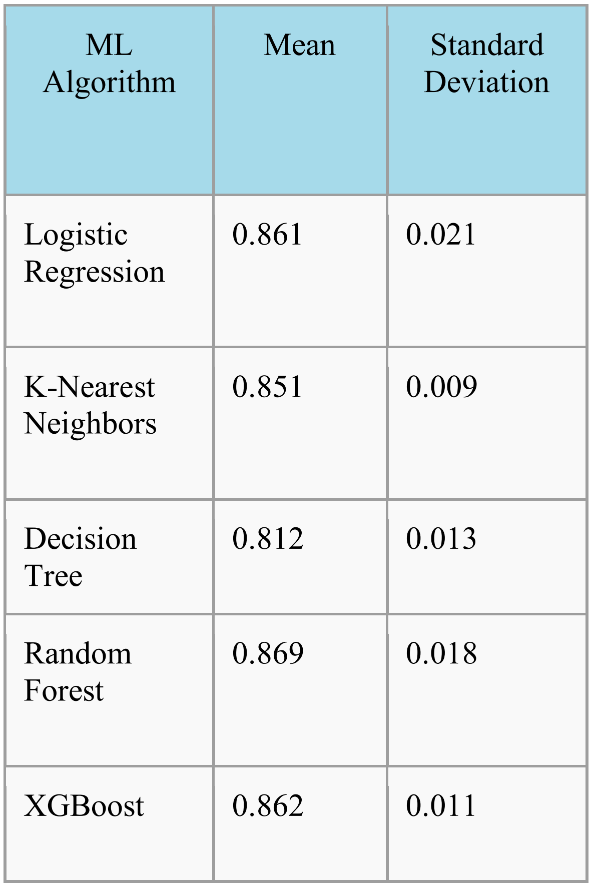
Fig. 6. Table summarizing the ML algorithms and hyperparameters used.

I also used GridSearchCV and set cv to five. I looked at the validation accuracy score for each ML algorithm across the five folds.

.

IV. Results

The means and standard deviations of the different ML algorithms are summarized in Fig. 7 below. Random Forest had the highest mean of 0.869 and K-Nearest Neighbors had the lowest standard deviation of 0.009.



ML Algorithm	Mean	Standard Deviation
Logistic Regression	0.861	0.021
K-Nearest Neighbors	0.851	0.009
Decision Tree	0.812	0.013
Random Forest	0.869	0.018
XGBoost	0.862	0.011

Fig. 7. Table summarizing the means and standard deviations for the various ML algorithms.

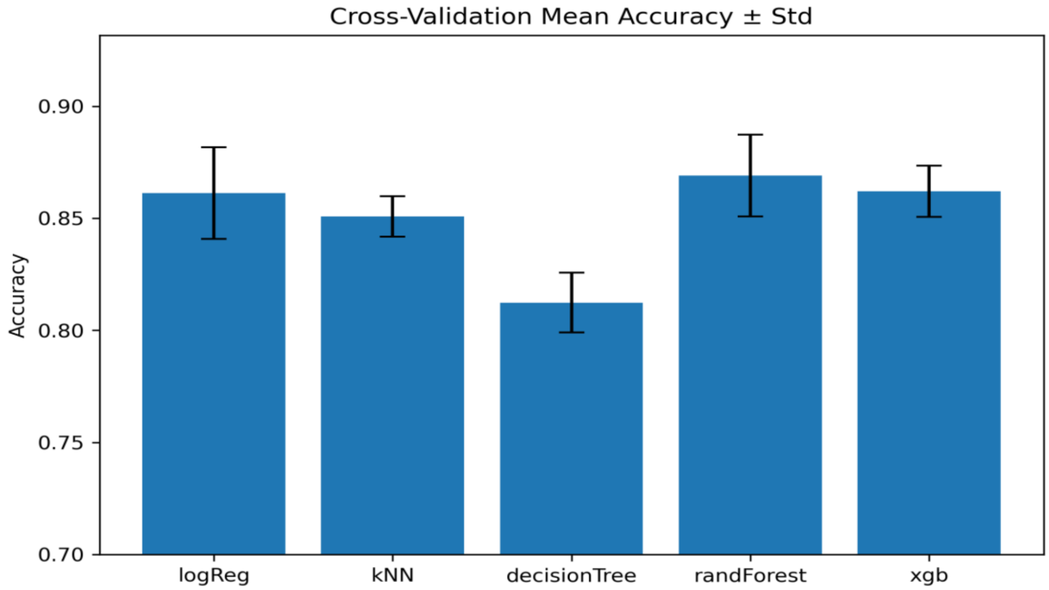


Fig. 8. The plot summarizing the means and standard deviations for the various ML algorithms.

The best model was Random Forest which had a test accuracy of 0.864 which is higher than the baseline accuracy of 0.602 so the model improves over the baseline by about 26.2%. It is also evident that the model performed well, with a true negative rate of 0.83 and a true positive rate of 0.88, indicating that it is able to correctly identify negative and positive cases the majority of the time (Fig. 9). The best parameters for Random Forest were max_depth = 15, min_sample_split = 10, n_estimators = 200.

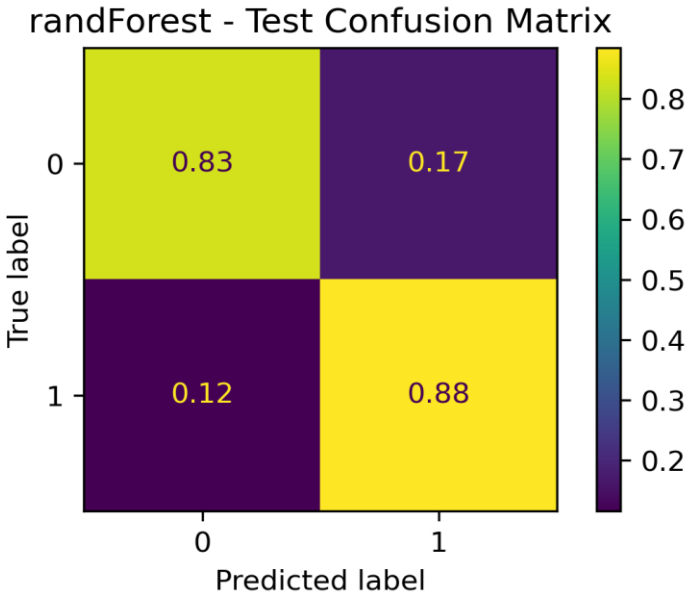


Fig. 9. The confusion matrix for the Random Forest model.

Permutation feature importance was used to measure how much the model's accuracy decreases when each feature is randomly permuted, determining the model's reliance on features for prediction. The top 10 features in the permutation importance plot (Fig.10) are those whose permutation causes the largest decrease in accuracy.

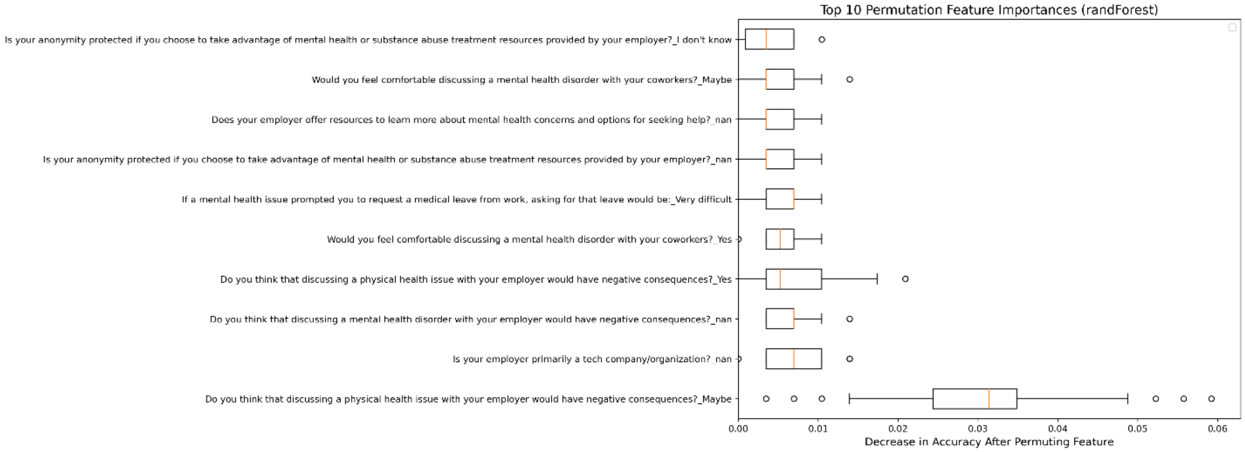


Fig. 10. The plot of the top 10 permutation feature importances.

SHAP was used to quantify the contribution of individual features to the model's prediction (Fig. 11). The three most influential features were related to a prior mental health diagnosis. "Have you been diagnosed with a mental health condition by a medical professional?" with a "Yes" response contributing positively to the output, suggesting that having been diagnosed contributes to an individual's likelihood to seek professional help. Not responding to specify what conditions the respondent might have contributes negatively to the output. This suggests that not being diagnosed with a condition or not knowing what condition one has makes a respondent less likely to seek treatment. A "No" response to "Have you been diagnosed with a mental health condition by a medical professional?" is the third most impactful to the output and contributes negatively to an individual's likelihood to seek professional help.

Putting aside responses relating to interfering with the respondent's work, the only features that were workplace specific were, first, being aware or not being aware of the mental health care provided by previous employers and second, having a bad experience at the previous or the current workplace ("Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?"). Being aware contributes positively to the output. An individual is more likely to seek professional treatment while being unaware contributes negatively, while an individual is less likely to seek professional treatment. Having a bad workplace experience (whether past or current), contributes positively to the output suggesting that respondents are more likely to seek professional help.

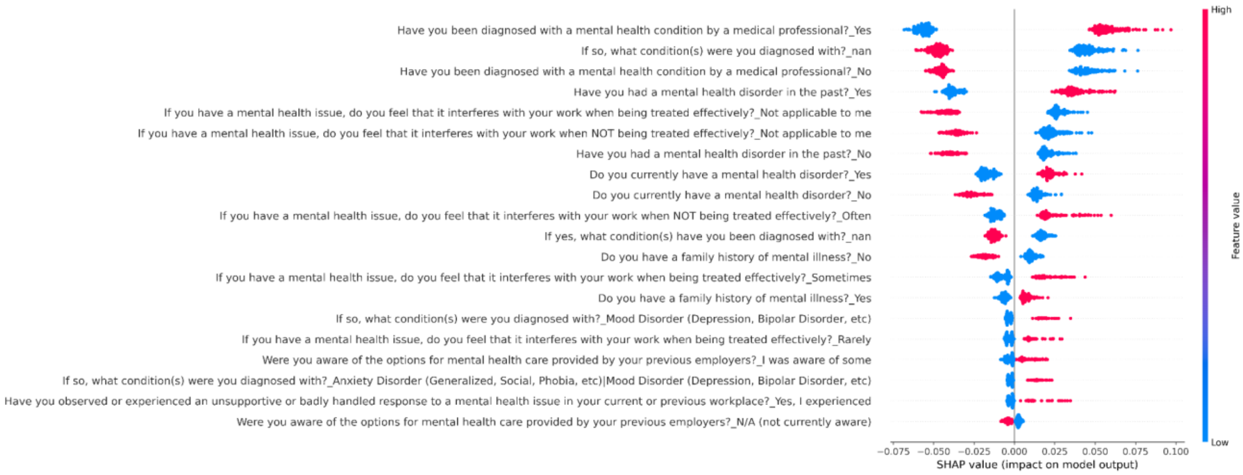


Fig. 11. SHAP summary plot of the top 20 features contributing to the prediction.

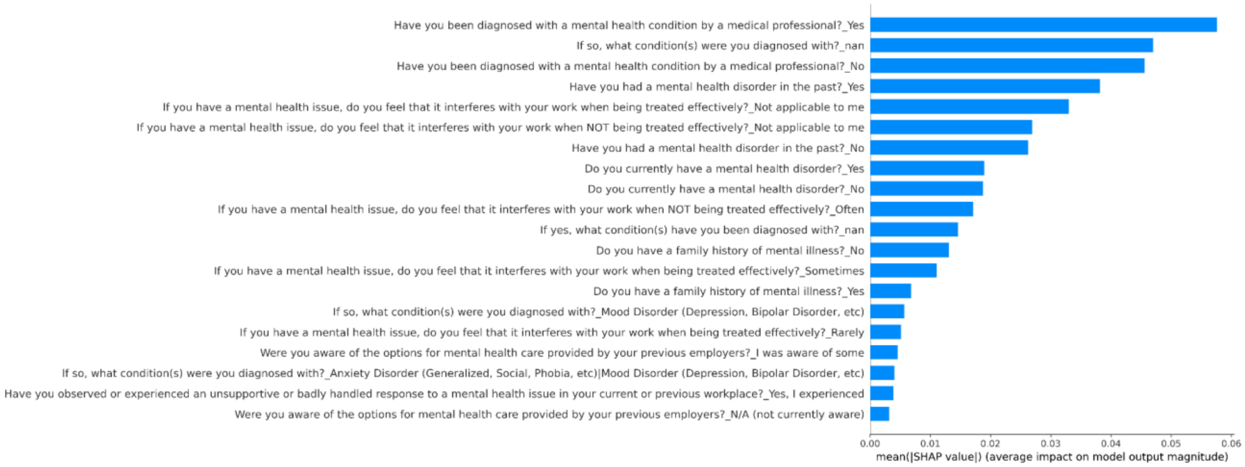


Fig. 12. Mean absolute SHAP value plot of the top 20 features contributing to the prediction.

Top features between permutation importance and SHAP values may differ because the permutation importance measures how the model's overall performance relies on a feature while SHAP values measure the individual contribution of a feature to the prediction. Features that influence the prediction may not be as crucial for the overall accuracy which explains the different feature rankings. Correlated features can also distort permutation importance while SHAP is able to capture their contribution.

The Top-10 feature importance plot shows results consistent with the SHAP value plots.

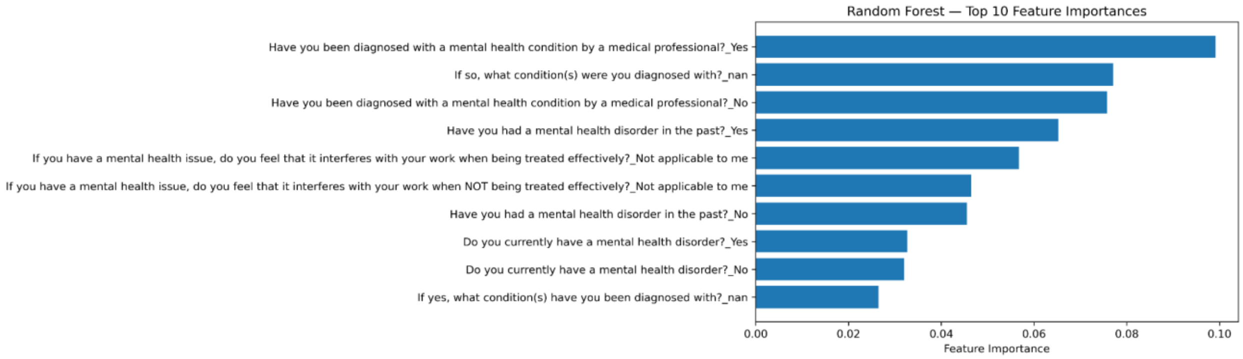


Fig. 13. The plot of top 10 feature importances.

V. Outlook

The model performed well overall. The features contributing most to an individual's likelihood to seek professional treatment were primarily related to the respondent having a prior mental health diagnosis, a history of mental illness, or whether being treated effectively interfered with their work. As mentioned previously, the only features that were workplace specific were being aware or not being aware of the mental health care provided by previous employers and also about a bad experience in the current or previous workspace. This indicates that attitudes towards mental health in the workspace may not have a large impact on an individual's likelihood to seek professional treatment. To better understand the effect of workplace attitudes, we could train the model using only responses to workplace-specific questions. This would isolate their contribution and might improve predictive power for that aspect of the prediction.

VI. References

Data Source: https://www.kaggle.com/datasets/osmi/mental-health-in-tech-2016