Mental Health in the Tech Industry: Analysis and Insights

Laura Abood
axu8rx

Kelly Gu
hnj4jk

Edison Huang
hmc4zu

Alisha Qian
sta2fu

Carlie Stewart
ayu6cp

Hana Wang
muf5cb

December 10, 2024

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

GitHub Repository: https://github.com/kkellygu/ds3001Project/tree/main

1 Project Context and Goals

The tech industry is known for its fast-paced and demanding environments, which can significantly impact mental well-being. As future professionals in this field, understanding the mental health landscape is crucial to fostering healthier workplaces. Using the Kaggle dataset "Mental Health in Tech 2016," this project aims to identify trends and patterns related to mental health among tech employees, examine how workplace conditions and resources influence mental health outcomes, provide actionable insights to improve employee well-being, and reduce stigma surrounding mental health in the tech sector.

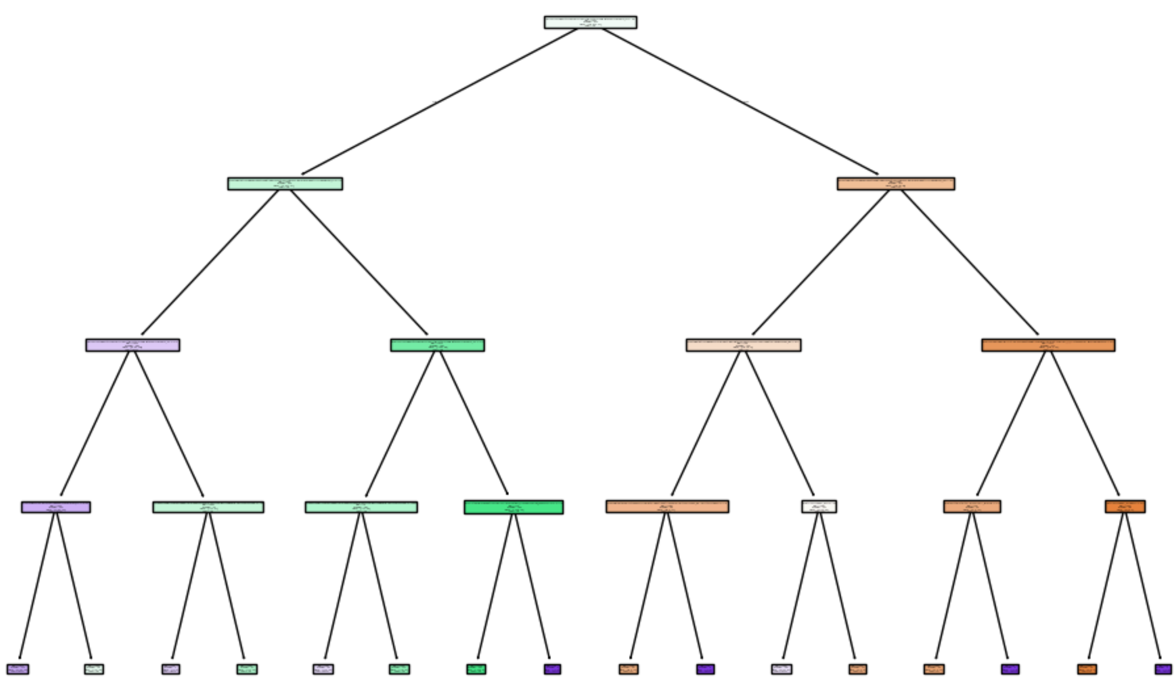
The target variable we are predicting is “Would you feel comfortable discussing a mental health disorder with your direct supervisor(s)?” Our predictive question asks, “Can workplace mental health resources predict the self-reported likelihood of employees seeking help for mental health issues within the tech industry?”

2 Results

Model Performance Overview

Model	Accuracy	Class	Precision	Recall	F-1 Score
Logistic Regression	0.60	No (0)	0.61	0.55	0.58
		Yes (1)	0.67	0.72	0.69
		Maybe (2)	0.50	0.51	0.51
Decision Tree	0.58	No (0)	0.69	0.46	0.55
		Yes (1)	0.64	0.73	0.68
		Maybe (2)	0.46	0.53	0.49
Random Forest	0.63	No (0)	0.67	0.72	0.70
		Yes (1)	0.64	0.64	0.69
		Maybe (2)	0.54	0.41	0.47

Decision Tree



Overall, the logistic regression model provided balanced performances across all the classes with an accuracy of 0.60. It performed reasonably well for “Yes” predictions with an F-1 score of 0.69, and the predictions for “Maybe” were slightly lower with an F-1 score of 0.51. This suggests room for improvement when predicting more ambiguous responses. As a baseline model, the logistic regression provided a suitable baseline. The decision tree model had an accuracy of 0.58, which was slightly lower than the logistic regression. It achieved its highest F-1 score of 0.68 with the “yes” class, but performed worse for the “maybe” class with 0.49. Lastly, the random forest model outperformed both the logistic regression and the decision tree model with an accuracy of 0.63. Its strongest performance was for the “no” class, which had an F-1 score of 0.70, due to the high recall value of 0.72. However, it struggled with the “maybe” class, which provided the lowest F-1 score out of all the models at 0.47. Despite the overall higher accuracy for the random forest class, its difficulties in predicting the “maybe” class suggest that ensemble learning methods alone are not sufficient to address the complexity of ambiguous responses. However, this framework performed best for our outcomes of interest.

Feature Importance Overview

Feature	**Importance**
Do you think that discussing a mental health disorder with your employer would have negative consequences? (No)	0.106479
Do you think that discussing a mental health disorder with your employer would have negative consequences? (Yes)	0.068377
Do you feel that your employer takes mental health as seriously as physical health? (Yes)	0.060997
Do you know the options for mental health care available under your employer-provided coverage? (No)	0.050659
Have you ever sought treatment for a mental health issue from a	0.050477

mental health professional?	
Do you work remotely? (Sometimes)	0.046187
Do you have a family history of mental illness? (Yes)	0.046149
Does your employer offer resources to learn more about mental health concerns and options for seeking help? (No)	0.045763
Do you feel that your employer takes mental health as seriously as physical health? (No)	0.044742
Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer? (Yes)	0.043685

Based on feature importance, we found that the most important feature by far was: “Would you feel comfortable discussing a mental health disorder with your coworkers?_No”, followed by: “Do you think that discussing a mental health disorder with your employer would have negative consequences?_No”, and finally: “Would you feel comfortable discussing a mental health disorder with your coworkers?_Yes”. Thus, the effect of feeling comfortable discussing mental health with coworkers is very highly correlated with feeling comfortable discussing mental health with supervisors. This is likely because coworkers contribute a lot to the workplace environment. With a work environment that feels unsafe and doesn’t foster open communication, it is unlikely that an employee would be comfortable being vulnerable to their supervisor.

Another way to look at this is that the supervisor also sets the tone for the team, so the employees on the team will likely follow this dynamic. Looking at the second most important feature, it also makes sense that people who believe that discussing mental health with employers will have negative consequences would not feel comfortable discussing these issues with their direct supervisors perhaps for fear of putting their jobs in jeopardy or just because of the associated stigma. Therefore, it seems that the thing that contributes most to an open discussion about mental health in the workplace is the team environment (accounting for both coworkers and employers/supervisors) and psychological safety, knowing they are cared for. Other issues like whether or not the company has good healthcare benefits were not deemed as important by the model. Again, this makes sense as talking to a supervisor about mental health does not necessarily mean the person will be required to receive treatment for it.

3 Limitations and Challenges

While our model performed with accuracy scores ranging from 0.58 to 0.63, they had a low F1 score averaging at 0.60. This means that while the models were somewhat effective at classifying mental health outcomes, they may face challenges balancing precision and recall, likely due to data imbalances and feature complexities. The categorical nature of the data primarily contained binary yes/no responses, which may oversimplify complex workplace and mental health dynamics, limiting the model's ability to adequately capture nuanced relationships. Additionally, there was a lot of missing data which may have affected the model's ability to make predictions. For our purposes, our model could be used as a starting point for future research. Another limitation was the possibility of multicollinearity among features specific to workplace conditions, such as employer support and mental health resources. These variables may potentially overlap and reduce the distinct predictive power of individual features. Generalizability was also limited because the dataset primarily represented tech industry workers from specific regions, meaning the findings may not generalize across different industries or countries.

4 Conclusion and Recommendations

The results of this analysis indicate that while logistic regression, decision tree, and random forest models provide helpful insights into identifying trends related to mental health patterns among employees, the performance of this project fell below the desired threshold of an F-1 score of >0.85. This suggests that further improvements in data preprocessing and feature engineering are required to make up for any missing data that may be affecting the model's predictions. While our accuracy is low, the purpose of our project does not warrant a very accurate prediction. We can use our model as a benchmark to give employers base insights as to whether their employees feel comfortable discussing mental health with them and what employers can do to improve in that regard.

From the results of this project, the recommendation that we came up with for moving forward was for employers to implement workplace initiatives that focus more on fostering a supportive culture of trust and openness where discussing mental health is normalized and supported. Additionally, employers should increase access to awareness of mental health resources, which can enhance employee comfort in seeking help. To address the ambiguity in our predictions for "Maybe" responses, incorporating additional qualitative data, like interviews or surveys, could increase the predictive accuracy of future analyses. It could also be more insightful to obtain quantitative metrics to further support our efforts to measure employees' affinity for seeking mental health resources.

By combining predictive insights with actionable recommendations, this project offers many ways for tech companies to foster an environment where the mental health of their employees is prioritized and recognized. When mental health is taken as seriously as physical health, the workplace environment ultimately progresses toward a healthier, more productive workforce, fostering more harmony between leadership and employees.

This project can be used as a starting point for future research. With improved data collection, the model can help employers determine if their employees feel that they are in a safe environment that encourages open discussions about mental health, and if not, how they can improve on achieving this goal.