

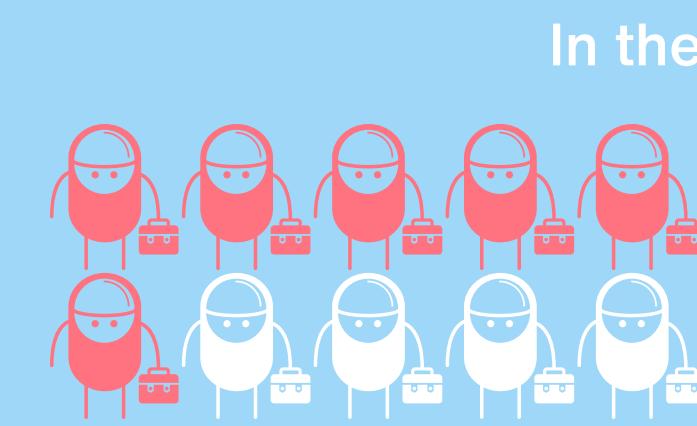
Modeling Indicators of Mental Health Openness in Tech Workspaces

Abdallah AbuHashem, Grace Hu, Crystal Zheng. Mentored by: Hancheng Cao

Motivation



In over 80% of the cases, treatment is effective, but it takes **8-10 years** for individuals to seek proper help



217 mil workdays lost due to depression each year

Problem

What **factors** make a person **more likely** to discuss their own mental health in the work space?

Data

We obtained, preprocessed, and standardized OSMI data collected via surveys conducted between 2014-2018^[2]. The data included personal and workplace features. Our target was predicting whether a person is willing to open up about mental health to the supervisor and other coworkers.

Methods

In this supervised learning classification problem, the focus was to achieve good classification results and interpretable coefficients of the features. As a result we chose Logistic Regression as a baseline. For better classification results, we also used models from table 1.

Table 1: Classifier models

Models		
Mixed	KNN	SVM
1. Log Reg	10. Minkowski	18. Linear
2. Lin Dis Analysis	11. Chebyshev	19. RBF
3. Gaussian Process	12. Euclidean	20. Sigmoid
4. Random Forest	13. Manhattan	21. Poly
5. Multi Percep	14. Matching	
6. AdaBoost	15. Jaccard	
7. Quad Dis Analysis	16. Dice	
8. Decision Tree	17. Kulsinski	
9. Naïve Bayes		

Results

Classifier Models Results

After running the models above on the dataset and plotting accuracy ranges in figure 1, we can see that **Logistic Regression and SVM were the two top predictors with a median accuracy of 86%**. (Models order follows table 1)

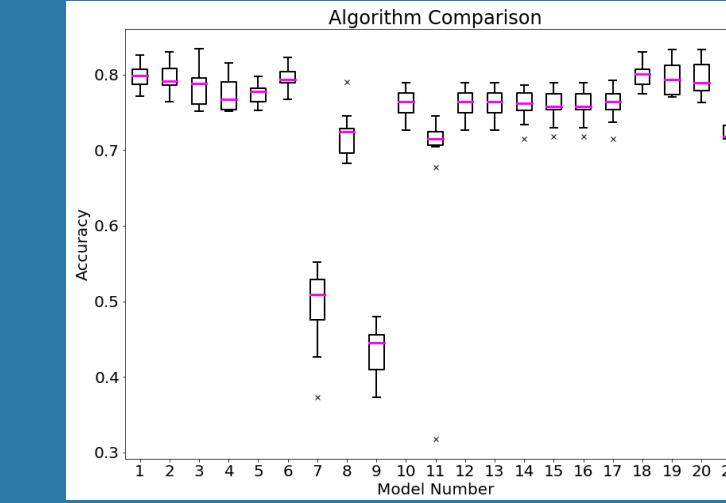


Figure 1: Classifier models accuracy

Neural Net Results

In addition to the classifier models mentioned above, we implemented neural nets. In figure 2, both plots show learning curves by comparing accuracy and loss in the training and the validation sets. The first plot shows a neural net with Stochastic Gradient Descent as the optimizer, while the second plot uses Adam optimizer. **The accuracy in SGD was similar to that of the classifier models while Adam significantly overfitted on the training set.**

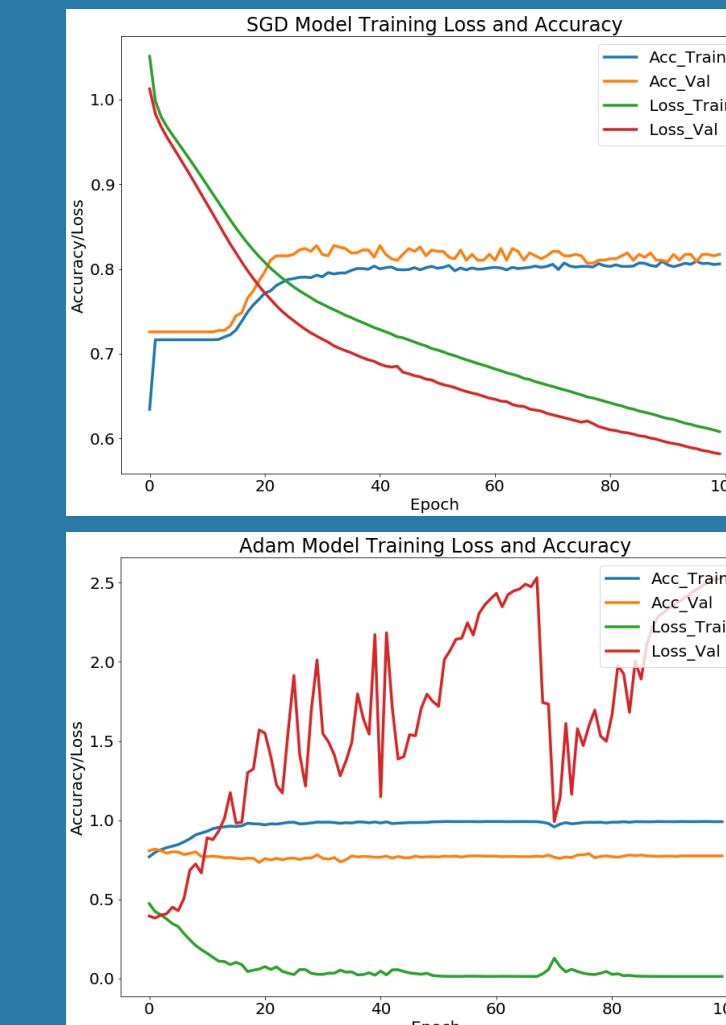


Figure 2: Neural nets learning curve

Smaller Dataset Results

While working with the data, we noticed that newer versions of the data had more features (2017 and 2018 survey). Running the same methods, we got similar results for the classifier models, but **for the NN (shown in figure 3), the accuracy consistently reached 90%**.

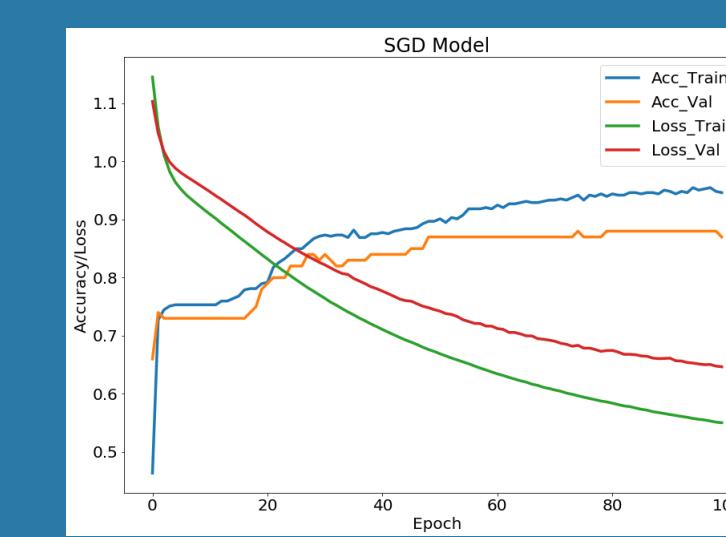


Figure 3: Neural net smaller dataset

Smaller Dataset PCA Results

The new smaller dataset had fewer entries by whittling down to ~500. As a result, we were interested in seeing the performance after using PCA. **The performance for Logistic Regression improved to consistently around 90% after fitting the model to data after PCA with 2 principal components.**

Discussion

Part of the real value of this project is understanding what qualities the workers would contribute to the decision of a worker of being open about mental health. After running Logistic Regression on the small dataset, we got the following weights:

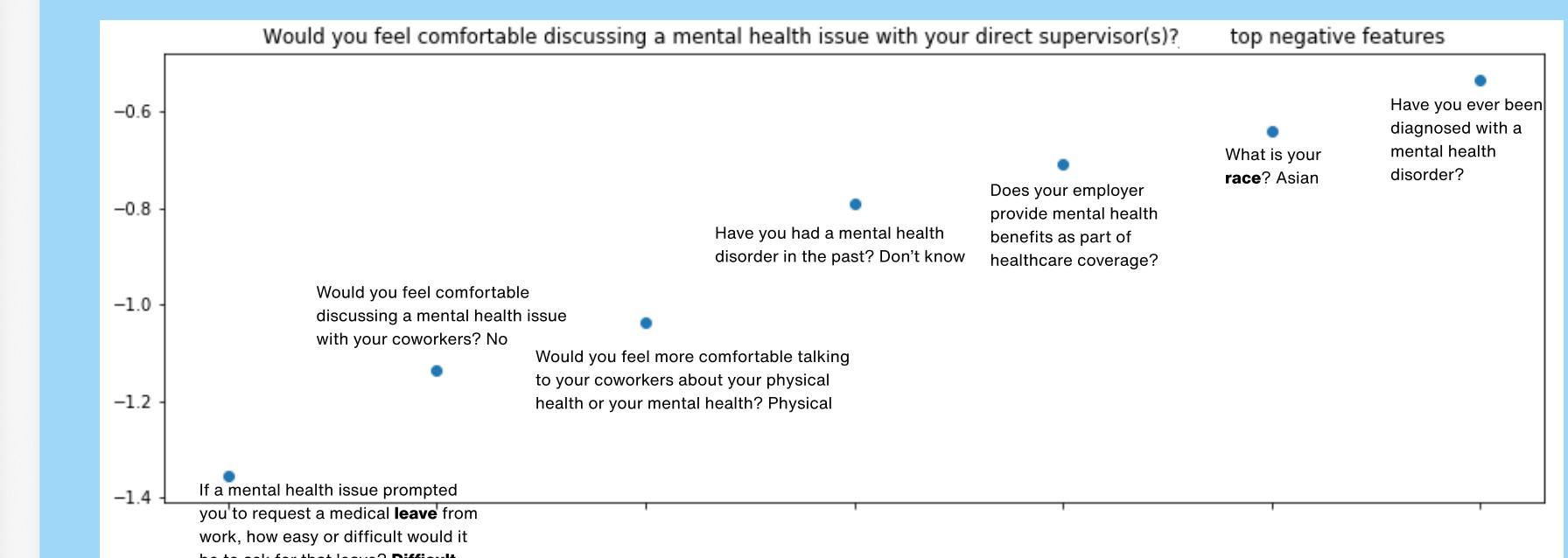
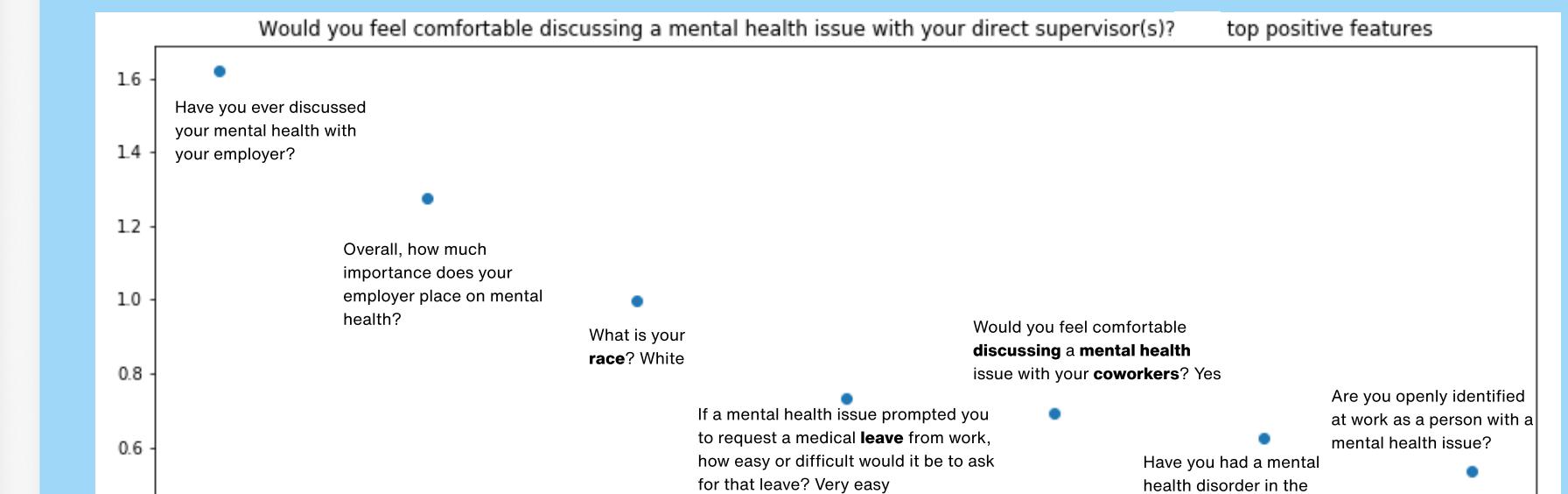


Figure 4: Top positive and negative feature coefficients

Although some features were expected to be strong indicators, there were some surprises. **For instance, being white is a strong predictor of being open, while being Asian is the opposite. Additionally, when comparing within features, we found that companies with 1000+ employees are less likely to be open about mental health**, which makes us wonder why that's the case as companies grow and have resources to promote wellbeing?

Future Investigation

In this project, we showed correlation between worker's perception of the workplace and their openness about mental health. Future work includes investigating why these correlations exist and how they may be used to design work environments. Quantitatively, more data is needed to get a clearer picture of correlations. Qualitatively, there needs to be a deep dive into the different features to understand how to better design work.

[1] Mental Health at Work. Retrieved from <https://www.mindsharepartners.org/mentalhealthatworkreport>

[2] Open Sourcing Mental Illness (OSMI). 2014-2018 OSMI Mental Health in Tech Survey. Retrieved from <https://osmihelp.org/research>

Acknowledgements: We would like to thank our mentor Hancheng Cao for providing guidance in dataset training, as well as Professors Dorsa Sadigh and Percy Liang for a great quarter in teaching Introduction to Artificial Intelligence!