A Comparative Study of Mental Health in Tech within in the US and UK

ABBY NEWBURY, SHIVANI DAS, JON SAKSVIG

Background Information

United States	United Kingdom
Less stigma discussing mental health	More stigma discussing mental health
Mental health treatment is not affordable	Mental health treatment is free
Mental health treatment is not easily accessible	Mental health treatment is easily accessible
23% of adults reported mental health diagnosis in 2016	11% of adults reported mental health diagnosis in 2016
26% of adults reported experiencing emotional distress in last 2 years	17% of adults reported experiencing emotional distress in last 2 years
23% of adults reported not wanting to see a professional when experiencing emotional distress	41% of adults reported not wanting to see a professional when experiencing emotional distress

HYPOTHESES

- **Response Variable:** whether employees seek treatment
- **Null Hypothesis:** Workplace mental health benefits and care options will be less than twice as important in the United States classification model predicting if patients seek treatment in comparison to the United Kingdom classification model.
- **Alternate Hypothesis:** Workplace mental health benefits and care options will be two or more times as important in the United States classification model predicting if patients seek treatment in comparison to the United Kingdom classification model.
- **Reasoning:** In the United Kingdom, the National Health Service (NHS) provides universal health care, meaning mental health care is free for all so citizens do not need to rely on workplace mental health benefits .In contrast, U.S. employees may need to rely more heavily on their workplace for mental health care due to the mostly privatized insurance system.

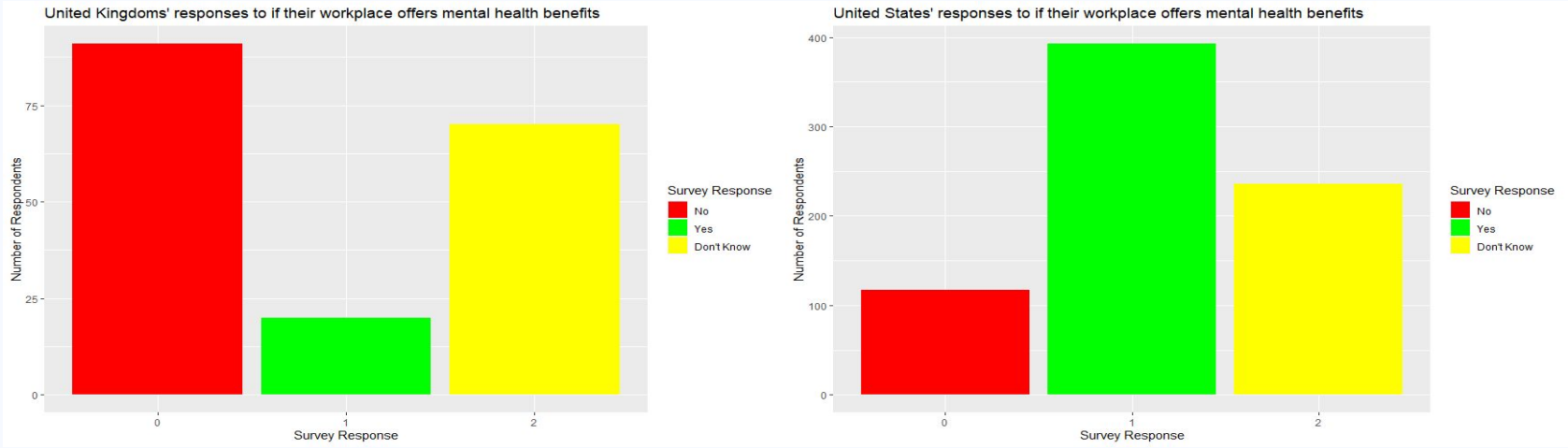
EDA HIGHLIGHTS

Exploring response proportions from sample to identify relationships between variables

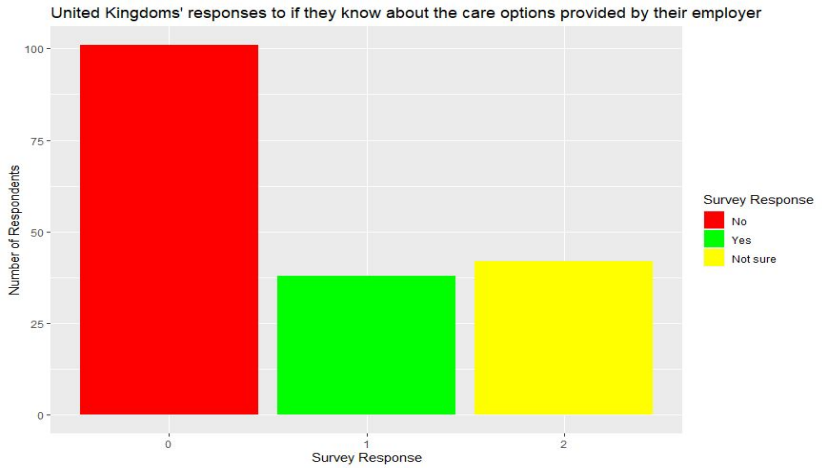
- The proportion of respondents who sought treatment is similar among US and UK respondents.
- The United States proportions for seeking treatment, knowing care options, and knowing mental health benefits are much closer in value than those of the UK.

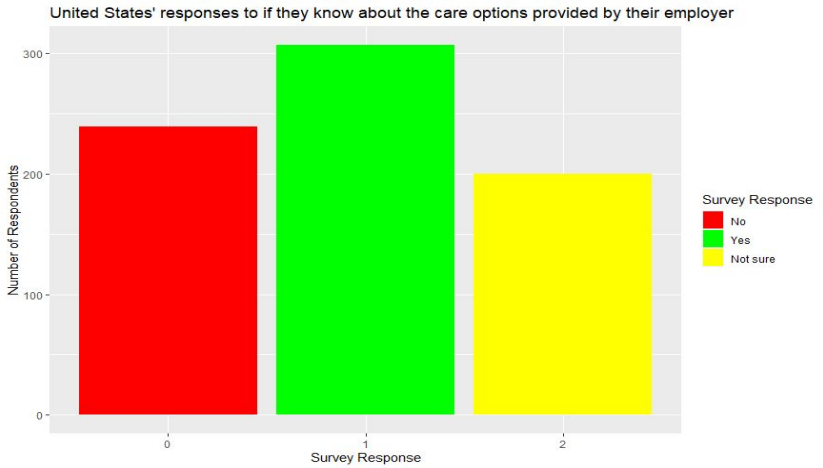
	United States	United Kingdom
Sought Treatment	.544	.492
Knows the care options provided by employer	.411	.209
Mental health benefits are provided by their employer	.526	.110

Difference in workplace benefits



Difference in workplace information on care options





Model Construction

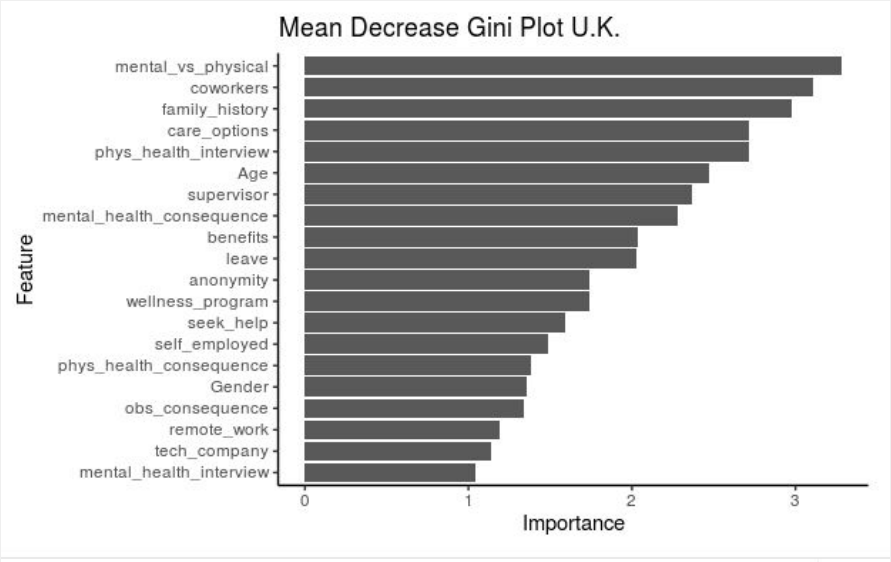
- Initial RF Model creation for U.S. and U.K.
- Feature Engineering: binning age variable, minimizing categories in leave variable
- Oversampling U.K., U.S. dataset to 1000
- Hyperparameter tuning: Changing test/train split for U.K., adjust mtry level
- Final model

Base Rates

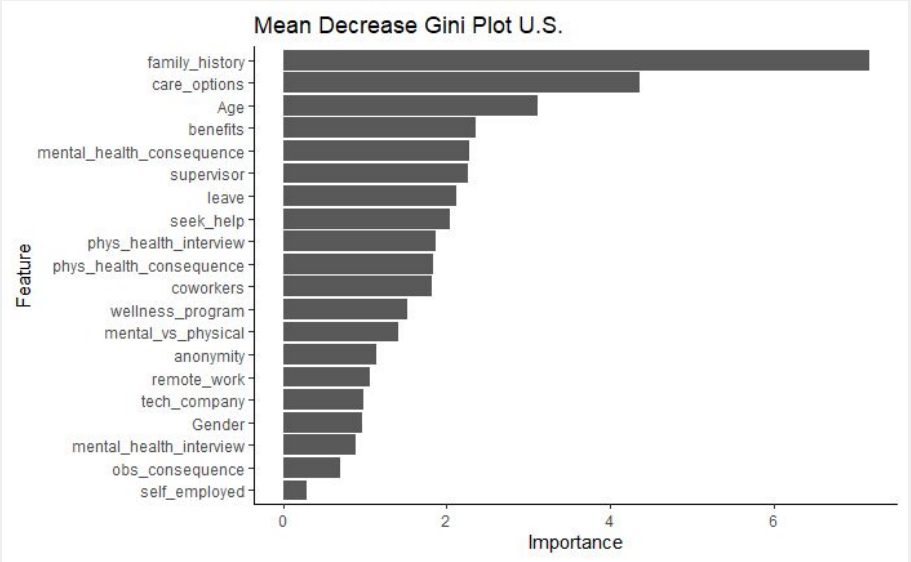
	United States	United Kingdom
Yes	0.53	0.486
No	0.47	0.514

FINAL MODEL

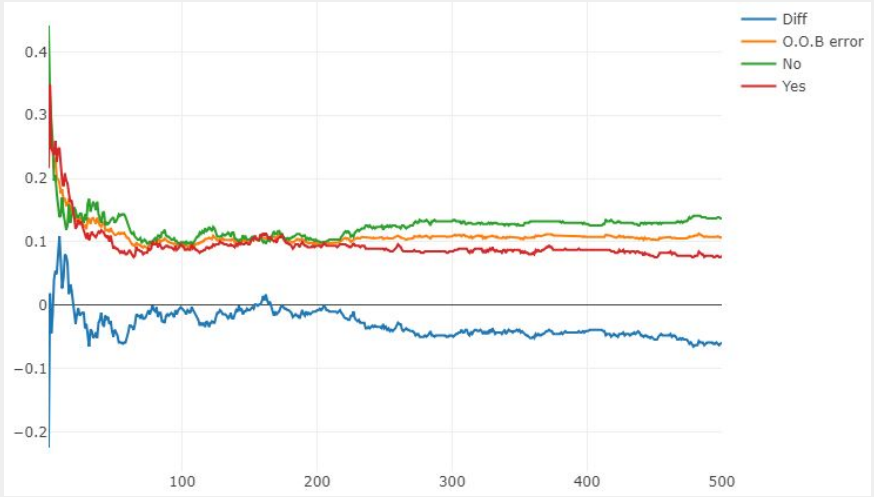
United Kingdom Hypertuned Model with feature engineering: Feature Importance



United States Hypertuned Model with feature engineering: Feature Importance



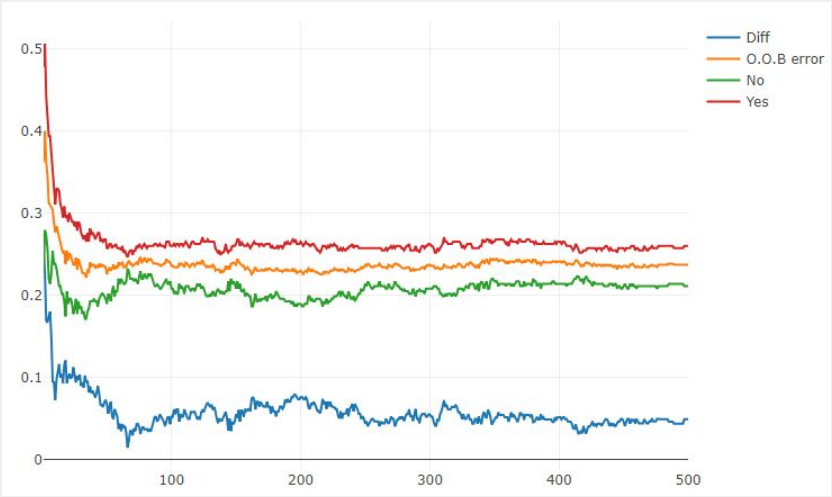
United Kingdom Hypertuned Model with feature engineering: plot



Actual		
Prediction	0	1
0	443	56
1	37	364

	U.K.
Accuracy	.913
F1 Score	.7876
ROC AUC	0.96

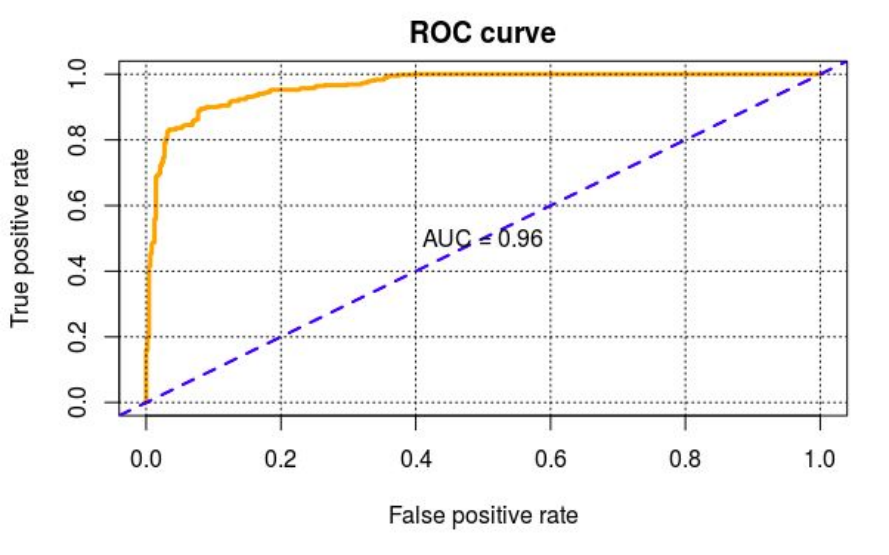
United States Hypertuned Model with feature engineering: Error plot



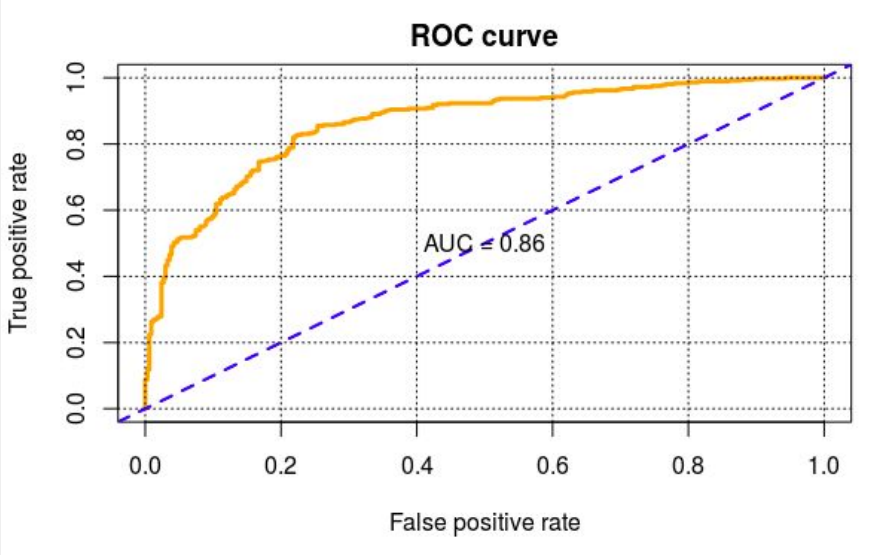
Actual		
Prediction	0	1
0	292	94
1	119	395

	U.S.
Accuracy	.7633
F1 Score	.7688
ROC AUC	.86

United Kingdom Hypertuned Model with feature engineering: Feature Importance



United States Hypertuned Model with feature engineering: Feature Importance



Comparison of Initial and Final Models

	Initial U.K.	Initial U.S.	Final U.K.	Final U.S.
Benefits Variable Importance	1.561	2.176	2.034	2.3619
Care Options Variable Importance	2.161	4.494	3.412	4.3538
Accuracy	0.5565	0.7402	0.913	0.7633
F1 Score	0.5299	0.7757	0.7876	0.7688
O.O.B Error Convergence	0.4	0.27	0.1	0.23
Better Predicted Class	Positive	Positive	Positive	Negative
ROC AUC	0.63	0.79	0.96	0.86

INSIGHTS

1. Disproving the base rate fallacy, model performance improved in comparison to the model solely predicting based on the frequency of each class.
2. Oversampling both data sets to have 1000 observations improved the accuracy and ROC AUC of both the U.S. and U.K. models
3. Benefits variable importance higher in U.S. model for both initial and final models
4. Care options variable importance higher in U.S. model for both initial and final models
5. Care options is on of the most important variables in the final U.S. model (Family History is #1) while employers opinion on mental vs. physical health is the most important variable in the final U.K. model

FINAL MODEL COMPARISON

	U.K.	U.S.
Hit Rate	0.506	0.514

- Performed two sample test for population proportions
  - Met all requirements for test
  - Did not perform continuity correction
- Results: p-value 0.7205, cannot reject the null hypothesis that the two proportions of correctly predicted positive classes are equal at any reasonable significance level

	Final U.K. Model	Final U.S. Model
Benefits Variable Importance	2.034	2.36
Care Options Variable Importance	3.412	4.35

HYPOTHESIS CONCLUSION

- While workplace benefits and care options mean decrease in gini was larger for the U.S. model as compared to the U.K. model,, both were less than twice as large and thus we cannot reject the null hypothesis
- After our two sample test for population proportions, we found that the two models were not different to any significant degree.
- Overall, there was not enough quantitative evidence to suggest any difference between the United States and United Kingdom models.

WEAKNESSES

1. **Oversampling:** creation of synthetic data points
   - Would recommend collecting more U.K. samples to confirm model performance
2. **Overfitting:**
   - Our feature engineering graphs for OOB error show that our US model is prone to overfitting since the error differential doesn't converge as the UK models does.
   - Overfitting in leave category could be because there are 5 options, very easy, somewhat easy, very difficult, somewhat difficult, and don't know. So, we changed it to: Easy, Difficult, Don't know.

3. Assumption of representative sampling:

- Majority of survey respondents come from tech-heavy, urban locations in each country - workplaces in these areas may have more incentives to discuss/provide mental health care options
- Majority of survey respondents were male.
- Age was right-skewed since majority of survey respondents were younger, which is intuitive since tech-industry often has a younger age group for its employees

FUTURE WORK/NEXT STEPS:

- Resample UK for more respondents, since creating synthetic data points in order to oversample will skew the data points and produces a better performing model, however, this doesn't reflect the data.
- Create future models based on synthetic data from the collected survey data to protect anonymity of the survey respondents since some survey respondents left comments that includes potentially sensitive information.
- Utilize a single model, to compare US and UK within one model and see if this differs from the conclusions reached with the two separate models.
- Publish results to the employers of respondents, with the respondents' consent, in order to improve mental health treatment rates in the tech-workplace.
- Ask respondents for suggestions to improve the survey.

Sources for Background Research

- https://www.talkspace.com/blog/us-versus-uk-comparing-mental-health-care-stigma/
- https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-018-1837-1
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426609/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC554041/
- https://www.displayr.com/how-is-variable-importance-calculated-for-a-random-forest/
- https://www.commonwealthfund.org/publications/issue-briefs/2020/may/mental-health-conditions-substance-use-comparing-us-other-countries
- https://www.r-bloggers.com/2017/05/an-interesting-study-exploring-mental-health-conditions-in-the-tech-workplace/