15.076 Final Report
Ankita Devasia & Lydia Yu

# ML for Mental Health Awareness

## Background

Mental health is an issue that is often not given enough attention or well understood. Individuals struggling with their mental health may feel that they do not know whether or not their condition is severe enough to require treatment or counseling. They may find it helpful to have a more objective, customized algorithm that is able to predict their need for treatment based on their personal wellbeing history.

## Significance

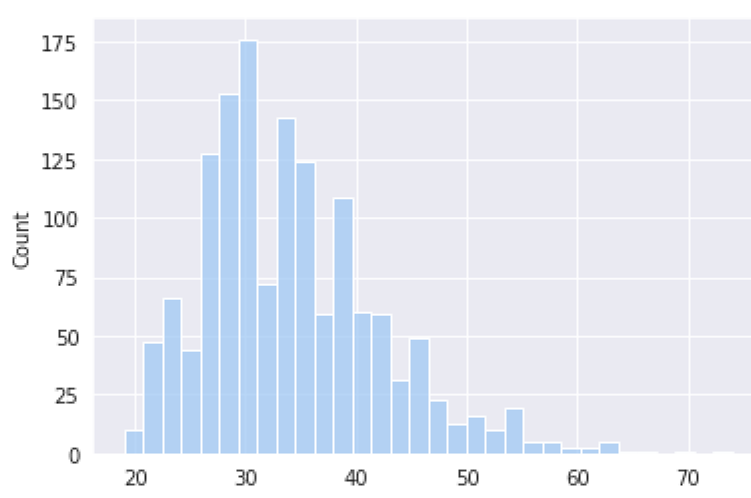
Symptoms of mental illness frequently go unnoticed until they are severe, and even then, many people are uncertain of the best treatment for their condition. Attitudes towards mental health and misperceptions of the prevalence of mental illness has resulted in a lack of discussion of the topic, worsening the situation for those struggling with their wellbeing. Having an accessible tool that individuals can use by inputting their personal health background can bring awareness to mental health treatment.

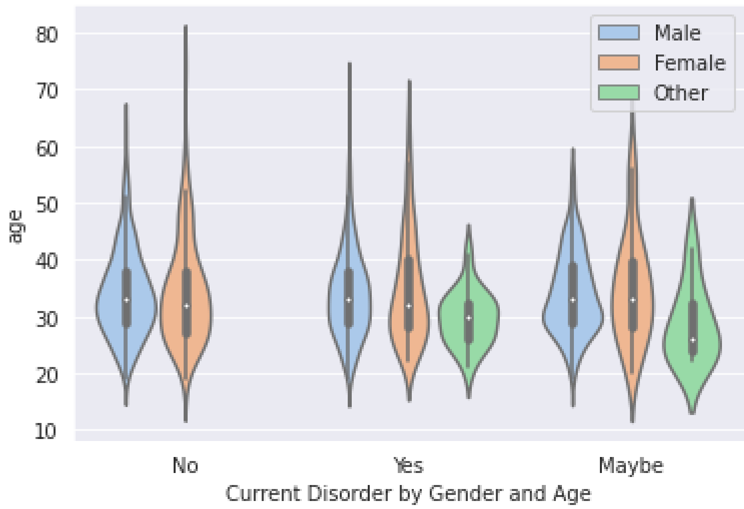
## Data

We are using a dataset obtained from Kaggle that measures attitudes towards mental health and the frequency of mental health disorders in the workplace. The data was sourced from the 2016 OSMI Mental Health in Tech Survey, which consists of 1433 observations of 63 variables. Some examples of questions asked in the survey include “Does your employer offer resources to learn more about mental health concerns and options for seeking help?”, “Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?”, and “If you have a mental health issue, do you feel that it interferes with your work when being treated effectively?”.

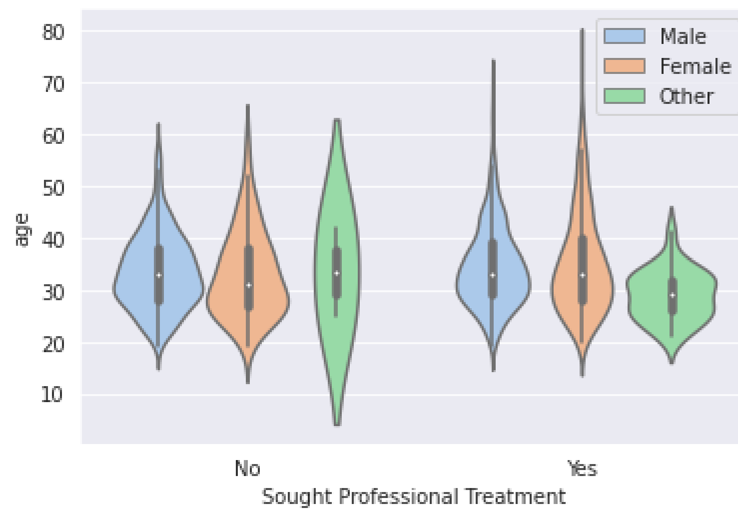
### EDA

74% of individuals in this dataset identify as male and 24% identify as female. Ages are concentrated between 25 and 40, with the highest frequencies of individuals being between 29 and 30.

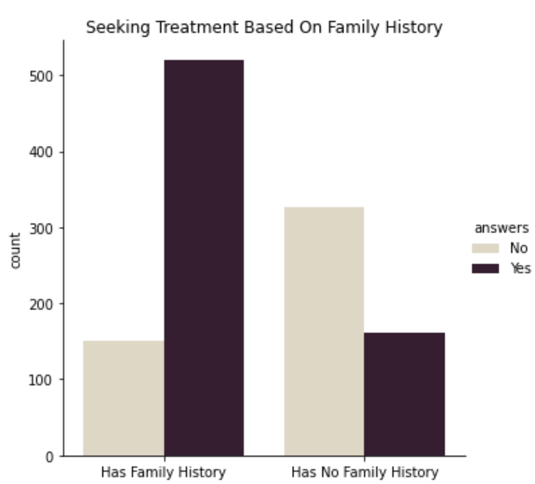





For individuals with a current mental health disorder, the median age within groups is similar across all genders. For individuals who are unsure, the median age for people who do not identify as male or female is lower. In addition, the distribution of ages for men and women with mental health disorders is similar.

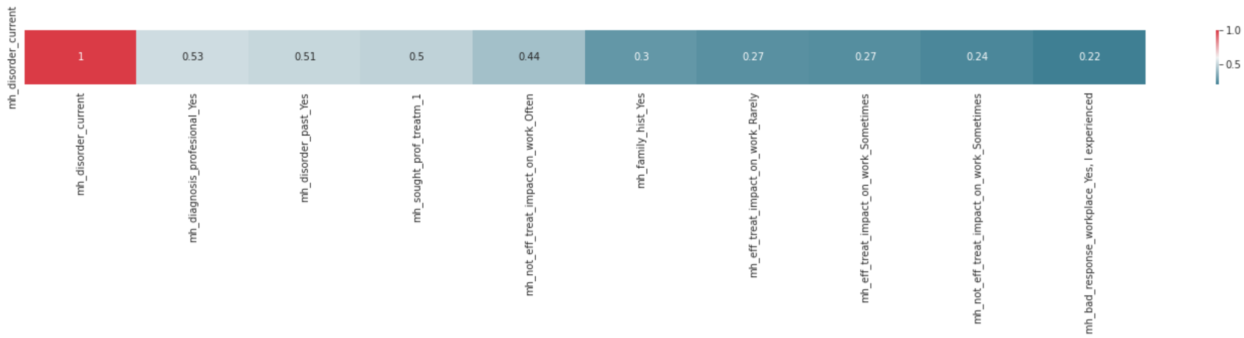



While 75% of women and 92% of individuals identifying with other genders sought professional treatment for mental health issues, only 52% of men did the same. In addition, the median age of individuals identifying with other genders who sought professional treatment was less than that of men and women, indicating that nonbinary people more frequently take action on mental health issues earlier than men and women.




233% more individuals with a family history of mental health illness sought professional treatment than individuals with no family history.

A correlation plot with the variable indicating incidence of a current mental health disorder is highly correlated with having been diagnosed by a professional, having a disorder in the past, or having sought professional treatment, which make intuitive sense. It also is correlated with having a family history and with work productivity being impacted. These variables are self-explanatory and do not particularly indicate revealing new information about incidence of mental health disorders.




## Data Preprocessing

Several steps, listed below, were taken to make the data appropriate for fitting a model to.

1. Renaming columns: because the majority of the survey questions were formulated as sentences, we chose to rename the columns to allow for easier data handling. For example, the column "Is your employer primarily a tech company/organization?" was renamed to "tech_comp_flag".
2. Encoding the gender variable: the gender survey question contained an open ended text box for respondents to type their gender. We decided to map the various responses to three categories: Male, which mapped to a value of 1; Female, which mapped to a value of 2, and Nonbinary, which mapped to a value of 3. For example, text responses of "Male" and "M" would both be mapped to 1.
3. Removing age outliers: because the survey was given to full-time employees, there were a few age values that didn't make sense and were likely typos (either unreasonably high or low). We chose to restrict the dataset to just the observations where the age was greater than 17 years and less than 76 years.
4. Removing unimportant features: some survey questions in the data were missing most of their values. These questions were unlikely to be useful in the models because they did not contain enough non-NA values. We chose to drop these columns.
5. Imputing other NA values: for all other missing values, we chose to impute their value with the most frequent value from that column. Because most of the columns contain non-numerical data, we felt that this was a more reasonable way to impute the data rather than taking a value such as a mean or median. We recognize that this may have limitations in that for variables in which there are many possible values that are relatively uniformly distributed, it may not be representative to impute all missing values with the most frequent value because all values have similar frequencies.

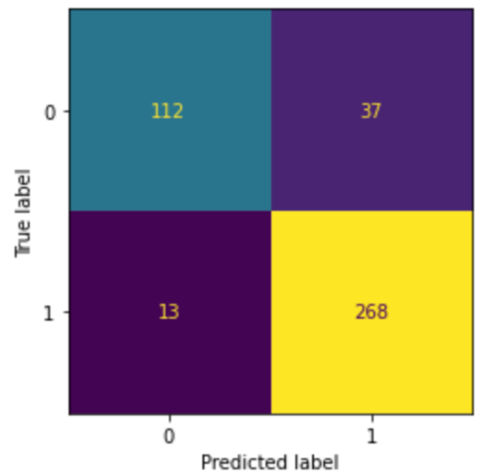
6. Encoding categorical variables: to work with categorical variables such as “Does your employer provide mental health benefits as part of healthcare coverage?” in which the possible responses are “Yes”, “No”, “I don’t know”, or “Not eligible for coverage”, we used one-hot encoding.
7. Encoding the dependent variable “current_mh_disorder”: the possible values for this variable are Yes, Maybe, and No. We encoded Yes = 1, Maybe = 1, and No = 0 because we felt that even if an individual is unsure as to whether they have a mental health disorder, they should seek treatment regardless.
8. Filling in missing values for “tech_flag”: the “tech_flag” variable, which indicates whether or not the respondent works for a tech company, had many missing values. To get around this for missing values, we flagged this variable as True if the “work_position” that the respondent recorded contained keywords such as “Back-end” and “Dev” related to positions that are commonly found in tech companies.

# Classification Models

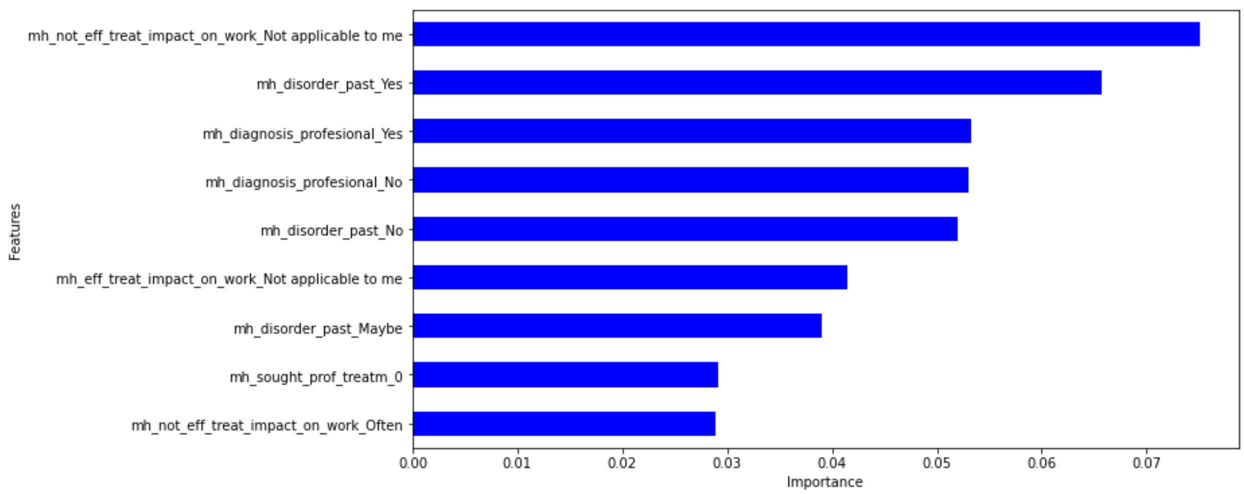
As a proxy for predicting whether or not an individual should seek mental health treatment, we chose to build a few different classification models where “current_mh_disorder” (whether or not the individual currently has a mental health health disorder) was the dependent variable. The data has three possible values for this variable (Yes, No, and Maybe), but as mentioned above, we chose to encode this as a binary classification problem where both Yes and Maybe responses are classified as 1. If an individual is classified as type 1, then we would suggest that they do seek mental health treatment. The independent variables were the rest of the (encoded) survey questions that were not dropped due to high levels of missing values. To judge the accuracy of these models, we looked at in-sample and out-sample $R^2$ as well as the number of false negatives (people who actually did have or might have had a mental health disorder, but who were classified as not having one). Ideally, we would want a model that has high $R^2$ values and a low false negative rate because it is dangerous to not suggest mental health treatment for someone who actually does need it. For all of the following models, the data was randomly split into 70% train and 30% test.

## Random Forest

We build a random forest classifier with n_estimators = 0 and max_depth = 10. This resulted in an in-sample $R^2$ of 0.969 and an out-sample $R^2$ of 0.884. Based on the confusion matrix for the test data, there were 13 individuals who were predicted to not have a mental health disorder (and therefore would not be suggested to seek treatment) but who actually did/might have had one and should seek treatment.



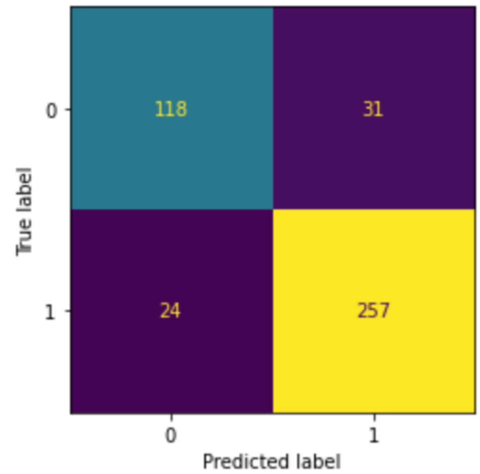

Feature importance:




The most important feature was a NA response to the survey question "If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?". Other important features are whether or not the individual had a mental health disorder in the past, whether or not they have been diagnosed with a mental health disorder by a professional, and whether or not they have ever sought treatment from a mental health professional. Most of these variables seem relatively obvious in determining whether or not an individual currently has a mental health disorder.

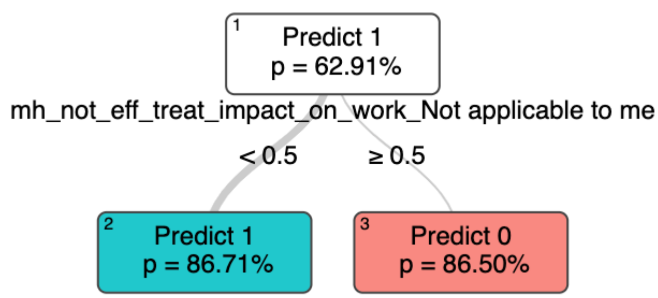
## XGBoost

The next model we tested was an XGBoost classifier with n_estimators = 100, learning_rate = 1.0, and max_depth = 1. This resulted in an in-sample $R^2$ of 0.879 and an out-sample $R^2$ of 0.872. Based on the confusion matrix for the test data, there were 24 individuals who were classified as not having a mental health disorder but who actually did/might have had one.




## Optimal Classification Tree

The final model we tested was an Optimal Classification Tree without hyperplanes with a max_depth parameter ranging from 1 to 7 and a minbucket parameter ranging from 1 to 5. The resulting tree is shown below:




While this tree had a test AUC of 0.832, we felt that it was overly simple and not useful for making predictions because the only variable it splits on is whether or not the individual answered NA to the question “If you have a mental health issue, do you feel that it interferes with your work when NOT being treated effectively?”. If they did answer NA to this question, which implies that a mental health issue or treatment for one does not apply to them, the model predicts that they do not have a mental health disorder. If they did not answer NA to this question, which implies that they do believe they have a mental health issue and their treatment may or may not be effective, the model predicts that they do have a mental health disorder. These predictions feel somewhat obvious and not holistic enough to be useful.

## Comparing the Models

Between the Random Forest and XGBoost models, Random Forest had higher $R^2$ values as well as a lower rate of false negatives. The variables that were most important in the random forest model, however, do not give much new insight into what may be affecting the likelihood of a mental health disorder aside from obvious factors such as a previous disorder or a current professional diagnosis of one. The OCT was certainly the most interpretable of the models, but as mentioned above, the resulting tree was too simple to be useful in providing insight into the predictors of mental health disorders.

## Limitations and Further Work

We recognize that this data is not the best for the given problem – the model is built on whether or not people currently know/feel like they have a mental health disorder, which is self-reported and may not be accurate. However, there were not enough people in the survey who actually went to a professional to get a diagnosis. Also, just because someone doesn't currently have a mental health disorder, doesn't mean that they shouldn't seek treatment. In addition, the very nature of it being survey data means that the responses have the potential to give biased or inaccurate data as the scaling of responses between different people may vary (e.g. a situation that may feel serious to one person may feel less so to a different person). The survey questions used resulted in models that either gave somewhat obvious results (for example, a professional diagnosis predicting a mental health disorder), and/or results that were not holistic enough because they only accounted for a few very obvious variables. It would be helpful to have data on the actual physical health of each person as well, since physical health is often an indicator of mental health, although this data may be difficult to obtain accurately. Finally, this data is limited to employees in tech companies, which is a small subset of the general population. Ideally, we would like to be able to have enough data to build a model that is representative of the general population so that a more diverse set of individuals may benefit from it.

### Optimal Policy Trees

One extension of this project would be to apply the concept of Optimal Policy Trees for the prescription of a mental health treatment that is individualized to each person's personal needs. In the data that we have, we only know whether or not someone actually had a mental health disorder, along with whether or not they sought treatment from a professional. However, we don't have an exact answer for the question of whether or not they should actually seek treatment. In addition, there are many different types of treatments available, and we have not yet looked into the possibility of prescribing a specific treatment for different people. By applying a policy tree and its use of counterfactuals, we can prescribe whether or not someone should seek treatment by looking at the outcome of how much their mental health improved, say a year after the assigned treatment, through follow-up questions on their physical and mental wellbeing, after applying the treatment of suggesting them to seek professional help or the opposite treatment of suggesting them to not seek professional help. By extension, if we had more information on different types of mental health treatments, the policy tree could incorporate specific types of treatments for different disorders and severities that gives the best outcome for each person given their personal features.

## Conclusion

Unfortunately, we were not able to accomplish our goals of understanding the factors that indicate whether or not an individual should seek mental health treatment. This is partially due to the difficulty in assessing mental health in general, as well as the limitations of the dataset we used in this study. The results of our analyses were somewhat obvious and not indicative of why an individual may need to seek professional treatment, perhaps due to the self-reported nature of the study. The breadth of the survey questions was limited to assumptions the respondent made, and in cases where they were not asked to self-diagnose, the state of their mental health was most often based on the diagnosis of a professional. In addition, factors related to the individual's work environment or position in tech did not correspond with our dependent variable, which was surprising. With different data, that accounts for a multitude of facets in the respondent's life, in addition to physical health information and lifestyle factors such as stress, perhaps our models would yield more revealing results.