PREDICTIVE MODELLING FOR MENTAL HEALTH IN THE TECH WORKPLACE

July 31, 2020

Zafrina Somani 501025767

CKME 136 | Ryerson University

Table of Contents

Introduction

The fast-paced nature of the tech industry is what drives innovation and competition, which can also be seen as a breeding ground for stress. With the shifting focus and importance of IT and technology in the workplace, it is imperative that individual mental health is taken into consideration.

The purpose of this project will be to identify workplace factors that contribute to employees seeking treatment. Moreover, the goal will be to predict individuals with mental health diagnoses based on their feelings toward workplace mental health. This will help to determine the attributes that contribute most towards determining whether or not an individual has been diagnosed with a mental health condition, as well as, whether or not they have sought treatment for a mental health condition.

Through data mining techniques, such as classification, the dataset will be used to determine the prevalence of mental illness in the workplace, awareness of mental health issues, and stigma (perceived and any stigma held by the individual) present in workplace. The dataset will also be examined using regression analysis to determine the relationship between the variables. This will ensure human resource departments have concrete support with creating programs and benefits aimed at reducing absenteeism and the costly loss of productivity.

Literature Review

Islam, M. R., Miah, S. J., Kamal, A. R. M., & Burmeister, O. (2019). A Design Construct of Developing Approaches to Measure Mental Health Conditions. *Australasian Journal of Information Systems*, *23*.

Islam et al. focus on creating a solution design for detecting and measuring the impact of mental health conditions using machine learning techniques. The OSMI dataset was used to summarize and measure the prevalence of mental health issues in the workplace. The analysis was further substantiated through identifying affecting factors using open access data through a modified API from Facebook. Machine learning was used to examine detection performance of IT workers with mental health issues to determine the significance of different types of features. The study found that back-end developers are suffering with mental health issues the most. Four key factors affecting mental health for IT professionals were ‘minimal risk’, ‘workplace concern’, ‘activity concern’, and ‘emotional stresses’.

Sadia, A. (2019). A Comparative Analysis of Four Classification Algorithm for Mental Health Analysis basis on technical People.

The purpose of the study was to formulate a model to determine the cause of mental health disorders and establish the most effective algorithm to predict a mental health disorder. The classification algorithms utilized in the study were Decision Tree, Naïve Bayes, Artificial Neural Networks (ANN), and k-Nearest Neighbours (kNN). K-means clustering was used as the clustering technique to identify the relationship between mental illness and working position in the tech workplace. ANN was found to be the best classification method for detecting a mental health condition, whereas Naïve Bayes has the best accuracy. Tech workers affected by mental health conditions were between the ages of 26-46. According to the encoding label, these individuals generally worked as systems administrators, front-end developers, back-end developers, executive leadership, team leaders, or developers.

Ahlmann-Eltze, C., & Yau, C. (2018, October). MixDir: Scalable Bayesian Clustering for High-Dimensional Categorical Data. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 526-539). IEEE.

Ahlmann-Eltze and Yau discuss the challenges of handling high-dimensional categorical datasets because as the number of attributes increases the correlation structure grows exponentially. Analysis on high-dimensional data is currently executed using univariate analysis where responses are associated to individual questions with the class attribute. This limits the ability to identify complex, multivariate response patterns. The study found that workers are less likely to expect negative consequences for being open about mental health issues if mental health is covered under employer-provided coverage. The clustering technique proposed can deal with large datasets, utilizes a Bayesian framework to handle missing data, and can handle a dataset where the number of latent classes is unknown.

Appiah, S., Barnard, S., & Deiven, J. (2017, October). Density-based Clustering of Workplace Effects on Mental Health.

The purpose of this paper was to use machine learning algorithms to find out which workplace factors affect employee mental health and how these factors impact mental health. More specifically, the objective was to determine workplace factors that contribute to the likelihood of the employee seeking help. The study found that females are more likely to seek and receive treatment to mental health conditions than males. Density-based spatial clustering of applications with noise (DBSCAN) was used to create clusters that are not necessarily linearly separable and more organic without a pre-specified number of clusters. Gender was found to have the largest impact on employee's likelihood to seek and receive mental health treatment. The study was unable to identify all the factors learning to an individual seeking mental health treatment and required better attribute optimization with a more relevant dataset.

Patel, P. (2018). Perceived Workplace Factors and their Influence on Self-Reported Mental Health Service Seeking Among Technology Workers.

The objective of this paper was to determine workplace-related perceptions that contribute to individuals who self-report receiving mental health treatment. A frequency table analysis was performed to determine which predictor variables in the survey accurately predicted the outcome response. Treatment-seeking behaviour was found to be associated with whether participants believed mental health issues interfered with their work. Gender was also found to have an impact on individuals seeking treatment – females were more likely to use mental health services over males. The following variables were seen to have an effect on whether or not a tech worker would seek treatment: family history, work interference, offered benefits, observed negative consequences, and age. Results from the regression model indicated that company's should invest in employee benefits to encourage the use of mental health services and prevent loss of productivity.

Reddy, U. S., Thota, A. V., & Dharun, A. (2018, December). Machine Learning Techniques for Stress Prediction in Working Employees. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*(pp. 1-4). IEEE.

Reddy, Thota, and Dharun developed a model to predict the risk of stress experienced by individuals in the workplace and if treatment is required by the individual. The model was trained on whether or not an employee was treated for a stress-related disorder in the past. Gender was found to have the highest influence on stress, where women were found to be under greater mental stress than men. The research also found that individuals working in tech companies were more at risk for developing mental health issues, regardless of whether or not their role was based in technology. Random Forest classifier was found to be the most stable model, whereas k-Nearest Neighbours was found to be the most unreliable (highest false-positive rate). Gender, family history of mental illness, and whether an employee provides mental health benefits to employees had the most impact on whether or not a person developed a mental health issue.

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124-136.

Lessmann, Baesens, Seow, and Thomas examine predictive methods and performance evaluation methods for credit scoring. Lessmann et al. create a new baseline for which predictive methods and performance evaluation methods shows the most efficacy. The study found that for the purpose of credit scoring, multi-layer perceptron artificial neural networks, logistic regression, random forest, and HCES with bootstrap sampling should be the new industry standard in credit scoring. The study speaks to the importance of assessment and shifting focus from logistic regression, which was the current industry standard. When discussing performance measure, Lessmann et al. found that AUC (area under the receiver operation characteristics curve) is a suitable measure to compare classifiers for retail scorecards. Kendall's rank coefficient was used to compare performance measure, where it was found that AUC, partial Gini index, and Brier score were strong evaluators. The study also spoke to the importance of using at least three performance evaluators and that AUC and H-measure are interchangeable as they show strong correlation.

Dataset Description

The clean dataset consists of 1128 observations of 55 variables.

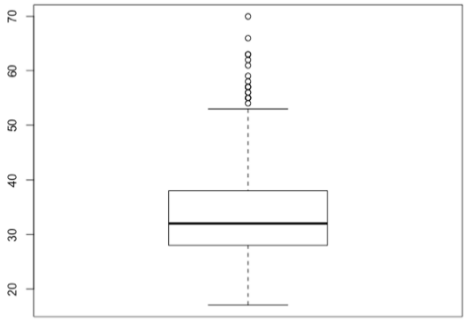
The dataset was cleaned to represent employees who are not self-employed as the focus of the project is workplace factors influencing the mental health of employees. The individuals who identified as self-employed were asked a different set of questions, these attributes were removed from the dataset. As this dataset focuses on individuals working at tech companies or individuals whose roles are related to tech or IT, any individuals who do not work at a tech company or have a role related to IT/tech were removed from the dataset.

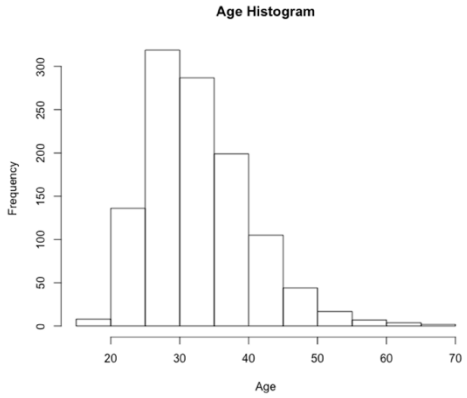
Missing values were evaluated to determine cause and were kept if they were determined to be missing at random or missing not a random. The missing values were determined to be missing at random for questions pertaining to previous employment, as individuals who did not have previous employers were not asked any questions about it. Observations were determined to be missing not at random due to their dependency on other variable's values.

The uncleaned age attribute had outliers which showed individuals working at the ages of 3, 99, and 323. Though these appeared to be data entry errors, these rows of data were removed. The cleaned attribute's boxplot and histogram are shown below. There still appear to be outliers in age, however, these outliers seem logical as individuals may still be working at the age of 70. It does appear that most individuals working in tech companies or in tech roles are in their 30s, this is exemplified by the age histogram being skewed slightly to the left.

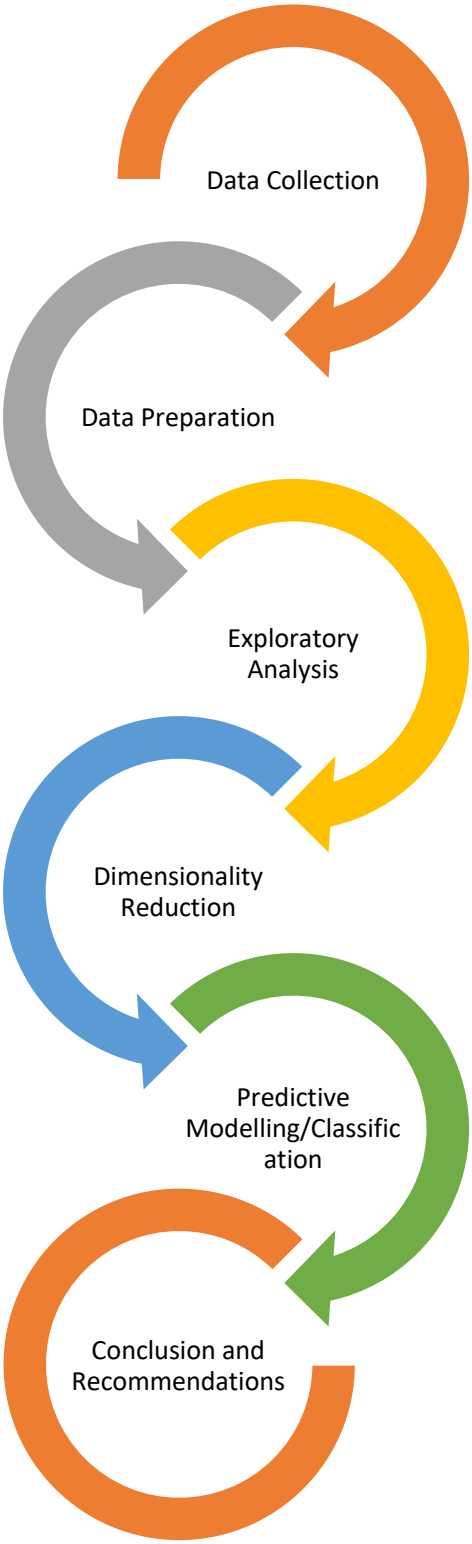
The gender data was cleaned to create three categories for the data. Gender was a freeform answer, so responses were varied. The categories decided on were female, male, and other. Other corresponds to individuals with gender expression that was neither male nor female. For example, these individuals identified as genderqueer, agender, androgynous, bigender, non-binary, genderfluid, and transgender.

Summary statistics for categorical has been attached in the appendix to show the frequencies for each attribute.





Approach


Data Collection
Data Preparation
Exploratory Analysis
Dimensionality Reduction
Predictive Modelling/Classification
Conclusion and Recommendations

Step 1: Data Collection

- Download data from Open Sourcing Mental Illness Ltd., via Kaggle and import into R

Step 2: Data Preparation

- Inspect data frame in R using head(), tail(), str(), summary(), dim()
- Ensure database is normalized – columns and rows are organized with each variable in a column and each observation in a different row
- Check and assign correct data types class(), as.factor()
- Pre-process character data
    - Standardize gender data
    - Inspect freeform answers
- Detect inconsistencies and cross-check variable relations
- Evaluate missing values
    - Determine whether to remove missing values or replace with variable (i.e. mean)
- Detect outliers and determine cause of outliers to decide whether or not to remove

Step 3: Exploratory Analysis

- Explore relationships among attributes in the data:
    - Most common mental health issue tech employees are diagnosed with
    - Most common mental health issues tech employees think they have

Step 4: Dimensionality Reduction

- Utilize various feature selection techniques to reduce dimensionality

Step 5: Predictive Modelling/Classification

- Conduct classification utilizing Logistic Regression, Artificial Neural Network, and Decision Trees
    - Identify workplace factors contributing to employees seeking treatment
    - Predict whether or not an individual has been diagnosed with mental health disorder based on workplace mental health factors

Methods

Feature Selection

The statistical data selection process' for this assignment includes various feature selection algorithms. Dimensionality reduction was utilized to improve both the accuracy, efficiency, and efficacy. The feature selection algorithms utilized in this project include chi-square, information gain, correlation-based feature selection, and consistency feature selection.

The Chi-square test for feature selection tests the independence of two variables. This statistical test aids in determining if a relationship exists between two variables (Gajawada, 2019). The Chi-square statistic is calculated using the following formula:

$$\chi = \sum \frac{(Observed - Expected)^2}{Expected}$$

$$Expected = \frac{RowTotal \times ColumnTotal}{OverallTotal}$$

The Chi-square test for feature selection targets features that are highly dependent on the response variable (Gajawada, 2019). A higher Chi-square value indicates that a feature is more dependent on the response variable and can be selected for model training (Gajawada, 2019). Chi-square tests are especially used for feature selection when the input variables are categorical (Brownlee, 2019).

Information gain is used for feature selection by evaluating the gain of each variable in relation to the class variable (Brownlee, 2019). Information gain measures how much entropy is reduced by based on a given value of a random variable (Brownlee, 2019). A large information gain indicates lower entropy (Brownlee, 2019). Information states how surprising and event is and is measure in bits; therefore, if an event has a lower probability, it will have more information (Brownlee, 2019). Entropy measures the amount of information in a random variable; moreover, a skewed distribution has lower entropy and a distribution where events have equal probability has a large entropy (Brownlee, 2019). The formula for information gain is:

$$Entropy(S) \equiv \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Correlation feature selection determines whether attributes are linearly dependent, which would imply they have the same effect on the dependent variable (R, 2020). If attributes are found to be highly correlated, one of the features can be dropped as they will have the same effect on the class variable. The correlation-based feature selection algorithm utilized found a subset of attributes using correlation and entropy measures (Kotthoff). The correlation-based feature selection determines the best subset of attributes that share a strong relationship with the class variable, while having low correlation amongst the other attributes (Jiarpakdee, Tantithamthavorn and Treude, 2018).

Consistency-based feature selection utilized the inconsistency rate to evaluate the attributes. The optimal subset of attributes is selected based on whether the inconsistency rate estimates the inconsistency rate of all metrics (Jiarpakdee, Tantithamthavorn and Treude, 2018). Attributes are considered inconsistent if they "have the same feature values but different class labels" (Lin, 2009). "The inconsistency rate is the summation of all the inconsistent counts over all patters divided by the total number of instances in the data" (Lin, 2009). The formula for determining the inconsistency rate is:

$$INCR = \frac{\sum_{i=1}^{h} INCi}{M}$$

Predictive Modelling Classification

Logistic regression determines the probability of the categorical class attribute based on the selected dependent variables (Lavalley, 2008). This probability is classified by:

$$f(x) = \frac{e^x}{1 + e^x}$$

Logistic regression is a linear regression on the logit transformation of f(x), where f(x) is the probability of success at each value of x (Babaoglu, 2020). In the above formula, x can be replaced with the equation of the line in the form of ax+b (Babaoglu, 2020).

An artificial neural network (ANN) comprises a group of neurons that process information in a parallel processing fashion (Lin, Chu, Wu and Verburg, 2011). An ANN consists of the input layer, which is made up of the neurons or nodes, hidden layer, which contains the connection weights, and the output layer, which contains the desired output of the model (Lin, Chu, Wu and Verburg, 2011). An ANN goes through a learning and recall process which is where the connection weights are adapted to produce the desired output (Lin, Chu, Wu and Verburg, 2011).

A decision tree is a non-linear, supervised classification method. The decision tree is made up of the root node, internal node, leaf nodes, and branches (Song and Ying, 2015). The decision tree starts with a binary target value found in the root node, it is followed by internal nodes that represent choices that the tree faces in trying to predict the target value, and leaf nodes are the final result on the bottom of the tree (Song and Ying, 2015). The branches of a decision tree represent outcomes from root and internal nodes (Song and Ying, 2015). Conditional inference trees were used to estimate the regression relationship and works in the following steps: 1) variables are tested with the null hypothesis that input variables are independent from the response variable, 2) if null hypothesis cannot be reject, process is stopped and the input variable with the strongest association to the response is selected, 3) input variable is split into binary, 4) above steps are repeated (Hothorn).

Evaluation Methods

A confusion matrix observes the relationship between the output of the predicted classifier and actual class (Diez, 2018). The following is an example of the 2-class matrix that this project will utilize for performance evaluation of predictive modelling:

Class		Predicted	
		Yes	No
Actual	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

A true positive is the number of correctly classified rows, whereas, a false positive is the number of incorrectly classified rows (Diez, 2018). A false negative shows the number of falsely rejected rows, whereas, a true negative depicts the number of correctly rejected rows (Diez, 2018). Precision and recall can be calculated with the confusion matrix. Recall is the ratio of correct predictions that were made of the records that were actually in the 'yes' class (Bhowmick, 2020). Precision measures the number of correct predictions made of the records that were predicted to be in the 'yes' class (Bhowmick, 2020). The formula for precision and recall, respectively, are as followed:

$$Recall/Sensitivity = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP}$$

The Brier score measures the accuracy of predictions and operates under the assumption that that the model's probability estimate is perfect (Glen, 2017). The Brier score is given by the following formula where o represents the binary outcomes that are being predicted and p represents the prediction probability:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2.$$

Results

The cleaned dataset was examined to gain a better understanding of the population that is being examined. The most common current diagnoses of individuals working the tech field was found to be anxiety disorders, which was followed closely by mood disorder. Individuals identified one current diagnoses; however, some individuals stated that they had a maximum of 8 current diagnoses. The most common combination of current diagnoses was found to be a mix of an anxiety and mood disorder. The same was found for undiagnosed and medically confirmed diagnoses.

Demographically, most individuals were found to be working and living in the United States; however, the second largest country that respondents were found to be living and working in is the United Kingdom. Most individuals were found to be working and living in the same country, with less than 1% of the population were living and working in different countries. The gender of the population was predominately found to be male; moreover, 23% identified as female, and 2% identified other (a combination of male and female or neither male nor female). Individuals working in the tech industry were mostly found to be between the ages of 20-49, with individuals mostly being between the ages of 30 and 39.

Four dimensionality reduction techniques were utilized to select the most important features to build the model. Chi-square, information gain, correlation, and consistency feature selection methods were used. The features that were selected using information gain, and correlation were found to be subsets of Chi-square; therefore, the chi-square features were selected to build the models with.

The data was divided using a 70% train and 30% test split.

The logistic regression model created for whether or not an individual was diagnosed by a medical professional the following variables were found to have the most significance on the model: the presence of a mental health condition in the past, the individual has sought treatment, the individual was currently diagnosed with a mental health condition. Whereas, for the logistic regression model created for whether or not an individual has sought treatment, the following variables were found to have the most significance: individual has been diagnosed by a medical professional, and the individual did not feel that a mental health condition negatively impacted their work if it was treated effectively.

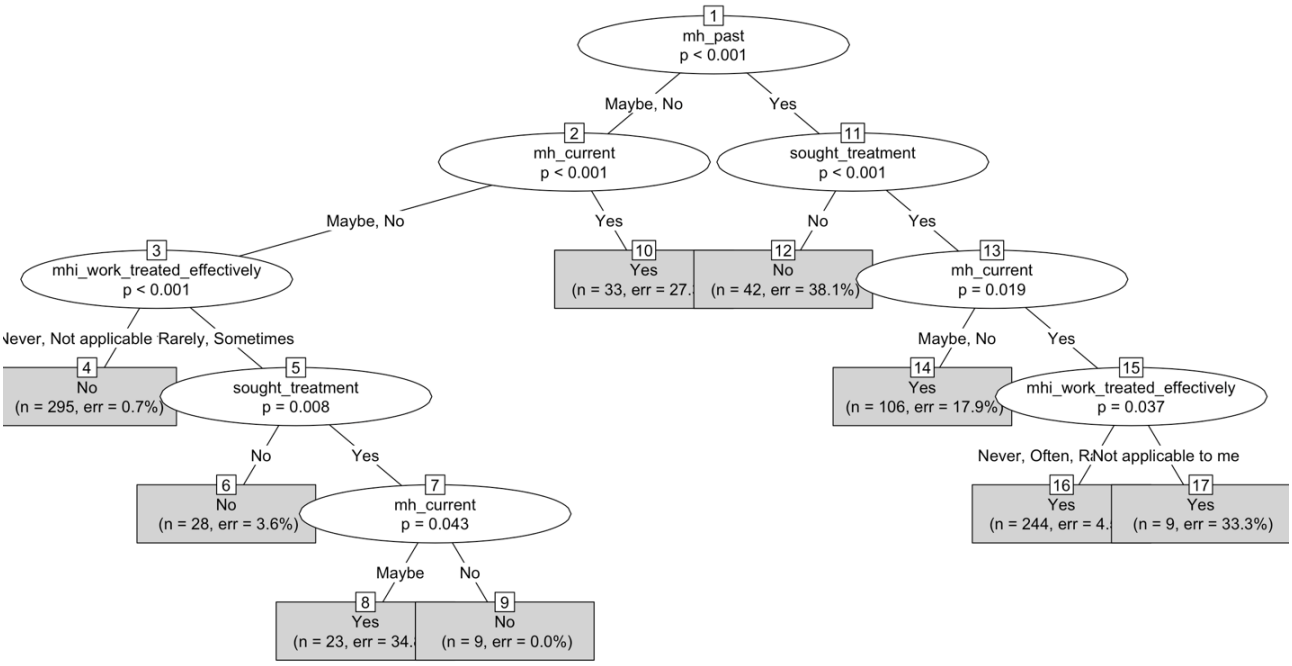
The following is the confusion matrix for the logistic regression models:

Confusion Matrix			
Logistic Regression	Diagnosed Medpro		
	Actual		
	Class	No	Yes

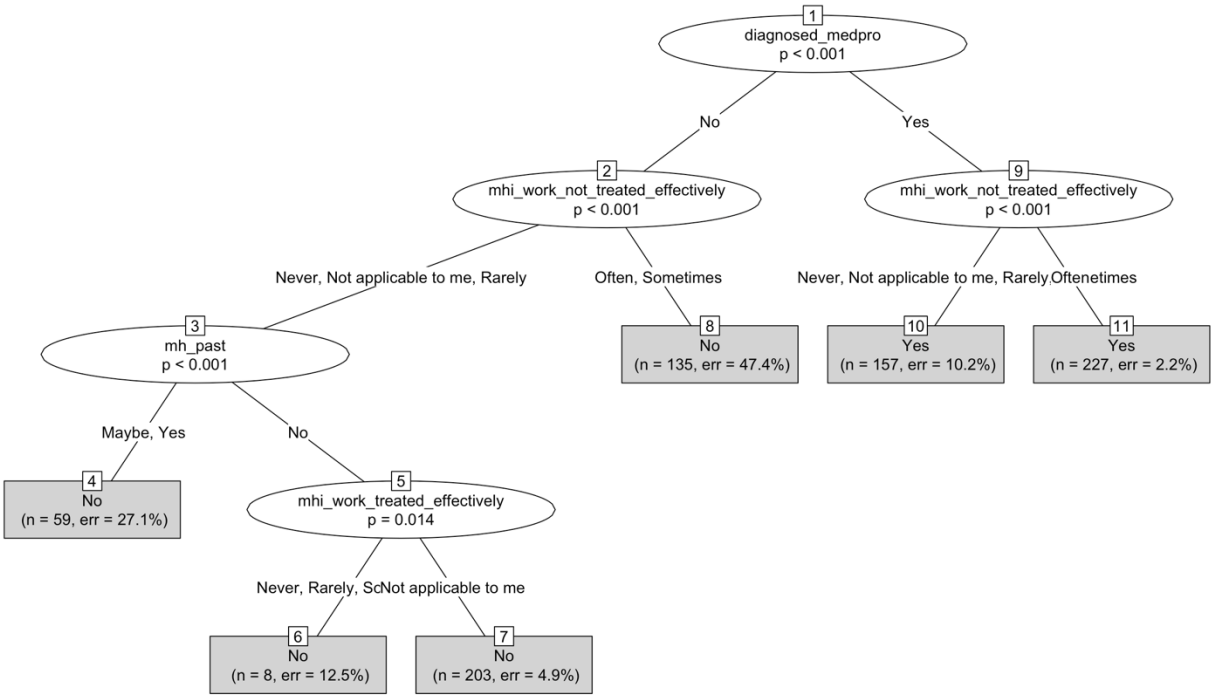
Predicted	No	150	17
	Yes	14	158
Accuracy	**0.9086**		
Logistic Regression	Sought Treatment		
	Actual		
	Class	No	Yes
Predicted	No	123	22
	Yes	23	171
Accuracy	**0.8673**		

The Brier score for the logistic regression model on diagnosed by a medical professional was 0.06880312. The Brier score for the logistic regression model on sought treatment was 0.08916919.

The decision tree for diagnosed by medical professional is:



The decision tree for sought treatment is:

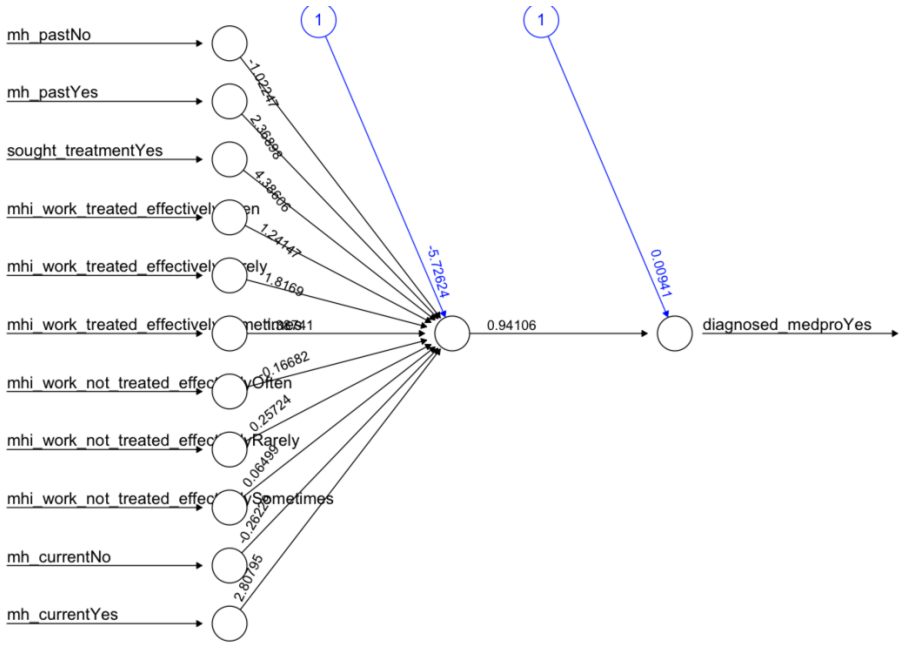


The confusion matrix for the above models are as follows:

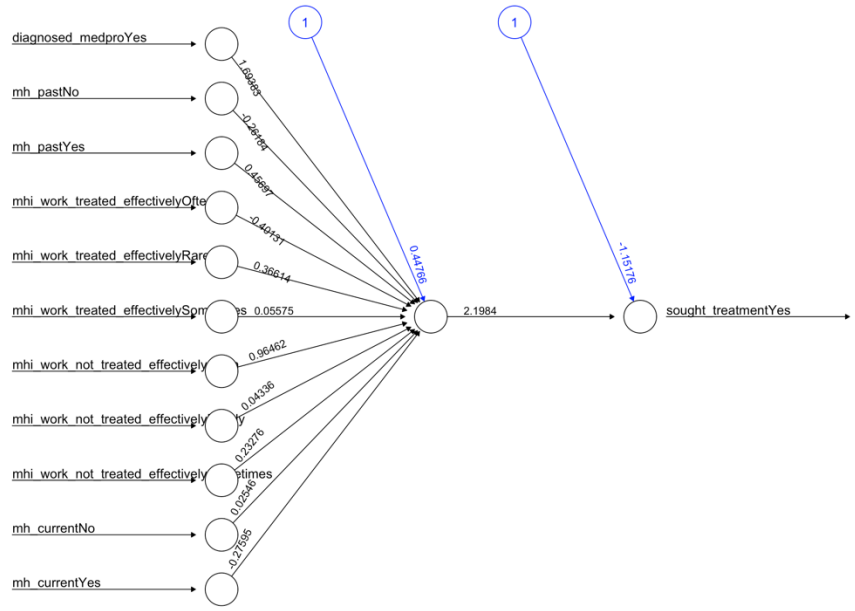
Conditional Inference Tree (Decision Tree)	Diagnosed Medpro		
	Actual		
	Class	No	Yes
Predicted	No	145	13
	Yes	19	162
Accuracy	**0.9056**		
Conditional Inference Tree (Decision Tree)	Sought Treatment		
	Actual		
	Class	No	Yes
Predicted	No	135	29
	Yes	11	164
Accuracy	**0.882**		

The Brier score for the decision tree pertaining to diagnosed by a medical professional is -0.9734513. The Brier score for the decision tree pertaining to seeking treatment is -0.9144543.

The artificial neural network created for diagnosed by medical professional is shown by the following diagram:



The artificial neural network created for seeking treatment is shown by the following diagram:



The confusion matrix for the above models is as follows:

Artificial Neural Network	Diagnosed Medpro		
	Actual		
Predicted	Class	No	Yes
	No	146	18
	Yes	13	162

Accuracy	**0.9086**		
Artificial Neural Network	Sought Treatment		
	Actual		
Predicted	Class	No	Yes
	No	120	24
	Yes	22	173
Accuracy	**0.8643**		

The Brier score for the diagnosed by medical professional model is 0.09144543. The Brier score for whether or not an individual sought treatment is 0.1356932.

Conclusion

The results showed the strongest model according to both the confusion matrix, accuracy and Brier score was found to be the logistic regression model for both whether or not an individual was diagnosed by a medical professional and whether or not an individual sought treatment.

The features determined to have the most significance on the diagnosed by a medical professional target variable were: whether or not an individual had a mental health condition in the past, whether the individual sought treatment for a mental health condition, whether a mental health condition was found to impact work if treated or not treated effectively, and whether or not an individual currently has a mental health condition. Whether or not an individual has had a mental health issue in the past would impact if they were diagnosed, as well as, an individual who has been diagnosed is likely to have been given treatment options by the medical professional.

The features determined to have the most significance on whether or not an individual sought treatment were as follows: whether or not an individual was diagnosed by a medical professional, whether an individual had a mental health condition in the past, whether a mental health condition was found to impact work if treated or not treated effectively, and whether or not an individual currently has a mental health condition. An individual would seek treatment if they have been diagnosed by a medical professional and if they have had a mental health issue in the past they are more likely to seek treatment to manage this condition.

Overall, the logistic regression models showed the most stability as determined by the performance measures utilized. The second strongest model was found to be the artificial neural networks with the decision trees showing the lowest Brier score.

The features found were more related to an individual's history with mental health and their ability to effectively manage their condition. Interestingly, whether or not an individual has mental health benefits was not found to have a strong impact on either class variable.

For further research, models could be formed using the features selected by the consistency measure and compared against those formed by the Chi-squared measure. The model could be further complicated by creating models that predict the class of mental health disorder an individual is diagnosed with (i.e., anxiety disorder, mood disorder, or a combination of disorders). The default settings were used for algorithms in R, to further understanding, these models could be complicated by comparing different activation function for the artificial neural network, and comparing different decision trees. Furthermore, creating a subset of the variables and classifying them as relating the an individual's mental health history, and workplace factors can be used to further identify any specific factors in each category.

Bibliography

Ahlmann-Eltze, C., & Yau, C. (2018, October). MixDir: Scalable Bayesian Clustering for High-Dimensional Categorical Data. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 526-539). IEEE.

Appiah, S., Barnard, S., & Deiven, J. (2017, October). Density-based Clustering of Workplace Effects on Mental Health.

Babaoglu, C. (2020). Statisitcal Learning Methods [Powerpoint slides].

Bhowmick, A. (2020). Natural Language Processing and Machine Learning [Powerpoint slides].

Brownlee, J. (2019, October 30). A Gentle Introduction to the Chi-Squared Test for Machine Learning. Retrieved July 24, 2020, from https://machinelearningmastery.com/chi-squared-test-for-machine-learning/

Brownlee, J. (2019, November 04). Information Gain and Mutual Information for Machine Learning. Retrieved July 23, 2020, from https://machinelearningmastery.com/information-gain-and-mutual-information/

Diez, P. (2018). Confusion Matrix. Retrieved July 01, 2020, from https://www.sciencedirect.com/topics/engineering/confusion-matrix

Gajawada, S. (2019, October 20). Chi-Square Test for Feature Selection in Machine learning. Retrieved July 24, 2020, from https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223?gi=18c3d47b8b78

Glen, S. (2017, October 31). Brier Score: Definition, Examples. Retrieved July 20, 2020, from https://www.statisticshowto.com/brier-score/

Hothorn, T. (n.d.). Party. Retrieved July 15, 2020, from https://www.rdocumentation.org/packages/party/versions/1.35/topics/Conditional%20Inference%20Trees

Islam, M. R., Miah, S. J., Kamal, A. R. M., & Burmeister, O. (2019). A Design Construct of Developing Approaches to Measure Mental Health Conditions. *Australasian Journal of Information Systems*, *23*.

Jiarpakdee, J., Tantithamthavorn, C., & Treude, C. (2018). Autospearman: Automatically mitigating correlated software metrics for interpreting defect models. In *IEEE International Conference on Software Maintenance and Evolution 2018* (pp. 92-103). IEEE, Institute of Electrical and Electronics Engineers.

Kotthoff, L. (n.d.). FSelector. Retrieved July 22, 2020, from https://www.rdocumentation.org/packages/FSelector/versions/0.31/topics/cfs

Lavalley, M. P. (2008). Logistic Regression. *Circulation, 117*(18), 2395-2399. doi:10.1161/circulationaha.106.682658

Lin, Y. P., Chu, H. J., Wu, C. F., & Verburg, P. H. (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling–a case study. *International Journal of Geographical Information Science*, *25*(1), 65-87.

Patel, P. (2018). Perceived Workplace Factors and their Influence on Self-Reported Mental Health Service Seeking Among Technology Workers.

Pengpeng Lin. (2009). *A Framework For Consistency Based Feature Selection*. Western Kentucky University
Research. https://digitalcommons.wku.edu/cgi/viewcontent.cgi?article=1062&context=theses

R, V. (2020, February 23). Feature selection - Correlation and P-value. Retrieved July 24, 2020, from https://towardsdatascience.com/feature-selection-correlation-and-p-value-da8921bfb3cf

Reddy, U. S., Thota, A. V., & Dharun, A. (2018, December). Machine Learning Techniques for Stress Prediction in Working Employees. In *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*(pp. 1-4). IEEE.

Sadia, A. (2019). A Comparative Analysis of Four Classification Algorithm for Mental Health Analysis basis on technical People.

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130.