Depression and its Association with Sleep and Other Features

Dilan Bharadwa
CICS
University of Massachusetts, Amherst
Amherst, USA
dbharadwa@umass.edu

Jonas Matos
CICS
University of Massachusetts, Amherst
Amherst, USA
jrmatos@umass.edu

Zenry Padua
CICS
University of Massachusetts, Amherst
Amherst, USA
zpadua@umass.edu

***Abstract*—This study was conducted in order to analyze the prevalence of sleep deprivation and its association with depression. The CDC conducts the National Health and Nutrition Examination Survey (NHANES) every year in to study adults' and children's state of health and nutrition in the United States. Within the NHANES survey there are questionnaires that contain data on patients mental health, more specifically 9 questions that when combined create a measure of depression severity (PHQ-9). The objective of the study is to show how sleep deprivation and other health features could possibly lead to a patient having or developing depression. Machine learning, regression models, and a T-test were the methods used to analyze patient health features and PHQ-9 measure. A two-tailed T-test was conducted on PHQ-9 measures and found that the PHQ-9 scores differed between those who had the recommended hours of sleep to those who had less. The results of the ridge regression found that the model did not fit the Actual PHQ-9 values from the validation set. After conducting a Gaussian process regression the results showed poor model performance. Finally, a random forrest classifaction model was created and the results showed that patient BMI had the greatest importance in depression severity. The findings from these methods show that there was no significant indication that the limited health features chosen from the NHANES data affected the outcome of a patient's depression severity.**

I. Introduction

Based on previous research, problems with sleep have been closely linked with symptoms of depression (Stickley, et al, 2019). Previous studies have also linked the prevalence of sleep disorders with the usage of substances like drugs or alcohol (National Institute on Drug Abuse, 2021). The fact that many people develop sleep and mental health disorders is alarming, given that these disorders have been associated with fatal outcomes (Stickley, et al, 2019). Thus, our study aims to analyze the prevalence of sleep deprivation and their association with depression and substance usage.

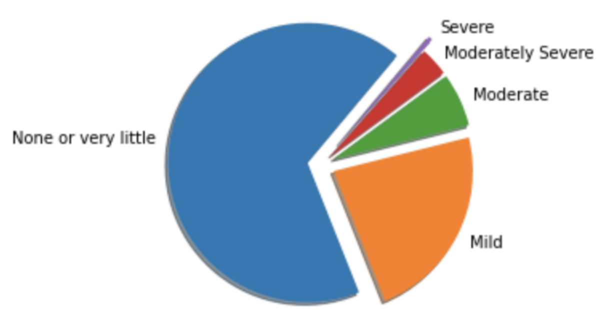
II. Objective

The objective of this study is to show how exposure to sleep deprivation and other physical features can ultimately lead to the outcome of a patient having depression. The data would be collected through 2017-2018 NHANES datasets and machine learning algorithms would be utilized to predict depression using sleep-related and other health-related features. Three different predictive models were constructed, along with a Two-Tailed T-test in order to properly analyze sleep and other health-related features and their association with depression.

III. Methods

NHANES 2017-2018 datasets were used to build machine learning models that used patient's attributes as parameters to predict what level of depression they are based on PHQ-9 measures. The attributes used were: The number of hours of sleep a patient had, Alcohol use, Marijuana use, Physical activity, BMI, Smoking, and Hours working at their job. PHQ-9 measures were used when the depression data was collected with the intention of using said measures as a way to judge the severity of depression of the patient. PHQ-9 measures is based on a 9 question questionnaire given to patients where each question's response ranged from 0 to 3, with the maximum score being 27. The cutoffs used for severity are 5 for Mild, 10 for Moderate, 15 for Moderately Severe, and 20 for Severe Depression. A pie chart was created to show the overall breakdown of PHQ-9 scores to have a better understanding of the NHANES data used in our models.

Fig. 1. Pie chart showing PHQ-9 score breakdown



A quick overview of the PHQ-9 measures show that there is a majority of patients who have mild depression to no depression.

The data was processed through Jupyter notebooks in Python by merging the datasets on patient's individual sequence numbers and then dropping any missing values from the combined dataset. A limitation at this point was that by dropping rows that had any missing values greatly reduced the number of patients in the dataset.

A. Two Tailed T-test

Two Tailed T-test measures if the expected summed PHQ-9 scores between two groups differ. The null hypothesis being that the expected PHQ-9 scores being equivalent to each other. And the alternative hypothesis being that the expected PHQ-9 scores are not equivalent to each other.

The population is split into two groups according to the CDC recommended hours of sleep, which would be 7 hours or greater. The first group would be the patients that received 7 or more hours of sleep and the second group would be the patients who received less than 7 hours of sleep. Once the groups were separated from the population a sample of 50 patients each were randomly chosen.

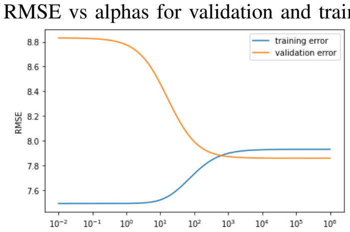
A T-test was conducted between the two groups which resulted in a test statistic of -2.53 and p-value of .013. From the t-statistic we can see that group 1's expected value is actually less compared to group 2's. This means that group 1, who had the CDC recommended amount of sleep, suffered less depressive symptoms than group 2. And if we were to use a significance level of 5%, we can confirm this significance. As the p-value shows that the null hypothesis should be rejected and we should accept the alternative hypothesis. Thus we can conclude that there is a significant change of depression severity between patients that sleep the recommended amount of hours and the patients that don't.

B. Ridge Regression

Ridge regression is a machine learning regression model that works best in instances where there is multicollinearity, which is when variables are independent but highly correlated with each other. Ridge regression uses a penalty function to prevent multicollinearity from affecting the results of the regression. In this study ridge regression was chosen because there are six variables that are being analyzed which could be correlated linearly with one another.

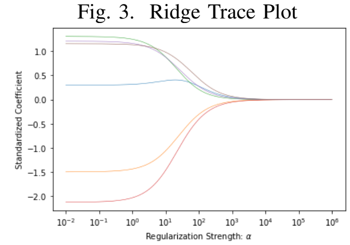
To start the ridge regression the data needs to be standardized and then split into training and testing sets. 70% of patients are included in the training set and the remaining patients are in the testing set. Next the optimum alpha value needs to be determined which was done by creating a plot of root mean square error vs alphas for validation and training error.

Fig. 2. RMSE vs alphas for validation and training error



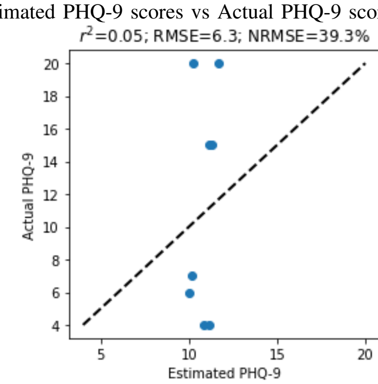
Looking at this plot the error settles when alpha is 10^2. Another way that was used to determine the alpha value for the ridge regression was by using a ridge trace plot. This chart is the standardized coefficient of the six features vs alphas.

Fig. 3. Ridge Trace Plot



The curves all converge towards zero when alpha is about 10^2. Therefore the alpha value for the ridge regression is 10^2. Next is to fit the ridge regression model from the sci-kit library on the patients in the training set and predict PHQ-9 of the testing set. To see the performance and accuracy of the ridge regression a plot of the actual PHQ-9 values for the patients in the testing set can be compared to the predicted PHQ-9 values of those patients.

Fig. 4. Estimated PHQ-9 scores vs Actual PHQ-9 scores of testing set



C. Gaussian Process Regression

Gaussian Process Regression is a machine learning algorithm that implements Gaussian Processes in order to solve regression problems. This type of regression model typically utilizes a kernel function, a function used to measure the similarity between data points, in order to predict the values for unseen data points from the training data. Multiple Gaussian Process Regression models were created in order to predict the PHQ-9 scores of subjects using sleep and other health-related features.

Before creating the Gaussian Process Regression model, the data was split into training and test sets, with 70% of subjects included in the training set and the remaining 30% of subjects in the test set. The data also needs to be standardized, so a standard scaler was used in order do so. The standard scaler essentially removes the mean and scales the features to have unit variance. After standardizing the data, scatter plots were created in order to show the relationship between the estimated and actual PHQ-9 scores.

Fig. 5. Estimated PHQ-9 Scores vs Actual PHQ-9 Scores of Testing Set (with most ”important” features)

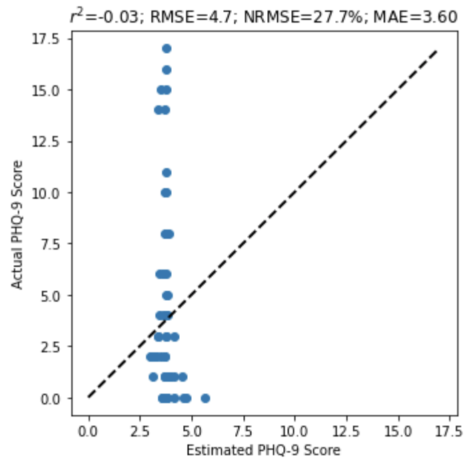
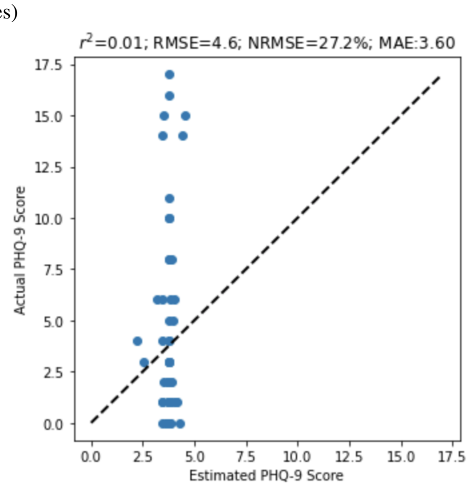


Fig. 6. Estimated PHQ-9 Scores vs Actual PHQ-9 Scores of Testing Set (with all features)



As can be seen above, the regression model that only included the four most ”important” features performed worse than the regression model that included all of the original features. This conclusion can be reached by looking at the scatter plots for the respective models and noticing that the second model has a higher R^2 score value and lower RMSE and NRMSE values.

Along with the scatter plots, useful metrics like the R^2 score, Root Mean Squared Error, Normalized Root Mean Square Error, and the Mean Absolute Error were computed in order to appropriately evaluate the performance of the regression models.

D. Random Forest Classification

In the case of the Random Forest Classification, the PHQ-9 summed score was converted into the labels of the appropriate PHQ-9 category. A Random Forest Classifier is made up of many decision trees that make splits depending on the features of the dataset. So the random forests have treated sleep as a feature when determining the PHQ-9 category. The impact each feature has upon determining the category can be measured in a value called Feature Importance. Feature Importance essentially shows how a feature is weighted in the model.

Forest 1 used all the features and had an accuracy of 68% and, interesting enough, BMI had the highest Feature Importance. In Forest 2, features below a Feature Importance of .15 were removed to potentially get rid of some noise from the model. However, removing the features of Marijuana use, physical activity and alcohol use and creating another random forest based on the remained features resulted in a lower accuracy rating of 65%. This indicated that some of those features were necessary in the classification of the PHQ-9 categories. And so in Forest 3, a random forest was made by only excluding the drug use feature as it is the only feature that had a Feature Importance less than .1. The model ended up performing similarly to Forest 1 with an accuracy of 68%. An interesting observation in Forest 2 and 3, is that OCQ, which stands for hours working at a job, had received a higher Feature Importance score than sleep.

Fig. 7. Random Forest Classification Model 1

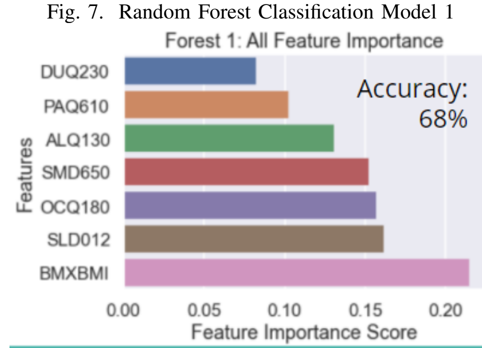
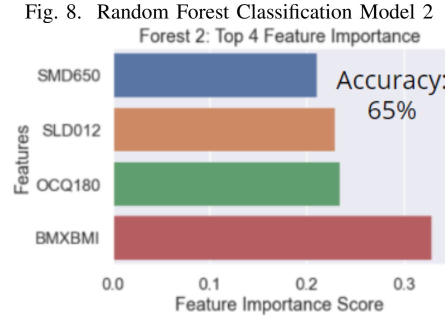


Fig. 8. Random Forest Classification Model 2



IV. Results

The results from the ridge regression model indicate that the model failed to fit the data. The r^2 value that was calculated is 0.05 meaning that the regression was only 5% accurate indicating that the predicted PHQ-9 scores were not similar

Fig. 9. Random Forest Classification Model 3

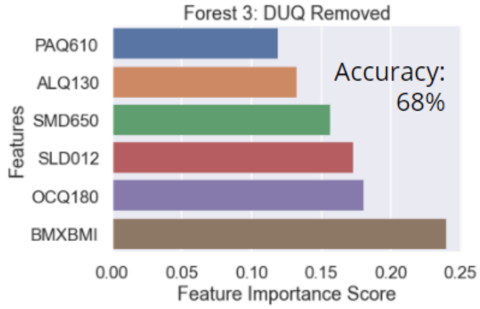
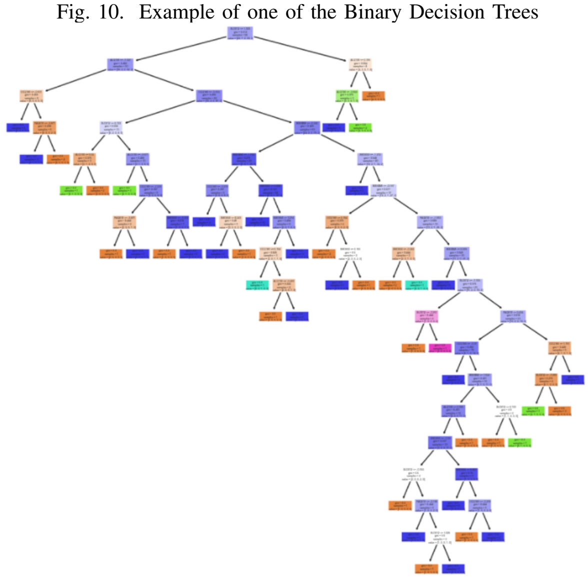


Fig. 10. Example of one of the Binary Decision Trees



to the actual PHQ-9 scores of the patients. Figure 4 suggests that the six features chosen to predict depression severity do not make an impactful difference. The root mean square error was 6.3 showing that the model contained a decent amount of error. Overall, ridge regression failed to show any correlation between the six patient health features and PHQ-9 depression severity.

Based on the results of the Gaussian Process Regression model, the model had bad performance indicating that the features included did not have a significant effect on a subject's PHQ-9 score (severity of depression). The R^2 value that was calculated is -0.03 for the first regression model and 0.01 for the second regression model. These low R^2 values indicate that the models did a poor job of explaining the variability in the outcome data.

From the Random Forest model that accounted for all features, BMI had the highest Importance at .214, followed by Sleep, Working Job Hours, and Smoking at .161, .157 and .152, respectively.

V. Conclusion

With the features included, the Gaussian Process Regression model did not perform well. This is an indication that more informative features need to be considered in order to determine whether sleep deprivation has a strong impact on individuals who have depression.

The Random Forest model demonstrates that sleep deprivation does not have the most impact upon whether an individual has depressive symptoms within our chosen features. Removing some of the lesser ranked features below .15 from the model decreases accuracy and changes some of the features weight. Removing features ranked below .1 kept the original accuracy and redistributed the weights. It was interesting to find that among the features that had been selected, sleep was sometimes a distant second place in having the highest feature importance in determining the depression severity groups.

The ridge regression model failed to perform well, suggesting that there may not be enough features that have an effect on the outcome. Another possible issue for why the model failed could be that the data may not have been multicollinear which is the optimal type of features for this type of regression.

In conclusion, it seems that our models lacked enough impactful features from the NHANES data in determining a patient's depression severity. All that can be determined is that depression cannot be predicted by a small amount of features about a person's life.

References

[1] Huyett, P., Siegel, N., Bhattacharyya, N. (2020, July 18). "Prevalence of Sleep deprivation and Association With Mortality: Results From the NHANES 2009–2010". PubMed. https://pubmed.ncbi.nlm.nih.gov/32681735/.

[2] National Institute on Drug Abuse. (2021, March 09). Connections between sleep and substance use deprivation. Retrieved April 01, 2021, from https://www.drugabuse.gov/about-nida/noras-blog/2020/03/connections-between-sleep-substance-use-deprivation.

[3] NHANES 2017–2018 Questionnaire Data. (2020) https://wwwn.cdc.gov/nchs/nhanes/Search/DataPage.aspx?Component=Questionnaire&CycleBeginYear=2017.

[4] Stickley, A., Leinsalu, M., DeVylder, J., Inoue, Y., Koyanagi, A. (2019, August 19). Sleep problems and depression AMONG 237 023 community-dwelling adults in 46 low- and MIDDLE-INCOME COUNTRIES. Retrieved April 01, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6700183/.