Depression and its Association with Patient Health

By: Dilan Bharadwa, Jonas Matos, and Zenry Padua

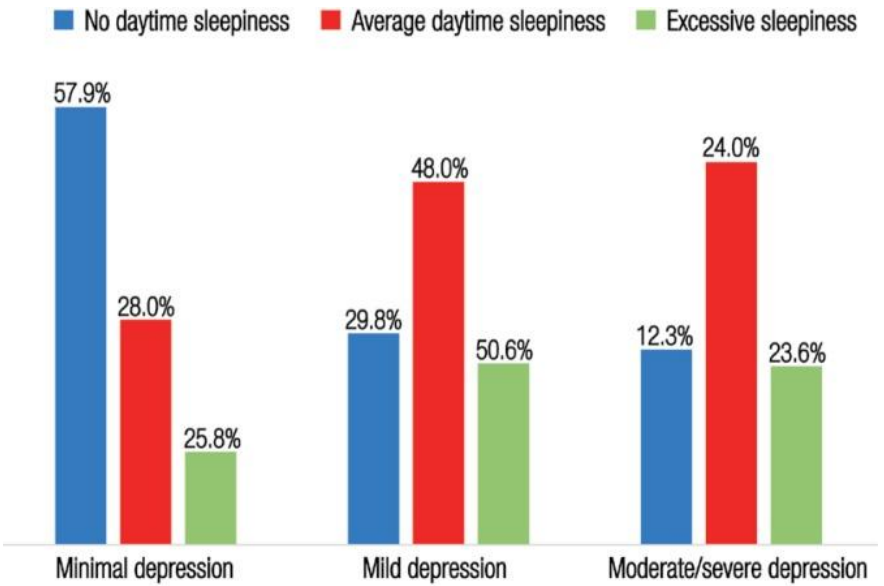
Introduction/Motivation

- Based on previous years of research, sleep problems have been heavily associated with symptoms of depression (Stickley, et al, 2019).
- There have also been various studies that have linked the prevalence of sleep disorders with the usage of substances like drugs and alcohol (National Institute on Drug Abuse, 2021).
- It is concerning to see that many individuals develop sleep and mental health disorders, given that these disorders have been associated with mortal outcomes (Stickley, et al, 2019).
- Consequently, our study aims to analyze the prevalence of sleep deprivation and its association with depression using NHANES datasets.

Background + Previous study

Prevalence of Sleep Deprivation and Relation with Depressive Symptoms among Medical Residents in King Fahd University Hospital, Saudi Arabia (2015)

- Measures used: Epworth Sleepiness Scale and Beck Depression Inventory
- Results: Significant associations between sleep deprivation and depressive symptoms, and depressive symptoms and sleepiness, depressive symptoms and being a female resident

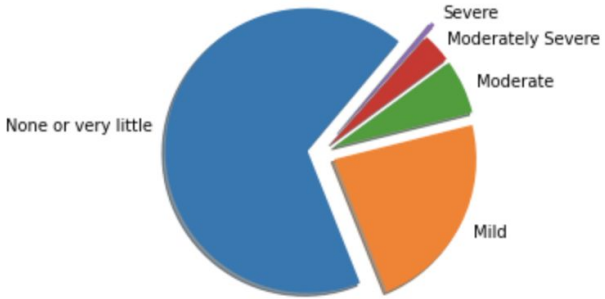


Al-Maddah EM, Al-Dabal BK, Khalil MS. Prevalence of Sleep Deprivation and Relation with Depressive Symptoms among Medical Residents in King Fahd University Hospital, Saudi Arabia. Sultan Qaboos Univ Med J. 2015 Feb;15(1):e78-84. Epub 2015 Jan 21. PMID: 25685390; PMCID: PMC4318611.

PHQ-9 Measures

The PHQ-9: validity of a brief depression severity measure

- Patients answer 9 depression screening questions scored 0-3
- Summing these scores creates a measure for depression severity, 0-27
- Cutoffs specifying severity of depression
  - 5, 10, 15, and 20 → mild, moderate, moderately severe, and severe depression
- PHQ-9 was used as the target values in our models
- PHQ-9 score percentage breakdown:
  - 67.14% - None or very little (0-4)
  - 22.86% - Mild (5-9)
  - 6.19% - Moderate (10-14)
  - 3.33% - Moderately severe (15-19)
  - 0.48% - Severe (20-27)



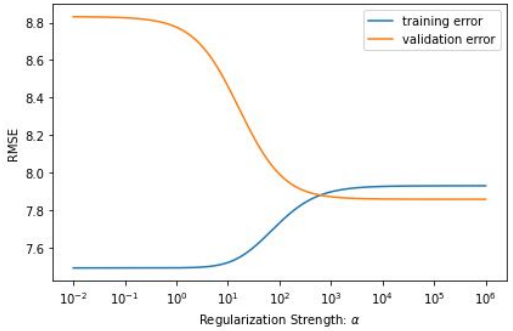
Methods - Two Tailed T-test

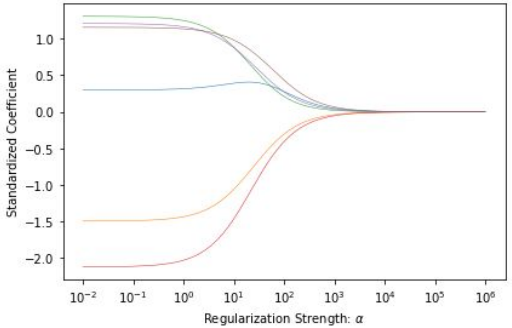
- A T-test measures if the expected summed PHQ-9 scores between the two groups differ. The CDC recommends at least 7 hours of sleep per day for Adults.
  - Group 1: Good sleep group (Sleep >= 7 Hours)
  - Group 2: Bad sleep group (Sleep < 7 Hours)
- Null Hypothesis: $\mu_1 = \mu_2$
- Alternative Hypothesis: $\mu_1 \neq \mu_2$

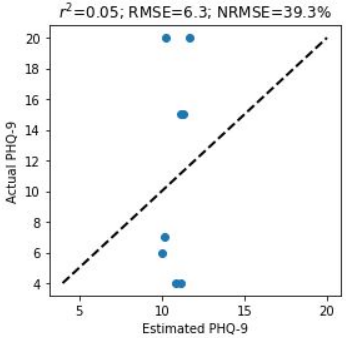
- T-statistic: -2.53
- P-value: 0.013

Methods - Ridge Regression

- Ridge regression work best in situations where there are many highly correlated independent variables
- Ridge regression was used to predict PHQ-9 scores using sleep related and other features
- Regularization Strength (α) was chosen by creating a ridge trace plot and MSE plot
- Estimated PHQ-9 vs Actual PHQ-9 was plotted

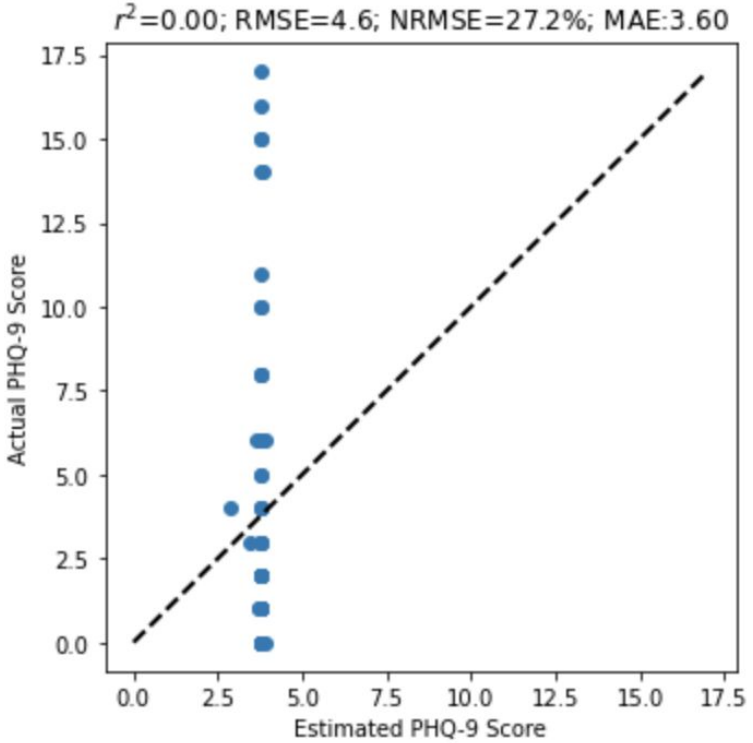






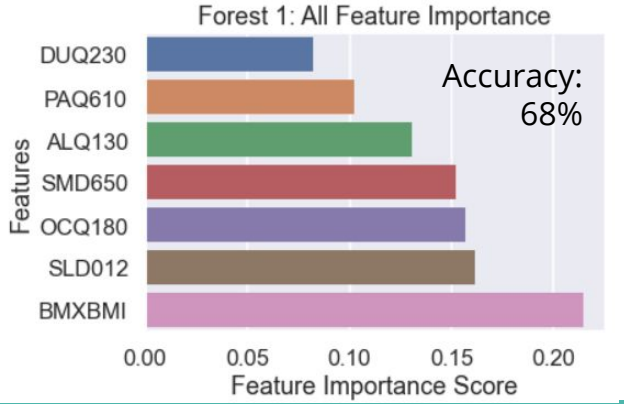
Methods - Gaussian Process Regression

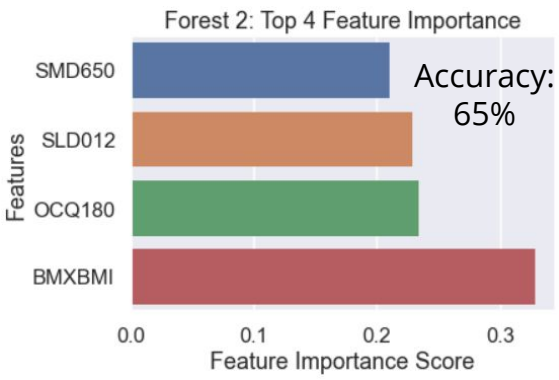
- Gaussian Process Regression is a machine learning algorithm that implements Gaussian Processes in order to solve regression problems.
- It involves the use of a **kernel function**, a measure of the similarity between data points, in order to predict the values for unseen data points from training data.
- Created a Gaussian Process Regression model in order to predict the PHQ-9 scores of subjects using sleep-related and other features.
- A scatter plot was created to show the relationship between the estimated and actual PHQ-9 scores, along with useful metrics like R^2, RMSE, NRMSE and MAE.

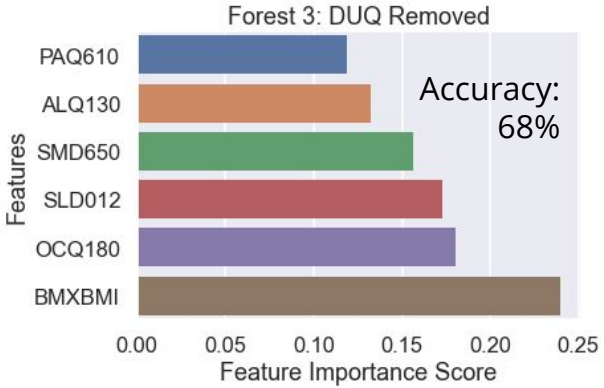


Methods - Random Forest Classification

- A Random Forest Classification is essentially creating many decision trees where the best splits are found from all input features or a random subset of the features.
- The focus of the model was to explore how sleep and other features relate to depression on the PHQ-9 scale.







Results

- 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.
- Looking at the results of the Ridge Regression, the features included do not have a strong effect on a patient's depression severity
  - R^2 = 0.05 → The model fails to fit the data, RMSE = 6.3, NRMSE = 39.3%
- Based on the results of the Gaussian Process Regression model, we can see that 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)
  - R^2 = 0.00, RMSE = 4.6, NRMSE = 27.2%, MAE = 3.60
- One thing to note is that even though the RMSE and MAE values were low, the model still performed relatively bad. In other words, metrics like RMSE and MAE can be misleading and don't tell the whole story!

Conclusion

- 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.
- With the features we 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 Ridge Regression model created did not perform well. This is most likely due to the fact that there were not enough impactful features included in the model

Discussion Questions

- What other features could have been included to help improve our model’s accuracy and overall performance?