Machine Learning: Determining Risk of Suicide From Nocturnal Sleep Behavior

Team: Natalie Fung, Maya Ramde, & Ellyse Robert

Problem Definition

Sleep quality is known to affect emotional regulation, with poor sleep quality being strongly associated with depression (Goldstein and Walker, 2014). Depression has become increasingly common in recent years and with that comes an uptick in suicidal thoughts and actions, especially in young adults. In a study conducted at Emory University in 2007, 11.1% of undergraduate students reported current suicidal ideation (Garlow, et al., 2008). We found a study entitled *Assessing Nocturnal Sleep/Wake Effects on Risk of Suicide* from the National Sleep Research Resource (NSRR) which collected data from a survey pertaining to quality of sleep and mental illness amongst undergraduate students. This survey is a collection of smaller surveys covering topics of sleep behavioral patterns, mental health issues, physical health issues, substance abuse and more. Based on each subject's scores in the individual sub-surveys, NSRR researchers were able to use the diagnostic medical questionnaires to objectively score participants on their sleep/behavioral patterns. With sleep behavior and mental health disorders being an ever-present problem in college campuses along with suicidal ideation among young adults seemingly on the rise, our team decided to explore the correlation between risk of suicide and nocturnal sleep behavior.

Description of Background

Cerel, et al. (2016) found that 48% of participants in their study reported lifetime exposure to suicide which means that they knew someone who lost their life to suicide. With almost half of the population being personally affected by this tragedy, exploring factors that could contribute to an increased risk of suicide may help decrease suicidal attempts or behaviors in the future. Our team is looking specifically whether sleep behaviors have an effect on risk of suicide and even further, which behaviors are the most predictive of having suicidal thoughts. Disordered sleep has been known to affect emotional regulation and that it often predates depression (O'Leary, et al., 2016). Furthermore, nearly all mood and anxiety disorders co-occur with one or more sleep abnormalities (Goldstein and Walker, 2014). We are interested in seeing if there is a way to predict the risk of suicide on a binary scale based on sleep factors.

Description of Dataset

For our project, we requested data file access from the aforementioned study which was granted to our team for academic use only. The study's collection of .csv files is downloaded as a zip file, which contains participants' survey answers and three data dictionary files that map variable names to their respective survey questions. In total, the 'participant' data set contains answers for over 150 survey questions and multiple diagnostic questionnaire scores. Each survey question [variable] is a part of the domain of a medical diagnostic questionnaire and the survey response is used to calculate a value which contributes towards a specific medical diagnostic score index. The following medical diagnostic tests/questionnaires are represented in the dataset: Pittsburgh Sleep Quality Index [PSQI], Brief Inventory of Sleep Control [BISQ], Insomnia Severity Index [ISI], Short UPPS-P measure of impulsivity [UPPS], Center for Epidemiologic Studies Depression Scale [CESD], Generalized Anxiety Disorder – 7 item scale [GAD-7], Sleep Disorders Symptoms Check List – 25 [SDS], Columbia Suicide Severity Ratings Scale [CSSRS], and Disturbing Dreams and Nightmares Severity Index [DDNS]. We would like to make note of the fact that each of the smaller surveys has been verified for their credibility and statistical significance as they are widely used and accepted by healthcare providers on a global scale.

Dataset Construction: *Mental Health Scores*

Using SQL, we built a table called '*Mental Health Scores*' that contains participants' scores on mental health disorder questionnaires, their resulting diagnoses from the questionnaires, and data on whether each participant had been clinically diagnosed with the given mental health disorder. The overall NSRR survey tested for depression via the CESD test, anxiety via the GAD-7 questionnaire, impulsivity via the UPPS questionnaire, and suicide ideation severity via the CSSRS test.

After compiling relevant data into the 'Mental Health Scores' table, we decided to use the remaining sleep study data to classify participants' risk of suicide. Using two suicide ideation severity test scores and participants' reported suicide attempts [CSSRS], we created a metric called *Risk of Suicide*, which is a binary classification denoting whether the survey participant is at risk of suicide (= 1) or not (= 0). If the participant received a score of 3 or greater on the suicide ideation severity indices (total range of 0 - 4) or the participant attempted suicide in their lifetime and/or the last 3 months, then the participant is determined to be 'at risk of suicide.'

Participant ID	Suicidal ideation severity in last 3 months	Suicidal ideation severity in lifetime (Scale 1-5)	Suicide attempt in past 3 mos	Suicide attempt in lifetime	Risk of Suicide
2	0	1	0	0	0
3	0	0	0	0	0
4	1	2	0	0	0
5	0	0	0	0	0
6	0	4	0	0	1

Figure 1. The image above shows the fields of the *Mental Health Scores* tables used to compute the target variable of our project.

The calculated *Risk of Suicide* metric will be used as the target variable in our machine learning classification model. Using participants' sleep study data, we aim for the classification model to determine whether the participant is 'at risk' or 'not at risk.'

Dataset Construction: *Sleep Study*

Next, we built another SQL table called 'Sleep Study*'* which holds the input variables for our data science project.

- 2 sleep behavior diagnostic index scores: *Sleep Control Score* and *Nightmare Severity Score*. The *Sleep Control Score* is the overall output from the Brief Inventory of Sleep Control test. The *Nightmare Severity Score* is the overall output from the Disturbing Dreams and Nightmares Severity Index.
- 8 sleep quality subscores: *Global Sleep Quality Score, Subjective Sleep Quality Score, Sleep Latency, Sleep Duration, Sleep Efficiency, Use of Sleep Medication, Daytime Dysfunction, Fatigue, Excessive Daytime Sleepiness, Sleep-related TMJ, Average Sleep Duration, REM Sleep Behavior Disorder,* and *Number of Awakenings During Night*. Each of these variables' values represent a score from 0 - 4.
- 3 demographic variables: *Age, Sex,* and *Race*. We chose to include demographic details as they may later serve to stratify trends possibly found during analysis.
- 3 survey questions related to substance use prior to bedtime: *Tobacco Use, Cannabis Use,* and *Alcohol Use*. The variables incorporate self-medication as a possible factor toward decreased mental health and possible risk of suicide. The values for these variables are categorical string variables

Data Pre-Processing

We observed that the *Age* column had some null values. Since >90% of the study was conducted on undergraduate students, we imputed the missing *Age* values with the mean age of the population [Mean Age = 20]. Additionally, four categorical variables contained text of the string data type as their values. To standardize all input variables to have numeric values, we transformed the following variables: *Sex, Tobacco Use, Cannibis Use,* and *Alcohol Use.* For the *Sex* variable, we equated 'Female' = 0 and 'Male' = 1. The other three variables had the same possible values, which we translated into numeric representation: 'Never' = 0, 'Rarely' = 1, and 'Often' = 2. Our final dataset is the entire *Sleep Study* table with the *Risk of Suicide* field appended as the target variable.

Description of Methods Used

Moving forward, we wanted to take this cleaned and prepared dataset into some preliminary analysis to see which of the 21 available input features would provide the best predictive Machine Learning model. First, we computed the average risk of suicide for each input feature and its corresponding values to obtain a granular interpretation of our data. Then, we visualized the relationships between the input variables and the target variable by creating frequency stratified by risk of suicide. Input variables with skewed frequency distributions can cause the data set to be imbalanced and impact model accuracy. We observed that *Age, Sex, Race, Sleep Duration, Use of Sleep Meds, Daytime Dysfunction, Excessive Daytime Sleepiness, REM Sleep Behavior Disorder, Sleep-Related TMJ, Cannabis Use, Tobacco Use, and Alcohol Use* were imbalanced variables due to their skewed frequency distributions. Figure 2 reveals that there is a significantly higher number of people who reported 'Never' to using tobacco or alcohol to help as opposed to those who responded with 'Often' or 'Rarely.'

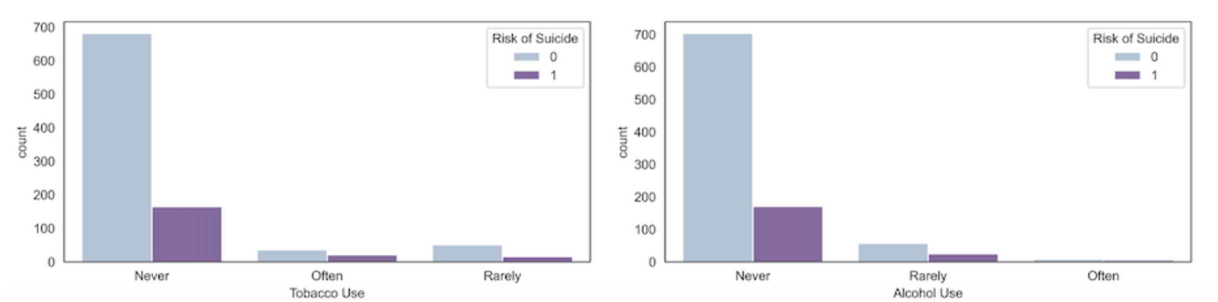

Figure 2. The images above show the skewed frequency distribution [stratified by *Risk of Suicide*] of *Tobacco Use* and *Alcohol Use* stratified with *Risk of Suicide.*

We also observed that our output variable, *Risk of Suicide,* is imbalanced, but that is expected in this study due to the actual percentage of suicide being very small in comparison to population size. However, we still aim to construct an accurate model given the imbalance.

After dropping the aforementioned input variables from the data set, we constructed a correlation matrix to view which features had the highest association with having a *Risk of Suicide* score of 1. A preliminary correlation matrix revealed that *Sleep Disturbance* and *# of Mid-night Awakenings* had low correlation to *Risk of Suicide*, resulting in being dropped from our data. After this removal, we developed a correlation heatmap to visualize our finalized input variables.

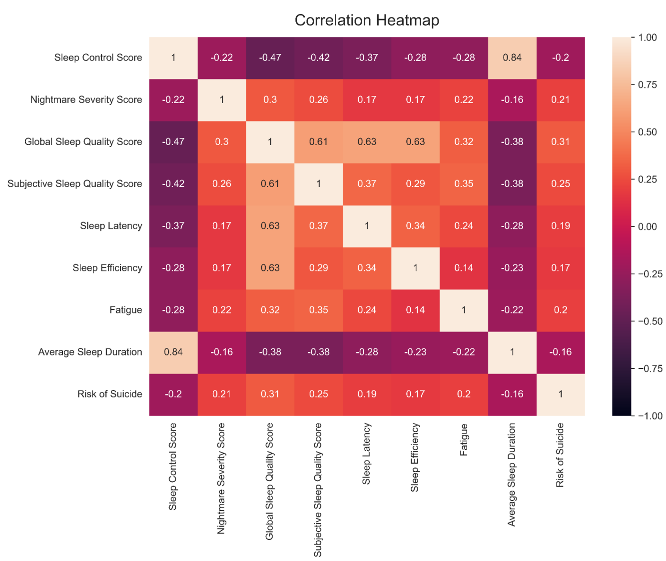

Figure 3. The image above shows the correlation matrix heatmap.

After we removed the skewed variables and variables with low correlation, we were left with a final dataset consisting of seven input variables. These variables include *Sleep Control Score, Nightmare Severity Score, Global Sleep Quality Score, Subjective Sleep Quality Score, Sleep Latency, Sleep Efficiency, Fatigue,* and *Average Sleep Duration.*

- *Sleep Control Score:* Score assigned based on averaged responses to a 4-item questionnaire pertaining to control over subjects' duration and quality of sleep (0: No control - 4: Complete control)

- *Nightmare Severity Score:* Score assigned based on summed responses to a 5-item questionnaire pertaining to frequency and intensity of the subjects' nightmares (Score out of 37).
- *Global Sleep Quality Score:* Overall score of Pittsburgh Sleep Quality Index Survey. (Score out of 21)
- *Subjective Sleep Quality Score:* Self-reported score based on subjects' opinions of their quality of sleep (0: Very good - 3: Very bad)
- *Sleep Latency:* Self-reported number of minutes it takes each subject to fall asleep at night. Values are then assigned depending on ranges of time (0-15 min = 0, 15-30 min = 1, 30-60 min = 2, >60 min = 3).
- *Sleep Efficiency:* Score based on ranking of percentage of time subject is actually asleep in bed over total time spent in bed (>85% = 0, 75-85% = 1, 65-75% = 2, <65% = 3)
- *Fatigue:* Number of times a subject has daytime sleepiness (0 = "Never", 1 = "Once a month", 2 = "1-3 times a week", 3 = "3-5 times a week", 4 = ">5 times a week")
- *Average Sleep Duration:* Number of hours of sleep per night on average

Experiment: Experiment Setup and Analysis Results

For our experiment, we decided to build a function that runs six machine learning and outputs their associated training accuracy. The ML models we used for this consisted of Naive Bayes Classifier, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine (linear and RBF classifier), Logistic Regression, and K-Nearest Neighbor. After running this function, we found that the training accuracies were as follows:

```
[0]Logistic Regression Training Accuracy: 0.8157216494845361
[1]K Nearest Neighbor Training Accuracy: 0.8337628865979382
[2]Support Vector Machine (Linear) Training Accuracy: 0.8041237113402062
[3]Support Vector Machine (RBF) Training Accuracy: 0.8234536082474226
[4]Gaussian Naive Bayes Training Accuracy: 0.7603092783505154
[5]Decision Tree Classifier Training Accuracy: 0.9987113402061856
[6]Random Forest Classifier Training Accuracy: 0.9819587628865979
```

Figure 4. The image above shows the training accuracy for all seven ML models.

After seeing these values, the Decision Tree Classifier and Random Forest Classifier fit the training data best. Since the input variables are very specific and the output variable is a simple binary classification, these training accuracy results were not surprising. However, we were wary that overfitting may have occurred, so we conducted K-fold cross validation and built a function to cross-validate each model. Our cross validation function gave each machine learning model an F1 score and a Precision Score. We specifically wanted to calculate the F1 score for each model since our target variable had an uneven class distribution.

```
Model[3] F1 Scores = [0.         0.16666667 0.16666667 0.         0.        ]
Model[3] Average F1 Score = 0.067
Model[3] Precision = [0.56276608 0.45468618 0.45477969 0.44068826 0.44060543]
Model[3] Average Precision = 0.471

Model[4] F1 Scores = [0.63157895 0.4        0.5        0.56       0.25      ]
Model[4] Average F1 Score = 0.468
Model[4] Precision = [0.72353038 0.41517544 0.4320649  0.4961261  0.43053175]
Model[4] Average Precision = 0.499

Model[5] F1 Scores = [0.45454545 0.33333333 0.25       0.24       0.14285714]
Model[5] Average F1 Score = 0.284
Model[5] Precision = [0.33653846 0.29198718 0.24377289 0.23948718 0.25576923]
Model[5] Average Precision = 0.274

Model[6] F1 Scores = [0.30769231 0.375      0.21052632 0.13333333 0.15384615]
Model[6] Average F1 Score = 0.236
Model[6] Precision = [0.4207604  0.39739446 0.35392791 0.27151721 0.31160594]
Model[6] Average Precision = 0.351
```

Figure 5. The image above shows the cross-validation F1 scores, and precisions of each ML model.

```
[[141   4]
 [ 41   9]]
Model[0] Testing Accuracy = 0.7692307692307693

[[133  12]
 [ 42   8]]
Model[1] Testing Accuracy = 0.7230769230769231

[[145   0]
 [ 50   0]]
Model[2] Testing Accuracy = 0.7435897435897436

[[137   8]
 [ 44   6]]
Model[3] Testing Accuracy = 0.7333333333333333

[[117  28]
 [ 25  25]]
Model[4] Testing Accuracy = 0.7282051282051282

[[117  28]
 [ 33  17]]
Model[5] Testing Accuracy = 0.6871794871794872

[[133  12]
 [ 42   8]]
Model[6] Testing Accuracy = 0.7230769230769231
```

Figure 6. The image above shows the confusion matrices and testing accuracies of each ML model.

After conducting K-fold cross validation on our data models, we also tested for model accuracy using our test data set, which is 20% of the whole dataset. The results from cross validation scores and the testing accuracy scores were used to validate observations drawn from the training accuracy scores.

Observation and Conclusion

Overall, the training and test accuracies were both relatively strong, however we noticed that those with a higher training accuracy tended to have a lower testing accuracy, most likely due to overfitting. This can be seen with the Decision Tree Classifier which had a 99.9% training accuracy but a 68.7% testing accuracy. In order to choose our most optimal prediction model, we decided to go with one that had a more consistent accuracy values for training and test since that is more indicative of a strong model. By using this logic, we chose the Naive Bayes classifier which has a training accuracy of 76.0% and a testing accuracy of 72.8%. We also wanted to factor in the F1 score since the original dataset was imbalanced and data for our output variable is imbalanced. The Naive Bayes classifier also had the highest F1 Score of all the models at 46.8%. In conclusion, we found that the Naive Bayes classifier was the best model overall for predicting the risk of suicide amongst undergraduate students. We found that the most predictive sleep variables for an increased suicidal risk were *Sleep Control Score, Nightmare Severity Score, Global Sleep Quality Score, Subjective Sleep Quality Score, Sleep Latency, Sleep Efficiency, Fatigue,* and *Average Sleep Duration.* In the future, we hope that this test can be utilized as a resource for helping people gain insight into potential mental health problems relating to sleep.

References

Cerel, J., Maple, M., van de Venne, J., Moore, M., Flaherty, C., & Brown, M. (2016). Exposure to Suicide in the Community: Prevalence and Correlates in One U.S. State. *Public health reports (Washington, D.C. : 1974)*, *131*(1), 100–107. https://doi.org/10.1177/003335491613100116

Garlow, S.J., Rosenberg, J., Moore, J.D., Haas, A.P., Koestner, B., Hendin, H. and Nemeroff, C.B. (2008), Depression, desperation, and suicidal ideation in college students: results from the American Foundation for Suicide Prevention College Screening Project at Emory University. Depress. Anxiety, 25: 482-488. https://doi.org/10.1002/da.20321

Goldstein, A. N., & Walker, M. P. (2014). The role of sleep in emotional brain function. Annual Review of Clinical Psychology, 10, 679–708. doi: 10.1146/annurev-clinpsy-032813-153716

O’Leary, K., Bylsma, L.M., & Rottenberg, J. (2017). Why might poor sleep quality lead to depression? A role for emotion regulation, Cognition and Emotion, 31:8, 1698-1706, DOI: 10.1080/02699931.2016.1247035