FINAL REPORT
STUDENT STRESS MONITORING PROJECT
IS 507- Data, Statistical Methods, and Information

Yuyang Liu, Xuyi Wu, Mohamad Afrillian Ramadhan, Lexxie Liu, Le Kim Ngan Hoang

1. Introduction:

Our group shared an interest in understanding the factors that shape student stress. We aim to explore approaches that can help universities identify and address these challenges. Prior work shows that college students face many stressors during this stage of life (Buchanan, 2012), and supporting their coping abilities is essential for maintaining mental health (Kumaraswamy, 2013). Our project examines student stress using the publicly available dataset from Kaggle. We selected this dataset for its usability and alignment with our project goal. Its prior academic use makes it a credible source for developing data-driven insights that may inform stress-management initiatives at UIUC.

2. Dataset overview:

Our dataset is sourced from the comprehensive Student Stress Monitoring Datasets from Kaggle (https://www.kaggle.com/datasets/mdsultanulislamovi/student-stress-monitoring-datasets/data).
It contains 1,100 observations, 21 features, aiming to predict the dependent variable, “stress_level”.
The predictors are structured across five scientifically identified domains: Psychological, Physiological, Environmental, Academic, and Social, which primarily consist of quantitative ratings.

3. Experiment Pipeline

3.1 EDA

There are no missing values, and all variables fall within their expected rating scales. We then examined the distribution of the outcome variable “stress_level“, which encodes three categories: 0 = no stress, 1 = moderate/normal stress, and 2 = high stress. According to Figure 1, the counts are 373 students with no stress (33.9%), 358 with moderate stress (32.5%), and 369 with high stress (33.5%).

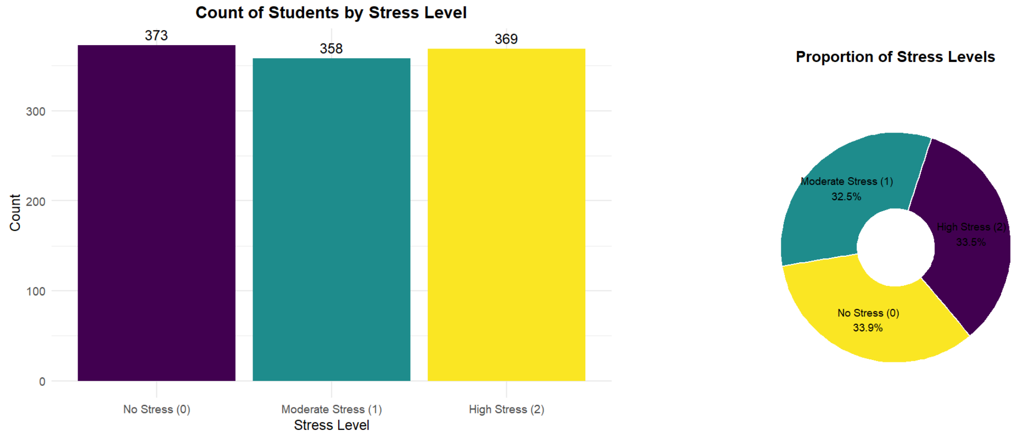


Figure 1. Stress level distribution

3.2 Feature selection

Feature Distribution

Good distribution features with high granularity are shown in Figure 1. However, the dataset also includes binary variables, such as “mental_health_history” (values 0 and 1), and low-cardinality features like “blood_pressure”. They can only indicate the presence or absence of a factor, failing to quantify its intensity or duration. Those less informative examples are illustrated in Figure 2.

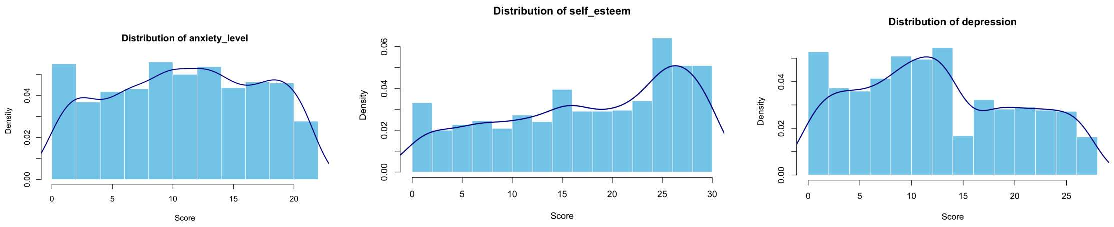


Figure 2. Good distributions

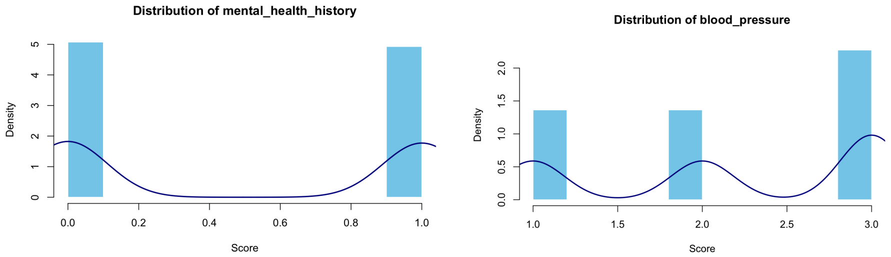


Figure 3. Less informative Distributions

Feature Importance Ranking (ANOVA)

To systematically rank the features, a one-way Analysis of Variance (ANOVA) was conducted. This statistical test compares the means of each feature across the distinct groups defined by the target variable, “stress_level”. The primary metric for this evaluation was the F-statistic as shown in Figure 3 with top 10 highlighted.

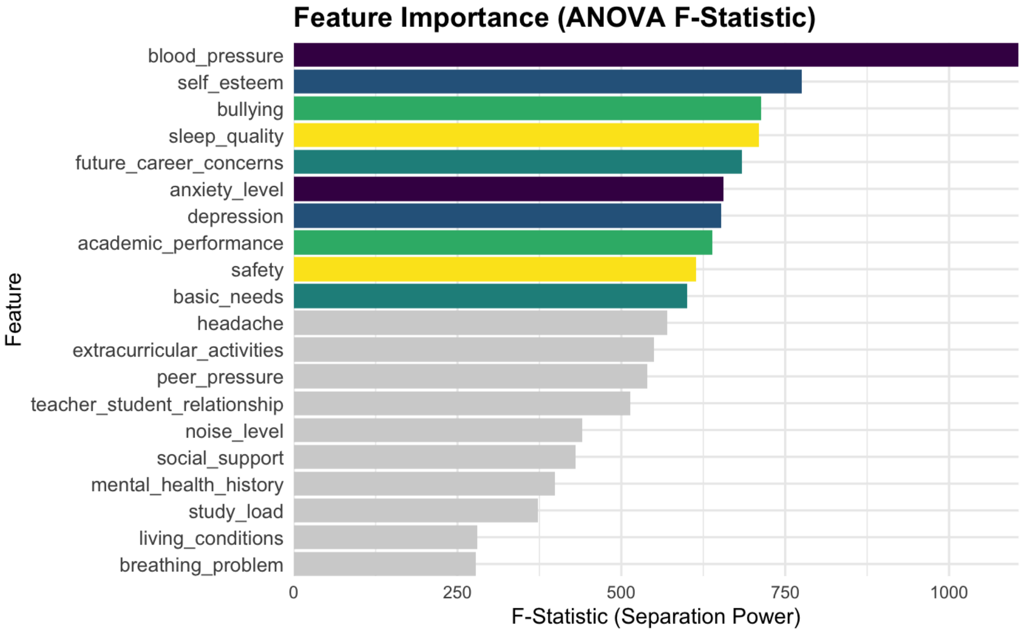


Figure 4. Feature Importance Ranking

From Figure 3, the low granular feature “blood_pressure” yielded the highest F-statistic (F≈1106), far larger than the second rank “self_esteem”. This suggests that while it lacks the nuance of a continuous scale as we discussed in feature distribution, the distinct categories it provides are exceptionally strong separators of stress levels. The granular features self_esteem (F≈775) and depression (F≈652) ranked highly.

Statistical Significance Verification (P-Value Analysis)

To validate the ANOVA findings, the statistical significance of the differences in group means was assessed using p-values. Even the feature with the lowest discriminatory power, “breathing_problem”, produced a p-value of approximately 1.77□10^{-98} (Figure 4). This near-zero probability indicates an extremely high level of confidence that the feature varies meaningfully with stress levels.

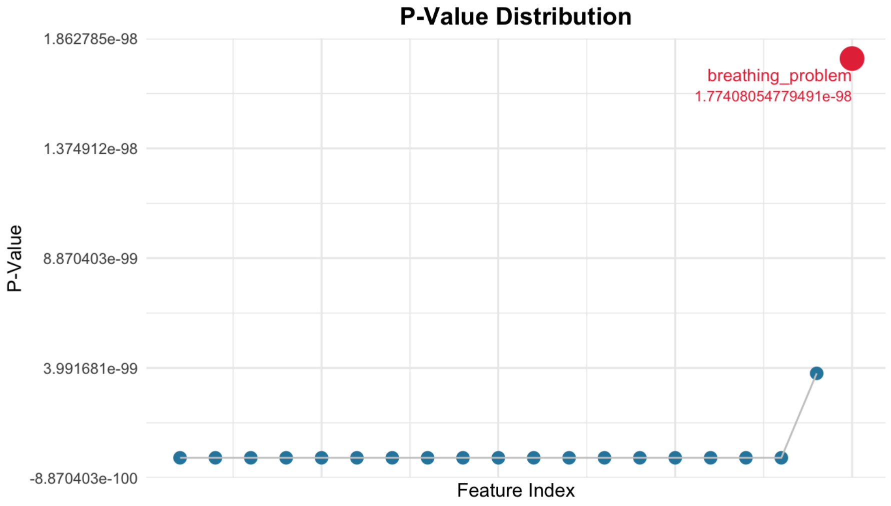
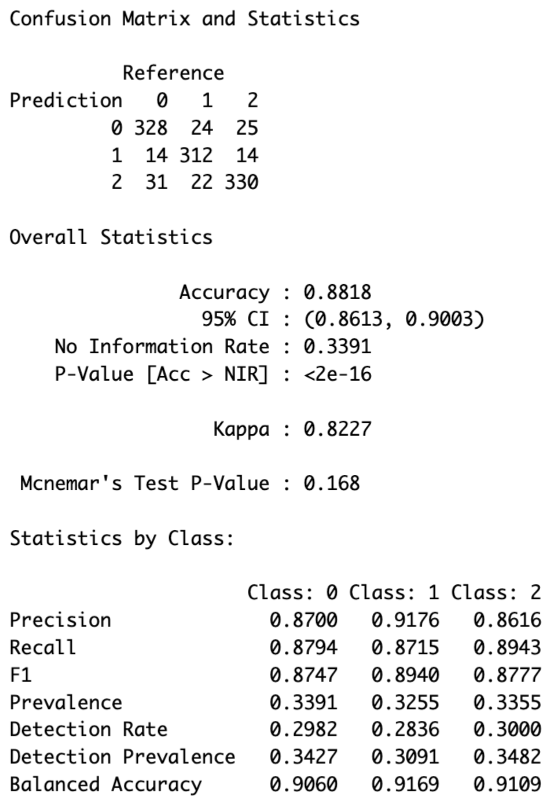


Figure 5. P-value Distribution

3.3 Model Selection, Training, & Testing

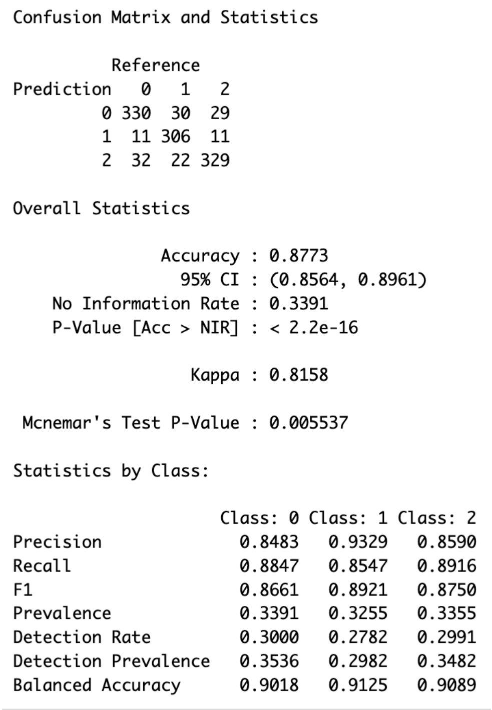
Model selection: K-Fold Cross-Validation, Logistic Regression (LR), Random Forest (RF)
Reasoning in detail is in **Appendix A**



```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 328  24  25
         1  14 312  14
         2  31  22 330

Overall Statistics

               Accuracy : 0.8818
                 95% CI : (0.8613, 0.9003)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8227

 Mcnemar's Test P-Value : 0.168

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.8700   0.9176   0.8616
Recall                 0.8794   0.8715   0.8943
F1                     0.8747   0.8940   0.8777
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.2982   0.2836   0.3000
Detection Prevalence   0.3427   0.3091   0.3482
Balanced Accuracy      0.9060   0.9169   0.9109
```

Figure 6. LR Result

Based on the LR result on figure 2, using all 20 features, the model could predict the student stress level with overall accuracy of **88.18%**.



```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 330  30  29
         1  11 306  11
         2  32  22 329

Overall Statistics

               Accuracy : 0.8773
                 95% CI : (0.8564, 0.8961)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8158

 Mcnemar's Test P-Value : 0.005537

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.8483   0.9329   0.8590
Recall                 0.8847   0.8547   0.8916
F1                     0.8661   0.8921   0.8750
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.3000   0.2782   0.2991
Detection Prevalence   0.3536   0.2982   0.3482
Balanced Accuracy      0.9018   0.9125   0.9089
```

Figure 7. RF Result

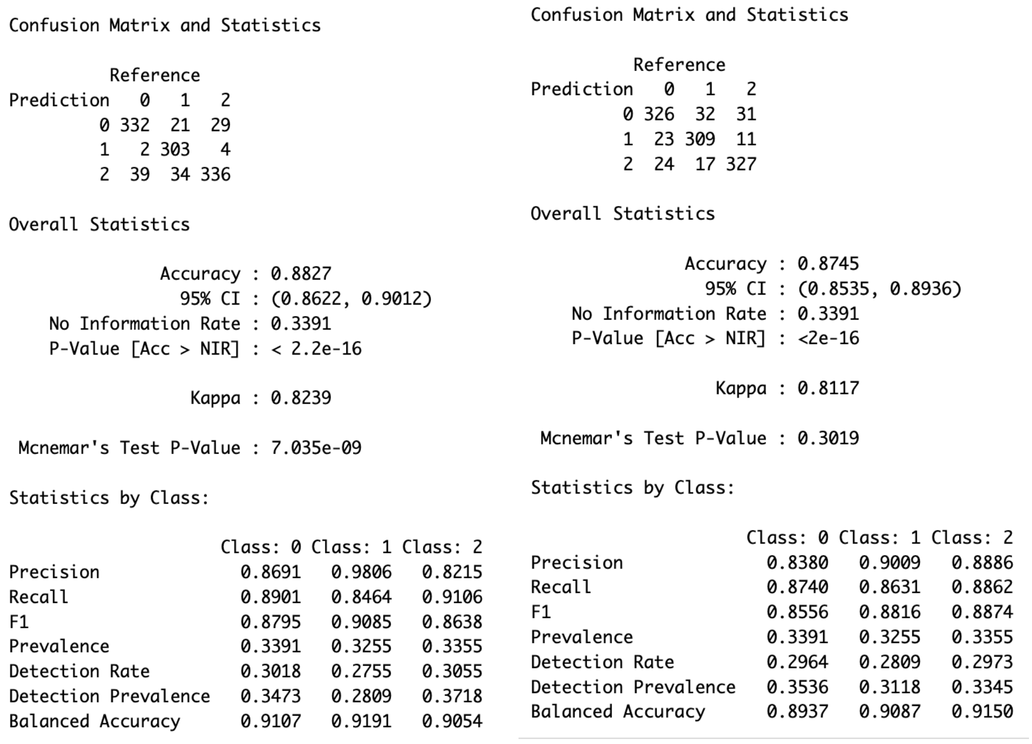
According to the RF result on Fig 6 using all 20 features, we can infer that the model could predict student stress level with overall accuracy of **87.73%**.

3.4 Testing the Models with Selected Features

In the EDA section, we used the ANOVA F-statistic to measure group variance. This metric helped us determine which features most significantly influence student stress level's prediction. We selected the top 10 features based on ANOVA results to test both models. The goal was to determine if feature selection improves overall accuracy.

These features include :

1. Blood Pressure
2. Self Esteem
3. Bullying
4. Sleep Quality
5. Future Career Concerns
6. Anxiety Level
7. Depression
8. Academic Performance
9. Safety
10. Basic Needs



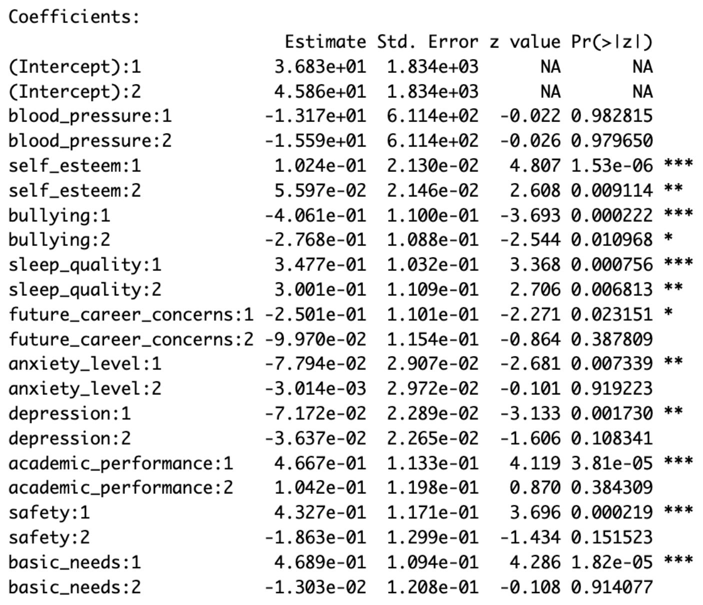
```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 332  21  29
         1   2 303   4
         2  39  34 336

Overall Statistics

               Accuracy : 0.8827
                 95% CI : (0.8622, 0.9012)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8239

 Mcnemar's Test P-Value : 7.035e-09

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.8691   0.9806   0.8215
Recall                 0.8901   0.8464   0.9106
F1                     0.8795   0.9085   0.8638
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.3018   0.2755   0.3055
Detection Prevalence   0.3473   0.2809   0.3718
Balanced Accuracy      0.9107   0.9191   0.9054
```

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 326  32  31
         1  23 309  11
         2  24  17 327

Overall Statistics

               Accuracy : 0.8745
                 95% CI : (0.8535, 0.8936)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8117

 Mcnemar's Test P-Value : 0.3019

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.8380   0.9009   0.8886
Recall                 0.8740   0.8631   0.8862
F1                     0.8556   0.8816   0.8874
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.2964   0.2809   0.2973
Detection Prevalence   0.3536   0.3118   0.3345
Balanced Accuracy      0.8937   0.9087   0.9150
```

Figure 8. LR & RF after Feature Selection

With the selected features, the LR model achieved an overall accuracy of **88.27%**, whereas the RF model reached **87.45%**.

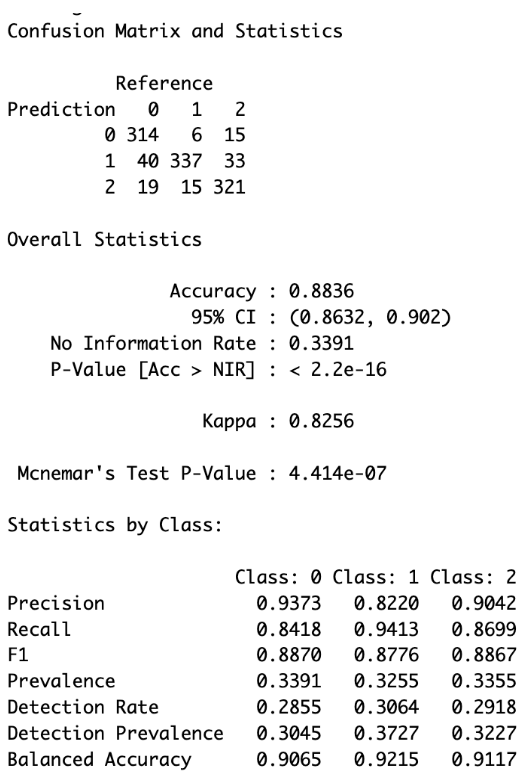
To help us narrow down to even more features, we employ LR and RF feature importance. The analysis is as follows :



```
Coefficients:
                           Estimate Std. Error z value Pr(>|z|)    
(Intercept):1             3.683e+01  1.834e+03      NA       NA    
(Intercept):2             4.586e+01  1.834e+03      NA       NA    
blood_pressure:1         -1.317e+01  6.114e+02  -0.022 0.982815    
blood_pressure:2         -1.559e+01  6.114e+02  -0.026 0.979650    
self_esteem:1             1.024e-01  2.130e-02   4.807 1.53e-06 ***
self_esteem:2             5.597e-02  2.146e-02   2.608 0.009114 ** 
bullying:1               -4.061e-01  1.100e-01  -3.693 0.000222 ***
bullying:2               -2.768e-01  1.088e-01  -2.544 0.010968 *  
sleep_quality:1           3.477e-01  1.032e-01   3.368 0.000756 ***
sleep_quality:2           3.001e-01  1.109e-01   2.706 0.006813 ** 
future_career_concerns:1 -2.501e-01  1.101e-01  -2.271 0.023151 *  
future_career_concerns:2 -9.970e-02  1.154e-01  -0.864 0.387809    
anxiety_level:1          -7.794e-02  2.907e-02  -2.681 0.007339 ** 
anxiety_level:2          -3.014e-03  2.972e-02  -0.101 0.919223    
depression:1             -7.172e-02  2.289e-02  -3.133 0.001730 ** 
depression:2             -3.637e-02  2.265e-02  -1.606 0.108341    
academic_performance:1    4.667e-01  1.133e-01   4.119 3.81e-05 ***
academic_performance:2    1.042e-01  1.198e-01   0.870 0.384309    
safety:1                  4.327e-01  1.171e-01   3.696 0.000219 ***
safety:2                 -1.863e-01  1.299e-01  -1.434 0.151523    
basic_needs:1             4.689e-01  1.094e-01   4.286 1.82e-05 ***
basic_needs:2            -1.303e-02  1.208e-01  -0.108 0.914077    
```

Figure 9. LR Feature Importance

Based on the results on Fig 8, we excluded the features with the largest p-values Blood Pressure and Future Career Concern (feature with no star and one star), and retrained the LR model. After the testing we achieved an overall accuracy of **88.36%**.



```
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 314   6  15
         1  40 337  33
         2  19  15 321

Overall Statistics

               Accuracy : 0.8836
                 95% CI : (0.8632, 0.902)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8256

 Mcnemar's Test P-Value : 4.414e-07

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.9373   0.8220   0.9042
Recall                 0.8418   0.9413   0.8699
F1                     0.8870   0.8776   0.8867
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.2855   0.3064   0.2918
Detection Prevalence   0.3045   0.3727   0.3227
Balanced Accuracy      0.9065   0.9215   0.9117
```

Figure 10. LR after feature selection 2

The next step is we use the feature importance of RF model, the result is as follows :

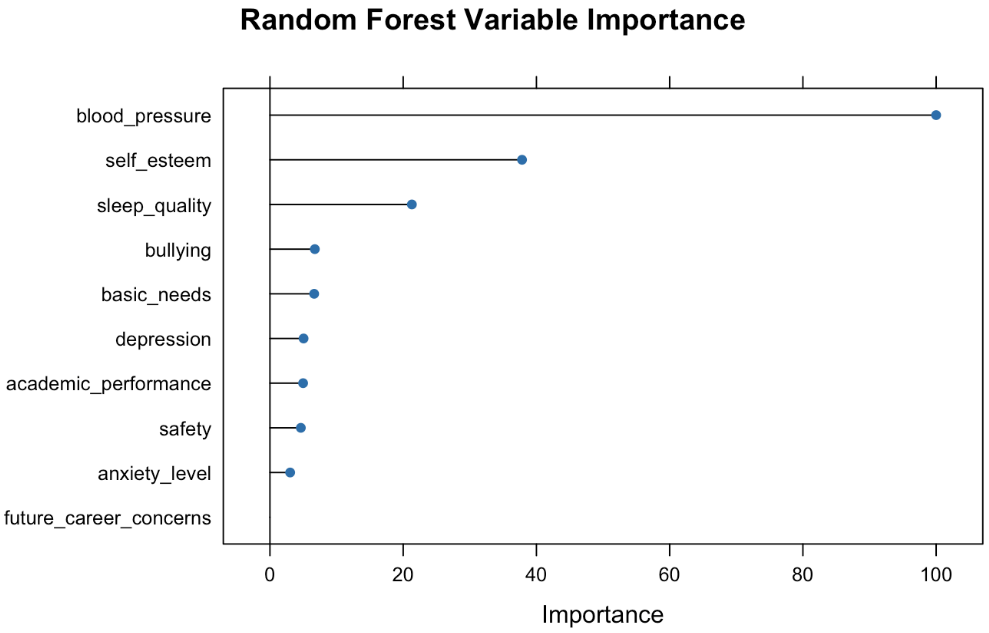
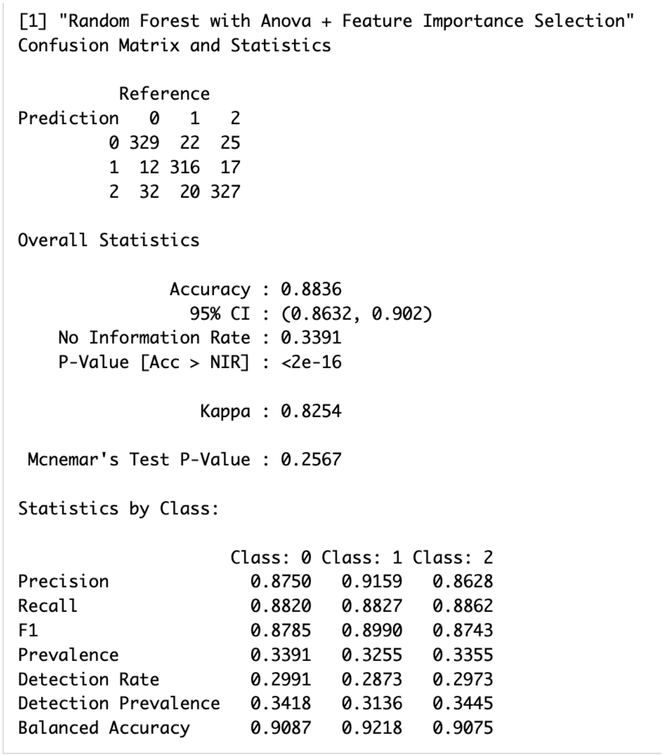


Figure 11. Random Forest Feature Importance

Based on the RF feature importance analysis, we excluded the three least significant features future career concerns, anxiety level, and safety and subsequently retrained the model.



```
[1] "Random Forest with Anova + Feature Importance Selection"
Confusion Matrix and Statistics

          Reference
Prediction   0   1   2
         0 329  22  25
         1  12 316  17
         2  32  20 327

Overall Statistics

               Accuracy : 0.8836
                 95% CI : (0.8632, 0.902)
    No Information Rate : 0.3391
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8254

 Mcnemar's Test P-Value : 0.2567

Statistics by Class:

                     Class: 0 Class: 1 Class: 2
Precision              0.8750   0.9159   0.8628
Recall                 0.8820   0.8827   0.8862
F1                     0.8785   0.8990   0.8743
Prevalence             0.3391   0.3255   0.3355
Detection Rate         0.2991   0.2873   0.2973
Detection Prevalence   0.3418   0.3136   0.3445
Balanced Accuracy      0.9087   0.9218   0.9075
```

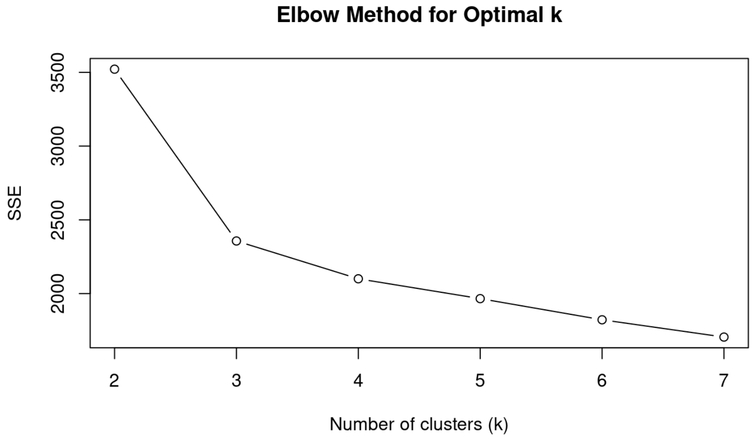
Figure 12. RF after feature selection 2

According to the testing result as can be seen in Figure 11, the RF model reached an overall accuracy of **88,36%.**

3.5 Clustering with Selected Features

The K-Means clustering method is used to help identify student lifestyle profiles with different stress levels. 5 lifestyle-related features were selected: 'sleep_quality','extracurricular_activities','noise_level', 'living_conditions', 'basic_needs' based on literature review.

The Elbow Method provides a pattern which the largest drop in SSE occurs between k = 2 and 3, after which the rate of improvement diminishes. It suggests that increasing the number of clusters beyond 3 reduces within-cluster variation. Therefore, we chose k = 3 as the most appropriate number of clusters.



The table presents the cluster centroids for the three-group K-Means solution. Each row represents a cluster, and each column shows the average value of that feature among all students in that cluster. These average values describe the lifestyle characteristics and stress levels that define each group.

	sleep_quality	extracurricular_activities	noise_level	living_conditions	basic_needs	stress_level
1	4.452888	1.553191	1.458967	3.534954	4.458967	0.04559271
2	1.090361	4.406627	3.984940	1.530120	1.584337	1.96084337
3	2.503417	2.437358	2.530752	2.503417	2.407745	0.97949886

4. Research questions

RQ1: **What proportion of students experience no stress at all, some stress, and high stress?**
Our EDA result indicates that roughly two-thirds of respondents report at least moderate stress, and about one-third fall into the high-stress category. The balanced class distribution is beneficial for later classification models because it reduces the risk that algorithms will be biased toward a majority class and allows performance metrics to reflect model quality rather than class imbalance.

RQ2: **How are different parts of student life—like sleep, workload, or social life—linked to stress levels?**
From section 3.2, the ANOVA results highlighted a hierarchy of feature importance based on their ability to separate stress categories, with the top 10 F-statistics including 30% psychological factors, 30% academic factors, 20% physiological factors, and 20% social factors. Moreover, since our p-values are extremely small, it proves that every single aspect of student life we measured shows a statistically significant change as stress increases. This indicates that different parts of student life are linked to stress universally and significantly.

RQ3: Can we predict the student stress level from the given factors?
According to section 3.3, the results indicate that the LR and RF models could predict the student stress level with high accuracy (88,18% and 87,73%) respectively. Moreover, when we examine F1 score, it shows balance across classes with a solid score around 0.87-0.89, indicating that the model achieves a strong balance between precision and sensitivity. This demonstrates that the classifier is robust and effective at distinguishing between stress levels.

RQ4: What types of student profiles can be identified based on their lifestyle factors and stress levels?
According to section 3.5, three student profiles can be concluded from the table:

(1) Balanced lifestyle students: Moderate scores across lifestyle measures and moderate stress levels

(2) Low-stress students: High sleep quality, perfect living conditions (low noise, well-met basic needs), and the lowest stress level)

(3) Overloaded students: Poor sleep, inadequate living conditions (suffered from noise, limited basic needs,) and the highest stress level)

RQ5: **If we only look at the most important causes of stress, does it make our predictions more accurate?**
As shown in the figures, LR's accuracy is improved after selecting only the top 10 features from the ANOVA test while the RF model experiences a decrease in accuracy. After excluding even more features with the LR & RF feature importance, the accuracy of both models improved, even though, as for LR, it also exhibits lower recall for classes 0 and 2, and as for the RF model, it further improved the recall for all classes 0, 1 and 2.

5. Project adjustments

The final analysis pipeline closely follows the proposed methods. Several improvements were made, including excluding KNN for RQ3 and DBSCAN for RQ4. Reasoning in detail is in **Appendix B**.
After the presentation, 3 feedback are chosen. Feedback detail is attached in **Appendix C**

All of those have been addressed directly in our analysis pipeline and analysis.

6. Insights & future recommendations for UIUC

Our findings show that physical factors and psychological factors are stronger drivers of stress than academic workload alone, implying that stress among students is shaped more by living conditions and personal health. Interventions from UIUC should therefore concentrate on strengthening living environments. Our study also suggests that with recent supply from UIUC, initiatives should be taken to spread awareness of basic-needs resources, which could help regulate stress. Additionally, to boost a healthy mentality, UIUC can expand approachable wellness programming and promote the mental health services from McKinley to more students.

7. Appendix

Appendix A

Model selection reasoning
Fold Cross-Validation: For this class project, 5-fold cross-validation is used to evaluate all models and prevents results from depending on a particular data split, providing a reliable estimate of how well each model generalizes.

Logistic Regression (LR): LR is an interpretable baseline model. The sample size is sufficient for estimating relationships between the target variable and 20 predictors, allowing the model to quantify each variable's influence on stress.

Random Forest (RF): RF was selected as it is a non-linear classifier, known to mitigate overfitting, handle mixed variable types, and produce feature-importance measures, which is extremely necessary for our RQ to identify the most influential stress-related factors.

K-Means: K-Means clustering is an unsupervised machine learning method. It can divide n points into k clusters when each point belongs to the cluster whose mean (i.e., the cluster center) is closest to it.

Appendix B

Method exclusion reasoning
KNN was excluded for RQ3 because it relies on a meaningful distance metric, which is hard to define with many categorical predictors. Additionally, KNN offers less interpretability compared to LR and RF. DBSCAN assumes that clusters form dense regions in continuous space; however, our predictors are mostly categorical and discrete, making it hard to define meaningful density and distance thresholds. In practice, this method also produces unstable and less interpretable results.
With the consideration that our stakeholders are management initiatives at the university, for this project scope, we preferred highly interpretable methods.

Appendix C

Lifestyle-related features selection reasoning

Sleep quality and physical activity are widely recognized as core lifestyle components that directly influence mental well-being (Sharma et al., 2024; Mammen & Faulkner, 2013; Benca et al., 2017). In addition, basic needs, living conditions, and environmental factors like noise can be incorporated into lifestyle and living-condition frameworks because they shape daily routines, stress exposure, and overall behavioral patterns (Serio et al., 2023; Velten et al., 2018). Therefore, we selected variables:

'sleep_quality', 'extracurricular_activities','noise_level', 'living_conditions', 'basic_needs' as lifestyle features in the analysis.

Appendix D

Feedback 1: Gawon Lim
Target: No Stress, Moderate, High Stress
It says distribution is pretty balanced out. Just to make sure that the data you got from kaggle is truly full survey result instead of already cleaned data.
I like how you guys used both supervised and unsupervised models. My suggestion would be regarding clustering. When we try to identify optimal number of k-means clusters, while silhouette score is one of the go-to method, if you use “elbow method” it would reassure your findings.

Feedback 2: Chantrice Santiago
It would be good to show comparative boxplots or grouped bar charts to show differences between groups for post-clustering results. You can also consider radar charts if applicable.
For ANOVA results (bar chart), it would be good to highlight the top n significant features (ex. maybe gray out the less significant features). Showing all features with all bars with nearly the same color makes it hard to focus on the main insight especially since the presentation of the visual was very quick.

Feedback 3: Anisha Kango
There was a clear explanation of ANOVA results and how different stress groups vary significantly, very easy to follow. The team can include one slide on practical usage (e.g.,sleep routines, academic counselling) to show how insights can translate into action.

8. References

1. Benca, R. M., Cirelli, C., & Tononi, G. (2017). Basic science of sleep. In *Kaplan & Sadock’s Comprehensive Textbook of Psychiatry* (10th ed., pp. 339–354). Wolters Kluwer.
2. Cohen, S. (2004). Social relationships and health. *American Psychologist, 59*(8), 676–684.
3. Mammen, G., & Faulkner, G. (2013). Physical activity and the prevention of depression: A systematic review of prospective studies. *American Journal of Preventive Medicine, 45*(5), 649–657.
4. Serio, F., De Donno, A., & Valacchi, G. (2023). Lifestyle, nutrition, and environmental factors influencing health benefits. *International Journal of Environmental Research and Public Health, 20*(7), 5323. https://doi.org/10.3390/ijerph20075323
5. Sharma, I., Marwale, A. V., Sidana, R., & Gupta, I. D. (2024). Lifestyle modification for mental health and well-being. *Indian Journal of Psychiatry, 66*(3), 219–234.
6. Umberson, D., & Montez, J. K. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior, 51*(Suppl.), S54–S66.
7. Velten, J., Bieda, A., Scholten, S., et al. (2018). Lifestyle choices and mental health: A longitudinal survey with German and Chinese students. *BMC Public Health, 18*, 632. https://doi.org/10.1186/s12889-018-5526-2