Analyzing Student Stress Factors through Machine Learning Techniques

Manan Patel and Zachary Perry

Dataset

- Dataset was sourced from Kaggle
  - *"Student Stress Factors: A Comprehensive Analysis"*
- 1,100 students, ages 15-24 (high school - college)
  - Dharan, Nepal
- 5 Major Factors
  - Psychological
  - Physiological
  - Social
  - Environmental
  - Academic
- Objectives
  1. *Can we predict the stress level of a new student based on the similarity of their stress factors to those of known students?*
  2. *Which factors contribute most significantly to predicting stress levels in students?*

	min	max
anxiety_level	0.0	21.0
self_esteem	0.0	30.0
mental_health_history	0.0	1.0
depression	0.0	27.0
headache	0.0	5.0
blood_pressure	1.0	3.0
sleep_quality	0.0	5.0
breathing_problem	0.0	5.0
noise_level	0.0	5.0
living_conditions	0.0	5.0
safety	0.0	5.0
basic_needs	0.0	5.0
academic_performance	0.0	5.0
study_load	0.0	5.0
teacher_student_relationship	0.0	5.0
future_career_concerns	0.0	5.0
social_support	0.0	3.0
peer_pressure	0.0	5.0
extracurricular_activities	0.0	5.0
bullying	0.0	5.0
stress_level	0.0	2.0

Methodology

Research Question: *Can we predict the stress level of a new student based on the similarity of their stress factors to those of known students?*

- **K-Nearest Neighbors**
  - Simple and explainable method
  - Evaluated based on accuracy score
  - **Hyperparameter Tuned**: K-fold cross validation (n_neighbors)
    - 10 splits
    - Tested all odd numbers, 1-30, found 27 neighbors was best

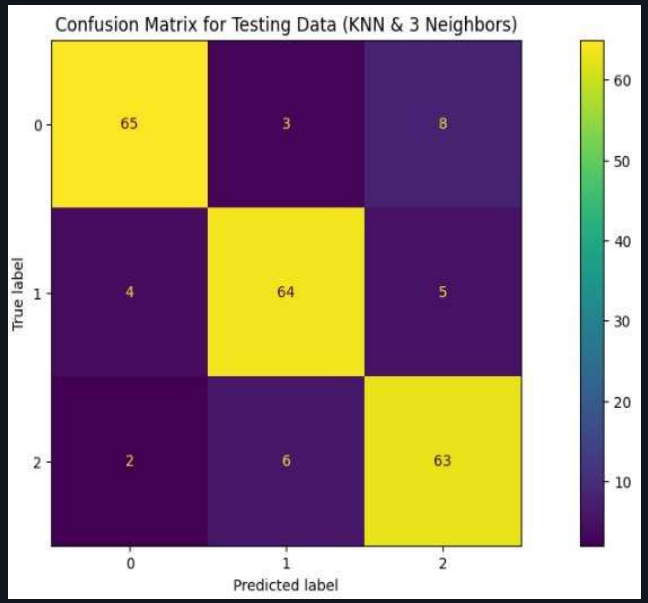
- **Multi-Layer Perceptron**
  - More complex, good for classification tasks
  - Evaluated based on accuracy score
  - max_iter=1000
  - **Hyperparameter Tuned**: K-fold cross validation (hidden_layer_sizes, learning_rate_init)
    - 3 splits
    - Testing multiple combinations of *hidden_layer_sizes* and *learning_rate_init* values
    - Tested to find best *network structure* for *hidden_layer_sizes*
    - Found learning_rate_init = 0.01 and hidden_layer_sizes = 500 to perform best
- Also tested **undersampling** for both

Methodology (cont.)

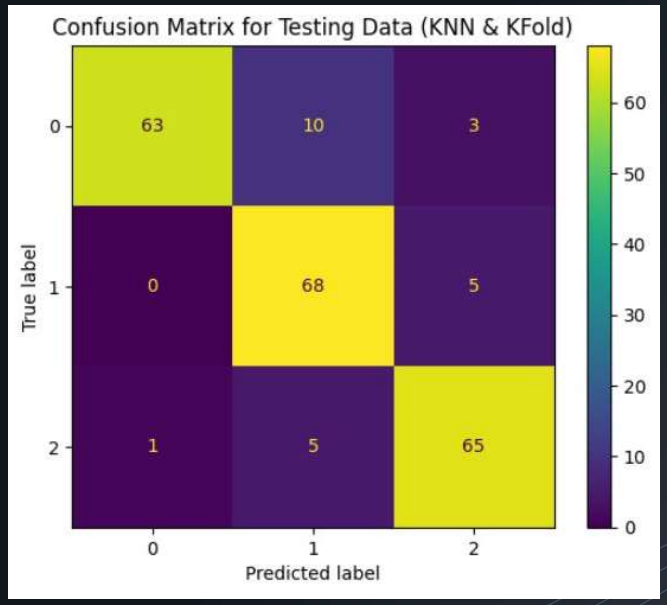
Research Question: *Which factors contribute most significantly to predicting stress levels in students?*

- **Decision Trees**
  - Interpretability and ranking of feature importance
  - Evaluated based on accuracy score
  - **Hyperparameter Tuned**: Max Depth of Tree (max_depth)
    - Tested several values to find optimal depth of tree
- **Random Forest**
  - Robust ensemble technique that leverages multiple decision trees
  - Evaluated based on accuracy score
  - **Hyperparameter Tuned**: Number of trees (n_estimators)
    - Tested various n_estimators values to find optimal number of trees in the forest

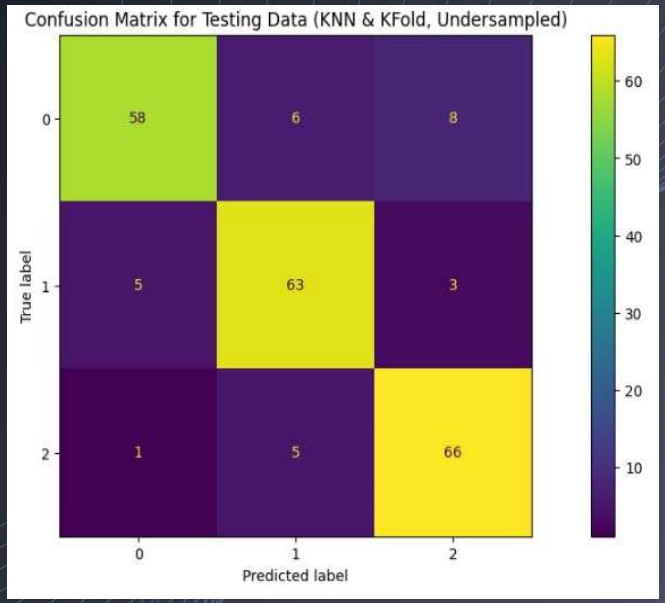
KNN Results



- Testing Accuracy: 87.2%
- *Neighbors*: 3

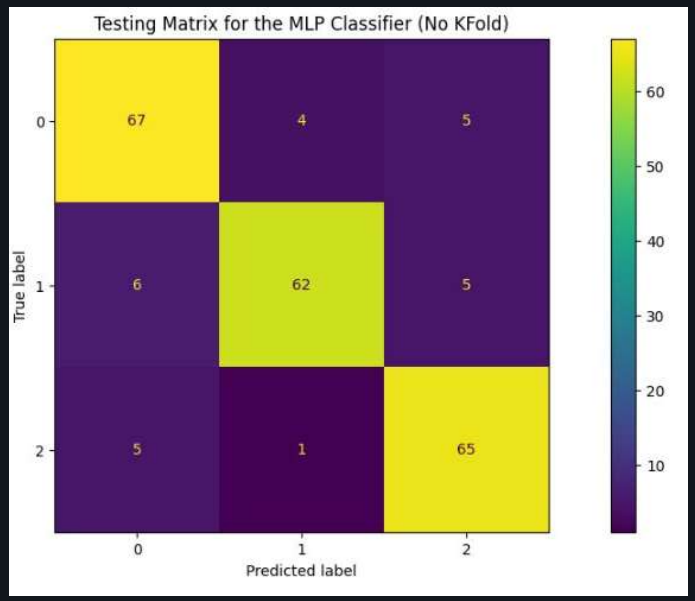


- Testing Accuracy: 89.1%
- *Neighbors*: 27

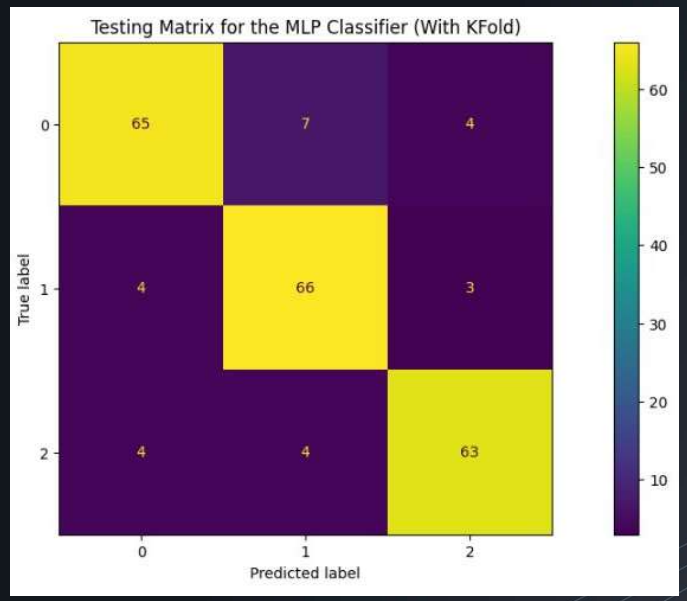


- Testing Accuracy: 86.98%
- *Neighbors*: 1

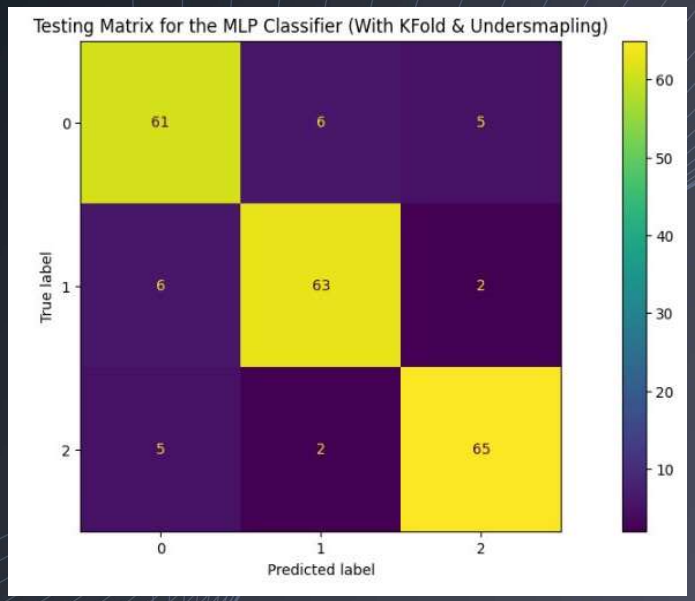
MLP Results



- Testing Accuracy: 88%
- *hidden_layer_sizes*: 50
- *learning_rate_init*: 0.0001



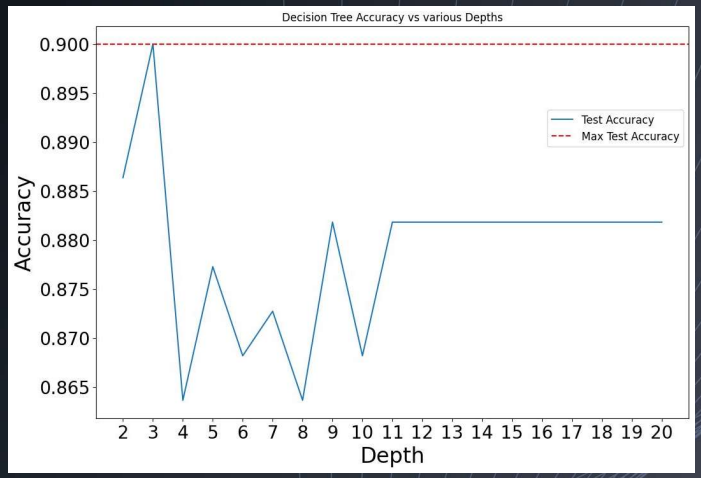
- Testing Accuracy: 88%
- *hidden_layer_sizes*: 500
- *learning_rate_init*: 0.01



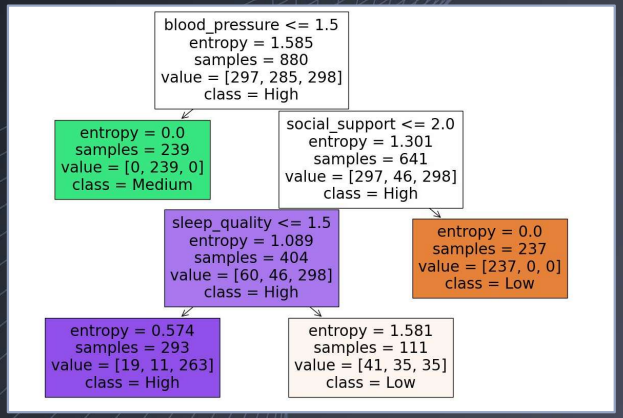
- Testing Accuracy: 87.9%
- *hidden_layer_sizes*: (100,100)
- *learning_rate_init*: 0.1

Decision Trees Results

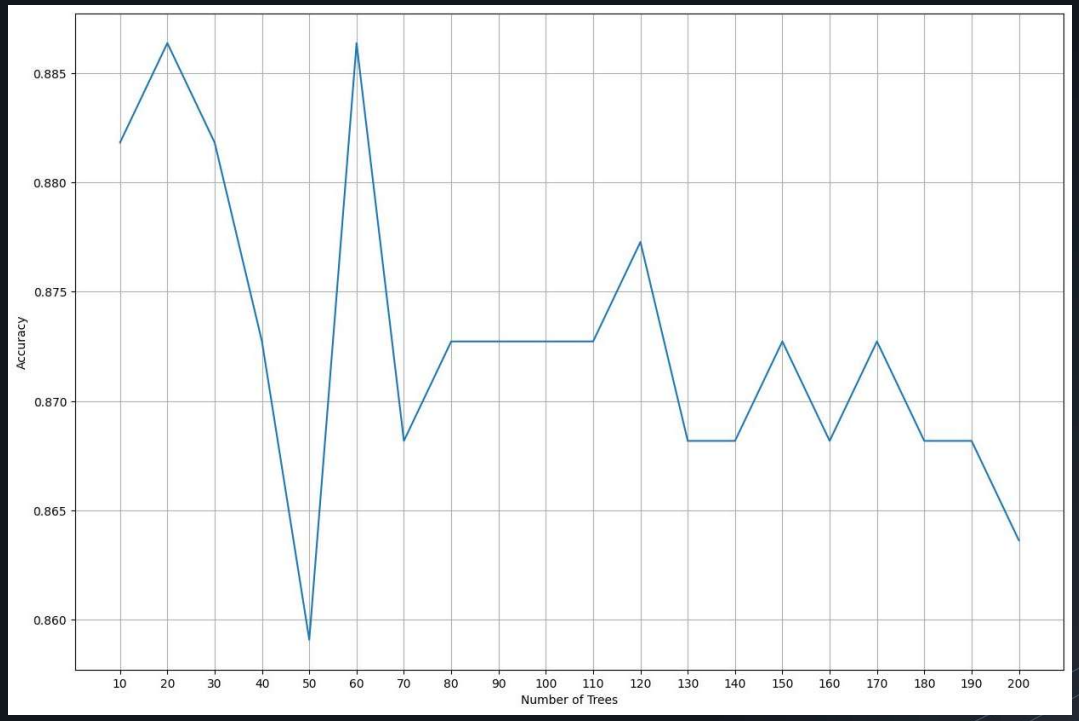
Rank	Feature	Importance
1	blood_pressure	40.21%
2	social_support	29.07%
3	sleep_quality	9.81%
4	bullying	4.37%
5	breathing_problem	4.03%
6	academic_performance	2.12%
7	depression	2.09%
8	extracurricular_activities	1.97%
9	teacher_student_relationship	1.51%
10	peer_pressure	0.88%
11	self_esteem	0.84%
12	noise_level	0.81%
13	study_load	0.70%
14	anxiety_level	0.70%
15	safety	0.34%
16	future_career_concerns	0.23%
17	living_conditions	0.20%
18	basic_needs	0.14%
19	headache	0.00%
20	mental_health_history	0.00%



Testing Accuracy: 90%
Max-Depth: 3



Random Forest results



Rank	Feature	Importance
1	sleep_quality	14.46%
2	blood_pressure	12.17%
3	safety	8.62%
4	basic_needs	8.11%
5	self_esteem	6.96%
6	anxiety_level	6.51%
7	teacher_student_relationship	6.40%
8	extracurricular_activities	6.17%
9	bullying	5.11%
10	headache	4.68%
11	noise_level	3.71%
12	social_support	3.59%
13	depression	3.53%
14	academic_performance	3.27%
15	study_load	1.43%
16	peer_pressure	1.40%
17	future_career_concerns	1.32%
18	living_conditions	1.15%
19	breathing_problem	0.79%
20	mental_health_history	0.61%

Testing accuracy: 89%
Number of Trees (n_estimators): 20

Discussion

- **KNN & MLP**
  - Both are capable of predicting the stress level of a student based on the characteristics and factors of other students
  - KNN is overall the better model

- **Decision Trees & Random Forest**
  - Physiological factors such as Blood pressure and sleep quality are most impactful along with social support
  - Decision Trees showcased better accuracy compared to Random Forest

Future Work

More time?

- Extensive exploration of hyperparameter tuning for all models

Additional Research?

- Are there certain combinations of features from different domains (Psychological, Physiological, Social, Environmental, Academic) that collectively have more influence on predicting stress levels?

Additional work would have done?

- Explore variations of ensemble approaches for models
  - Bagging or Boosting for Decision Trees and Random Forests
  - Different distance metric for KNN
  - Varied architectures for MLP

Team Contributions

Zachary Perry

· Focused on predicting stress levels of new students based on similarity to known students.
· Conducted in-depth research and testing of KNN and MLP models and hyperparameter tuning of each model for optimal performance
· Created appropriate visualizations

Manan Patel

· Worked on answering the factors that contributed significantly to predicting stress levels in students
· Explored Decision Trees and Random Forest models, conducting thorough testing and hyperparameter tuning for each
· Created appropriate visualizations

Both

· Collaboratively discovered and selected the dataset for the project
· Formulated the research questions jointly
· Shared responsibilities in creating a presentation and co-authored the paper

References

[1] “Student Stress Factors: A Comprehensive Analysis,” www.kaggle.com. Https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis

[2] R. Spitzer, “GAD-7 Anxiety,” 1999. Available: https://adaa.org/sites/default/files/GAD-7_Anxiety-updated_0.pdf

[3] M. Rosenberg, "ROSENBERG SELF-ESTEEM SCALE," fetzer.org, 1965. Available: https://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures\/Self_Measures_for_Self-Esteem_ROSENBERG_SELF-ESTEEM.pdf

[4] “PATIENT HEALTH QUESTIONNAIRE (PHQ-9),” Oct. 2005. Available: https://med.stanford.edu/fastlab/research/imapp/msrs/_jcr_content/main/accordion/accordion_content3/download_256324296/file.res/PHQ9%20id%20date%2008.03.pdf

[5] American Psychological Association, "APA Dictionary of Psychology," dictionary.apa.org, 2020. https://dictionary.apa.org/social-support

[6] C. Schuman. NearestNeighbors [PowerPoint slides]. University of Tennessee, Knoxville, 2023. Available: https://utk.instructure.com/courses/179872/files/17767847?module_item_id=3771387

[7] C. Schuman. TeamProject_NeuralNetworks_Part2 [PowerPoint slides]. University of Tennessee, Knoxville, 2023. Available:https://utk.instructure.com/courses/179872/files/18465674?module_item_id=3820936