

## Machine Learning Analysis of Stress Factors in College Students: Predictive Modeling and Feature Importance

Analysis of stress among college students aged 18–21 uses a dataset from a nationwide survey, capturing 20 features across psychological, physiological, environmental, academic, and social categories. Data, collected via anonymized Google Forms on a five-point Likert scale, includes variables like anxiety, sleep quality, and academic performance. Machine learning models such as Stacking Classifier, Voting Classifier, Gradient Boosting, and MultiLayer Perceptron evaluate stress levels (Eustress, Distress, No Stress). Feature selection through permutation importance and mutual information identifies blood pressure, sleep quality, and self-esteem as key predictors. The MultiLayer Perceptron achieves the highest accuracy at 91.51%, followed by the Stacking Classifier at 91.21%. Cross-validation with RepeatedKFold ensures robust evaluation. Findings reveal complex interactions between emotional, physical, and academic stressors, offering insights for targeted interventions to improve student well-being. The study emphasizes the critical role of physiological and psychological factors in stress prediction, guiding the development of data-driven support strategies.

**Keywords:** Stress; Machine learning; Feature selection; Student well-being; Predictive modeling