

Interpretation & Discussion

The results of the Logistic Regression model show that it can be an effective model in classifying students according to their stress levels based on various lifestyle factors presented in this dataset. The resultant confusion matrix showed that this model was best in predicting students for a moderate level of stress while committing fewer errors than in the low and high categories. This could mean that the features provided are more indicative of average stress conditions, whereas extreme stress could require additional or more specific indicators for improvement in classification accuracy.

The 5-Fold Cross-Validation results further demonstrated that the model maintained consistent performance across different subsets of data. The accuracy values across all five folds showed only slight variations, indicating that the model generalizes well and is not overly dependent on a single training sample. This stability suggests good reliability and robustness of the Logistic Regression model when applied to similar data.

The learning curve analysis gave further insight into the training behavior of the model. With a smaller training dataset, the model had high training accuracy and slightly lower validation accuracy, indicating slight overfitting. As the training size increases, the curves gradually converge, which means the model benefited from having more data and got a better generalization. This pattern confirms that the Logistic Regression model performs more effectively while it is exposed to larger datasets, although the improvement plateaus beyond a certain point.

The Logistic Regression model was a good baseline classifier, which had decent accuracy and was very consistent in its validation results. Yet, there is still room for improvement: better feature selection or transformation may help to capture more complex relationships between lifestyle habits and stress levels; tuning hyperparameters, such as regularization strength (C) or solver type, may refine the performance of the model. Besides, other models, like Decision Tree, KNN, or SVM, can be more appropriate with regard to nonlinear patterns in data. Yet, Logistic Regression, being simple, interpretable, and reliable, is a very good starting point for stress level prediction, serving as a solid foundation for further model optimization and comparison.