Week 2: Literature Review and Dataset Selection

We did a literature review during Week 2 to determine the background and relevance of predicting student academic performance with statistics and prediction methods. We made sure to get data that is related to what we want to study.

Literature Review

1. In this study, Kotsiantis, S. B., Pierrakeas, C., & Pintelas, P. (2004). Using machine learning to guess how students will do in a distance learning setting. Applied Artificial Intelligence, 18(5) pages 411-426.

Several machine learning algorithms, like decision trees, Naive Bayes, and k-nearest neighbors, were compared to predict how students would perform. It points out that academic and behavioral information can help predict performance.

2. Adekitan, A. I. and Salau, O. (2019). How the first three years' grades of engineering students affect their final CGPA, as seen through predictive analytics. Heliyon, (Vol. 5, Issue 2), e01250.

This document looks at the connection between initial academic outcomes and the final results, using regression models to help predict what happens. This method pays attention to ongoing examination of assessment data.

3. Al-Barrak, M. A. and Al-Razgan, M. (2016). Predicting how students will perform in the end using decision trees: A study done here. Journal of Information and Education Technology, 6(7): 528–533.

The authors use decision trees on student records to see if GPA can be predicted, and the findings highlight that attendance and past academic history matter.

Dataset Selection

We chose the 'Student Performance Dataset' from Sling Academy, which shows the marks for various subjects and personal data of many students. Because this dataset supports ANOVA, regression, and classification, it is ideal for the project to analyze and discover elements affecting academic performance.