Cheyanne Gardner

MGT 665

April 12, 2025

Lab 2: Classification Problem

GitHub repository Link:

https://github.com/CGardner1998/Lab-2-Classification-Problem

Introduction:

For this project, I chose a dataset focused on student depression from Kaggle, which has information about academic pressure, sleep habits, financial stress, and other personal and lifestyle factors.

For lab 2 I need to see if a student is experiencing depression using three popular classification algorithms: Logistic Regression, k-Nearest Neighbors (k-NN), and Decision Trees. Each model was measured using accuracy, precision, recall, and F1-score to compare their performance. This analysis not only highlights how different models perform on real-world data but also explores how machine learning can be applied to better understand mental health challenges among students.

Comparison:

Looking at the evaluation metrics, Logistic Regression performed really well overall, with high accuracy and precision. It's a great choice when you need something that's easy to understand and gives balanced results. It works best when the data has clear, linear patterns, but it might not do as well with more complex relationships.

The k-Nearest Neighbors model had okay results across the board. It's simple to use and doesn't assume much about the data, but it can be sensitive to how features are scaled and doesn't handle large datasets efficiently. Its lower recall and F1-score mean it might miss some true positives. On the other hand, the Decision Tree model had high recall and F1-score, so it's good at catching actual positive cases, something that's especially useful for things like mental health screening. But it can overfit and have slightly lower precision, which means more false alarms.

Overall, Logistic Regression is a solid, well-rounded option, Decision Trees are great when catching positives matters most, and k-NN works best with smaller, cleaned-up datasets.