Executive Summary: Student Mental Health K-Means Clustering Analysis

Submitted by: Abdullahi Mohamed Jibril
Course: Applied Machine Learning for Analytics
Date: August 19, 2025

Project Overview

This project successfully implemented a K-Means clustering algorithm to analyze patterns in student mental health data, identifying distinct groups of students based on their mental health characteristics, academic performance, and demographic information. The application was developed using Python in Google Colab environment, utilizing advanced machine learning techniques to uncover meaningful insights about student wellbeing.

Dataset and Methodology

The analysis utilized a comprehensive student mental health dataset containing 101 student records with 11 key features including age, gender, academic year, CGPA, marital status, and critical mental health indicators (depression, anxiety, panic attacks, and treatment-seeking behavior). The methodology involved:

- Data Preprocessing: Handled mixed datetime formats, addressed missing values using median imputation, and encoded categorical variables using Label Encoding

- Feature Engineering: Extracted temporal features (hour and day of week) from timestamp data

- Standardization: Applied StandardScaler to normalize features for optimal clustering performance

- Cluster Optimization: Used both Elbow Method and Silhouette Score analysis to determine the optimal number of clusters (4)

Key Findings and Cluster Profiles

The K-Means algorithm successfully identified four distinct student mental health profiles:

1. Cluster 0 (27 students, 27%): High-anxiety students with low treatment-seeking behavior

2. Cluster 1 (22 students, 22%): Well-adjusted students with minimal mental health concerns

3. Cluster 2 (43 students, 43%): Students experiencing depression but actively seeking help

4. Cluster 3 (9 students, 9%): At-risk students with panic attacks but no treatment engagement

Challenges and Solutions

Technical Challenges:

- Mixed datetime formats: Resolved by implementing a flexible datetime parsing function

- Categorical data encoding: Successfully handled using LabelEncoder for consistent numerical representation

- Feature scaling: Addressed through StandardScaler to ensure equal feature contribution

Data Quality Issues:

- Missing age values were imputed using median values to maintain data integrity

- Pandas future warnings were addressed through proper DataFrame manipulation techniques

Actionable Insights and Recommendations

The clustering analysis reveals critical intervention opportunities:

1. Immediate Attention Required: Cluster 3 students (at-risk group) need targeted counseling and crisis intervention protocols

2. Preventive Measures: Cluster 0 students would benefit from anxiety management workshops and stress reduction programs

3. Support Enhancement: Cluster 2 students, while seeking help, may need improved access to mental health resources

4. Maintenance Programs: Cluster 1 students require wellness maintenance and prevention strategies

Technical Implementation and Quality Assurance

The application demonstrated robust technical implementation with:

- Clean, well-documented code following professional Python standards

- Comprehensive unit testing with 100% test success rate covering data integrity, cluster assignment, and feature scaling

- Effective visualization using matplotlib and seaborn for cluster analysis and pattern identification

- Proper error handling and data validation throughout the pipeline

Business Value and Impact

This clustering analysis provides educational institutions with data-driven insights to:

- Allocate mental health resources more effectively based on student risk profiles

- Implement targeted intervention programs for specific student groups

- Develop predictive models for early identification of at-risk students

- Enhance student support services through personalized approaches

Conclusion

The K-Means clustering application successfully segmented the student population into meaningful mental health profiles, providing actionable insights for educational stakeholders. The robust technical implementation, combined with comprehensive testing and clear visualization, demonstrates the practical value of machine learning in addressing student mental health challenges. The identified clusters offer a foundation for developing targeted intervention strategies that can significantly improve student wellbeing and academic success.