Mental Health Risk Prediction using Machine Learning

Author: Sumeet Yadav
Program: B.Sc. Computer Science
Semester: IV
Year: 2026

Abstract
Mental health has become an important concern among students due to academic pressure and lifestyle changes. This study explores the use of machine learning techniques to predict mental health risk based on behavioral and lifestyle factors. A synthetic dataset was generated containing attributes such as sleep duration, screen time, study hours, physical activity, and stress level. Logistic regression was used as a baseline classification model and compared with a decision tree classifier. The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrix. The proposed system is intended for decision support and does not perform clinical diagnosis.

1. Introduction
Mental health plays a significant role in an individual's overall well-being and academic performance. Students often experience increased stress due to academic workload, prolonged screen usage, reduced physical activity, and irregular sleep patterns. Machine learning techniques can help identify patterns in behavioral data and support early risk detection.

2. Problem Statement
Traditional mental health assessment methods are manual and do not scale efficiently. There is a need for an automated data-driven approach to classify individuals into mental health risk categories.

3. Objectives
- Analyze lifestyle factors related to mental health risk
- Develop a machine learning-based prediction system
- Compare logistic regression and decision tree models
- Evaluate model performance using multiple metrics

4. Hypotheses
H0: Lifestyle factors do not significantly contribute to mental health risk prediction.
H1: Lifestyle factors significantly contribute to mental health risk prediction.

5. Literature Review

6. Dataset Description

A synthetic dataset with 150 records was generated. Features include sleep hours, screen time, study hours, physical activity, and stress level. Synthetic data ensures ethical compliance.

7. Methodology

The dataset was split into training and testing sets. Logistic regression was implemented as the baseline model and compared with a decision tree classifier. Evaluation metrics included accuracy, precision, recall, F1-score, and confusion matrix.

8. Results

9. Conclusion and Future Scope