

Clustering and Classification of Lifestyle Habits for Identifying Stress Levels

Rykelle Wei J. Ruelan

Department of Computer, Information Sciences and Mathematics, University of San Carlos, Cebu City, Philippines

James Matthew Garcia

Department of Computer, Information Sciences and Mathematics, University of San Carlos, Cebu City, Philippines

Abstract

Stress is a growing concern in modern society, often influenced by daily lifestyle habits. This study identifies and categorizes individuals' stress levels based on behavioral factors using a machine learning approach applied to a simulated Mental Health and Lifestyle dataset containing approximately 11,000 records. The methodology involves two main steps: (1) applying K-Means clustering to discover lifestyle behavior groups using seven lifestyle habits, specifically sleep hours, work hours, physical activity hours, social media usage, diet quality, smoking habits, and alcohol consumption; and (2) using Multinomial Logistic Regression to classify individuals into Low, Medium, or High stress levels. The study also examined whether incorporating demographic variables such as age, gender, occupation, and country improved classification performance. Three model configurations were evaluated, with the baseline lifestyle-only model achieving 93.42% accuracy and a macro AUC-ROC of 0.9915, demonstrating that lifestyle habits alone are sufficient for accurate stress level classification. Adding cluster labels and demographic variables produced negligible improvement, confirming the sufficiency of lifestyle habits as predictors. A web-based application was also developed to allow users to input their lifestyle data and receive personalized stress predictions and actionable recommendations.

Keywords

Stress Level Classification, Lifestyle Habits, K- Means Clustering, Multinomial Logistic Regression, Machine Learning.