JOURNAL INFORMATIC, EDUCATION AND MANAGEMENT (JIEM)
Vol. 7, No. 2, August 2025, pp. 258 ~ 275
ISSN: 2716-0696, DOI: 10.61992/jiem.v7i2.142

Machine Learning Models for Predicting Mental Health Indicators Using Digital Physical Activity Data: A Systematic Literature Review

Dirga Febrian[1]*, Rolly Maulana Awangga[1]
[1] Universitas Logistik dan Bisnis Internasional

Article Info

Article history:
Received 24 July 2025
Revised 27 July 2025
Accepted 30 July 2025

Keywords:
Machine Learning, Mental Health, Digital Physical Activity, Feature Selection

ABSTRACT

This systematic literature analysis examines 40 studies (2020–2025) on the use of machine learning to predict mental health using digital activity data. Two research questions are presented: algorithm performance comparison and model effectiveness factor. Data surveys (43,9%) are a more widely used data collection method. Because of its interpretability, Logistic Regression is the most popular (29.3%), whereas Random Forest (26.8%) is best for performance-interpretability. With a rata-rata accuracy of 80.1% ± 4.2% and an AUC of 87.1% ± 1.8%, XGBoost provides superior performance. The best study achieves an AUC >0,98 through feature engineering that canggih using SHAP and recursive feature elimination. Critical success factors include cermat fitur selection, temporal dinamika, cross-validation, and clinical interpretability. Although machine learning has significant potential, there are still challenges with standardization, generalizability, and real-world implementation. Research in the long term requires longitudinal studies, external validation, and standard protocols to realize this technology's potential in improving mental health outcomes.

Corresponding Author:
Dirga Febrian | Universitas Logistik dan Bisnis Internasional
Email: dirgafebrian@ulbi.ac.id