

Early Detection of Stunting in Indonesian Toddlers: A Machine Learning Approach

Herjanto Janawisuta¹, Putu Harry Gunawan², Indwiarti³

^{1,2}School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: ¹herjanto@student.telkomuniversity.ac.id, ²phgunawan@telkomuniversity.ac.id,
³Indwiarti@telkomuniversity.ac.id

Abstract

Stunting, or growth retardation, is a condition in which a child's physical growth is hindered due to prolonged inadequate nutrition, often accompanied by illness. Between 2000 and 2020, the global prevalence of stunting decreased from approximately 30.3% to 22%. Despite a significant reduction, in 2019, an estimated 21.3% of children under five were still projected to experience stunting worldwide. One contributing factor to stunting in young children is the need for more parental supervision and attention. Therefore, it is crucial to conduct research to predict the likelihood of stunting and, ideally, prevent it in toddlers. The use of accurate prediction models can aid in identifying at-risk toddlers, enabling timely interventions. Random Forest is commonly chosen for classification tasks due to its widespread popularity and robust performance. On the other hand, Logistic Regression is often preferred as a straightforward classification algorithm, being the simplest among its counterparts. Its appeal lies in its ability to achieve high accuracy without extensive hyperparameter tuning efforts. This study utilized a stunting dataset from the Bojongsoang Health Center, comprising 6677 entries collected in August 2022. The research shows that the Random Forest model consistently outperforms Logistic Regression, even in the presence of class imbalance. While Logistic Regression demonstrates strong performance, the Random Forest model stands out with near-perfect accuracy, particularly in oversampling scenarios where Logistic Regression is 89.4%, and Random Forest is an impressive 99.8%. This underscores the robustness of Random Forest in situations where achieving an imbalanced representation of classes is feasible.

Keywords: Stunting, Machine Learning, Random Forest, Logistic regression
