Classifying Stunting Status In Toddlers Using K-Nearest Neighbor And Logistic Regression Analysis

Alvin Tolopan Armando Sibuea[1], Putu Harry Gunawan[2]

[1,2]School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: [1]alvintoasibuea@student.telkomuniversity.ac.id, [2]phgunawan@telkomuniversity.ac.id

Abstract

Stunting is a critical child growth disorder, characterized by a height below the norm for one's age group. Despite a notable decrease in stunting prevalence in Indonesia from 37% in 2014 to 21.6% in 2022, achieving the targeted reduction to 14% by 2024 remains imperative. This study contributes to this national health goal by developing a robust predictive model for stunting in toddlers using machine learning. The research employs two models, K-Nearest Neighbors (KNN) and Logistic Regression. The dataset used for this research shows a big gap of data imbalance, for the majority of its data is significantly higher than the minority. Both of the chosen method is focusing on mitigating data imbalance through oversampling and undersampling techniques. The KNN model is particularly suited for this study due to its effectiveness in handling the complex, non-linear patterns often found in multifaceted health data like stunting indicators. It consistently demonstrated high accuracy, averaging 0.980, and reaching 0.987 for F1-Score. Logistic Regression, chosen for its ability to provide clear interpretability, especially useful in understanding the impact of various health indicators, also performed well with an average accuracy of 0.877 and an F1-Score of 0.894. This study highlights the significance of machine learning in addressing child stunting, providing effective tools for prediction. The combination of KNN's ability to handle complex data and Logistic Regression's interpretability, along with data balancing, contributes to the goal of reducing stunting prevalence. In summary, this research tackles child stunting, a pressing issue in Indonesia. By leveraging machine learning techniques, it develops predictive models to aid in stunting prevention. KNN excels in capturing complex patterns, while Logistic Regression offers interpretability. These models offer promise in reaching the vital goal of reducing stunting to 14% by 2024.

Keywords: Machine Learning, Stunting, K-Nearest Neighbor, Logistic Regression