39. A hybrid ensemble boosting model for enhanced blood donor retention

Nahashon Kiarie, [1,*] Amos Chege[1] and Mary Mwadulo[1]

[1]School of Computing and Informatics, Meru University of Science and Technology

*Corresponding author email: ercmutwiri@gmail.com

Subtheme: Computing and Informatics - Leveraging Computing and informatics Technologies for Climate adaptation and resilience

Abstract

Emergency situations such as accidents, medical operations and diseases necessitate regular blood transfusion. In Kenya, seven people require a blood transfusion every 10 minutes yet the country suffers a shortage of blood supply. One of the reasons for the acute shortage is lack of proper donor retention strategies. Existing predictive models for blood donor retention are often based on single algorithms, which suffer from their inherent weaknesses and most of them have low predictive accuracies. The Light GBM algorithm employs leaf-wise growth strategy, excels in loss reduction and hence improves accuracy. However, this may lead to potential overfitting, on the other hand, the XGBoost algorithm employs a level-wise growth strategy, which is computationally intensive but incorporates a robust mechanism for combating overfitting, such as the regularization parameter, column sampling, and weight reduction on new trees. This study aims to develop a hybrid ensemble gradient boosting model based on XGboost and Light GBM. The hybrid ensemble model aims to leverage on the strengths of both algorithms to enhance robustness, stability, and generalization while mitigating their individual biases and reducing overfitting hence leading to more accurate and consistent predictions. The base models were trained in parallel with data obtained from blood banks in Kenya. A weighted ensemble model was created by assigning weights to the respective prediction results of each model, the ensemble model was then evaluated and the accuracy compared with the accuracy achieved by the individual algorithms. The ensemble model achieved a performance accuracy of 89.6% better than XGBoost (88.3%) and the Light GBM (87.8). This study will enable blood agencies to accurately predict blood donor retention, reduce constant donor recruitment efforts and hence save time and costs. Additionally, it will provide insights for targeted retention strategies, ensuring a steady blood supply, ultimately saving lives.

Key words: *Blood transfusion, Predictive Models, Blood bank.*