

Predicting Wealth Status of Households Using Ensemble Machine Learning Algorithms

Authors : Habtamu Ayenew Asegie

Abstract : Wealth, as opposed to income or consumption, implies a more stable and permanent status. Due to natural and human-made difficulties, households' economies will be diminished, and their well-being will fall into trouble. Hence, governments and humanitarian agencies offer considerable resources for poverty and malnutrition reduction efforts. One key factor in the effectiveness of such efforts is the accuracy with which low-income or poor populations can be identified. As a result, this study aims to predict a household's wealth status using ensemble Machine learning (ML) algorithms. In this study, design science research methodology (DSRM) is employed, and four ML algorithms, Random Forest (RF), Adaptive Boosting (AdaBoost), Light Gradient Boosted Machine (LightGBM), and Extreme Gradient Boosting (XGBoost), have been used to train models. The Ethiopian Demographic and Health Survey (EDHS) dataset is accessed for this purpose from the Central Statistical Agency (CSA)'s database. Various data pre-processing techniques were employed, and the model training has been conducted using the scikit learn Python library functions. Model evaluation is executed using various metrics like Accuracy, Precision, Recall, F1-score, area under curve-the receiver operating characteristics (AUC-ROC), and subjective evaluations of domain experts. An optimal subset of hyper-parameters for the algorithms was selected through the grid search function for the best prediction. The RF model has performed better than the rest of the algorithms by achieving an accuracy of 96.06% and is better suited as a solution model for our purpose. Following RF, LightGBM, XGBoost, and AdaBoost algorithms have an accuracy of 91.53%, 88.44%, and 58.55%, respectively. The findings suggest that some of the features like 'Age of household head', 'Total children ever born' in a family, 'Main roof material' of their house, 'Region' they lived in, whether a household uses 'Electricity' or not, and 'Type of toilet facility' of a household are determinant factors to be a focal point for economic policymakers. The determinant risk factors, extracted rules, and designed artifact achieved 82.28% of the domain expert's evaluation. Overall, the study shows ML techniques are effective in predicting the wealth status of households.

Keywords : ensemble machine learning, households wealth status, predictive model, wealth status prediction

Conference Title : ICMLC 2025 : International Conference on Machine Learning and Cybernetics

Conference Location : Vancouver, Canada

Conference Dates : January 11-12, 2025