

Development of a Lexicographic Multi-Objective Optimization-Based Heterogeneous Ensemble Regression Analysis: A Case of Cost-Benefit Analysis of a Risk Control Measure in a Steel Plant

Ajeet Kumar Shukla, Banerwar Sarker, J Maiti

Indian Institute of Technology, Kharagpur

In this study, a novel approach to developing an ensemble-based regression model using lexicographic multi-optimization is proposed with the goal of achieving improved prediction performance. Four different base models—Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Network (ANN), and K-Nearest Neighbours (KNN)— were implemented as the foundation for the ensemble. The prediction performance of the four models were checked in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) separately. Then, in order to minimize both RMSE and MAE, a lexicographic multi-objective optimization technique was implemented. The methodology centres on prioritizing these objectives using lexicographic ordering and subsequently deriving the optimal weights of the base models through lexicographic multi-objective optimization methodologies. The approach is validated using a benchmark cost-benefit analysis dataset in a steel plant and demonstrates that the proposed ensemble model outperforms individual base models and conventional ensemble techniques, marking a significant advancement in ensemble learning for improved predictive capabilities across various domains.