ABSTRACT

Customer churn is a phenomenon in which customers discontinue the use of a company's services or products, potentially impacting business stability and revenue. In the fitness industry, such as gyms, churn occurs when members choose not to renew their memberships. This study aims to develop a predictive model for gym membership churn using the Random Forest algorithm, known for its capability in handling high-dimensional data and providing stable accuracy estimates through ensemble techniques. Hyperparameter optimization was conducted using RandomizedSearchCV and GridSearchCV to obtain the best parameter combination that enhances model performance. The identification of members likely to churn enables more targeted preventive strategies, supporting optimal efforts to retain customer loyalty. The dataset used, titled *Gym Customers Features and Churn*, was sourced from Kaggle and consists of 4,000 observations. The analysis includes data exploration, preprocessing (missing value detection, multicollinearity, and outlier identification), and splitting the data into training and testing sets. The study shows that the model tuned using GridSearchCV performs better than the one using RandomizedSearchCV, achieving an AUC score of 95.4% on the testing data, indicating that the model falls into the excellent classification category.

Keywords: Customer churn, Random Forest, GridSearchCV, RandomizedSearchCV, Hyperparameter tuning, Gym membership