

ABSTRACT

This research aims to develop a sales prediction model for Big Mart using machine learning techniques. The sales data used includes 1,559 products across 10 stores in 2013 obtained from Vidhya Analytics. The problem faced is how to optimize hyperparameters to improve prediction accuracy. Most of the previous studies used conventional methods or simple hyperparameter optimization, while this study tries to fill the gap by comparing two hyperparameter optimization methods, Grid Search Optimization (GSO) and Bayesian Optimization (BO), in the context of XGBoost model for sales prediction. The objectives of this study are to assess the performance of the XGBoost model with default hyperparameter settings, analyze the accuracy improvement of hyperparameter optimization using GSO, and compare the time efficiency and accuracy between GSO and BO. The sales data used comes from the Big Mart sales dataset which includes various product and store attributes, with model performance evaluation performed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. The default XGBoost using $\text{max_depth} = 6$, $\text{learning_rate} = 0.3$, $\text{n_estimators} = 100$, and $\text{subsample} = 1$ produces an RMSE value of 1213 and MAE of 850 and is used as a baseline. GSO on the XGBoost model produces the best hyperparameters with $\text{max_depth} = 3$, $\text{learning_rate} = 0.01$, $\text{n_estimators} = 50$, and $\text{subsample} = 0.9$ giving an RMSE value of 1070 and MAE of 754. Meanwhile, BO focused on RMSE minimization produces Hyperparameter $\text{max_depth} = 3$, $\text{learning_rate} = 0.0126$, $\text{n_estimators} = 357$, and $\text{subsample} = 0.6$ resulting in the best RMSE value of 1069 and MAE of 755. BO which is focused on MAE minimization produces Hyperparameter $\text{max_depth} = 3$, $\text{learning_rate} = 0.0668$, $\text{n_estimators} = 100$, and $\text{subsample} = 0.9$ achieving an RMSE value of 1071 and MAE of 753. The results showed that BO performed slightly better than GSO in terms of prediction accuracy, but BO was much more efficient in terms of total computation time at 530 seconds compared to GSO at 1481 seconds. This research contributes to the selection of a more efficient and effective Hyperparameter optimization method for predictive models in the retail industry.

Keywords: XGBoost, *Hyperparameter Optimization*, *Grid Search Optimization*, *Bayesian Optimization*, sales prediction