Optimisation-based weighted ensemble algorithm for predicting prices of spices

Ankit Kumar Singh[1], Md Yeasin[2], Ranjit Kumar Paul[2,*], H. S. Roy[2], Prakash Kumar[2], A. K. Paul[2] and Anita Sarkar[1]

[1]The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India
[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

India plays a significant role in global agriculture production, including spices. Spices play an important role in cultural significance and economic trade relations while providing nutritional and medicinal benefits. The volatility and complexity of the price of spices require improved forecasting methods to support informed decision-making in agricultural markets. Recently, researchers have focused on using the traditional time series model as well as machine learning (ML) model to forecast the price of agricultural commodities. Using a standalone model struggles to capture the complex pattern in time series data. To overcome this challenge, ensemble machine learning approaches based on fixed weight (FW-ensemble) have been proposed. The ML models like artificial neural networks (ANN), random forest (RF), *k*-nearest neighbours (kNN), extreme gradient boosting (XGBoost), support vector regression (SVR) and the stochastic model, e.g. autoregressive integrated moving average (ARIMA) model have been used. The outputs of these models are ensembled using optimised fixed weights. In this study, the prices of two important spices, namely turmeric and coriander, from 2010 to 2024, collected from AGMARKNET (https://agmarknet.gov.in/) were considered. The MCS algorithm was used to select the better-performing model. The empirical performance of the ensemble method was compared with that of the stochastic model (ARIMA), ML techniques (ANN, RF, *k*NN, XGBoost, SVR) and deep learning techniques, e.g. long short-term memory (LSTM) and gated recurrent unit (GRU), based on several accuracy measures. It revealed that the FW-ensemble approach significantly outperformed the other candidate models in terms of prediction accuracy.

Keywords: Accuracy, ensemble, forecasting, machine learning, spices.

regulation and maintaining national food security stability. In recent times, researchers have emphasised on accurate price forecast of agricultural commodities[2–4]. Brandt and Bessler[5] discussed various approaches such as exponential smoothing, econometric models, autoregressive integrated moving average (ARIMA) models, expert judgment and composite forecasting methods for predicting time series data and conducted a comparative analysis to evaluate the effectiveness of these methods in predicting the increased price volatility of agricultural products. Alternative time series forecasting models like exponential and trend-seasonal component models are also considered due to limitations of traditional statistical models in capturing omplex patterns, nonlinear relationships and high-dimensional data in real-world applications[6,7].

In contrast, machine learning (ML) algorithms are generally more efficient in identifying dynamic patterns and automatically extracting relevant features. These advantages lead to growing trend towards utilising ML algorithms for predicting nonlinear time series data in recent times[8–11]. Sinta *et al.*[12] utilised the ensemble k-nearest neighbours (kNN) method for forecasting the prices of rice crops. Shengwei *et al.*[13] explored factors influencing agricultural price fluctuations and developed a prediction model using least squares support vector regression (SVR) model for agricultural prices. Paul *et al.*[14] explored the efficacy of various ML algorithms, like SVR, artificial neural network (ANN), random forest (RF) and gradient boosting machine (GBM), in predicting wholesale prices of brinjal, demonstrating better performance of ML techniques with respect to other benchmark models.

While ML models generally exhibit higher accuracy than traditional statistical models, predicting agricultural commodity prices remains challenging due to numerous influencing factors. These include supply and demand dy-