Applied Soft Computing Journal 86 (2020) 105837

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series

Matheus Henrique Dal Molin Ribeiro [a,b,*], Leandro dos Santos Coelho [a,c]

[a] Graduate Program in Industrial & Systems Engineering (PPGEPS), Pontifical Catholic University of Parana (PUCPR), 1155, Rua Imaculada Conceicao, Curitiba, Parana, 80215-901, Brazil
[b] Department of Mathematics, Federal Technological University of Parana (UTFPR), Via do Conhecimento, Km 01 - Fraron, Pato Branco, Parana, 85503-390, Brazil
[c] Department of Electrical Engineering, Federal University of Parana (UFPR), 100, Avenida Cel. Francisco Heraclito dos Santos, Curitiba, Parana, 81530-000, Brazil

ARTICLE INFO

Article history:
Received 15 September 2018
Received in revised form 6 September 2019
Accepted 5 October 2019
Available online 11 October 2019

Keywords:
Bagging
Boosting
Agricultural commodity
Ensemble regression
Stacking
Time series

ABSTRACT

The investigation of the accuracy of methods employed to forecast agricultural commodities prices is an important area of study. In this context, the development of effective models is necessary. Regression ensembles can be used for this purpose. An ensemble is a set of combined models which act together to forecast a response variable with lower error. Faced with this, the general contribution of this work is to explore the predictive capability of regression ensembles by comparing ensembles among themselves, as well as with approaches that consider a single model (reference models) in the agribusiness area to forecast prices one month ahead. In this aspect, monthly time series referring to the price paid to producers in the state of Parana, Brazil for a 60 kg bag of soybean (case study 1) and wheat (case study 2) are used. The ensembles bagging (random forests — RF), boosting (gradient boosting machine — GBM and extreme gradient boosting machine — XGB), and stacking (STACK) are adopted. The support vector machine for regression (SVR), multilayer perceptron neural network (MLP) and *K*-nearest neighbors (KNN) are adopted as reference models. Performance measures such as mean absolute percentage error (MAPE), root mean squared error (RMSE), mean absolute error (MAE), and mean squared error (MSE) are used for models comparison. Friedman and Wilcoxon signed rank tests are applied to evaluate the models' absolute percentage errors (APE). From the comparison of test set results, MAPE lower than 1% is observed for the best ensemble approaches. In this context, the XGB/STACK (Least Absolute Shrinkage and Selection Operator-KNN-XGB-SVR) and RF models showed better performance for short-term forecasting tasks for case studies 1 and 2, respectively. Better APE (statistically smaller) is observed for XGB/STACK and RF in relation to reference models. Besides that, approaches based on boosting are consistent, providing good results in both case studies. Alongside, a rank according to the performances is: XGB, GBM, RF, STACK, MLP, SVR and KNN. It can be concluded that the ensemble approach presents statistically significant gains, reducing prediction errors for the price series studied. The use of ensembles is recommended to forecast agricultural commodities prices one month ahead, since a more assertive performance is observed, which allows to increase the accuracy of the constructed model and reduce decision-making risk.

1. Introduction

The operations inherent to agribusiness constitute an important network of activities that have a strong influence on the local, regional, and national economy. Generally, in times of crisis, agribusiness has been one of the areas that have contributed to the growth of different indicators related to the economy, such as the Gross Domestic Product (GDP). In view of the current Brazilian economic scenario, the GDP volume of agribusiness increased by 1.9%, facilitating national GDP growth and inflation control [1]. At the state level, according to recent reports from the Parana Institute for Economic and Social Development [2], the Parana state ended 2017 with a GDP of R$ (BRL) 415.8 billion, equivalent to 6.35% of the Brazilian economy. Regarding production, according to the Ministry of Agriculture, Livestock and Food Supply (MAPA) [3], the area planted with grains is expected to increase by 14.9% in the next ten years. Indeed, this is an important area

* Corresponding author at: Department of Mathematics, Federal Technological University of Parana (UTFPR), Via do Conhecimento, Km 01 - Fraron, Pato Branco, Parana, 85503-390, Brazil.
E-mail addresses: matheus.dalmolinribeiro@gmail.com (M.H.D.M. Ribeiro), lscoelho2009@gmail.com (L. dos Santos Coelho).

https://doi.org/10.1016/j.asoc.2019.105837
1568-4946/

for the development of the regions, and it also contributes to family income. Faced with this, it is important to know future price quotations, since this may affect the economic planning of small, medium, and large producers.

In this aspect, the forecast for time series of agricultural commodities prices plays an important role in the economic scenario. Also, the precision of the forecasts is as important as the expected results, and it is directly linked to the accuracy of the adopted model. Its importance permeates all areas of knowledge, as forecasts can be used as the basis for revision/implementation of public policies, development of strategic planning in companies, and decisions in the corporate world. In general, forecasting can be understood as a process of idealizing a more probable result, assuming a set of assumptions inherent in several aspects such as technology, environment, pricing, marketing, and production.

To improve the predictive capacity of regression models, among which we can highlight the time series models, several approaches have been proposed in the literature. One of these are regression ensembles [4,5]. This principle is based on machine learning, whose essence is associated with the idea of divide-to-conquer and which seeks to overcome the limitation of a machine learning model that operates in isolation [6,7]. Models characterized as ensembles refer to multiple base models (base learners, similar or not) developed to solve the same problem, each of which learns characteristics from data and accordingly makes a prediction. Hence, the final prediction is presented as a combination of the individual predictions. Through this approach we want to obtain a robust system that incorporates the prediction from all base learners. Among the basic ways to build ensembles for regression problems is the grouping by average and weighted average. In general, this approach achieves better results than single prediction models, since it has a better capacity to generalize and adapt in different scenarios [8]. The purpose of using regression ensembles is to generate a set of models to improve the prediction accuracy of the adopted models, whose target variables are numerical [9]. In this aspect, it is a methodology that has been used extensively over years and has been successful in different applications such as energy [10], climate [8], market forecasting [11], finance [12,13], and chemistry [14].

From this perspective, this paper aims to explore the predictive capacity of regression ensembles by comparing ensembles among themselves, as well as with approaches that consider a single model (reference models), to forecast agricultural commodities prices one month ahead. The main motivation for the choice of this forecast horizon for the area of agribusiness is associated with the fact that short term forecasting allows small, medium and large producers to develop strategic planning for a short period of time to meet their instantaneous necessities. This paper conducts two case studies in which historical data of monthly prices (R$) paid to the producers in the state of Parana, Brazil, are considered for 60 kg bags of soybean (case study 1) and wheat (case study 2). The specific objectives are stated as follows: (i) Identify, by related works, the main variables that act as price drivers in the soybean and wheat cases. In addition, extract the importance of them for each case study; (ii) Present a theoretical revision of the regression ensembles based on bagging (random forests — RF), boosting (gradient boosting machine — GBM and extreme gradient boosting machine — XGB) and stacking (STACK) and (iii) verify if the ensemble approaches outperform single models (K-Nearest Neighbor — KNN, Multilayer Perceptron Neural Network — MLP and Support Vector Regression with linear Kernel— SVR), as well as which structure is better in the adopted cases.

The main contribution of this paper to the soft computing and machine learning literature lies in the application, evaluation and addition of discussions not previously developed regarding the behavior of GBM, RF, STACK and XGB to forecast (in the short-term) agricultural commodities prices, with the use of the price drivers identified in Section 3.2. While the papers identified in Section 2 focus on the study of price volatility by hybrid models, as well as combining decomposition, optimization and machine learning models, this paper proposes the investigation of the performance of bagging, boosting and stacking ensemble approaches in the presence of different features (price drivers) regarding short-term forecasting for the soybean and wheat commodities.

Through these contributions, it is expected that once a few time series models using regression ensembles to predict the agribusiness theme are introduced in the literature, this work will fill a gap in an area of the economy essential to the composition of economic indicators. Seeking to illustrate the superior performance of the ensemble models vis-à-vis reference approaches to forecasting in time series associated with agribusiness, it is expected that this study will stimulate the interest of financial market researchers to use this approach, since it performs well and enables managers to minimize their losses and leverage their earnings.

The remainder of this paper is structured as follows: Section 2 presents a set of related works on the subject of regression ensembles for commodities forecasting. Section 3 describes the data sets used in this paper, as well as the predictors employed in the modeling process. Section 4 introduces the theoretical aspects necessary for the development of this paper. Section 5 presents the adopted methodology. Section 6 details the results and discussions. Finally, Section 7 concludes the paper with general considerations and directions for future research.

2. Related works

Commodities comprise an important chain of products that directly influence the global economy. These products are grouped into categories as follow: energy, metal, livestock, and agriculture [15]. In this section, some recent works are presented regarding the prediction of prices, volatility, and future returns for the aforementioned sectors, using regression ensembles, as well as the main gaps are pointed out. Table 1 summarizes some related works with regard to objective (ensemble adopted), commodity category and findings/results.

Considering the works cited in Table 1, it is possible to identify that different approaches, based on regression ensembles, have been used to predict commodities prices, return and volatility as well as have achieved good results against traditional methodologies, in special over models that considering a single prediction. However, it is also possible to verify that, for the most part, these studies do not use the methodologies adopted here, such as bagging, boosting, and stacking, especially in approaches with focus on short-term forecasting for agricultural commodities area. Faced of this, this study seeks to fill this gap, as well as to contribute to the dissemination of the machine learning models approach in time series forecasting field.

3. Cases studies

The objective this section is to present the cases studies adopted in this paper. The first subsection presents the cases, and the second subsection presents the predictive variables adopted in the modeling process as well as some considerations.

3.1. Soybean and wheat cases

To meet the objective of this research, monthly series of the prices received by the producers in the state of Parana, Brazil, for soybean and wheat agricultural commodities are used. These

Table 1
Selection of publications with different ensemble learning approaches used to forecast commodities prices.

Publication	Objective	Commodity	Finding
He et al. [16]	Developing linear ensembles associated with wavelets, based on multiple-resolution analysis to forecast copper, gold and silver prices.	Metal	Predictions generated by proposed model are more accurate than autoregressive integrated moving average (ARIMA) generalized autoregressive conditional heteroskedasticity models.
Gabralla et al. [17]	Proposing an ensemble combining three types of artificial neural networks (ANN) (Feedforward — FNN, Recurrent and Radial Basis Function), by weighted average to forecast the West Texas intermediate oil prices.	Energy	Showed that the neural network ensemble surpassed the individual (neural networks) models evaluated.
Pierdzioch et al. [18]	Applying L_2-boosting algorithm to forecast silver and gold price returns.	Metal	Errors and biases of the predictions are significantly reduced. Moreover, comparisons with other models are not made.
Pierdzioch et al. [19]	Evaluating boosting approach with different loss functions to forecast gold volatility.	Metal	When the results are compared to an autoregressive (AR) models, the boosting showed better results.
Yu et al. [20]	Developing a decomposition-and-ensemble (EEMD) learning approach and extreme learning machine (ELM) to forecast crude oil prices.	Energy	Proposed models showed better results regarding individual ELM, least squares support vector regression (LSSVR), ANN and ARIMA, as well as when associated to decomposition paradigm.
Yu et al. [21]	Applying an ensemble of LSSVR with results aggregated by weighted average to forecast crude oil price.	Energy	Proposed model showed better results when compared to classic prediction methodologies and individual LSSVR.
Zhao et al. [22]	Developing ensemble of deep learning (stacked denoising autoencoders associated with bagging — SDAE-BAG) to forecast oil gross price per barrel.	Energy	This approach is tested against some competing methodologies (random walk, Markov regime switching, FNN, SVR and SDAE) and showed superior forecasting capability.
Wang et al. [23]	Combining backpropagation neural network optimized by particle swarm with four decomposition approaches: empirical mode decomposition (EMD), wavelet packet transformed, intrinsic time-scale decomposition, and variational mode decomposition to forecast wheat, corn, and soybean prices.	Agricultural	The approach that combines variational mode decomposition, particle swarm and backpropagation neural network showed better results than other combinations of proposed approaches.
Yang et al. [24]	Adopting bagging associated with heterogeneous autoregressive (BHAR) model to forecast the volatility of agricultural commodities futures contracts such as soy, cotton, wheat, and corn in China.	Agricultural	Predictions made by the BHAR model and principal components analysis produce the lowest metrics in relatively longer forecast horizons.
Xiong et al. [25]	Hybridizing EEMD, ARIMA and artificial intelligence tools such as FNN, SVR, and ELM for short-term forecast of hog prices in China.	Livestock	Results showed that the enhanced EEMD–ARIMA-ELM-SVM outperforms the selected counterparts.
Pierdzioch and Risse [26]	Evaluating multivariate RF to forecast future returns of gold, silver, platinum, and palladium.	Metal	Proposed approach outperform univariate models.
Tang et al. [27]	Combining EEMD with predictions generated by random vector functional link algorithm to forecast crude oil prices.	Energy	Results from proposed approach outperform EEMD-ELM.
Wang et al. [28]	Adopting EEMD to study the global volatility of the corn, wheat, and rice commodities.	Agricultural	Results showed that decomposition used allows more accurate evaluation of the different components present in the price series and identification of different agents in the price changes.
Bonato et al. [29]	Applying quantile-boosting approach to forecast future returns for gold.	Metal	Results obtained from proposed model are better than results from AR models.
Ding [30]	Combining EEMD process associated with ANN to forecast oil prices.	Energy	Results suggest that the EEMD-ANN method presents better predictions than other EMD-ANN.

commodities are chosen because they are part of an important productive chain, being a source of income for many producers, as well as they contribute to different aspects of the economic sector. The general dataset will be divided into two parts, the first one being used for training the machine learning models (70% of the data) and the second to test the performance of the trained model (30% of the data) [31]. Continuous periods are separated for the two sets, so that the temporal dynamics of the data are considered. In this setting, out-of-sample forecasts tend to be more accurate [32]. The period divisions of the training and test sets, as well as observations number (n), for each case study, are presented in Table 2.

Table 2
Periods used for training and test set (Month/Year — n).

Commodity	Data	Period (n)
Case study 1 (Soybean price)	All	07/2001 to 05/2018 (203)
	Training	01/2001 to 05/2013 (143)
	Test	06/2013 to 05/2018 (60)
Case study 2 (Wheat price)	All	02/2004 to 05/2018 (172)
	Training	02/2004 to 02/2014 (121)
	Test	03/2014 to 05/2018 (51)

From Fig. 1(a), it is possible to visualize the study series. It can be seen that there are well defined periods of increase

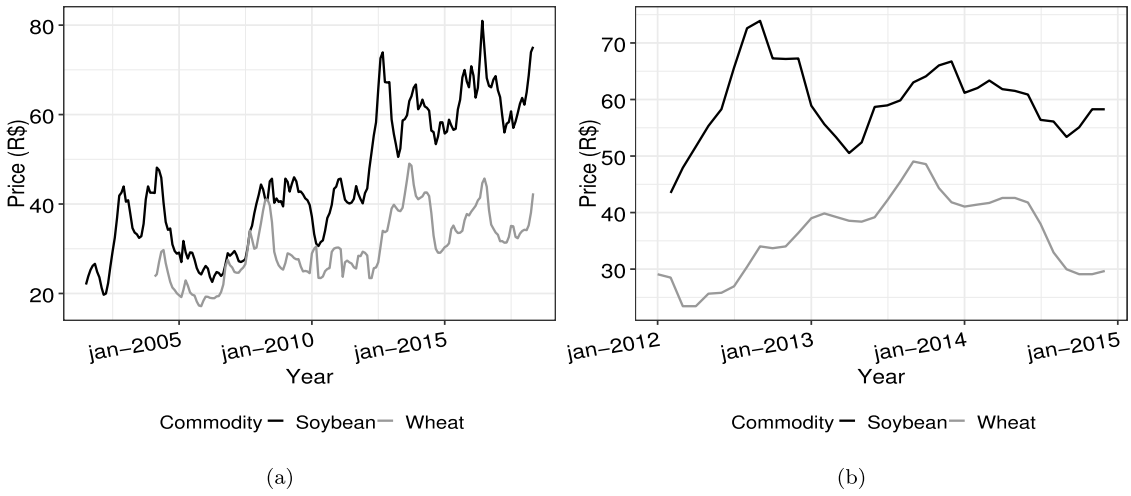


Fig. 1. Agricultural commodities prices received by the producer in Parana, Brazil in entire period (left) and high volatility period (right).

Table 3
Price drivers for soybean and wheat commodities.

Publications	Price Driver	Period
Bini et al. [34]	EC	2000–2012
Fernandez-Perez et al. [35]	EC	2006–2016
Paris [36]	EC	2001–2014
Caldarelli and Bacchi [37]	ICP	1967–2008
Alves et al. [38]	ICP	1995–2005
Minga et al. [39]	ICP	2003–2012
Bodart et al. [40]	ER	1980–2012
Baffes and Haniotis [41]	ER	1960–2014
Pereira et al. [42]	ER	2002–2002
Moraes et al. [43]	FM	2003–2012
Rezende et al. [44]	FM	2004–2017

and decrease in the paid prices, suggesting the presence of seasonality and cyclicity in prices. Hence, it is possible to identify nonlinear patterns behavior of the series. Faced with this, using machine learning approaches to forecast prices may bring a gain over single models. In addition, for these commodities, between 2012 and 2015 the prices show high volatility as illustrated in Fig. 1(b). This, perhaps, may be associated to combination of some factors, such as weather, politics, reduced stock and growth in consumption. Moreover, in this period there was growth of exchange rate (R$/US$) [33]. This paper did not focus on the insertion/description of this events as predictors, once the variables described in Section 3.2 may be able to add this behavior to the data.

3.2. Predictive variables

Understanding which agents influence commodity pricing policies is of the utmost importance. This makes it possible to evaluate the timeliest moment in the financial market against these elements, rendering decision making more assertive. With regard such price driver in relation to agricultural commodities such as soybean and wheat, there is no consensus regarding which agents play a role in price determination. Thus, in the literature it is possible to find a range of references that sought to evaluate the association of their prices with factors such as: prices of energy commodities (EC), exchange rate (ER), interaction between commodities prices in domestic (ICP) and foreign markets (FM). Indeed, Table 3 presents some references which guide the choose of adequate predictors to forecasting soybean and wheat prices.

From Table 3 information, to forecast the prices of commodities adopted in this paper, the response (R) and predictors (P) variables are described in the sequence. The responses are the price (R$) of a 60 kg bag of soybean (R_1) and wheat (R_2) received by producers in state of Parana, price (R$) of a 60 kg bag of soybean (P_2) and wheat (P_3) from the CEPEA/Esalq-USP (*Centro de Estudos Avançados em Economia Aplicada - Escola Superior de Agricultura Luiz de Queiroz/Universidade de São Paulo*, in Portuguese), exchange rate (US$ for R$) ($P_4$), average pump price (R$) per liter of ethanol (state PR) (P_5), price of a 50 kg (R$) bag of wheat flour (state PR) (P_6) and future price of soybean (P_7) and wheat (P_8) contract (US$) quoted on the Chicago Board of Trade (CBOT). The historical series for the variables R_1, R_2, P_2, P_4, and P_6 are obtained from the website of the Institute of Applied Economics Research (IPEA) (*Instituto de Pesquisa Econômica Aplicada*, in Portuguese).[1] The information for the variables P_1 and P_3 are obtained from the website of the Center for Advanced Studies in Applied Economics — CEPEA/Esalq-USP.[2] Information regarding the variable P_5 is collected from the website of the National Petroleum Agency (ANP) (*Agência Nacional de Petróleo*, in Portuguese).[3] The variables P_7 and P_8 are obtained from the website of the CME Group's[4] using the Application Programming Interface (API) Quandl [45]. Initially, the source code of the desired data extraction is obtained from the website https://www.quandl.com and subsequently used in R Language. In this context, Table 4 summarizes the response variables and their predictive variables, based on previous studies mentioned above in this section.

Table 4
Distribution of the response variables of case studies 1 and 2 with the respective predictive variables.

Case study	Response variable	Predictive variable
1	R_1	P_1, P_2, P_4, P_5, P_7
2	R_2	P_3, P_4, P_5, P_6, P_8

The Table 5 introduces the descriptive measures (DM) for the response and predictive variables for both case studies. Faced with this, in general the measures for soybean (R_1) and wheat price (R_2) follow the results from CEPEA indicators (P_1 and P_3), as well as from CBOT indicators (P_7 and P_8). In the same way,

1 http://www.ipeadata.gov.br/Default.aspx
2 https://www.cepea.esalq.usp.br/br/consultas-ao-banco-de-dados-do-site.aspx
3 http://www.anp.gov.br/precos-e-defesa-da-concorrencia/precos/levantamento-de-precos
4 http://www.cmegroup.com/

Table 5
Descriptive measures for the variables of case studies 1 and 2.

DM	R_1	R_2	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
1st quartil	31.81	25.32	35.62	14.76	27.46	1.97	1.43	57.63	14.10	9.38
3th quartil	58.56	35.13	64.77	22.17	40.06	3.02	2.05	87.45	26.38	14.57
Mean	45.11	30.41	49.87	18.91	34.00	2.48	1.78	72.38	21.23	12.10
Median	42.53	29.69	46.32	18.20	31.70	2.31	1.67	68.12	21.24	11.27
SD	15.19	7.24	16.67	5.81	9.05	0.63	0.52	18.70	7.10	3.67
CV (%)	33.66	23.79	33.43	30.75	26.60	25.62	29.30	25.84	33.45	30.35

this happen to other predictors (P_2, P_4, and P_5), although with less intensity. Alongside this, the variations (CV) for all variables are similar. In this way, this indicates that the price paid to the producer tends to follow the oscillation of possible prices that are directly linked to them.

4. Theoretical aspects

Decision-making is an essential step in different situations, especially in data modeling. However, since the results obtained will be the basis for decision-making, they should be as accurate as possible. In this aspect, the description and implementation of techniques that allow forecasting with greater accuracy should be the object of study in many situations, in order to make decision-making more concise and assertive. Therefore, in the following Subsections we describe the main theoretical aspects necessary for understanding the development of this study.

4.1. Ensemble learning for regression tasks

In the context of predicting a time series, an ensemble performs the partitioning of the original series into simpler series whose purpose of decomposition is to apply prediction algorithms in each decomposed series. The general prediction is obtained by aggregating, on average, the results of the individual predictions, referring to the decomposed series [46]. In general, according to Allende and Valle [47], the advantages of combining models for time series forecasting can be credited to three facts, namely: (i) it increases overall forecast accuracy to a large extent by using appropriate aggregation techniques, (ii) there is often a great uncertainty about the optimal forecast model and, in such situations, combination strategies are the most appropriate alternatives, and (iii) the combination of multiple forecasts can efficiently reduce errors. In addition, using ensembles to carry out forecasting can be justified in different ways, for example, it may be associated with greater ability to generalize results, reduced model variance, and greater tolerance to data noise [9]. Therefore, regression ensemble techniques for forecasting with data observed over time must be explored.

From this perspective, this section seeks to present the main aspects inherent to these methodologies, as well as the concepts necessary for understanding the development of this paper. Alongside this, a brief of the models adopted as reference (KNN, MLP and SVR) is presented.

4.1.1. Bagging

Bagging, or bootstrap aggregating, is a classical technique for the creation of an ensemble proposed by Breiman [48]. It was initially proposed for use in problems of classification. However, it can be used in problems that aim to perform data regression. It is characterized by the creation of multiple samples, with refitting by means of the bootstrap technique, from the same set of data, so that it is possible to build multiple distinct trees for the same predictor and use them to generate an aggregate prediction. The final prediction from this process can be obtained by voting or average, for classification and regression problems respectively [49]. This allows the generation of multiple samples for the same set. The advantage of using this technique for ensemble generation is that it allows for the reduction of error in the baseline predictors, which can be considered unstable before certain perturbations and that it can provide its own estimates of predictive performance, which correlate with estimates of cross-validation or estimates of test sets [50].

When using the bagging technique, the built trees may have similar characteristics, thus implying an absence of diversity. A way of circumventing this problem is to use RF [51], which can be understood as modeling using bagging together with a random selection of predictors to compose each node of the decision tree. This method, together with the selection of the best set of characteristics that explain the data, represents an important step in this approach [52].

In accordance with the method described by Breiman [51], the RF algorithm performs in the following way: (i) Given a training set with N observations for a variable response and M predictors, B samples are generated via the bootstrap technique, (ii) beginning from the previous item, B regression trees are built, one for each sample, each node containing a randomly chosen subset of M predictor variables, (iii) with the best subset obtained in step (i), an estimate of the response variable is generated from each regression tree and (iv) a final prediction is generated via average considering estimates obtained in step (iii). In general, the main objective of this approach is to improve the performance of regression trees through the reduction of their variance [53,54].

4.1.2. Boosting

Boosting is a technique used to solve problems of classification and regression, which combines a set of models (decision trees), whose individual performance can be considered weak, whereas when the individual predictions are aggregated, a model of high accuracy is obtained. The central focus of this approach is bias reduction, that is, improving the suitability of the model to data [55]. Several machine learning models have been proposed in this idea, using decision trees with individual learners, but we highlight the GBM and XGB.

The GBM algorithm is based on the boosting principle, proposed by Friedman [56]. It uses the gradient descent approach to build models in the negative sense of the partial derivative of the loss function in relation to the set of predictions. This process can be understood from a numerical optimization perspective, which seeks to find an additive model that minimizes the loss function [57]. From this perspective, a model (decision tree) is initially adjusted to the data and, on this basis, predictions are performed and the initial residue is obtained. A new model is adjusted to the previous residuals, a new prediction, to which the initial forecast is added, and then a new residue is obtained. This process is repeated until a convergence criterion is reached, and the final prediction is obtained by average in regression problems. In each iteration, a new model is adjusted to the data, aiming to compensate for the weaknesses of the previous model. However, this process tends to generate overfitting. To avoid this problem, weights are assigned to observations predicted with the biggest errors [58]. Other of GBM's abilities is to provide an importance rank for the characteristics used in the model from the adjusted model. In this way, one can identify and classify the influences of

independent variables on forecasts of response. This tool is used in the case studies presented herein and further details can be found in [59,60].

Other variation that has received attention from the academic community is the XGB algorithm [61,62]. It is an extension of the GBM approach, and its objective is to prevent overfitting and optimize the available computational resources. Aiming at this, a regularization term with weights, responsible for controlling the complexity of the model, is added to the loss function. The iterative learning process of the algorithm is identified as an iteration continues and the loss function decreases gradually, but the regular terms continue to expand [63]. Considering the range of possibilities for the loss function and the possibility of the lack of analytically forms for the derivatives of these functions, the loss function is replaced by a Taylor polynomial of order 2 and the objective function is optimized [64]. Within this context, the XGB derives an analytically expression to evaluate the performance of the model in terms of the original objective function, whereas if the gain for that iteration is negative, that subdivision must not be considered. According James et al. [65], we note that the advantage of adding a regularization term to the loss function in the objective function has its basis in the trade-off of the bias and variance. To the extent that the regularization increases, there are gains in reduction of variance in detriment to the increase of bias of the estimates.

As regard as the GBM and XGB approaches, several parameters should be defined. This step plays an important role for modeling process, once that an incorrect choice for these parameters tends to make the training process costly and the models developed less accurately.

4.1.3. Stacked generalization

Stacked generalization or simply STACK is other approach that combines different prediction models in a single model, working at levels or layers. This approach introduces the concept of meta-learning, represents an asymptotically optimal learning system, and aims to minimize the errors of generalization by reducing the bias of its generalizers [66,67].

Considering a STACK approach using two levels or layers (level-0 and level-1). In level-0, different models are trained, and subsequently the prediction of the response variable for each one is performed. These forecasts are used as an input set for the level-1 model. This model, too called of meta-model, is trained, and its prediction is the desired end result. In general, the model of the subsequent level, learns from the previous level models, where each model of the level immediately below provides its best estimate [68–70]. In general lines, the model in level-1 learn with the predictions of the models of the level-0. Special attention should be given to the fact that the number of levels need not be restricted to two. The rule is that the n level model must use the predictions of the $n-1$ level models [71].

The main improvement in the predictive results, when STACK is applied, is apparent when there is diversity among the models of the different levels. This occurs because models with different principles of generalization tend to yield different results [9]. Diversity can be included in the modeling process by using models that follow different learning strategies, using different data characteristics and/or observations. In this way, the objective is that models of different levels disagree with each other, introducing a natural diversity in the set that allows for the modeling of different dynamic patterns in a scenario of time series forecasting [72,73].

4.1.4. Advantages and disadvantages of ensembles in regression tasks

Per the previous sections, regression ensembles comprise a set of tools that can solve different types of problems and are responsible for creating predictive systems with high accuracy. On the one hand, the following can be considered as this method's advantages: better capacity of generalization compared to individual models, flexibility to adapt to different tasks; ability to model nonlinear patterns, handling missing data, and identifying the importance of regressor variables. On the other hand, it has the following disadvantages: difficulty in determining values for the set of hyperparameters, computational costs, difficulty in definition of models for the composition of the STACK approach; and difficulty in results interpretability [74,75].

4.2. Reference models

In this section characteristics and structures of reference models are described.

4.2.1. K-nearest neighbor

Such as several models for supervised learning, KNN approach can be applied in classification and regression problems [76]. Indeed, in the following lines, the model structure adopted in this paper as well as the interpretation for time series problem are presented.

In this context, let $\mathbf{y} = \{y_1, \ldots, y_M\}$ a training set composed by M prices (M = 1, ..., 143 for case study 1 and $M = 1, \ldots, 121$ for case study 2), in which each y_M is associated to N features ($N = 1, \ldots, 4$) or in this paper, the price drivers. Giving a set of training prices, the algorithm will maps k nearest past similar prices to new prices drives, in which these k values are namely nearest neighbors. In this aspect, to find the nearest values, a similarity measure is adopted. In this context, the k-nearest neighbors are those that similarity measure between past prices and new prices is the smallest. Considering that the set of k-nearest neighbors are defined, the future prices will be obtained by average of past similar values. Adjacent to the simplicity of this approach, computational cost may be a disadvantage.

Therefore, considering the above mentioned, in this paper, for both case studies the euclidean distance is adopted as similarity measure, the new prices are obtained as average of k past similar prices. Alongside, the k value for each case study is obtained in training process by grid search during cross-validation step and these information are summarized in Table 6.

4.2.2. Multilayer perceptron neural network

The MLP is a kind of feedforward neural networks proposed to overcome the limitations of single perceptron [77]. Faced with this, it is able to solve nonlinear problems, and then suitable for problems associated with time series. Basically, this architecture have three basic features stated as follow: (i) Each neuron in the network includes a differentiable activation function, (ii) the networks exhibits high degree of connectivity, (iii) contains one or more hidden layers between the input and output layers. In order to training the MLP several approaches can be adopted, such as the backpropagation algorithm. This approach proceeds in two phases, namely: forward and backward [78].

In addition, the mathematical expression for this network can be stated as follows:

$$y = f_o\left\{\sum_{j=1}^{N} \mathbf{w}_j f_H\left[\sum_{i=1}^{n} \mathbf{h}_{ij}\mathbf{X}_i + \mathbf{b}_j\right] + \mathbf{b}_o\right\} \tag{1}$$

in which y is the output, $\mathbf{X}$ is the input vector, $\mathbf{h}_{ij}$, $\mathbf{b}_j$ and f_H are the, respectively, the weight matrix, the bias vector and the activation function of the hidden layer, and $\mathbf{w}_j$, $\mathbf{b}_o$ and f_o are the weight vector, the bias scalar and the activation function of the output layer [79]. In this paper, the MLP structure is composed by three feedforward layers, that is, one input layer with four inputs

(price drivers), one hidden layer, in which the neurons number is obtained by grid search during cross-validation step, and one output layer (commodities prices). In addition, the weights are initialized randomly in interval (−0.3, 0.3) and backpropagation algorithm is used to train the MLP architecture. Finally, the identity activation function is used in hidden and output layers. Alongside the number of neurons used in the hidden layer for each case study are presented in Table 6.

4.2.3. Support vector regression

Initially, support vector machines were proposed to solve classification tasks, and then were extended to solve regression problems (SVR) [80]. The main idea behind this methodology consists in determining points close to a hyperplane (support vectors) that maximize the margin between two point classes (points greater and lower than the target variable) obtained from the difference between the target value and a threshold. In other words, this approach aims to minimize the structural risks in regression problems, seeking to minimize the upper bound of the generalization error, rather than minimize the training error [81].

Due to the fact of most of the real problems have nonlinear characteristics, the kernel concept can be add to SVR approach. By using kernel functions it is possible to map the inherent characteristics of the data. In the literature, several kernels functions are used, such as Gaussian, radial basis, polynomial and linear. In this context, in this paper the linear kernel is used for both case studies, which can be stated as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i' \mathbf{x}_j, \tag{2}$$

which is equivalent to an inner product between $\mathbf{x}_i'$ and $\mathbf{x}_j$ observations in some feature space. When this kernel is used, the parameter cost should be obtained. In this context, this parameter is obtained by grid search during cross-validation step, and for each case study these information are summarized in Table 6.

On the one hand, the main advantages of the use of SVR for the case studies adopted in this paper, especially for layer-1 in the STACK approach, concern in its capacity to capture the predictor nonlinearities and then use it to improve the predictions of the prices. In the same direction, it is advantageous to employ this perspective in the case studies adopted since the samples are small [80]. On the other hand, the main disadvantage for the data set adopted is related to the choice of kernel function to be used. The use of an incorrect function can lead to erroneous conclusions.

5. Methodology for agribusiness time series modeling

With the purpose to clarify the modeling steps, this section describes the main steps adopted in the data analysis. First, the data preprocessing is presented by the Box–Cox transformation. Second, the hyperparameters overview is presented, and then the control hyperparameters for adopted models are described in Table 6 contained in Section 6. Third, the time series cross-validation process is described. Fourth, an overview of STACK modeling is shown, as well as the performance measures. Finally, an infographic is presented in Fig. 2.

5.1. Data preprocessing

One of the main steps of modeling is data preprocessing, which permeates through different stages and highlights the transformation of the data as an essential step in the construction of intelligent systems. In short, different models have different sensitivities according to their characteristics, so that transformations can reduce the impact of characteristics such as asymmetry, thus leading the model to present better results [82]. This is a modeling phase that allows for the extraction of the maximum potential from the characteristics of a data set, increasing the flexibility, simplicity, and accuracy of the models. Regarding the forecasting in time series, applications of the Box–Cox transformation (BC) to the data are common. It uses the method of maximum likelihood estimation to estimate the parameter λ of (3),

$$y^* = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0 \end{cases} \tag{3}$$

where y^* is the transformed value. It is easy to observe that when $\lambda = 0$, this transformation is reduced to the logarithm transformation [83]. In this way, BC transformation is applied.

5.2. Hyperparameters

The different models used herein have different sets of hyperparameters, so that, depending on the set of values used, the predictive performance can be maximized [84]. Therefore, models of greater parsimony and accuracy can be constructed. As a result, because such parameters control the complexity of the model, wrong choices can result in models prone to improper inferences, as well as models prone to overfitting [11]. In the boosting approach, due to the amount of parameters to be defined, different combinations tend to make the training process lengthy. Therefore, the step of setting such values should be undertaken with caution. In this way, Table 6 presents grid values, as well as the values for the hyperparameters adopted for the models used in case studies 1 and 2.

5.3. Cross-validation in time series

Once that the data sets are preprocessed, in this paper the models structure assumed for training phase are stated as follow

$$\hat{R}_1(t) = f\{P_1(t-1), P_2(t-1), P_5(t-1), P_7(t-1)\}, \tag{4}$$

$$\hat{R}_2(t) = f\{P_3(t-1), P_5(t-1), P_6(t-1), P_8(t-1)\}, \tag{5}$$

where f is a function of the model adopted to modeling, t is the time of evaluation, $R_c(t), (c = 1, 2)$, are the output for case studies 1 and 2, $P_p(t-1), (p = 1, 2, 3, 5, 6, 7, 8)$ are the predictors values lagged in one month, for case studies 1 (Eq. (4)) and 2 (Eq. (5)).

In the sequence, according to each method presented in Section 4.1, it is clear that different parameters must be defined for each situation. Therefore, a way to determine the appropriate combination of parameters of the models discussed is by grid search using cross-validation. Through this, it is possible to minimize forecasting errors associated with test samples in order to find the set of hyperparameters that maximize the performance of the models and avoid overfitting [85]. Since the case studies addressed here refer to time series data, the approach of leave-one-out cross-validation (LOOCV) is suitable [86]. This approach can also be understood as a cross-validation procedure in sequential blocks and a particular case of k-fold cross-validation.

In this perspective, training is conducted in an iterative way, in which training and validation sets are used. The process is carried out several times and in each iteration the training set contains more observations than in the previous iteration and the validation set has less. The corresponding training set consists only of observations that occurred prior to the observation that forms the test set [87]. For the case studies addressed here, the *trainControl* function contained in *caret* package is used. The main parameters adopted are: (i) method equals *timeslice*, (ii) initial

window equals 70% of training set, (iii) horizon, equals 10% of training set to compose the validation set and (iv) fixed window equals *true*. In this process, the training window moves ahead with fixed horizon. The process is carried out until entire training set is used. The training performance is obtained by the average of performance measures of iterations. Further details can be obtained in Bergmeir et al. [88].

5.4. Structure of stacked generalization modeling

In this paper, the STACK modeling will be performed considering two levels. In this context, SVR with linear kernel and least absolute shrinkage and selection operator (LASSO) regression are adopted as meta-learners for case studies 1 and 2. The follow steps are adopted.

1. After training the (1) GBM, (2) SVR/LASSO, (3) KNN, (4) MLP, (5) RF and (6) XGB models, the predictions are combined (2 in 2, 3 in 3, 4 in 4 and 5 in 5) with the purpose of building a STACK layer-0, and then a meta-learner is trained. The model adopted in STACK layer-1 is not used in layer-0;
2. In step 1, for each case study, 56 models are evaluated, and the choice of the best is made using results from the test set;
3. For the results presented in Appendix, the models numbered between 1–15 refer to the combination of the models above taken 2 in 2, between 16–35 refer to the combination of 3 in 3, 36–50 refer to the combination of 4 in 4, and between 51–56 refer to the combination of models taken from 5 in 5 in the order mentioned in step 1;
4. After training each STACK model, the performance measures for the training and test set are obtained.

5.5. Performance measures

After adjusting the various models for each response variable, a model that best explains the variability of the data must be chosen, and, consequently, in the case of time series prediction, this is the one that has a greater accuracy in the generalization of the results. Faced with this aspect, different performance measures have been used to assess the models performance, in which the most frequent being mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), and root mean squared error (RMSE). In this aspect, for the evaluation of the models adopted in this paper, such performance measures will be used, and can be stated in Eq. (6)–(9).

$$MAE = \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{n} \right|, \tag{6}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{7}$$

$$MAPE = 100 \times \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{9}$$

where n represents the number of observations of the modeled time series, y_i is the i-th value observed for series and $\hat{y}_i$ is the i-th value predicted by the model adopted. In case study 1, training set $i = 1, \ldots, 143$ and test set $i = 144, \ldots, 203$ and in case study 2, training set $i = 1, \ldots, 121$ and test set $i = 122, \ldots, 172$ are adopted.

In addition to the performance measures mentioned above, the Friedman test is applied to assess whether there is a statistically significant difference between the absolute percentage error (APE) of the models evaluated. Given the significance, the Wilcoxon signed rank test (lower tail) is used to verify whether the APE of the models are lower in relation to each other [89]. The null hypothesis of Wilcoxon test say there is no reduction in APE of model 1 against model 2, and the alternative hypothesis say the APE of model 1 is lower than of the model 2. Ljung–Box test is used to evaluate the standard residuals of final models for each case study [90].

The results presented in Section 6 are generated using the processor Intel(R) Core(TM) i5-4200U, central processing unit with 1.6 Hz, 8 GB Random Access Memory with Windows 10 operating system. We used the R software [91] with the aid of Application Programming Interface (API) Quandl [45] for the extraction of the series of prices outside Brazil and the *caret* package [92] is used to train the models. For each model, only numeric parameters are chosen by grid search throughout the training phase. For the XGB model, a linear booster is used, and for the MLP model one hidden layer is adopted. Moreover, default parameters adopted by the *caret* package for each model are used.

Fig. 2 presents an overview for the steps adopted to conduced the data modeling.

6. Results and discussions

In this section, the exploratory analysis of the time series used in each case study is presented (Section 6.1), followed by the results of performance measures (Section 6.2), statistical tests for the errors of test set and residual analysis. After each Table and Figure, discussions are made. The Tables A.1 and A.2, which contain the performance measures of the 56 generated models, are presented in the Appendix.

6.1. Exploratory analysis

In Fig. 3, the correlation matrix for the explanatory variables in each case study is presented for the training set observations, after applying the Box–Cox transformation. Faced with this, the λ value is close to 0, and thus the log transformation is adopted. In addition to the variables whose price is quoted in dollars (P_7 and P_8), the Dollar-Real exchange rate (P_4) is used for the conversion. Then, the exchange rate is combined with soybean and wheat prices in the CBOT, giving rise to the variables P_7 (CBOT soybean price in R$) and P_8 (CBOT wheat price in R$) that are employed in case studies 1 and 2, respectively. In this context, only the resultant variable P_7 and P_8, in R$, are inserted in the correlation analysis and not P_4. The color scale in relation to its correlation is presented at the bottom of the Fig. 3. More intense colors indicate a close correlation of 1 or -1, whereas lighter colors indicate a correlation close to 0. In both case studies, the indicator CEPEA/Esalq-USP (P_1 and P_3) and response variable (R_1 and R_2) are positively correlated. In addition, this relation also occurs between predictors. In this aspect, this suggests that the increase or decrease in the value of one tends to increase or decrease those that are correlated with each other. In this way, this indicates that the price paid to the producer tends to follow the oscillation of possible prices that are directly linked to them.

6.2. Evaluation of proposed models

For each case study, the models cited in Section 4.1 are trained and the set of hyperparameters that maximize the predictive performance of each are obtained by grid search. Table 6 refers to hyperparameters grid and the best values adopted for the models.

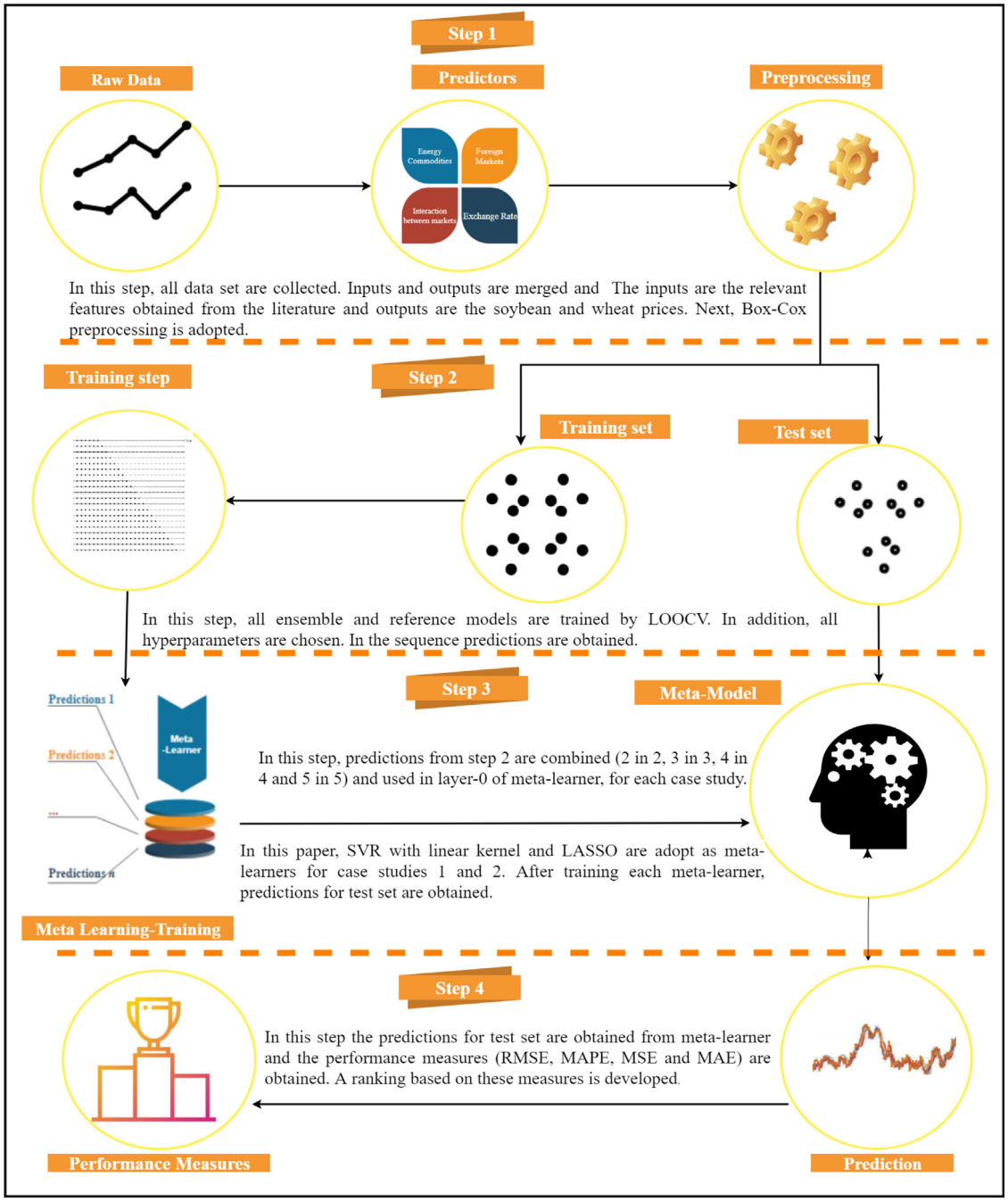
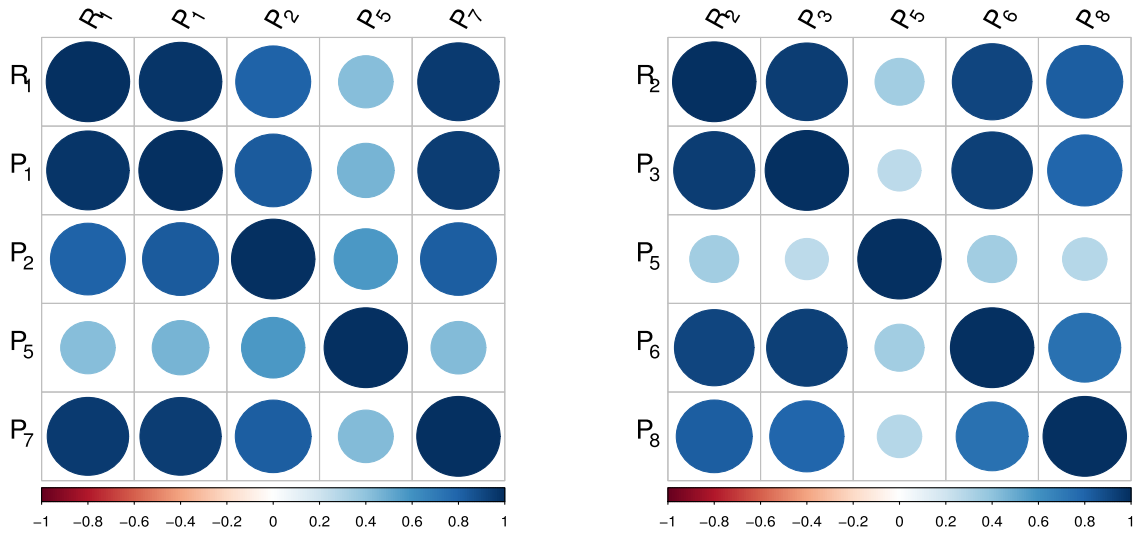


Fig. 2. Roadmap for the proposed data modeling.

Table 7 presents the performance measures adopted for the models evaluation. In addition, the results are sorted regarding lower MAPE for training and test results. Finally, the best models are those that presented lower metrics in the test set. For the STACK approach, regarding results from test set, in case study 1, level-0 is composed of models KNN, LASSO and XGB with SVR in level-1. In case study 2, SVR and XGB compose level-0 with LASSO in level-1. In Table A.2, the models numbered 9, 19, 30, 31, 40, 41, 49, 54 show the same performance. Layer-0 of these is composed of SVR and XGB in all cases, with some variations for base learners combination. This means that including other base learners, in this case study, did not improve the STACK performance. In this aspect, model 9 is chosen for the STACK approach, since its complexity is lower than other configurations.

Evaluating the results by test set, as presented in Tables 7 and 8, it can be stated in general that the methods based on the ensemble approach present good performance for short-term forecasting in the agribusiness area in relation to other models used. According to Qureshi et al. [75], in relation to single models, this is mainly due to the fact that the ensemble approach makes the predictive model more robust. In this context, with regard to the comparison of regression ensembles with different classes of models, the results presented here corroborate with



(a) Case study 1 (b) Case study 2

Fig. 3. Correlation matrix for the prices in case studies 1 (left) and 2 (right).

Table 6
Selected hyperparameters, by grid search, for models trained with LOOCV in each case study (CS).

Model	Hyperparameters	Grid Search	CS 1	CS 2
GBM	Number of boosting iterations	{10, 20, 50, 100, 500, 1000}	1000	500
	Max tree depth	{1, 5, 10}	10	5
	Shrinkage	{0.01, 0.05, 0.1, 0.5}	0.1	0.01
	Minimum terminal node size	{3, 5, 10}	5	3
XGB	Number of boosting iterations	{50, 100, 200, 300}	50	300
	L1 regularization	{0, 0.01, 0.5, 1}	0	0
	L2 regularization	{0, 0.01, 0.5, 1}	0.01	1
	Learning rate	{0.5, 0.3, 0.25, 0.01, 0.125, 0.0625, 0.03125}	0.03125	0.01
RF	Number of randomly selected predictors	{2, 3, 4}	3	4
	Number of trees	500	500	500
SVR	Cost	{0.025, 0.5, 0.75, 1, 1.15, 1.25, 2, 3, 4}	0.025	4
	Kernel	–	Linear	Linear
KNN	Number of neighbors	{5, 7, …, 17, 19}	7	5
	Distance	–	Euclidean	Euclidean
	Aggregation results	–	Average	Average
MLP	Number of hidden units	{2, 3, 4}	3	4
$STACK_1$	Cost	{0.025, 0.5, 0.75, 1, 1.15, 1.25, 2, 3, 4}	0.5	–
	Kernel	–	Linear	-
$STACK_2$	L1 penalty	{0.001, 0.003, 0.005, 0.009}	–	0.001

Table 7
Performance measures models evaluation.

Case	Training set					Test set				
Study	MAPE	RMSE	MSE	MAE	Model	MAPE	RMSE	MSE	MAE	Model
	0.0093	0.0013	0.0001	0.0004	XGB	0.9787	0.0554	0.0031	0.0407	XGB
	0.0229	0.0010	0.0001	0.0008	GBM	0.9855	0.0504	0.0025	0.0408	STACK
	0.5870	0.0270	0.0007	0.0210	MLP	1.0867	0.0617	0.0038	0.0453	GBM
1	0.5973	0.0277	0.0008	0.0213	RF	1.1315	0.0612	0.0037	0.0472	MLP
	1.1202	0.0524	0.0027	0.0401	KNN	1.1549	0.0619	0.0038	0.0482	RF
	1.3050	0.0579	0.0033	0.0468	SVR	1.3138	0.0664	0.0044	0.0546	SVR
	1.3933	0.0611	0.0037	0.0496	STACK	1.3737	0.0758	0.0057	0.0574	KNN
	MAPE	RMSE	MSE	MAE	Model	MAPE	RMSE	MSE	MAE	Model
	0.0262	0.0012	0.0000	0.0009	XGB	0.7394	0.0360	0.0013	0.0265	RF
	0.6387	0.0281	0.0008	0.0211	RF	0.8833	0.0419	0.0018	0.0314	XGB
	0.8061	0.0330	0.0011	0.0265	GBM	0.9626	0.0426	0.0018	0.0342	STACK
2	1.0669	0.0510	0.0026	0.0353	STACK	0.9842	0.0411	0.0017	0.0348	GBM
	1.3146	0.0562	0.0032	0.0437	KNN	1.5135	0.0612	0.0037	0.0536	SVR
	1.5296	0.0697	0.0049	0.0506	SVR	1.7912	0.0717	0.0051	0.0631	MLP
	1.6354	0.0680	0.0046	0.0537	MLP	2.8993	0.1302	0.0170	0.1016	KNN

those published in the literature for results from other areas of knowledge.

Regarding the performance of adopted models, it can be seen in the training phase that the XGB presented a great capacity to learn from the data. Its MAPE is lower, between 146% and 2337.79%, in relation to the ensemble models for case studies 1 and 2, respectively. Regarding the single approach models, this percentage reduction is even higher, being at minimum 6211.83% and 4917.56% lower for case studies 1 and 2. Concerning the RMSE of models, the XGB performance is also better in relation to the others, that is, it had lower observed RMSE values and greater variations are observed when comparing the results of XGB in relation to MLP, SVR and KNN, in both case studies. In this scenario, this allows high-capacity predictions of observations of this approach, with small errors, thus being stable with respect to the time series forecasting [93].

In respect of the results for the test phase, according to the Table 7, the approach based on ensemble learning presented results consistent with minimized errors. In both case studies, GBM, STACK, XGB and RF presented lower performance measures compared to the other approaches. In percentage terms, in case study 1, the XGB and STACK models had similar behaviors. Relating the XGB model with the others models, except STACK, percentage reduction of the performance measures are observed, which range between 11.04%–40.36%, 10.47%–36.82%, 19.35%–83.87% and 11.30%–41.03% for MAPE, RMSE, MSE and MAE respectively, in relation to the other models. When the performance of XGB is evaluated in relation to STACK, GBM and RF, smaller variations are observed compared to XGB in relation to MLP, SVR and KNN. These results demonstrate the generalization capacity and robustness of the XGB approach when applied to forecasts one month ahead of the price paid to the producer of soybean. The XGB and STACK models are competitive, with similar performances. This perhaps may be associated to the fact that in the STACK model, only the XGB, LASSO and KNN models form the level-0. In addition, the good performance of the boosting-based approach is associated with the fact that such a model acts by correcting errors committed by its previous basic models, thus potentially improving prediction accuracy [94].

With regard to the results of case study 2, similar considerations can be made. However, the RF approach is more effective than other ensemble and reference models. If we compare the results of the RF model with STACK, XGB and GBM, the reduction in the performance measures of this model range from 19.46%–33.11%, 14.17%–16.39%, 30.77%–38.46%, 18.49%–31.32% for MAPE, RMSE, MSE and MAE respectively. This reduction is greater when comparing the RF performance against reference models. It can be stated that the RF model reduced the predicted errors, showing to be a suitable tool for short-term forecasting, which should be used as an aid in decision making, since an accurate modeling can be developed [95].

Besides, for both training and test sets, the layers-0 elements number of the STACK approach did not statistically interfere in the predictive performance of the final model ($\chi^2_3 = 0.1232–2.99$, p-value > 0.05), that is, MAPE, RMSE, MSE and MAE did not have statistically significant gains with this variation. This information confirms the hypothesis that for an ensemble modeled in levels, it tends to achieve good results depending on the diversification of the initial layer models, and not according to the number of elements. Alongside this, Tables A.1 and A.2 show that 23.21% and 57.14% of the models show better performance than single models for case studies 1 and 2. These results corroborate with the effectiveness of this methodology.

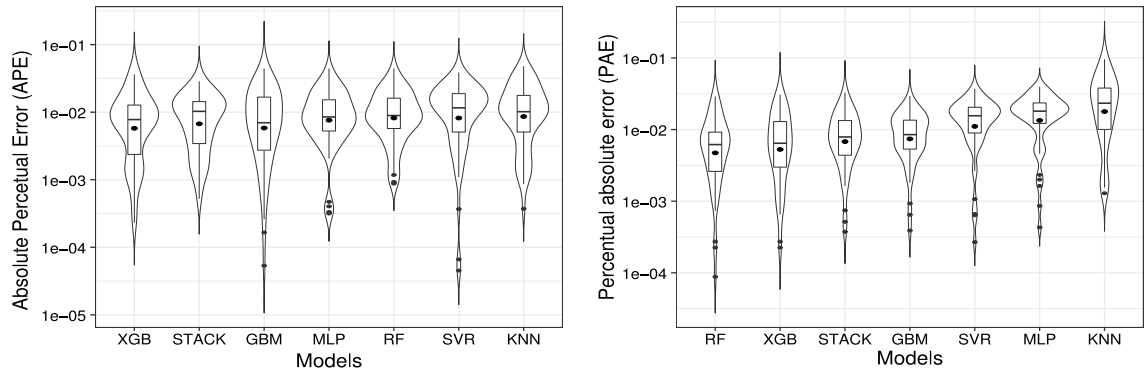
The Fig. 4 presents the violin graph for the APE distribution of each model used for predictions in the test set. The central box represents the box-plot of the APE obtained and the black dot represents the mean APE. It is illustrated that ensemble-based models reduce the APE in relation to other models. In this direction, it is possible to observe that the model (for the test set) with smaller metrics values in Table 7 has an APE with a smaller variation, showing to be stable. Using the Friedman test, it is confirmed that the APE in the test set for adopted models in both case studies differs from each other ($\chi^2_6 = 15.17–84.45$, p-value < 0.05). This suggests that there are models whose observed values for the APE are equal to or smaller than the others. Alongside this, Table 8 presents the Wilcoxon signed rank test (lower tail) results for the evaluation of the APE reduction of the evaluated models, in the test set, given the existence of a significant difference pointed out by the Friedman test in both case studies ($\chi^2_6 = 15.17–84.45$, p-value < 0.05).

For case study 1, it is possible to verify from these results that the APE of the XGB model is inferior to the APE of the SVR and KNN models at the 5% significance level. When comparing the XGB model with other ensemble models, the errors are statistically the same at the 5% level. A similar situation is observed for the case study 2. However, the APE of the RF and XGB models are statistically equal to the 5% level. In general, this indicates that those predicted values are very close to the observed values. On the whole, ensemble models have lower APE than SVR and KNN. An exception is the MLP model and RF with SVR, in which there are errors statistically equal to the ensemble models in case study 1. Faced with this, it is possible to identify the great capacity to learn with the data of this approach, with smaller prediction errors, and smaller differences observed between the ensemble approaches than with the others, indicating the consistency of this perspective [96].

The Figs. 5(a) and 5(b) show the observed and predicted series from the smaller performance measure models reported (for the test set) in Table 7. The best performing model to forecast agricultural commodities prices one month ahead are XGB for case study 1 (soybean price) and RF for case study 2 (wheat price). In this aspect, possible to identify that the behavior of the data is learned by the models, which are able to perform predictions compatible with the observed values. The good performance obtained in the training phase persists in the test stage, thus indicating, that the ensemble methodology is robust in the accomplishment of the developed predictions. This is justified by the fact that machine learning modeling is capable of accommodating nonlinearities, as well as modeling complex relationships between dependent and independent variables, such as the pricing relationships relevant to agribusiness [93].

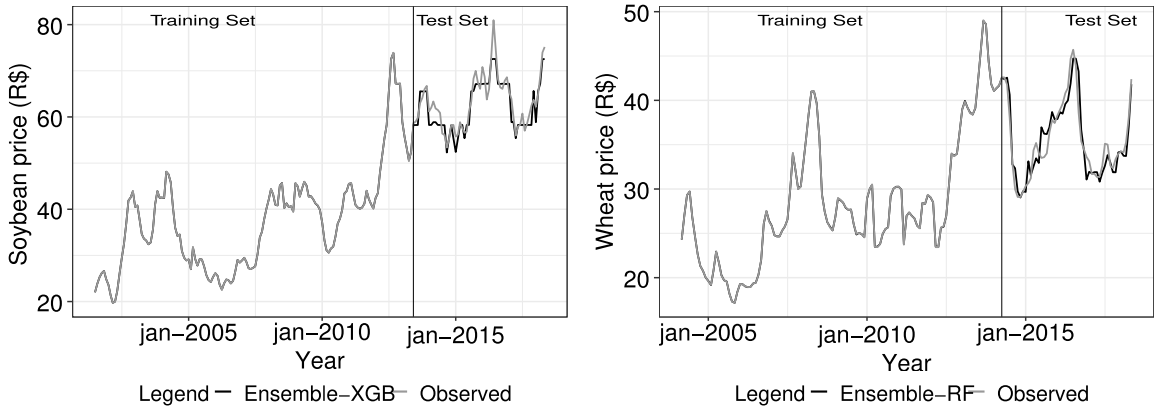
The Fig. 6 illustrates the importance of each variable in the process of predicting the prices paid to producers in case study 1 (left) and case study 2 (right). This score is obtained using the GBM model, regarding results from training set.

For both case studies, the indicator of CEPEA/Esalq-USP (P_1 and P_3) showed high contribution to the pricing of these commodities. Alongside this, the ICP (P_2, P_3 and P_6), FM (P_7 and P_8) and EC (P_5) influence was less intense in both case studies, and these results corroborate with those in Figs. 3(a) and 3(b). It should be noted that the price of ethanol (P_5) was excluded from the set of features, in both cases. However, the results obtained in the new modeling are worse than those presented here, so it was decided to maintain this feature. This confirms that even with a small importance in the formation of prices, the price of ethanol ends up influencing. These results suggest that agricultural commodities prices are extremely affected by reference indicators such as CEPEA/Esalq-USP and prices of the main markets, such as the CBOT. In addition, feature engineering plays an important role in machine learning modeling. Initially this approach was used with the purpose to improve the performance of proposed ensembles. In this context, statistical features based on lags (5 for example)



(a) Case study 1 (b) Case study 2

Fig. 4. Violin plot for the models APE in case studies 1 (left) and 2 (right).



(a) Case study 1 (b) Case study 2

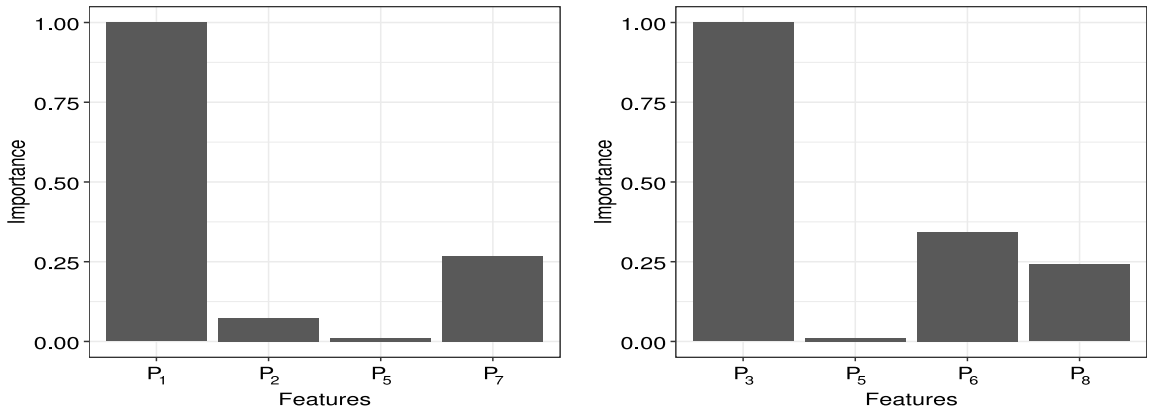
Fig. 5. Observed and predicted prices by best models reported in Table 7, in case studies 1 (left) and 2 (right).

Table 8
Wilcoxon signed rank test statistic (W) (lower tail) and *p*-value for APE comparisons.

Case study 1 - Soybean case								
Model 1 vs Model 2	W	*p*-value	Model 1 vs Model 2	W	*p*-value	Model 1 vs Model 2	W	*p*-value
XGB vs STACK	1590	>0.05	STACK vs GBM	1752	>0.05	GBM vs MLP	1599	>0.05
XGB vs GBM	1698	>0.05	STACK vs MLP	1627	<0.01	GBM vs RF	1569	>0.05
XGB vs MLP	1546	>0.05	STACK vs RF	1611	>0.05	GBM vs KNN	1463	<0.05
XGB vs RF	1508	>0.05	STACK vs KNN	1494	<0.05	GBM vs SVR	1460	<0.01
XGB vs KNN	1362	<0.05	STACK vs SVR	1388	<0.05	MLP vs RF	1698	>0.05
XGB vs SVR	1295	<0.05	RF vs KNN	1603	<0.05	MLP vs KNN	1577	>0.05
KNN vs SVR	1816	>0.05	RF vs SVR	1526	>0.05	MLP vs SVR	1509	<0.05
Case study 2 - Wheat case								
RF vs XGB	1125	>0.05	XGB vs STACK	1112	>0.05	STACK vs GBM	1184	>0.05
RF vs STACK	969	<0.05	XGB vs GBM	1061	>0.05	STACK vs SVR	758	<0.05
RF vs GBM	903	<0.05	XGB vs SVR	730	<0.05	STACK vs MLP	627	<0.05
RF vs SVR	585	<0.05	XGB vs MLP	607	<0.05	STACK vs KNN	586	<0.05
RF vs MLP	500	<0.05	XGB vs KNN	535	<0.05	GBM vs SVR	757	<0.05
RF vs KNN	468	<0.05	SVR vs MLP	1043	>0.05	GBM vs MLP	616	<0.05
MLP vs KNN	967	<0.05	SVR vs KNN	843	<0.05	GBM vs KNN	599	<0.05

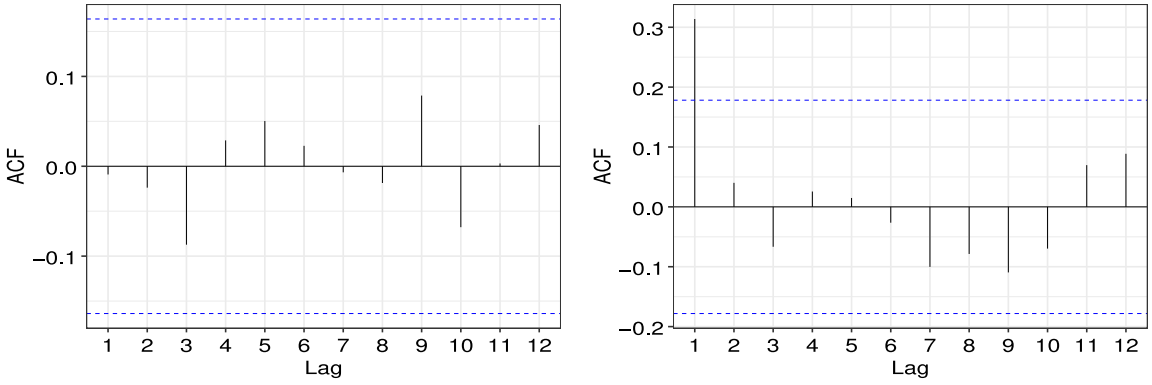
of inputs (average, standard deviation, skewness, kurtosis for five past values) are adopted. Besides, features such as hyperbolic tangent, square root, P_i^2, P_i^3 (predictive variable $i = 1, \ldots, 8$) and several combinations such as sum, rate and difference between variables from Section 3.2, were adopted. In this context, the feature extraction technique (filter based on correlation) was used to extract a features set. However, the improvement is not significant regarding the use of initial features alone (predictive variables of Table 4). Hence, in this context, a model with less components is preferable.

From Fig. 7 it is possible to evaluate the residuals (autocorrelation function) of final models adopted in each case study, from training set results. The Box–Ljung test is applied to the residuals from the final ensemble models to determine whether residuals are random. In both cases, this test shows that the first 12 lag autocorrelations among the residuals are zero ($\chi^2_{12} = 3.87$–19.28, p-value > 0.05), indicating that the residuals are random and that the model provides an adequate fit to the data [97]. These results are supported by Fig. 7, once all lags are between confidence interval extremes.



(a) Case study 1 (b) Case study 2

Fig. 6. Importance of the variables used to forecast prices one month ahead in case studies 1 (left) and 2 (right).



(a) Case study 1 (b) Case study 2

Fig. 7. Autocorrelation to standard residuals of final models in case studies 1 (left) and 2 (right).

7. Conclusions and future research

This paper aimed to compare the predictive performance of the GBM, XGB, RF and STACK regression ensembles, as well as the MLP, SVR and KNN models in two case studies related to agribusiness, namely: case 1, the soybean price and case 2, the wheat price paid to the producer from the state of Parana, for short-term forecasting. Explanatory variables were chosen based on studies found in the literature. Different models were combined, in the stacking approach, to form the level-0 of the meta-modeling. The SRV with a linear kernel and LASSO regression were adopted as meta-learners for case studies 1 and 2. The Friedman and the Wilcoxon signed rank tests (lower tail) were adopted to verify the existence of a difference between the models' APE. Additionally, residual analysis was conducted in order to check the residuals adequacy of the trained model.

Regarding results, to forecast one month ahead, three models can be adopted as follows: XGB and the second, called LASSO-KNN-XGB-SVR STACK for case study 1 and RF for case study 2. In both cases, the Box–Cox transformation ($\lambda = 0$) is applied to predictors and response variables. The results, regarding the test set, allows to ensure that the ensemble approaches perform better than single models, especially the XGB model, to forecast agricultural commodities in the short-term. In addition, the approaches based on boosting for ensemble generation proved to be robust for the problems discussed here. This methodology presented good performance for both case studies, with small prediction errors (MAPE $< 1\%$) and high accuracy.

The information contained in this research is relevant for decision-making aid, as well as in synthesizing the main factors that affect soybean and wheat agricultural commodities prices. Based on the results presented here, small producers and managers of cooperatives are supported by concrete information that allows them to choose when to deliver their products to the wholesaler, as well as to draw market strategies that maximize their profits. Therefore, the findings in this paper allow those responsible to carry out deliberations with reduced risk and maximized assertiveness. In this paper there are no considerations about weather changes or political aspects for agribusiness area. According to Thompson et al. [98], these aspects are able to modestly increase the price volatility. In this context, the prediction assertiveness can be reduced. Since, in this moment, these factors are not considered, this is assumed to be a study limitation. Nevertheless, even with satisfactory results, the decision-making process must be parsimonious, since agricultural commodities prices are volatile and tend to show large fluctuations.

As future research, it is intended to: (i) Design experiments to define the hyperparameters of the models, (ii) use other predictive variables as well as weather and political factors, (iii) adopt decomposition techniques, (iii) perform cross-correlation analysis for the features and (iv) optimize hyperparameters using metaheuristics.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2019.105837.

Acknowledgments

The authors would like to thank National Council of Scientific and Technologic Development of Brazil — CNPq (Grants number: 303908/2015-7-PQ, 204893/2017-8 and 404659/2016-0-Univ) for its financial support of this work. Furthermore, the authors wish to thank the Editor and anonymous referees for their constructive comments and recommendations, which have significantly improved the presentation of this paper.

Appendix. Performance measures to evaluate the STACK models used in case studies 1 and 2 to forecast agricultural commodities prices one month ahead.

See Tables A.1 and A.2.

Table A.1
Performance measures to evaluate the STACK models used to forecast prices one month ahead when in case study 1 (soybean price) the meta-learner is SVR with linear kernel.

	Training set				Test set			
Model	MAPE	MAE	RMSE	MSE	MAPE	MAE	RMSE	MSE
1	1.2894	0.0461	0.0573	0.0033	1.1392	0.0471	0.0574	0.0033
2	3.3279	0.1174	0.1444	0.0209	6.2475	0.2583	0.2721	0.0741
3	2.9525	0.1058	0.1302	0.0170	5.9080	0.2439	0.2648	0.0701
4	3.1042	0.1111	0.1361	0.0185	6.1873	0.2555	0.2772	0.0769
5	2.3028	0.0821	0.0977	0.0096	4.2165	0.1740	0.1963	0.0385
6	1.3585	0.0484	0.0599	0.0036	1.0158	0.0420	0.0519	0.0027
7	1.3318	0.0475	0.0590	0.0035	1.0487	0.0434	0.0534	0.0028
8	1.3313	0.0475	0.0590	0.0035	1.0655	0.0441	0.0542	0.0029
9	1.3247	0.0473	0.0586	0.0034	1.0330	0.0427	0.0527	0.0028
10	1.3000	0.0471	0.0585	0.0034	1.5045	0.0621	0.0715	0.0051
11	1.2777	0.0463	0.0575	0.0033	1.4584	0.0602	0.0695	0.0048
12	1.5507	0.0558	0.0679	0.0046	1.5039	0.0621	0.0709	0.0050
13	1.1923	0.0430	0.0530	0.0028	1.9253	0.0795	0.0929	0.0086
14	1.1841	0.0428	0.0538	0.0029	1.7773	0.0734	0.0867	0.0075
15	1.1667	0.0421	0.0540	0.0029	1.6844	0.0696	0.0823	0.0068
16	1.1361	0.0408	0.0517	0.0027	1.2912	0.0534	0.0642	0.0041
17	1.1560	0.0415	0.0520	0.0027	1.2666	0.0524	0.0629	0.0040
18	1.1554	0.0415	0.0519	0.0027	1.2903	0.0533	0.0639	0.0041
19	1.2498	0.0447	0.0556	0.0031	1.1168	0.0462	0.0559	0.0031
20	2.6154	0.0948	0.1143	0.0131	5.1149	0.2110	0.2332	0.0544
21	2.7047	0.0980	0.1169	0.0137	5.2910	0.2184	0.2412	0.0582
22	2.4579	0.0875	0.1048	0.0110	4.5242	0.1867	0.2091	0.0437
23	3.1107	0.1114	0.1364	0.0186	6.1972	0.2559	0.2776	0.0771
24	2.6587	0.0952	0.1166	0.0136	5.2390	0.2162	0.2408	0.0580
25	2.7370	0.0981	0.1196	0.0143	5.4138	0.2234	0.2488	0.0619
26	1.2394	0.0444	0.0561	0.0031	1.1732	0.0485	0.0591	0.0035
27	1.2354	0.0442	0.0560	0.0031	1.1649	0.0482	0.0588	0.0035
28	1.3933	0.0496	0.0611	0.0037	0.9855	0.0408	0.0504	0.0025
29	1.3321	0.0476	0.0591	0.0035	1.0543	0.0436	0.0536	0.0029
30	1.3589	0.0485	0.0601	0.0036	1.0247	0.0424	0.0521	0.0027
31	1.3594	0.0485	0.0601	0.0036	1.0238	0.0424	0.0521	0.0027
32	1.2720	0.0460	0.0574	0.0033	1.4484	0.0598	0.0690	0.0048
33	1.3761	0.0496	0.0615	0.0038	1.3105	0.0542	0.0645	0.0042
34	1.3972	0.0503	0.0620	0.0038	1.2306	0.0509	0.0611	0.0037
35	1.1202	0.0406	0.0525	0.0028	1.6250	0.0671	0.0795	0.0063
36	1.1432	0.0410	0.0520	0.0027	1.2740	0.0527	0.0634	0.0040
37	1.1430	0.0410	0.0522	0.0027	1.2934	0.0535	0.0643	0.0041
38	1.1367	0.0408	0.0515	0.0026	1.2241	0.0506	0.0609	0.0037
39	1.1571	0.0415	0.0520	0.0027	1.2549	0.0519	0.0624	0.0039
40	1.1778	0.0422	0.0527	0.0028	1.1713	0.0484	0.0584	0.0034
41	1.1797	0.0423	0.0528	0.0028	1.1919	0.0493	0.0594	0.0035
42	2.7551	0.0995	0.1191	0.0142	5.4045	0.2231	0.2444	0.0598
43	2.3772	0.0861	0.1031	0.0106	4.6500	0.1918	0.2152	0.0463
44	2.5146	0.0911	0.1081	0.0117	4.9064	0.2024	0.2265	0.0513
45	2.6813	0.0960	0.1173	0.0138	5.2847	0.2181	0.2429	0.0590
46	1.2355	0.0442	0.0560	0.0031	1.1559	0.0478	0.0584	0.0034
47	1.2562	0.0450	0.0570	0.0032	1.1148	0.0461	0.0562	0.0032
48	1.2575	0.0450	0.0572	0.0033	1.0934	0.0452	0.0553	0.0031
49	1.3576	0.0485	0.0601	0.0036	1.0310	0.0427	0.0524	0.0027
50	1.3994	0.0504	0.0622	0.0039	1.2546	0.0519	0.0617	0.0038
51	1.1417	0.0410	0.0520	0.0027	1.2716	0.0526	0.0634	0.0040
52	1.1380	0.0408	0.0521	0.0027	1.2497	0.0517	0.0618	0.0038
53	1.1350	0.0407	0.0521	0.0027	1.2598	0.0521	0.0623	0.0039
54	1.1827	0.0424	0.0529	0.0028	1.1674	0.0483	0.0582	0.0034
55	2.4781	0.0897	0.1068	0.0114	4.8312	0.1993	0.2233	0.0499
56	1.2552	0.0449	0.0571	0.0033	1.1027	0.0456	0.0557	0.0031

Table A.2
Performance measures to evaluate the STACK models used to forecast prices one month ahead when in case study 2 (wheat price) the meta-learner is LASSO.

	Training set				Test set			
Model	MAPE	MAE	RMSE	MSE	MAPE	MAE	RMSE	MSE
1	1.3855	0.0457	0.0627	0.0039	1.2322	0.0438	0.0506	0.0026
2	1.6542	0.0544	0.0649	0.0042	3.1348	0.1105	0.1267	0.0160
3	1.0232	0.0342	0.0425	0.0018	1.8729	0.0668	0.0750	0.0056
4	0.9627	0.0322	0.0400	0.0016	1.5587	0.0558	0.0653	0.0043
5	0.9077	0.0299	0.0358	0.0013	1.5006	0.0534	0.0598	0.0036
6	1.6660	0.0549	0.0743	0.0055	2.1257	0.0748	0.0880	0.0078
7	1.2781	0.0421	0.0569	0.0032	1.3817	0.0491	0.0571	0.0033
8	1.2488	0.0413	0.0566	0.0032	1.1976	0.0427	0.0510	0.0026
9	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
10	1.3719	0.0455	0.0555	0.0031	2.7900	0.0987	0.1103	0.0122
11	1.3800	0.0458	0.0561	0.0031	2.7381	0.0968	0.1081	0.0117
12	1.3977	0.0462	0.0552	0.0030	3.0613	0.1076	0.1288	0.0166
13	1.0333	0.0347	0.0439	0.0019	1.9334	0.0689	0.0778	0.0061
14	0.9277	0.0312	0.0396	0.0016	1.7843	0.0636	0.0725	0.0053
15	0.8617	0.0291	0.0376	0.0014	1.4266	0.0512	0.0626	0.0039
16	1.4467	0.0476	0.0644	0.0042	1.5027	0.0532	0.0611	0.0037
17	1.3313	0.0438	0.0596	0.0035	1.3056	0.0464	0.0536	0.0029
18	1.3380	0.0441	0.0603	0.0036	1.2345	0.0439	0.0509	0.0026
19	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
20	1.4014	0.0463	0.0558	0.0031	2.7344	0.0967	0.1081	0.0117
21	1.4409	0.0476	0.0573	0.0033	2.7499	0.0972	0.1088	0.0118
22	1.4436	0.0476	0.0568	0.0032	2.9905	0.1053	0.1234	0.0152
23	1.0276	0.0343	0.0427	0.0018	1.8765	0.0669	0.0752	0.0056
24	0.9545	0.0320	0.0401	0.0016	1.8028	0.0643	0.0727	0.0053
25	0.9309	0.0312	0.0390	0.0015	1.5289	0.0547	0.0643	0.0041
26	1.3200	0.0435	0.0585	0.0034	1.5010	0.0533	0.0613	0.0038
27	1.3249	0.0437	0.0594	0.0035	1.4382	0.0511	0.0587	0.0035
28	1.0779	0.0357	0.0514	0.0026	1.0265	0.0364	0.0449	0.0020
29	1.2781	0.0421	0.0569	0.0032	1.3817	0.0491	0.0571	0.0033
30	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
31	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
32	1.3719	0.0455	0.0555	0.0031	2.7900	0.0987	0.1103	0.0122
33	1.2444	0.0413	0.0502	0.0025	2.6555	0.0938	0.1070	0.0114
34	1.2632	0.0419	0.0510	0.0026	2.6931	0.0950	0.1093	0.0119
35	0.9277	0.0312	0.0396	0.0016	1.7843	0.0636	0.0725	0.0053
36	1.3770	0.0453	0.0613	0.0038	1.4410	0.0511	0.0586	0.0034
37	1.4117	0.0465	0.0630	0.0040	1.4580	0.0517	0.0592	0.0035
38	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020
39	1.3313	0.0438	0.0596	0.0035	1.3056	0.0464	0.0536	0.0029
40	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
41	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
42	1.4014	0.0463	0.0558	0.0031	2.7344	0.0967	0.1081	0.0117
43	1.2506	0.0415	0.0503	0.0025	2.6566	0.0938	0.1069	0.0114
44	1.3118	0.0434	0.0525	0.0028	2.7141	0.0958	0.1096	0.0120
45	0.9545	0.0320	0.0401	0.0016	1.8028	0.0643	0.0727	0.0053
46	1.3200	0.0435	0.0585	0.0034	1.5010	0.0533	0.0613	0.0038
47	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020
48	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020
49	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
50	1.2444	0.0413	0.0502	0.0025	2.6555	0.0938	0.1070	0.0114
51	1.3770	0.0453	0.0613	0.0038	1.4410	0.0511	0.0586	0.0034
52	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020
53	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020
54	1.0669	0.0353	0.0510	0.0026	0.9626	0.0342	0.0426	0.0018
55	1.2506	0.0415	0.0503	0.0025	2.6566	0.0938	0.1069	0.0114
56	1.0767	0.0356	0.0514	0.0026	1.0164	0.0360	0.0445	0.0020

References

[1] Cepea, Centro de Estudos Avançados em Economia Aplicada. PIB do Agronegócio Brasileiro, 2018, https://www.cepea.esalq.usp.br/upload/kceditor/files/Relatorio (Access in 20 ago. 2018) (in Portuguese).

[2] Ipardes, Instituto Paranaense de Desenvolvimento Econômico e Social. Economia do Paraná cresceu 2,5% - mais que o dobro do Brasil, 2018, http://www.ipardes.gov.br/index.php?pg_conteudo=1&cod_noticia=925 (Access in 02 ago. 2018) (in Portuguese).

[3] Brasil, Ministério da Agricultura, Pecuária e Abastecimento. Projeções do agronegócio 2017/2018 a 2027/2028: Projeções de Longo Prazo, 2018, http://www.agricultura.gov.br/assuntos/politica-agricola/todas-publicacoes-de-politica-agricola/projecoes-do-agronegocio/banner_site-03-03-1.png/view (Access in 20 ago. 2018) (in Portuguese).

[4] S. Haykin, Neural Networks: A Comprehensive Foundation, International edition, Prentice Hall, ISBN: 9780132733502, 1999, https://dl.acm.org/citation.cfm?id=541500.

[5] T.G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, Springer, ISBN: 978-3-540-45014-6, 2000, pp. 1–15, Heidelberg, Germany, https://link.springer.com/chapter/10.1007/3-540-45014-9_1.

[6] Y. Ren, P. Suganthan, N. Srikanth, Ensemble methods for wind and solar power forecasting—A state-of-the-art review, Renew. Sustain. Energy Rev. 50 (2015) 82–91, http://dx.doi.org/10.1016/j.rser.2015.04.081.

[7] F. Divina, A. Gilson, F. Goméz-Vela, M. García Torres, J.F. Torres, Stacking ensemble learning for short-term electricity consumption forecasting, Energies 11 (4) (2018) http://dx.doi.org/10.3390/en11040949.

[8] E. Soares, P. Costa, B. Costa, D. Leite, Ensemble of evolving data clouds and fuzzy models for weather time series prediction, Appl. Soft Comput. 64 (2018) 445–453, http://dx.doi.org/10.1016/j.asoc.2017.12.032.

[9] J.a. Mendes-Moreira, C. Soares, A.M. Jorge, J.F.D. Sousa, Ensemble approaches for regression: A survey, ACM Comput. Surv. 45 (1) (2012) 10:1–10:40, http://dx.doi.org/10.1145/2379776.2379786.

[10] A. Torres-Barrán, Álvaro Alonso, J.R. Dorronsoro, Regression tree ensembles for wind energy and solar radiation prediction, Neurocomputing 326–327 (2019) 151–160, http://dx.doi.org/10.1016/j.neucom.2017.05.104.

[11] X.-d. Zhang, A. Li, R. Pan, Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine, Appl. Soft Comput. 49 (2016) 385–398, http://dx.doi.org/10.1016/j.asoc.2016.08.026.

[12] C. Krauss, X.A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500, European J. Oper. Res. 259 (2) (2017) 689–702, http://dx.doi.org/10.1016/j.ejor.2016.10.031.

[13] B. Weng, W. Martinez, Y.-T. Tsai, C. Li, L. Lu, J.R. Barth, F.M. Megahed, Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models, Appl. Soft Comput. 71 (2018) 685–697, http://dx.doi.org/10.1016/j.asoc.2018.07.024.

[14] A. Peimankar, S.J. Weddell, T. Jalal, A.C. Lapthorn, Multi-objective ensemble forecasting with an application to power transformers, Appl. Soft Comput. 68 (2018) 233–248, http://dx.doi.org/10.1016/j.asoc.2018.03.042.

[15] F.J. Fabozzi, Handbook of Finance, Financial Markets and Instruments, vol. 1, John Wiley & Sons, ISBN: 978-0-470-07814-3, 2008, p. 852, http://dx.doi.org/10.1002/9780470404324.

[16] K. He, K.K. Lai, J. Yen, Ensemble forecasting of value at risk via multi resolution analysis based methodology in metals markets, Expert Syst. Appl. 39 (4) (2012) 4258–4267, http://dx.doi.org/10.1016/j.eswa.2011.09.108.

[17] L.A. Gabralla, H. Mahersia, A. Abraham, Ensemble neurocomputing based oil price prediction, in: A. Abraham, P. Krömer, V. Snasel (Eds.), Afro-European Conference for Industrial Advancement, Springer International Publishing, Cham, ISBN: 978-3-319-13572-4, 2015, pp. 293–302, http://dx.doi.org/10.1007/978-3-319-13572-4_24.

[18] C. Pierdzioch, M. Risse, S. Rohloff, Forecasting gold-price fluctuations: a real-time boosting approach, Appl. Econ. Lett. 22 (1) (2015) 46–50, http://dx.doi.org/10.1080/13504851.2014.925040.

[19] C. Pierdzioch, M. Risse, S. Rohloff, A boosting approach to forecasting the volatility of gold-price fluctuations under flexible loss, Resour. Policy 47 (2016) 95–107, http://dx.doi.org/10.1016/j.resourpol.2016.01.003.

[20] L. Yu, W. Dai, L. Tang, A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting, Eng. Appl. Artif. Intell. 47 (2016) 110–121, http://dx.doi.org/10.1016/j.engappai.2015.04.016.

[21] L. Yu, H. Xu, L. Tang, LSSVR ensemble learning with uncertain parameters for crude oil price forecasting, Appl. Soft Comput. 56 (2017) 692–701, http://dx.doi.org/10.1016/j.asoc.2016.09.023.

[22] Y. Zhao, J. Li, L. Yu, A deep learning ensemble approach for crude oil price forecasting, Energy Econ. 66 (2017) 9–16, http://dx.doi.org/10.1016/j.eneco.2017.05.023.

[23] D. Wang, C. Yue, S. Wei, J. Lv, Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting, Algorithms 10 (3) (2017) http://dx.doi.org/10.3390/a10030108.

[24] K. Yang, F. Tian, L. Chen, S. Li, Realized volatility forecast of agricultural futures using the HAR models with bagging and combination approaches, Int. Rev. Econ. Finance 49 (2017) 276–291, http://dx.doi.org/10.1016/j.iref.2017.01.030.

[25] T. Xiong, L. Chongguang, B. Yukun, An improved EEMD-based hybrid approach for the short-term forecasting of hog price in China, Agricult. Econ. 63 (3) (2017) http://dx.doi.org/10.17221/268/2015-AGRICECON.

[26] C. Pierdzioch, M. Risse, Forecasting precious metal returns with multivariate random forests, Empir. Econom. (2018) 1–18, http://dx.doi.org/10.1007/s00181-018-1558-9.

[27] L. Tang, Y. Wu, L. Yu, A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting, Appl. Soft Comput. 70 (2018) 1097–1108, http://dx.doi.org/10.1016/j.asoc.2017.02.013.

[28] L. Wang, W. Duan, D. Qu, S. Wang, What matters for global food price volatility? Empir. Econom. 54 (4) (2018) 1549–1572, http://dx.doi.org/10.1007/s00181-017-1311-9.

[29] M. Bonato, R. Demirer, R. Gupta, C. Pierdzioch, Gold futures returns and realized moments: A forecasting experiment using a quantile-boosting approach, Resour. Policy 57 (2018) 196–212, http://dx.doi.org/10.1016/j.resourpol.2018.03.004.

[30] Y. Ding, A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting, Energy 154 (2018) 328–336, http://dx.doi.org/10.1016/j.energy.2018.04.133.

[31] F. Anifowose, A. Khoukhi, A. Abdulraheem, Investigating the effect of training–testing data stratification on the performance of soft computing techniques: an experimental study, J. Exp. Theor. Artif. Intell. 29 (3) (2017) 517–535, http://dx.doi.org/10.1080/0952813X.2016.1198936.

[32] V. Cerqueira, L. Torgo, J. Smailović, I. Mozetič, A comparative study of performance estimation methods for time series forecasting, in: 2017 IEEE International Conference on Data Science and Advanced Analytics, DSAA, 2017, pp. 529–538, http://dx.doi.org/10.1109/DSAA.2017.7, Toky, Japan.

[33] R. Trostle, Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food Commodity Prices, USDA Economic Research Service, Report WRS-0801, 2008, http://dx.doi.org/10.2172/1218372.

[34] D.A. Bini, M.D. Canever, A.A. Denardim, Correlação e causalidade entre os preços de commodities e energia, Nova Econ. 25 (2015) 143–160, http://dx.doi.org/10.1590/0103-6351/1985, (in Portuguese).

[35] A. Fernandez-Perez, B. Frijns, A. Tourani-Rad, Contemporaneous interactions among fuel, biofuel and agricultural commodities, Energy Econ. 58 (2016) 1–10, http://dx.doi.org/10.1016/j.eneco.2016.05.014.

[36] A. Paris, On the link between oil and agricultural commodity prices: Do biofuels matter? Int. Econ. 155 (2018) 48–60, http://dx.doi.org/10.1016/j.inteco.2017.12.003.

[37] C.E. Caldarelli, M.R.P. Bacchi, Fatores de influência no preço do milho no Brasil, Nova Econ. 22 (2012) 141–164, http://dx.doi.org/10.1590/S0103-63512012000100005, (in Portuguese).

[38] L.R.A. Alves, C.E.a.L. Cardoso, F.I. Felipe, M.T. Campion, Causalidade e transmissão entre preços de mandioca, trigo, milho e seus derivados no Paraná, Rev. Econ. Agronegócio 4 (3) (2015) http://dx.doi.org/10.25070/rea.v4i3.85, (in Portuguese).

[39] E. Minga, A.F. Alves, J.L. Parré, Especulação afeta o preço das commodities agrícolas? Rev. Espac. 37 (20) (2016) (in Portuguese).

[40] V. Bodart, B. Candelon, J.-F. Carpantier, Real exchanges rates, commodity prices and structural factors in developing countries, J. Int. Money Finance 51 (2015) 264–284, http://dx.doi.org/10.1016/j.jimonfin.2014.11.021.

[41] J. Baffes, T. Haniotis, What explains agricultural price movements? J. Agric. Econ. 67 (3) (2016) 706–721, http://dx.doi.org/10.1111/1477-9552.12172.

[42] E.S. Pereira, J.R. Silva, S.F. Maia, Os efeitos da taxa de câmbio e dos preços do petróleo nos preços internacionais das commodities brasileiras, Rev. Econ. Nordeste 48 (1) (2017) 131–144 (in Portuguese).

[43] B.M.M. Moraes, R. Bender Filho, K.M. Vieira, P.S. Ceretta, Análise de causalidade de Preços no mercado internacional da soja: O caso do Brasil, Argentina e Estados Unidos, Desenvolv. Questão 14 (34) (2016) 301–319 (in Portuguese).

[44] J.D.S. Rezende, O.J. de Oliveira Neto, K.A. Silva, Volatilidade e transmissão dos preços internacionais do trigo para os preços domésticos do trigo e derivados no Brasil, Future Stud. Res. J.: Trends Strateg. 10 (1) (2018) 132–159 (in Portuguese).

[45] R. McTaggart, G. Daroczi, C.L. Leung, Quandl: API wrapper for Quandl.com, 2016, https://CRAN.R-project.org/package=Quandl, R package version 2.8.0.

[46] W.J.P. Villanueva, Comitê de máquinas em previsão em séries temporais, Mestrado em Engenharia Elétrica e Computação, Faculdade de Engenharia Elétrica e de Computação, Campinas - São Paulo, 2006, (in Portuguese).

[47] H. Allende, C. Valle, Ensemble methods for time series forecasting, in: Claudio Moraga: A Passion for Multi-Valued Logic and Soft Computing, vol. 349, Springer International Publishing, ISBN: 978-3-319-48317-7, 2017, pp. 217–232, http://dx.doi.org/10.1007/978-3-319-48317-7_13, chap. 13.

[48] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, http://dx.doi.org/10.1023/A:1018054314350.

[49] H. Erdal, İlhami Karahanoğlu, Bagging ensemble models for bank profitability: An emprical research on Turkish development and investment banks, Appl. Soft Comput. 49 (2016) 861–867, http://dx.doi.org/10.1016/j.asoc.2016.09.010.

[50] S. Hamze-Ziabari, T. Bakhshpoori, Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5 and CART algorithms, Appl. Soft Comput. 68 (2018) 147–161, http://dx.doi.org/10.1016/j.asoc.2018.03.052.

[51] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[52] M. Thakur, D. Kumar, A hybrid financial trading support system using multi-category classifiers and random forest, Appl. Soft Comput. 67 (2018) 337–349, http://dx.doi.org/10.1016/j.asoc.2018.03.006.

[53] D. Assouline, N. Mohajeri, J.-L. Scartezzini, Large-scale rooftop solar photovoltaic technical potential estimation using random forests, Appl. Energy 217 (2018) 189–211, http://dx.doi.org/10.1016/j.apenergy.2018.02.118.

[54] H. He, W. Zhang, S. Zhang, A novel ensemble method for credit scoring: Adaption of different imbalance ratios, Expert Syst. Appl. 98 (2018) 105–117, http://dx.doi.org/10.1016/j.eswa.2018.01.012.

[55] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227, http://dx.doi.org/10.1007/BF00116037.

[56] J.H. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Statist. 29 (5) (2001) 1189–1232, http://www.jstor.org/stable/2699986.

[57] C. Persson, P. Bacher, T. Shiga, H. Madsen, Multi-site solar power forecasting using gradient boosted regression trees, Sol. Energy 150 (2017) 423–436, http://dx.doi.org/10.1016/j.solener.2017.04.066.

[58] S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, Energy Build. 158 (2018) 1533–1543, http://dx.doi.org/10.1016/j.enbuild.2017.11.039.

[59] G. Ridgeway, gbm: generalized boosted regression models, 2017, https://CRAN.R-project.org/package=gbm, R package version 2.1.3.

[60] C. Ding, X.J. Cao, P. Naess, Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo, Transp. Res. A 110 (2018) 107–117, http://dx.doi.org/10.1016/j.tra.2018.02.009.

[61] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, ISBN: 978-1-4503-4232-2, 2016, pp. 785–794, http://dx.doi.org/10.1145/2939672.2939785.

[62] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, Xgboost: Extreme gradient boosting, 2017, https://CRAN.R-project.org/package=xgboost R package version 0.6-4.

[63] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, X. Niu, Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning, Electron. Commer. Res. Appl. 31 (2018) 24–39, http://dx.doi.org/10.1016/j.elerap.2018.08.002.

[64] P. Carmona, F. Climent, A. Momparler, Predicting failure in the U.S. banking sector: An extreme gradient boosting approach, Int. Rev. Econ. Finance (2018) http://dx.doi.org/10.1016/j.iref.2018.03.008.

[65] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, seventh ed., Springer-Verlag New York, 2017, p. 426, http://dx.doi.org/10.1007/978-1-4614-7138-7.

[66] D.H. Wolpert, Stacked generalization, Neural Netw. 5 (2) (1992) 241–259, http://dx.doi.org/10.1016/S0893-6080(05)80023-1.

[67] M.J. Van der Laan, E.C. Polley, A.E. Hubbard, Super learner, Stat. Appl. Genet. Mol. Biol. 6 (1) (2007) http://dx.doi.org/10.2202/1544-6115.1309.

[68] E. Shamaei, M. Kaedi, Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions, Appl. Soft Comput. 45 (2016) 187–196, http://dx.doi.org/10.1016/j.asoc.2016.03.009.

[69] G. Serbes, B.E. Sakar, H.O. Gulcur, N. Aydin, An emboli detection system based on dual tree complex wavelet transform and ensemble learning, Appl. Soft Comput. 37 (2015) 87–94, http://dx.doi.org/10.1016/j.asoc.2015.08.015.

[70] A. Petropoulos, S.P. Chatzis, V. Siakoulis, N. Vlachogiannakis, A stacked generalization system for automated forex portfolio trading, Expert Syst. Appl. 90 (2017) 290–302, http://dx.doi.org/10.1016/j.eswa.2017.08.011.

[71] A. Pernía-Espinoza, J. Fernandez-Ceniceros, J. Antonanzas, R. Urraca, F.M. de Pison, Stacking ensemble with parsimonious base models to improve generalization capability in the characterization of steel bolted components, Appl. Soft Comput. 70 (2018) 737–750, http://dx.doi.org/10.1016/j.asoc.2018.06.005.

[72] V. Cerqueira, L. Torgo, M. Oliveira, B. Pfahringer, Dynamic and heterogeneous ensembles for time series forecasting, in: 2017 IEEE International Conference on Data Science and Advanced Analytics, DSAA, 2017, pp. 242–251, http://dx.doi.org/10.1109/DSAA.2017.26, Tokyo, Japan.

[73] V. Cerqueira, L. Torgo, F. Pinto, C. Soares, Arbitrated ensemble for time series forecasting, in: M. Ceci, J. Hollmén, L. Todorovski, C. Vens, S. Džeroski (Eds.), Machine Learning and Knowledge Discovery in Databases, Springer International Publishing, Skopje, Macedonia, ISBN: 978-3-319-71246-8, 2017, pp. 478–494, http://dx.doi.org/10.1007/978-3-319-71246-8_29.

[74] F. Anifowose, J. Labadin, A. Abdulraheem, Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines, Appl. Soft Comput. 26 (2015) 483–496, http://dx.doi.org/10.1016/j.asoc.2014.10.017.

[75] A.S. Qureshi, A. Khan, A. Zameer, A. Usman, Wind power prediction using deep neural network based meta regression and transfer learning, Appl. Soft Comput. 58 (2017) 742–755, http://dx.doi.org/10.1016/j.asoc.2017.05.031.

[76] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1) (1991) 37–66, http://dx.doi.org/10.1007/BF00153759.

[77] P.R.G. David E. Rumelhart, Parallel Distributed Processing, Vol. 1: Foundations, vol. 1, A Bradford Book, 1987, ISBN: 026268053X, 9780262680530.

[78] S. Mabu, M. Obayashi, T. Kuremoto, Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems, Appl. Soft Comput. 36 (2015) 357–367, http://dx.doi.org/10.1016/j.asoc.2015.07.020.

[79] N. Messikh, S. Bousba, N. Bougdah, The use of a multilayer perceptron (MLP) for modelling the phenol removal by emulsion liquid membrane, J. Environ. Chem. Eng. 5 (4) (2017) 3483–3489, http://dx.doi.org/10.1016/j.jece.2017.06.053.

[80] H. Drucker, C.J.C. Burges, L. Kaufman, A.J. Smola, V. Vapnik, Support vector regression machines, in: M. Mozer, M.I. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems 9, MIT Press, 1997, pp. 155–161, http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf.

[81] R. Chen, C.-Y. Liang, W.-C. Hong, D.-X. Gu, Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, Appl. Soft Comput. 26 (2015) 435–443, http://dx.doi.org/10.1016/j.asoc.2014.10.022.

[82] M. Kuhn, K. Johnson, Applied Predictive Modeling, first ed., Springer-Verlag New York, 2013, http://dx.doi.org/10.1007/978-1-4614-6849-3, ISBN: 1461468485, 978-1-4614-6848-6, 978-1-4614-6849-3.

[83] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2) (1964) 211–243, http://dx.doi.org/10.1111/j.2517-6161.1964.tb00553.x.

[84] J. Wang, R. Hou, C. Wang, L. Shen, Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting, Appl. Soft Comput. 49 (2016) 164–178, http://dx.doi.org/10.1016/j.asoc.2016.07.024.

[85] P. Shine, M. Murphy, J. Upton, T. Scully, Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms, Comput. Electron. Agric. 150 (2018) 74–87, http://dx.doi.org/10.1016/j.compag.2018.03.023.

[86] B. Weng, L. Lu, X. Wang, F.M. Megahed, W. Martinez, Predicting short-term stock prices using ensemble methods and online data sources, Expert Syst. Appl. 112 (2018) 258–273, http://dx.doi.org/10.1016/j.eswa.2018.06.016.

[87] G. Athanasopoulos, R.J. Hyndman, Forecasting: Principles and Practice, second ed., OTexts, ISBN: 9780987507112, 2018, https://otexts.org/fpp2/ (Access in 06 sep. 2019).

[88] C. Bergmeir, R.J. Hyndman, B. Koo, A note on the validity of cross-validation for evaluating autoregressive time series prediction, Comput. Statist. Data Anal. 120 (2018) 70–83, http://dx.doi.org/10.1016/j.csda.2017.11.003.

[89] B.E. Flores, The utilization of the wilcoxon test to compare forecasting methods: A note, Int. J. Forecast. 5 (4) (1989) 529–535, http://dx.doi.org/10.1016/0169-2070(89)90008-3.

[90] B. Kedem, E. Slud, On goodness of fit of time series models: An application of higher order crossings, Biometrika 68 (2) (1981) 551–556, http://dx.doi.org/10.2307/2335604.

[91] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018, http://www.R-project.org/.

[92] M. Kuhn, Building predictive models in r using the caret package, J. Stat. Softw. Artic. 28 (5) (2008) 1–26, http://dx.doi.org/10.18637/jss.v028.i05.

[93] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of support Vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China, Energy Convers. Manage. 164 (2018) 102–111, http://dx.doi.org/10.1016/j.enconman.2018.02.087.

[94] Y. Zhang, A. Haghani, A gradient boosting method to improve travel time prediction, Transp. Res. C 58 (2015) 308–324, http://dx.doi.org/10.1016/j.trc.2015.02.019.

[95] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, S. Shearer, Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield, Comput. Electron. Agric. 153 (2018) 213–225, http://dx.doi.org/10.1016/j.compag.2018.07.016.

[96] H.T. Pedro, C.F. Coimbra, M. David, P. Lauret, Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts, Renew. Energy 123 (2018) 191–203, http://dx.doi.org/10.1016/j.renene.2018.02.006.

[97] P.A. Morettin, C. Toloi, Análise de séries temporais, second ed., Blucher, 2006, p. 564, (in Portuguese).

[98] W. Thompson, Y. Lu, S. Gerlt, X. Yang, J.E. Campbell, L.M. Kueppers, M.A. Snyder, Automatic responses of crop stocks and policies buffer climate change effects on crop markets and price volatility, Ecol. Econom. 152 (2018) 98–105, http://dx.doi.org/10.1016/j.ecolecon.2018.04.015.