

Article

Short-Term Forecasting for Energy Consumption through Stacking Heterogeneous Ensemble Learning Model

Mergani A. Khairalla ^{1,2,*} , Xu Ning ¹, Nashat T. AL-Jallad ¹ and Musaab O. El-Faroug ³ 

¹ School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China; xuning@whut.edu.cn (X.N.); jallad@whut.edu.cn (N.T.A.-J.)

² School of Science and Technology, Nile Valley University, Atbara 346, Sudan

³ Faculty of Engineering, Elimam Elmahdi University, Kosti 11588, Sudan; musaabgaffar@mahdi.edu.sd

* Correspondence: mergani@whut.edu.cn; Tel.: +86-188-7609-0760

Received: 20 May 2018; Accepted: 13 June 2018; Published: 19 June 2018



Abstract: In the real-life, time-series data comprise a complicated pattern, hence it may be challenging to increase prediction accuracy rates by using machine learning and conventional statistical methods as single learners. This research outlines and investigates the Stacking Multi-Learning Ensemble (SMLE) model for time series prediction problem over various horizons with a focus on the forecasts accuracy, directions hit-rate, and the average growth rate of total oil demand. This investigation presents a flexible ensemble framework in light of blend heterogeneous models for demonstrating and forecasting nonlinear time series. The proposed SMLE model combines support vector regression (SVR), backpropagation neural network (BPNN), and linear regression (LR) learners, the ensemble architecture consists of four phases: generation, pruning, integration, and ensemble prediction task. We have conducted an empirical study to evaluate and compare the performance of SMLE using Global Oil Consumption (GOC). Thus, the assessment of the proposed model was conducted at single and multistep horizon prediction using unique benchmark techniques. The final results reveal that the proposed SMLE model outperforms all the other benchmark methods listed in this study at various levels such as error rate, similarity, and directional accuracy by 0.74%, 0.020%, and 91.24%, respectively. Therefore, this study demonstrates that the ensemble model is an extremely encouraging methodology for complex time series forecasting.

Keywords: time series forecasting; ensemble learning; heterogeneous models; SMLE; oil consumption

1. Introduction

In Machine Learning (ML), ensemble methods combine various learners to calculate prediction based on constituent learning algorithms [1]. The standard Ensemble Learning (EL) methods include bootstrap aggregating (or bagging) and boosting. Random Forest (RF) [2]; for instance, bagging combines random decision trees and can be used for classification, regression, and other tasks. The effectiveness of RF for regression has been investigated and analyzed in [3]. The boosting method, which builds an ensemble by adding new instances to emphasize misclassified cases, yields competitive performance for time series forecasting [4]. As the most generally utilized usage of boosting, Ada-Boost [5] has been compared with other ML algorithms such as support vector machines (SVM) [6] and furthermore combined with this algorithm to additionally enhance the forecasting performance [7]. Also, stacking [8] is an instance of EL multiple algorithms. It combines the yield which is produced by various base learners in the first level. In addition, by utilizing a meta-learner, it tries to combine the outcomes from these base learners in an ideal method to augment the generalization ability [9].

Although multistep predictions are desired in various applications, they are more difficult tasks than the one-step, due to lack of information and accumulation of errors. In some universal forecasting rivalries held lately, different forecasting methods were proposed to solve some genuine issues. In numerous studies, authors compared the performance of hybrid model on long-term forecasting, for instance, in [10], comparison results demonstrated that an ensemble of neural networks, such as multilayer perceptron (MLP), performed well in these competitions [10]. Also, Ardakani et al. [11] proposed optimal artificial neural networks (ANN) models based on improved particle swarm optimization for long-term electrical energy consumption. Regarding the same aspect this study, [12] introduced a model named the hybrid-connected complex neural network (HCNN), which is able to capture the dynamics embedded in chaotic time series and to predict long horizons of such series. In [13], researchers combined models with self-organizing maps for long-term forecasting of chaotic time series.

On the other hand, in short-term forecasting models, such as ANN and SVM, provide excellent performance for one-step forecasting task [14,15]. However, these models perform poorly or suffer severe degradation when applied to the general multistep problems. As well, the long-term forecasting models are designed for long time prediction tasks (for instance monthly or weekly time series prediction). That means they may perform better in multistep forecasting, while worse in one-step ahead than other methods. In general, the performance of combined forecasting models (e.g., mixing short-term and long-term approaches) is better when compared to single models [16]. Therefore, a forecasting combination can be benefit from performance advantages of short-term and long-term models, while avoiding their disadvantages. Furthermore, major static combination approaches [17–19] depend on assign a fixed weight for each model such as (average, inverse mean), while dynamic combinations methods such as bagging and boosting investigated to combine the results of complementary and diverse models generated by actively perturbing, reweighting, and resampling training data [20,21]. Therefore horizon dependent weights used to avoid the shortcoming of a static and dynamic combination for short- and long-term forecasts [14].

Oil Consumption (OC) is a significant factor for economic development, while the accuracy of demand forecasts is an essential factor leading to the accomplishment of proficiency arranging. Due to this reason, energy analysts are concerned with how to pick the most suitable forecasting methods to provide accurate forecasts of OC trends [22]. However, numerous techniques contribute to estimating the oil demand in future. The field of energy production, consumption, and price forecasting have been gaining significance as a current research theme in the entire energy sectors. For instance, numerous studies investigated for electricity price forecasting such as Rafał [23], this review article aims to explain and partition the primary methods of electricity price forecasting. Furthermore, Silvano et al. [24] analyzed electricity spot-prices of the Italian power exchange by comparing traditional methods and computational intelligence techniques NN and SVM models. Also, Nima and Farshid [25] proposed a hybrid method for short-ahead price forecasting composed of NN and evolutionary algorithms.

Several studies discussed the issue of time series prediction using different methodologies including statistical methods, single machine learning models, soft computing on ensemble, and hybrid modeling.

Statistical methods have been investigated for time series prediction in the energy consumption area, such as moving average [26], exponential smoothing [27,28], autoregressive moving average (ARMA) [29], and autoregressive integrated moving average (ARIMA) models [30]. For instance, the ARIMA model has been introduced for natural gas price forecasting [31]. However, these statistical techniques do not yield convincing results for complicated data patterns [32,33]. In this context, the Gray Model (GM) forecast accuracy was enhanced by using a Markov-chain model. The outcome of this study demonstrated that the hybrid GM-Markov-chain model was more accurate and had a higher forecast accuracy than GM (1, 1) [34].

In fact, neural networks offer a promising tool for single machine learning model in time series analysis due to their unique features [35]. To further improve the generalization performance, ANN models were investigated for forecasting future OC [36]. Another study experimented with ANN models to predict the long-term energy consumption [37]. For the same purpose, an ANN model was applied to forecast load demands in future [38].

However, ANNs yield mixed results when dealing with linear patterns [39], it is difficult to obtain high accuracy rates of predictors by using the single method, either statistical or ML techniques individually. In order to avoid the limitations associated with the individual models; researchers suggested a hybrid model which combines linear and nonlinear methods to yield high prediction accuracy rates [32,39]. Several studies investigated hybrid modeling to optimize the parameters of the ANN [40]. Hence the improved performance of artificial bee colony (ABC-LM) over other alternatives has been demonstrated on both benchmark data and OC time series.

Similarly, an NN, combined with three algorithms in a hybrid model, then optimized by using a genetic algorithm was used to estimate OC; the outcome demonstrated the efficiency of the hybrid model overall benchmark models [41]. Moreover, a researcher in [42] proposed a genetic algorithm—gray neural network (GA-GNNM) hybrid model to avoid the problem of over-fitting, by examining hybrid versus a total of 26 combination models. Authors concluded that the hybrid models provided desirable forecasting results, compared to the conventional models. Also, the GA has more flexibility in adapting NN parameters to overcome the performance instability of neural networks [22].

In the same context, hybrid models were investigated to solve prediction intervals and densities problems, and have become more common. As shown in Hansen [43] fuzzy model combined with neural models, this combination increased the computation speed, and the coverage is extended. Thus, the problem of the narrow prediction intervals is resolved. Similarly, in [44] the prediction interval also concerned with blend of neural networks and fuzzy models to determine the optimal order for the fuzzy prediction model and estimate its parameters with greater accuracy. Since prediction intervals and forecast densities have become more popular, many types of research have been done about how to determine the appropriate input lag, for this purpose, the fuzzy time series model suggested increasing accuracy by solving the problems of data size (sampling) and the normality [45]. Regarding the same aspect, Efendi and Deris extended a new adjustment of the interval-length and the partition number of the data set, this study discussed the impact of the proposed interval length in reducing the forecasting error significantly, as well as the main differences between the fuzzy and probabilistic models [46].

Finally, as a conclusion from the above studies, hybrid methods give off an impression of being an astounding way to combine predictions of several learning algorithms. The hybrid regression models give preferred predictive accuracy over any single learner. Nonetheless, there was no distinctive way to merge the outcome forecasts of individual models.

In this paper, the goal is to introduce a novel EL framework that can reduce model uncertainty, enhance model robustness, and enhance forecasting accuracy on oil datasets, improve model accuracy, being defined as having a lower measure of forecasting error. The most important motivation for combining different learning algorithms is based upon the assumption that diverse algorithms using different data representations, dissimilar perceptions, and modelling methods are expected to arrive at outcomes with different prototypes of generalization [47]. In addition, to date, comparatively few researches have addressed ensembles for different regression algorithms [48].

We demonstrate that the OC framework can significantly outperform the current methodologies of utilizing the single and classic ensemble forecasting models in single and multistep performance. Although the idea is straightforward, it is yet a robust approach, as it can outperform the average model, as one does not know a priori which model will perform best. The merits of this proposed methodology are analyzed empirically by first describing the exact study design and after that, assessing the performance of various ensembles of different OC models on the GOC. These outcomes are then compared to the classical approach in the literature, which takes the calibrated model with the lowest measure of forecasting error on the calibration dataset at the horizon (1-ahead) to OC of the same dataset at the horizon $t = n$ (10-ahead).

In summary, the developed ensemble model takes full advantage of each component and eventually achieves final success in energy consumption forecasting. The major contributions of this paper come therefore from three dimensions as follows:

1. In this study, we develop a new ensemble forecasting model that can integrate the merits of single forecasting models to achieve higher forecasting accuracy and stability.
We have introduced a novel theoretical framework how to predict OC. Although the ensemble concept is more demanding regarding computational requirements, it can significantly outperform the best performing model (SVR) of individual models. While the idea is straightforward, it is yet a robust approach, as it can outperform the linear combination methods, as one does not know a priori which model will perform best.
2. The proposed ensemble forecasting model aims to achieve effective performance in multi-step oil consumption forecasting.
Multi-step forecasting can effectively capture the dynamic behavior of oil consumption in the future, which is more beneficial to energy systems than one-step forecasting. Thus, this study builds a combined forecasting model to achieve accurate results for multi-step oil consumption forecasting, which will provide better basic for energy planning, production and marketing.
3. The superiority of the proposed ensemble forecasting model is validated well in a real energy consumption data.
The novel ensemble forecasting displays its superiority compared to the single forecasting model and classic ensemble models, and the prediction validity of the developed combined forecasting model demonstrates its superiority in oil consumption forecasting compared to classical ensemble models (AR, Bagging) and the benchmark single models (SVR, BPNN and LR) as well. Therefore, the new developed forecasting model can be widely used in all temporal data application prediction.
4. A perceptive discussion is provided in this paper to further verify the forecasting efficiency of the proposed model.
Four discussion aspects are performed, which include the significance of the proposed forecasting model, the comparison with single models, and classical ensemble methods, the superiority of the developed forecasting model's stability, which bridge the knowledge gap for the relevant studies, and provide more valuable analysis and information for oil consumption forecasting.

The structure of the paper is organized into five sections: Section 2 is devoted to describing proposed methods design. Section 3 presents the experimental results. Section 4 offers the consumption prediction analysis and discussion. Section 5 describes the conclusion and further suggestion for future works.

2. Materials and Methods

2.1. Proposed Framework

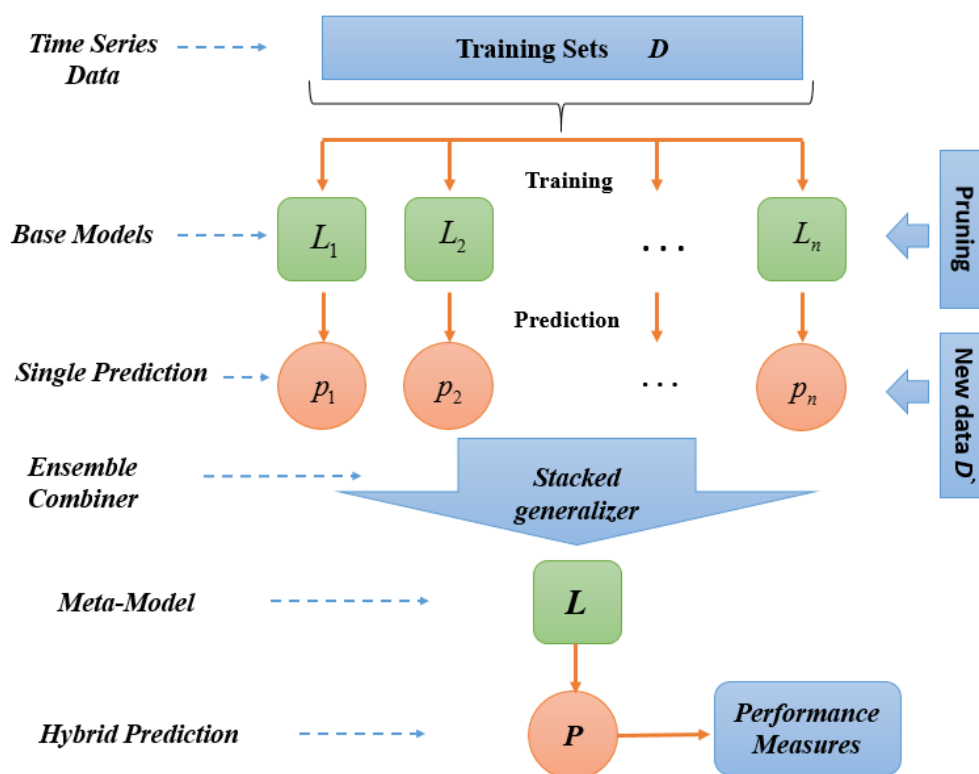
In Section 1, reviewed the literature in three different areas (i.e., single, hybrid, and soft computing on ensemble). While the hybrid modeling literature advanced significantly over the last 20 years, the research on minimizing forecast error, model uncertainty, and hybrid methods is still relatively limited so far. To the best of our knowledge, no attempts exist yet of combining these different areas, by using EL methods to reduce the issues of OC tasks (see Table 1 for a summary).

In particular, we will outline a very general theoretical framework to calibrate and combine heterogeneous ML models using ensemble methods. Its modularity is displayed in Figure 1 and allows for flexible implementation regarding base models, forecasting techniques, and ensemble architecture. For a practical application of this method, we have split the Stacking Multi-Learning Ensemble (SMLE) framework into four main phases and will describe them including their sub-steps in further detail as follows.

Table 1. Summary of related studies on forecasting OC between 2009 and 2017.

Reference	Method	Type	Duration	Region	Horizon
[36]	ANN	Single	1965–2010	Turkey	Long-term
[37]	MLP	Single	1992–2004	Greek	Long-term
[46]	FTS ¹ , RTS ²	Hybrid	1965–2012	Malaysia and Indonesia	Long-term
[45]	FTS, RTS	Hybrid	1965–2012	Malaysia	Long-term
[40]	ABCLM ³	Hybrid	1981–2006	Jordan, Lebanon, Oman, and Saudi Arabia	Short-term
[41]	ABCNN ⁴ , CSNN ⁵ , GANN ⁶	Hybrid	1980–2006	Middle East region	Short-term
[22]	GANN, ABCNN	Hybrid	1980–2006	OPEC ¹⁰	Short-term
[34]	GM ⁷	Hybrid	1990–2002	China	Short-term
[44]	ANFIS ⁸	Hybrid	1974–2012	U.S.	Short-term
[42]	GA, GNNM ⁹	Hybrid	2000–2010	China	Short-term
SMLE *	SVR, BPNN, LR	Ensemble	1965–2016	GOC ¹¹	Long-term

¹ Fuzzy Time Series; ² Regression Time Series; ³ Artificial Bee Colony Algorithm; ⁴ Artificial Bee Colony Neural Network; ⁵ Cuckoo Search Neural Network; ⁶ Genetic Algorithm Neural Network; ⁷ Grey Markov; ⁸ Adaptive Neuro-Fuzzy Inference Systems; ⁹ Genetic Algorithm—Gray Neural Network; ¹⁰ Organization of the Petroleum Exporting Countries; ¹¹ Global Oil Consumption; * Proposed Method.

**Figure 1.** Stacking Multi-Learning Ensemble (SMLE) Framework.

2.2. Ensemble Generation

In the original data set, the initial training data, represented as D , had m observations and n features, so that it is $m \times n$. The modeling procedure can be realized by setting different parameters of the base learners. In this level, some heterogeneous models were trained on D using one method

of training (i.e., cross-validation). Moreover, each model offered prediction results $p_i (i \in 1, 2, \dots, n)$ which were then cast into a second level data; the outcome became the input for the second level as training data.

2.2.1. Ensemble Pruning

Ranking-based subset selection method ranks the candidate models according to criteria, such as the mean absolute percentage error (MAPE), directional accuracy (DA), and Euclidean Distance (ED), and included only the top n models from all candidate models.

2.2.2. Ensemble Integration

This step describes how the selecting models were combined into ensemble forecast. In this context, the stacking method is used to build the second level data, stacking uses a similar idea to K-folds cross-validation to solve two significant issues: Firstly, to create out-of-sample predictions. Secondly, to capture distinct regions, where each model performs the best. The stacking process investigates by inferring the biases of the generalizers concerning the provided base learning set. Then, stacked regression using cross-validation was used to construct the 'good' combination. Consider a linear stacking for the prediction task. The basic idea of stacking is to 'stack' the predictions f_1, \dots, f_m by linear combination with weights $a_i, \dots, (i = 1, \dots, m)$:

$$f_{\text{stacking}}(x) = \sum_{i=1}^m a_i f_i(x), \quad (1)$$

where the weight vector a is learned by a meta-learner.

2.2.3. Ensemble Prediction

The second level learner model(s) can be trained on the D' data to produce the outcomes which will be used for final predictions. In addition, to select multiple sub-learners, stacking allows the specification of alternative models to learn how to best combine the predictions from the sub-models. Because a meta-model is used to combine the predictions of sub-models best, this method is sometimes termed blending, as in mixing the final predictions.

In brief, Figure 1 demonstrated the general structure of SMLE framework, which consisted of various learning steps, after applying this scheme, three SMLE models were generated, while the difference between the SMLE models were not in structure, but in the type of base model in level #0 and the differences between the three models in the part of base model can be explained as follows:

- 1st SMLE in base layer used SVR learner and in Meta layer LR used as meta learner.
- 2nd SMLE in base layer used BPNN learner and in Meta layer LR used as meta learner.
- 3rd SMLE in base layer used SVR and BPNN learners and in Meta layer LR used as meta learner.

2.3. Experiment Study Design

2.3.1. Data

The GOC data were used as benchmark data; this dataset was downloaded from the website: <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-worldenergy.html>. The data represented total OC in the world; the data was yearly type and had a duration from 1965 to 2016. The data consisted of two factors, thus dependent variable oil consumption (in Million Tonnes), which was a feature over time, and date (in years) was the independent variable in this case study. Therefore, the OC time series for this experiment had 52 data points. For a better explanation, we visualized whole actual time series in Figure 2, with a blue circle in curve.

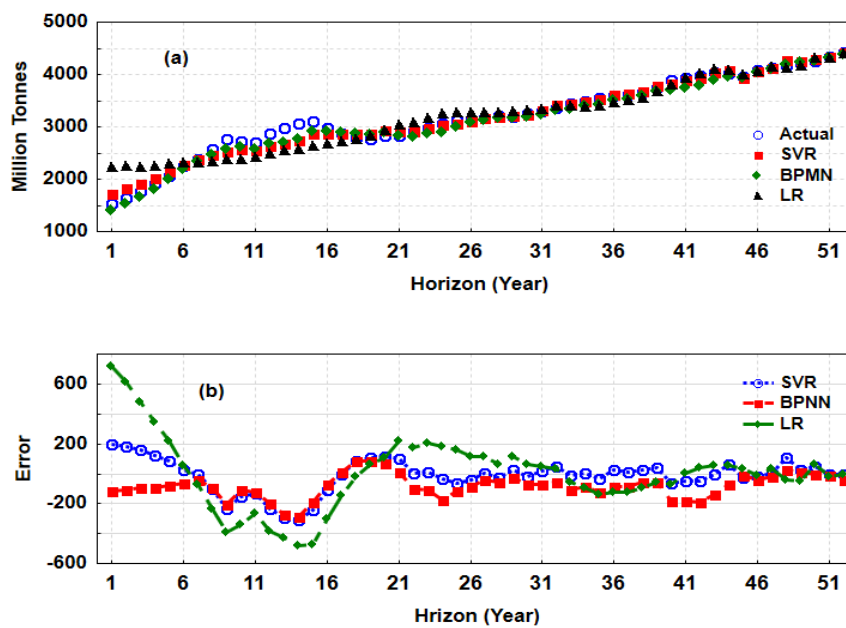


Figure 2. Comparison of (a) the actual and predicted consumption with the use of the SVR, BPNN and LR single learners (b) errors of all single models.

2.3.2. Models

As above-mentioned, we applied the ensemble SMLE model to predict the GOC data set after combining the heterogeneous models, Table 2 lists the learners' parameters that have been investigated in this paper. These related methods are presented briefly as follows:

1. The BPNN algorithm consists of multiple layers of nodes with nonlinear activation functions and can be considered as the generalization of the single-layer perceptron. It has been demonstrated to be an effective alternative to traditional statistical techniques in pattern recognition and can be used for approximating any smooth and measurable functions [49]. This method has some superior abilities, such as its nonlinear mapping capability, self-learning and adaptive capabilities, and generalization ability. Besides these features, the ability to learn from experience through training makes MLP an essential type of neural networks and it is widely applied to time series analysis [50].
2. The SVM algorithm is always considered a useful tool for classification and regression problems due to the ability to approximate a function. Furthermore, the kernel function is utilized in the SVR to avoid the calculations in high-dimensional space. As a result, it can perform well when the input features have high dimensionality. It separates the positive and negative examples as much as possible by constructing a hyperplane as the decision surface. The support vector regression (SVR) is the regression extension of SVM, which provides an alternative and promising method to solve time series modeling and forecasting [51,52].
3. LR is a popular statistical method for regression and prediction. It utilizes the ordinary least-squares method or generalized least-squares to minimize the sum of squares of errors (SSE) for obtaining the optimal regression function [53].

Table 2. Summary of parameters setting for all learners.

Model	Parameters
LR	Attribute method selection = Md5, batch Size = 100, and ridge = 1.0×10^{-8}
SVR	Kernel = (Poly), C = 1, exponent = 2 and epsilon = 0.0001.
BPNN	MLP(1-3-1)
Bagging	Base learner = REPTree, bagSizePercent = 100%, No. iteration = 10.
AR	Base learner = linear regression, No. iteration = 10, Shrinkage = 1.0.
1st SMLE	Base learner (SVR(Kernel = (Poly), C = 1, exponent = 2, epsilon = 0.0001)), meta learner (LR), Combination method= Stacked generalization
2nd SMLE	Base learner (MLP(1-3-1)), meta learner (LR), Combination method = Stacked generalization
3rd SMLE	Base learner (SVR (Kernel = (Poly), C = 1, exponent = 2, epsilon = 0.0001) and MLP (1-3-1), meta learner (LR), Combination method= Stacked generalization

2.3.3. Evaluation Measure

This subsection describes several aspects of the evaluation of the different models; the evaluation aspects include the estimation of error rates and pairwise comparisons of classifiers/ensembles.

1. Performance Evaluation

In terms of performance error estimation, the mean absolute percentage error (MAPE) was adopted as an indicator of accuracy for all forecasting methods. The accuracy is expressed as a percentage value, and is defined by the Formula (2) as below:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad (2)$$

where y_i is the actual value and \hat{y}_i is the forecast value.

2. Time Series Similarity

The distance between time series can be measured by calculating the difference between each point of the series. The Euclidean Distance (ED) between two-time series $Q = \{q_1, q_2, \dots, q_n\}$ and $S = \{s_1, s_2, \dots, s_n\}$ is defined as:

$$D(Q, S) = \sqrt{\sum_{i=1}^n (q_i - s_i)^2}. \quad (3)$$

This method is moderately easy to calculate, and has complexity of $O(n)$ [54].

3. Continuous Growth Rates (CGR)

Calculating change growth rate in data is useful for average annual growth rates that steadily change. It is famous because it relates the final value in series to the initial value in the same series, rather than just providing the initial and final values separately—it gives the ultimate value in context [30]. The CGR value calculated according to Formula (4) as follows:

$$k = \frac{\ln\left(\frac{y_{t+1}}{y_t}\right)}{t}, \quad (4)$$

where k represents the annual growth rate y_t represents the initial population size, t represents the future time in years and k is CGR.

2.4. The Algorithm for Stacking Multi-Learning Ensemble (SMLE)

In this study, SMLE offers a dynamic EL method. The SMLR method depends on the sequence characteristic of OC data. For accurate OC prediction, we express the algorithm of SMLE when predicting the next m th moment OC at the time t . The general design of the proposed model considered both diversity management and accuracy enhancement for base models. Here the algorithm of SMLE is described below as pseudocode in Algorithm 1:

Algorithm 1: Stacking Multi-Learning Ensemble (SMLE).

Input: Dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$;
 First-level learning algorithms L_1, L_2, \dots, L_n ;
 Second-level learning algorithm L ;

Process:

%Train a first-level individual learner h_t by applying the first-level learning algorithm L_t to the original dataset D

for $t = 1, \dots, T$:

$h_t = L_t(D)$

end;

% generate a new data set

$D' = \phi$;

for $i = 1, \dots, m$:

for $t = 1, \dots, T$:

$z_{it} = h_t(x_i)$ % Use h_t to predict training example x_i

end;

$D' = D' \cup \{(z_{i1}, z_{i2}, \dots, z_{iT}), y_i\}$

end;

% Train the second-level learner h' by applying the second-level learning algorithm L to the new data set D'

$h' = L(D')$.

Output: $H(x) = h'(h_1(x_1), \dots, h_T(x_T))$

3. Results

In this section, we evaluated various models on GOC 52-year data sets using BPNN, SVR, and LR as the base models to demonstrate their predictability of both single and EL forecasting. Hence, there were single models used as benchmark model compared to ensemble predictors. In the second experiment, we tested two classic ensemble models include bagging and additive regression (AR). Moreover, the third experiment tested three ensemble models based on SMLE scheme. To establish the validity of the evaluated method, a further procedure was done by comparing the obtained results of single models with the outcome of the ensemble models. Evaluation criteria were used to compare and analyze the prediction, such as T-Time, DA, MAPE, and ED, which are excellent methods for predicting GOC. Meanwhile, we compare the evaluation criteria of multistep (10-ahead) with single step (1-ahead) forecasting to find the better SMLE model for predicting GOC in both short-term and term-long horizon situations. Finally, consumption growth rate evaluated for all prediction outcome.

3.1. Single Models Results

Regarding the experiment design and the overall steps described in Section 2.2, the first test in this experiment was to compare the performance of all base models separately. The output of 10-fold cross-validation tests run on the initial training were used to determine whether each model was sufficient for OC data to make the forecasting results more stable. Figure 2a presents the comparison of the best-obtained results from all base models with the real OC data. It is evident that the results obtained according to the SVR method for the 52 known years (1965–2016) were close to the actual ones and comparable to those produced by the BPNN and LR models.

Similarly, Figure 2b demonstrated the residual errors of the prediction, to make a reliable comparison to quantitatively analyze the performances of the base models; we considered the MAPE measure indices for performance accuracy processes, which are listed in Table 3.

In brief, as seen in Table 3, the MAPE between predicted and actual values for the SVR model is 1.24% given by relative accuracy (DA) 89.9% which indicates clearly that the SVR model is well working and has acceptable accuracy. Regarding the same aspect, we can observe that the SVR had superiority in both run time and similarity (0.01 and 0.034, respectively). However, it is worth mentioning that the LR models scored poor performance compared to other single models. The similarity between actual and predicted data was measured using the Euclidean Distance (ED), as shown in Table 3; the BPNN score 0.034, which was small indicates the best predictive performance, while LR scores 0.074 was the worst similarity across the models.

3.2. Classic Ensemble Models Results

In the second experiment, we empirically tested two classical ensemble models, included bagging and additive regression (AR). To illustrate the behavior of all classical ensemble fitting, they were compared with actual data in Figure 3a,b, for visual comparison of the residual error of each model. The evaluation matrix of single learning, classical ensemble methods, and proposed SMLE models are summarized in Table 3. As observed from Table 3 and Figure 3, the bagging model performed better than the AR model in all evaluation measures, except in DA. Similarly, the bagging model performed better than the best single model (SVR) in performance and similarity while SVR perform best in DA and has least training time. For this dataset, we accordingly developed homogeneous and heterogeneous ensembles of individual models rather than using their hybrid versions.

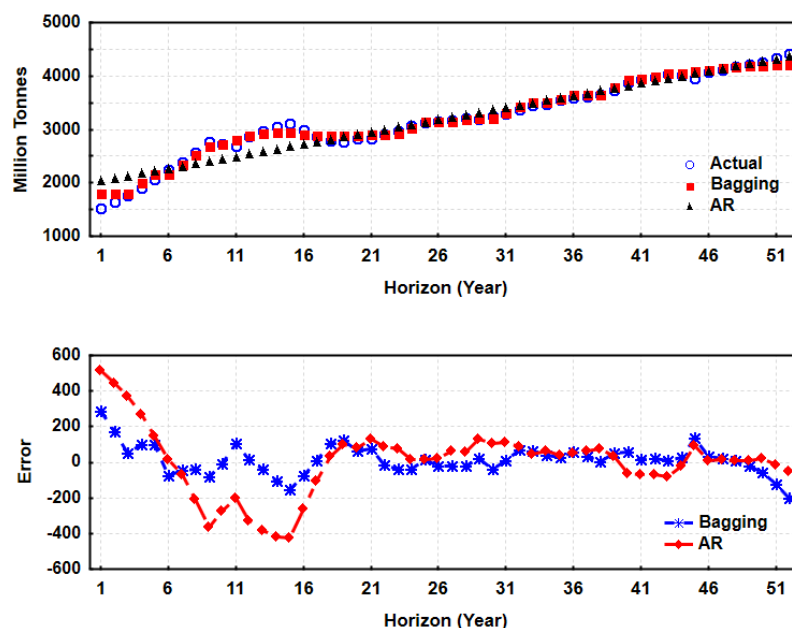


Figure 3. Illustrated (a) actual and predicted consumption using classic ensemble learners (b) error of classic ensemble models.

3.3. The SMLE Results

In the third experiment, we empirically tested three heterogeneous stacking models, each model was composed of a combination of base and meta-models. The first ensemble model consists of SVR as a base learner and LR as meta-learner. To illustrate the behavior of all SMLE for fitting, they were compared with actual data in Figure 4a,b, for visual comparison of the residual error of each model. The evaluation matrix of single learning methods, and proposed framework is summarized in Table 3.

The outcome of this model, as presented in Table 3, enhanced the forecasting accuracy by 34% when it was compared to the best base learner, SVR. Moreover, the second ensemble model was a mix of BPNN as the base learner and LR as the meta-learner, the combined model increased the forecasting accuracy by decreasing the error by 46%, compared to the best single model as mentioned previously.

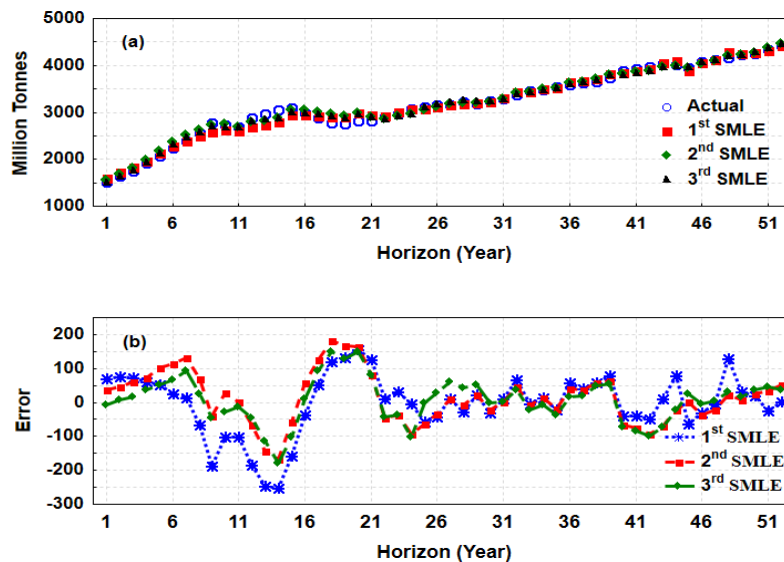


Figure 4. Illustrated (a) actual and predicted consumption using SMLE learners (b) error of SMLE models.

Table 3. Summary of different evaluating measures among all models on GOC data.

Measures	Single Models			Classic Ensemble Models		SMLE Models		
	LR	SVR	BPNN	Bagging	AR	1st SMLE	2nd SMLE	3rd SMLE
MAPE (%)	6.77	2.82	3.15	2.52	5.19	2.27	2.07	1.65
DA (%)	82.59	89.03	89.90	66.17	82.59	88.50	90.69	91.24
ED	0.074	0.035	0.034	0.026	0.059	0.028	0.024	0.020
T-Time	0.04	0.05	0.03	0.06	0.07	0.09	0.13	0.17

Bold number indicates the best value in all measures.

Finally, the third ensemble model was a combination of SVR and BPNN as base learners and LR as the meta-learner. The forecasting result of this model indicates that the accurate predictive model decreased the error of the best base model by 50%, which led to proof of the superiority of the third model over both the single and combination models. The similarity between actual and predicted data is shown in Table 3, the 3rd SMLE based (SVR-BPNN) model score was 0.020, while 1st SMLE based (SVR) score was 0.028, the worst similarity in across all the models. Also, it can be observed that all ensemble model had less distance compared to single models.

In the same aspect, the 3rd SMLE performed better than the best classic ensemble model (bagging) in all measures, except for training time (T-Time); this was due to the ensemble model learning level which consumed more time and calculation.

4. Discussion

In this subsection, we practically used all single and ensemble forecasters to solve the problem of how to estimate the future OC. For further evaluation of SMLE scheme stability, all models were examined in 1-ahead and 10-ahead horizon predictions.

From Figure 5 and Table 4, it is easy to find that the proposed SMLE method was the best one for OC forecasting in all prediction horizons (i.e., 1, 3, 5, 7, and 10-step-ahead), relative to other

models considered in this study. In all the models, the SMLE-based BPNN-SVR model did not only accomplish the highest accuracy at the level estimation, which was measured by the MAPE criteria, it additionally got the highest hit rate in direction prediction, which was estimated by the DA criterion. Then again, among the majority of the models utilized as a part of this investigation, the single LR model performed the poorest in all progression ahead forecasts. LR model not only had the lowest level accuracy, which was measured by MAPE, but also acquired the worst score in direction accuracy, which was measured by the DA criteria. The main reason might be that LR was a class of the typical linear model and it could not capture the nonlinear patterns and occasional characteristics existing in the data series. Apart from the SMLE-based BPNN-SVR and LR models, which performed the best and the poorest, respectively. All models listed in this study produce some interestingly blend results, these outcomes were analyzed by using four estimation criteria (i.e., MAPE, DA, *T*-test, and CGR).

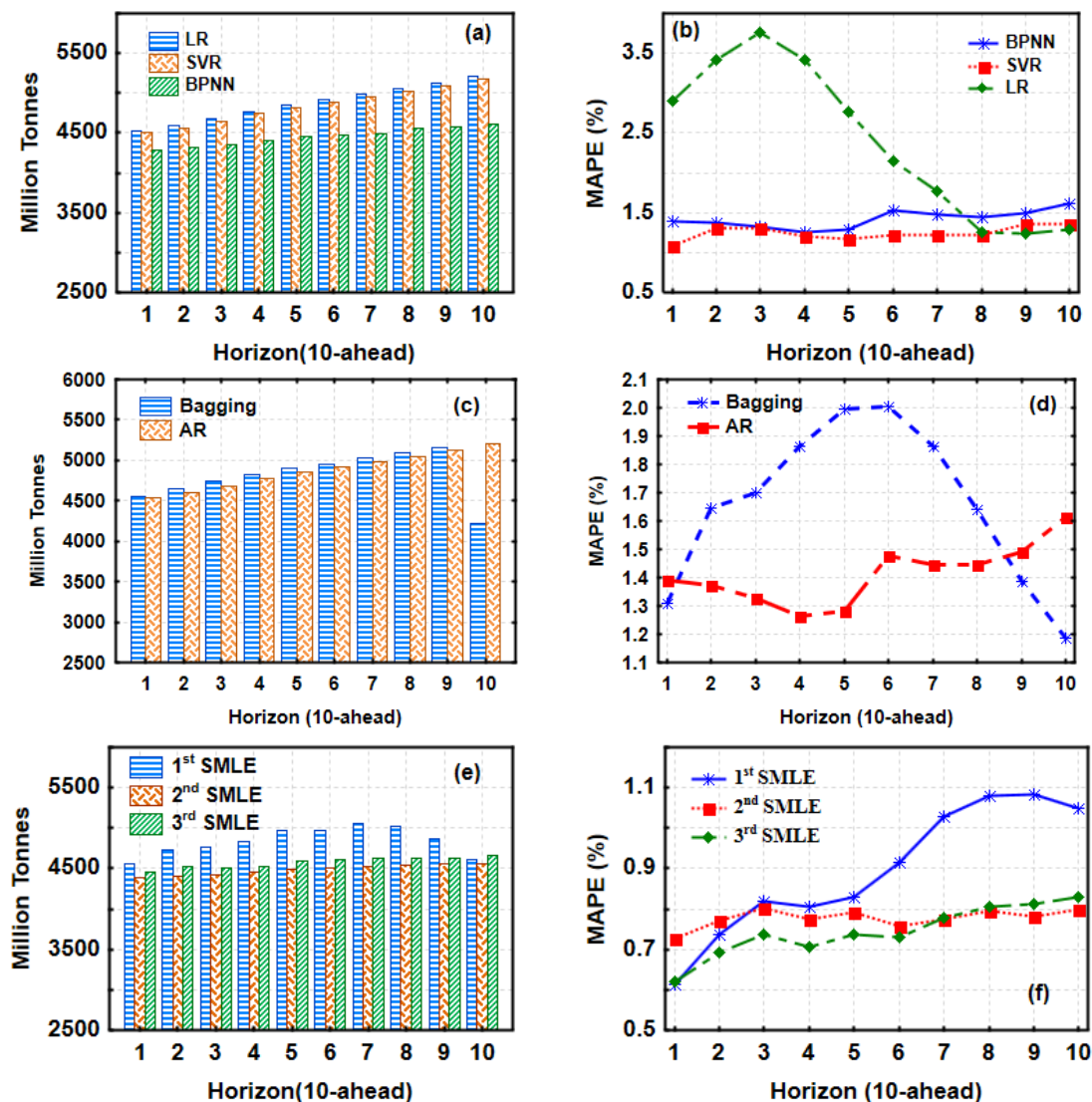


Figure 5. Illustrated 10-ahead consumption prediction and MAPE. (a) Single models. (b) Error of single models. (c) Classic ensemble models. (d) Error of classic ensemble models. (e) SMLE models. (f) error of SMLE models.

Firstly, in the case of level accuracy, the results of the MAPE measure demonstrated that the SMLE-based BPNN-SVR performed the best, followed by SMLE-based BPNN, SMLE-based SVR models, SVR and BPNN, and the weakest model was LR as shown in Figure 5b,f. Moreover,

from Table 4, the MAPE values of the SMLE-based BPNN-SVR model were 0.61 in 1-ahead and 0.74 as an average of the 10-step-ahead predictions, which was less than other methods. Also, in the short-term prediction step, better performance was observed when comparing ensemble methods with single models, the results indicate that the ensemble methods outperformed the single and classic ensemble methods in all cases. The principle reason could be that the cross-validation decomposition methodology did efficiently enhance the forecast execution. Interestingly, the 1-step-ahead and multi-step-ahead prediction horizon of single model forecasts were inferior to ensemble models. Focusing on the single methods and classic ensembles, all the ML models outperformed the LR model; the reason may be that LR is a typical linear model, which is not suitable for capturing the nonlinear and seasonal characteristics of OC series. In ML models (i.e., SVR, BPNN), it can be seen that SVR performed slightly better than BPNN in all 10-step-ahead predictions and BPNN perform poorest in all the step prediction. The main reason leading to this may be the parameter selection. The MAPE values of LR were from 2.91 to 2.40, which were slightly inferior to SVR and BPNN models. The possible reason was that the prediction results of LR, which was under the linear hypothesis were more volatile than those of the ML models.

Second, the high-level exactness does not necessarily imply that there was a high hit rate in forecasting direction of OC. The correct forecasting direction is essential for the policy manager to make an investment plan in oil-related operations (production, price, and demand).

Table 4. 10-ahead forecasting performance among all models on GOC data.

Model	MAPE (%) over 10-Ahead Horizon					Avg.
	1-ahead	3-ahead	5-ahead	7-ahead	10-ahead	
LR	2.91	3.76	2.76	1.77	1.30	2.40
SVR	1.08	1.30	1.17	1.23	1.36	1.24
BPNN	1.39	1.33	1.28	1.48	1.61	1.42
Bagging	1.31	1.70	1.99	1.86	1.19	1.66
RF	1.39	1.33	1.28	1.45	1.61	1.41
1st SMLE	0.62	0.82	0.83	1.03	1.05	0.90
2nd SMLE	0.73	0.80	0.79	0.77	0.80	0.78
3rd SMLE	0.61	0.74	0.74	0.78	0.83	0.74

Therefore, the DA comparison is necessary. In Figure 6a–c, some similar conclusions can be drawn regarding DA criterion. (i) The proposed 3rd SMLE model performed significantly better than all other models in all cases, followed by the other two ensemble models, then two of the single ML models (i.e., SVR, BPNN), (LR, AR) had equal values, and bagging model had the worst values. Individually, the DA values of all SMLE-based ensembles were similar 92.31% for the 1 step-ahead predictions and showed superiority with 91.24% for average 10-ahead step forecasts for the 3rd SMLE model. (ii) The three ensemble methods mostly outperformed the single prediction models. Furthermore, among the ensemble methods, the SMLE-based BPNN-SVR model performed the best, and SMLE-based BPNN model outperformed SMLE-based SVR model, except for the 2-ahead forecast. (iii) SVR model outperformed other methods, BPNN had the similar performance as SVR in the 2, 3, 5 step-ahead forecasts, except that SVR exceeded BPNN in both 1-ahead and average ahead prediction. The possible reason leading to this phenomenon may be the choice of optimal parameters for the models. We also found that bagging model had the lowest directional accuracy of 66.17%.

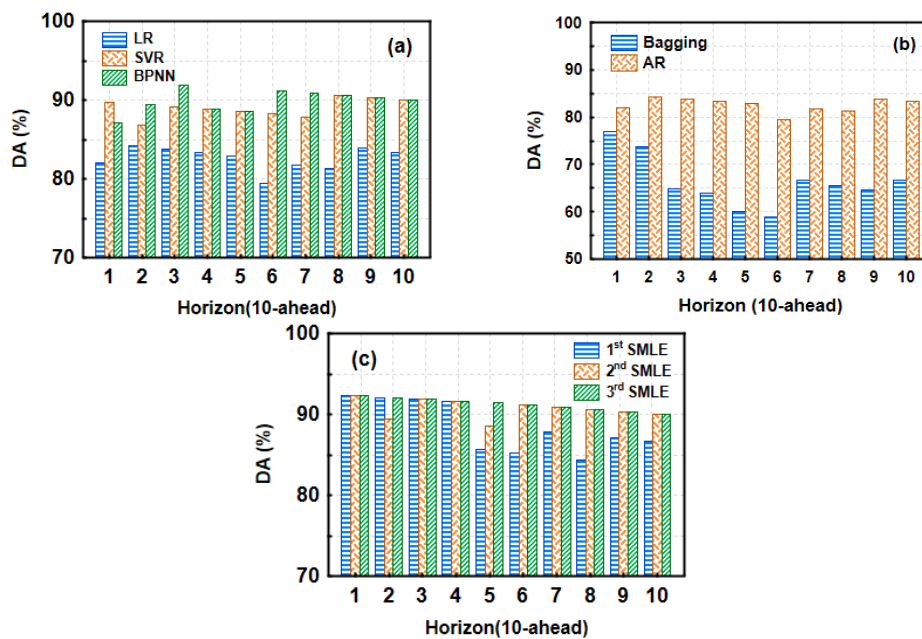


Figure 6. Illustrated 10-ahead consumption directional accuracy (DA) of (a) single models, (b) classic ensemble models, and (c) SMLE ensemble models.

Also, comparing different prediction horizons, the short-term prediction horizon showed better performance for in all the model see Table 4. Taking 1-step-ahead forecasting and an average of the 10-step-ahead predictions for example, for all the SMLE-based ensemble, BPNN, SVR, bagging, AR models, the 1-step-ahead forecasting outperformed the average of the 10-step-ahead forecast, no matter the level accuracy or directional accuracy. Apart from the models mentioned above, SMLE-based ensembles and ML models and classic ensembles performed better in 1-step-ahead prediction given directional accuracy. However, from the point of level accuracy, both these approaches only had slight superiority in 6-step-ahead prediction. Except for the LR, which performed almost poorer in the 1-step-ahead compared to the average of 10-step-ahead prediction as shown in Figure 7a–c.

Third, to further validation of SMLE models forecasting, the *t*-test was used to test the statistical significance of the prediction performance. The *t*-test results presented in Table 5, for all ensemble models under this study were not significant ($df = 51$, p -value > 0.05). Based on the detailed statistical test, no significant differences were observed between the actual OC and that predicted by the SMLE models. The mean differences in the last column of Table 5, indicate that in the population from where the sample models were drawn, the actual and predicted OC was statistically semi-equal. Therefore, it was possible to prove that the SMLE model was useful in predicting OC based on the heterogeneous models with excellent levels of accuracy (see Table 5). So, we can conclude that the model developed structure is sufficient with more parameters setting (i.e., kernels, neuron) for OC prediction.

Table 5. The *t*-test results of actual and predicted oil consumption using SMLE models.

Model	<i>t</i>	<i>p</i> -Value	Mean Difference
1st SMLE	0.227	0.823	0.8081
2nd SMLE	1.320	0.193	0.6803
3rd SMLE	0.728	0.470	0.6178

The forecasted values for each model and total OC growth rate from 2017 to 2026 is summarized in Table 6. As seen from the table, all models will still be increasing in the period from 2017 to

2026. However, the average annual rates will decrease in all. For the period between 1965 and 2016, the rate of increases was 2.2% for BPNN, 1.3% for LR, 1.8% for SVR, 1.5% for bagging 1.6%, for AR 2.0% for SMLE-based SVR, 2.0% for SMLE-based BPNN, and 2.1% for SMLE-based BPNN-SVR. Additionally, for the forecasted period between 2017 and 2026 the rates were expected to be 0.74%, 1.39%, 1.38%, 1.42%, 1.39%, 0.13%, 0.38%, and 0.44, respectively. On the other hand, the average annual rate of total oil demand decreased from 1.8% between 1965 and 2016 to 0.91% between 2017 and 2026.

Lastly, the summarized results in Table 6 demonstrate that the annual growth rates of 1-ahead OC were more significant than the total average OC in 10-ahead years. Figure 8, shows the apparent rise in the 1-ahead in both single and classic ensemble models, and for the SMLE models there was a sudden drop from 1- to 2-ahead years, also note the stability in the growth from 2-ahead to 10-ahead, with close values in all models, except for SMLE-based BPNN where there was a few decreasing in the 9-, 10-ahead, sequentially. The decrease in the rate of oil demand may be interpreted as there being other alternative energies that affect oil demand, this will be achieved in the coming decades, as compared with all other energy type consumption. Rates of changes and reserves in the OC of all the models indicate that the SMLE scheme was the best to determine the actual demand of energy globally, which facilitates the planning process, associated with the issue OC prediction. Based on these study findings, we suggested some recommendations.

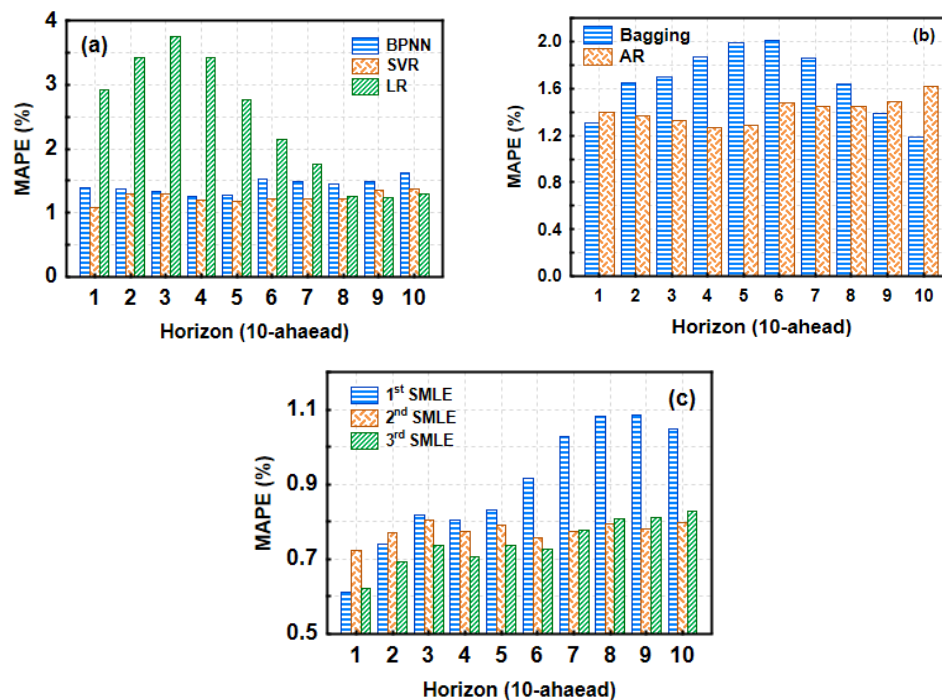


Figure 7. Illustrated 10-ahead consumption prediction errors of (a) single models (b) classic models (c) SMLE ensemble models.

Table 6. Summary of forecasted values and CGR for OC using all models from 2017 to 2026.

Years	Models							
	BPNN	LR	SVR	Bagging	RF	1st SMLE	2nd SMLE	3rd SMLE
2017	4279.16	4531.93	4504.13	4559.73	4531.93	4554.73	4395.00	4459.94
2018	4316.27	4595.00	4567.34	4646.02	4595.00	4737.14	4409.57	4523.12
2019	4349.24	4677.23	4650.48	4748.19	4677.23	4764.54	4426.70	4511.95
2020	4401.42	4771.37	4746.31	4819.88	4771.37	4835.67	4454.33	4530.88
2021	4448.96	4857.41	4821.93	4896.70	4857.41	4967.25	4482.28	4592.90

Table 6. Cont.

Years	Models							
	BPNN	LR	SVR	Bagging	RF	1st SMLE	2nd SMLE	3rd SMLE
2022	4471.17	4915.79	4889.67	4956.50	4915.79	4967.25	4501.68	4602.70
2023	4496.30	4980.78	4954.70	5025.85	4980.78	5056.31	4519.38	4633.88
2024	4555.86	5052.61	5023.62	5095.53	5052.61	5029.93	4536.88	4635.57
2025	4578.78	5128.13	5097.75	5163.26	5128.13	4861.17	4553.77	4628.92
2026	4606.92	5205.40	5170.20	5215.23	5205.40	4615.31	4566.33	4659.36
2017–2026	0.74%	1.39%	1.38%	1.42%	1.39%	0.13%	0.38%	0.44%
1965–2016	2.2%	1.3%	1.8%	1.5%	1.6%	2.0%	2.0%	2.1%

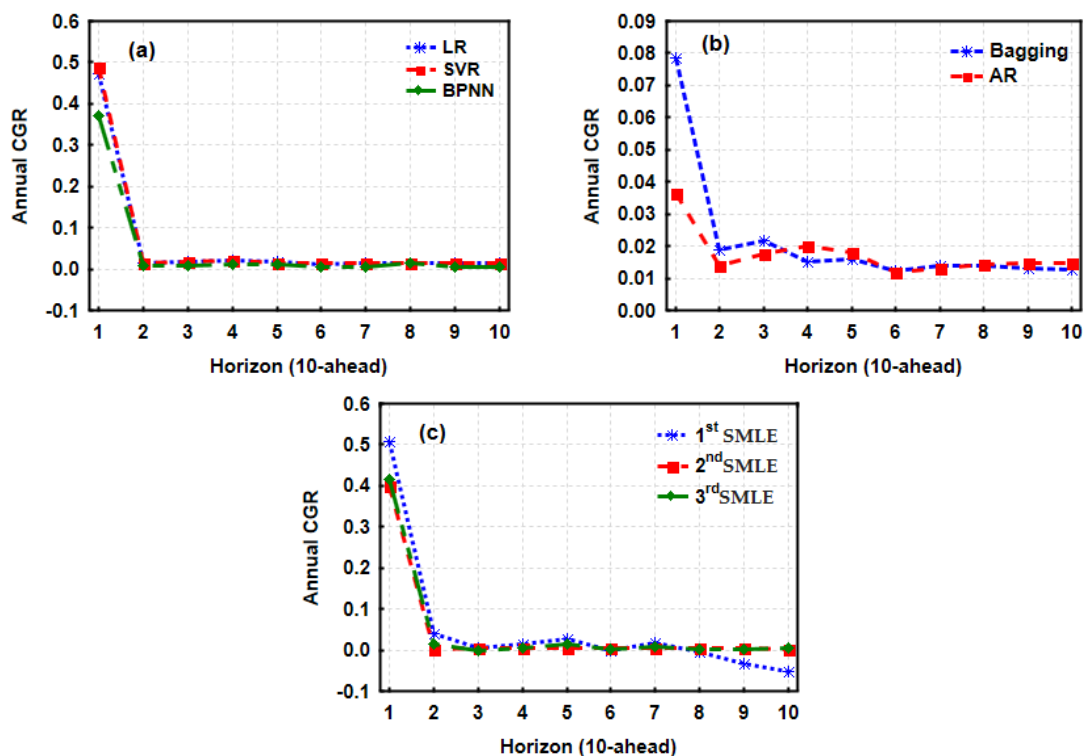


Figure 8. Illustrated annual CGR for 10-ahead consumption prediction using (a) single models (b) classic ensemble models (c) SMLE models.

We summarized all of the above results in Table 7 and Figure 9. In general, combining the forecasters using SMLE will significantly improve the final prediction. Generally, from the analysis of the experiments presented in this study, we can draw several important conclusions as follows: Firstly, the SMLE-based BPNN-SVR model was significantly superior to all models in this study regarding similarity, level accuracy, and direction accuracy. Through performance enhancement, the SMLE-based BPNN-SVR outperformed other models at the 1.17 statistical significance level, compared to the best benchmark models SVR and bagging, respectively. Secondly, the prediction performance of the SMLE-based BPNN-SVR, SMLE-based SVR and SMLE-based BPNN models were better than the single and classic ensemble methods. These results indicate that the hybrid, based on stacking method, can efficiently improve the prediction performance in the case of OC. Thirdly, nonlinear models, with seasonal adjustment, were more suitable as base learners for the ensemble to predict the time series with annual volatility than linear methods, due to properties above of OC (i.e., nonlinear and non-stationary). However, computationally, the new method consumed more time because of its way of segmenting inputs and the use of the ensemble. Fourthly, the average annual rate of total oil demand decreased from 1.8% between 1965 and 2016 to 0.91% between 2017 and 2026. Finally, on one

hand, short-term forecasting models, such as BPNN and SVM, provided excellent performance for one-step forecasting task. However, these models performed poorly or suffered severe degradation when applied to the general multistep problems. In general, the performance of ensemble forecasting models (e.g., combining short-term and long-term approaches) was better when compared to single models. Therefore, a forecasting combination can benefit from performance advantages of short-term and long-term models, while avoiding their disadvantages. Furthermore, to overcome the shortcoming of a static combination approach, a dynamic combination of short- and long-term forecasts can be employed by using horizon dependent weights.

Table 7. Summary of evaluation measures among all models on GOC data.

NO.	Model	Evaluation Matrix				Score ¹	Indexed Rank ²
		T-Time	MAPE (%)	ED	DA (%)		
1	LR	0.04	2.4	0.074	82.59	29	8
2	BPNN	0.05	1.42	0.035	89.03	20	6
3	SVR	0.03	1.24	0.034	89.90	19	5
4	Bagging	0.06	1.66	0.026	66.17	18	4
5	AR	0.07	1.41	0.059	82.59	22	7
6	1st SMLE	0.09	0.9	0.028	88.50	15	3
7	2nd SMLE	0.13	0.78	0.024	90.69	9	2
8	3rd SMLE	0.17	0.74	0.020	91.24	8	1

¹ Score: sum of rank values from (1–8) for each model depends on performance in related measure. ² Order value for each model depending on total score, for example rank no 1 means the first model.

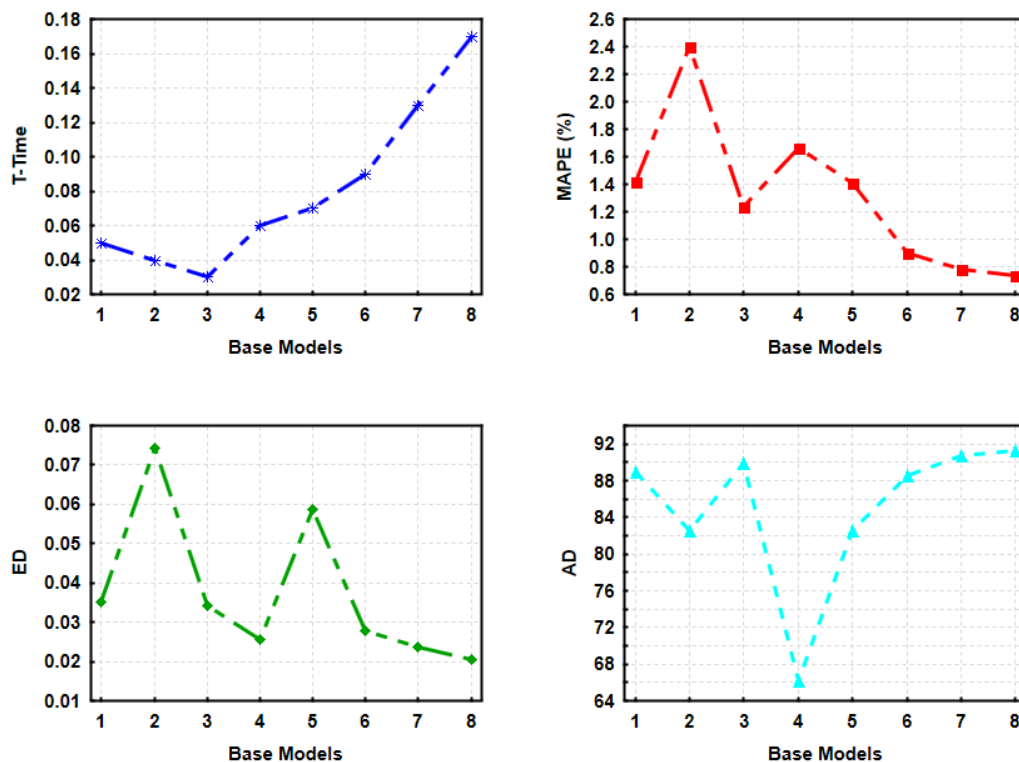


Figure 9. Illustrated T. Time MAPE, ED, and AD evaluation measures of all models on 10-ahead GOC prediction. (The order of models were arranged from 1–8 according to Table 6, as LR, BPNN SVR, bagging, AR, 1st, 2nd, and 3rd SMLE, respectively).

5. Conclusions

Forecasting time series data is considered as one of the most critical applications and has concerned interests of researchers. In this study, we discussed the problem of combining heterogeneous forecasters and showed that ensemble learning methods could be readily adapted for this purpose. We have introduced a novel theoretical ensemble framework integrating BPNN, SVR, and LR, based on the principle of stacking; which was proposed for the GOC forecasting. This framework has been able to reduce uncertainty, improve forecasting performance, and manage the diversity of learning models in empirical analysis.

According to the experimental results and analyses, the proposed ensemble models have been able to outperform the classical ensemble and single models on OC data analyzed results. Furthermore, all ensemble models have been able to exceed the best performing individual models on single-ahead, as well as the multi-ahead horizon.

The advantages of proposed model to the knowledge comes therefore along three aspects as follows:

Firstly, in methodology part, we have introduced a novel theoretical framework based on ensemble learning for OC forecasting. Although the ensemble concept is more demanding regarding computational requirements, it can significantly outperform single models and classical hybrid models. While the idea is straightforward, it is yet a robust approach, as it can outperform linear combination methods, as one does not know a priori which model will perform best.

Secondly, theoretically we have demonstrated that ensemble methods can be successfully used in the context of OC forecasting due to the ambiguity decomposition.

Thirdly, we have conducted a very extensive empirical analysis of advanced machine learning models, as well as ensemble methods. Just the calibration alone of such a wide range of ensemble models is very rare in the literature, considering that the ranking of some evaluation measures per model to run, which was not only due to the limited.

This study has two limitations including: the consideration of the integration of heterogeneous algorithms (SVR, BPNN and LR) without using ensemble pruning for internal hyper-parameters; and the evaluation process investigated on single data set, so that this model can verified in different data sets. All these limitations could be interesting future research.

In future work, homogeneous ensemble model based SVR with different kernels can be developed and evaluated. In addition, investigating ensemble pruning by using evolutionary algorithms that provides an automatic optimization approach to SVR hyper-parameters, could be an interesting future research work in the hybrid-based energy forecasting field. Another direction of future work is to apply ensemble models in other energy prediction problems, such as pricing, production, and load forecasting.

Author Contributions: M.A.K. performed the experiments, analyzed the data, interpreted the results and wrote the paper. X.N. supervised this work, all of the authors were involved in preparing the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was funded by the Chinese Scholarship Council, China (under Grants CSC No. 2015-736012).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Seni, G.; Elder, J.F. Ensemble methods in data mining: Improving accuracy through combining predictions. In *Synthesis Lectures on Data Mining and Knowledge Discovery*; 2010; Volume 2, pp. 1–126.
2. Segal, M.R. *Machine Learning Benchmarks and Random Forest Regression*; Center for Bioinformatics and Molecular Biostatistics: San Francisco, CA, USA, 2004.
3. Grömping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **2009**, *63*, 308–319. [[CrossRef](#)]

4. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [[CrossRef](#)]
5. Shrestha, D.L.; Solomatine, D.P. Experiments with AdaBoost. RT, an improved boosting scheme for regression. *Neural Comput.* **2006**, *18*, 1678–1710. [[CrossRef](#)] [[PubMed](#)]
6. Morra, J.H.; Tu, Z.; Apostolova, L.G.; Green, A.E.; Toga, A.W.; Thompson, P.M. Comparison of AdaBoost and support vector machines for detecting Alzheimer’s disease through automated hippocampal segmentation. *IEEE Trans. Med. Imaging* **2010**, *29*, 30–43. [[CrossRef](#)] [[PubMed](#)]
7. Guo, L.; Ge, P.-S.; Zhang, M.-H.; Li, L.-H.; Zhao, Y.-B. Pedestrian detection for intelligent transportation systems combining adaboost algorithm and support vector machine. *Expert Syst. Appl.* **2012**, *39*, 4274–4286. [[CrossRef](#)]
8. Aldave, R.; Dussault, J.-P. Systematic ensemble learning for regression. *arXiv*, 2014.
9. Lemke, C.; Gabrys, B. Meta-learning for time series forecasting and forecast combination. *Neurocomputing* **2010**, *73*, 2006–2016. [[CrossRef](#)]
10. Crone, S.F.; Hibon, M.; Nikolopoulos, K. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *Int. J. Forecast.* **2011**, *27*, 635–660. [[CrossRef](#)]
11. Ardakani, F.J.; Ardehali, M.M. Novel effects of demand side management data on accuracy of electrical energy consumption modeling and long-term forecasting. *Energy Convers. Manag.* **2014**, *78*, 745–752. [[CrossRef](#)]
12. Gómez-Gil, P.; Ramírez-Cortes, J.M.; Hernández, S.E.P.; Alarcón-Aquino, V. A neural network scheme for long-term forecasting of chaotic time series. *Neural Process. Lett.* **2011**, *33*, 215–233. [[CrossRef](#)]
13. Fonseca-Delgado, R.; Gomez-Gil, P. Selecting and combining models with self-organizing maps for long-term forecasting of chaotic time series. In Proceedings of the 2014 International Joint Conference on Neural Networks, Beijing, China, 6–11 July 2014; pp. 2616–2623.
14. Simmons, L. Time-series decomposition using the sinusoidal model. *Int. J. Forecast.* **1990**, *6*, 485–495. [[CrossRef](#)]
15. Abdoos, A.; Hemmati, M.; Abdoos, A.A. Short term load forecasting using a hybrid intelligent method. *Knowl.-Based Syst.* **2015**, *76*, 139–147. [[CrossRef](#)]
16. De Gooijer, J.G.; Hyndman, R.J. 25 years of time series forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [[CrossRef](#)]
17. Stock, J.H.; Watson, M.W. Combination forecasts of output growth in a seven-country data set. *J. Forecast.* **2010**, *23*, 405–430. [[CrossRef](#)]
18. Khairalla, M.; Xu, N.; Al-Jallad, N. Modeling and optimization of effective hybridization model for time-series data forecasting. *J. Eng.* **2018**. [[CrossRef](#)]
19. Hsiao, C.; Wan, S.K. Is there an optimal forecast combination? *J. Econom.* **2014**, *178*, 294–309. [[CrossRef](#)]
20. Barrow, D.; Crone, S. Dynamic model selection and combination in forecasting: An empirical evaluation of bagging and boosting. *Med. Phys.* **2011**, *25*, 435–443.
21. Barrow, D.K.; Crone, S.F.; Kourentzes, N. An evaluation of neural network ensembles and model selection for time series prediction. In Proceedings of the International Joint Conference on Neural Networks, Barcelona, Spain, 18–23 July 2010; pp. 1–8.
22. Chiroma, H.; Abubakar, A.I.; Herawan, T. Soft computing approach for predicting OPEC countries’ oil consumption. *Int. J. Oil Gas Coal Technol.* **2017**, *15*, 298–316. [[CrossRef](#)]
23. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [[CrossRef](#)]
24. Cincotti, S.; Gallo, G.; Ponta, L.; Raberto, M. Modeling and forecasting of electricity spot-prices: Computational intelligence vs. classical econometrics. *AI Commun.* **2014**, *27*, 301–314.
25. Amjady, N.; Keynia, F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. *Int. J. Electr. Power Energy Syst.* **2008**, *30*, 533–546. [[CrossRef](#)]
26. Azadeh, A.; Ghaderi, S.; Sohrabkhani, S. Forecasting electrical consumption by integration of neural network, time series and ANOVA. *Appl. Math. Comput.* **2007**, *186*, 1753–1761. [[CrossRef](#)]
27. Bianco, V.; Manca, O.; Nardini, S.; Minea, A.A. Analysis and forecasting of nonresidential electricity consumption in Romania. *Appl. Energy* **2010**, *87*, 3584–3590. [[CrossRef](#)]

28. Ediger, V.Ş.; Tatlıdil, H. Forecasting the primary energy demand in Turkey and analysis of cyclic patterns. *Energy Convers. Manag.* **2002**, *43*, 473–487. [[CrossRef](#)]
29. Manera, M.; Marzullo, A. Modelling the load curve of aggregate electricity consumption using principal components. *Environ. Model. Softw.* **2005**, *20*, 1389–1400. [[CrossRef](#)]
30. Ediger, V.Ş.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708. [[CrossRef](#)]
31. Mishra, P. Forecasting natural gas price-time series and nonparametric approach. In Proceedings of the World Congress on Engineering, 2012 Vol I WCE 2012, London, UK, 4–6 July 2012; Available online: http://www.iaeng.org/publication/WCE2012/WCE2012_pp490-497.pdf (accessed on 19 June 2018).
32. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [[CrossRef](#)]
33. Lai, K.K.; Yu, L.; Wang, S.; Huang, W. Hybridizing exponential smoothing and neural network for financial time series predication. In *Proceedings of the International Conference on Computational Science, Reading, UK, 28–31 May 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 493–500.
34. Ma, H.; Zhang, Z. Grey prediction with Markov-Chain for Crude oil production and consumption in China. In *Proceedings of the Sixth International Symposium on Neural Networks (ISNN 2009), Wuhan, China, 26–29 May 2009*; Springer: Berlin, Germany, 2009; pp. 551–561.
35. Niska, H.; Hiltunen, T.; Karppinen, A.; Ruuskanen, J.; Kolehmainen, M. Evolving the neural network model for forecasting air pollution time series. *Eng. Appl. Artif. Intell.* **2004**, *17*, 159–167. [[CrossRef](#)]
36. Turanoglu, E.; Senvar, O.; Kahraman, C. Oil consumption forecasting in Turkey using artificial neural network. *Int. J. Energy Optim. Eng.* **2012**, *1*, 89–105. [[CrossRef](#)]
37. Ekonomou, L. Greek long-term energy consumption prediction using artificial neural networks. *Energy* **2010**, *35*, 512–517. [[CrossRef](#)]
38. Buhari, M.; Adamu, S.S. Short-term load forecasting using artificial neural network. In Proceedings of the International Multi-Conference of Engineers and Computer Scientist, Goa, India, 19–22 January 2012.
39. Nochai, R.; Nochai, T. ARIMA model for forecasting oil palm price. In Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Penang, Malaysia, 13–15 June 2006; pp. 13–15.
40. Chiroma, H.; Abdulkareem, S.; Muaz, S.A.; Abubakar, A.I.; Sutoyo, E.; Mungad, M.; Saadi, Y.; Sari, E.N.; Herawan, T. An intelligent modeling of oil consumption. In *Advances in Intelligent Informatics*; Springer: Berlin, Germany, 2015; pp. 557–568.
41. Chiroma, H.; Khan, A.; Abubakar, A.I.; Muaz, S.A.; Gital, A.Y.U.; Shuib, L.M. Estimation of Middle-East Oil Consumption Using Hybrid Meta-Heuristic Algorithms. Presented at the Second International Conference on Advanced Data and Information Engineering, Bali, Indonesia, 25–26 April 2015.
42. Xia, Y.; Liu, C.; Da, B.; Xie, F. A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Syst. Appl.* **2018**, *93*, 182–199. [[CrossRef](#)]
43. Hansen, B.E. Interval forecasts and parameter uncertainty. *J. Econom.* **2006**, *135*, 377–398. [[CrossRef](#)]
44. Rubinstein, S.; Goor, A.; Rotshtein, A. Time series forecasting of crude oil consumption using neuro-fuzzy inference. *J. Ind. Intell. Inf.* **2015**, *3*, 84–90. [[CrossRef](#)]
45. Efendi, R.; Deris, M.M. Forecasting of Malaysian oil production and oil consumption using fuzzy time series. In *Proceedings of the International Conference on Soft Computing and Data Mining, San Diego, CA, USA, 11–14 September 2016*; Springer: Berlin, Germany, 2016; pp. 31–40.
46. Efendi, R.; Deris, M.M. Prediction of Malaysian–Indonesian oil production and consumption using fuzzy time series model. *Adv. Data Sci. Adapt. Anal.* **2017**, *9*, 1750001. [[CrossRef](#)]
47. Aho, T.; Enko, B.; Eroski, S.; Elomaa, T. Multi-target regression with rule ensembles. *J. Mach. Learn. Res.* **2012**, *13*, 2367–2407.
48. Xu, M.; Golay, M. Survey of model selection and model combination. *SSRN Electron. J.* **2008**. [[CrossRef](#)]
49. Wang, J.-Z.; Wang, J.-J.; Zhang, Z.-G.; Guo, S.-P. Forecasting stock indices with back propagation neural network. *Expert Syst. Appl.* **2011**, *38*, 14346–14355. [[CrossRef](#)]
50. Ebrahimpour, R.; Nikoo, H.; Masoudnia, S.; Yousefi, M.R.; Ghaemi, M.S. Mixture of MLP-experts for trend forecasting of time series: A case study of the Tehran stock exchange. *Int. J. Forecast.* **2011**, *27*, 804–816. [[CrossRef](#)]

51. Miranian, A.; Abdollahzade, M. Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 207–218. [[CrossRef](#)] [[PubMed](#)]
52. Kao, L.-J.; Chiu, C.-C.; Lu, C.-J.; Yang, J.-L. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing* **2013**, *99*, 534–542. [[CrossRef](#)]
53. Tofallis, C. Least squares percentage regression. *J. Mod. Appl. Stat. Methods* **2009**, *7*, 526–534. [[CrossRef](#)]
54. Kianimajd, A.; Ruano, M.G.; Carvalho, P.; Henriques, J.; Rocha, T.; Paredes, S.; Ruano, A.E. Comparison of different methods of measuring similarity in physiologic time series. *IFAC-PapersOnLine* **2017**, *50*, 11005–11010. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).