UTM
UNIVERSITI TEKNOLOGI MALAYSIA

IJBES

International Journal of Built Environment and Sustainability
Published by Penerbit UTM Press, Universiti Teknologi Malaysia
IJBES 12(1)/2025, 9-21

Enhancing Electricity Consumption Forecasting in Limited Dataset: A Simple Stacked Ensemble Approach Incorporating Simple Linear and Support Vector Regression for Malaysia

Zun Liang Chuan
Statistics & Data Analytics Cluster, Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan Pahang, Malaysia
chuanzl@umpsa.edu.my

Ong Shao Jie; Tham Yim Hin
Faculty of Chemical Process Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan Pahang, Malaysia

Siti Nur Syamimi Binti Mat Zain; **Yunalis Amani Binti Abdul Rashid**; **Ainur Naseiha Binti Kamarudin**
Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakbo, 26300, Kuantan Pahang, Malaysia

ABSTRACT

Rapid population growth and urbanization, coupled with technological advancements, have driven higher electricity demand, predominantly sourced from contributors to climate change. This article introduces a novel artificial intelligence (AI) time-series algorithm, a simple stacked ensemble of simple linear regression (SLR) and Support Vector Regression (SVR), designed to forecast Malaysia's annual electricity consumption, particularly in scenarios with limited datasets utilizing the Cross Industry Standard Process for Data Mining (CRISP-DM) data science methodology. Analysis revealed that this simple stacked ensemble SVR-based time-series algorithm, employing an ε-insensitive loss function with a third-degree polynomial kernel, outperformed 71 other SVR-based algorithms, including four time-series algorithms from the previous study. The algorithm's forecasting insights from the formulated algorithm could guide policymakers in establishing more effective regulations aligned with Sustainable Development Goals (SDGs) such as affordable and clean energy (SDG7), decent work and economic growth (SDG8), industry, innovation and infrastructure (SDG9), sustainable cities and communities (SDG11), responsible consumption and production (SDG12), and climate action (SDG13), which benefit economic, environmental, human, and social.

Article History

Received : *22 December 2023*
Received in revised form : *19 August 2024*
Accepted : *5 September 2024*
Published Online : *10 January 2025*

Keywords:

Data Science Methodology, Support Vector Regression, Stacked Ensemble Time-Series Algorithm, Electricity Forecasting, Sustainable Development Goals.

Corresponding Author Contact:

chuanzl@umpsa.edu.my

DOI: 10.11113/ijbes.v12.n1.1254

1. Introduction

Energy plays a substantial role in socio-economic development worldwide. This is due to energy being a predominant component of the Sustainable Development Goals (SDGs), composed of four principal pillars such as economic, environmental, human, and social (Tan *et al.*, 2013). According to the International Energy Agency's 2007 World Energy Outlook, the global energy consumption demand was projected to grow by 55% between 2005 and 2030 (Khanna & Rao, 2009). Based on this projection, it was expected that three-fourths of the demand for global energy consumption would rise in developing countries, including Malaysia. The principal root causes of the rapid growth of electricity consumption in Malaysia are due to rapid population growth, expansion of living quarters, and technological advancements. Specifically, Figure 1 summarizes the annual electricity generation and consumption for 1978-2021 in Malaysia, conveying that total electricity consumption has experienced rapid compounded growth over the years.

In light of the pressing climate change issue and the need for environmental sustainability, the global energy landscape is shifting toward cleaner and more sustainable energy resources. The transition to clean and renewable energy, such as biomass, geothermal, hydropower, and wind, has gained global attention as an essential step toward reducing greenhouse gas emissions and mitigating environmental impacts (Solaun & Cerdá, 2019). This is because embracing clean energy solutions offers a path to combat climate change, fosters socio-economic development, and ensures energy security for future generations. Consequently, formulating an AI time-series algorithm for forecasting annual electricity consumption is much needed to support informed decision-making for clean energy transition, socio-economic progress, energy conservation, and environmental protection.

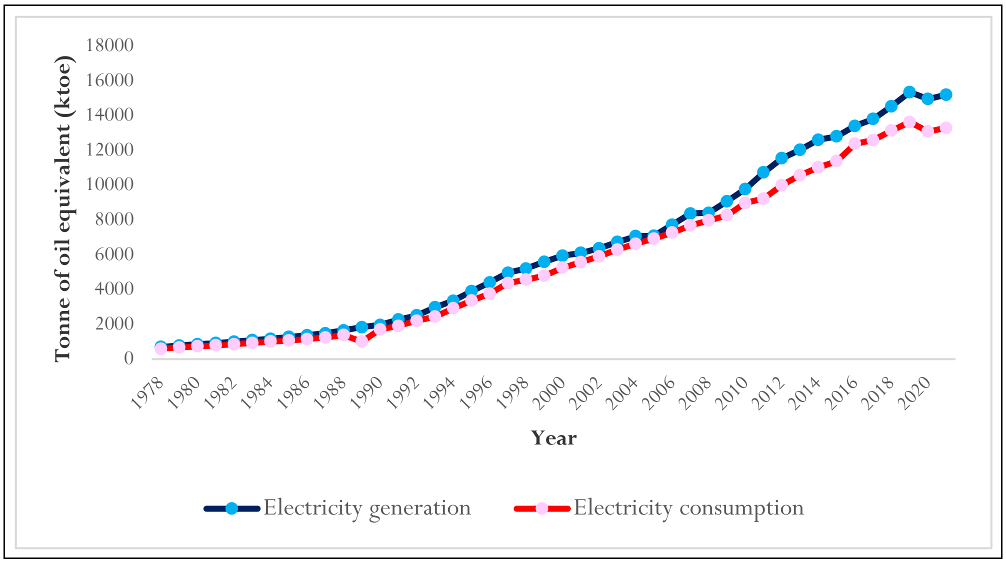

Figure 1. Electricity generation and consumption trends in Malaysia (1978-2021) (Energy Commission, 2024)

From the perspective of statistical machine learning, time-series algorithms formulated in previous studies can generally be categorized into three different approaches such as conventional time-series algorithms (Guo *et al.*, 2020; He *et al.*, 2016; Hong *et al.*, 2013; Jifri *et al.*, 2017; Kamisan *et al.*, 2018; Kandananond, 2011; Lee & Ko, 2011; Massaoudi *et al.*, 2021; Miswan *et al.*, 2016a, 2016b; Muneer *et al.*, 2022; Ping & Kamarudin, 2022; Razak *et al.*, 2009; Sulandari *et al.*, 2022), AI time-series algorithms (Abad *et al.*, 2020; Chong *et al.*, 2017; Guo *et al.*, 2020; He *et al.*, 2016; Her *et al.*, 2022; Jifri *et al.*, 2017; Kamisan *et al.*, 2018; Kandananond, 2011; Lee & Ko, 2011; Massaoudi *et al.*, 2021; Miswan *et al.*, 2016a; Muneer *et al.*, 2022; Nagi *et al.*, 2011; Shapi *et al.*, 2021; Sulandari *et al.*, 2022) and hybrid AI time-series algorithms (Abad *et al.*, 2020; Guo *et al.*, 2020; He *et al.*, 2016; Kamisan *et al.*, 2018; Lee & Ko, 2011; Massaoudi *et al.*, 2021; Miswan *et al.*, 2016a; Nagi *et al.*, 2011; Razak *et al.*, 2009; Sulandari *et al.*, 2022). In literature, conventional time-series algorithms, such as Autoregressive Integrated Moving Average-based (ARIMA-based) and multiple linear regression-based (MLR-based), have been extensively applied worldwide for short-term electricity consumption forecasting. These applications span various countries, including the Republic of Poland (Khanna & Rao, 2009), China (Guo *et al.*, 2020; He *et al.*, 2016; Hong *et al.*, 2013), Canada (He *et al.*, 2016), Singapore (He *et al.*, 2016), Malaysia (Razak *et al.*, 2009; Miswan *et al.*, 2016a, 2016b; Jifri *et al.*, 2017; Kamisan *et al.*, 2018; Massaoudi *et al.*, 2021; Muneer *et al.*, 2022; Ping & Kamarudin, 2022; Sulandari *et al.*, 2022), Thailand (Kandananond, 2011), Taiwan (Lee & Ko, 2011), the United of America (USA) (Massaoudi *et al.*, 2021), and Indonesia (Sulandari *et al.*, 2022).

However, ARIMA-based and MLR-based time-series algorithms have notable limitations. These algorithms require extensive time-series datasets, especially when utilizing the maximum likelihood estimation (MLE) method. They are primarily suitable for linear association environments and may not effectively capture complex non-linear associations. Additionally, the conventional time-series algorithms, such as the ARIMA-based and MLR-based time-series algorithms are essential to satisfy certain assumptions. Specifically, the residuals are required to fulfill the stationary univariate process and have an independent and identically Gaussian distribution with mean zero and constant variance.

Moreover, analysis results from worldwide literature revealed that AI and hybrid AI time-series algorithms outperformed conventional time-series algorithms in predicting electricity consumption, including Malaysia (Jifri *et al.*, 2017; Kamisan *et al.*, 2018; Massaoudi *et al.*, 2021; Miswan *et al.*, 2016b; Sulandari *et al.*, 2022). For instance, Jifri *et al.* (2017) proposed utilizing a multivariable regression-based AI time-series algorithm to predict the electricity load demand in Johor state, Malaysia. Their analysis results revealed that the proposed regression-based AI time-series algorithm outperformed conventional time-series algorithms such as exponential smoothing (ES), univariate Box-Jenkins (UBJ), autoregressive autoregressive (ARAR), Errors, Trends, and Seasonal (ETS) algorithms.

Meanwhile, Kamisan *et al.* (2018) proposed a hybrid AI time-series algorithm that integrated multivariable linear regression (MLR) and ANN in predicting the electricity load, principally focusing on the commercial area of Johor State, Malaysia. To authenticate the effectiveness, this study compared the performance of the proposed hybrid AI time-series algorithm with the MLR and ANN time-series algorithms. The analysis results of this study also revealed that the proposed hybrid AI time-series algorithm outperformed the MLR and ANN time-series algorithms on average.

Furthermore, Muneer *et al.* (2022) proposed utilizing the long short-term memory (LSTM) time-series algorithm to predict the load consumption of the residential sector. The analysis also revealed that LSTM outperformed conventional time-series algorithms such as ES and UBJ. Meanwhile, Sulandari *et al.* (2022) proposed hybridizing the clustering-based bootstrap aggregation time-series algorithms in predicting electricity load, primarily focusing on Malaysia, the Republic of Poland, and Indonesia. Specifically, this article proposed to hybridize the bootstrap aggregation respectively with conventional time-series algorithms such as Seasonal UBJ, neural network autoregression (NNAR), trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components (TBATS), and Double-Seasonal Holt-Winters (DSHW) time-series algorithms. The analysis results from this study showed that the number of bootstrap series does not affect the forecasting accuracy. Contrarily, the analysis results in this study revealed that the forecasting accuracy merely improved when the hybridization of the bootstrap aggregation with the appropriate aforementioned conventional time-series algorithms was utilized.

Nevertheless, several recent studies merely focus on effectively comparing of AI and hybrid AI time-series algorithms in predicting the electricity load in Malaysia. For instance, Chong *et al.* (2017) proposed to utilize ε-Support Vector Regression (ε-SVR) for load prediction. In this study, the researchers also compared the effectiveness corresponding to the kernel function of SVR such as linear, polynomial, radial basis function (RBF), and sigmoid. Their analysis results revealed that ε-SVR with a linear kernel function outperformed SVR with 2-degree polynomial, RBF, and sigmoid kernel functions, as well as Bayes-regularization ANN time-series algorithms. Shapi *et al.* (2021) studied the effectiveness of comparing of three different cloud-based AI time-series algorithms in predicting the energy consumption of smart commercialized buildings. This article has compared the predictive performance of three different AI time-series algorithms such as ANN, k-Nearest Neighbours (kNN), and ε-SVR utilizing the RBF kernel function. The principal analysis results of this research revealed that ε-SVR time-series algorithms outperformed ANN and kNN. In contrast, Her *et al.* (2022) proposed employing the ANN to forecast the short-term electricity load demand. This study compared and evaluated the superiority of four non-linear activation functions of the hidden layer such as exponential, sigmoid, softsign, and tanh in predicting the electricity load demand. The analysis results of this study revealed that the tanh activation function of the hidden layers outperformed other activation functions.

On the other hand, Nagi *et al.* (2011) proposed a novel hybrid of self-organizing map (SOM) and ε-SVR (SOM-ε-SVR) time-series algorithm in predicting the electricity load, focusing on the electricity load dataset of EUNITE Competition 2001, Peninsular Malaysia, and the USA with taking into account the external determinants such as peak load, temperature, type of day, and annual holidays. The analysis results revealed that the proposed hybrid AI time-series algorithm outperformed ε-SVR, Backpropagation Neural Network (BPNN) time-series algorithms, and also a series of regression-based, SVR-based, and ANN-based time-series algorithms in the literature periods 2005-2009.

Moreover, Massaoudi *et al.* (2021) proposed another novel advanced hybrid AI time-series algorithm for predicting the load demand principally focusing on Malaysia and the USA. In particular, the proposed hybrid AI time-series algorithm stacked three different AI algorithms such as extreme gradient boosting machine (XGB), light gradient boosting machine (LGBM), and multilayer perceptron (MLP) (Stacked XGB-LGBM-MLP). The simulation results of this study revealed that the proposed stacked AI time-series algorithm outperformed 11 comparison benchmark AI time-series algorithms, which were based on forecasting 24 hours ahead. Specifically, the comparison benchmarks of the AI time-series algorithm employed in this study include hybrid Convolutional Neural Networks Fuzzy Time-Series (FTS-CNN), Seasonal UBJ, Probabilistic Weighted Fuzzy Time-Series (PWFTS), Weighted Fuzzy Time-Series (WFTS), Integrated Weighted Fuzzy Time-Series (IWFTS), LSTM, Random Forest (RF), kNN, XGB, LGBM and MLP.

The principal limitations of the AI time-series algorithm, such as ANN-based including the LSTM, MLP-based, and SOM-based are these algorithms require high computational cost, the need for a physically interpretable determination of the optimal number of neurons, hidden layers, and appropriate activation function, and required a very large sample size to result in a reliable and accurate prediction (Farsi *et al.*, 2021). However, Malaysia's annual electricity consumption data is limited and characterized by a small sample size (< 50). Despite the previous studies (Chong *et al.*, 2017; Shapi *et al.*, 2021) revealing that the SVR-based time-series algorithm performed well in predicting Malaysia's annual electricity consumption data. However, the SVR-based time-series algorithm is unsuitable for forecasting future short-term electricity consumption. Furthermore, the appropriateness of the kernel function applied in the SVR-based time-series algorithm is highly dependent on the inherent characteristics of the load consumption dataset, which inherently vary globally due to the differences of the electricity grid and are not comprehensively investigated in the Malaysia dataset, including the research carried out by Chong *et al.* (2017), and Shapi *et al.* (2021).

To address these issues, the principal objective of this article is to formulate a novel simple stacked ensemble AI time-series algorithm for forecasting annual electricity consumption in Malaysia utilizing a limited dataset. Specifically, this study aims to carry out a mathematical analysis by comprehensively comparing the effectiveness of the hybrid of simple linear regression (SLR) and various well-known kernel functions utilizing ε-insensitive and ν-insensitive loss functions of simple stacked ensemble SVR-based time-series algorithms. The kernel functions taken into account in this study include linear $(\boldsymbol{\eta}_{\mathrm{LK}})$, polynomial $(\boldsymbol{\eta}_{\mathrm{PK}})$, with a maximum attainable degree of six, RBF $(\boldsymbol{\eta}_{\mathrm{RK}})$, and sigmoid $(\boldsymbol{\eta}_{\mathrm{SK}})$, Kernel functions. The maximum attainable value for the kernel function parameter (Δ) is five for the $(\boldsymbol{\eta}_{\mathrm{PK}})$ and $(\boldsymbol{\eta}_{\mathrm{SK}})$, respectively. To pursue the principal objective of this study, the rest of this article is organized as follows: Section 2 provides a brief overview of the schematic methodology of the ε-SVR, ν-SVR, kernel functions, and goodness-of-fit (GoF) statistics. Section 3 presents the results of analysis and the corresponding discussion. Finally, the concluding remarks are given in Section 4.

2. Data Science Methodology

This section provides an overview of the schematic methodology of the formulated novel simple stacked ensemble SVR-based time-series algorithm, associated with its theoretical backgrounds. Fundamentally, the schematic method of the formulated simple stacked ensemble SVR-based time-series algorithm is developed based on the Cross Industry Standard Process for Data Mining (CRISP-DM) data science methodology, as depicted in Figure 2. The data science methodology presented in Figure 2 is tailored to this research. The CRISP-DM data science methodology was selected for this study primarily due to its flexibility and effective application across various studies (Solano *et al.*, 2022; Brzozowska *et al.*, 2023; Chuan et al., 2024; Liang *et al.*, 2024; Chuan *et al.*, in press).

2.1 Business Understanding and Data Understanding

The first two predominant phases in the CRISP-DM data science methodology are business and data understanding. The primary focus of business understanding is to comprehend the principal objective of the project and its requirements. Meanwhile, the second phase, data understanding, builds upon the foundation of business understanding. The primary focus of this phase is to profile the acquired annual electricity consumption time-series dataset. Specifically, the principal objective of data mining in this article is to formulate a novel simple stacked ensemble SVR-based time-series algorithm for predicting the annual electricity consumption in Malaysia from 1978 to 2021. The formulated simple stacked ensemble SVR-based time-series algorithm in this study does not rely on any statistical assumption, unlike conventional time-series algorithms. As the time-series dataset involved in this study is a univariate secondary dataset from the Energy Commission of Malaysia, Microsoft Excel and an open-source R statistical software is utilized to carry out the statistical analysis; therefore, there is no financial cost imposed.

Meanwhile, the insightful predictive results from the formulated simple stacked ensemble SVR-based time-series algorithm could primarily benefit policymakers in establishing more effective regulations and actions in the energy sector. This, in turn, would aid in effective decision-making for socio-economic development, energy saving, and environmental protection. On the other hand, the acquired dataset is profiled utilizing appropriate univariate graphical and non-graphical exploratory data analysis tools, such as Boxplot and the first four statistical L-moments.

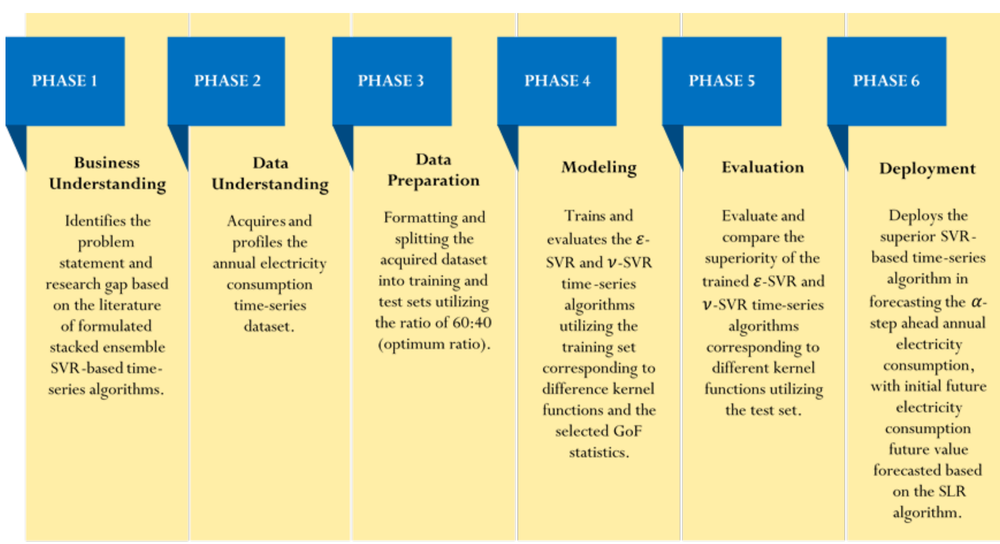

Figure 2. Stacked ensemble SVR-based time-series algorithms with CRISP-DM data science methodology

2.2 Data Preparation

The data preparation phase generally focuses on preparing and cleaning the acquired dataset, including tasks such as imputing missing values, outlier identification and correction, data integration, and data formatting. Owing to the acquired univariate time-series dataset from the official website of the Energy Commission of Malaysia being a complete dataset without missing values, the imputation of missing values and data integration is not required in this study. Furthermore, outlier identification and correction are not performed in this article. This is due to the principal objective of this study, which is to formulate a novel simple stacked ensemble SVR-based time-series algorithm for predicting the compound growth non-linear characteristics of annual electricity consumption in Malaysia as depicted in Figure 1. Contrarily, the acquired time-series dataset is formatted by splitting it into training and test sets (hold-out cross-validation) with a 60:40 ratio selected based on the optimum ratio among 60:40, 70:30, 80:20 and 90:10. Specifically, the optimum ratio of 60:40 for training and test sets in this article is comprehensively selected based on the minimum absolute error of the average between the GoF statistics for the training and test sets. As a result, selecting the optimum ratio of the training and test sets imposed in this study allowed for minimizing the risk of overfitting.

2.3 Modeling

The principal objective of the modeling phase in the CRISP-DM data science methodology is to build and assess trained AI algorithms, which is consistent with the principal purpose of this article. In particular, the principal aim of this article is to formulate a novel simple stacked ensemble SVR-based time-series algorithm trained utilizing the split training set, which includes both simple stacked ensemble ε-SVR-based and ν-SVR-based time-series algorithms. This study also aims to comprehensively investigate the superiority of various kernel functions and the parameters of the kernel functions applied in the formulated simple stacked ensemble SVR-based time-series algorithms for effectively forecasting the compound growth non-linear characteristics of electricity consumption. Consequently, a brief overview of the theoretical backgrounds of ε-insensitive and ν-insensitive loss functions of SVR, the kernel functions, and the GoF statistics are also presented in this sub-section.

2.3.1 Support Vector Regression Time-Series Algorithms

Support vector machines (SVM) are widely recognized as powerful supervised AI algorithm that can be utilized for both classification and regression tasks. SVM is particularly well-suited for small sample sizes or limited dataset environments. Since this section limits the discussion to the regression task, the SVR will be utilized instead of SVM. The stem of this preliminary study has been motivated to carry out a comprehensive mathematical analysis of the SVR-based time-series algorithm study owing to its effectiveness highlighted in Malaysia's literature (Chong *et al.*, 2017; Nagi *et al.*, 2011; Shapi *et al.*, 2021). However, these studies do not discuss the effectiveness of ν-SVR time-series algorithms. Consequently, this study desired to bridge this gap by providing a comprehensive mathematical analysis to compare the effectiveness of various kernel functions utilized in ε-insensitive and ν-insensitive loss functions and the corresponding parameters of the kernel functions for SVR are utilized to predict the annual electricity consumption in Malaysia utilizing a limited dataset.

Suppose that $\left\langle \left((\mathbf{X}_{obs})_k, (\mathbf{Y}_{obs})_k \right) \middle| k = 1, 2, \ldots, n \right\rangle$ is a set of data points, such that $\mathbf{X}_{obs} \in \mathbb{R}^d$. Fundamentally, the concept of SVR is to map the nonlinearly $\mathbf{Y}_{obs}$ into high-dimension feature space, ξ, utilizing a nonlinear mapping, $\boldsymbol{\eta}$, which can be denoted as Eq. (2.1).

$$f\left(\delta_k, \delta_k^*, \mathbf{X}_{obs}, \mathbf{X}\right) = \left(\sum\left(\delta_k - \delta_k^*\right)\boldsymbol{\eta}\left(\mathbf{X}_{obs}, \mathbf{X}\right)\right) + \lambda \qquad (2.1)$$

where $\delta_k \geq 0$ and $\delta_k^* \geq 0$ are unknown coefficients, and the kernel function, $\boldsymbol{\eta}(\mathbf{X}_{obs}, \mathbf{X}) = \varsigma(\mathbf{X}_{obs}) \cdot \varsigma(\mathbf{X})$ represents the inner product of two vectors, $\mathbf{X}_{obs}$ and $\mathbf{X}$, in the features space, which $\boldsymbol{\eta} \in \mathbb{R}^d \to \xi, \sum\left(\delta_k - \delta_k^*\right)\boldsymbol{\eta}(\mathbf{X}_{obs}) \in \xi$, and λ represents the support vector (SV) bias. In mathematics, the unknown coefficients of δ_k, δ_k^*, and λ for the ε-SVR and ν-SVR regression in Eq. (2.1) is attainable by minimizing the regularized ε-insensitive and ν-insensitive loss function in Eqs. (2.2)-(2.3), respectively.

$$\min\left(2^{-1}\left\|\sum\left(\delta_k - \delta_k^*\right)\right\|^2 + C\left(n^{-1}\sum\left(\delta_k - \delta_k^*\right)\right)\right) \qquad (2.2)$$

$$\min\left\{2^{-1}\left\|\sum\left(\delta_k - \delta_k^*\right)\right\|^2 + C\left(\nu\varepsilon + n^{-1}\sum\left(\delta_k - \delta_k^*\right)\right)\right\} \qquad (2.3)$$

where $\varepsilon \geq 0$, $0 \leq \nu \leq 1$, and C represents the regularized parameter, which is the well-recognized default value of $\varepsilon = 0.1$ is employed in this article. The root motivation of $\varepsilon = 0.1$ is employed due to its ability to balance exploration and exploitation in various applications (He, 2018; Rätz *et al.*, 2019). From the mathematical theoretical perspective, the primary difference between simple stacked ensemble ε-SVR-based and ν-SVR-based time-series algorithms is the ν, the parameter in Eq. (2.3) allowed the simple stacked ensemble ν-SVR time-series algorithm to control the number of SVs available in the resulting algorithm, while vice versa for the simple stacked ensemble ε-SVR-based time-series algorithm. As a result, the simple stacked ensemble ν-SVR time-series algorithm imposed less computational complexity than the simple stacked ensemble ε-SVR time-series algorithm.

2.3.2 Kernel Functions and Goodness-of-Fits Statistics

In SVR, kernel functions are utilized to train and predict the AI algorithm. Based on the literature (Chong *et al.*, 2017; Nagi *et al.*, 2011; Rohmah *et al.*, 2021), there are four well-known kernel functions utilized in SVR, such as linear kernel $(\boldsymbol{\eta}_{\mathrm{LK}})$, polynomial kernel $(\boldsymbol{\eta}_{\mathrm{PK}})$, RBF kernel $(\boldsymbol{\eta}_{\mathrm{RK}})$, and sigmoid kernel $(\boldsymbol{\eta}_{\mathrm{SK}})$. Mathematically, these four well-known kernel functions of SVR, which are employed in this article can be expressed as Eqs. (2.4)-(2.7), respectively.

$$\boldsymbol{\eta}_{\mathrm{LK}}\left(\mathbf{X}_{obs},\mathbf{X}\right)=\mathbf{X}_{obs}^{'}\mathbf{X} \tag{2.4}$$

$$\boldsymbol{\eta}_{\mathrm{PK}}\left(\mathbf{X}_{obs},\mathbf{X}\right)=d^{-1}\left(\mathbf{X}_{obs}^{'}\mathbf{X}+\Delta\right)^{\gamma} \tag{2.5}$$

$$\boldsymbol{\eta}_{\mathrm{RK}}\left(\mathbf{X}_{obs},\mathbf{X}\right)=\exp\left(-d^{-1}\left\|\mathbf{X}_{obs}-\mathbf{X}\right\|^{2}\right) \tag{2.6}$$

$$\boldsymbol{\eta}_{\mathrm{SK}}\left(\mathbf{X}_{obs},\mathbf{X}\right)=\tanh\left(d^{-1}\left(\mathbf{X}_{obs}^{'}\mathbf{X}\right)+\Delta\right) \tag{2.7}$$

where $\mathbf{X}_{obs}$ is a vector of the years, Δ is the parameter of the kernel function, γ is the degree of polynomials, and $\tanh(\cdot)$ represents the hyperbolic tangent function. Mathematically, the Δ in $\boldsymbol{\eta}_{\mathrm{PK}}$ and $\boldsymbol{\eta}_{\mathrm{SK}}$ allows for adjusting the independent term in the kernel function, creating non-symmetric kernel functions. Particularly, the effectiveness comparison of Δ ranging from 0 to 3 is employed in $\boldsymbol{\eta}_{\mathrm{PK}}$, and Δ ranging from 0 to 5 is employed in $\boldsymbol{\eta}_{\mathrm{SK}}$, respectively. Moreover, the effectiveness comparison of γ ranging from 0 to 6 is also presented in this article.

Due to the difficulty in identifying the well-suited kernel function that is best suited for the time-series dataset acquired in this study, an effective comparison of these kernel functions is carried out. This comparison utilizes internal validation based on the training set and selected GoF statistics, including root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). In mathematics, RMSE (A_1), MAE (A_2), and MAPE (A_3) can be expressed as Eqs. (2.8)-(2.10), respectively.

$$A_1=\left(n^{-1}\sum\left(vec\left(\mathbf{Y}_{obs}\right)-vec\left(\mathbf{Y}_{pred}\right)\right)^{2}\right)^{0.5} \tag{2.8}$$

$$A_2=n^{-1}\sum\left|vec\left(\mathbf{Y}_{obs}\right)-vec\left(\mathbf{Y}_{pred}\right)\right| \tag{2.9}$$

$$A_3=100n^{-1}\sum\left|\left(vec\left(\mathbf{Y}_{obs}\right)-vec\left(\mathbf{Y}_{pred}\right)\right)/vec\left(\mathbf{Y}_{obs}\right)\right| \tag{2.10}$$

where $vec(\cdot)$ represents the vectorization.

2.4 Evaluation and Deployment

In the context of the CRISP-DM data science methodology, the principal focus of the evaluation phase is to assess the superiority of the trained AI time-series algorithms utilizing the test set and GoF statistics in Eqs. (2.8)-(2.10). To identify the best-performing simple stacked ensemble SVR-based time-series algorithm, this article proposes ranking the superiority based on three criteria, including the average of GoF statistics respectively acquired from the training and test sets and the absolute error of the average of GoF statistics between the training and test sets. This article desires to highlight that the multi-criteria decision-making (MCDM) method, including the Technique for Order of Preference by Similarity to Ideal Solution-based (TOPSIS-based) (Chuan *et al.*, 2018) and VIseKriterijuska Optimizacija I Komoromisno Resenje-based (VIKOR-based) (Chuan *et al.*, 2020) method are not taken into account in this article due to the irrationality of the analysis results acquired from these MCDM method. Specifically, the simple stacked ensemble SVR-based time-series algorithms with high GoF statistics are ranked as the superior AI time-series algorithm. In contrast, those with low GoF statistics are ranked lower.

Meanwhile, the principal focus of the deployment phase is to deploy the proposed superior AI time-series algorithm in forecasting the α-step ahead annual electricity consumption, which insightful predictive results may principally beneficial for the policymakers to establish more effective regulations, and take more effective actions for the energy sector that can be beneficial to economic, environmental, human, and social, which are the principal pillars of the Sustainable Development Goals (SDGs).

3. Analysis Results and Discussions

In this article, all the statistical analyses were entirely carried out utilizing R statistical software, described in the business understanding phase. In particular, Figure 3 and Table 1 respectively depict the graphical and numerical summary of annual electricity consumption in Malaysia from 1978 to 2020, corresponding to the economic activity sectors. These economic activity sectors include industrial ($S1$), commercial ($S2$), residential ($S3$), agriculture ($S4$), and transport ($S5$). The time-series utilized to train the proposed simple stacked ensemble SVR-based time-series algorithm in this study aggregated the annual electricity consumptions from $S1$, $S2$, $S3$, $S4$ and $S5$. Figure 3 and Table 1 revealed that the $S1$ economic activity sector has the highest electricity consumption on average, and conversely for the $S5$ economic activity sector. In Malaysia, the industrial economic sector, such as manufacturing, is the main contributor to Malaysia's economic growth, accounting for 9.5% of the Gross Domestic Product (GDP) in 2021, which decreased to -2.7% in 2020 (Department of Statistics Malaysia, 2022). This analysis result leads to the highest demand for electricity consumption in $S1$ to sustain industrial operations due to the high demand for industrial products (Ministry of Finance Malaysia, 2023), while the demand for electricity in $S2$, $S3$, $S4$ and $S5$ is lower. In contrast, most of the vehicles utilized in Malaysia consumed petrol vehicles (PV) instead of electrical vehicles (EV) due to the market price of EVs being much higher than PV and various obstacles such as battery capacity, charging station availability, and limited traveling distance (Muzir *et al.*, 2022; Yean, 2022). The lowest electricity consumption in $S5$ can be attributed based on the total number of PVs sold, which is much higher than EVs, especially for Perodua Myvi (Yean, 2022). Consequently, the transport economic activity sector had the lowest average electricity consumption. However, the transport economic activity sector is one of Malaysia's leading contributors to CO2 emissions, given that transport is a fundamental and essential infrastructure for economic development (Solaymani, 2022).

On the other hand, Figure 3 reveals that $S3$ and $S1$ respectively have the highest and lowest variability in terms of electricity consumption. However, this analysis result provided an inaccurate representation due to the high variation among the central measurements of $S1-S5$. Consequently, the L-Coefficient of Variation (L-CV) (Table 1) has been utilized to describe the variation corresponding to the economic activity sectors considered in this study. Specifically, Table 1 depicts that $S4$ and

*S*2 respectively have the highest and lowest relative variation in electricity consumption. These variations arise due to the dependency on usage demand in various economic activity sectors and socioeconomic growth (Pei *et al.*, 2016).

Meanwhile, Figure 3 also revealed that the shape of the distribution for all acquired time-series datasets is positively skewed, which conflicts subjectively with the analysis results of L-skewness and L-kurtosis presented in Table 1. This discrepancy arises due to the fact that all the results of the analysis of L-skewness and L-kurtosis are closer to zero. To address this conflict, the Shapiro-Wilk statistical hypothesis test has been employed, revealing that none of the acquired time-series datasets are normally distributed, including the annual electricity time-series data (total). However, the non-normality of the acquired time-series datasets does not affect the formulation of the proposed simple stacked ensemble SVR-based time-series algorithms in this article.

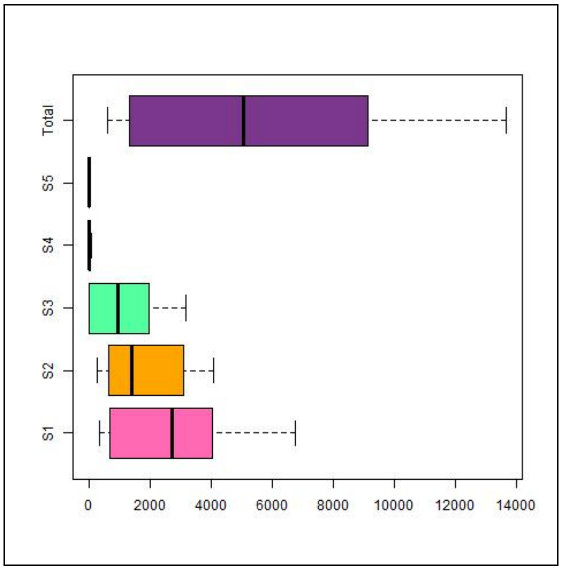

Figure 3. Exploring electricity consumption: Boxplot summaries by economic activity sectors

Table 1. Exploring electricity consumption: first four L-moments summaries by economic activity sectors

Economic Activity	Numerical summary (ktoe)				Shapiro-Wilk (p-value)
	L-mean	L-CV	L-skewness	L-kurtosis	
*S*1	**2769.1591**	0.4255	0.1427	-0.0090	0.0020
*S*2	1787.1136	**0.4069**	0.1617	-0.0647	0.0003
*S*3	1128.7955	0.5120	0.1563	-0.0340	0.0012
*S*4	13.0682	**0.7482**	0.5072	0.0967	0.0000
*S*5	**9.6591**	0.6857	0.4208	0.0606	0.0000
Total	**5707.7955**	**0.4366**	**0.1514**	**-0.0401**	**0.0010**

Table 2. Prediction performance of simple stacked ensemble ε-SVR and ν-SVR time-series algorithms for electricity consumption utilizing hold-out cross-validation

Algorithm	Insensitive Loss Function	Kernel	Training			Test			Rank
			A_1	A_2	A_3	A_1	A_2	A_3	
1	ε-SVR	Linear	591.0884	511.1750	36.7663	2640.0560	2397.9111	21.6525	15
2	ε-SVR	Polynomial00	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
3	ε-SVR	Polynomial01	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
4	ε-SVR	Polynomial02	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
5	ε-SVR	Polynomial03	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47

Algorithm	Insensitive Loss Function	Kernel	Training			Test			Rank
			A_1	A_2	A_3	A_1	A_2	A_3	
6	ε-SVR	Polynomial10	591.0884	511.1750	36.7663	2640.0560	2397.9111	21.6525	15
7	ε-SVR	Polynomial11	591.0885	511.1749	36.7663	2640.0555	2397.9105	21.6525	12
8	ε-SVR	Polynomial12	591.0885	511.1749	36.7663	2640.0558	2397.9108	21.6525	13
9	ε-SVR	Polynomial13	591.0885	511.1749	36.7663	2640.0558	2397.9108	21.6525	13
10	ε-SVR	Polynomial20	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
11	ε-SVR	Polynomial21	217.1114	172.8019	11.5263	1979.1126	1686.5619	15.0757	11
12	ε-SVR	Polynomial22	217.1162	172.7977	11.5246	1978.0538	1685.5748	15.0667	9
13	ε-SVR	Polynomial23	217.0537	172.6915	11.5050	1978.6242	1685.9581	15.0695	10
14	ε-SVR	Polynomial30	1029.0344	896.0925	66.7161	21152.3693	16796.5460	142.3327	61
15	ε-SVR	Polynomial31	187.3186	145.9430	10.0355	525.5437	453.0548	4.6980	6
16	ε-SVR	Polynomial32	187.4254	146.4372	10.1270	505.4873	441.5153	4.5874	4
17	ε-SVR	Polynomial33	187.4141	146.4299	10.1256	505.7985	441.6708	4.5887	5
18	ε-SVR	Polynomial40	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
19	ε-SVR	Polynomial41	155.1138	132.7465	9.2822	20830.1119	15505.8065	127.8584	62
20	ε-SVR	Polynomial42	154.9608	132.5594	9.2582	20837.5412	15511.0810	127.9011	63
21	ε-SVR	Polynomial43	155.1933	132.8801	9.2913	20857.5503	15526.1261	128.0261	64
22	ε-SVR	Polynomial50	1300.9035	1130.1258	72.6544	118587.7846	84547.5708	689.8737	71
23	ε-SVR	Polynomial51	144.6513	123.6100	8.8444	4242.8651	3704.4560	33.1435	28
24	ε-SVR	Polynomial52	144.6396	123.5543	8.8350	4299.0482	3754.8410	33.5529	30
25	ε-SVR	Polynomial53	144.5739	123.4804	8.8336	4267.1875	3726.3429	33.3201	29
26	ε-SVR	Polynomial60	2111.9294	1667.3889	76.6441	8780.9444	8485.0294	80.2194	47
27	ε-SVR	Polynomial61	152.3023	134.0926	9.8875	72628.2640	44766.7664	351.1898	67
28	ε-SVR	Polynomial62	147.1339	126.5195	8.7871	56588.4324	34616.0963	271.0826	65
29	ε-SVR	Polynomial63	147.1142	126.4821	8.7873	56822.6782	34770.6001	272.3103	66
30	ε-SVR	Radial Basis	174.8601	143.6850	9.1589	7113.0083	6283.7111	55.5757	46
31	ε-SVR	Sigmoid0	2139.8140	1799.8970	105.3810	13241.0080	12661.3057	118.4837	56
32	ε-SVR	Sigmoid1	5114.1474	4547.4612	281.2719	8876.3756	8851.4261	87.5944	31
33	ε-SVR	Sigmoid2	4638.9704	3792.7226	252.7749	15337.4716	15282.7544	151.3223	57
34	ε-SVR	Sigmoid3	2012.1558	1609.9946	84.4725	8143.6855	8038.5693	80.9168	43
35	ε-SVR	Sigmoid4	1873.3513	1460.2094	60.6649	4467.0280	3738.4668	35.0662	26
36	ε-SVR	Sigmoid5	2077.6942	1638.7820	74.4339	4828.2902	4478.1790	47.5062	27
37	ν-SVR	Linear	591.4326	514.3722	36.4659	2714.7401	2475.1795	22.4048	19
38	ν-SVR	Polynomial00	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
39	ν-SVR	Polynomial01	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
40	ν-SVR	Polynomial02	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
41	ν-SVR	Polynomial03	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
42	ν-SVR	Polynomial10	591.4326	514.3722	36.4659	2714.7401	2475.1795	22.4048	19

Algorithm	Insensitive Loss Function	Kernel	Training			Test			Rank
			A_1	A_2	A_3	A_1	A_2	A_3	
43	v-SVR	Polynomial11	591.4326	514.3722	36.4659	2714.7396	2475.1789	22.4048	17
44	v-SVR	Polynomial12	591.4326	514.3722	36.4659	2714.7400	2475.1794	22.4048	18
45	v-SVR	Polynomial13	591.4326	514.3722	36.4659	2714.7406	2475.1799	22.4048	21
46	v-SVR	Polynomial20	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
47	v-SVR	Polynomial21	204.5962	154.5386	8.7482	2386.2121	2060.6638	18.4476	24
48	v-SVR	Polynomial22	204.7640	155.0126	8.8096	2382.2093	2057.1981	18.4172	22
49	v-SVR	Polynomial23	204.8097	155.2215	8.8341	2386.0071	2060.8847	18.4514	23
50	v-SVR	Polynomial30	1018.1176	917.7205	61.7556	16923.3082	13173.4292	110.6201	60
51	v-SVR	Polynomial31	187.9912	141.3769	9.9572	415.4350	352.3095	3.5486	1
52	v-SVR	Polynomial32	188.1437	141.5263	9.9956	417.2176	351.8043	3.5419	2
53	v-SVR	Polynomial33	188.0339	141.4812	9.9807	416.5891	352.3169	3.5487	3
54	v-SVR	Polynomial40	1974.0579	1766.9259	113.6069	7980.3367	7653.5294	71.8555	37
55	v-SVR	Polynomial41	134.6000	76.5773	4.2875	12626.3062	9238.3976	75.5675	55
56	v-SVR	Polynomial42	133.0966	76.9926	4.2838	13323.6259	9752.4481	79.7976	59
57	v-SVR	Polynomial43	133.1872	76.9833	4.2820	13288.3393	9724.9978	79.5669	58
58	v-SVR	Polynomial50	1289.1320	1147.3179	78.5027	128795.5779	92146.4689	753.0239	72
59	v-SVR	Polynomial51	124.5840	79.1025	5.1283	4606.1172	3737.0180	31.5878	33
60	v-SVR	Polynomial52	124.4859	79.2116	5.1418	4561.4802	3707.1155	31.3581	32
61	v-SVR	Polynomial53	124.2459	79.1910	5.1452	4632.6211	3760.0786	31.7891	34
62	v-SVR	Polynomial60	1974.0579	1766.9259	113.6069	7980.33667	7653.52941	71.85545	36
63	v-SVR	Polynomial61	109.0720	53.1058	3.3695	84919.0835	53961.4693	426.8698	69
64	v-SVR	Polynomial62	108.8123	53.1381	3.3808	85222.2912	54165.8336	428.5077	70
65	v-SVR	Polynomial63	109.9402	55.2473	3.4825	77215.1257	49033.6687	387.8482	68
66	v-SVR	Radial Basis	161.6960	84.7390	4.2626	7056.4205	6214.1145	54.8612	45
67	v-SVR	Sigmoid0	1017.2201	826.2256	69.6440	1607.9273	1249.052	10.5878	7
68	v-SVR	Sigmoid1	3748.6288	3358.6702	215.8664	8811.2744	8661.4520	83.4359	35
69	v-SVR	Sigmoid2	3588.8270	3068.2610	194.0320	12617.2081	12497.0746	121.7557	54
70	v-SVR	Sigmoid3	1842.2663	1568.1254	88.0983	8179.5092	8049.5779	77.8916	44
71	v-SVR	Sigmoid4	1814.4390	1633.4990	101.4700	3222.3784	2984.1346	32.2298	8
72	v-SVR	Sigmoid5	1940.7084	1740.7857	111.4437	4127.6143	3848.2153	40.7347	25

***Note:** Polynomial32 represents third degrees η_{PK} with the corresponding value of $\Delta = 2$, while sigmoid0 represents η_{SK} with the corresponding value of $\Delta = 0$.

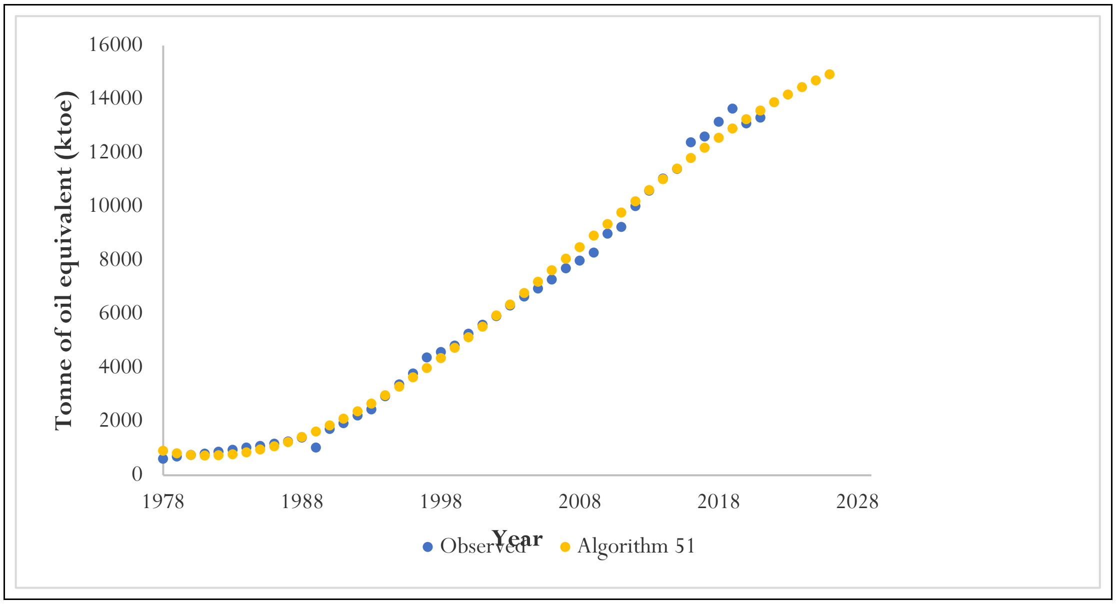

Figure 4. Predicted performance and forecast of electricity consumption period 1978-2026 utilizing training and test sets

In pursuit of this article's primary objective of data mining, Table 2 depicts the analysis results for the prediction performance of 72 simple stacked ensemble SVR-based time-series algorithms in predicting annual electricity consumption. Among the 72 time-series algorithms, Algorithms 1, 10, 30, and 31 proposed in the literature (Chong *et al.*, 2017; Shapi *et al.*, 2021) have been employed as benchmark comparisons. These algorithms were selected because of limited previous studies that utilized AI-based time-series algorithms for electricity consumption prediction in Malaysia. The analysis results presented in Table 2 revealed that simple stacked ensemble ν-SVR-based (Algorithms 51, 52, and 53) and simple stacked ensemble ε-SVR-based (Algorithms 15, 16, and 17) time-series algorithms, utilizing the $\boldsymbol{\eta}_{\mathrm{PK}}$ is outperformed other kernel functions such as $\boldsymbol{\eta}_{\mathrm{LK}}$, $\boldsymbol{\eta}_{\mathrm{RK}}$, and $\boldsymbol{\eta}_{\mathrm{SK}}$. These algorithms ranked in the top 6 for prediction performance among the 72 simple stacked ensemble SVR-based time-series algorithms considered in this article. Furthermore, the analysis results demonstrate that the formulated simple stacked ensemble ε-SVR-based time-series algorithm utilizing $\boldsymbol{\eta}_{\mathrm{PK}}$, improved the prediction performance of the formulated simple stacked ensemble SVR-based time-series algorithm in the literature.

In the literature, Chong *et al.* (2017) and Shapi *et al.* (2021) conveyed that the ε-SVR time-series algorithm, respectively associated with $\boldsymbol{\eta}_{\mathrm{RK}}$ and $\boldsymbol{\eta}_{\mathrm{LK}}$ is the superior time-series predictive algorithm for predicting the electricity consumption in smart commercialized buildings in Johor State and the annual electricity consumption in Malaysia. However, the analysis results depicted in this article were discrepant with those presented in the literature (Chong *et al.*, 2017; Shapi *et al.*, 2021). Nevertheless, the analysis results in this article remain rational as the annual electricity time series in Malaysia for the period 1978-2021 revealed a compound growth non-linear characteristic (Figure 1). This characteristic is consistent with the employed kernel function from a mathematical perspective. As a result, this divergent finding highlights that the utilization of the appropriate kernel function in the simple stacked ensemble SVR-based time-series algorithm is highly dependent on the intrinsic attributes of the acquired time-series dataset. By acknowledging the inherent characteristics of the acquired time-series data, researchers can enhance the accuracy and reliability of their predictive algorithms, thus contributing to more effective forecasting in the context of electricity consumption in smart commercialized buildings.

On the other hand, this study has deployed the top-ranked of the formulated simple stacked ensemble SVR-based time-series algorithm (Algorithm 51) to forecast 5-year future annual electricity consumption for 2022-2026 as depicted in Figure 4. This article highlighted that the initial value of the future value period 2022-2026 has been filled in utilizing the SLR predicted algorithm formulated based on the acquired time-series dataset for the period 1978-2021, in which the statistical analysis revealed a statistically significant $(P-value \approx 0.0000)$ linear association between the annual electricity consumption and the year. This step is to ensure the forecasting of 5-year future annual electricity consumption utilizing Algorithm 51 can be performed. Based on Figure 4, the annual electricity consumption is expected to experience continuous growth in the future. In other words, this forecasted 5-year future annual electricity consumption is rational and valid due to the expectation of the continual development of the Malaysian population in 2023 (Department of Statistics Malaysia, 2023) and accumulated technological advancements due to the arrival of the Industrial Revolution 4.0 (IR4.0). Therefore, policymakers establish more effective regulations and undertake beneficial actions for benefit for economic, environmental, human, and social.

4. Conclusions and Policy Implications

The short-term annual energy consumption is predominantly utilized for effective socio-economic development, energy saving, and environmental protection decision-making. Consequently, this study aims to formulate a novel simple stacked ensemble SVR-based time-series algorithm for forecasting annual electricity consumption in Malaysia for a limited dataset scenario utilizing CRISP-DM data science methodology. To authenticate the superiority of formulated AI time-series algorithms, a total of 72 simple stacked ensemble SVR-based time-series algorithms, including the benchmark comparison SVR-based time-series algorithms formulated in previous studies, were comprehensively evaluated and presented in this article. The principal analysis results revealed that the formulated simple stacked ensemble SVR-based time-series algorithm, based on the ν-insensitive loss function and the third-degree polynomial kernel outperformed the other 71 simple stacked ensemble SVR-based time-series algorithms taken into account in this article, including four benchmark comparison SVR-based time-series algorithms formulated in the literature. In conclusion, the insightful forecasted results acquired from the superior formulated simple stacked ensemble ν-SVR time-series algorithm can primarily be beneficial for policymakers in establishing more effective regulations and undertaking actions that can be beneficial to economic, environmental, human, and social, which are the principal pillars of the Sustainable Development Goals (SDGs) such as affordable and clean energy (SDG7), decent work and economic growth (SDG8), industry, innovation and infrastructure (SDG9), sustainable cities and communities (SDG11), responsible consumption and production (SDG12), and climate action (SDG13). To further enhance the accuracy and the superiority of the formulated simple stacked ensemble SVR-based time-series algorithm, this study proposes extending this work to incorporate other atmospheric, meteorological, and socioeconomic exogenous variables, such as wind, temperature, rainfall amount, population, economic trading activities, as potential exogenous variables in the future. This is because the aforementioned exogenous variables potentially affected the annual electricity consumption in Malaysia.

Declaration of Conflict of Interest

The author declared that there is no conflict of interest.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used ChatGPT in order to improve the readability and language of this work. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Acknowledgement

The author would like to thank the Energy Commission of Malaysia (Ministry of Science, Technology, and Innovation) for providing the annual electricity consumption dataset utilized in this study. The authors are also grateful for the valuable comments and suggestions from the reviewers. Furthermore, the authors would like to express their appreciation to the Ministry of Higher Education for the financial support under the Fundamental Research Grant Scheme (Grant No. FRGS/1/2024/SS01/UTEM/02/2).

Awards

This project achieved a place among the top five finalists in the Malaysian Industrial Mathematical Modelling Challenge (MIMMC2020) hosted by Universiti Teknologi Malaysia (UTM) Centre for Industrial and Applied Mathematics (UTM-CIAM). Our team was honored with the consolidation prize in the ultimate round among these leading teams. Further information is available at https://research.utm.my/ciam/mimmc2020/.

References

Abad, L. A., Sarabia, S. M., Yuzon, J. M., & Pacis, M. C. (2020). A short-term load forecasting algorithm using support vector regression & artificial neural network method (SVR-ANN). In Proceedings of the 11th IEEE Control and System Graduate Research Colloquium, Shah Alam, Malaysia. https://doi.org/10.1109/ICSGRC49013.2020.9232630

Brzozowska, J., Pizon, J., Baytikenova, G., Gola, A., Zakimova, A., & Piotrowska. (2023). Data engineering in CRISP-DM process production data-case study. Applied Computer Science, 19(3), 83-95. https://doi.org/10.35784/acs-2023-26

Chong, L. W., Rengasamy, D., Wong, Y. W., & Rajkumar, R. K. (2017). Load prediction using support vector regression. In Proceedings of the 2017 IEEE Region 10 Conference, Penang, Malaysia. https://doi.org/10.1109/TENCON.2017.8228016

Chuan, Z. L., Ismail, N., Yusoff, W. N. S. W., Fam, S-F., & Romlay, M. A. M. (2018). Identifying homogeneous rainfall catchments for non-stationary time series using TOPSIS algorithm and bootstrap k-sample Anderson darling test. International Journal of Engineering & Technology (UAE), 7(4), 3228-3237.

Chuan, Z. L., Deni, S. M., Fam, S-F., & Ismail, N. (2020). The effectiveness of a probabilistic principal component analysis model and expectation maximisation algorithm in treating missing daily rainfall data. Asia-Pacific Journal of Atmospheric Sciences, 56, 119-129. https://doi.org/10.1007/s13143-019-00135-8

Chuan, Z. L., Japashov, N., Yuan, S. K., Qing, T. W., & Ismail, N. (2024). Analyzing enrolment patterns: modified stacked ensemble statistical learning-based approach to educational decision-making. Akademika, 94(2), 232-251. https://doi.org/10.17576/akad-2024-9402-13

Chuan, Z. L., Wei, D. C. T., Aminuddin, A. S. B., Fam, S-F., & Ken, T. L. (in press). Comparison of multiple linear regression and multiple nonlinear regression models for predicting rice production. AIP Conference Proceedings.

Department of Statistics Malaysia. (2022). Gross Domestic Product (GDP) By State 2021. Retrieved from https://www.dosm.gov.my/portal-main/release-content/gross-domestic-product-gdp-by-state-2021 (accessed August 8, 2024).

Department of Statistics Malaysia. (2023). Demographic Statistics Malaysia First Quarter. Retrieved from https://www.dosm.gov.my/uploads/release-content/file_20230510164730.pdf (accessed August 8, 2024).

Energy Commission. (2024). Malaysia energy information hub [Statistics]. Retrieved from https://meih.st.gov.my/statistics

Farsi, B., Amayri, M., Bouguila, N., & Eicker, U. (2021). On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach. IEEE Access, 9, 31191-31212. https://doi.org/10.1109/ACCESS.2021.3060290

Guo, X., Zhao, Q., Zheng, D., Ning, Y., & Gao, Y. (2020). A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. Energy Reports, 6, 1046-1053. https://doi.org/10.1016/j.egyr.2020.11.078

He, X. J. (2018). Crude oil prices forecasting: time series vs. SVR models. Journal of International Technology and Information Management, 27(2), 25-42. https://doi.org/10.58729/1941-6679.1358

He, Y., Xu, Q., Wan, J., & Yang, S. (2016). Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. Energy, 114, 498-512. https://doi.org/10.1016/j.energy.2016.08.023

Her, O. Y., Mahmud, M. S. A., Abidin, M. S. Z., Ayop, R., & Buyamin, S. (2022). Artificial neural network based short-term electrical load forecasting. International Journal of Power Electronics and Drive Systems, 13(1), 586-593. https://doi.org/10.11591/ijpeds.v13.i1.pp586-593

Hong, W-C., Dong, Y., Zhang, W. Y., Chen, L-Y., & Panigrahi, B. K. (2013). Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. International Journal of Electrical Power & Energy Systems, 44(1), 604-614. https://doi.org/10.1016/j.ijepes.2012.08.010

Jifri, M. H., Hassan, E. E., & Miswan, N. H. (2017). Forecasting performance of time series and regression in modeling electricity load demand. In Proceedings of the 7th IEEE International Conference on System Engineering and Technology 2017, Shah Alam, Malaysia. https://doi.org/10.1109/ICSEngT.2017.8123412

Kamisan, N. A. B., Lee, M. H., Suhartono, S., Hussin, A. G., & Zubairi, Y. Z. (2018). Load forecasting using combination model of multiple linear regression with neural network for Malaysian city. Sains Malaysiana, 47(2), 419-426. http://dx.doi.org/10.17576/jsm-2018-4702-25

Kandananond, K. (2011). Forecasting electricity demand in Thailand with an artificial neural network approach. Energies, 4(8), 1246-1257. https://doi.org/10.3390/en4081246

Khanna, M., & Rao, N. D. (2009). Supply and demand of electricity in the developing world. Annual Review of Resource Economics, 1, 567-595. https://doi.org/10.1146/annurev.resource.050708.144230

Lee, C-M., & Ko, C-N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. Expert Systems with Applications, 38(5), 5902-5911. https://doi.org/10.1016/j.eswa.2010.11.033

Liang, C. Z., Sern, A. L. B., Cheng, T. C., Luen, D. L. K., Japashov, N., & Hiae, T. E. (2024). Empowering Industry 5.0: nurturing STEM tertiary education and careers through Additional Mathematics, in S. N. S. Al-Humairi. (Eds), Utilizing Renewable Energy, Technology, and Education for Industry 5.0. IGI Global, Pennsylvania, pp. 124-125. https://doi.org/10.4018/979-8-3693-2814-9

Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., & Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. Energy, 214, 118874. https://doi.org/10.1016/j.energy.2020.118874

Ministry of Finance Malaysia. (2023). Section 1: Economic Performance and Outlook. Retrieved from https://budget.mof.gov.my/pdf/belanjawan2023/economy-fiscal/section1.pdf (accessed August 8, 2024).

Miswan, N. H., Said, R. M., Anuar, S. H. H. (2016). ARIMA with regression model in modelling electricity load demand. Journal of Telecommunication, Electronic and Computer Engineering, 8(12), 113-116.

Miswan, N. H., Said, R. M., Hussin, N. H., Hamzah, K., & Ahmad, E. Z. (2016). Comparative performance of ARIMA and DES models in forecasting electricity load demand in Malaysia. International Journal of Electrical & Computer Sciences, 16(1), 6-9.

Muneer, A., Ali, R. F., Almaghthawi, A., Taib, S. M., Alghamdi, A., & Ghaleb, E. A. A. (2022). Short term residential load forecasting using long short-term memory recurrent neural network. International Journal of Electrical and Computer Engineering, 12(5), 5589-5599. https://doi.org/10.11591/ijece.v12i5.pp5589-5599

Muzir, N. A. Q., Mojumder, M. R. H., Hasanuzzaman, M., & Selvaraj, J. (2022). Challenges of electric vehicles and their prospects in Malaysia: A comprehensive review. Sustainability, 14(14), 8320. https://doi.org/10.3390/su14148320

Nagi, J., Yap, K. S., Nagi, F., Tiong, S. K., & Ahmed, S. K. (2011). A computational intelligence scheme for the prediction of the daily peak load. Applied Soft Computing, 11(8), 4773-4788. https://doi.org/10.1016/j.asoc.2011.07.005

Pei, T. L., Shaari, M. S., & Ahmad, T. S. T. (2016). The effects of electricity consumption on agriculture, service and manufacturing sectors in Malaysia. International Journal of Energy Economics and Policy, 6(3), 401-407.

Ping, Y. Y., & Kamarudin, A. N. (2022). Forecasting the electricity demand in Malaysia using ARIMA model. In Proceedings of Science and Mathematics, Johor, Malaysia.

Rätz, M., Javadi, A. P., Baranski, M., Finkbeiner, K., & Müller, D. (2019). Automated data-driven modeling of building energy systems via machine learning algorithms. Energy and Buildings, 202, 109384. https://doi.org/10.1016/j.enbuild.2019.109384

Razak, F. A., Shitan, M., Hashim, A. H., & Abidin, I. Z. (2009). Load forecasting using time series models. Jurnal Kejuruteraan, 21, 53-62. https://doi.org/10.17576/jkukm-2009-21-06

Rohmah, M. F., Putra, I. K. G. D., Hartati, R. S., & Ardiantoro, L. (2021). Comparison four kernels of SVR to predict consumer price index. Journal of Physics: Conference Series, 1737, 012018. https://doi.org/10.1088/1742-6596/1737/1/012018

Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: case study in Malaysia. Developments in the Built Environment, 5, 100037. https://doi.org/10.1016/j.dibe.2020.100037

Solaun, K., & Cerdá, E. (2019). Climate change impacts on renewable energy generation. A review of quantitative projections. Renewable and Sustainable Energy Reviews, 116, 109415. https://doi.org/10.1016/j.rser.2019.109415

Solano, J. A., Cuesta, D. J. L., Ibáñez, S. F. U., & Coronado-Hernández, J. R. (2022). Predictive models assessment based on CRISP-DM methodology for students performance in Colombia-Saber 11 Test. Procedia Computer Science, 198, 512-517. https://doi.org/10.1016/j.procs.2021.12.278

Solaymani, S. (2022). CO_2 emissions and the transport sector in Malaysia. Frontiers in Environmental Science, 9, 774164. https://doi.org/10.3389/fenvs.2021.774164

Sulandari, W., Yudhanto, Y., & Rodrigues, P. C. (2022). The use of singular spectrum analysis and k-means clustering-based bootstrap to improve multistep ahead load forecasting. Energies, 15(16), 5838. https://doi.org/10.3390/en15165838

Tan, C. S., Maragatham, K., & Leong, Y. P. (2013). Electricity energy outlook in Malaysia. IOP Conference Series: Earth and Environmental Science, 16, 012126. https://doi.org/10.1088/1755-1315/16/1/012126

Yean, T. S. (2022). Electric Vehicles in Malaysia: Moving Uphill from Niche to Mass Market. Retrieved from https://fulcrum.sg/electric-vehicles-in-malaysia-moving-uphill-from-niche-to-mass-market/ (accessed August 8, 2024).