Engineering and Applied Science Letters

PSR Press

Article

On the Robustness of the Olanrewaju-Olanrewaju Regression Kernel-Based to Nonparametric Kernels for Support Vector Regressor (SVR)

Rasaki Olawale Olanrewaju[1,*], Toyin Omoyeni Oguntola[2], Lukman Abiodun Nafiu[3] and Sodiq Adejare Olanrewaju[4]

1 Business Analytics Value Networks (BAVNs), Africa Business School (ABS), Mohammed VI Polytechnic University (UM6P), X4JH+QJR Avenue Mohamed Ben Abdellah Regragui, Rabat 10112, Morocco
2 Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria
3 School of Graduates and Research at Metropolitan International University (MIU), Kampala, Uganda
4 Department of Statistics, University of Ibadan, Ibadan, University of Ibadan, 900001, Oyo State, Nigeria
* Correspondence: farooqhamdani@gmail.com

Received: 06 December 2023; Accepted: 20 February 2024; Published: 30 June 2024.

Abstract: In this article we studied and juxtaposed nonparametric Least Square and the Olanrewaju-Olanrewaju regression-type $L_{(O-O)\lambda_{\gamma(|\theta|)}}$ kernels for supervised Support Vector Regressor (SVR) machine learning of hyperplane regression in a bivariate setting. The nonparametric kernels used to expound the SVR were Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular. Lagrangian multiplier estimation technique was adopted in estimating the involved SVR hyperplane regression coefficients as well as other embedded coefficients in each of the stated kernels. In addition, point estimate of the Euclidean distance (r) and error margin (d) in each of the SVR kernels were carved-out. In demonstration to the annual birthrate and its percentage change ($\Delta\%$) of the Nigeria populace from 1950 to 2023, the Olanrewaju-Olanrewaju regression-type kernel for SVR robustly outperformed the nonparametric and Least Square kernel-based SVRs with a miniature Cross-Validation index of -1205.49. 5.9% and 3.2% hyperplane estimated regression coefficients from the Olanrewaju-Olanrewaju kernel-based SVR were recorded for the annual birthrate and its percentage change ($\Delta\%$) respectively. Interpretably, this connotes that for every one percent increment in the annual birthrate per 1000, the mean rate of the Nigeria populace from 1950 to 2023 increased by 5.9% while other variables were held constant. Similarly, its percentage change per 1000 increased by 3.2% while other variables were held constant. In recommendation, the nonparametric and Olanrewaju-Olanrewaju regression-type SVRs as well as the Least Square SVR were pinpointed for future consideration of categorical, missing and zero bivariate observations.

Keywords: Kernels, Lagrangian Multiplier, Least Square, Olanrewaju-Olanrewaju Regression-Type, Nonparametric Regression, Support Vector Regressor (SVR).

Support Vector Machine (SVM) is a Machine Learning (ML) method first proposed by [1] for capturing the scope of statistical theory, structural noise (risk) minimization, forecasting, construction of intelligent machines, regression estimation problems of classifying classifiers, feature spaces, dependency estimation problem, etc. [1,2]. SVMs are classified as one of the evolutional trends in designing statistical model techniques for modern, conventional, and operational ML methods [3]. SVMs were designed for extraction of trends and deduction from empirical observations, especially dependency observations that usually emanate from pattern recognition, demographic studies, neural networks etc. [4,5]. Whenever kernels like Gaussian, Radial Basis (RB), Gaussian additive, polynomial, sigmoid, hyperbolic are applied to SVM, such SVM is usually regarded as Kernel Machine (KM) [6].

According to [7], SVMs made it possible to deal with any form of non-linear classification and transformation identified to be a new generation set of ML. In many ways, incorporation of kernels makes sub-symbolic ML approaches like artificial neural networks deepening. However, in many instances these sub-symbolic learning approaches are new formation of algorithms with common statistical methods rather than the ordinary classical algorithms [8,9]. These new generations set of ML methods makes it possible to find

solutions to analogical methods peculiar to large-scale SVMs, Sequential Minimal Optimization (SMO) for solving quadratic programming problem that arises in parallel analogical arrangement of solving large-scale SVMs [10].
Another uniqueness of the SVMs is that they are new techniques eligible for dichotomous classifications that comprise components of ML, neural networks, statistical inference, and non-parametric methods [11]. Similarly to mathematical financial stochastic processes, SVMs as well possess the ability to assort financial firms to be either insolvent or solvent based on their score values. These score values by SVMs are usually in relation to some financial ratios and indexes (It is to be noted that this relationship is neither parametric nor linear). SVMs are sometimes used in evaluating financial ratios that can be transformed before evaluation by classical inference techniques [12]. The introduction of kernels to SVM makes it gain pliability in the selection of the signifier that separates insolvent from solvent companies. This makes it to usually demand for either linearity or non-linearity against the functionality of any data, since kernels are distribution-free kind of ML technique that operates in local optima [13]. Notably, SVMs also make it possible to estimate reliable out-of-sample predictions when trained and untrained data are combined. These predictions are usually more reliable whenever the associated smoothness and penalty parameters in the SVM setting are appropriately chosen [14,15]. This implies that by choosing appropriately the smoothness and penalty parameters; SVMs will be robust even when the trained data is biased. This makes SVMs to always deliver a unique solution because the optimality problem is always convex. This is a singular merit compare to neural networks, that usually have many solutions associated to its local minima. The numerous solutions associated to neural networks' local minima make it to lose its robustness over different samples [16].

[17] pinpointed that ML methods designed for SVMs do pose the lacuna of combining vector machine classifiers. This is because each SVM model clutches different set of attributes in multiclass classifications. Consequently, whenever SVM is used for regression analysis, it is usually referred to as Support Vector Regressor (SVR) [18]. Elaborately, SVR is a kernel-based regression technique that maps both linear and non-linear separable data (either in a univariate, bivariate or multivariate setting) in a real number space to higher-dimensional space with the use of kernels. SVR is being controlled by two parameters, namely, the regularization parameter (λ) and the gamma parameter (γ) [19]. The regularization parameter (λ) controls biasedness and variance trade-off in SVR. Smaller values of the regularized parameter usually lead to model under-fitting, while larger values do lead to model over-fitting.

In this article we shall subject the SVR to a transformable bivariate data framework of high-dimensional space of nonparametric kernels of Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular and juxtapose with the Least Square and Olanrewaju-Olanrewaju regression-type kernels. Lagrangian multiplier estimation technique shall be adopted to estimate involved coefficients. The Euclidean distance (r) and error margin (d) in the parallel hyperplane in each of these SVRs kernel-based shall be worked-out, as well as the theoretical asymptotic generalization of biasedness and variance trade-off from kernel perspective. Demonstratively, this write-up shall be applied to annual birthrate and its percentage change $(\Delta\%)$ of the Nigeria populace from 1950 to 2023 for deductive inference, plans and policymaking.

1. Materials and Methods

This section defines some basic notions and symbols needed for transforming and mapping nonlinearized Support Vector Regression (SVR) problem to linearized Support Vector Regression (SVR). This section also provides the needed formulae for estimating Euclidean distance between nearest points of hyperplane, as well as the error margin between parallel bounding planes. In addition, the optimality condition needed for Least Square-SVR parameter estimation via Lagrangian multiplier effect will be provided in this section.

According to [20], SVR classifies linear and non-linear ML regression on hyperplane, such that linearized SVR solves any regression problems even while dealing with non-linearized regression problems. It maps the original vector of observations, say X, into a higher-dimensional feature space via mapping function; say "Ω", in order to make the higher-dimensional feature space linear. Mathematically, let "d" be the higher-dimensional feature of finitely many space "Ω", such that, $\Omega \subseteq \mathbb{R}^{g,n}$ and $X \in \mathbb{R}^d$ is the random covariate vector through some many and many-to-one mapping $\alpha : \mathbb{R}^d \to \Omega$, otherwise $\alpha : \mathbb{R}^d \to \mathcal{H}$ and "ng" of $\mathcal{H}$ dimensional space. $\mathcal{H}$ is an Hilbert space, while "n" and "g" are the sample size and Lebesgue Probability Density Function (PDF) respectively. The linear function below generalizes SVR for regression:

$$G_{\Omega,n} = \Big\{ g : g(X) = w^T \alpha(X) + \theta;\ \alpha : \mathbb{R}^d \to \Omega;\ \theta \in \mathbb{R};\ \omega \in \mathbb{R}^{n_g} \tag{1}$$

Where "θ" and "ω" are the regression coefficients and mapping observational matrix respectively. Given a trained bivariate dataset of $D_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$, where $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ with regression model class of $G_{\Omega,n}$ defined in equation (1), then the Least Square-SVR model can be mathematically represented as:

$$\begin{cases} \widehat{\theta} = \min\limits_{\theta,\varepsilon,\omega}(\omega,\theta,\varepsilon)\dfrac{\omega^T\omega}{2} + \dfrac{\gamma}{2}\sum\limits_{i=1}^{n}\varepsilon_i^2 \\ s.t \quad Y_i = \omega^T\alpha(X_i) + \theta + \varepsilon_i \ \Rightarrow\ \varepsilon_i = Y_i - \omega^T\alpha(X_i) - \theta \qquad i = 1,\ldots,n \end{cases} \tag{2}$$

Where "γ " is the regularized parameter. For transformed data (e.g. transformation required in exponentiated and Poisson regression) or kernel-based function (e.g. Polynomial kernel, Gaussian Radial Basis Function (Gaussian RBF)). D_n could be transformed into $D_n^* = \{(\alpha(X_1), Y_1), (\alpha(X_2), Y_2), \ldots, (\alpha(X_n), Y_n)\}$. It is to be noted that $\frac{\omega^T\omega}{2} \approx \frac{1}{2}\|\omega\|^2$. The Euclidean distance between nearest points and hyperplane which is:

$$r = \left| \frac{\min\limits_{\theta,\varepsilon,\omega}(\omega,\theta,\varepsilon)}{\|\omega\|} \right| \tag{3}$$

such that the error margin between parallel bounding planes, say "d", can be defined as $d = \frac{2}{\|\omega\|} = 2r$. "$\alpha$" is usually unknown and can be estimated using the Lagrangian multiplier set-up. The Lagrangian effect of equation (2) is:

$$ML(\theta,\omega,\varepsilon:\delta) = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\delta_i\Big\{\theta + \varepsilon_i + \omega^T\alpha(X_i) - Y_i\Big\} \tag{4}$$

Then the condition for optimality for Least Square-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is:

$$\begin{cases} \sum\limits_{i=1}^{n}\delta_i = 0 & \Leftarrow \dfrac{\partial ML}{\partial\theta} = 0 \ \ \forall\ i = 1,\ldots,n \\ \omega - \sum\limits_{i=1}^{n}\delta_i\alpha(X_i) & \Leftarrow \dfrac{\partial ML}{\partial\omega} = 0 \ \ \forall\ i = 1,\ldots,n \\ \theta + \omega^T\alpha(X_i) - Y_i + \varepsilon_i = 0 & \Leftarrow \dfrac{\partial ML}{\partial\delta_i} = 0 \ \ \forall\ i = 1,\ldots,n \\ \delta_i = \varepsilon_i\gamma & \Leftarrow \dfrac{\partial ML}{\partial\varepsilon_i} = 0 \ \ \forall\ i = 1,\ldots,n \end{cases} \tag{5}$$

1.1. The Nonparametric Kernels-Based for Support Vector Regressor (SVR)

This section presents the optimality condition needed for parameter estimation of the nonparametric SVR-kernels of Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular via Lagrangian multiplier effect. The condition for optimality for Bisquare Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is:

$$ML(\theta,\omega,\varepsilon:\delta) = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\delta_i\left\{\theta + \varepsilon_i + \omega^T\alpha\left(K\left(\frac{X_i - x}{h_l}\right)I_{\{F_i=l\}}\right) - Y_i\right\} \tag{6}$$

where:

$$K\left(\frac{X_i - x}{h_l}\right)I_{\{F_i=l\}} = \frac{\frac{15}{16}\left\{1 - 2\|X_i - X_j\|^2 + \|X_i - X_j\|^4 I_{[-1,\,1]}\|X_i - X_j\|\right\}}{h_l} \qquad \forall\ \{i,j\} = 1,\ldots,n \tag{7}$$

h_l is the smoothing parameter otherwise known as bandwidth of "l"the individual covariate.

$$\begin{cases} \sum_{i=1}^{n} \delta_i = 0 & \Leftarrow \frac{\partial ML}{\partial \theta} = 0 \ \forall \ i = 1, \ldots, n \\ \omega - \sum_{i=1}^{n} \delta_i \alpha \left(\frac{15}{16} \left\{ 1 - 2 \left\| X_i - X_j \right\|^2 + \left\| X_i - X_j \right\|^4 I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\} \right) & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \ \forall \ \{i, j\} = 1, \ldots, n \\ \theta + \omega^T \alpha \left(\frac{15}{16} \left\{ 1 - 2 \left\| X_i - X_j \right\|^2 + \left\| X_i - X_j \right\|^4 I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\} \right) - Y_i + \varepsilon_i = 0 & \\ & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \ \forall \ \{i, j\} = 1, \ldots, n \\ \delta_i = \varepsilon_i \gamma & \Leftarrow \frac{\partial ML}{\partial \varepsilon_i} = 0 \ \forall \ i = 1, \ldots, n \end{cases} \tag{8}$$

The condition for optimality for Gaussian Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$K\left(\frac{X_i - x}{h_l}\right) I_{\{F_i = l\}} = (2\pi)^{-d}/2 \left(- \left\| X_i - X_j \right\|_2^2 / 2h_l^2 \right) \tag{9}$$

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \left\{ (2\pi)^{-d}/2 \left(- \left\| X_i - X_j \right\|_2^2 / 2h_l^2 \right) \right\} & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \ \forall \ \{i, j\} = 1, \ldots, n \\ \theta + \omega^T \alpha \left\{ (2\pi)^{-d}/2 \left(- \left\| X_i - X_j \right\|_2^2 / 2h_l^2 \right) \right\} - Y_i + \varepsilon_i = 0 & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \ \forall \ \{i, j\} = 1, \ldots, n \end{cases} \tag{10}$$

The condition for optimality for Triweight Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$K\left(\frac{X_i - x}{h_l}\right) I_{\{F_i = l\}} = \frac{\frac{35}{36} \left\{ 1 - 3 \left\| X_i - X_j \right\|^2 + 3 \left\| X_i - X_j \right\|^4 - \left\| X_i - X_j \right\|^6 \right\} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} \tag{11}$$

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \left(\frac{\frac{35}{36} \left\{ 1 - 3 \left\| X_i - X_j \right\|^2 + 3 \left\| X_i - X_j \right\|^4 - \left\| X_i - X_j \right\|^6 \right\} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} \right) \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \\ \theta + \omega^T \alpha \left(\frac{\frac{35}{36} \left\{ 1 - 3 \left\| X_i - X_j \right\|^2 + 3 \left\| X_i - X_j \right\|^4 - \left\| X_i - X_j \right\|^6 \right\} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} \right) - Y_i + \varepsilon_i = 0 \\ \qquad \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \forall \ \{i, j\} = 1, \ldots, n \end{cases} \tag{12}$$

The condition for optimality for Uniform Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$K\left(\frac{X_i - x}{h_l}\right) I_{\{F_i = l\}} = \frac{\frac{1}{2} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} \tag{13}$$

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \frac{\frac{1}{2} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \ \forall \ \{i, j\} = 1, \ldots, n \\ \theta + \omega^T \alpha \frac{\frac{1}{2} \left\{ I_{[-1,\,1]} \left\| X_i - X_j \right\| \right\}}{h_l} - Y_i + \varepsilon_i = 0 & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \quad \forall \{i, j\} = 1, \ldots, n \end{cases} \tag{14}$$

The condition for optimality for Epanechnikov Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$K\left(\frac{X_i - x}{h_l}\right) I_{\{F_i = l\}} = \frac{\frac{3}{4}\left\{1 - \|X_i - X_j\|^2\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l} \tag{15}$$

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \left(\frac{\frac{3}{4}\left\{1 - \|X_i - X_j\|^2\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l}\right) & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \quad \forall\ \{i,\, j\} = 1, \ldots, n \\ \theta + \omega^T \alpha \left(\frac{\frac{3}{4}\left\{1 - \|X_i - X_j\|^2\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l}\right) - Y_i + \varepsilon_i = 0 & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \quad \forall\ \{i,\, j\} = 1, \ldots, n \end{cases} \tag{16}$$

The condition for optimality for Triangular Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$K\left(\frac{X_i - x}{h_l}\right) I_{\{F_i = l\}} = \frac{\left\{1 - \|X_i - X_j\|\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l} \tag{17}$$

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \left(\frac{\left\{1 - \|X_i - X_j\|\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l}\right) & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \quad \forall\ \{i,\, j\} = 1, \ldots, n \\ \theta + \omega^T \alpha \left(\frac{\left\{1 - \|X_i - X_j\|\right\}\left\{I_{[-1,\,1]}\ \|X_i - X_j\|\right\}}{h_l}\right) - Y_i + \varepsilon_i = 0 & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \quad \forall\ \{i,\, j\} = 1, \ldots, n \end{cases} \tag{18}$$

Noted: It is to be noted that equation (8), (10), (12), (14), (16), and (18) are the Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular nonparametric kernels specified for the SVR via Lagrangian multiplier effect.

1.2. The Olanrewaju and Olanrewaju Regressor-Type Kernel for Support Vector Regressor (SVR)

This section presents the fundamental notion of convexity of the Olanrewaju-Olanrewaju regression-type estimator in its real-value function and its embedded coefficients. The asymptotic limiting of the Olanrewaju-Olanrewaju regression-type estimator will also be presented in this section. Additionally, the optimality condition needed for the Olanrewaju-Olanrewaju-SVR parameter estimation via Lagrangian multiplier effect will be provided in this section.

Theorem 2.1

Let $f : \mathbb{R} \to \mathbb{R}$ be a real-value function that satisfies the appropriate characterization of the Olanrewaju-Olanrewaju penalized regression-type estimator of

$$f(x) = \lambda\left(\gamma + \frac{2}{\pi}\right) \arctan \frac{|\theta|}{\gamma} \frac{e^{\lambda} - e^{-\lambda}}{e^{\lambda} + e^{-\lambda}} \tag{19}$$

as proposed by [21]. Then the kernel penalty for the convex regression type-estimator of the Olanrewaju-Olanrewaju Regression-Type Estimator for SVR is

$$L_{(O-O)\lambda_{\gamma(|\theta|)}} = K\left(\frac{X_i - x}{h_l}\right) I_{\gamma\{F_i = l\}} = \frac{\lambda\left(\gamma + \frac{2}{\pi}\right) \arctan\left(\frac{|X_i - x|}{\gamma}\right) \tanh(\lambda)}{h_l} \tag{20}$$

$$= \frac{\lambda\left(\gamma + \frac{2}{\pi}\right) \arctan\left(\frac{|X_i - x|}{\gamma}\right) \frac{\sinh(\lambda)}{cosh(\lambda)}}{h_l}$$

$$= \frac{\lambda\left(\gamma+\frac{2}{\pi}\right)\arctan\left(\frac{|X_i-x|}{\gamma}\right)\left(\frac{e^{\lambda}-e^{-\lambda}}{e^{\lambda}+e^{-\lambda}}\right)}{h_l}$$

$\ni \lambda \geqslant 0$ (λ a non-negative value), usually 0.005, where "γ" is usually peg at 3.5 according to [22]. The condition for optimality for Olanrewaju-Olanrewaju Kernel-SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is SVR with $\delta_i \in \mathbb{R}$ as the Lagrangian multiplier effect is

$$\begin{cases} \omega - \sum_{i=1}^{n} \delta_i \alpha \left(\frac{\lambda\left(\gamma+\frac{2}{\pi}\right)\arctan\left(\frac{|X_i-x|}{\gamma}\right)\frac{\sinh(\lambda)}{cosh(\lambda)}}{h_l} \right) & \Leftarrow \frac{\partial ML}{\partial \omega} = 0 \quad \forall\ \{i,\, j\} = 1,\ldots,n \\ \theta + \omega^T \alpha \left(\frac{\lambda\left(\gamma+\frac{2}{\pi}\right)\arctan\left(\frac{|X_i-x|}{\gamma}\right)\frac{\sinh(\lambda)}{cosh(\lambda)}}{h_l} \right) - Y_i + \varepsilon_i = 0 & \Leftarrow \frac{\partial ML}{\partial \delta_i} = 0 \quad \forall\ \{i,\, j\} = 1,\ldots,n \end{cases} \tag{21}$$

such that $\sum_{i=1}^{n} \delta_i = 0$ and $\delta_i = \varepsilon_i \gamma$ for $\frac{\partial ML}{\partial \theta} = 0$ and $\frac{\partial ML}{\partial \varepsilon_i} = 0$ for all respectively and that $\forall\ i = 1,\ldots,n$. Coefficients "θ" and "δ" can be estimated via the following system after elimination of "ω" and "δ" in all the mentioned kernels.

$$\begin{bmatrix} 0 \\ \hline y \end{bmatrix} = \begin{bmatrix} \frac{\langle 0 | 1_n^T \rangle}{\left\langle 1_n \middle| \Psi + \frac{1}{\gamma} I_n \right\rangle} \end{bmatrix} \begin{bmatrix} \theta \\ \hline \delta \end{bmatrix} \tag{22}$$

With $\delta = (\delta_1,\ldots,\delta_n)^T$, $y = (Y_1,\ldots,Y_n)^T$ and $1_n = (1,\ldots,1)^T$. Using the Mercer's condition, the ij-element of Ψ is given by

$$\Psi_{ij} = \delta(X_i)^T \delta(X_j) = K\left(X_i,\, X_j\right) \quad i,j = 1,\ldots,n \tag{23}$$

Where Ψ is the positive semi-definite matrix with ij^{th}-element, such that $\Psi_{ij} = K\left(X_i X_j\right)$.

1.3. Algorithm

Input: Observations/samples/simulations stored in the trained bivariate dataset machine learning of either $\{D_1,\ldots,D_n\}$ or $\{D_1^*,\ldots,D_n^*\}$ with outer and inner number of iterations say "q" and "T", smooth function from Hilbert space "H", bandwidths $\{h_1,\ldots,h_n\}$ and regularization parameter "γ" that is usually 3.5 in magnitude.

1. For "g" that runs from outer dimension "q" proceed to
2. If $g = 1$, then
3. Compute the initial estimator based on D or D^*: $\widetilde{\theta}_0 = \min_{\theta,\varepsilon,\omega}(\omega,\theta,\varepsilon)\frac{\omega^T\omega}{2} + \frac{\gamma}{2}\sum_{i=1}^{n}\varepsilon_i^2$
4. Else
5. $\widetilde{\theta}_0 = \widetilde{\theta}^{(g-1)}$
6. end if
7. $\widetilde{\theta}^{(g,o)} = \widetilde{\theta}_0$ for $t = 1,\ldots,T$
8. $\widetilde{\theta}^{(g,o)} \Rightarrow \{D_1,\ldots,D_n\}$ or $\widetilde{\theta}^{(g,o)} \Rightarrow \{D_1^*,\ldots,D_n^*\}$
9. Compute $\widetilde{\theta} = \frac{1}{2}\omega^T\omega + \frac{\gamma}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}\delta_i\left\{\theta + \varepsilon_i + \omega^T\alpha\left(K\left(\frac{X_i-x}{h_l}\right)I_{\{F_i=l\}}\right) - Y_i\right\}$ and transfer it to the machine.
10. Then the central machine computes condition for optimality $\frac{\partial ML}{\partial \widetilde{\theta}}, \frac{\partial ML}{\partial \omega}, \frac{\partial ML}{\partial \delta_i}, \frac{\partial ML}{\partial \varepsilon}$
11. Estimates of $\widetilde{\theta}$, ω, δ_i and ε are extracted from each of the stated kernels $K\left(\frac{X_i-x}{h_l}\right)$

1.4. Asymptotic Biasedness and Variance from Kernel Perspective

$$\mathbb{E}\left[\widehat{g}(x)\right] = \int \frac{1}{h_l} K\left(\frac{X_i - x}{h_l}\right) g(x)\partial x \tag{24}$$

$$\text{Var}\left(\widehat{g}(x)\right) = \frac{1}{nh_l{}^2} Var\left(K\left(\frac{X_i - x}{h_l}\right)\right) = \frac{1}{nh_l{}^2}\mathbb{E}\left[K\left(\frac{X_i - x}{h_l}\right)^2\right] - \frac{1}{nh_l{}^2}\mathbb{E}\left[K\left(\frac{X_i - x}{h_l}\right)\right] \tag{25}$$

$$= \left[\int \frac{1}{h_l{}^2} K\left(\frac{X_i - x}{h_l}\right)^2 g(x)\partial x - \frac{1}{n}\left(\int \frac{1}{h_l} K\left(\frac{X_i - x}{h_l}\right)\right)^2\right] \times \frac{1}{n} \tag{26}$$

So,

$$\text{Bias}(x) = \int \frac{1}{h_l} K\left(\frac{X_i - x}{h_l}\right) g(x)\partial x - g(x) \tag{27}$$

Let

$$\rho = \frac{(X_i - x)}{h_l}$$

$$\text{Bias}(x) = \int K(\rho) g(X_i + h_l\rho) - g(x) = \int K(\rho)\left(g(X_i + h_l\rho) - g(x)\right)\partial\rho \tag{28}$$

Assuming a Taylor expansion of "g" and that "g" is sufficiently smooth such that

$$g(X_i + h_l\rho) = g(x) + h_l\rho\partial g(x) + \frac{1}{2}h_l^2\rho^2\partial^2 g(x) + \dots \tag{29}$$

Inserting equation (29) into equation (28), biasedness of equation (28) could be written as

$$\text{Bias}(x) = h_l\partial g(x)\int \rho K(\rho)\partial\rho + \frac{1}{2}h_l^2\partial^2 g(x)\int \rho^2 K(\rho)\partial\rho + \dots \tag{30}$$

$$\int \rho K(\rho)\partial\rho \approx 0$$

This implies that

$$\text{Bias}(x) = \frac{1}{2}h_l^2\partial^2 g(x)\int \rho^2 K(\rho)\partial\rho + \text{higher order terms of the bandwidth} \tag{31}$$

The variance from equation (25) becomes

$$Var\left(\widehat{g}(x)\right) = \frac{1}{n}\int \frac{1}{h_l{}^2} K\left(\frac{X_i - x}{h_l}\right)^2 g(x)\partial x - \frac{1}{n}\left(g(x) + \text{Bias}(x)\right)^2 \tag{32}$$

$$= \frac{1}{nh_l}\int g(X_i - \rho h_l) K(\rho)^2\partial\rho - \frac{\left(g(x) + \text{Bias}(x)\right)^2}{n} \tag{33}$$

$$= \frac{1}{nh_l}\int g(X_i - h_l\rho) K(\rho)^2\partial\rho + O(n^{-1}) \tag{34}$$

Where,

$$O(n^{-1}) = \frac{\left(g(x) + \text{Bias}(x)\right)^2}{n}$$

$$Var\left(\widehat{g}(x)\right) = \frac{1}{nh_l} g(x)\int K(\rho)^2 + O(n^{-1}) \tag{35}$$

If "g" is smooth, hence $g\left(X_i - h_l\rho\right) \to g(X_i)$ as $h_{l_n} \to 0$. In fact, $h_{l_n} \to 0$, as $h_l \to 0$, $n \to \infty$ for sample size (n). This literally connotes that

$$\text{Bias}(x) = h_l^2\partial^2 g(x)\int \frac{\rho^2 K(\rho)\partial\rho}{2} + O\left(h_l^2\right);$$

$$Var\left(\widehat{g}(x)\right) = \frac{g(x)}{nh_l}\int K(\rho)^2\partial\rho + O\left((nh_l)^{-1}\right) \quad \text{as} \quad n \to \infty$$

2. Result

The dataset used to validate the nonparametric SVRs via its stated kernels, as well as the Least Square and Olanrewaju-Olanrewaju SVRs is the annual birthrate per 1000 of the Nigeria populace from 1950 to midyear of 2023 simultaneously with its annual percentage change (Δ%) per 1000. The simultaneously merged dataset was extracted from the National Bureau of Statistics (NBS), from her National Data Archive (NADA) Centre. Consequently, the annual birthrate per 1000 of the Nigeria populace and its annual percentage change (Δ%) will be analyzed as covariates to its dependent annual year of record.

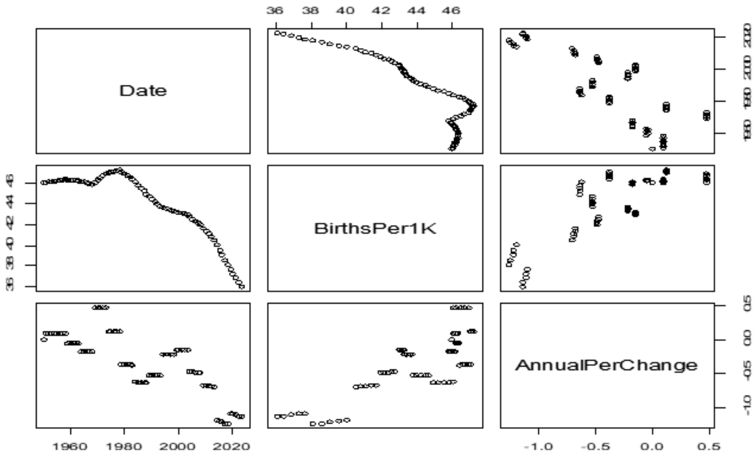


Figure 1. Scatter Plot of the Annual Birthrate and its Percentage Change per (Δ%) 1000 of the Nigeria Populace.

Scatterplot shows the relationship between two continuous (quantitative: interval and ratio scaled measured) variables. The data points of one of the studied variables appear on the horizontal axis, while the data points of the other variable appear on the vertical axis. Each numeracy appears as a point on the graph. For the annual birthrate per 1000 trend on the left hand side of the mid plot, the trend possessed an overall pattern that strikes departure from above in a downward decremented manner. Consequently, the annual birthrate per 1000 trend of the Nigeria populace from 1950 to midyear of 2023 possessed a highly negative correlation with respect to the annual years of studied. In a similar manner, the annual percentage change(Δ%) of the birthrate per 1000 with respect to the annual years of studied possessed an evenly spread-out in a decremented downward manner (highly negative correlation with respect to the annual birth rate). It is to be noted that the closer the annual birthrate per 1000 data points lie together to make a line, the higher the correlation.

Table 1. Nonparametric, Least Square, and Olanrewaju-Olanrewaju Kernel-Based SVRs

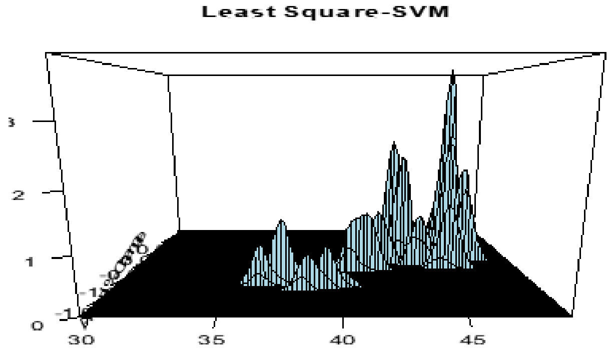
	Least Square-SVR	Bisquare Kernel-SVR	Gaussian Kernel-SVR	Triweight Kernel-SVR
h_{x_1}	0.273(0.238)	0.283(0.247)	0.273(0.238)	0.273(0.238)
h_{x_2}	0.007(0.033)	0.013(0.061)	0.007(0.033)	0.007(0.033)
$C.V$	10.777	-7.098	10.777	10.777
Bias	0.034(0.0051)	0.0534(0.002)	0.045(0.002)	0.037(0.0089)

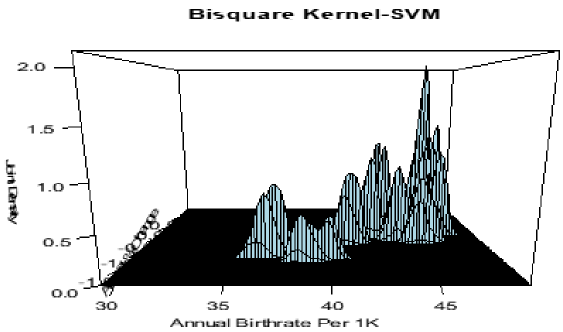
Epanechnikov Kernel-SVR	Triangular Kernel-SVR	Olanrewaju-Olanrewaju–SVR
0.268(0.233)	0.283(0.247)	0.059(0.242)
0.018(0.081)	0.014(0.061)	0.005(0.032)

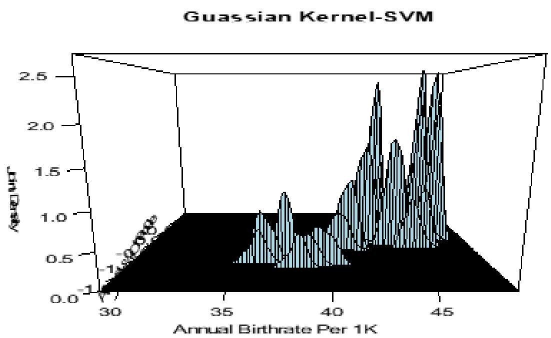
Keys: C.V. = Cross Validation; h_{x_1} = bandwidth of x_1; h_{x_1} = bandwidth of x_2.

Cross-Validation (C.V.) is one of the key indexes for testing ML performance. Alternatively, C.V. is a resampling procedure used to evaluate ML models on a limited data sample. C.V. at times is sometimes referred to as rotational estimation of various similar model validation techniques for assessing statistical analysis of generalized independent data set. It is basically meant for estimating the pragmatic performance of a predictive model, the stability of the involved parameters as well setting the goals of prediction. Notably, Bisquare, Uniform, Epanechnikov, Triangular, and Olanrewaju-Olanrewaju Kernel-SVRs gave a negative C.V.'s of $-7.098, -48.779, -30.711, -7.0984,$ and $-$ 7.0984 respectively for the distributional-free machine-learning SVR analyzes via the SVM techniques. It is an already established fact that the smaller (a reduced negative value) the C.V. the betterment and ideal the model that produced the smallest C.V. index, similarly to that of the interpretation of other continuous-type model performance indexes like AIC, BIC, and its variants. The Olanrewaju-Olanrewaju kernel-based for the SVR produced the miniature C.V. of -1205.49 with bandwidths 0.059 (0.242) and 0.005 (0.032) for the covariates annual birthrate (x_1) and percentage change (Δ%) per 1000 (x_2) respectively, while in brackets are standard errors of estimates. Interpretably from the estimated coefficients of the Olanrewaju-Olanrewaju kernel-based SVR, for every one percent increment in the annual birthrate per 1000, the mean rate of the Nigeria populace from 1950 to midyear of 2023 increased by 5.9% such that other variables were held constant. Its percentage change (Δ%) per 1000 increased by 3.2% such that other variables were held constant.

Another notable point of model performance is the asymptotic kernel's biasedness and variability. It is noted that the Bisquare, Uniform, Epanechnikov, Triangular, and the Olanrewaju-Olanrewaju kernel-based SVRs produced the smallest asymptotic biasedness of 0.0534, 0.0013, 0.0031, 0.0534, and 0.0001 with corresponding variances as (0.002), (0.0001), (0.0003), (0.0089), and (0.0000) respectively. However, the Olanrewaju-Olanrewaju kernel-based SVR yielded the smallest combine asymptotic point estimate of biasedness and variability of 0.0001 and (0.0000) respectively. In other words, the Olanrewaju-Olanrewaju kernel-based gave a robust account of analyzing SVR higher dimensional hyperplane application wise for continuous type bivariate.


Least Square-SVM
30
35
40
45


Bisquare Kernel-SVM
2.0
1.5
1.0
0.5
0.0
30
35
40
45
Annual Birthrate Per 1K


Guassian Kernel-SVM
2.5
2.0
1.5
1.0
0.5
0.0
30
35
40
45
Annual Birthrate Per 1K

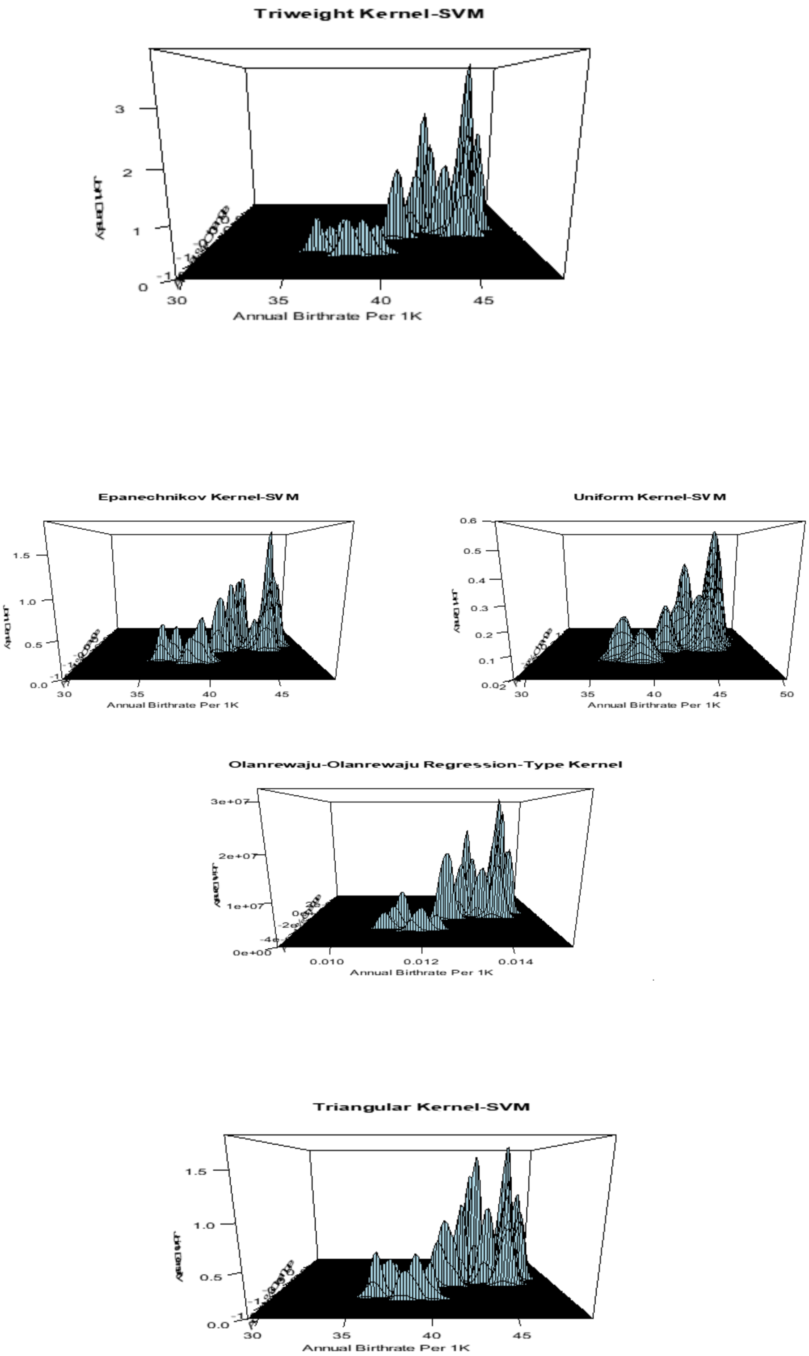


Figure 2. Joint densities of the nonparametric, least square, and Olanrewaju-Olanrewaju kernel-based SVR.

The joint densities of nonparametric triweight and Epanechnikov kernel-based SVRs yielded near heavy-flat tail for-all for the jump annual birthrate series and its percentage change (Δ%) per 1000 of the Nigeria Populace from 1950 to midyear of 2023. However, the nonparametric triangular and the Olanrewaju-Olanrewaju kernel-based SVRs notably yielded a mixture of flat heavy tail, normal "Bell-shape", and sided tails for-all density curves for the jump annual birthrate series and its percentage change (Δ%) per 1000. Least Square, Bisquare, Gaussian, and Uniform kernel-based SVRs yielded a fewer mixture of flat heavy tail, normal "Bell-Shape", and sided tails for-all density curve compare to other categories initially mentioned.

From table 2 above, the Euclidean distance (r) between nearest points and hyperplane of the bivariate kernel observations was minima for triweight and the Olanrewaju-Olanrewaju kernel-based SVRs with 0.7281 and 0.7008 estimates respectively. In collaboration, their parallel error margins (d) were miniature with 1.4562 and 1.4015 respectively. Consequently, this affirmed the robustness of the Olanrewaju-Olanrewaju

Table 2. Euclidean Distance & Error Margin of Parallel Hyperplane of the Nonparametric, Least Square, and Olanrewaju-Olanrewaju Kernel-Based SVRs

	Least Square-SVR	Bisquare Kernel-SVR	Gaussian Kernel-SVR	Triweight Kernel-SVR
d	1.6031	1.6545	1.5378	1.4439
r	0.8016	0.8273	0.7689	0.7212

Uniform Kernel –SVR	Epanechnikov -SVR	Triangular -SVR	Olanrewaju-Olanrewaju–SVR
1.4562	1.5756	1.5248	1.4015
0.7281	0.7878	0.7624	0.7008

kernel-based SVR for better explanation of the changes and negative changes accrued to the annual birthrate series and its percentage change (Δ%) per 1000 of the Nigeria populace respectively from 1950 to midyear of 2023 for deductive inference and economic future plans.

3. Conclusion

This study provides a variety of nonparametric regression kernels; the Least Square and the Olanrewaju-Olanrewaju regression-type kernels as augmentation for machine learning performance method. The SVR was used to classify hyperlanes of linearized non-linear regression mapping of high performance machine-learning kernels in a bivariate setting. The nonparametric SVR kernels treated were Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular and juxtaposed with the Olanrewaju-Olanrewaju regression-type ($L_{(O-O)\lambda_{\gamma(|\theta|)}}$) and Least Square kernels. The Lagrangian multiplier estimation technique was adopted for estimation of the involved regression coefficients, Euclidean distance (r), error of margin (d) between nearest points of parallel hyperplane, as well as other coefficients. In demonstration to a real-life problem, we applied the nonparametric SVR kernels, the Least Square and the Olanrewaju-Olanrewaju kernel to annual birthrate series and its percentage change (Δ%) of the Nigeria populace from 1950 to midyear of 2023. In conclusion, the Olanrewaju-Olanrewaju kernel robustly outperformed the nonparametric and Least Square kernels for SVR with a miniature Cross-Validation. Interpretably, for the estimated coefficients of the Olanrewaju-Olanrewaju kernel-based SVR, for every one percent increment in the annual birthrate per 1000, the mean rate of the Nigeria populace from 1950 to midyear of 2023 increased by 5.9% such that other variables were held constant. Its percentage change (Δ%) per 1000 increased by 3.2% such that other variables were held constant.

In recommendation, the nonparametric kernels of Bisquare, Gaussian, Triweight, Uniform, Epanechnikov, and Triangular; and that of the Olanrewaju-Olanrewaju and Least Square kernels for SVR could be fine-tuned for consideration of categorical data, missing observations and zero bivariate observations.

Acknowledgment

We would like to acknowledge the support of the National Bureau of Statistics (NBS), through her National Data Archive (NADA) Centre.

Author Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflicts of Interest: We the authors certify that there is no conflict of interest whatsoever with any affiliation, or involvement with any organization, financial and non-financial entity.

References

[1] Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer, New York.

[2] Cortes, C. & Vapnik, V. (1995). Support-vector networks, *Machine Learning, 20*(3), 273–297.

[3] Jijo, B. T. & Abdulazeez, A. M. (2021) Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends,2* (1), 20-28. ISSN: 2708-0757.

[4] Cattaneo, M. D, & Jansson M. (2018). Kernel-based semiparametric estimators: small bandwidth asymptotics and bootstrap consistency. *Econometrica, 86*, 955–995, https://doi.org/10.3982/ECTA12701.

[5] Fox, J. & Weisberg S. (2018). *Nonparametric Regression in R*. An Appendix to An R Companion to Applied Regression, Regression, 3rd Edition.

[6] Tibshirani, R. (2023). Nonparametric Regression: Nearest Neighbors and Kernels, Advanced Topics in Statistical Learning, Spring.

[7] Sun, F. & Lian S. (2022). Improved nonparallel support vector machine, *Sci. Rep., 12*, 17855.

[8] Sparapani, R. Spanbauer, C. & McCulloch, R. (2021). Nonparametric machine learning and efficient computation with modeling additive regression trees: The BART R Package. *Journal of Statistical Software, 97*(1), 1-66. Doi:10.18637/jss.v097.i01.

[9] Roy, A. & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A Review. *Reliability Engineering and System Safety, 2*(33), 109-126. ISSN 0951-8320.

[10] Liu, Y. & Yang, S. (2022). Application of decision tree-based classification algorithm on content marketing. *Journal of Mathematics,10*, Article ID 6469054.

[11] Sarker, I.H. (2021). Machine learning: algorithms, real world applications, and research directions. *SN Computer Science, 2*(160), 2-21.

[12] Ghanbari, M. & Goldani, M. (2022). Support vector regression parameters optimization using golden sine algorithm and its application in stock market.*Advances in Mathematical Finance and Applications,7*(2), 477-487. Doi:10.22034/AMFA.2021.1936352.1623.

[13] López, O.A.M., López, A.M. & Crossa, J. (2022). Support Vector Machines and Support Vector Regression Multivariate Statistical Machine Learning Methods for Genomic Predictions. ISBN:978-3-030-89009-4.

[14] Arun K.T.D., Yogendrab, B.E., Prashanth, J. & Prema, N.S. (2024). Water quality index forecasting using machine learning. *Journal of Propulsion Technology, 45*(1), ISSN:1001-4055.

[15] Shahzad, U. (2023). Prediction of probabilistic transient stability using support vector Regression. *Australian Journal of Electrical and Electronics Engineering, 20*(1), 35-49.

[16] Desai, S.S. & Kashid, D.N. (2020). Support vector machine-based modified Sp statistic for subset selection with non-normal error terms. *Journal of Modern Applied Statistical Methods, 218*(1), 1-21.

[17] Ding, S. F., Zhang, Z.C., Guo, L. .L., Zhang, J., & Xu, X. (2023). Survey on Twin support vector regression. *Acta Electonica Sinica, 51*(4), 1117-1134.

[18] Yuan, D. D., Li, M., Li, H.Y., Lin, C.J., &. Ji, X. (2022). Wind power prediction method: support vector regression optimized by improved jellyfish search algorithm, *Energies,15* 6404.

[19] Rodríguez-Pérez, R. & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *J. Comput. Aided Mol. Des*, 36, 355–362.

[20] De Brabanter, K., Suykens, J. & De Moor, B. (2013). Nonparametric regression via StatLSSVM. *Journal of Statistical Software, 55*(2), 1-21.

[21] Olanrewaju, R.O. & Olanrewaju, S.A. (2024). On the efficiency of the newly proposed Olanrewaju- Olanrewaju $L_{(O-O)\lambda_{\gamma(|\theta|)}}$ penalized regression-type estimator via GLMs technique. *Studies of Applied Economics 42*(1), 1-23. Doi:1025115/sae.v42i1.9532.

[22] Olanrewaju, R.O. (2020). Penalized likelihood estimation of gamma distributed response variable via corrected solution of regression coefficients.*Journal of Modern Applied Statistical Methods,19*(1). https://doi.org/10.22237/jmasm/1608552720.