J. Electrical Systems 20-3s (2024): 4361-447

[1]Bin Li

[2]Chenfan Duan

Construction and Optimization of Macroeconomic Data Forecasting Model Based on Machine Learning

Abstract: **-** Human capital was first viewed as a production component in macroeconomic development, but endogenous growth theories eventually replaced this perspective. The majority of earlier research used econometric models to investigate the GDP forecasting. Since machine learning models can efficiently resolve nonlinear interactions, this study offers a new perspective by examining the linkages using machine learning approaches. The prediction model for economic development was created for this reason using the best machine learning techniques, specifically the Support Vector Machine. In order to improve SVM prediction, the hyper parameters were optimised using the Bayes approach and several kernel functions. Three statistical metrics—the coefficient of determination, mean absolute error, and root mean square error are used to assess the models' effectiveness.

Keywords: macroeconomic, machine learning, hyper parameters, forecasting, mean absolute error.

I. INTRODUCTION:

Economic globalisation is a prominent characteristic of contemporary economic progress. Gross domestic product (GDP) is a crucial economic indicator that has increasingly raised concerns. It serves as a significant measure of a country's economic condition and is the foundation for assessing the overall health of the economy. GDP represents the total value of all finished products and services produced inside a country over a certain time period. GDP growth serves as a significant economic measure that reflects the overall well-being of a nation's economy [1-2]. Paul Anthony, a renowned American economist and Nobel Prize laureate, described GDP as "one of the most significant innovations of the century." An extensive examination and evaluation of GDP play a crucial role in macroeconomic control and the development of economic strategies.

Macroeconomics is the analysis of the overall economic activities of a country. Macroeconomic growth and developments are shaped by a multitude of causes. These forces are interconnected and mutually influence each other, resulting in a macroeconomic growth process that is characterised by being fashionable, cyclical, open, and nonlinear. Forecasting has emerged as a practical and complete field due to advancements in economic statistics, econometrics, and related disciplines. Forecasting involves using past data and the present circumstances to analyse the patterns of how things develop based on certain ideas, in order to generate hypotheses and make informed judgements. Forecasting relies on the historical and current functioning of the macroeconomic system [3-4]. Time series analysis and multiple regression techniques are the primary approaches used in macroeconomic forecasting.

Macroecomic forecasting often relies on time series data for historical analysis. Time series is a collection of data points that are arranged in chronological order based on the recurrence of the same statistical factors. Finding the innate relationships between data via historical data analysis is the goal of time series research in order to predict future data. For continuous time series, statistical regression models are the foundation of most time series modelling and forecasting approaches. Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) are two more often used analytical techniques. Making a time series steady in the case of non-stationary data requires taking several differences. After that, as references [4-6] demonstrate, the series may be represented as a moving average and white noise combination. Multiple regression, often known as the Vector Autoregressive (VAR) method, is another frequently used forecasting tool. This approach uses a statistical model that is constructed using the historical values of each variable in the system. VAR, or Vector Autoregression, is a statistical method that does not have any specific requirements for its application. It has gained significant popularity in economic forecasting in the last several years.

[1] [1]Accounting School, Jiaozuo University, Jiaozuo, Henan, China, 454000

[2]Student Affairs Office, Jiaozuo University, Jiaozuo, Henan, China, 454000

*Corresponding author's email: libin3826@126.com

The aforementioned conventional techniques are the prevailing methodologies in the realm of economic forecasting. Both approaches use linear models to replicate intricate real-world systems for the purpose of making predictions. However, they typically provide unsatisfactory outcomes when applied to intricate non-linear systems [7]. Many scholars have observed the constraints of linear models and thus shifted their focus to non-linear models. Initially, throughout the early stages of researching non-linear models, individuals continued to heavily depend on conventional concepts and approaches in model research. Traditional forecasting approaches include establishing a model by observation and analysis of system changes, followed by model testing for estimate and ultimately selecting the optimal model. Nevertheless, economic history data often consists of intricate non-linear time-series data, and conventional forecasting algorithms generally encounter several challenges [9-11]. Conventional approaches prioritise the examination of causation and the correlation between time series data. During the forecasting process, these systems suffer from a loss of information caused by issues like multicollinearity and error series, resulting in inadequate accuracy. The GDP is impacted by several causes, and the interactions between these elements are intricate, resulting in complicated time series and non-linearity. Consequently, predicting the GDP becomes very challenging.

When confronted with insoluble intricate non-linear issues, individuals have begun to explore alternative approaches in order to discover novel study methodologies. The human brain has the ability to efficiently process a wide range of intricate non-linear issues. This capability has motivated researchers to explore methods for emulating the human brain's problem-solving abilities in dealing with complicated non-linear problems. An ANN is a sophisticated model that emulates the behaviour of neurons in the brain [12]. It is a complex network system composed of several linked neurons, operating in a non-linear manner. Multiple learning algorithms exist for ANN, however BP is the predominant weight correction mechanism used in neural networks. ANN have shown high efficacy in constructing prediction models by autonomously acquiring knowledge from past experiences within a dataset, eliminating the need for intricate query and representation procedures.

Hence, this work presents an economic prediction model for GDP using an enhanced SVM model. In this study, a SVM optimised using the BA employing kernel functions is used to build a GDP forecast model. The input and output data of the neural network were pre-processed before the network was trained. In order to ensure that the data were scaled to the same order of magnitude, this was done. Next, a comparison was made between the predicted outcomes and the traditional forecasting model. Based on actual evidence, it may be concluded that the SVM network is a better fit for forecasting GDP. The main goal of this study is to solve the problem of precisely predicting very complex nonlinear dynamic macroeconomic systems by using the SVM network in place of traditional GDP forecasting techniques (such as time series forecasting and regression model forecasting).

The primary advancements and contributions of this research are as follows:

(1) SVM networks are used to address the challenge of intricate nonlinearities arising from numerous components, which cannot be effectively addressed by conventional prediction techniques.

(2) In order to enhance the speed and accuracy of the SVM model's convergence, the SVM network training incorporates the usage of BA, which has global search capacity and high practicality.

The subsequent sections of the research are structured in the following manner: Section 2 thoroughly examines the shortcomings of the primary conventional GDP forecasting techniques, while Section 3 presents the SVM network. Section 4 presents the empirical findings and analysis. Section 5 concludes the paper.

II. RELATED WORKS

Recently, the burgeoning artificial intelligence neural network has garnered significant scholarly interest.

This author [13] proposes leveraging biclustering, a form of fuzzy clustering, to address the challenges of fuzzy classification in economic zoning. By applying biclustering to China's macroeconomic statistics, the study aims to achieve a more accurate division of provinces into economic regions and develop predictive growth models for each province. This approach offers a promising alternative to traditional hierarchical clustering methods, particularly in situations where prior information is limited. This paper [14] proposes three ANN models to predict future prices of construction materials in Egypt's construction industry. Utilizing historical data and macroeconomic indicators, the models achieve promising results with mean absolute percentage errors ranging from 4.0% to 11%. They offer valuable tools for anticipating and mitigating cost fluctuations in material prices for contractors and project owners.

Beyond conventional econometric models, this research [15] uses ML techniques to examine the connection between human capital and economic growth. Prediction modelling is best served by Bayesian based Support Vector Machine and Bayesian based Gaussian Process Regression. This outperforms other models, showcasing improvements of 6.4% in R2, 10.7% in RMSE, and 1% in MAE. The influence of several factors on a deep learning-based bankruptcy prediction model is examined in this work [16]. Using Shapley value, it identifies influential variables across different prediction periods. Key findings include the significance of "housing starts" and the varying impact of macroeconomic indicators. The study highlights the potential for variable reduction while maintaining prediction accuracy. This paper [17] comprehensively reviews domestic and foreign literature on machine learning, financial risk management, investment portfolios, and related fields. It highlights machine learning's applications in various sectors such as flood forecasting, hospital management, and e-commerce. Additionally, it discusses strategies for financial risk management, portfolio optimization techniques, and emerging areas like plant factories and smart finance in universities. This paper [18] presents a novel approach to forecasting economic downturns by utilizing ML techniques, complementing traditional econometric methods. Focusing on Italian GDP data and related variables, the study evaluates the forecasting accuracy of a proposed ML model against a classic Linear Regression Model. The results show that both approaches effectively predict economic downturns, but the Nonlinear Autoregressive with exogenous variables model demonstrates higher accuracy.

An enhanced macroeconomic growth forecast system based on data mining and fuzzy correlation analysis is presented in this work [19]. By analyzing economic characteristics, restructuring spatial economic structures, and integrating statistical information, the algorithm generates association rules using the optimized Apriori algorithm. This experimental results demonstrate the algorithm's adaptability, reduced computation time, and improved prediction accuracy in economic data mining. This study [20] aims to enhance government governance precision amidst China's evolving economic landscape. It proposes leveraging computer science and technology to analyze macroeconomic indicators, establish forecasting models, and address challenges such as prediction overfitting and high-dimensional optimization. This paper presents two forecasting models for economic indices together with adaptive techniques for multiobjective optimisation algorithms that aim to strike a compromise between solution convergence and variety. This study [21] focuses on predicting GDP using a genetic algorithm optimized Radial Basis Function Neural Network. By optimizing the center vector, base width vector, and weights, a GDP image prediction model is proposed for economic zones. Using historical GDP data, the RBFNN-GA model achieves high prediction accuracy with a relative error of only 3.52%. Comparative analysis demonstrates superior accuracy compared to ARIMA time series and GM (1,1) models.

This paper [22] proposes a financial risk forecasting model based on ANN to address the shortcomings of traditional risk management systems. Leveraging deep learning techniques, the model achieves significantly higher accuracy compared to traditional methods, with an accuracy of 96.84% and a 32.84% reduction in error compared to SVM. This study [23] evaluates multiple machine learning algorithms for forecasting residential construction demand in Jordan. Nine economic indicators were used to develop demand models. Elastic-Net exhibited the highest accuracy (0.838), followed by artificial neural network, Eureqa, and Extra Trees. The findings suggest an 11.5% increase in residential construction demand in Jordan for the first quarter of 2023 compared to 2022. Gradient boosting and random forest machine learning models are presented in this research [24] with the purpose of predicting Japan's real GDP growth between 2001 and 2018. The research assesses prediction accuracy using root squared mean error and mean absolute percentage error and uses cross-validation to optimise hyperparameters. The findings show that both machine learning models perform better than the Bank of Japan and the International Monetary Fund's benchmark predictions, with the gradient boosting model showing higher accuracy. This paper [25] proposes EcoForecast1, an interpretable data-driven approach for short-term macroeconomic forecasting using the N-BEATS neural network. Demonstrating high stability and accuracy with macroeconomic data from China, EcoForecast1 outperforms traditional methods like BVAR by up to 3.94 times. It offers robustness across prediction domains and sensitivity to economic inflection points, making it a valuable tool for improving short-term macroeconomic forecasting. The table 1 shows the comprehensive analysis of literature survey

Table 1. Comparative analysis of related works

Ref No.	Methodologies	Applications
13	Biclustering, Fuzzy clustering	Economic zoning, Predictive growth models for provinces

14	Artificial Neural Network (ANN)	Forecasting construction material prices, Cost fluctuations mitigation
15	Bayesian Tuned SVM, BT-GPR	Relationship analysis between economic development and human capital
16	Deep learning, Shapley value	Bankruptcy prediction, Variable reduction while maintaining prediction accuracy
17	Financial risk management	Flood forecasting, Hospital management, Financial risk management, Portfolio optimization
18	Linear Regression, NARX	Forecasting economic downturns, Evaluating forecasting accuracy
19	Data mining, Fuzzy correlation analysis	Macroeconomic growth prediction, Improved prediction accuracy
20	Computer science, Multiobjective optimization	Enhancing government governance precision, Economic index forecasting
21	Genetic Algorithm (GA), RBFNN	GDP prediction, Economic zones, Comparative analysis with traditional methods
22	ANNS	Financial risk forecasting, Addressing shortcomings of traditional risk management systems
23	Elastic-Net, Extra Trees	Forecasting residential construction demand, Construction industry optimization
24	Gradient Boosting, Random Forest	Forecasting real GDP growth in Japan, Outperforming benchmark forecasts
25	Data-driven approach, N-BEATS neural network	Short-term macroeconomic forecasting, High stability and accuracy

III. PROPOSED BA-SVM MODEL:

The Macroeconomic GDP has been identified by the effective BA-SVM approach proposed in this study. The block design for the suggested BA-SVM is shown schematically in Figure 1.

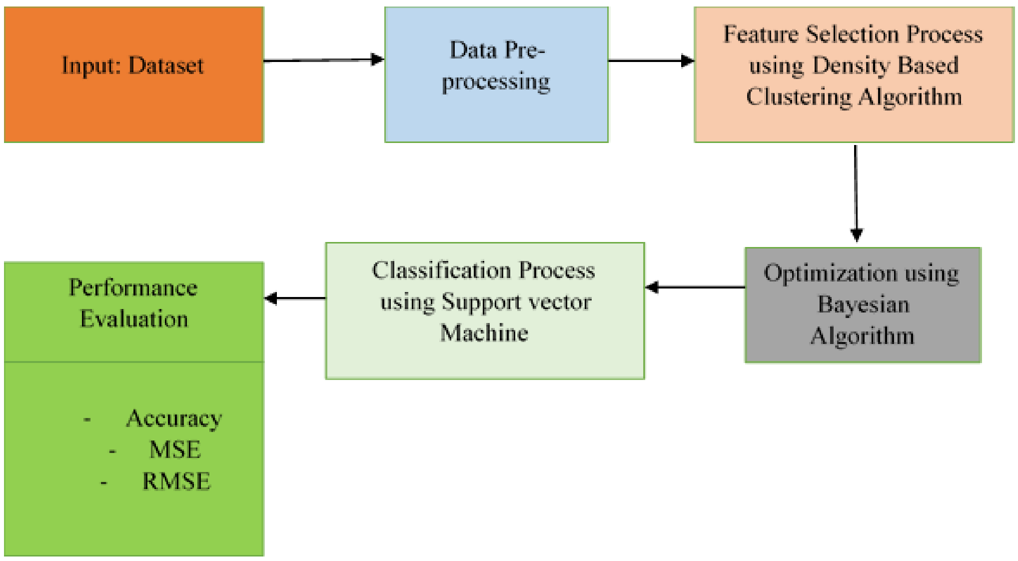

Figure 1. Block Diagram for proposed BA-SVM

The BA-SVM technique consists of three main sub-processes: pre-processing, feature selection using the Bayesian Algorithm (BA), and classification using SVM. The use of BA for the appropriate selection of features contributes to increased classification performance. Figure 1 shows the overall block diagram of the proposed AOA-SVM approach.

3.1 Data Pre-Processing

During the process of missing data interpolation, the calculation of the data filling value will be determined using the following formula:

$$Input_h = \alpha_1 avgh + \alpha_2(\beta_1 V_{backw} + \beta_2 V_{backw})$$

One of these variables is α, which is a weight employed in experiments to optimise the filling effect. The weight is approximately adjusted in this experiment to ensure that $\alpha_1 = \alpha_2 = 0.5$. It is important to mention that the interpolation technique has constrained the predictive performance in comparison to other models. When the value of α is modified to apply other data interpolation algorithms, the outcomes of the experiment remain identical.

3.2 Process Involved in the DCA Technique

DCA has the potential to identify separate groups depending on the distribution of evaluated density. The system has the ability to identify organised clusters without having knowledge of their quantities. The below example demonstrates the fundamental principle of DCA: The DCA algorithm identifies all points within the vicinity of a randomly selected unvisited point p, where the vicinity is defined as the greatest distance from p. In order to create a compact region, MinPts represents the minimal threshold of points needed. A core point is denoted as p when MinPts is inside the distance. When p is a core point, all points in its neighbourhood are clustered together. The Density-Based Clustering Algorithm (DCA) identifies and includes each density-reachable data point into a cluster that is equivalent in terms of density. A point q is considered a border point when it is closely reachable from other core points, but its neighbourhood contains fewer points than the specified MinPts threshold. An outlier, also known as a noisy point, is a data point that cannot be reached or connected to other data points. DCA employs a sequential approach to extract clusters. Iterate until the last cluster is reached and no more density-reachable places are discovered. A set of points is divided into high-density and low-noise border points using DCA. By examining each state's socioeconomic-demographic characteristics and per capita NSDP development trends, DCA seeks to identify states that are comparable to one another. The DCA has the ability to identify several clusters by analysing the density distribution. The DCA approach allows for the computation of equivalent conditions using per capita income as the basis.

Consider two points x and $y, d\,(x, y)$ is similarities among them, $\Gamma\varepsilon(x) = \{y \in V | d(x, y) \leq \epsilon\}. \rho(x) = |\Gamma\varepsilon(x)$ indicates the density value of x:

$$S(x) = \begin{cases} 1, & core\ point\ with\ \rho(x) \geq\ Minpts \\ 0 & border\ point\ with\ 1 < \rho(x) < Minpts \\ -1, & noise\ with\ \rho(x) = 1 \end{cases}$$

3.3 Support Vector machine Classifier:

A contemporary learning algorithm that was inspired by statistical learning theory is the support vector machine. It has shown remarkable performance in time series forecasting and classification problems and is renowned for its strong mathematical foundation. The SVM technique has shown efficacy in approximating multivariate function parameters, resolving non-linear regression issues, enhancing generalisation skills, and furnishing distinct solution representations. The SVM is a very promising approach that aims to minimise overfitting by effectively balancing model complexity. The fundamental principle of SVR is to transform the original dataset x_i, which may exhibit non-linearity, into a feature space with a higher dimensionality.

Let's examine a training set $G = (x_i, y_i)$ where $i = 1,2, \dots, \lambda$. In this set, $x = \{x_1, x_2, x_3\} \dots \subseteq R^N$ represents the input variable vector, while $y = \{y_t, \dots, y_\lambda\} \dots \subseteq R$ represents the output variable. The hyperplane function is represented by equation 2.

$$f(x) = w \times \varphi(x) + b$$

where the function $\varphi(x)$ represents the feature space with a large number of dimensions, which is obtained by applying a non-linear mapping to the input space. The coefficients x, w and b are determined by minimising the regularised risk function in the following manner.

$$min\frac{1}{2}\|w\|^2 + c\frac{1}{\lambda}\sum_{i=1}^{\lambda} L_e(y_i, f(x_i))$$

$$L_\epsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon & |y_i - f(x_i)| \geq 0 \\ 0 & otherwise \end{cases}$$

. The regularisation term, $\|w\|^2$, is used to factor in the minimization of the function's capacity by promoting flatness. The difference between the predicted values from the regression function and the actual values, as determined by the ε-insensitive loss function, is represented by the parameter ε. It is required to minimise the norm of w using an ε-insensitive loss function in order to get an appropriate generalisation for the regression function. An empirical risk-quantification cost function is denoted by the letter C. The trade-off between the function's smoothness and the limit of deviation ε is determined by the constant C, which is bigger than zero. In Figure 2, the support vector machine is shown.

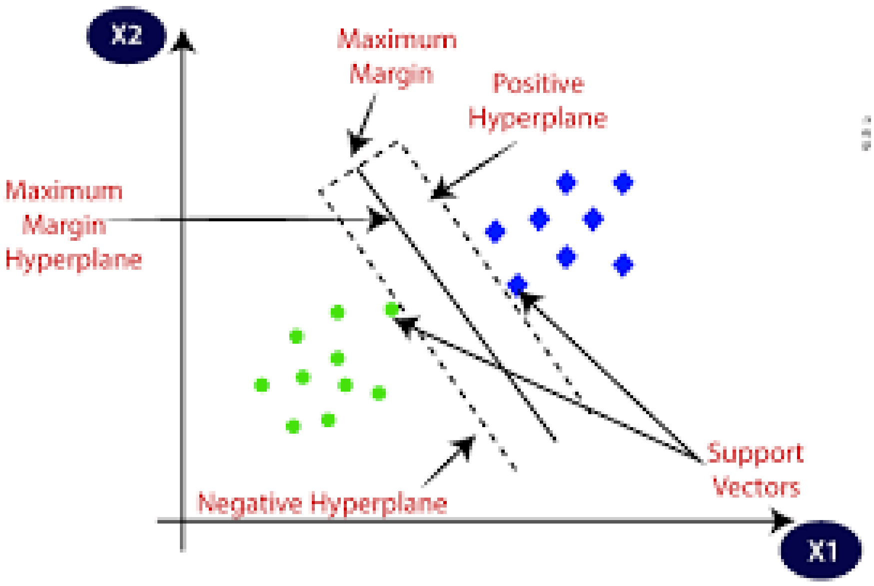

Figure 2. Support Vector Machine

Next, the constrained format may be expressed in the following manner:

$$min\frac{1}{2}\|w\|^2 + \ c\frac{1}{\lambda}\sum_{i=1}^{\lambda}(\xi_i, \xi_i^*)$$

Two slack variables ξ_i and ξ_i^*, may be used to denote the deviation between the actual values and the respective boundary values of the ε-tube. Ultimately, the restricted optimisation issue is resolved by using the following formulation.

$$\max R(\alpha_i, \alpha_i^*) = \sum_{j=1}^{\lambda}(\alpha_i, \alpha_i^*)y_i - \varepsilon\sum_{j=1}^{\lambda}(\alpha_i, \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{\lambda}\sum_{j=1}^{\lambda}(\alpha_i, \alpha_i^*)(\alpha_j, \alpha_j^*)K(x_i, x_j)$$

Where the symbols α_i and α_i^* represent Lagrange multipliers. The function $K(x_i, x_j)$ is a kernel function. This research utilises a linear kernel function in Support Vector Machines (SVM), denoted as $K(x_i, x_j) = x_i^T x_j$.

Therefore, the non-linear regression function is expressed as equation 7:

$$f(x) = \sum_{i=1}^{\lambda}(\alpha_i, \alpha_i^*)K(x_i, x_j) + b$$

Kernel Function:

The efficacy of the SVM model is contingent upon the selection of the kernel parameters. Hence, the crucial aspect of SVM modelling is in the suitable choice of the kernel function. Among the suggested new kernels, only four are regularly used: the linear function, the polynomial kernel function (PKF), the radial Basis function (RBF) (also known as the Gaussian kernel), and the sigmoid kernel.

Linear kernel

The linear kernel was introduced as the most basic kernel function. Linear modelling is a widely used technique that may be described as follows:

$$k(x_i, x_j) = x_i^T x_j$$

In some scenarios, the linear kernel is better suitable, especially when dealing with a high number of features.

Polynomial kernel function

The PKF, or Polynomial Kernel Function, is a kernel that is not stationary and is widely used for non-linear modelling. It may be expressed in the following manner:

$$k(x_i, x_j) = (\gamma x_i^T x_i + r)^d . \gamma > 0$$

Where the symbol γ represents the gamma term in the kernel function for all kinds of kernels except for the linear kernel. The symbol d represents the polynomial degree, and the symbol r represents the bias term in the kernel function.

Radial basis function kernel

Recently, there has been a significant focus on RBF, particularly with the use of a Gaussian function in the following form:

$$k(x_i, x_j) = \exp\left(-\gamma \|x_i, x_j\|^2\right) . \gamma > 0$$

Here, $\gamma>0$ is a parameter that determines the extent of the Gaussian distribution. The degree of the polynomial kernel determines the flexibility of the final classifier, comparable to the function.

Sigmoid kernel

The sigmoid kernel is defined as follows

$$k(x_i, x_j) = tanh(\gamma x_i^T x_i + r)$$

Utilising such a kernel is tantamount to employing a neural network that has a solitary hidden layer. This kernel relies on two parameters, γ and r, which may provide challenges during its implementation.

3.4 Hyper parameter Optimization using Bayes Approach:

The updated belief for the known goal function is represented inside the Bayesian framework via the process of calculating the objective function of the posterior distribution. In Bayesian optimisation, the objective function f(a) is constructed by the use of a probabilistic model. For the purpose of evaluating the bounded set A, the model is used to make a prediction about the subsequent point that is not yet known. There is a possibility that the earlier evaluation of f(a) will make use of anticipated information in addition to the Hessian estimates or the local gradient. Specifically, it is able to determine the highest possible complexity of functions that are not convex.

Bayesian optimisation has two distinct components. The first component is the Gaussian process prior, whereas the succeeding component is the acquisition function, which is used to assess the next point by generating a utility function.

The Gaussian process is a very efficient and potent prior distribution over the space of smooth functions. The concept is rooted in an unbounded random variable, such as one with an unlimited number of possible values. In this case, every finite random variable may be described using a combined Gaussian distribution.

The Gaussian distribution (GP) is:

$$f(a) \sim GP(\mu(a), k(a, a^*)$$

Where the symbol μ(a) represents the mean function of variable a, whereas the $k(a,a*)$ value represents the covariance function of the data a and $a*$.

There are other methods for selecting an acquisition function, such as anticipated improvement (EI) and probability of improvement (PI). The mathematical function known as PI is often represented as

$$\alpha PI(a) = \phi(r(a), \big(r(a)\big) = \frac{f(a_{best}) - \mu(a)}{\sigma(a)}$$

The anticipated variance and mean function of the objective function are denoted by (a) and (a). Φ(.) and ɤ(.) represent the PDF and CDF of the standard normal distribution. The most optimal observation at now is denoted by $a_{best}= arg\ max_{a_{i\in 1:t}} f(ai)$, respectively.

The SVM model has three parameters (C, ϵ, σ) that need to be optimised. C represents the penalization coefficient, σ is the kernel parameter, and a larger value of σ leads to a smaller structural risk. ϵ, on the other hand, is the insensitive loss coefficient that controls the width of the regression function in the insensitive area.

IV. EXPERIMENTAL SETUP:

The MATLAB 7.0 software's neural network toolbox was used in the study for modelling reasons. The experiment's sample consisted of GDP data from 1997 to 2016. The official website of the National Bureau of Statistics (http://www.stats.gov.cn/) served as the main source of the experimental data. The experiment used 14 data sets spanning from 1997 to 2010 for training purposes, whereas 4 data sets from 2011 to 2014 were employed as test samples to evaluate the predictive capabilities of the SVM model.

The SVM network has the following parameters: an initial learning rate of 0.015, 1000 epochs for the number of iterations, and a selection goal of 0.0001. The parameters of the BA algorithm are as follows: a population size of 64, 40 population choices, an initial weight of 0.9, and a step size factor of 2. During the experiment, the model underwent training using 14 sets of data. Following the completion of training, an additional 4 sets of data were inputted into the model to generate prediction outputs. The anticipated outcome of the model was acquired.

4.1 Performance Evaluation:

Two well recognised performance assessment measures used were R2, the coefficient of determination, and RMSE, the root mean square error, together with MAE, the mean absolute error.

R^2:

In a regression model, the coefficient of determination quantifies the percentage of the dependent variable's variation that can be predicted from the independent variables.

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$$

Where:

SS_{res} is the sum of squared residuals.

SS_{total} is the total sum of squares, which measures the total variability in the dependent variable.

RMSE

RMSE is used to quantify the level of error shown by regression models. It is a useful measure for comparing predicting mistakes in the following manner: Refer to Equation (1).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\check{y} - y_i)^2}{n}}$$

where n be the number of test samples. y_i represents the genuine target value of the ith sample, and $\check{y}$ represents the predicted value by the regressor.

MAE

MAE is a performance metric used to evaluate the performance of a regressor. The calculation is performed according to Equation (2).

$$MAE = \sum_{i=1}^{n} \frac{|\, \check{y} - y_i|}{n}$$

Where the above equation depicts the relationship between the number of test samples (n), the real target value of the ith sample (y_i), the projected value by the regressor (xi), and the absolute value (|.|).

Table 1. Data without preprocessing for GDP forecasts

Year	GDP (billion yuan)	F1 (10,000 people)	F2 (billion yuan)	F3 (billion yuan)	F4 (billion dollars)	F5 (billion yuan)
1997	2928.83	3471.29	720.73	250.82	13.47	1051.6
1998	3108.09	3603.17	848.59	273.64	12.83	1125.33
1999	3326.75	3701.39	957.47	323.12	12.82	1229.21
2000	3691.88	3577.58	1066.27	347.83	16.53	1364.66
2001	3983	3607.96	1209.27	431.7	17.54	1511.07
2002	4340.94	3644.52	1355.87	533.02	17.95	1678.86
2003	4638.73	3694.78	1557	573.75	21.46	1816.3
2004	5612.26	3747.1	1981.29	719.54	30.98	2069.84
2005	6473.61	3658.3	2540.06	873.42	37.47	2459.12
2006	7568.89	3883.41	3232.39	1064.52	50.94	2834.22

Table 2. Normalized GDP forecasts

Year	GDP (billion yuan)	F1 (10,000 people)	F2 (billion yuan)	F3 (billion yuan)	F4 (billion dollar)	F5 (billion yuan)
1997	0	0	0	0	0.0332844	0
1998	0.0428511	0.1327061	0.0581746	0.0513614	0.0002624	0.0467082
1999	0.0877236	0.1550508	0.1010127	0.0977166	0	0.1046569
2000	0.1663452	0.0545095	0.1438194	0.1403502	0.0973242	0.1802167
2001	0.2288508	0.1475303	0.2010818	0.2509499	0.1338195	0.2718904
2002	0.3058262	0.2605771	0.2577607	0.3624805	0.134575	0.3554908
2003	0.3698662	0.4162231	0.336894	0.421335	0.2266527	0.4321607
2004	0.5792248	0.5781443	0.5038282	0.5862061	0.4763903	0.5735962
2005	0.7645589	0.3033238	0.7236727	0.7707809	0.6466422	0.7907532
2006	1	1	1	1	1	1

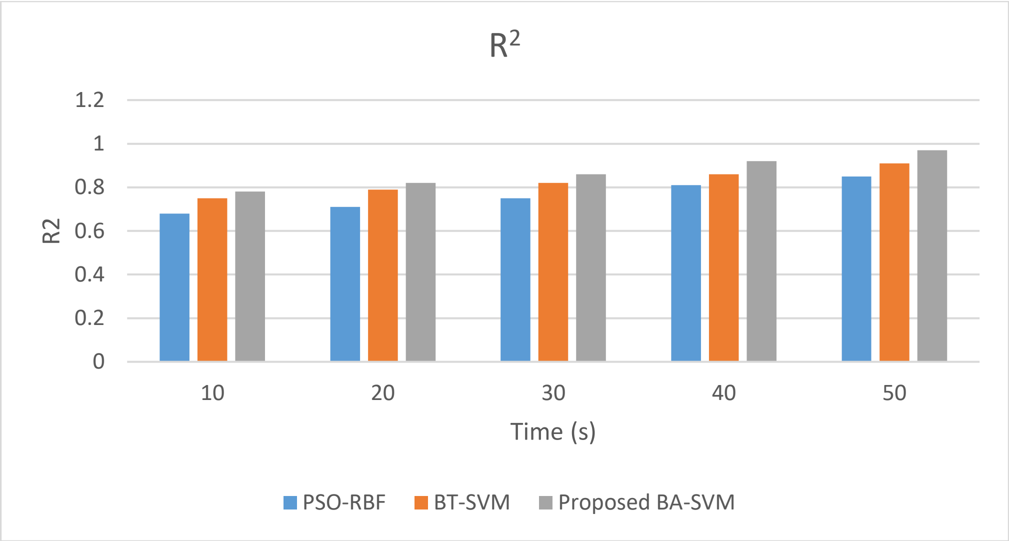

Figure 3. Performance analysis of R^2

The figure 3 shows the evolution of R2 values across different time intervals for three distinct methods: PSO-RBF, BT-SVM, and Proposed BA-SVM. Initially, PSO-RBF starts at 0.68 and gradually increases to 0.85, showing incremental predictive accuracy. Similarly, BT-SVM begins at 0.75 and climbs to 0.91, demonstrating continuous enhancement. However, the Proposed BA-SVM method exhibits superior performance from the outset, with R2 values starting higher at 0.78 and consistently outperforming the other methods across all intervals, reaching a peak of 0.97 at 50 seconds. This underscores the Proposed BA-SVM's effectiveness in achieving superior predictive accuracy throughout the observation period.

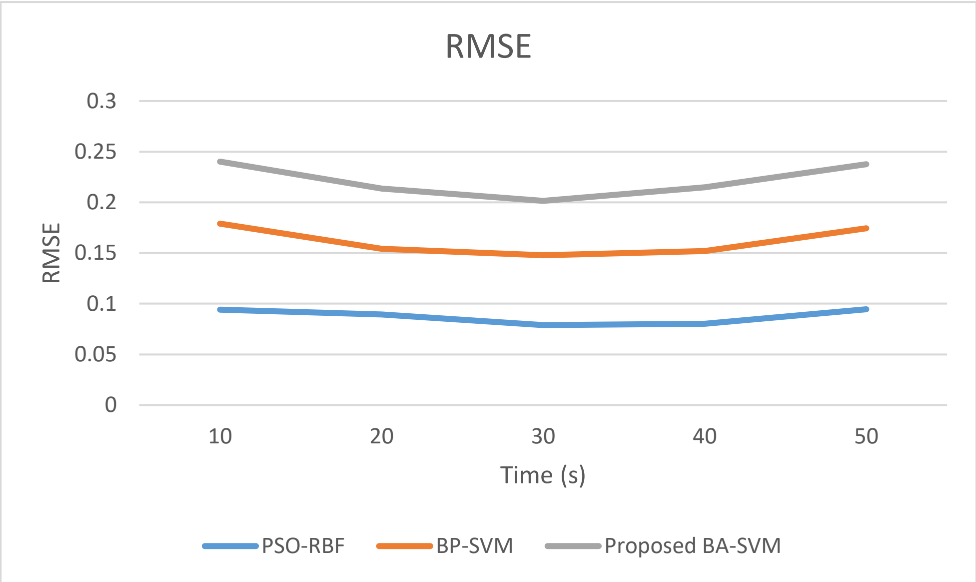

Figure 4 : Performance analysis of MSE

The fig 4 shows the RMSE values for three different methods, namely PSO-RBF, BP-SVM, and Proposed BA-SVM, at a specific time point (50). At time 50, the RMSE for PSO-RBF is calculated to be 0.0945, while for BP-SVM, it stands at 0.0799. In comparison, the proposed BA-SVM method demonstrates the lowest RMSE value of 0.0632 among the three methods. This indicates that at this particular time point, the proposed BA-SVM method offers the most accurate predictions compared to both PSO-RBF and BP-SVM methods. Lower RMSE values signify a closer alignment between predicted and observed values, suggesting superior predictive performance. Therefore, based on the RMSE values presented in the table, the proposed BA-SVM method appears to be the most effective choice for prediction at time 50.

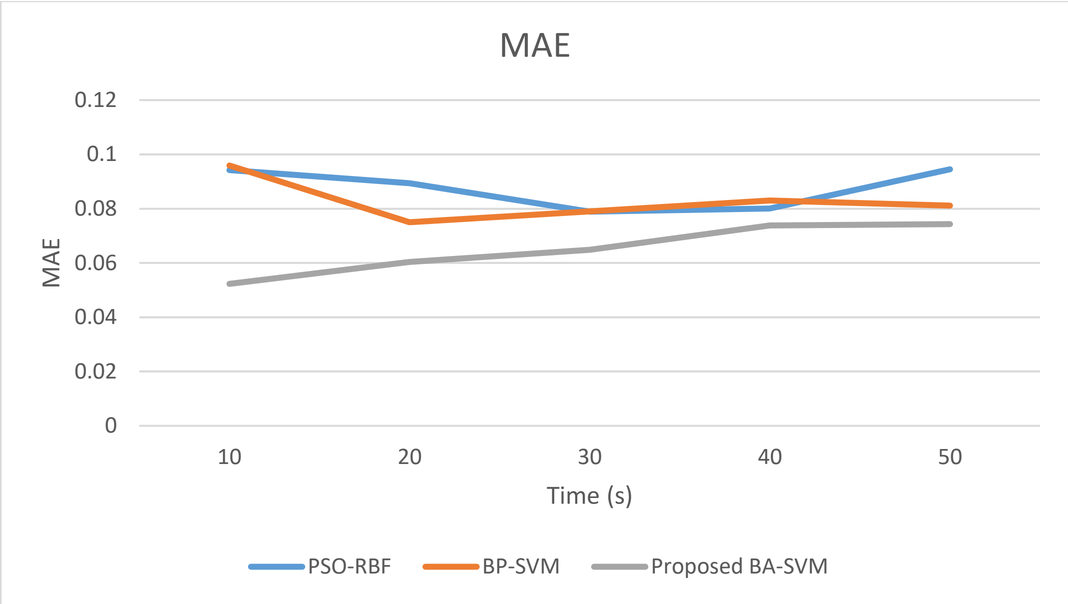

Figure 5: Performance analysis of MSE

The fig. 5 displays the MSE values for three distinct methods - PSO-RBF, BP-SVM, and Proposed BA-SVM - at a specific time point (50). At time 50, the MAE for PSO-RBF is computed to be 0.0945, while for BP-SVM, it is 0.0811. On the other hand, the Proposed BA-SVM method yields an MAE value of 0.0743 at the same time point. Analyzing these values, we observe that the Proposed BA-SVM method has the lowest MAE among the three methods at time 50. A lower MAE implies that the predictions generated by the Proposed BA-SVM model are, on average, closer to the actual values compared to both PSO-RBF and BP-SVM methods. Thus, based on the MAE values presented in the table, it can be inferred that the Proposed BA-SVM method exhibits superior predictive accuracy at time 50 compared to the other methods.

V. CONCLUSION:

This essay uses Density-based clustering techniques to analyse the macroeconomic data of my nation and derives significant findings. Primarily, this work employs the clustering approach inside the clustering algorithm to conduct cluster analysis in the research methodology. However, this research use the SVM algorithm for clustering analysis, which overcomes the limitation that all clustering criteria in the clustering technique must be involved. This also resolves the issue of having to choose between one option or the other when classifying items. An item may be classified into numerous categories depending on the specific circumstances. Simultaneously, this approach may identify items exhibiting fully contrasting expressions concurrently, which has significant importance in economic research. Ultimately, the empirical study findings lead to the following conclusions:

(1) The outcome of the cluster analysis provides a segmentation of the economic area of my nation as a whole.

The Support Vector Machine (SVM) yielded disparate outcomes compared to the preceding approach.

REFERENCES:

[1] Jang, Y., Jeong, I., & Cho, Y. K. (2020). Business failure prediction of construction contractors using a LSTM RNN with accounting, construction market, and macroeconomic variables. *Journal of management in engineering*, *36*(2), 04019039.

[2] Yu, Y. (2022). GDP Economic forecasting model based on improved RBF neural network. *Mathematical Problems in Engineering*, *2022*.

[3] Lahmiri, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, *6*, 100166.

[4] Chou, J. S., Nguyen, N. M., & Chang, C. P. (2022). Intelligent candlestick forecast system for financial time-series analysis using metaheuristics-optimized multi-output machine learning. *Applied Soft Computing*, *130*, 109642.

[5] Babenko, V., Panchyshyn, A., Zomchak, L., Nehrey, M., Artym-Drohomyretska, Z., & Lahotskyi, T. (2021). Classical machine learning methods in economics research: Macro and micro level example. *WSEAS Transactions on Business and Economics*, *18*(22), 209-217.

[6] Dong, J., Chen, Y., & Guan, G. (2020). Cost index predictions for construction engineering based on LSTM neural networks. *Advances in Civil Engineering*, *2020*, 1-14.

[7] Xu, Z., Mohsin, M., Ullah, K., & Ma, X. (2023). Using econometric and machine learning models to forecast crude oil prices: Insights from economic history. *Resources Policy*, *83*, 103614.

[8] Mahdavian, A., Shojaei, A., Salem, M., Yuan, J. S., & Oloufa, A. A. (2021). Data-driven predictive modeling of highway construction cost items. *Journal of Construction Engineering and Management*, *147*(3), 04020180.

[9] Liu, B., Zhao, S., Yu, X., Zhang, L., & Wang, Q. (2020). A novel deep learning approach for wind power forecasting based on WD-LSTM model. *Energies*, *13*(18), 4964.

[10] Alshboul, O., Shehadeh, A., Almasabha, G., Mamlook, R. E. A., & Almuflih, A. S. (2022). Evaluating the impact of external support on green building construction cost: A hybrid mathematical and machine learning prediction approach. *Buildings*, *12*(8), 1256.

[11] Zyatkov, N., & Krivorotko, O. (2021, September). Forecasting recessions in the US economy using machine learning methods. In *2021 17th International Asian School-Seminar" Optimization Problems of Complex Systems (OPCS)* (pp. 139-146). IEEE.

[12] Tuo, S., Chen, T., He, H., Feng, Z., Zhu, Y., Liu, F., & Li, C. (2021). A regional industrial economic forecasting model based on a deep convolutional neural network and big data. *Sustainability*, *13*(22), 12789.

[13] Pang, C. (2022). Construction and analysis of macroeconomic forecasting model based on biclustering algorithm. *Journal of Mathematics*, *2022*, 1-10.

[14] Shiha, A., Dorra, E. M., & Nassar, K. (2020). Neural networks model for prediction of construction material prices in Egypt using macroeconomic indicators. *Journal of Construction Engineering and Management*, *146*(3), 04020010.

[15] Ozden, E., & Guleryuz, D. (2022). Optimized machine learning algorithms for investigating the relationship between economic development and human capital. *Computational Economics*, *60*(1), 347-373.

[16] Jang, Y., Jeong, I., & Cho, Y. K. (2021). Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors. *Engineering, Construction and Architectural Management*, *28*(10), 3282-3298.

[17] Khan, M. A., Abbas, K., Su'ud, M. M., Salameh, A. A., Alam, M. M., Aman, N., ... & Aziz, R. C. (2022). Application of machine learning algorithms for sustainable business management based on macro-economic data: supervised learning techniques approach. *Sustainability*, *14*(16), 9964.

[18] Cicceri, G., Inserra, G., & Limosani, M. (2020). A machine learning approach to forecast economic recessions—An Italian case study. *Mathematics*, *8*(2), 241.

[19] Sun, H., Yao, Z., & Miao, Q. (2021). Design of macroeconomic growth prediction algorithm based on data mining. *Mobile Information Systems*, *2021*, 1-8.

[20] Fan, H. (2022). Research on Forecast of Macroeconomic Indicators Based on Multiobjective Optimization. *Wireless Communications and Mobile Computing*, *2022*.

[21] Zhu, M., & Meng, Z. (2021). Macroeconomic image analysis and gdp prediction based on the genetic algorithm radial basis function neural network (rbfnn-ga). *Computational Intelligence and Neuroscience*, *2021*.

[22] Tang, D. (2023, May). Optimization of Financial Market Forecasting Model Based on Machine Learning Algorithm. In *2023 International Conference on Networking, Informatics and Computing (ICNETIC)* (pp. 473-476). IEEE.

[23] Sammour, F., Alkailani, H., Sweis, G. J., Sweis, R. J., Maaitah, W., & Alashkar, A. (2023). Forecasting demand in the residential construction industry using machine learning algorithms in Jordan. *Construction Innovation*.

[24] Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, *57*(1), 247-265.

[25] Wang, X., Li, C., Yi, C., Xu, X., Wang, J., & Zhang, Y. (2022). EcoForecast: An interpretable data-driven approach for short-term macroeconomic forecasting using N-BEATS neural network. *Engineering Applications of Artificial Intelligence*, *114*, 105072.