



International Review of Applied Sciences and Engineering

12 (2021) 1, 50-56

יוחם

10.1556/1848.2020.00132 © 2020 The Author(s)

ORIGINAL RESEARCH PAPER





Volatility forecast of CIDC Construction Cost Index using smoothing techniques and machine learning

P. Velumani* and N.V.N. Nampoothiri

Kalasalingam Academy of Research and Education, Tamil Nadu, India

Received: May 12, 2020 • Accepted: August 7, 2020 Published online: January 30, 2021

ABSTRACT

The Construction Industry Development Council (CIDC) of India has been calculating and publishing the Construction Cost Index (CCI), monthly, since 1998. Construction cost variations interrogate different kinds of projects such as roads, power plants, buildings, industrial structures, railways and bridges. The success rate of completion of construction project is diminished due to the lack of prediction knowledge in CCI. Predicting CCI in greater accuracy is quite difficult for contractor and academicians. The following factors are influenced higher in CCI such as population, unemployment rate, consumer price index (CPI), long term interest rate, domestic credit growth, Gross Domestic Product (GDP) and money supply (M4). CCI can be used to forecast the construction cost. The relevant resource data was collected across the nation between 2003 and 2018. As outcome-based, non-econometric tools such as smoothing techniques, artificial neural network (ANN) and support vector machines (SVMs) have produced a better outcome. Among these, smoothing techniques have given the notable low error and high accuracy. This accuracy has measured by Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Root Mean Square Error (RMSE). The major objective of this research is to help the cost estimator to avoid underestimation and overestimation.

KEYWORDS

Construction Industry Development Council, Construction Cost Index, smoothing techniques, artificial neural network, support vector machine, prediction

1. INTRODUCTION

The Construction Industry Development Council (CIDC) of India has been calculating and publishing the Construction Cost Index (CCI), monthly, since 1998. Construction cost variations interrogate different kinds of projects such as roads, power plants, buildings, industrial structures, railways and bridges. Different types of resources are required for each sector to complete the task. The relevant resource data is collected across the nation. While framing the CCI, the following things are considered such as construction materials, labours, fuel, oil and lubricants data.

In the construction industry, cost trends are measured with the help of CCI records [1]. During construction, it is often used to identify the time-lapse between estimation and execution of site operations [2]. CCI is most important to provide a comparison between cost changes, which differ from period to period for all the construction goods or services [3]. Usually, clients need a good and precise estimation cost for better acquisition of the project proposed within the period. This cost estimation becomes an important rule to satisfy the requirements of clients [4]. Hence to benefit-cost estimators, CCI forecasting becomes a vital thing to predict. Commonly, CCI increases over the long term, but it performs well on short term variation. Hence to avoid estimation, it is essential to predict the varying CCI so that cost estimators can manage their execution of the project [5]. Though there are numbers of prediction tools giving better accuracy, another prediction tool, namely, SVM method uses

*Corresponding author. E-mail: velumani@klu.ac.in



kernel technique which can solve certain disadvantages faced by previous techniques. For the prediction of CCI, support vector machine (SVM) technique is implemented as a new idea. Hence to provide an accurate estimation cost, bidding operation and construction investments in India, the further work deals with the prediction of CCI using smoothing techniques, neural network (NN) and SVM. A good comparison carried out to show the best model that delivers less error in prediction is studied in further work.

2. LITERATURE SURVEY

Author	Tools	Description
[6]	wavelet transformation and neural network method	Using a wavelet transformation and neural network method, CCI is forecasted by creating new models and compared for a better result
[7]	Time series models.	To predict CCI accurately, the information of CCI along with explanatory variables can be used for multivariate time series models. On comparing univariate time series models for forecasting CCI, it is shown that multivariate time series models are more accurate
[8]	K-nearest neighbour (NN)	It is found difficult to predict accurate results of CCI, especially during mid and long-term conditions. Hence some machine learning algorithms are used to predict CCI in mid and long-term operations namely K-nearest
[9]	Neural Network, Time Series analysis and Regression methods	neighbour (NN) and perfect random tree ensembles (PERT) in USA It is difficult to find a source to collect CCI in Egypt; hence with the help of previous data of key construction cost, a formula is driven to calculate CCI. They are
[10]	K-nearest neighbours (KNN)	then using neural network, time series analysis and regression methods for predicting CCI Modified K-nearest neighbours (KNN) algorithm has been used to forecast CCI of USA from (continued)

Continued

Author	Tools	Description	
[11]	Artificial Neural Network	betterment in forecasting. For analysing the variations in building CCI	
		based on the average price level, historical data of economic variable are collected, and the relationship between CCI and economic factors are revealed in Turkey ANN is very frequently used software for accurate results as a future prediction tool since it uses back propagation algorithm to show better	
[12]	Support Vector Machine, Artificial Neural Network	results. It records better prediction because it programmed to follow the same function done by neurons in human brain. Support Vector Machine (SVM) is used in the prediction, and it derives precise output on forecasting solar and wind energy resources. SVM is compared with Artificial Neural Networks (ANN). The results of SVM play a better role	

2.1. Construction Cost Index

This part of the paper explains a deriving formula for the CCI, and how to fix the value of CCI [9]. In India, concrete structures are prevalent as compared with other kind of structures such as steel and lumber. The Construction materials such as steel, cement, sand and bricks are the key elements to fix the CCI. Weights, production and unit price of the key elements are involved in fixing the value of CCI. Steel has the highest impact to fix the CCI (54%); followed by cement, sand and bricks, which are 30, 10, 6%, respectively. Cement and steel have a significant role in ruling the CCI, which is around 84% of CCI.

$$CCI = W_B \times Ui_B + W_S \times Ui_C + W_G \times Ui_G$$

where W_B : Weight of bricks (1,000 nos) in CCI W_S : Weight of Steel (ton) in CCI W_C : Weight of Cement (ton) in CCI W_G : Weight of Sand (m³) in CCI Ui_B : Unit price of bricks for i year Ui_S : Unit price of steel for i year Ui_C : Unit price of



cement for i year U_{iG} : Unit price of sand for i year 'i: particular base year.

The weights of CCI calculated formulas are shown below

$$W_{B} = \frac{\frac{(PB \times 100)}{(PB + PS + PC + PG)}}{UbB}$$

$$W_{S} = \frac{\frac{(PS \times 100)}{(PB + PS + PC + PG)}}{UbS}$$

$$W_{C} = \frac{\frac{(PC \times 100)}{(PB + PS + PC + PG)}}{UbC}$$

$$W_{G} = \frac{\frac{(PG \times 100)}{(PB + PS + PC + PG)}}{UbG}$$

where *PB*: Production of Bricks at the *i. PS*: Production of Steel at the *i. PC*: Production of Cement at the *i. PG*: Production of Sand at the *i. UbB*: Unit Price of the Bricks at the *i*/1,000 nos. *UbS*: Unit Price of the Steel at the *i*/ton. *UbC*: Unit Price of the Cement at the i/ton. UbG: Unit Price of the Sand at the *i*/m³.

2.2. Research methodology

By considering the objectives of this research, the work starts with the identification of problem related to the prediction of construction cost. There are a lot of criteria that control the flow of construction cost; here (Fig. 1) CCI is picked up to predict problem identification, which is the primary process in this methodology. The problem identification is proposed to forecast CIDC's CCI with the help of predicting models. The next step in the methodology is to identify through previous literature to find a clear idea to take over the research work further. The existing works were analysed, and a few works referred for this research by considering the limitations, and the next step followed.

The next process is to collect the raw data for analysing those data. These are some factors, which influence CCI that are collected and analysed for predicting with the help of forecasting models. Those factors are considered as variables. Each variable is collected through official government websites in India.

These collected factors are then subjected to correlation method using the standard formula. The correlation should be done between CCI and individual factors that are collected, and correlation factor, which results that a value greater than 0.75 must be taken into account whereas factors less than 0.75 are ignored. Then, correlated factors that are chosen for further research are proposed to involve in analysing them. The prediction tools such as ANN, SVM and smoothing techniques are used to analyse the correlated factors and are compared for better results. The best prediction result can be derived from this analysis. The comparison can be made with the help of errors obtained in each model, namely Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Based on the error comparison, the prediction can be analysed, and it selects the best one to be applied for CCI further. The accurate results are used to predict future CCI such that constructs cost can be controlled, and project budget flow should be normal even in cases of mid-term and long-term forecasts.

2.3. Data collection

Based on the last 23 months Construction cost index reports, infrastructure development expenses are rapidly increasing in Indian state, Delhi when compared to the leading states' cities such as Bangalore, Chennai and Kolkata (AravindJayaraman2014-website link). Delhi building CCI data are collected as a predicted variable. To make the analysis of raw data, the group of data should be collected for each variable. Usually, eight explanatory variables (Consumer Price Index (CPI),

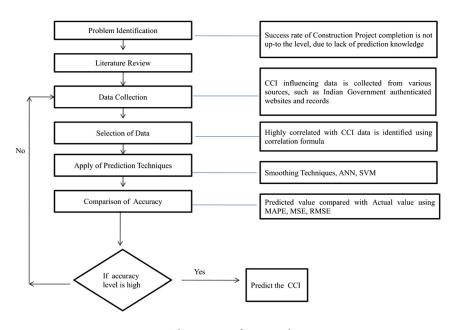


Fig. 1. The process of CCI prediction



employment level in construction, housing starts, money supply, Project Price Index (PPI), Gross Domestic Product (GDP), building permits and crude oil price) were used to forecast CCI in United States with the help of Vector (VEC) models [13]; Forecasting Construction Cost Index using Multivariate Time Series Models). Forty-one potential independent variables were used for predicting CCI using regression techniques and out of that five variables were selected in each of the six models using the chosen variables. Using selective empirical test, it is proved that CPI, Construction and Operation Plan (COP), Gross Domestic Product (GDP) and Building Plan (BP) are the most leading indicators of predicting CCI [10]. The relationship strength, between BCCI and aforementioned factors, is analysed with the help of historical data [11]. Consumer price index, crude oil price and production price index are considered as most consistent explanatory variables on applying for the vector of CCI using correction models [7].

From the literature review, it is decided to select eight numbers of different variables for the further work related to predicting CCI to calculate for future trends. The selected factors that are to be used as variables are listed as follows; population, unemployment rate, national income, GDP, construction price, CPI, CCI cement and CCI all commodities. Every variable is directly collected from government websites of India. Population, unemployment rate, money supply, GDP, and Interest rate as the important variables in developing of Construction Cost prediction model [14].

2.4. Population

Commonly in India, population values are calculated every ten years once. Based on the ten years value, it is converted to monthly value from 2003 to 2018 by using the linear regression techniques. Considerably, the population immediately impacts CCI due to the fact that the increase in population will obviously increase the need for lodging.¹

2.5. Unemployment Rate

Unemployment is all around perceived as bothersome. While market analysts and scholastics make persuading contentions that there is a certain common degree of joblessness that cannot be eradicated, raised unemployment forces considerable expenses on the individual, the public and the nation. More terrible yet, most of the expenses are of the dead misfortune assortment, where there are no counterbalancing increases to the costs that everybody must bear. Contingent upon how it is deliberate, the joblessness rate is available to translation. It is understood that increased unemployment reduces CCI.

2.6. Money supply (M4)

Here, M4 indicates the 4th type of money.² The money supply is categorised in three types by Reserve bank of India

since 1977, such as Reserve bank money (M0), Narrow supply (M1 and M2) and Broad supply (M3 and M4)

M4 = M3 + all kind of Post office deposits

M3 = M1 + timely deposits of banking system

M2 = M1 + all kind of Post office deposits

M1 = Public currency + All kind of bank accounts such as current account and saving account + RBI

M0 = Reserve bank linked money, and it is called High-powered money.

The money supply is the complete load of cash that is available for use in an economy on a particular day. This incorporates every one of the notes, coins and requests stores held by the general population on such a day. For example, cash demand and cash supply are a stock variable. One significant point to note is that the supply of cash kept with the administration, national bank, and so on is not considered in the money supply. This cash is not in real flow in the economy, and subsequently does not frame a piece of the fiscal inventory.

2.7. Consumer Price Index (CPI)

A CPI estimates changes in the value level of a weighted normal market bushel of purchaser merchandise and enterprises bought by families. The computation associated with the estimation of CPI is extensive. Different classifications and sub-classifications have been made for characterising utilisation things and based on customer classifications like urban or rural. Given these files and sub files, the last generally speaking file of cost is determined for the most part by measurable national organisations. It is one of the most significant measurements for an economy and is commonly founded on the weighted normal of the costs of wares. It gives a thought of the typical cost for essential items.

2.8. Long term interest rate

Long term interest alludes to government bonds developing in ten years. Rates are principally controlled by the value charged by the loan specialist, the hazard from the borrower and the fall in the capital worth. Long term interest is for the most part midpoints of daily rates, estimated as a rate. These financing costs are suggested by the costs at which the administration securities are exchanged on money related markets, not the financing costs at which the advances have been given. In all cases, they allude to bonds whose capital reimbursement is ensured by governments. Long term interest is one of the determinants of business speculation. Low long-term interest empowers interest in new gear and high financing costs decrease it. Speculation is a significant wellspring of financial development.

2.9. Domestic credit growth

The expression "domestic credit" refers to loan or credit that a nation or domain's national bank makes accessible to



¹https://www.ceicdata.com/en/indicator/india/population

²http://ncert.nic.in/textbook/textbook.htm?leec1=3-6

borrowers inside a similar region. This may incorporate business banks and even include the administration itself. As a rule, an administration, regardless of whether this is on a nearby, city or national scale, needs to obtain cash to support its tasks and offer administrations to its constituents. It subsequently acquires a government obligation. This may either be outside obligation, which is that owing to outer money-related substances, or inward obligation, which is owed to loan specialists inside a similar nation. Obviously, local credit falls under the subsequent classification.

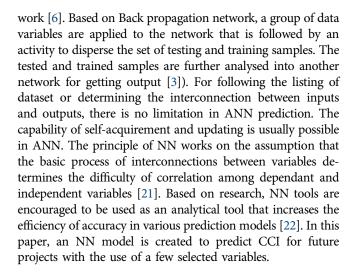
2.10. Smoothing techniques

Smoothing techniques are continued to be used for prediction since 1950, which only can replicate a framework or model. The model achieves a long term by including a plan designing. After that, this technique was extended and modified further [15-17]. In this research, the smoothing technique is used for forecasting CCI in India. From literature reviews, it is revealed that due to relative uniformity and perfection in execution of data, smoothing techniques are generally preferred for forecasting on consideration of patterns, regular and alternate features of the statistical data needful to human intervention [15]. Recent data are given more preference than past observation for analysis in smoothing technique and hence new data tend to be used for better impact. The exponential function is generated for assigning exponentially decreasing weights over a while and based on that constant is chosen on relatively decreasing weights of previous data [18, 19]. Smoothing techniques are run with the help of smoothing constant α . The size of that decides the amount of error in earlier prediction. The value of smoothing constant ' α ' has been found to be from 0 to 1, and the value '0' indicates that the case is insensitive, while the value '1' indicates that case is to be balanced between 20 and 30% of error with that of earlier one [20]. There are some limitations on running the smoothing technique when data with only random variation is used. Also, there should be no trends or seasonality in data. In addition to that, this technique shows perfection only in short term forecasting. Hence the earlier data have a steady design, smoothing techniques are preferred here. The forecast value is derived as follows, selection of α is one of the important parts in smoothing techniques. Generally, it is selected based on the least MAPE or least sum of mean square.³ In this analysis, 0.1 is considered as the value of α .

2.11. Artificial neural network

In future, any type of given historical data can be predicted by using NNs. The algorithms and calculating process of the NN are found to be similar to the neural system structure in human brain [9]. Back propagation network is the basic criteria to be used in the NN for analysing the input data and further predicting the future data, i.e. CCI in the present research

 3 https://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ForecaSmo. htm



2.12. Support vector machine (SVM)

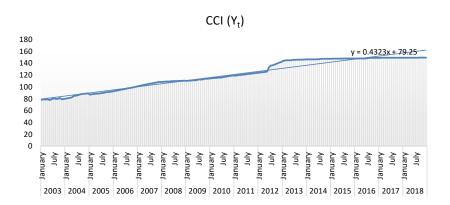
SVM is a proposed model of statistical method, and the learning process is used to find prediction function [23], SVM is a tool basically assigned to the class of kernel methods. The historical data are used in time series analysis [24]. ANN is usually adopted to forecast outputs in different criteria for its accuracy and better results, and there are some cases where the result of ANN has inaccuracy and hence to overcome those limitations SVM method has been introduced [12]. Based on functions of augmentation problem, SVM can be classified into two types, namely least square SVM and ESVM [24]. SVM and ANN methods are compared for precision in predicting wind power, and SVM has shown better accuracy [25]. Till now, SVM technique has not been applied in the field of construction issues. This paper inputs an appropriate result to predict CCI with greater accuracy by the help of SVM and those results are compared with that of ANN. In the field of construction, the methods showing betterment are going to be revealed in further work.

3. RESULT AND DISCUSSION

Allocation of validating part from whole data is not easy. Akindele [26] has suggested that 20% of data is to be evaluated from the entire data after trained of 80% data. Generally, a smaller number of validating data process gives the better result as expected. The square root value of the whole data has taken as validating data. From the 192 available data, 14 data were involved for validation, and the remaining 178 data were used for calibration. These 178 data were separated into two parts such as training and testing, one-third of 178 data were allocated for testing, and the rest of data were used for training.

Raw data and modified data had been taken for error analysing of the CCI. Two kinds of modified data and one raw data, totally three types of data were involved in this error analysis. From Graph 1, Y_t is a raw data of CCI, linear function formula y = 0.43x + 79.25 gets it from the trend





Graph 1. Raw data of CCI and time

Table 1. Error values of different tools

		Error comparison		
Sl.No	Name of tool	MAPE	MSE	RMSE
1	Smoothing techniques	0.008	0.0052	0.072
2	Artificial neural network	0.78	1.87	0.37
3	Support vector machine	5.78	112.28	10.60

analysis. The difference between y and Y_t data are taken as one kind of modified data. The ratio between the differences of original value, minimum value and maximum value, and the minimum value is known as another modified data. After the detailed evaluation, the first kind of modified data gave the competitive results when compared to the other data. Prediction tools such as Smoothing techniques, Artificial Neural Network (ANN) and Support Vector Machine (SVM) are involved in finding the error, and these error values are tabulated below. Low value of MAPE, MSE and RMSE is declared by the best prediction tools. When compared to the modified data, raw data is given the best output in the present research.

As per Table 1 values, smoothing techniques have given the low error value when compared to the other methods. The future CCI value can be predicted by smoothing techniques in low errors.

4. CONCLUSION

Prediction of Construction Cost Index is not an easy task, because volatile variances are there between the influenced factors throughout the year. MSEs are having some limitations that are all kinds of prediction tools is assumed that the variance between the influenced factors is constant for non-available data. Further, one of the major constraints is collecting data in India. Some of the non-available data are managed by the econometric method such as time series analysis, because useful outputs can get it from minimum data. After interrogating the data carefully, predictions are done by a non-econometric method such as smoothing techniques, ANN and SVM. As mentioned in the output, the

smoothing technique is bestowing and provides a better result when compared to the other forecasting methods. The predicted CCI volatility protects quantity surveyor from underestimation and overestimation. The present study also recommends that some of the other market available prediction tools can be used in future to predict the CIDC's CCI. It may be interrogated to identify; how different models perform while predicting the accuracy in the index.

ABBREVIATIONS

CIDC	Construction Industry Development Council		
CCI	Construction Cost Index		
MAPE	Mean Absolute Percentage Error		
MSE	Mean Square Error		
RMSE	Root Mean Square Error		
W_B	weight of bricks (1,000 nos) in CCI		
W_S	weight of steel (ton) in CCI		
W_C	weight of cement (ton) in CCI		
W_G	weight of sand (m ³) in CCI		
Ui_B	unit price of bricks for i year		
Ui_S	unit price of steel for i year		
Ui_C	unit price of cement for i year		
Ui_G	unit price of sand for i year		
i	particular base year		
PB	Production of Bricks at the i		
PS	Production of Steel at the i		
PC	Production of Cement at the i		
PG	Production of Sand at the i		
UbB	Unit Price of the Bricks at the i/1,000 nos		
UbS	Unit Price of the Steel at the i/ton		
UbC	Unit Price of the Cement at the i/ton		
UbG	Unit Price of the Sand at the i/m ³		
ANN	Artificial Neural Network		
SVM	Support Vector Machine		
CPI	Consumer Price Index		
VEC	Vector		
PPI	Project Price Index		
GDP	Gross Domestic Product		
COP	Construction and Operation Plan		
BP	Building Plan		
ENR	Engineering News Record		



M0 Reserve Bank Money Supply
M1 and M2 Narrow Supply of Money
M3 and M4 Broad supply of Money
RBI Reserve Bank of India
NN neural network

PERT Program Evaluation Review Techniques

REFERENCES

- [1] J.-W. Xu, "Stochastic forecast of construction cost index using a cointegrated vector autoregression model," *J. Manag. Eng.*, pp. 10–8, 2013.
- [2] S. Hwang, "Time series models for forecasting construction costs using time series indexes," *J. Construction Eng. Manag.*, pp. 656–62, 2011.
- [3] P. Williams, "Predicting changes in construction cost indexes using neural networks," *J. Construction Eng. Manag.*, pp. 306–20, 1994.
- [4] D. J. Lowe, "Predicting construction cost using multiple regression techniques," *J. Construction Eng. Manag.*, pp. 750–8, 2006.
- [5] B. Ashuri, "Time series analysis of ENR construction cost index," J. Construct. Eng. Manag., pp. 1227–37, 2010.
- [6] H. Nam, Time Series Analysis of Construction Cost Index Using Wavelet Transformation and A Neural Network. Construction Automation Group, I.I.T. Madras, 2007, pp. 453–6.
- [7] S. Shahandashti, "Forecasting engineering news-record construction cost index using multivariate time series models," *J. Construction Eng. Manag.*, pp. 1237–43, 2013.
- [8] J. Wang, "Predicting ENR construction cost index using machine-learning algorithms," *Int. J. Construct. Edu. Res.*, pp. 1–17, 2016.
- [9] Y. Elfahham, "Estimation and prediction of construction cost index using neural network, time series, and regression," *Alex*andria Eng. J., pp. 499–506, 2019.
- [10] J. Wang, "Predicting ENR's construction cost index using the modified K nearest Neighbors algorithm," Construct. Res. Cong., pp. 2502–9, 2016.
- [11] N. Atabeyli, "Using historical data of economic variables in investigating variations in building construction cost index," Creative Construction Conference, Turkey, pp. 718–23, 2018.

- [12] A. Zendehboudi, "Application of support vector machine models for forecasting solar and wind energy resources: A review," *J. Clean. Prod.*, pp. 272–85, 2018.
- [13] B. Ashuri, "Forecasting construction cost index using multivariate time series models," *J. Construction Eng. Manag.*, pp. 202–12, 2013.
- [14] J. Heng, "Construction price prediction using vector error correction models," *J. Construction Eng. Manag.*, 2013.
- [15] R. J. Hyndman, "A state space framework for automatic fore-casting using exponential smoothing methods," *Int. J. Forecast.*, vol. 18, no. 3, pp. 439–54, 2002.
- [16] J. K. Ord, "Estimation and prediction for a class of dynamic nonlinear statistical models," J. Am. Stat. Assoc., 311-5, 1997.
- [17] J. W. Taylor, "Exponential smoothing with a damped multiplicative trend," *Int. J. Forecast.*, pp. 715–25, 2003.
- [18] Gardner, "Exponential smoothing: the state of the art," *J. Forecast.*, pp. 1–28, 1985.
- [19] A. L. Maia, "Holt's exponential smoothing and neural network models for forecasting interval-valued time series," *Int. J. Forecast.*, vol. 27, no. 3, pp. 740–59, 2011.
- [20] A. A. Homayoun Khamooshi, "Project duration forecasting using earned duration management with exponential smoothing techniques," *J. Manag. Eng.*, vol. 33, no. 1, pp. 04016032, 2016.
- [21] T. M. Elhag, "Cost Modeling: neural networks vs regression techniques," Int. Conf. on Construction Information Technology (INCITE), Langkawi, Malaysia, Construction Industry Development Board Malaysia (CIDB), 2004.
- [22] J. Cetkovic, "Assessment of the real estate market value in the European market by artificial neural networks application," *Complexity*, pp. 1–10, 2018.
- [23] O. Kramer, Short-Term Wind Energy Forecasting Using Support Vector Regression, University of California, Berkeley, International Computer Science Institute, 2011, pp. 1–10.
- [24] J. Zeng, "Short-term wind power prediction using a wavelet support vector machine," *IEEE Trans. Sust. Energ.*, pp. 255–64, 2012.
- [25] M. G. Giorgi, "Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN)," *Energies*, pp. 5251–72, 2014.
- [26] Akindele, "Corruption and economic retardation: A retrospective analysis of Nigeria's experience since independence," Readings in the Political Economy of Nigeria since Independence, 1990.

Open Access. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

