Developing a machine learning algorithm to investigate the role of energy consumption in sustainable development: A case study of China

Ali Hashemizadeh[1,a], Faezeh Zareian Baghdad Abadi[2,3*]

[1]College of Management, Shenzhen University, Shenzhen, Guangdong, 518060, PR China
[2]China Center for Special Economic Zone Research, Shenzhen University, Shenzhen, Guangdong, 518060, PR China
[3]College of Economics, Shenzhen University, Shenzhen, Guangdong, 518060, PR China

[a]E-mail address; ahz@szu.edu.cn
*E-mail address: zareian@emails.szu.edu.cn
*Corresponding Author

Abstract. Nowadays, the power and strengths of machine learning methodologies encourage all decision-makers to use the benefits of these approaches in their policy planning. The increasing complexity of economic planning in terms of various impactful factors and targets necessitates the comprehensive analysis of influential parameters in economic development using developed machine learning methods. While considering limited factors in economic investigations lead us to unreal conclusions, in this study, five indicators are considered (Total Energy Consumption, Birth Rate, Wastewater Treatment Rate, Number of Repairs Enterprises, and Number of Existing Environment and Public Facilities Management Enterprises) to represent circular economy, social, environmental and energy consumption aspects of sustainable development, respectively. A modern machine-learning algorithm, named XGBoost, had been developed to investigate the impact of mentioned indicators on the sustainable development of the study region. According to panel data of Guangdong Province, China, from 2010 to 2019, the findings state that the circular economy has the top priority in sustainable development. The repairing and reusing of resources as a circular economy solution greatly impact sustainable development. Based on these results, some policy recommendations are provided for appropriate economic development considering social and environmental issues in Guangdong Province.

Keywords: Machine learning, Energy consumption, Renewable energy, Sustainable development, Guangdong, China

1 Introduction

Sustainable development is a top priority, especially in light of the substantial environmental problems caused by the significant growth in the use of natural resources [1]. Nowadays, global society is focusing much more on sustainable development as one of

A. Rauf et al. (eds.), *Proceedings of the 3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, Atlantis Highlights in Engineering 20,
https://doi.org/10.2991/978-94-6463-262-0_20

the grand challenges for human beings in the 21st century, caused by environmental issues due to rapid economic growth [2]. The long-term growth of countries is the emphasis of sustainable development. Since sustainable development has different aspects, evaluating different items is necessary to reach practical results.

Several researchers have explored a broad range of indicators connected with sustainable development in this area, but there are still certain gaps that need to be addressed based on updated datasets [3]. The first gap is about considered indicators. While some indicators are proper for some case studies, they did not correctly represent other areas' conditions. The circular economy is one of the most significant paradigms that researchers and authorities have recently paid more attention to. Economic boom while lessening resource and power consumption may be possible thanks to the circular economy (CE) [4]. From a social aspect, China's government changed its child policy [5]. It has various social and economic impacts that should be investigated. Besides, while some regions are significant areas, they did not get enough attention in the existing literature. In this research, we are going to fulfill these gaps by considering appropriate indicators such as the circular economy and birth rate to cover all aspects of sustainable development, using updated data to reflect current conditions. Also, in this study, Guangdong Province, China as an important area of China, is chosen for analysis to propose practical policy implications. Among all the provinces in China, Guangdong Province has had the quickest economic growth. Figure 1 depicts Guangdong's geographic position and regional segmentation [6].

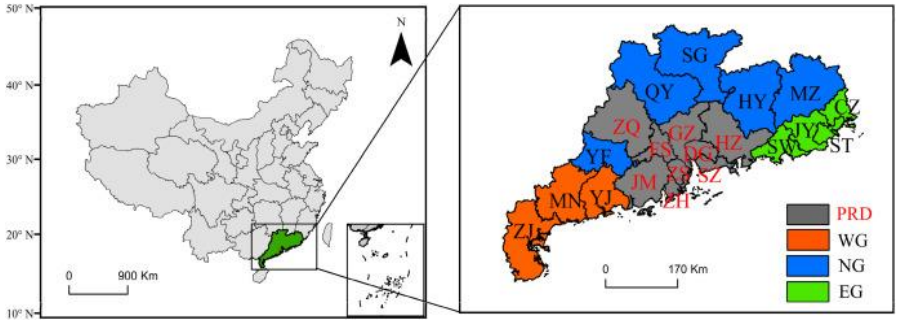

Fig. 1. - Guangdong's geographic area and regional classification [6].

2 Data collection and methodology

2.1 Collected data

This research selects six indicators to investigate their impact on sustainable development, representing energy, social, circular economy, and environmental aspects. The data used in this research were obtained from [7] from 2010 to 2019 for ten years in quarterly frequency. These indicators are introduced in Table 1 and Table 2.

Table 1. – The introduction of the impactful factors on sustainable development

Category	Indicator	Symbol
Economic development	Gross domestic product (GDP)	T
Energy aspect	Total Energy Consumption	F1
Social aspect	Birth Rate	F2
Circular aspect	Wastewater Treatment Rate	F3
	Repairs Enterprises	F4
Environmental aspect	Environment and Public Facilities Management Enterprises	F5

Table 2. – The definition of the impactful factor on sustainable development

Indicator	Definition
Gross domestic product (GDP)	It is the most comprehensive and accurate quantitative estimate of a region's overall economic development.
Total Energy Consumption	It states the sum of energy used for electricity, transport and heating in all industries.
Birth Rate	The birth rate is the number of live births per 1,000 in a population at a specific time.
Wastewater Treatment Rate	It demonstrates the share of the regional people connected to a wastewater treatment plant in percentage.
Repairs Enterprises	It shows the number of Existing Enterprises in Resident Services and Repairs field in Guangdong Province.
Environment and Public Facilities Management Enterprises	It reflects the number of Existing Environment and Public Facilities Management fields in Guangdong Province.

The descriptive statistics of used variables in this study are provided in Table 3, while the related scatter plots are illustrated in Figure 2.

Table 3. - Descriptive statistics of the used data

Indicator	T	F1	F2	F3	F4	F5
Observations	40	40	40	40	40	40
Min.	9129	26070	10	79	21551	6927
Max.	26690	34142	14	97	127651	30302
Average	18067	30067	12	90	57394	15748
Median	17482	29662	11	91	44522	13955
St. Dev.	4863.27	2241	1	5	35195	7297
Skewness	0.14291	0.221	1	0	0.7015	0.488

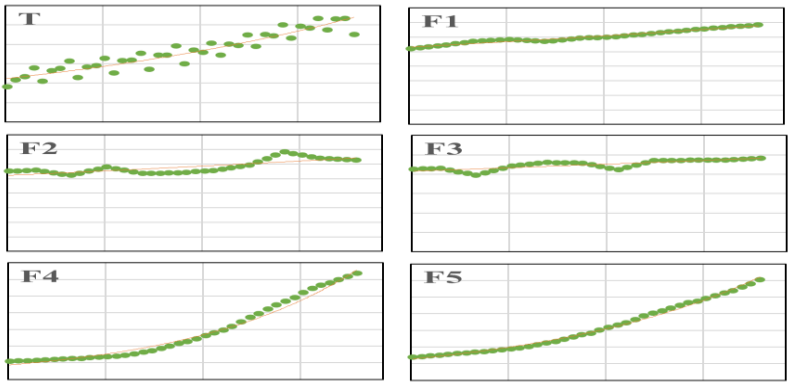

Fig. 2. - The scatter plots of the considered indicators.

2.2 Methodology: machine-learning algorithm

The XGBoost technique, created in 2016 for regression and classification problems, stands out as one of the effective solutions when the data obtained for a framework do not satisfy many of the constraints of standard regression analysis [8]. Compared to the conventional ones, this approach can more match intricate correlations between variables. XGBoost incorporates the prediction of several learners into a single model in an iterative procedure [9]. We used the perspective nonlinear effects of each of the provided factors in light of the strong connections among several indicators that might affect sustainable development. Hence, an XGBoost-based model was created. The schematic design of the XGBoost regression method is shown in Figure 3.

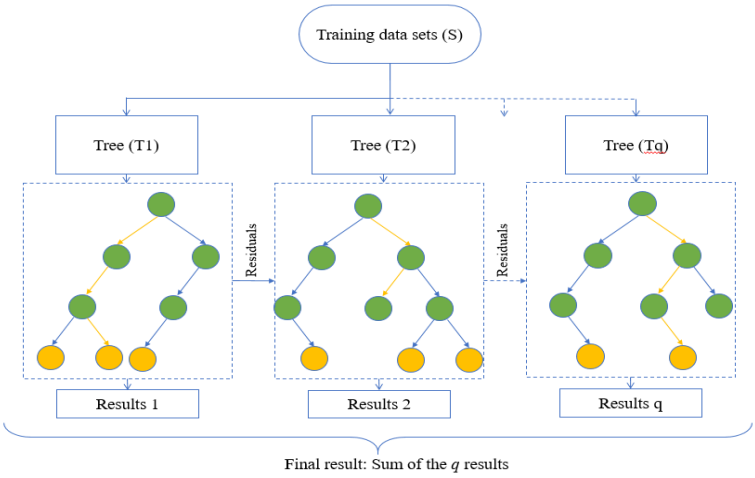

Fig. 3. - Schematic framework of XGBoost RA model.

The fundamental idea behind XGBoost is to use an avaricious method to learn each basis tree. By regularly developing fresh decision trees to fit the residuals of the previous prediction, the residual error between the anticipated value and the actual one is also continuously lowered, increasing precision in forecasting. The optimization objective is alternatively represented by the following calculation [10]:

$$P^{(t)} = \sum p\left(y_q, \widehat{y_q}^{(t-1)} + f_t(x_q)\right) + \Omega(f_t) \tag{1}$$

where p denotes the loss function, $f(x)$ represents the tree structure, and y_q corresponds to the ensemble tree's expected outcome. The previous prediction and a new feature must both be taken into consideration during each repeat of the process. The portion of the computation has been modified to look like this:

$$F^{(t)}(x) = F^{(t-1)}(x) + f_t(x) \tag{2}$$

Because the leaf node value of each tree reflects the modified output of the prior round fit, summing together all leaf nodes yields the final prediction outcomes of the tree structure.

3 Results and discussion

In this section, the obtained results from the implemented model are presented. This study considered Total Energy Consumption, Birth Rate, Permanent Residents, Wastewater Treatment Rate, Repairs Enterprises, and Environment and Public Facilities Management Enterprises as impactful factors in sustainable development, which GDP represented. Figure 4 shows the impact of mentioned indicators on the sustainable development of the study region.

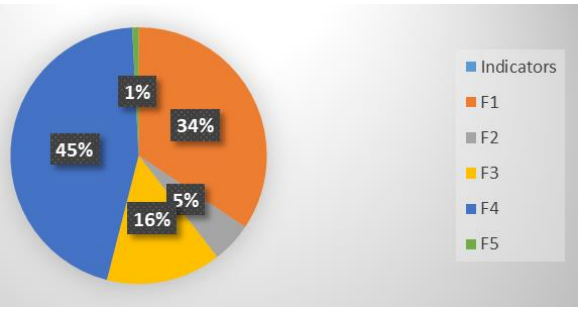

Fig. 4. - The influence of considered indicators in the economic development of the study region.

Based on the obtained results, the most significant indicator is F4 (Number of existing Repairs Enterprises). Based on the circular economy foundations, repairing and reusing goods are one of the main solutions to achieve sustainable development. At the same time, it improves the efficiency of goods consumption and reduces produced

waste of society. Then, it can boost economies appropriately. This result aligns with (Lahane and Kant, 2022), confirming our findings. After F4, the second most important indicator is F1 (Total Energy Consumption). Energy is the basement of all economic activities. Increasing energy consumption in efficient ways can be a prerequisite for economic development. Total energy consumption has different impacts on societies, and the importance of this parameter has been emphasized in this study, similar to the existing literature review [11]. The third indicator is F3 (Wastewater Treatment Rate). It belongs to the circular economy aspect of our model. The key role of water consumption and water treatment is clear, especially with the scarcity of water resources. The circular economy concept could improve sustainable development by improving the reuse of scarce resources. This result is aligned with [12].

4 Conclusion and policy implications

Machine learning is the most powerful tool for decision-making procedures in the current complex environment. Economic activity has recently raised several issues about unsustainable growth that might hurt people and the functioning of the environment. In order to address these difficulties, decision-makers must embrace creative and long-term solutions. The circular economy is seen as a different approach to conventional economic planning. For value gain, the circular economy aids in maintaining the closed-loop status of the resources, parts, and finished goods even beyond their end-of-life phase. Consequently, it lessens the amount of trash produced across the economy, lowers the amount of raw materials needed for production, and enhances the green environment. The current work suggests identifying and evaluating the influential variables in sustainable development and has expanded a model to consider circular economy using the XGBoost regression algorithm.

To consider economic development, energy aspect, social aspect, circular economy aspect, and environmental aspect of sustainable development, these indicators are selected respectively: Gross domestic product (GDP), Total Energy Consumption, Birth Rate, Wastewater Treatment Rate, Repairs Enterprises, and Environment and Public Facilities Management Enterprises. Our findings illustrate that number of repair enterprises is the most important factor among the selected factors, followed by total energy consumption, wastewater treatment rate, and birth rate. These results confirmed the significant role of the circular economy in sustainable development. The circular economy can potentially lessen the economy's reliance on nonrenewable resources while boosting sustainable development. Furthermore, the circular economy may help get us closer to sustainable development's overall aim by encouraging more sustainable practices like recycling and repairing.

Acknowledgment

Natural Science Foundation of China (NSFC) supported this work [grant number: 72250410376]. The authors would like to thank the editor for editing the manuscript and the reviewers for their detailed and helpful comments.

References

1. A. Dabbous and A. Tarhini, "Does sharing economy promote sustainable economic development and energy efficiency? Evidence from OECD countries," *J. Innov. Knowl.*, vol. 6, no. 1, pp. 58–68, 2021, doi: 10.1016/j.jik.2020.11.001.
2. United Nations, "Global Sustainable Development Report 2019: The Future is Now – Science for Achieving Sustainable Development," New York, NY, USA, 2019.
3. N. Ahmad, L. Youjin, S. Žiković, and Z. Belyaeva, "The effects of technological innovation on sustainable development and environmental degradation: Evidence from China," *Technol. Soc.*, vol. 72, p. 102184, 2023, doi: https://doi.org/10.1016/j.techsoc.2022.102184.
4. D. Knäble, E. de Quevedo Puente, C. Pérez-Cornejo, and T. Baumgärtler, "The impact of the circular economy on sustainable development: A European panel data approach," *Sustain. Prod. Consum.*, vol. 34, pp. 233–243, 2022, doi: https://doi.org/10.1016/j.spc.2022.09.016.
5. X. Chen, I. Calder, and B. Mak, "China's second-child baby boom and fertility tourism: Strategic considerations for Malaysia," *J. Destin. Mark. Manag.*, vol. 15, p. 100377, 2020, doi: 10.1016/j.jdmm.2019.100377.
6. H. Mo *et al.*, "Potential impact of industrial transfer on PM2.5 and economic development under scenarios oriented by different objectives in Guangdong, China," *Environ. Pollut.*, vol. 316, p. 120562, 2023, doi: 10.1016/j.envpol.2022.120562.
7. National Bureau of Statistics of China, "China statistical yearbook," 2021.
8. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
9. A. Shehadeh, O. Alshboul, R. E. Al Mamlook, and O. Hamedat, "Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression," *Autom. Constr.*, vol. 129, p. 103827, 2021, doi: 10.1016/j.autcon.2021.103827.
10. M. Ye *et al.*, "Estimation of the soil arsenic concentration using a geographically weighted XGBoost model based on hyperspectral data," *Sci. Total Environ.*, vol. 858, p. 159798, 2023, doi: 10.1016/j.scitotenv.2022.159798.
11. A. Hashemizadeh, Q. Bui, and S. A. H. Zaidi, "A blend of renewable and nonrenewable energy consumption in G-7 countries: The role of disaggregate energy in human development," *Energy*, vol. 241, p. 122520, 2022, doi: 10.1016/j.energy.2021.122520.
12. S. Lahane and R. Kant, "Investigating the sustainable development goals derived due to adoption of circular economy practices," *Waste Manag.*, vol. 143, pp. 1–14, 2022, doi: https://doi.org/10.1016/j.wasman.2022.02.016.