Proceedings of the International Conference on Industrial Engineering and Operations Management
Manila, Philippines, March 7-9, 2023

Carbon Footprint Prediction in the Philippines Using Time Series in Data Science

Sweethy Jhoden Lim and Ma. Mica Ella Cortez
College of Computer Studies
Angeles University Foundation
Pampanga, Philippines
lim.sweethyjhoden@auf.edu.ph, cortez.mamicaella@auf.edu.ph

Dr. Rosanna Esquivel
Professor of Computer Science
College of Computer Studies Program Chair
Angeles University Foundation
Pampanga, Philippines
esquivel.rosanna@auf.edu.ph

Abstract

This study generally aimed to develop a model that can predict carbon footprint emissions based on the datasets obtained. The parameters that the researchers utilized were population and energy production from 1990 to 2019. The data were all chosen from the vital drivers of CO2 emissions identified through the literature review and had been collected using the open-sourced program WEKA. The three algorithms used in this study are Linear Regression (LR), Multilayer Perceptron (MLP), and Sequential Minimal Optimization Regression (SMOreg) to test which among these most-used algorithms in time series data mining prediction delivers outstanding results. These were participated in the trials given to the researchers to gather sufficient data to draw conclusions. The researchers determined that the MLP model is able to estimate greenhouse gas emissions at error levels that are acceptable in comparison to the actual values and test results. If the Philippines maintains its current trajectory, the carbon footprint emissions will eventually rise and probably reach 255,595.03 kt in the year 2030. There is a powerful correlation between population and energy production that contributes to the carbon footprint emissions within the Philippines.

Keywords

Carbon Footprint, Time Series Forecasting, Linear Regression, Multilayer Perceptron, and SMOreg.

1. Introduction

Humanity has lived alongside nature and has been a part of that environment. However, as our global population and economy have grown, our impacts on the environment grew as well. The protection of the world we live in is vital for the sustainability of our existence. When people realized that environmental issues were putting their lives at risk because of economic growth, they started working to improve the quality of the environment. According to a study from Huang et al. (2017), the widely accepted definition of the carbon footprint refers to human production or consumption-related gaseous emissions that contribute to global warming. One key concept in the fight against climate change is carbon footprint, which refers to the amount of greenhouse gases produced by human activities such as transportation, energy use, and food production. The carbon footprint is a common notion in talks about responsibility and climate change mitigation. In 2016, experts affiliated with the World Economic Forum classified global warming as the number one hazard to society and the economy Cann (2016). There is a strong correlation between carbon footprint, energy consumption, and energy supply in the Information and Communication Technology (ICT) and Entertainment and Media (E&M) sectors, making it vital to investigate energy use concurrently. Numerous countries are establishing a green economy as part of long-term development strategies to leave future generations with a habitable world. To achieve these goals, various analytical approaches have been developed to predict carbon footprints and identify ways to reduce them. In recent years, data mining techniques have gained interest as a powerful tool for assessing, planning, and decision-making processes in many industries, including environmental management.

By using advanced data analysis to identify patterns and trends, researchers can gain new insights into energy use and identify strategies for reducing carbon emissions.

1.1 Objectives

This study seeks to address the pressing issue of carbon footprint emissions by developing a comprehensive model that employs advanced data science techniques to predict the impact of various factors on carbon footprint generation. To achieve this goal, a wide range of data will be utilized, including population and energy production, which will be analyzed using time series methods. This study aims to raise awareness among individuals, including students and professors, regarding the causes of carbon footprint emissions. By gaining a deeper understanding of the factors that contribute to carbon footprint emissions, individuals and organizations can devise effective strategies to minimize their environmental impact and promote sustainability. This study hopes to contribute to these efforts by providing valuable insights and offering solutions to reduce carbon footprint emissions and mitigate their effects on the environment.

2. Literature Review

Many researches on carbon footprint emissions have been conducted as the technological transition in the ecological literature has progressed. The majority of these research are geared toward lowering carbon emission rates through the use of renewable energy, and policy recommendations have been developed in this context.

According to Ahmed et al. (2019), global warming as a product of human emissions has resulted in climate change worldwide. As mentioned in the study of Bello et al. (2018), they had investigated the impact of hydroelectricity consumption on the environment. They have used several techniques to perform this study: ecological footprint, carbon footprint, water footprint, and CO2 emission. After running several tests, it is suggested that these techniques can help reduce the pollution in hydroelectricity consumption. As indicated in the study of Fan et al. (2018), Beijing is essential to the research of carbon emission reduction since its carbon emissions have consistently been among the highest in the country. The results indicate that the total direct carbon footprint and the urban and rural direct carbon footprints associated with inhabitants' consumption in the national capital increase. The natural carbon footprint of urban inhabitants is primarily caused by electricity, gasoline, and heating power.

According to the research conducted by Anusooya et al. (2019), forecasting the peak time load among data centers and dispersing the load will reduce the information center's power consumption and carbon emissions. Reducing carbon emissions through decreasing energy usage in a data center will have an effect on the environment, which can lead to a smaller carbon footprint. Xu et al. (2020) mentioned that numerous studies had been conducted to determine the elements that affect CO2 emissions. Multiple variables, including population activities, energy consumption patterns, economic growth, innovation and technology, urbanization processes, and government regulations, are used to explain their impact on carbon emissions. Wandana et al. (2021) stated that those with a higher population, economy, manufacturing, and process industries are often the highest emitters of greenhouse gases, implying they have a larger carbon footprint than countries with a smaller population, economy, manufacturing, and process industries.

Li et al. (2018) analyzed the carbon emissions of the Beijing-Tianjin-Hebei region, which relies heavily on fossil fuels. They developed a novel forecasting method called support vector - extreme learning machine (SVM-ELM) that utilized support vector machine kernels to optimize the outer learning machine algorithm. They demonstrated that SVM-ELM approaches were more accurate in predicting carbon emissions than support vector machines (SVM) or extreme learning machines (ELM). According to a study by Rezaei et al. (2018), researchers worldwide are interested in studies that attempt to forecast carbon emission patterns. Rezaei used the group method of data handling algorithm (GMDH) in the artificial neural network approach to predict the carbon emissions of four Scandinavian countries, including Finland, Sweden, Denmark, and Norway. Total primary energy consumption, the kind of fuels used, and GDP were used as indicators in this study from 1990 to 2016. In the study of Scherer and Milczarski (2021), they illustrated the management and evaluation of emission processes. They implemented unsupervised machine learning techniques to classify the operations into the following categories: optimal processes, near-optimal processes, far-from-optimal processes with low and high energy consumption, and processes with inaccurately input data due to human error. Sharma et al. (2021) studied the implementation of machine learning (ML) approaches in frozen vegetable production, where the authors used expert knowledge to evaluate the manufacturing process using classification methods such as support vector machine, random forest, and multilayer perceptron.

There have only been a few studies done on the estimation of greenhouse gases for the Philippines, as can be observed from the research that was discussed before. Furthermore, it has been seen that the rate of carbon emission is attempted to be approximated based on a few parameters in the majority of the research that has been conducted. Therefore, the goal of this in-depth study is to close this gap by using a number of parameters that are thought to affect the rate at which greenhouse gases are released.

3. Methods

For this study, the researchers decided to make use of the OSEMN methodology. The data included in the dataset used for this study was gathered from various sources and a new dataset was created. After creating the dataset, missing values were dealt by discarding excess data that did not correspond to the variables utilized in the study. The format of the dataset was also examined to verify that the application runs smoothly and without errors. Following the completion of the dataset, the correlation between variables was examined to ensure that the parameters selected were suitable for this project. To have a better grasp of the data, visualizations such as graphs were generated. Using WEKA's time series plugin, the data will then be prepared for training, testing, and modeling. The results of testing and modeling were analyzed to reach a conclusion for this study.

4. Data Collection

The study's variables were chosen from the vital drivers of CO2 emission identified through the literature review. The variables used are population and energy production. These will be used to predict the carbon footprint emissions of 2020-2030 based on the data from the 1990-2019 period. Different sources were utilized to obtain the variables used in this study. Table 1 shows a brief description of the variables that were used in this study.

Table 1. Details of the variables and sources.

Variables	Definitions	Sources
Population	Philippines' population from 1960-2020	The World Bank https://data.worldbank.org/indicator/SP.POP.TOTL?end=2020&locations=PH&most_recent_year_desc=false&start=1960
Energy Production	kWh	Energy Information Administration https://www.eia.gov/opendata/v1/qb.php?sdid=INTL.44-1-PHL-QBTU.A

The datasets used in this study was preprocessed through WEKA. WEKA is an open-source program that includes data preparation tools, implementation of numerous machine learning algorithms, and visualization tools to help you build machine learning approaches and apply them to real-world data mining situations. This program makes it simple to work with large amounts of data and train a computer with machine learning methods. The three most often used algorithms in time series data mining prediction models identified through the literature review will be used in this study. Mean Absolute Percent Error (MAPE) and Mean Squared Error (MSE) will be utilized in this study to assess the success of the models and choose the optimal model.

5. Results and Discussion

In the study, WEKA was used to estimate the future values of carbon footprint emissions. WEKA's Time Series Forecasting plugin was used to examine time series prediction, an essential component of the study's overall research methodology. To achieve accurate time series predictions, the study made use of three distinct algorithms that are commonly utilized in time series data mining models. These algorithms, namely Linear Regression (LR), Multilayer Perceptron (MLP), and Sequential Minimal Optimization Regression (SMOreg), have been identified as highly effective in predicting future trends in time series data.

5.1 Numerical Results

For this study, a new dataset was created using the collected data, which spans the years 1990 to 2019, and contains information on population and energy production. The dataset that was used for this study was divided into two sections. The data from 1990 to 2013 were used for training, and the data from 2014 to 2019 were used for testing estimation. In the study, the three models were initially trained with the data from the years 1990 to 2013.

Subsequently, beginning in 2014 and continuing through 2019, attempts were made to estimate the rates of carbon footprint emission. During these predictions, Mean Absolute Percent Error (MAPE) and Mean Squared Error (MSE) values were used as metrics to test the prediction success of each model. Table 2 shows the MAPE and MSE values obtained by each model.

Table 2. MAPE and MSE values obtained by each model.

	LR		MLP		SMOreg	
Year	**MAPE**	**MSE**	**MAPE**	**MSE**	**MAPE**	**MSE**
2014	1.3143	5921599.7	1.1129	4255522.35	1.6103	9785072.31
2015	1.3529	5778190.02	1.0139	4065794.02	1.5795	9254312.03
2016	1.3868	5965898.41	1.1832	5175571.81	1.7328	10606582.42
2017	1.4875	7063191.91	1.3642	6174840.59	1.7756	10857109.1
2018	1.5005	7486183.18	1.3337	6018063.49	1.6874	11148670.23
2019	1.5984	8263582.09	1.3807	6139552.81	1.4652	10501803.79

Following this step, the most successful model was identified by computing the average of the MSE and MAPE values acquired by each model. This allowed for the identification of the model with the best overall performance. Table 3 displays the derived MSE and MAPE values on an average for each model.

Table 3. Average MAPE and MSE values of each model.

	MLP	**LR**	**SMOreg**
Average MAPE	1.2314	1.4401	1.6418
Average MSE	5304890.85	6746440.89	10358924.98

According to the data in Table 3, the Multilayer Perceptron (MLP) has the best MAPE and MSE averages, which is why it is located in the first column of the table. Because it had the lowest average, this indicates that the MLP model had better success rates than the other models. The Linear Regression (LR) model had the second lowest average, and it was actually closer to the average of MLP than any other model. The third and final model, known as Sequential Minimal Optimization Regression (SMOreg), had the greatest average; hence, it was the least accurate of the three models that were utilized. In addition, the greenhouse gas emission rates estimated by each model for the years 2014 to 2019 were compared with the actual data in order to evaluate the success of the models. This evaluation was carried out in order to determine whether or not the models were accurate. Table 4 presents specific information regarding the actual data as well as the predictions of the rates of greenhouse gas emissions that were obtained from the model.

Table 4. Comparison of actual and forecasted values of each model.

Year	**GHG Actual Value**	**Forecasted Values**		
		LR	**MLP**	**SMOreg**
2014	183369.9951	179047.7149	**180754.7854**	175137.3063
2015	192300.0031	184816.1175	**187970.5795**	177449.0637
2016	204850.0061	191464.3611	**195633.1275**	180761.0503
2017	221059.9976	197829.4597	**200365.7098**	183762.214
2018	226940.0024	204768.4349	**206441.7592**	186783.7575
2019	234279.9988	**212641.6714**	209694.0034	191787.3927

The values that were highlighted in Table 4 were the estimated values that came the closest to matching the actual values for greenhouse gas emissions. As shown in Table 4, the MLP model was successful in obtaining values that are substantially quite similar to the actual values. As a result, the MLP model, which produced the best results overall, will be utilized in the study to forecast greenhouse gas emissions from the years 2020 through 2030.

5.2 Graphical Results

In order to provide accurate predictions for greenhouse gas emissions for the years 2020-2030, the study employed the Multilayer Perceptron (MLP) model. Figure 1 depicts the values that are forecasted in the years to come.

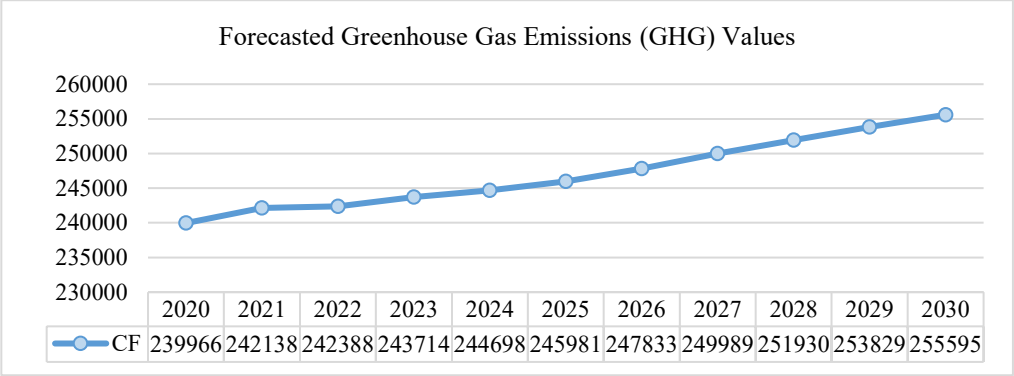

Figure 1. Forecasted greenhouse gas emission (GHG) values for 2020–2030.

As shown in Figure 1, the annual rates of greenhouse gas emissions in the Philippines have been steadily rising over the course of the last few years. In the process of predicting future carbon footprint emissions, population and energy production are considered to be essential factors. In the course of this study, it has been established that there is a connection between the levels of greenhouse gas emissions and population as depicted in Figure 2.

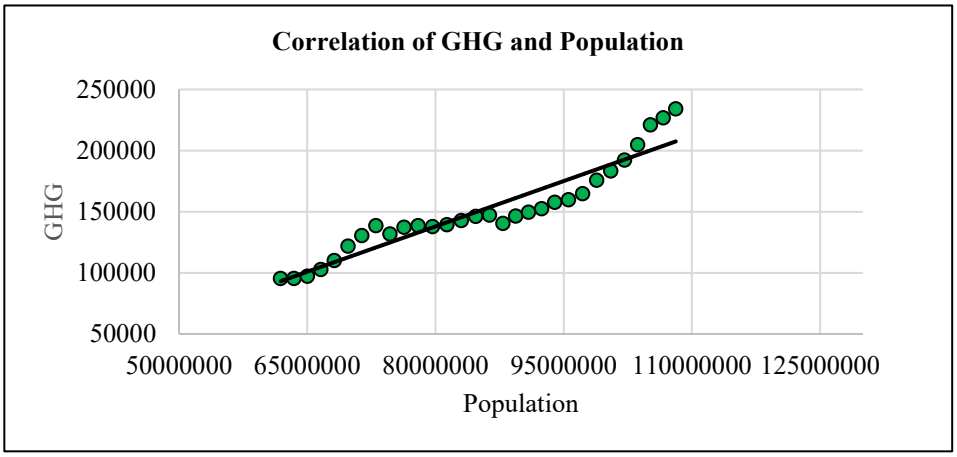

Figure 2. Correlation of GHG and population.

As shown in Figure 2, there is a strong positive correlation between greenhouse gas emissions and population. This study also demonstrated that there is a connection between the levels of greenhouse gas emissions and energy production as depicted in Figure 3.

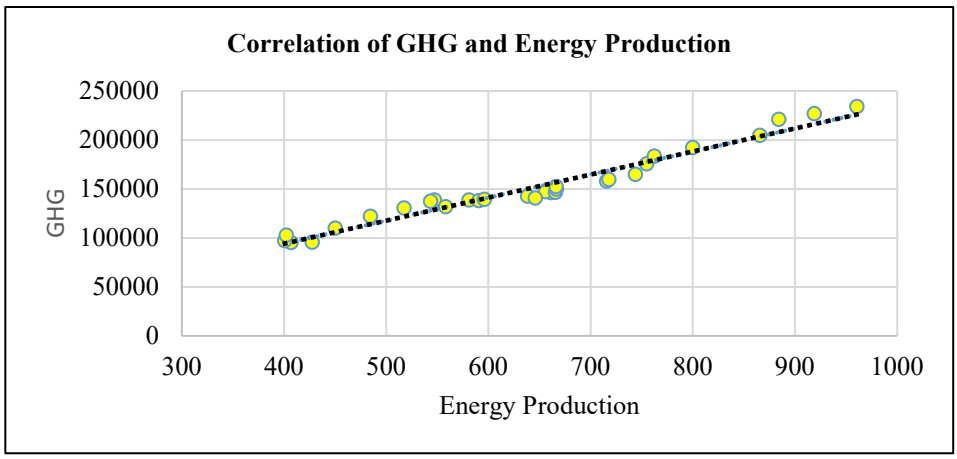

Figure 3. Correlation of GHG and energy production.

As shown in Figure 3, there is also a strong positive correlation between greenhouse gas emissions and energy production.

5.3 Proposed Improvements

Population and energy production were the only independent variables included in this study, so it is recommended to identify additional variables that affect carbon footprint emissions. The study made use of variables that are essential factors affecting carbon footprint emissions; it is recommended that the findings in this study be used to minimize CO2 emissions in the next few years. This study made use of data from 1990–2019; it is recommended to add new data to improve the success rate of the time series forecast.

5.4 Validation

Forecasts of greenhouse gas emissions for the Philippines were derived for the purpose of this study by the implementation of time series data mining techniques, using parameters including population and energy production variables. When looking at other studies that attempted to estimate emissions from carbon footprint, it was discovered that artificial neural network techniques were used in those studies. By employing artificial neural networks, Akyol and Ucar (2021) were able to make a prediction on Turkey's carbon footprint emissions. In spite of the fact that this study included substantially all of the same variables, it came to very different conclusions. It is presumed that the cause for these results was the utilization of only two variables and the incorporation of additional data into the dataset that was used for this study.

6. Conclusion

Based on the results obtained from the model, the following conclusions were drawn: If the Philippines continues on its current path, carbon footprint emissions will eventually rise and could reach 255,595.03 kt in 2030. There is a strong correlation between population and energy production, which contributes to the Philippines' carbon footprint emissions.

References

Abirami, S., and Chitra, P., Multilayer Perceptron. Multilayer Perceptron - an overview | ScienceDirect Topics. Available: https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron, March 16, 2022.

Ahmed, Z., Wang, Z., and Ali, S., *Investigating the Non-Linear Relationship between Urbanization and CO2Emissions: An Empirical Analysis, Air Quality, Atmosphere, and Health*, vol. 12, no. 8, pp. 945–953, 2019.

Akyol, M., and Uçar, E., Carbon footprint forecasting using time series data mining methods: the case of Turkey, *Environmental Science and Pollution Research*, vol. 28 no. 29, pp. 38552–38562, 2021.

Anusooya, G., Vijayakumar, V., and Narayanan, V. N., Reducing the carbon emission by early prediction of peak time load in a data center, *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4341–4348, 2019.

Bartlett, S., and Keir, S., Calculating the carbon footprint of a Geriatric Medicine clinic before and after COVID-19, *Age and Ageing*, vol. 51, no. 2, 2022.

Bello, M. O., Solarin, S. A., and Yen, Y. Y., The impact of electricity consumption on CO2 emission, carbon footprint, water footprint and ecological footprint: The role of hydropower in an emerging economy, *Journal of Environmental Management*, vol. 219, pp. 218–230, 2018.

Cann, O., What are the top global risks for 2016? World Economic Forum. Available: https://www.weforum.org/agenda/2016/01/what-are-the-top-global-risks-for-2016, February 5, 2020.

Chen, C., Liu, G., Meng, F., Hao, Y., Zhang, Y., and Casazza, M., Energy consumption and carbon footprint accounting of urban and rural residents in Beijing through Consumer Lifestyle Approach, *Ecological Indicators*, vol. 98, pp. 575–586, 2019.

Davis, E., Climate Change Analytics: Predicting Carbon Price and CO2 Emissions. *La Salle University Digital Commons*. Available: https://digitalcommons.lasalle.edu/analytics_capstones/1/, September 19, 2019.

Fan, Z., Lei, Y., and Wu, S., Research on the changing trend of the carbon footprint of residents' consumption in Beijing*, Environmental Science and Pollution Research*, vol. 26, no. 4, pp. 4078–4090, 2018.

Huang, W., Li, F., Cui, S. H., Li, F., Huang, L., and Lin, J. Y., Carbon Footprint and Carbon Emission Reduction of Urban Buildings: A Case in Xiamen City, China, *Procedia Engineering*, vol. 198, pp. 1007–1017, 2017.

Jin, H., Prediction of direct carbon emissions of Chinese provinces using artificial neural networks, *PLOS ONE*, vol. 16, no. 5, 2021.

Kumari, S., and Singh, SK., Machine learning-based Time Series Models for Effective CO2 Emission prediction in India, *Research Square*, 2022.

Kunda, D., and Phiri, H., An Approach for Predicting CO2 Emissions using Data Mining Techniques, *International Journal of Computer Applications*, vol. 172, no.3, pp. 7–10, 2017.

Lei, W., Alves, L.G.A. and Amaral, L.A.N. Forecasting the evolution of fast-changing transportation networks using machine learning*, Nat Common*, vol. 13, p. 4252, 2022.

Li M., Wang W., De G., Zi X., and Tan Z., Forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on grey prediction theory and extreme learning machine optimized by support vector machine algorithm, *Energies* , vol. 11, no.9, p. 2475, 2018.

Ma, L., and Wang, D., Construction of Game Model between Carbon Emission Minimization and Energy and Resource Economy Maximization Based on Deep Neural Network, *Computational Intelligence and Neuroscience*, pp. 1–12, 2022.

MabAlam, Time-Series Data Mining & Applications, Data Science Central, Available: https://www.datasciencecentral.com/time-series-data-mining-amp-applications, March 16, 2022.

Malmodin, J., and Lundén, D., The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015, *Sustainability*, vol. 10, no. 9, p. 3027, 2018.

Rezaei M.H., Sadeghzadeh M., Nazari M.A., Ahmadi M.H., and Astaraei F.R., Applying GMDH artificial neural network in modeling CO2 emissions in four Nordic countries, *Int J Low-Carbon Technol*, vol. 13, no. 3, pp. 266–271, 2018.

Rugani, B., & Caro, D., Impact of COVID-19 outbreak measures of lockdown on the Italian Carbon Footprint, *Science of The Total Environment*, vol. 737, p. 139806, 2020.

Scherer, M., and Milczarski, P., Machine-Learning-Based Carbon Footprint Management in the Frozen Vegetable Processing Industry, *Energies*, vol. 14, no. 22, p. 7778, 2021.

Seriño, M. N. V., Rising carbon footprint inequality in the Philippines, *Environmental Economics and Policy Studies*, vol. *22, no.* 2, pp. 173–195, 2021.

Sharma, R., Kamble, S.S., Gunasekaran, A., Kumar, V., and Kumar, A., A systematic literature review on machine learning applications for sustainable agriculture supply chain performance, *Computers & Operations Research,* vol. 119, p. 104926, 2020.

Wandana, L. S., Wadanambi, R. T., Preethika, D. D. P., Dassanayake, N. P., Chathumini, K. K. G. L., and Udara, S. P. R., Carbon Footprint Analysis: Promoting Sustainable Development*, Journal of Research Technology and Engineering*, vol. 2, no. 1, 2021.

Wu, W., Yang, H., et al., A Real-Time Recording Model of Key Indicators for Energy Consumption and Carbon Emissions of Sustainable Buildings, *Sensors*, vol. 14, no. 5, pp. 8465–8484, 2014.

Xu, S., Zhang, Y., and Chen, X., Forecasting Carbon Emissions with Dynamic Model Averaging Approach: Time-Varying Evidence from China, *Discrete Dynamics in Nature and Society*, vol. 2020, p. 14, 2020.

Zhang, D., Shen, J., Zhang, F., Li, Y., and Zhang, W., Carbon footprint of grain production in China, *Scientific Reports*, vol. 7, no. 1, 2017.

Biographies

Sweethy Lim is a student from Angeles University Foundation currently taking up a Bachelor of Science in Computer Science, specializing in Data Science. She is a consistent scholar and has been a part of different school organizations. She served as the secretary for the College of Computer Studies Honors Society during her freshman year. She also served as the treasurer for the College of Computer Studies - College Student Council during her sophomore year. She became a student ambassador for Microsoft Learn and is currently at the Beta milestone. Ms. Lim completed a thesis project entitled "Carbon Footprint Prediction in the Philippines Using Time Series in Data Science." Her research interests include machine learning, artificial intelligence, deep learning, and algorithms.

Ma. Mica Ella Cortez is from Samal, Bataan, and she attended college at Angeles University Foundation, taking up courses in Bachelor of Science in Computer Science with a Major in Data Science. The subjects of programming and algorithms have been tackled along the way. She enjoys watching movies, exploring other technologies on the web, and listening to music. She decided to pursue data science because, since she was in high school, she indulged herself in technology and software systems and wanted to learn deeply about the systems as she pursued a master's degree. She had also attended certifications and seminars related to her Azure course and at other companies that provide computer science seminars. Her research interests include architecture, artificial intelligence, database systems, graphics, human interaction, machine learning, natural language processing, programming languages, scientific computing, security, systems, and networking. She had her thesis subject and proposed a prediction model on Carbon Footprint on the field energy production and consumption in the Philippines. The results were alarming, as the prediction implies that the carbon footprint gets higher in the coming years.

Rosanna Esquivel is a Professor of Computer Science and Program Chair for the College of Computer Studies, BS Computer Science at Angeles University Foundation, where she also served as the Internal Quality Auditor (IQA). She was formerly the Program Chair for the College of Computer Studies (BS Information Technology) from 2019–2020 and Assistant Dean for the College of Computer Studies from 2014–2019. She is a part of the Regional Quality Assessment Teams (RQATs) and a member of the Philippine Society of IT Educators—Region III Chapter (PSITE-R3). She was an undergraduate thesis coordinator for the Department of Computer Studies at De La Salle University—Dasmariñas. She obtained her Master's Degree in Computer Science at De La Salle University, Philippines.