공학석사 학위논문

DNN 모델을 활용한 업종에 따른 온실가스 배출량 예측 모델 연구

A Study on the Prediction Model of Greenhouse Gas Emissions in Industries Using Deep Neural Network

2023 년 2 월

서울대학교 대학원

기계공학부

이 종 호

DNN 모델을 활용한 업종에 따른 온실가스 배출량 예측 모델 연구

A Study on the Prediction Model of Greenhouse Gas Emissions in Industries Using Deep Neural Network

지도교수 송 한 호

이 논문을 공학석사 학위논문으로 제출함

2022 년 10 월

서울대학교 대학원

기계공학부

이 종 호

이종호의 공학석사 학위논문을 인준함

2022 년 12월

위 원 장 : 고 승 환 (인)

부위원장 : 송 한 호 (인)

위 원 : 도 형 록 (인)

Abstract

A Study on the Prediction Model of Greenhouse Gas Emissions in Industries Using Deep Neural Network

Chonghoh Lee

Department of Mechanical Engineering

The Graduate School

Seoul National University

Recently, greenhouse gas emission management has begun internationally to respond to the crisis caused by rapid climate change. In the case of Korea, to respond to the climate crisis, 2050 carbon neutrality is declared, and various environmental policies are proposed to ensure that greenhouse gas emissions are being net-zero. It is also necessary to manage greenhouse gas emissions from product exports, as the EU and US Carbon Border Adjustment Mechanism (CBAM) systems are mentioned. Accordingly, companies have begun implementing various systems to manage greenhouse gas emissions. In the case of large corporations, they have begun to manage greenhouse gas emissions from their businesses as well as their suppliers, and suppliers must identify and report greenhouse gas emissions to meet the needs of large corporations. In addition, several small and medium-sized businesses whose main business is exports must also identify and evaluate greenhouse gas emissions in order to export their products. However, unlike large companies that can prepare for this trend, most s mall and medium-sized businesses do not have a system to manage greenhouse gas emissions, so they cannot even grasp data to calculate greenhouse gas emissions, let alone manage greenhouse gas emissions. Therefore, small and

medium-sized businesses are currently responding by identifying greenhouse gas emissions once through environment consulting at a high cost, but some of them often give up cooperative relationships with large companies or product exports because they need a burdensome level of cost and time to cover each time. To solve the problem of identifying greenhouse gas emissions of small and medium-sized businesses, this study developed a machine learning DNN model that predicts greenhouse gas emissions using corporate information, electricity usage, and gas usage data that are already secured by small and medium-sized business es. In the case of the greenhouse gas emission prediction model developed in this study, the machine learning regression model performance index, r2 score, was evaluated to be about 0.927, and it is believed that many small and medium-sized businesses can use it to understand the status of greenhouse gas emissions and respond to regulations.

Keyword : Greenhouse gas emissions · Deep neural network · Emission prediction model · Small and medium-sized businesses · Machine learning

Student Number : 2020-22808

Table of Contents

List of Figures

List of Tables

1. Introduction

Recently, various problems caused by rapid climate change have occurred, and international movements to solve the climate crisis are actively emerging. The main causes of climate change are greenhouse gas emissions, and examples of greenhouse gases are CO2, CH4, N2O, PFCs, HFCs, and SF6. The types of greenhouse gases managed by industry are different, and the total greenhouse gas emissions by industry are calculated by converting the greenhouse effect of each gas based on the global warming potential (GWP) of each gas, and various environmental policies are based on these calculated emission values.

The United Nations Framework Convention on Climate Change (UNFCCC), adopted in 1992, has launched an international effort to combat climate change. The agreement was an international agreement that was not legally binding, but it required advanced countries to restrict various greenhouse gas emissions, including carbon dioxide, in order to reduce global warming caused by greenhouse gases. Accordingly, the Kyoto Protocol, which came into effect in 2005, aims to reduce six types of greenhouse gas emissions in developed countries by 5.2% from 1900 levels from 2008 to 2012. However, as major advanced countries announced their withdrawal, only countries that account for 15% of total greenhouse gas emissions participated, and the effectiveness of the agreement faded somewhat. Since then, the Paris Agreement, called a historic turning point in responding to international climate issues, has been adopted in 2015. This agreement is an international agreement that aims to keep the global average temperature increase significantly below the preindustrial 2°C, and to limit the temperature increase to as close to 1.5°C as possible. Countries should set their own targets for reducing greenhouse gases, commit to the international community, and jointly verify their implementation. It became the first climate agreement to take effect as a comprehensive binding international law. Since then, the U.S. has officially withdrawn from the 2020 agreement, but more than 200 countries, which account for 87% of global carbon emissions, are still implementing the agreement.

As the Paris Agreement was adopted and applied as international law, countries and institutions established and strengthened policies on greenhouse gas emissions which are the main cause of climate change. In case of EU, the European Green

Deal was announced at the European Commission (EC) to achieve the net-zero emissions by 2050 target. Based on the climate law announced later, the Green Deal was legislated, and in addition, a legislative package called 55-year Conformity, which will reduce greenhouse gas emissions by 55% compared to 1990 levels by 2030. The package includes banning the use of internal combustion engines from 2035 onwards, presenting goals to expand alternative fuel infrastructure to promote the development, production and use of eco-friendly cars, and implementing a carbon tax based on greenhouse gas emissions on imports in the region starting in 2026.

In addition, the U.S., one of the major greenhouse gas emitters, officially withdrew from the Paris Agreement, but accelerated regulations on greenhouse gas emissions by announcing the National Greenhouse Gas Reduction Goal (NDC) by 2030 and the U.S. long-term strategy to achieve net-zero emissions by 2050. The five major changes to achieve net-zero include decarbonization of power, conversion of end-use units to electrification, and the adoption of other clean fuels, reduced energy consumption through technology development, reduced methane and other non-carbon greenhouse gas emissions, and increased CO2 absorption.

In the case of Korea, it is also actively managing greenhouse gas emissions by participating in these international movements. It agreed to ratify the Kyoto Protocol in 2002, starting with its membership in the United Nations Framework Convention on Climate Change (UNFCCC) in 1993. In addition, starting with the voluntary goal of 30% reduction compared to the 2020 Business As Usual (BAU) in 2009, the Framework Act on Low Carbon Green Growth was enacted in 2011 to lay the legal foundation for implementing the goal. Since then, it has continued to respond to climate change by managing companies' greenhouse gas emissions and implementing regulations and systems to reduce greenhouse gas emissions in 2012, establishing greenhouse gas reduction roadmaps in 2014, and implementing emission trading systems in 2015. Following the Paris agreement, it submitted and officially registered a national contribution plan (INDC) in June 2015, including a target of 37% reduction in greenhouse gas emissions compared to the 2030 Greenhouse Gas Emission Outlook (BAU) and actively participated in international climate change by submitting a 40% reduction from 2018 levels at

COP26 held in the UK in October 2021.

With the implementation of these environmental policies and regulations, various companies have also begun to present greenhouse gas reduction goals and roadmaps to respond to climate change regulations. Starting with the RE100 declaration that electric energy used in all global businesses will be replaced by renewable energy, global conglomerates such as Apple and Microsoft are proposing roadmaps for achieving carbon neutrality by 2050, which includes reducing greenhouse gas emissions from the value chain not only within the workplace but also across the entire business. In addition, the ESG report, which is published annually, discloses the current greenhouse gas emission level without adding or subtracting details on the roadmap implementation plan. Samsung Electronics, Hyundai Motor, and other major domestic companies are actively establishing roadmaps related to carbon neutrality, and through ESG reports, they disclose greenhouse gas emissions and roadmaps based on various carbon accounting standards such as TCFD, SASB, and CDP. In addition, large companies have begun to manage the greenhouse gas emissions of their partners, as indirect emissions from the supply chain account for 80% of a company's total emissions according to CDP. As more and more large companies take strong measures to exclude companies that do not manage greenhouse gas emissions from the supply chain, greenhouse gas emissions management is essential for partners, mostly small and medium-sized companies.

In addition, the EU and the United States have recently announced plans to implement a carbon border tax, a bill that imposes trade tariffs on exports from countries and companies that emit a lot of carbon dioxide beyond simple domestic regulations. In the case of the bill, not only large companies but also small and medium-sized companies that export carbon-intensive products will be applied. Now, not only large companies but also small and medium-sized companies that export carbon-intensive products must measure and manage greenhouse gas emissions.

However, most small and medium-sized companies do not have a greenhouse gas emission management system, do not manage data for calculation, and lack knowledge on calculating greenhouse gas emissions. As a result, many small and medium-sized companies rely on environmental consulting to measure their

emissions, but this is burdensome and difficult to respond actively due to high costs.

Many previous studies of greenhouse gas emissions prediction have focused on country or industrial total emission projections for setting business-as-usual scenarios or just setting reduction plans for reducing emissions. Furthermore, most recent research has focused on the case of European countries and the United States, where environmental regulations are strong.

Our goal for this study is to develop a prediction model that can predict greenhouse gas emissions of company by using only the power usage, gas usage, and basic information of company data managed by small and medium-sized companies. Now, unlike previous studies, it is important to predict greenhouse gas emissions of company because the new environmental regulations are focused on greenhouse gas emissions of company, not greenhouse gas emissions of country.

2. Methodology

2.1 Purpose of Machine Learning

Machine learning is a new method that learns data to find patterns and rules for each case, so that even if a new case occurs, it can respond without writing a separate code based on the previously learned pattern and rules. Because of these characteristics, it has the advantage of easy response to new variables and easy maintenance, also is known as a method suitable for application to complex problems. It also has the advantage of steadily improving performance through evaluation.

Machine learning is largely useful when performing two tasks. It is a classification that determines which group the new data is classified by learning several data, and a regression that predicts what numbers will come out when a new case is entered based on learning the data. This study used machine learning for regression because it aims to predict how much greenhouse gas emissions will be generated by small and medium-sized businesses based on their corporate information and electricity usage data and gas usage data.

2.2 Artificial Neural Network Model

The artificial neural network (ANN) model is a machine learning model modeled after the human neuron structure in software and is a model of artificial intelligence technology. Perceptron, the most basic of artificial neural networks, is a mathematical model of neurons, and when any input enters, weights are applied to it and transmitted to other neurons. The one that connects perceptron is called a layer, and the layers constituting the artificial neural network are divided into input layer, hidden layer, and output layer according to their characteristics. In addition, the artificial neural network model with multiple hidden layers is called Deep Neural Network (DNN), and the algorithm designed to learn is called Deep Learning. In the case of Deep Learning, unlike existing rule-based algorithms that require prior knowledge for product automation and observation and experiment to verify completed rules, automated factor extraction ability and model character-

istics that quickly find optimal rules make it easy to customize.

In this study, deep neural network (DNN) techniques are utilized, and these techniques are useful for modeling complex nonlinear relationships, just like existing artificial neural networks. In principle, the greenhouse gas emissions to be predicted in this study should be calculated mathematically based on greenhouse gas emissions caused by all energy sources usage. However, most small and medium-sized businesses do not know what energy usage data affects these greenhouse gas emissions and do not have the data themselves. It is hard to small and medium-sized businesses to start collecting and managing data that are not collected before, so this study aims to estimate greenhouse gas emissions using electricity usage, gas usage, regions, industries, and number of workers to solve this problem. In this study, DNN techniques are used to solve very complex nonlinear predictions caused by omission of many of the existing data needed for calculation and variable main cause of greenhouse gas emissions depending on the industry.

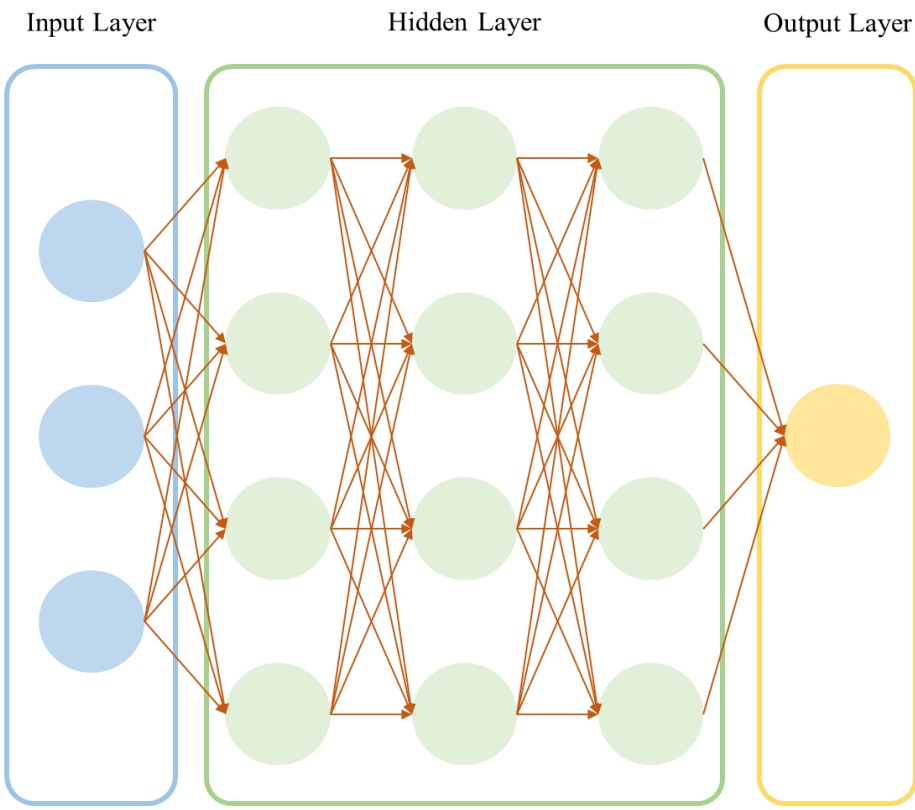

Figure 2.1 DNN model diagram

3. Data Preprocessing

Data preprocessing refers to a step of preprocessing data to be used for learning in machine learning model in a good form. Most of the data that is secured is organized according to the source of data and the method of collection, and it is essential to change it into a form that can be learned by computers. Since the outcomes of machine learning model vary depending on the way the data is preprocessed, it is said that data preprocessing is important to complete a successful model, and it takes time about 80 to 90 percent to preprocess the data rather than learning the model. No matter how nice a model is, if it learns with strange data, it outputs strange results, which means that a well-structured model must be learned with high-quality data to function properly. There are many kinds of data preprocessing such as a data filtering process, a data transformation process, and a data scaling process.

3.1 Data Analysis

The data used in this study are energy usage data and greenhouse gas emissions data for each company in various manufacturing sectors provided by the Korea Energy Agency (KEA) as an open API. The dataset consists of both numerical type and string type of data such as the industrial classification of company, the number of workers, energy name, region, industry code, year, greenhouse gas emissions, and energy sources usage. This dataset consists of about 900,000 sets of data, consisting of data from about 100,000 companies every year from 2010 to 2018. Considering that there were about 400,000 businesses in the manufacturing sector in Korea in 2018, this is the data of companies equivalent to about 25% of the total manufacturing industry in Korea.

In addition, about 10% of whole companies use only electricity and gas, while the remaining 90% have greenhouse gas emissions from the use of other energy sources as well as electricity and gas.

Table 3.1 Example of raw data of KEA open API

Company Code	Number of Workers	Industry Code	Region	Energy Source	Industry	Energy Usage	Year	Greenhouse Gas Emissions
FFFFB485A 8496D5516	5 - 9	18119	Seoul	Electricity	Other printing	1.4441	2010	7.5981
FVC3689D W453D5641	20 - 49	10742	Incheon	LNG	Mixed seasonings	43.5890	2011	22.9252
E16489X55 6D23F16S2	10 - 19	26429	Gyeonggi	LNG	Communic-ation equipment	0.1055	2013	0.2231
C4584V564 D16D12FA	5 - 9	26511	Chungnam	Electricity	Television manufactur-ing	0.3486	2012	1.8340
V15416C84 9D23135W	10 - 19	29223	Gwangju	Propane	Metal forming machine	72.4875	2017	173.7798
EFD549DV X5489D2V	Less than 5	29223	Gwangju	Diesel	Metal forming machine	1.2550	2018	3.5963
BDNB4658 B12AC5D6	Less than 5	14411	Gyeonggi	Electricity	Stockings and other socks	1.3756	2016	7.2378

3.2 Data Merging

The raw data of the KEA Open API includes data on energy usage and greenhouse gas emissions for one energy source per row. In other words, companies that use multiple energy sources are distributed and represented in multiple rows, and it is necessary to merge these data into one data per company. Data from the same company could be collected through the company's unique code, and the data was merged into one data containing total emissions and usage data for each energy source. There are a total of 30 types of fuel, and if there are multiple workplaces in the same company, the data was integrated by dividing it by workplace.

3.3 Data Filtering

Data filtering was conducted as a step of processing the original dataset provided by open API system into data that can be used for learning. First, the data of companies expressed in percentiles were removed due to the high amount of greenhouse gas emissions and energy use because these data are not disclosed the actual energy usage and greenhouse gas emissions, and the purpose of this study is to create a model for predicting greenhouse gas emissions for small and medium-sized businesses with poor data collection. Second, other data except for the industry code, number of workers, region, electricity usage, gas usage, and greenhouse gas emissions were removed which are data to be used to learn machine learning model. Third, if there is data that is not written to that data, it is excluded because it means null, not zero. Last, we identified outliers that could interfere with learning through the box plot function and removed data with greenhouse gas emissions of 800 tCO2e/yr which are the top 10% of greenhouse gas emissions per year or more and data from the bottom 1% that were inaccurate due to rounding at the minimum unit. After the data filtering process, 819,561 sets of data were left, and these data were used for learning DNN model.

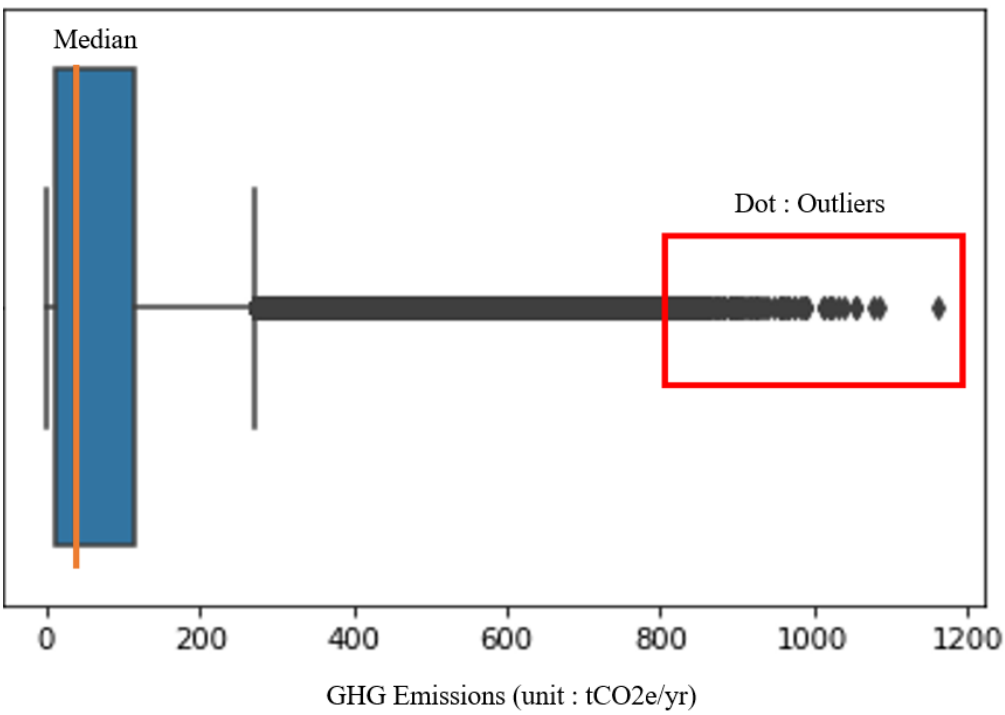

Figure 3.1 Boxplot of greenhouse gas emissions

Table 3.2 Example of filtered data of KEA open API

Number of Workers	Industry Code	Region	Electricity Usage	LNG Usage	Greenhouse Gas Emissions
5 - 9	18119	Seoul	1.4441	0.0000	7.5981
20 - 49	10742	Incheon	105.6284	43.5890	766.9155
10 - 19	26429	Gyeonggi	3.5962	0.1055	15.8441
5 - 9	26511	Chungnam	0.3486	1.7865	5.9683
10 - 19	29223	Gwangju	5.9280	0.7841	216.8445
Less than 5	29223	Gwangju	7.4442	0.2699	24.9663
Less than 5	14411	Gyeonggi	1.3756	3.5961	11.4721

3.4 Data Transformation

Data transformation was a step of transforming the data that transformed the string data of the dataset that completed data filtering into numbers. First, in the case of the 'number of workers', ordinal encoding method was used that can consider the order with ordered data with a large number of people and a small number of people depending on the category. In the case of "number of workers" variable is big, it is important to consider the order as variable that can reflect the characteristics that energy use and greenhouse gas emissions may be higher in most industries. Next, in the case of 'region' and 'industrial classification', encoding was applied in a way that gives unique numbers to data without a separate order such as priority. Label encoding is used in this case among one-hot encoding, target encoding, and label encoding. The reason that does not use one-hot encoding is this encoding technique that gives unique vectors to each category, and if there are many categories, the input of dataset itself becomes too large, so it is not suitable to apply the encoding technique to more than 10 "regions" and 40 "industry." Next, in the case of target encoding, feature meaning can be given to the characteristics in a way that the target value is reflected in the characteristics, but it was not used in this study due to the problem that data overfitting occurs frequently. In this study, the learning performance may not be good because label encoding has no significant characteristics compared to other encoding techniques, but it is utilized because several problems that can easily occur on such a large dataset do not occur.

Next, we checked the distribution of each data. For effective machine learning, it is recommended that data exist in a form close to a normal distribution. This is because the unilaterally biased data distribution has the effect of destroying the model by causing the machine learning model to overlearn the specific data distribution. To solve these problems, we use methods such as log transformation, square root transformation, and box-to-cox transformation to transform numbers, or sometimes copy the insufficient number of existing data, and shape the distribution of the data into a normal distribution. In this study, electrical data typically existed in a biased distribution, and distorted data conversion was performed using log transformation.

To understand the relationship between the input variable and the greenhouse gas

emission, after data transformation, the trend was confirmed using the pair plot function. As a result, the trend was relatively clear in the case of electricity and gas use, which is directly related to greenhouse gas emissions, and it was confirmed that it was difficult to grasp the trend by comparing the remaining variables one-on-one.

Table 3.3 Example of transformed data of KEA open API

Number of Workers	Industry Code	Region	Electricity Usage	LNG Usage	Greenhouse Gas Emissions
1	2	1	1.4441	0.0000	7.5981
3	1	3	105.6284	43.5890	766.9155
2	19	2	3.5962	0.1055	15.8441
1	17	7	0.3486	1.7865	5.9683
2	11	4	5.9280	0.7841	216.8445
0	11	4	7.4442	0.2699	24.9663
0	8	2	1.3756	3.5961	11.4721

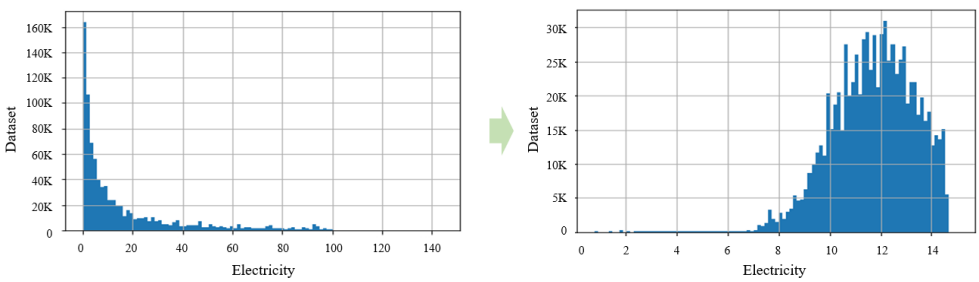

Figure 1.2 Skewed data transformation
(Electricity usage, Log transformation)

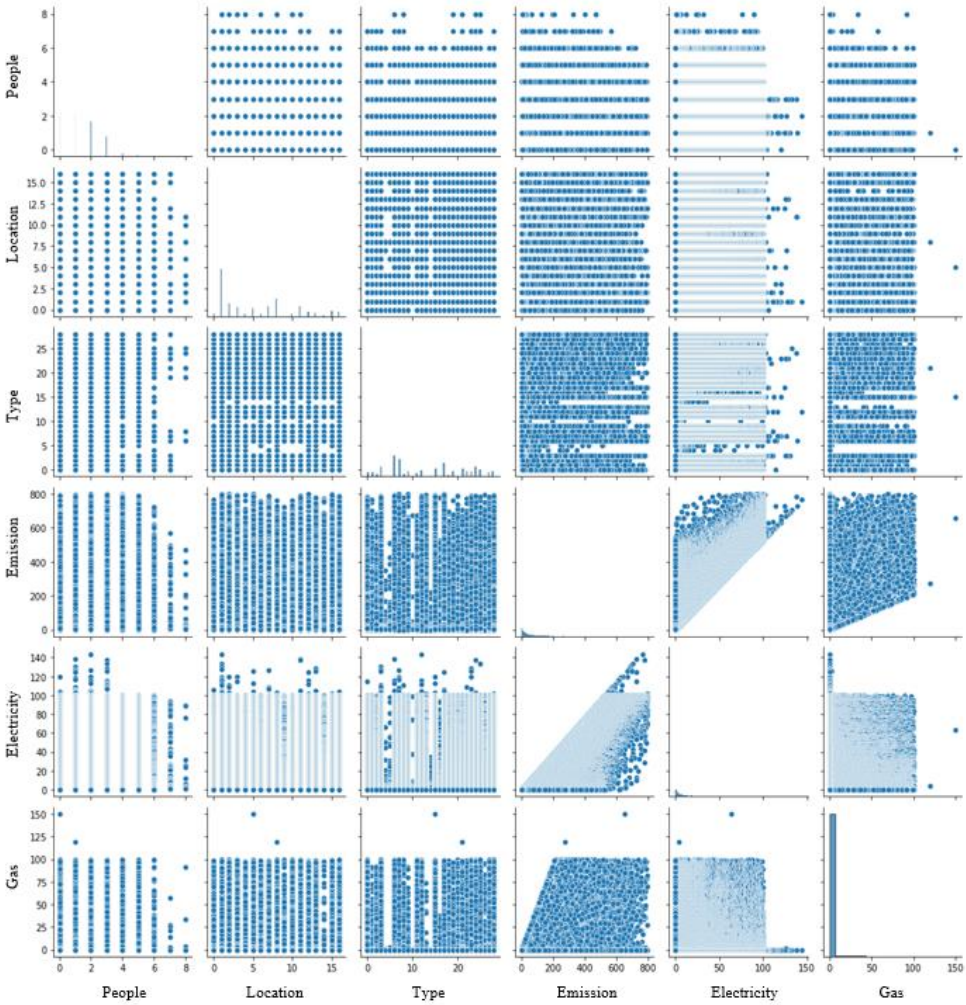

Figure 3.3 Pairplot of features

3.5 Data Scaling & Dataset Classification

As the data transformation was completed, all features were converted into numerical types and data scaling was performed because the scale was different depending on the feature. The Scaler for data scaling utilized the StandardScaler function, which ensures that data follows a standard normal distribution. In the case of the StandardScaler function, all features can be made on the same scale, so it is useful to prevent a phenomenon in which only specific features are weighted.

After completing data scaling, the correlation between the scaled-completed variables and greenhouse gas emissions was checked. As a result, it was confirmed that there was a relationship between the number of workers and greenhouse gas emissions not only electricity usage and gas usage.

Next, the scaled datasets were classified into train datasets for machine learning training, validation datasets for post-learning performance verification, and finally, test datasets for model performance tests. The random classification method was used for dataset classification, and 64% of the data was designated as the train dataset, 16% of the data was designated as the validation dataset, and 20% of the data was designated as the test dataset.

	Emission	Electricity	Gas	People	Location	Type
Emission	1.00	0.72	0.19	0.46	-0.03	-0.08
Electricity	0.72	1.00	0.06	0.44	-0.06	-0.12
Gas	0.19	0.06	1.00	0.13	-0.02	0.02
People	0.46	0.44	0.13	1.00	-0.05	0.00
Location	-0.03	-0.06	-0.02	-0.05	1.00	0.05
Type	-0.08	-0.12	0.02	0.00	0.05	1.00

Figure 3.4 Correlation coefficient of features

4. Deep Neural Network Model Design

The design of a machine learning model is crucial as it impacts the model's ability to learn from data and make accurate predictions. A well-designed model should be able to handle the complexity of the problem and effectively identify patterns in the data. The performance of a machine learning model is determined by two main factors, parameters learned from data and self-regulation of the learning process established before training. Hyperparameters, such as learning rates, loss functions, and batch sizes, fall under the latter category and need to be carefully selected before training. This process is known as hyperparameter tuning. The model used in this study had an input layer, three hidden layers with 64 nodes, and an output layer. Electricity usage, gas usage, number of workers, location, and industrial type were used as input variables and were designed to produce greenhouse gas emissions as output. The commonly used ReLU function was used as the activation function between nodes. The loss function chosen was 'mean squared error', mainly used in Keras, and an early stop function was applied to prevent overfitting of the training dataset. This method stores the model with the minimum validation loss and stops learning when validation loss is not improved for 10 consecutive times. The performance of the model was finally verified using the r2 score index, which evaluates the performance of a regression analysis model.

Table 4.1 Model design

Features	**Elements**
Input layer	5 inputs - Number of workers - Location - Industrial Code - Electricity Usage - Gas Usage
Hidden layer	3 layers (64 nodes each)
Output layer	1 output - Greenhouse gas emissions
Activation function	ReLU function
Loss function	Mean squared error
Additional function	Early stop
Performance Indicators	R2 score

5. Result

As a result of training using the model designed in this study, it was confirmed that learning was stopped at the 84th epoch. This was due to the interruption of learning caused by the previously set early stop function. Additionally, the change in the average absolute error and the average square error until the 84th epoch was also confirmed. Firstly, in the case of the average absolute error, the overall trend tends to decrease only slightly, making it difficult to confirm that the model itself has greatly improved. However, in the case of the mean square error, which is the verification loss, it was confirmed that the model was learned normally. It can be inferred that these two indicators indicate that the error amplitude of the predicted value has gradually decreased based on the specific error value.

Next, the r2 score, which is an indicator of verifying the performance of the model, was checked. The closer the result range is to 1 from 0 to 1, the better the prediction performance of the model. For the model used in this study, we scored 0.927 points based on the test dataset, 0.926 points based on the validation dataset, and 0.927 points based on the train dataset, which is the data used for direct learning. In general, considering that a level of 0.7 means a significant correlation or prediction model in engineering analysis, regression models with r2 scores above 0.9 are interpreted as very meaningful prediction models, indicating that they have developed sufficiently significant prediction models.

Finally, the relationship between the actual and predicted values was analyzed. A dot plot was used for the results analysis, and it was confirmed that the correlation was not clear despite the high r2 score, indicating excellent prediction performance. However, this data often overlapped with multiple data points when using a large number of data points. It was important to check how much data overlapped at a single point location. To confirm this, the alpha value was applied to the dot plot, and as a result, a clear linear correlation was confirmed. In addition to the linear correlation, unusual trends were identified in the dot plot. For most of the data that failed to make accurate predictions, the predicted value was found to be smaller than the actual emission. It was determined that this was caused by predicting small emissions for companies with zero or very small electricity usage, which showed the highest correlation between emissions and features previously analyzed. This

was further confirmed by extracting specific data with zero electricity usage from the raw data and predicting it through the model.

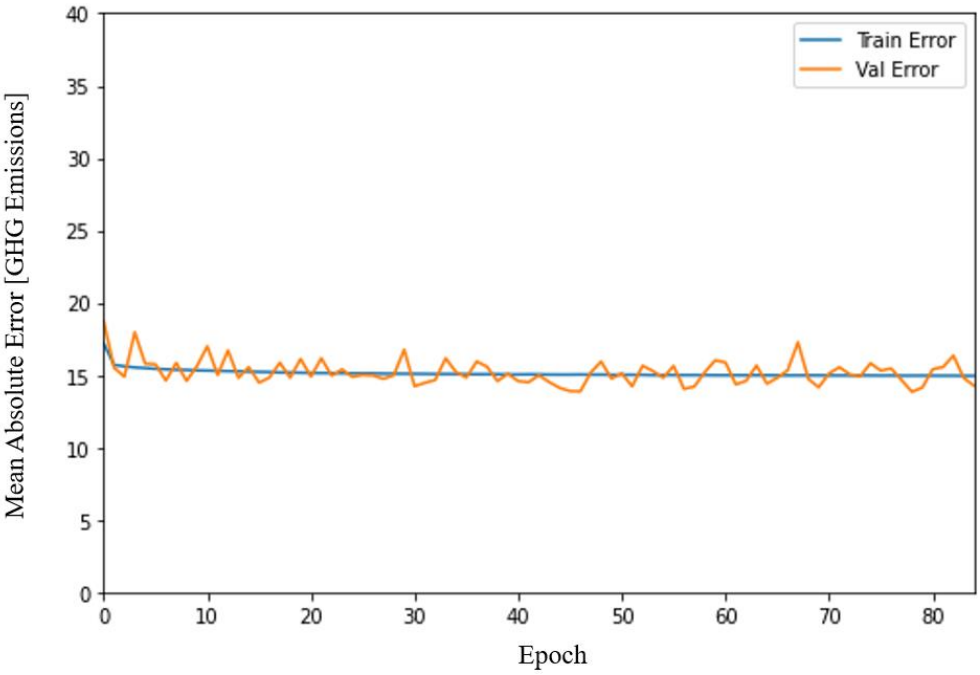

Figure 5.1 Mean absolute error of each epoch

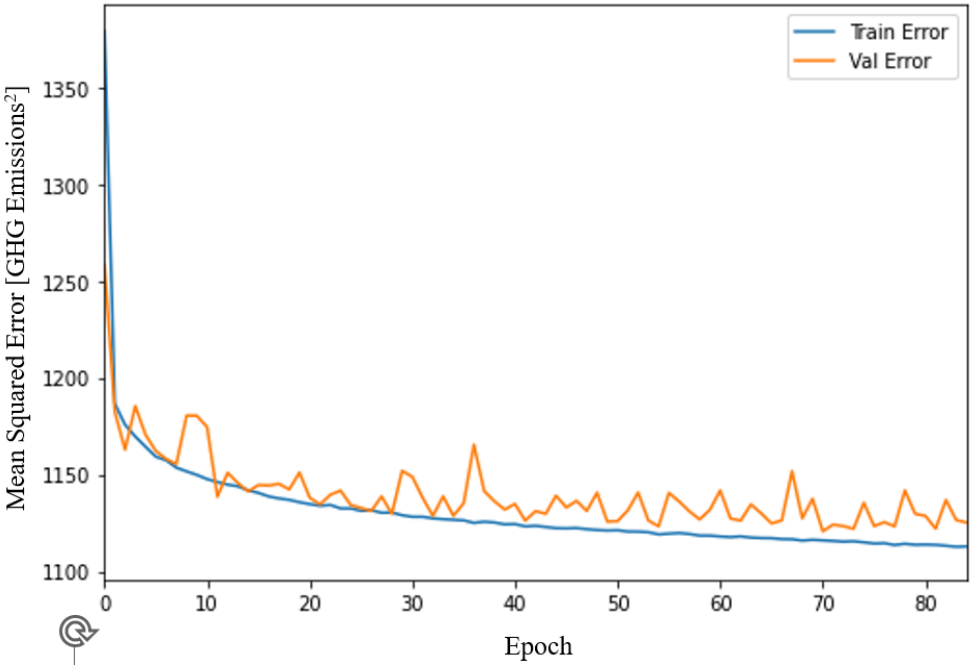

Figure 5.2 Mean squared error of each epoch

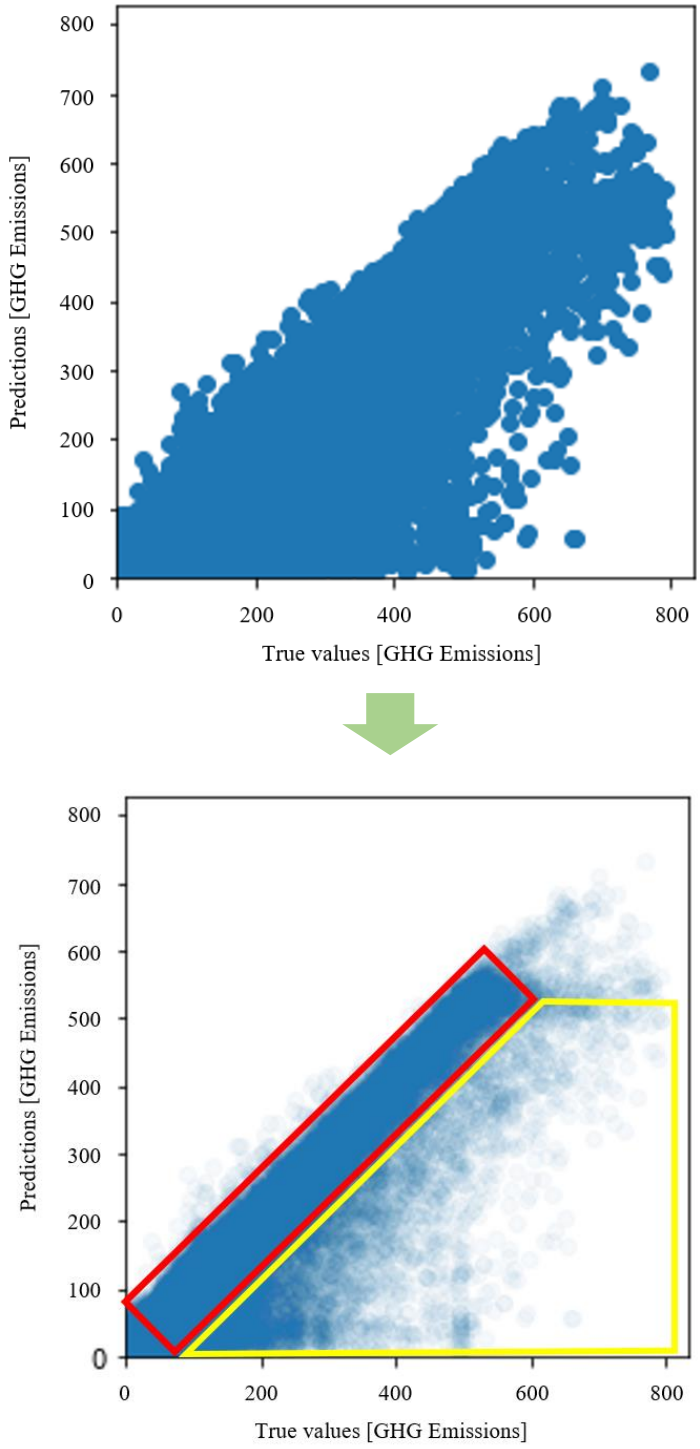

Figure 5.3 Dotplot of Predictions & True values

6. Conclusion

In this study, a greenhouse gas emission prediction model with a high prediction level (r2 score of 0.926) was developed using a machine learning DNN model. For this prediction model, most companies can predict the status of greenhouse gas emissions using only readily available data, such as corporate information, power usage, and gas usage. However, data from companies that do not use electricity tends to predict lower emissions compared to actual emissions, so it is necessary to further develop a greenhouse gas emission prediction model for these companies in the future.

In order to achieve carbon neutrality in countries that are striving for it internationally, it is a top priority to understand the current status of greenhouse gas emissions. Governments conduct total energy surveys every few years to secure the status of greenhouse gas emissions at the national level and calculate emissions by industry, which requires a lot of preparation, effort, and time. This study has succeeded in developing a DNN model that predicts greenhouse gas emissions of small and medium-sized businesses, which can help most small and medium-sized businesses cope with these national challenges more easily and quickly by making it easier to understand and manage their emissions.

This model can also be useful for business purposes. Large companies are preparing to introduce internal systems to manage greenhouse gas emissions from partner companies but have not been able to collect data due to data management problems. Applying greenhouse gas emission prediction model algorithms to internal systems of these large companies can help them secure emissions data from numerous partners with poor data management. Additionally, this will help small and medium-sized businesses maintain cooperative relationships with large companies, as greenhouse gas emissions data must be reported to them. Furthermore, small and medium-sized companies can easily and quickly identify their greenhouse gas emissions instead of consulting, which is very expensive. In addition, this prediction model can help small and medium-sized businesses that need to respond to carbon border taxes easily and quickly grasp the current status of their greenhouse gas emissions.

References

[1] Korean Statistical Information Service. 2018~2020 Greenhouse Gas Emissions by Manufacturing Sector. 2022. p. 2018.

[2] Data Portal. Korea Energy Agency - Statistics on Energy Use and Greenhouse Gas Emissions - Microdata.

[3] Kedar Potdar, Taher S. Pardawala, Chinmay D. Pa. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications 2017;175(4):7-9.

[4] Paul Covington, Jay Adams, Emre Sargin. Deep Neural Networks for YouTube Recommendations 2016.

[5] Leonardo Nascimento, Takeshi Kuramochi, Niklas Hohne. The G20 emission projections to 2030 improved since the Paris Agreement, but only slightly. Mitigation and Adaptation Strategies for Global Change 2022; 27:39.

Abstract in Korean

최근 급격한 기후 변화에 따라 이를 억제하기 위해 국제적으로 온실가스 배출량을 관리하기 시작하였다. 우리나라의 경우에도 2050 탄소중립을 선언하며 온실가스 배출량이 net-zero가 되도록 하는 여러 환경 정책들이 제시되고 있다. 또한, 탄소국경세 제도가 언급됨에 따라 제품 수출에 있어서도 온실가스 배출량의 관리가 필요하게 되었다. 그에 따라 기업들은 온실가스 배출량 관리를 위해 여러 제도를 시행하기 시작하였으며 대기업의 경우 본인들 사업장 온실가스 배출량을 넘어 협력사들의 온실가스 배출량 관리를 시작하였고, 협력사들은 대기업의 요구에 맞춰 온실가스 배출량을 파악 및 보고해야 하는 상황이다. 또한, 수출이 주 사업인 여러 중소/중견 기업들 역시 제품 수출을 위해 본인들의 온실가스 배출량을 파악해야 하는 상황이 되었다. 허나 이러한 흐름에 대해 비교적 미리 준비가 가능한 대기업과 달리 대부분의 중소/중견 기업들은 온실가스 배출량 관리를 위한 시스템이 정립되어 있지 않으며 관리는 물론 온실가스 배출량 현황 파악을 위한 데이터 자체가 관리되고 있지 않은 것이 실상이다. 그래서 현재 이러한 경우 큰 비용을 들여 컨설팅 통해 일회성으로 온실가스 배출량 파악을 하며 대응을 하고 있으나 대부분 중소/중견기업이 매번 감당하기에는 부담이 큰 수준의 비용과 시간이 필요하여 대기업과의 협력관계 혹은 제품 수출 등을 포기하는 경우도 많이 생기고 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 중소/중견기업들이 이미 확보하고 있는 데이터인 기업정보와 전기 사용량, 가스 사용량 데이터를 활용하여 그들의 온실가스 배출량을 예측하는 머신 러닝 DNN 모델을 개발하였다. 본 논문에서 개발한 온실가스 배출량 예측 모델의 경우 머신 러닝 회귀모델 성능 지표인 r2 score가 0.927 수준으로 굉장히 유의미한 예측을 할 수 있는 모델로 평가되었고 여러 중소/중견기업이 온실가스 배출량 관련 현황 파악 및 규제 대응에 유의미하게 활용할 수 있을 것으로 판단된다.