Predicting Solar Power Generation from Direction and Tilt Using Machine Learning XGBoost Regression

Yongjun Kim and Yungcheol Byun

Computer Engineering Department, Jeju National University, Jeju, Korea

yjkim0123@gmail.com, ycb@jejunu.ac.kr (Corresponding Author)

Abstract. Electricity production using photovoltaic panels is widely used all over the world. Electricity generation using solar power is helping a lot in reducing the Earth's carbon production. However, many photovoltaic panels are often installed improperly to produce maximum efficiency. This not only reduces the amount of electricity produced, but also wastes resources required to make solar panels. To solve this problem, the electricity production data using the longitude and latitude of the actual solar panel installation location in 5 regions in Korea, and the direction and tilt angle of the installation site are used to predict the electricity production by using the machine learning XGBoost regression model. A study was conducted to determine the direction and tilt angle of the installation site to generate maximum electricity production in longitude and latitude. Through this, the longitude and latitude of the solar panel installation location for maximum electricity production were analysed, and in the future, more local data will be used to get out of the limited area, and specific research efficiency will be brought.

1. Introduction

Recently, there has been a lot of interest in energy and research on next-generation energy around the world. Among them, energy research using sunlight is a representative example.

Solar power generation is a power generation method that directly converts the light energy of the sun into electrical energy. Photovoltaic power generation uses the light energy of the sun and directly converts it into electrical energy using a photoelectric converter called a solar cell. Because it partially uses light, it can be used even on cloudy days, and the efficiency of using solar energy is higher than that of thermal power generation [1].

Therefore, energy is produced by the installation of many solar power panels throughout the country in Korea, and the Korean government supports this to produce a large portion of electric energy. The energy produced in this way is sometimes consumed at home, and the surplus energy is sold back to KEPCO(Korea Electric Power Corporation).

In particular, the reason for producing electric energy using sunlight is closely related to the global environmental problem, and it is attracting attention worldwide as an eco-friendly energy because it has less air pollution than thermal power generation.

Carbon dioxide, one of the trace gases present in the earth's atmosphere, is one of the greenhouse gases that absorbs radiant energy in the infrared region emitted from the earth's surface and plays an important role in the energy balance of the earth system. become a supplier. The concentration of carbon dioxide in the atmosphere may change naturally according to long-term fluctuations in climate, or it may increase artificially due to increased use of fossil fuels, causing rapid changes in climate [2].

Energy production using solar energy can vary depending on how the solar panel is installed.

It is necessary to vary the installation angle of the panel according to latitude or longitude, and there is a large difference in energy production according to this.

Latitude is a location that indicates how far north or south you are from the equator. The equator becomes 0`, and if it goes up to the north, it is called north latitude and is indicated by N (North), and if it goes down to the south, it is called south latitude and is indicated by S (South). Longitude is a line that marks east and west vertically based on the old Greenwich Observatory in England. In Greenwich Observatory, the right side is called East Longitude and is marked with E (East), and the left side is called West Longitude and is marked with W (West). The 180’ point where the east and west longitudes meet is the line where the date changes [3].

In this paper, the part to be carried out in this study is based on five domestic sites (Jeongseon, 2 place of Gangneung, Sangju, and Yeongwol) with various features (latitude, longitude, azimuth, inclination, date, temperature, air volume, humidity, and sunlight [4]) were used to predict the energy production for a specified date.

A related study in Predicting Solar Generation from Weather Forecasts Using Machine Learning [5], which has a method to improve prediction using SVM-based prediction models, and Solar power generation forecasting using ensemble approach based on deep learning and statistical methods in [6], the ensemble method of LSTM and GRU was used.

2. Related Research

2.1 Solar power generation

It refers to a power generation method that converts light from the sun into electrical energy using the photovoltaic effect. It is distinguished from solar thermal power generation, which generates electricity using the thermal energy of light, in that it directly converts light energy into electrical energy.

The photovoltaic effect is almost like the photoelectric effect, but the circumstances are slightly different. While the photoelectric effect generally refers to the effect of a material receiving light and emitting electrons, the photovoltaic effect refers to the movement of electrons and holes generated because of the photoelectric effect inside the material to create a potential difference. That is, the photovoltaic effect can be said to be a secondary effect that can be considered because of the photoelectric effect.

As shown in figure 1, a device that converts light energy into electrical energy using the photovoltaic effect is a solar cell. Solar cells are largely composed of n-type semiconductors, p-type semiconductors, and anti-reflection plates and electrodes that help absorb light. Electrons in n-type semiconductors and holes in p-type semiconductors play a major role in transferring charges. When these two semiconductors are attached to each other, electrons diffuse from the n-type semiconductor, which has relatively many electrons, to the p-type semiconductor, and vice versa. Holes spread out into the n-type semiconductor. Therefore, positive holes with a positive charge are concentrated on the n-type semiconductor side, and electrons with a negative charge are concentrated on the p-type semiconductor side, centering on the junction. Accordingly, a built-in field is generated in the junction from the n-type semiconductor to the p-type semiconductor. When the electric field is sufficiently diffused, the strength of the electric field also becomes stronger. Therefore, the region where the electric field is generated is called the depletion region (space charge region).[7]

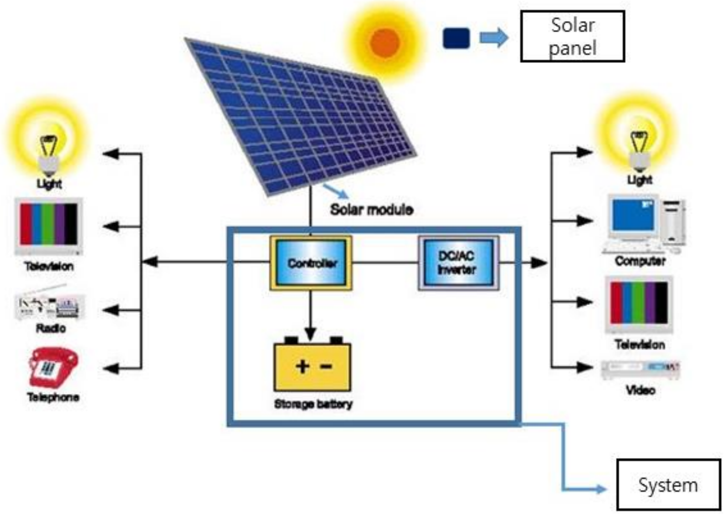

Figure 1. Solar power structure [7].

Electric energy using solar power refers to energy related to electricity. Ultimately, it is energy derived from the potential energy or kinetic energy of the electric charge. From the user's point of view, energy is divided into several forms and can be converted into each other. Among them, electrical energy is the most useful energy used in daily life. In effect, electricity is the energy carried or consumed by an electrical circuit because it moves through the circuit. Electrical energy is supplied by the combination of the electric current delivered by the circuit and the potential difference. It is electrical potential energy until it reaches the user and is converted into another form of energy. When converted from potential energy, electrical energy becomes other forms of energy such as heat, light, and motion [8].

When light with energy greater than the semiconductor's bandgap enters the solar cell, it excites the electrons inside the solar cell to a higher energy state and creates a hole in the place where the electron was, forming an electron-hole pair (electron-hole pair). hole pair) is created. At this time, under the influence of the internal electric field generated by the junction of the n-type semiconductor and the p-type semiconductor, the excited electrons move toward the n-type semiconductor and the holes move toward the p-type semiconductor. Since electrons and holes, each having a negative charge and a positive charge, move in opposite directions, current flows along the direction of movement of the holes in the solar cell. Therefore, when an external circuit is connected to this solar cell, current flows along this circuit, which is the basic principle of solar power generation. The advantage of solar power generation is that there are few maintenance elements and the maintenance cost is low. Because of these strengths, the number of households installing solar power generators for home use is increasing in Korea. A solar power plant capable of producing 10 W/m2 has an energy production efficiency per area 4 times that of wind power and 20 times that of biomass, showing high efficiency in new energy. In terms of solar energy, 100 ~ 250 W/m2, the efficiency of solar power generation is currently showing 10 ~ 20% efficiency, and the efficiency is expected to increase further in the future. However, photovoltaic power generation has a disadvantage in that the output variation according to the weather is large. The direction of sunlight changes with time, and the efficiency decreases depending on the weather, such as cloudy or rainy days. It also includes the issue of efficiency as a function of temperature. As of 2017, photovoltaic panels show an abominable efficiency at 25 degrees Celsius, but their efficiency declines above and below temperatures. In Korea, the reason that solar power generation efficiency is the highest in March-June and September-November usually comes from efficiency according to temperature. In the early days of photovoltaic power generation, CdTe solar cells were used. At this time, cadmium was a toxic substance and hazardous substances were detected in each manufacturing process [9].

2.2 Machine learning
Machine learning refers to the process of analysing given data to find regularities and patterns and extracting meaningful information using them [10]. For example, the task of machine learning is to find a characteristic or rule that can distinguish an apple from a banana from many pictures of apples or bananas, and then use this to determine whether a new picture is an apple or a banana when it comes in. In this case, the process of finding a rule from data is called training or training, and the rule determined by learning is called a model. And the process of passing new data as input to the trained model and judging the result is called prediction or inference [11].

Machine learning is largely divided into supervised learning and unsupervised learning. Supervised learning is a method of learning using data for which the correct answer is known. Unsupervised learning falls into the category of problems of figuring out how data is structured. Unlike supervised learning, this method does not give a target value for the input value.

It is an advanced form of machine learning that learns and predicts itself without inputting data for human learning. These models evolved from artificial neural networks that mimic human neural networks. Deep learning is one of the machine learning techniques, Feature Learning (or Representation Learning). Deep learning is a deep neural network (DNN) developed from artificial neural networks, and it is called hierarchical feature learning (or representation learning) because it is a model in which multiple neurons transmit signals to the next one [12].

3. Analytical and Research Environment
To derive the results of this study, the analysis and research environment was conducted with a method of inference by applying the existing power generation characteristic information obtained using XGBoost, a method of machine learning.

XGBoost was initially started as a research project by Tianqi Chen as part of the Distributed (Deep) Machine Learning Community (DMLC) group. Tree boosting is a highly effective and widely used machine learning method. In this paper, we describe a scalable end to-end tree boosting system called XGBoost, which is used widely by data scientists to achieve state-of-the-art results on many machine learning challenges[13]. After a while, Python and R packages were built, and XGBoost now has package implementations for Java, Scala, Julia, Perl, and other languages. This brought the library to more developers and contributed to its popularity in the Kaggle community, where it was used in numerous competitions [14][15].

And the OS environment of the study was made on Windows 10, and Anaconda Jupyter Notebook Tool was used based on Python 3.8.5 version.

4. Methodology and Result
As a general principle of installation of solar panels, the azimuth should face south in a place not affected by shadows, and it should be installed within the range of 45 degrees from east to west based on the site conditions depending on the site conditions. shall. However, the installation is slightly different depending on the surrounding conditions and the amount of sunlight. The amount of electricity produced varies depending on the installation direction, and such installation method and direction are very important factors for efficient electricity production in each environment.

The following is the content of the article that presents the problems caused by errors in the installation direction of the photovoltaic panel.

The Daily Mail reported on the 20th that most of the world's solar panels (panels) were installed in the wrong direction. Scientists at the Pecan Street Research Institute in Austin, Texas, in a report published in Quartz, found that solar panels installed in the west get more energy. So far, in the northern hemisphere, solar panels have been installed south to provide the best exposure to the sun during daylight hours. When installing solar panels in a private house, facing south has been considered essential. The report pointed out that west-facing solar panels produced 2% more electricity per day than south-facing solar panels. Households with west-facing solar panels produced more electricity in the afternoon at sunset than those with south-facing solar panels, reducing their dependence on external electricity by 65% during the afternoon, when electricity bills are the most expensive. South facing is 54%.

In the UK, the average solar power generation capacity for households is 3.5 to 4 kWp based on the maximum power, and the installation cost ranges from £5,500 (about 9.37 million won) to £9,500 (about 16.18 million won). A 4kWp solar panel produces 3700 kwh of electricity per year, which is equivalent to the electricity demand of a household. The UK government subsidizes households with solar panels in the amount of £78 a year [16].

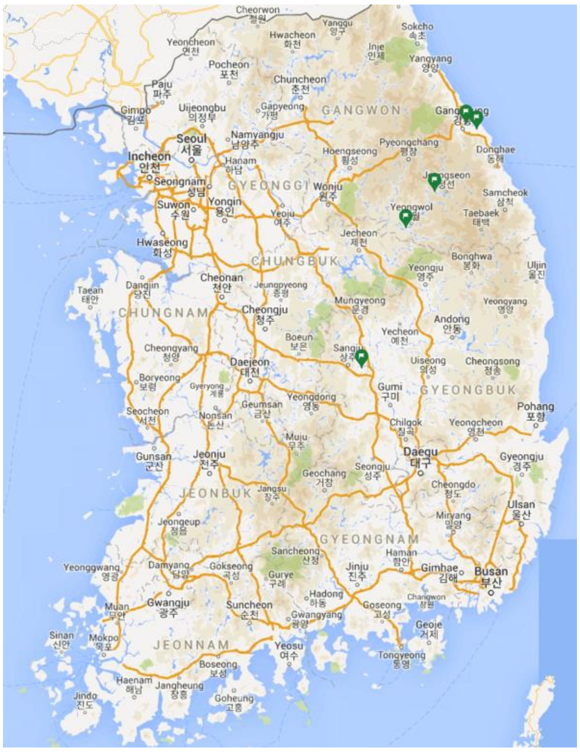

Figure 2. Five places of data used.

Therefore, in this study, based on these problems, a part of the country was randomly selected as a measure to maximize the production of solar electric energy, and the machine learning technique of artificial intelligence was used with the currently installed installation direction, especially the longitude and latitude data. It is intended to present an efficient installation direction by performing a primary analysis using

As shown in Figure 2, the data used for this study show the longitude of the actual solar panel installation location according to year, month, day, and time in 5 regions (Jeongseon, Gangneung1, Gangneung2, Sangju, and Yeongwol) in Korea. A study was conducted to find an efficient power production method with information on latitude and azimuth and inclination angle of the installation site.

In addition, an additional method for effective power production was conducted with temperature, airflow, humidity, and sunlight according to date information.

The power production according to the inclination and azimuth angles was analyzed with the obtained existing information.

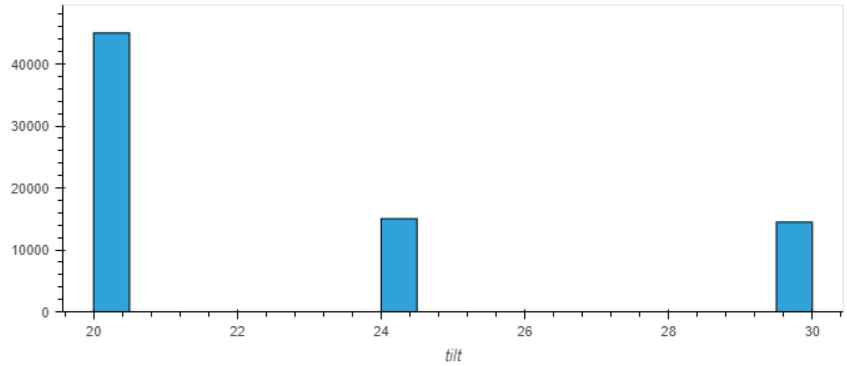

Figure 3. Tilt angle distribution.

As shown in Figure 3, the x-coordinate is the tilt angle, and the y-coordinate is the amount of data. In other words, looking at the distribution of inclination angles in Figure 2 data, the inclination angles are mainly represented by three parts: 20` ~ 20.4`, 24` ~ 24.4`, and 26.6` ~30`. It was found that most of them were distributed in 20` ~ 20.4`.

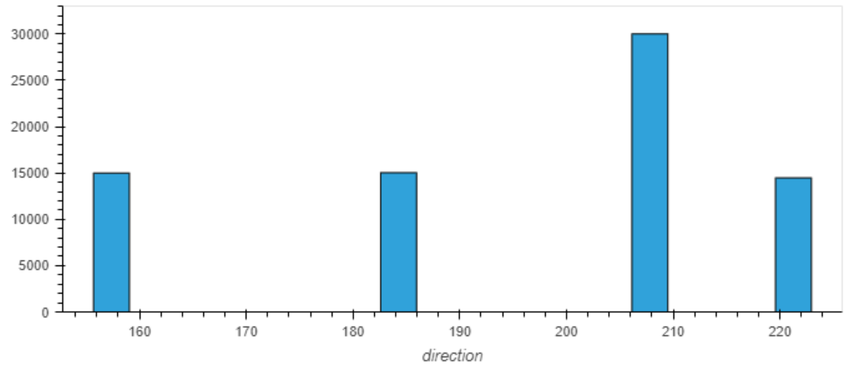

Figure 4. Direction distribution.

As shown in Figure 4, the x coordinate is the direction, and the y coordinate is the amount. That is, Figure looking at the distribution of the direction angle of the data in 2, the direction distribution map is divided into 4 parts. Although it is almost similar, it is widely distributed in 206` ~ 210`.

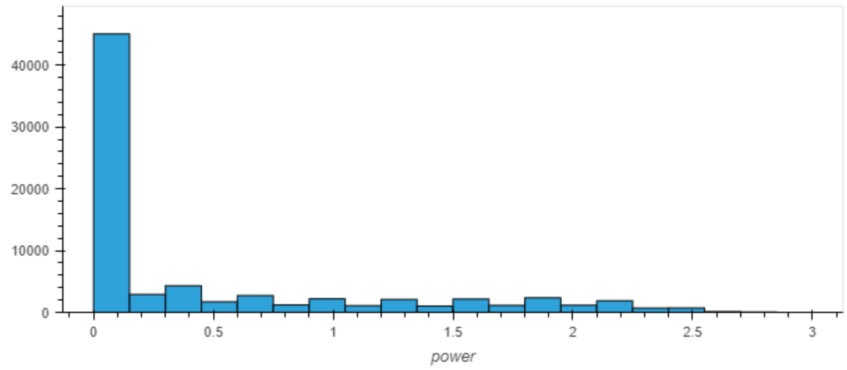

Figure 5. Distribution of electricity production.

As can be seen in Figure 5, the X-coordinate is electricity production, and the y-coordinate is the number. Also, as can be seen from the graph, most of the time shows the power production of 0KW ~ 0.2KW. Since it mainly produces power in proportion to the amount of time the sunlight shines, it cannot produce electricity from evening to night and dawn.

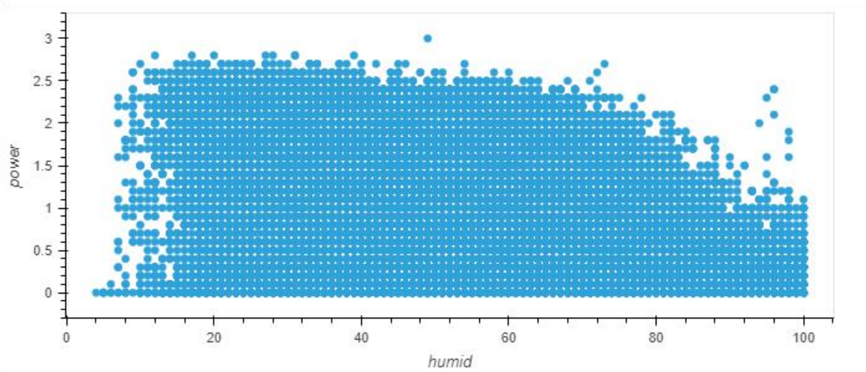

Figure 6. Electricity production according to humidity.

And the contents of the analysis of electricity production according to humidity are shown in Figure 6. That is, as shown in the graph, the x-coordinate represents humidity and the y-coordinate represents the amount of electricity produced, and this represents the amount of electricity produced by dividing the humidity between 0 and 100. As a result of the analysis, the maximum production is reduced when the humidity exceeds 10 and the humidity exceeds 80.

Figure 7 shows the relationship between each feature (latitude, longitude, direction, tilt, year, month, day, hour, temp, wind, humid, sun, power). The closer the result is to 1, the greater the correlation, and conversely, the closer the result is to -0.4, the greater the correlation. If the power to be obtained is compared with the amount of insolation (sun), there is a very large correlation with 0.8. Also, when comparing the power and humidity, it is -0.45, indicating that there is a very large correlation inversely.

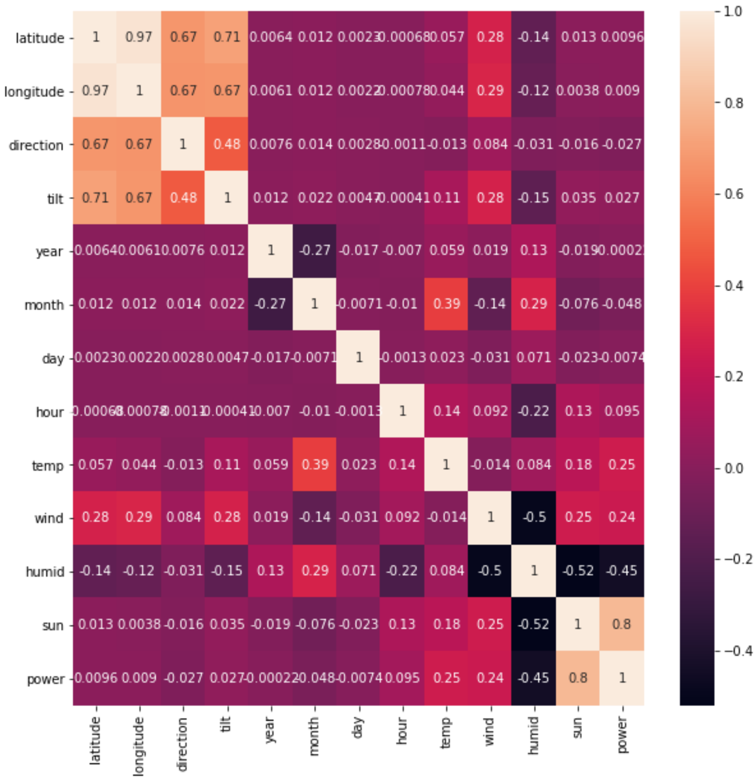

Figure 7. Relationship diagram of each feature of data.

As shown in figure 8, when using XGBoost, the x-axis in figure 8 is the alpha value, and 0 is 10 to the 0th power, 1 is 10 to the 1st power, and 2 is 10 to the 2nd power. Conversely, -1 is 10 to the -1 power, -2 is 10 to the -2 power, and -3 is 10 to the -3 power. The y-axis represents the R2 score, and the closer to 1, the more accurate the judgment. The graph line at the top shows the change in the training data, and the graph line at the bottom shows the change in the test data. Therefore, when the x-axis is 1 (10 to the 1st power), the R2 score of the training data is the highest at 0.986, and when the x-axis is 2 (10 to the 2nd power), the R2 score of the test data is the highest at 0.938. Through the above results, when alpha is 10, train 0.986 and test 0.928 gave excellent results.

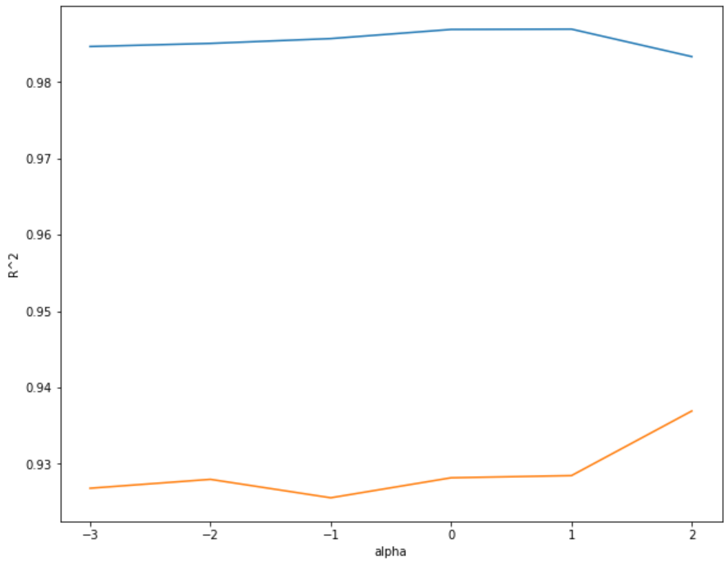

Figure 8. R2 score of training data and test data using XGBoost.

5. Conclusions

In this paper, a study to obtain the maximum efficiency of many photovoltaic panels installed in five regions of Korea. We used the longitude and latitude information of the photovoltaic panel installation location for obtaining electricity production. Also, we utilized the direction and inclination angle of the installation site. The maximum output according to latitude and longitude was generated by obtaining the direction and inclination angles by training the XGBoost regression model.

The data conducted in the study is a CSV file with 14 characteristics: name, latitude, longitude, azimuth, inclination, year, month, day, time, temperature, wind volume, humidity, sunlight, and electricity production. It is divided into Jeongseon, Gangneung 1, Gangneung 2, Sangju, and Yeongwol. As a result of using the XGBoost regression model, it was possible to predict electricity production by obtaining an excellent R2 score of 0.986 for train and 0.928 for test.

In the future, we will use more regional data to predict solar power generation in other regions to bring specific research efficiency, and we want to further expand the scope and method of research for this purpose.

Acknowledgement

This work was supported by the 2022 education, research and student guidance grant funded by Jeju National University.

6. References

[1] https://terms.naver.com/entry.naver?docId=2268719&cid =51331&categoryId=51333
[2] https://terms.naver.com/entry.naver?docId=5702156&cid =64656&categoryId=64656
[3] https://m.blog.naver.com/chammaja1/100194907067

[4] https://data.kma.go.kr/cmmn/main.do
[5] Navin, S., Pranshu S., David, I., Prashant, S. (2011). Predicting Solar Generation from Weather Forecasts Using Machine Learning
[6] Mariam, A., Imtiaz, A. (2019). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods
[7] https://news.skhynix.co.kr/1291
[8] https://terms.naver.com/entry.naver?docId=3537219&cid =60217&categoryId=60217
[9] https://terms.naver.com/entry.naver?docId=5741385&cid=60217&categoryId=60217, [Naver Knowledge Encyclopedia] Photovoltaic system, solar power system (Encyclopedia of Physics)
[10] Yongjun, K., (2021) A Study on the Analysis of Effective Learning Environment for Vehicle License Plate Recognition based on CNN Deep learning, Hannam University Master's Thesis, https://dl.nanet.go.kr/search/searchInnerDetail.do?searchType=INNER_SEARCH&resultType=INNER_SEARCH_DETAIL&searchMehtod=L&searchClass=S&controlNo=KDMT12021000019790&queryText=&zone=&fieldText=&prevQueryText=a+study+on+the+analysis+of+effective+learning+environment+for+vehicle%3AALL_NI_TOC%3AAND&prevPubYearFieldText=&languageCode=&synonymYn=&refineSearchYn=&pageNum=&pageSize=&orderBy=&topMainMenuCode=&topSubMenuCode=&totalSize=1&totalSizeByMenu=1&seqNo=&hanjaYn=Y&knowPub=&isdb=&isdbsvc=&tt1=&down=&checkedDbIdList=&baseDbId=&selectedDbIndexIdList=&caller=&asideState=true&dpBranch=ALL&journalKind=&selZone=ALL_NI_TOC&searchQuery=+a+study+on+the+analysis+of+effective+learning+environment+for+vehicle#none
[11] Sunkyu, H., (2020) Computer vision, machine learning, and Gilbut learning with openCV4
[12] http://blog.naver.com/jevida/221843366216
[13] Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
[14] https://en.wikipedia.org/wiki/XGBoost
[15] Eunsun, C., Som, A., Yujeong, J., Jinsil, K., Jaesung, K. (2020). A Study on the Comparison of Power Data Prediction Performance Using XGBoost and NGBoost
[16] https://www.etnews.com/201311201082, Solar panels should be installed to face west rather than south - EtNews, 2013.11.20.