Forecasting Greenhouse Gas Emissions Based on Different Machine Learning Algorithms

Ilayda Ulku[0000-0003-0464-7007] and Eyup Emre Ulku[2][0000-0002-1985-6461]

[1] Faculty of Engineering, Industrial Engineering Department, Istanbul Kultur University, Istanbul, 34156, Turkey
[2] Faculty of Technology, Computer Engineering Department, Marmara University, Istanbul, 34722, Turkey
i.karabulut@iku.edu.tr

Abstract. With the increase in greenhouse gas emissions, climate change is occurring in the atmosphere. Although the energy production for Turkey is increased at a high rate, the greenhouse gas emissions are still high currently. Problems that seem to be very complex can be predicted with different algorithms without difficulty. Due to fact that artificial intelligence is often included in the studies to evaluate the solution performance and make comparisons with the obtained solutions. In this study, machine learning algorithms are used to compare and predict greenhouse gas emissions. Methane (CH_4), carbon dioxide (CO_2), nitrous oxide (N_2O), and fluorinated gases (F-gases) are considered direct greenhouse gases originating from the agriculture and waste sectors, energy, industrial processes, and product use, within the scope of greenhouse gas emission statistics. Compared to different machine learning methods, support vector machines can be considered an advantageous estimation method since they can generalize more details. On the other hand, the artificial neural network algorithm is one of the most commonly used machine learning algorithms in terms of classification, optimization, estimation, regression, and pattern tracking. From this point of view, this study aims to predict greenhouse gas emissions using artificial neural network algorithms and support vector machines by estimating CO_2, CH_4, N_2O, and F-gases from greenhouse gases. The data set was obtained from the Turkish Statistical Institute and the years are included between 1990 and 2019. All analyzes were performed using MATLAB version 2019b software.

Keywords: Machine Learning Algorithm, Greenhouse Gases, Forecasting.

1 Introduction

Energy consumption and carbon footprint topics have been important issues in today's world in comparison with the past decade. With the increase in energy consumption, electricity generation including wind power, hydropower, and also solar power is the preferred option for electricity generation from renewables. However, renewable energy can meet only around half of the forecast of current trends of global electricity demand according to the IEA report [8].

According to economic trends about the electricity generation with the help of renewables grow significantly around all over the world by 8% in 2021 and also by more than 6% in 2022 where renewables include wind and solar PV and hydropower. Although, this significant increase in the electricity generation from the renewables can meet only half of the global electricity demand for the reported years 2021 and 2022, according to the new IEA report.

The electricity production information is presented by Turkey Electricity Distribution Inc. (TEDAŞ) on the basis of measurement of natural gas, renewable energy and wastes, hydroelectric energy, coal and coal derivatives, and liquid fuel data. The dataset covers the years between 1990 and 2020 and is obtained from TEDAŞ data [1]. According to the dataset represented in Fig. 1, the electricity production from coal decreases in Turkey, but it is not reflected in greenhouse gas emissions.

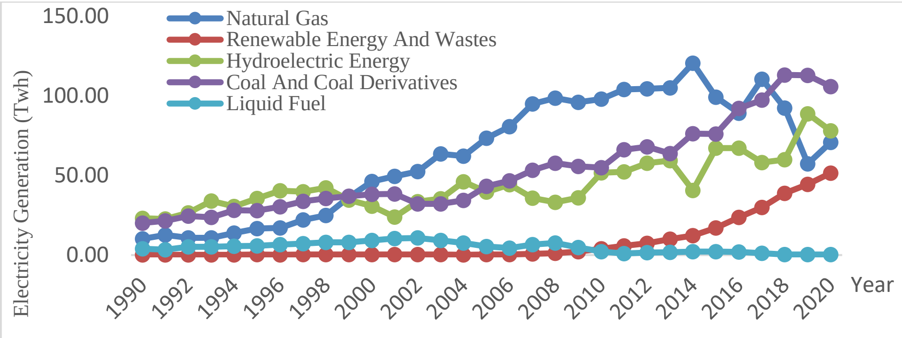

Fig. 1. Electricity generation in Turkey between 1990 and 2020.

On the other hand, there is a significant increase in renewable energy and waste since 2017. According to the dataset, electricity generation increased by 15.7% in 2020 compared to the previous year. Also, this electricity generation in renewable energy increased its share to 16.8% of total electricity generation with 51.54 TWh. Accelerating renewable energy investments will reduce the greenhouse gas emissions from coal power plants. Thus, renewable energy investments provide of great importance in preventing climate change in the atmosphere.

It is obvious that there are several studies in the literature for forecasting CO_2 emissions [5, 6, 9]. In this study, the energy production, number of road vehicles, and the amount of municipal waste collected in tonnes per year are considered. There is a strong correlation between each input pair in terms of CO_2 as an output parameter.

In order to analyze the efficiency of machine learning algorithms, the data set obtained from the estimation of the CO_2 emissions was first used. To train the algorithms, the data between 1990 and 2009 is used. Then, the last 5 years covering the years 2015 and 2019 were estimated with 2 algorithms. Among several statistical metrics the coefficient of determination, R^2 (R-Squared) is used to determine how well the data fit the developed model in terms of discussing the performance of the algorithms. Artificial neural network and support vector machine are used to predict greenhouse gas emissions by using year, electricity generation (Mwh), the number of road motor vehicles, Turkey's population, and the amount of municipal waste collected tonnes per year as input parameters.

The remainder of this study is organized as follows: In section 2 the descriptive statistics of the data set are represented. Also, mathematical models and machine learning algorithms are discussed with the relationship between input and output variables. Statistical comparisons, formulas, explanations, and success criteria are also presented under Section 2. The results of the research are given under Section 3. Finally, the paper concludes with the conclusion and recommendations in Section 4.

2 Methodology

In this study, the year, the population in million, the energy production, number of road vehicles, and the amount of municipal waste collected in tones per year are used as the input dataset. The CO_2 emissions are used as the output of the data. The main greenhouse gas that is responsible for the environmental pollution effect on the heat balance is CO_2.

The Climate Convention reported that Carbon dioxide (CO_2), Methane (CH_4), Nitrous Oxide (N_2O), and Fluorinated Gas (F-gases) are values that imply global warming [2]. The sum of CH_4, CO_2, N_2O, and F-gases between the 1990 and 2019 periods are used as CO_2 equivalent greenhouse gas emissions for this study.

There is a 72% of total greenhouse gas emissions in 2019 which had the largest share with energy-related CO_2 equivalent as represented in Fig. 2. On the other hand, there is a significant increase which is recorded as 161% in 2019 in emissions compared to 1990 and 2019. But, when the previous year is compared there is a decrease that is recorded as 2.3% [3].

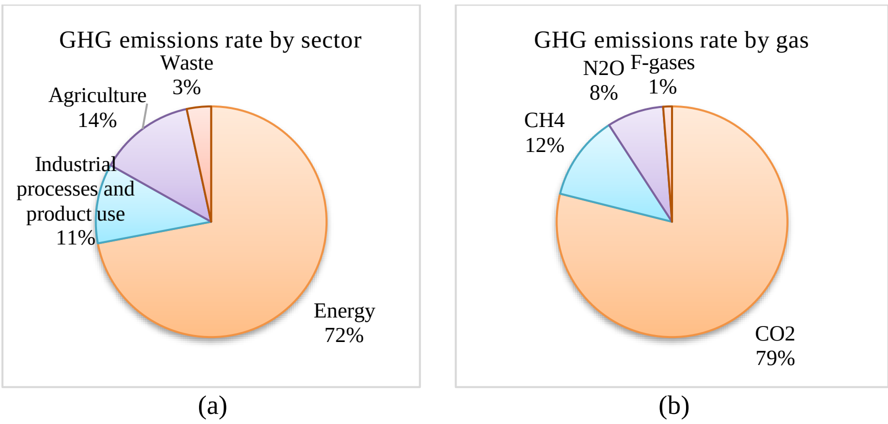

Fig. 2. Total greenhouse gas emissions in 2019 (a) by sector (b) by gas

To compare and predict greenhouse gas emissions machine learning algorithms are used in this study. Methane (CH_4), carbon dioxide (CO_2), nitrous oxide (N_2O), and fluorinated gases (F-gases) are considered direct greenhouse gases named as CO_2

equivalent (106 tonnes) as in the second column of Table 1. The descriptive statistics are represented in the dataset used in this study.

Table 1. Descriptive statistics of the dataset.

	CO_2 (10^6 tonnes)	Year	Electricity Generation (Mwh)	# of road motor vehicles	Turkey population	Collected municipal waste (10^3/year)
Std. Error	18.05	1.61	14333616.65	1121304.42	1527313.77	796855.73
Median	326.09	2004.50	156327250.00	10691091.50	68435377.00	25205500.00
Std. Dev.	98.89	8.80	78508451.68	6141637.26	8365442.04	4364558.58
Kurtosis	-1.25	-1.20	-1.17	-1.10	-1.13	-0.26
Skewness	0.36	0.00	0.29	0.44	0.07	-0.11
Range	305.41	29	247258885.00	19406297.00	28034997.00	14567472.38
Minimum	219.57	1990	57543000.00	3750678.00	55120000.00	17757000.00
Maximum	524.98	2019	304801885.00	23156975.00	83154997.00	32324472.38
Observation	30	30	30	30	30	30

As long as there are different units for each parameter, the data of each parameter must be scaled in order to obtain a normalized range that is between 0 and 1 as given in equation (1). X_1 is used as the actual data and the minimum and the maximum values in the dataset are given as $X_{1(\mathrm{min})}$ and $X_{1(\mathrm{max})}$, respectively. The normalized value is represented with $X_{1(normalized)}$ which takes a value between 1 and 0.

$$X_{1(normalized)} = \frac{X_1 - X_{1(min)}}{X_{1(max)} - X_{1(min)}} \quad (1)$$

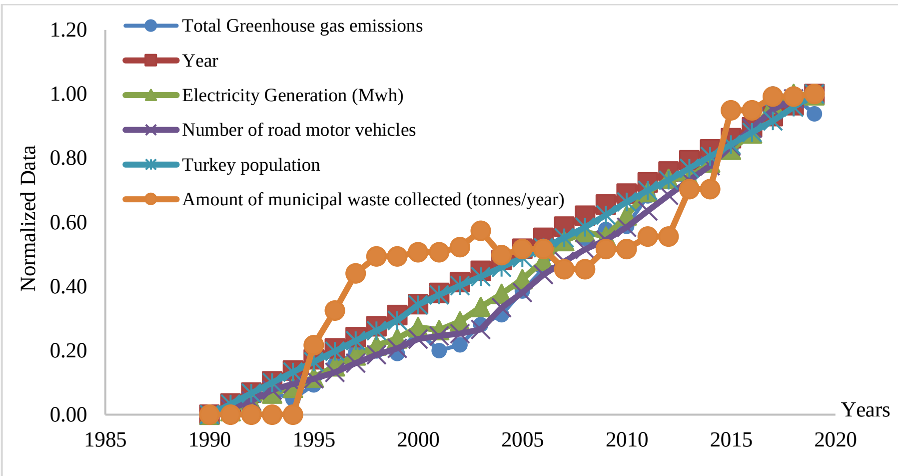

Fig. 3. Normalized values of each used dataset.

Also, normalized values of each used dataset are represented in Fig. 3. There is a significant increase over the years in each parameter.

In order to see if there is any linear relationship between variables correlation analysis is used as a statistical method. The correlation values of each dataset are represented in Table 2. When the coefficient is close to 1 there is a perfect correlation which means a change on the variable affects the other variable. According to the values in Table 2, these datasets are valid to be used in the machine learning algorithms to forecast CO_2 emissions.

Table 2. Correlation values of each attribute in the dataset.

	CO_2	Year	Electricity Generation	Number motor vehicles	Population	Municipal waste
CO_2	1					
Year	0.982	1				
Electricity Generation	0.996	0.993	1			
Number of motor vehicles	0.995	0.985	0.997	1		
Population	0.983	0.999	0.994	0.988	1	
Municipal waste	0.885	0.916	0.902	0.895	0.920	1

2.1 Machine Learning Algorithms

Artificial Neural Network (ANN) Algorithm. The process of a human neural system is used to determine if there is any connection between inputs and outputs with the help of Artificial Neural Networks (ANN). Each connection means a neuron of hidden layers that is used with a proper weight. Model training consists of setting values. There are target values whose weights are optimized to match output estimates and targets [6]. This is accomplished by calculating the error between the forecasts and targets and then adapting the weights via learning methods [5]. In this study, a feedforward backpropagation network type is used. As a training function gradient decent function is used that is commonly-used in order to train neural networks and machine learning models. There is a single layer that consists of 10 neurons in the model.

Support Vector Machine (SVM) Algorithm. To find the distance between data points of distinct groups Support Vector Machines (SVM) are used. In this algorithm, the Kernel distribution is considered to optimize the distance [7]. In this study, the Gaussian radial technique is considered.

In order to train the model to make proper predictions for new data, the kernel approximation classifiers are used in the model with several observations.

3 Numerical Results

The data set was obtained from the Turkish Statistical Institute and the years are included between 1990 and 2019. All analyzes were performed using MATLAB version 2019b software.

In forecasting models, R^2 is used as the proportion of the variance in order to predict the output by using the input variables. The fit quality of the proposed model is measured with R^2 [4] as represented by equation (2) where y_i is the actual value and $\hat{y}_i$ is the predicted value for all i from 1 to the m.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{2}$$

The following Fig. 4 represents the variance of the predicted output by using the proposed model. The ANN model explains 99.58 % of the variation when all predictors are considered.

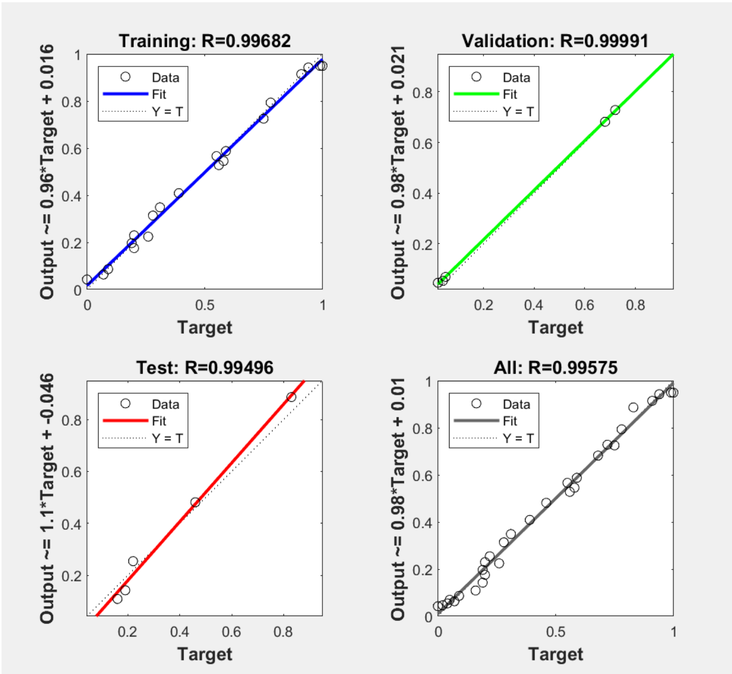

Fig. 4. Training regression.

The results of the proposed model performance and the actual data are represented in Fig. 5.

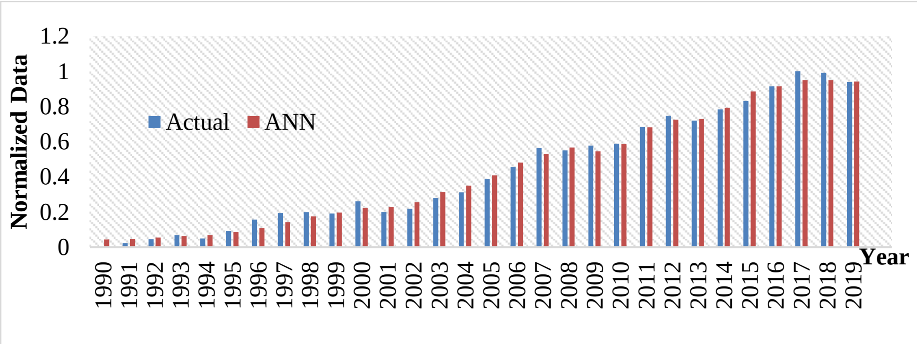

Fig. 5. Forecasting of CO_2 emissions.

The following Fig. 6 represents the plot to identify useful predictors for separating classes and to understand the relationships between features. The training data and misclassified points with dashed lines are illustrated on the parallel coordinates plot.

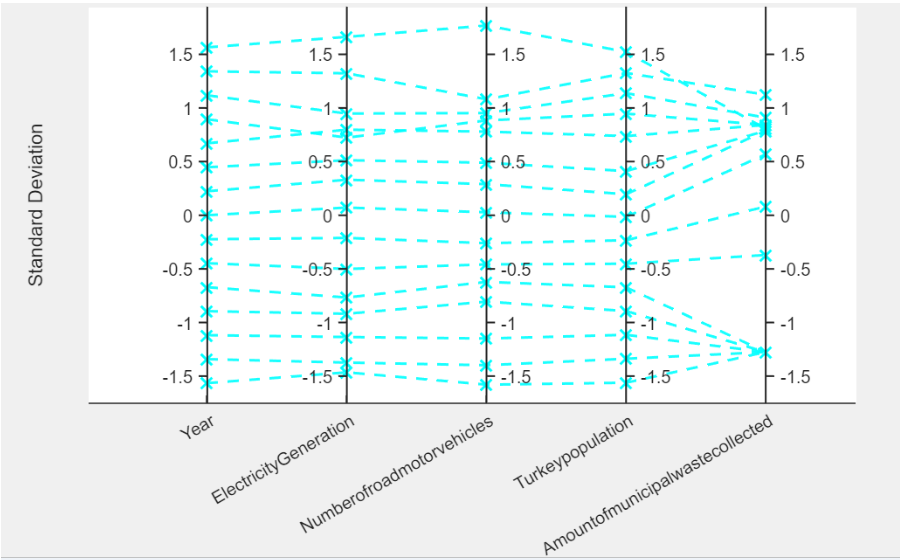

Fig. 6. Parallel coordinates plot for separating classes with predictors.

4 Conclusion

In this paper, among several machine learning algorithms, artificial neural network and support vector machine are used to predict greenhouse gas emissions. Methane (CH_4), carbon dioxide (CO_2), nitrous oxide (N_2O), and fluorinated gases (F-gases) are considered direct greenhouse gases. Year, electricity generation (Mwh), number of road motor vehicles, Turkey's population, and the amount of municipal waste collected tonnes per year are used to predict total greenhouse gas emissions. ANN model explains 99.58 % of the variation when all predictors are considered in terms of forecasting the CO_2 emissions. With the help of SVM, useful predictors can be identified for separate classes. Also, the relationships between features can be demonstrated clearly with the help of the algorithms. The results are satisfied for each machine learning algorithm in forecasting CO_2 emissions.

Machine learning algorithms represent convincing forecasting results. Therefore, in a future study, several machine learning algorithms will be studied and the performance success of the machine learning algorithms in the forecast will be evaluated.

References

1. Turkey Electricity Distribution Inc. (TEDAŞ), Energy Statistics, https://biruni.tuik.gov.tr/medas/?kn=147&locale=tr, last accessed 2022/03/24.
2. United Nations, Reporting from and review of Parties included in Annex I to the Convention Reports on national greenhouse gas inventory data from Parties included in Annex I to the Convention for 1990–2016, 1990–2017, 1990–2018 and 1990–2019, https://unfccc.int/sites/default/files/resource/sbi2021_11E.pdf, last accessed 2022/03/24.
3. Turkish Statistics Institution, Greenhouse Gas Emission Statistics, 1990-2019https://data.tuik.gov.tr/Bulten/Index?p=Sera-Gazi-Emisyon-Istatistikleri-1990-2019-37196, last accessed 2022/03/24.
4. Tao, Y., Yue, G., Wang, X.: Dual-attention network with multitask learning for multistep short-term speed prediction on expressways, Neural Comput. Appl. 33(12), 7103–7124 (2021).
5. Bakay, M., S., Agbulut, Ü.: Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. J. Clean. Prod. 285, 125324, (2021).
6. Platon, R., Dehkordi, V., R., and Martel, J.: Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. Energy Build 92: 10–18 (2015).
7. Lazos, D., Sproul, A., B., and Kay, M.: Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. Renew Sustain Energy Rev 39: 587–603 (2014).
8. Electricity Market Report, July 2021, https://iea.blob.core.windows.net/assets/01e1e998-8611-45d7-acab-5564bc22575a/ElectricityMarketReportJuly2021.pdf, last accessed 2022/03/24.
9. Cao, Y., Yin, K., Li, X., and Zhai, C.: Forecasting CO_2 emissions from Chinese marine fleets using multivariable trend interaction grey model. Applied Soft Computing 104: 107220 (2021).