

Data Mining For CO2 Emissions Prediction In Italy İtalya'daki CO2 Emisyon Tahmini için Veri Madenciliği

¹Saleh Abuzir , ²Yousef Abuzir 

¹Brescia University, Civil and Environmental Engineering, Brescia, Italy

²Al-Quds Open University, Faculty of Technology and Applied Sciences, Ramallah Palestine

¹s.abuzir@studenti.unibs.it, ²yabuzir@qou.edu

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received :15 January 2021

Accepted : 8 February 2021

Keywords:

CO2 emissions, Climate change, energy consumption, Data Mining, WEKA

ABSTRACT

This study is a preliminary evaluation of the situation of CO2 emissions in Italy, reviewing the international and national literature, using global datasets, and using data mining techniques for analysis and prediction. The study used descriptive methods. It focuses on finding the main potential parameters that effect the concentration of CO2 emissions based on energy resources in Italy. Sequential Minimal Optimization regression (SMOreg), Linear Regression, and Simple Linear Regression are used. Based on the analysis, the Liquid Fuel sector has had the highest rate of increase in CO2 emission 56.8%. R. Linear Regression algorithm gives us a better performance of the prediction for the CO2 emissions than the second algorithm Simple Linear Regression. These results are in line with the present condition in Italy due to the Italian National Program on Climate Change which focuses on reducing carbon dioxide emissions.

© 2020 BandırmaOnyediyül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

MAKALE BİLGİSİ

MakaleTarihleri

Gönderim:15 Ocak 2021

Kabul : 8 Şubat 2021

AnahtarKelimeler:

CO2 emisyonu, İklim değişikliği, Enerji tüketimi, Veri madenciliği, WEKA

ÖZET

Bu çalışmada ulusal ve uluslararası literatür çalışmaları gözden geçirilerek ve global veri kümeleri ve veri madenciliği teknikleri kullanılarak İtalya'daki CO2 emisyon durum tahmini ve değerlendirmesi yapılmıştır. Çalışmada tanımlayıcı teknikleri kullanılmış ve İtalya'daki enerji kaynaklarına dayalı olarak CO2 emisyonlarının konsantrasyonunu etkileyen ana potansiyel parametreleri bulmaya odaklanılmıştır. Sıralı Minimal Optimizasyon regresyonu (SMOreg), Doğrusal Regresyon ve Basit Doğrusal Regresyon kullanılmıştır. Analize göre, Sıvı Yakıt sektörü CO2 emisyonunda en yüksek artış oranı% 56,8 olmaktadır. R. Doğrusal Regresyon algoritması, CO2 emisyonları tahmininde ikinci algoritma Basit Doğrusal Regresyondan daha iyi bir performans sağlamaktadır. Elde edilen bu sonuçlar, karbondioksit emisyonlarını azaltmaya odaklanan İtalyan İklim Değişikliği Ulusal Programı nedeniyle İtalya'daki mevcut durumla uyumludur.

© 2020BandırmaOnyediyülÜniversitesi, MühendislikveDağaBilimleriFakültesi.Dergi Park tarafından yayınlanmaktadır. TümHaklarıSaklıdır.

1. INTRODUCTION

Climate change is one of the factors that threaten the environment, health, and the economy of international societies. In 1990, the United Nations issued a convention on climate change with the goal of limiting the emission of carbon dioxide [1]. Most of the countries of the world responded positively to this agreement and about 186 parties have ratified it. On December 15, 1993, the European Union accepted it. The most important goal of the climate change agreement was to stabilize greenhouse gases (GHG) [2].

Today, we can notice that greenhouse gas emissions (GHG) from human activity, especially from the use of fossil fuels in their various daily activities, the change of the land, and deforestation, have led to an increase in the concentrations of these gases in the atmosphere. An increase in carbon dioxide levels is expected to have clear economic, environmental and social consequences, and will have widespread negative impacts on human health and natural habitats[3]. Climate change will spread insect-borne diseases such as malaria and dengue fever, habitat loss, and security concerns, redraw the map of this universe, and affect the weather. The change in the climate will cause harm to nature and ecosystems.

Climate change has a negative impact on most countries of the world, including Italy. Follow the Kyoto Protocol; Italy has a legal commitment to reduce GHG emissions by an average of 6.5 percent below the base-year level 2008-2012 with an average of 485.7 MtCO₂ equivalents per year in 2008-2012. Taking into account that Italy has one of the lowest energy intensities among Organization for Economic Co-operation and Development (OECD) countries based on energy consuming per GDP unit, achieving this commitment is a challenge. Italy, however, has agreed to abide itself to such an ambitious goal because it knows that if climate change has not been seriously handled, the impact on the global environment, the economy, and security will be very negative [1].

It is important to understand the past, current, and future events related to the world's climate and use data mining and Artificial Intelligence (AI) research to analyze and identify trends and characteristics of climate data.

Using the Data Mining approach to study the characteristics of CO₂ emissions will effectively support scientific theory and should be central to any interdisciplinary process in order to measure or quantify climate prediction uncertainties to discover the relations between climate system disruption and its response to energy consumption. In order to understand how climate change influenced the environment in the ancient past and to find ways to foresee how things will change in the future, we investigated patterns from the records of CO₂. Historical data records for CO₂ emissions were studied and analyzed. The results showed that there are clear climate changes in the environment that occur over time. These changes and upheaval are repeated in a similar pattern most of the time and show a good relationship that can be used for predictive analysis of changes occurring in the environment. Because these patterns are repeated regularly, it is crucial to understand how the ecosystem deals with these fundamental processes and how the natural system response to climate change. Studying these patterns will provide us with an insight and a future vision of what the future may hold and what measures can be taken and implemented to lower the impact of negative changes on the environment [3-5].

This study is a preliminary evaluation of the situation of CO₂ emission in Italy, using international and national literature review, global and historical datasets, and using data mining to find useful information in the datasets.

This study identifies, analyze and evaluate potential parameters that affect carbon dioxide emissions in Italy using data mining and discuss the various measures, tools and procedures used to reduce the percentage of carbon dioxide emissions in Italy that have achieved a positive and real impact, to achieve the nation's goals regarding it concerns sustainable development..

In the following sections, key Data Mining challenges facing climate changing in general and CO₂ emissions in particular are discussed. The study gives an overview of recent efforts to apply Data Mining to this field with different cases of studies related to climate change. This paper began by a general overview of previous related studies to data mining and addressed different uses of data mining in climate change and CO₂ emissions predication. The paper then discusses the methodology, and the dataset used in this study. The last two sections report the experimental analysis and deal with a conclusion.

2. LITERATURE REVIEW

Jeong, et al., 2020 used data mining techniques to establish a benchmark in South Korea for CO₂ emission using a decision tree algorithm [6].

The paper of Jeslet and Jeevanandham, [7] used two machine learning algorithms, Artificial Neural Network (ANN) and Decision Tree in forecasting different weather forecasting. The parameters used in their study are speed of the wind, temperature degree of the air, the amount of rain, and water evaporation. A meteorological data model was generated and it was used to train the classifier. Based on their evaluation the study showed good performance in using these two algorithms for forecasting or predicting weather events and predictions of climate change in their area.

The work of Somu et al., 2020,[8] presents a deep learning framework kCNN-LSTM based on the following algorithms to KMeans clustering, Long Short Term Memory (LSTM) neural networks and Convolutional Neural

Networks (CNN) to provide accurate building energy consumption forecasts based the energy consumption data recorded at predefined intervals.

The main objective of the study of Farhate et al., 2018[9] was to use predictive analysis techniques of data mining to predict the amount of soil emissions of CO₂ induced by crop management in Brazilian sugarcane fields. Their approach showed high performance and accuracy in predicting CO₂ emissions and lower cost for computation using the Multilayer Perceptron classifier and the logistic regression Bagging classifier.

The Coronavirus disease was detected at the end of December 30, 2019, in a Chinese city named Wuhan, after the Chinese government announced an outbreak of the virus. On March 11, 2020, the World Health Organization declared the disease a global pandemic [10]. Studies results showed that there is an increase equal to 1% of emissions of CO₂ growth rate by the year 2019[11], The measurement showed that there is no increase in the percentage of emissions of carbon dioxide (CO₂) in late 2020, after COVID-19 [12]. Le Quéré et al., [13] in their work, they studied and estimated the daily change in CO₂ emissions and its implications during the COVID-19 confinement based on official data associate with the nature of energy recourses, daily activities, and policies. During the COVID-19 confinement and the government decisions, the percentage of the daily global CO₂ emissions decreased by -17% and fell to 83MtCO₂ (with a range of -11 to -25% for $\pm 1\sigma$) compared to the 100MtCO₂ released on an average day in 2019.

The size of the global level of CO₂ emissions reduced because most countries were under the lockdown. It is worth mentioning that this daily global decrease in carbon emissions is close or equal to half of the percentage resulting from daily global carbon emissions from surface transport.

Using of big data in climate change has limit applications. The paper of Hassani et al., 2019[14] investigates and summarizes the approaches of using the techniques of big data in climate change. The study focuses on promoting new scientific researches in climate change by using big data.

A data mining technique was used to study climate change results from vehicles by analyzing fuel consumption and passenger vehicle emissions [15], predict carbon prices [16], and forecast CO₂ emissions for China with consideration given to its population [17]. Another study used data mining to forecast the main influencing factors of CO₂ emissions is Grey Model (GM) [18]. Data mining used in low-carbon management, the discovery of evolving patterns of climate change, predict climate change impacts, remote-sensing, economic, technological, and management approaches to combat climate change, exploitation of trans-Arctic maritime transportation [19, 20].

While there is much research related to the literature of climate change, climate statistics or datasets and widely spread applications of data mining in health, business, law, education, recommendation, tourism as well as other areas, there is a lack of studies in climate change.

Data mining strategies for classification include analyzing the various attributes associated with different data types. Once we identify the main characteristics of these data types, we can classify related data. It is important for recognizing CO₂ emissions data that can be used for prediction [21].

Regression methods are useful in defining the essence of the relationship in a dataset between variables. In certain cases, these associations may be causal or merely correlate with others. Regression is a basic technique of the white box that simply shows how variables are related. Regression techniques used in aspects of forecasting and data modeling [21].

3. MATERIALS AND METHODS

3.1. Method

The study raises questions on how to apply data mining in climate change with a case study of CO₂ emissions in Italy, how to operate the process of assigning responsibility, if and how CO₂ emissions account for everyday life, and the wider moral and political implications of this allocation of government and individual responsibility.

The research consisted of four methodological steps:

- Literature review of current practical approaches using data mining in CO₂.
- Understand opportunities and challenges for CO₂ emissions reduction based on datasets to generate quantitative data.
- The analysis of data using Machine Learning (ML) algorithms to highlight key trends that are cross-referenced with the qualitative reflections.

The importance of research comes through its dealing with a contemporary and important topic, considering the world's interest in climate change. The framework requires the initiation and implementation of specific frameworks, based on the change and revise in human behavior that is the first cause of environmental problems. This framework follows a clear strategy consistent with the future, and with the needs of the present alike.

This study seeks to achieve a set of objectives, the most important are shows the key role that data mining plays in studying climate change to prompt future research, presents the different applications and case studies of using

data mining in the climate change. In this paper, a case study is presented to analyze the prediction and study of the percentage of carbon dioxide emissions in Italy using data mining techniques.

In the following sections, we focus on how to use data mining to analyze CO2 emissions using coupled with classification and prediction techniques to extract and analyze statistical interrelationships in fields of CO2 emissions in Italy.

Finally, the paper concludes how to bring significant implications for climate change based on predictive analysis, and patterns interpretation of CO2 emissions, and enactment of responsibility.

3.2. Statistical Datasets

One of the ways data scientists or researchers can understand how natural systems react to climate change is by looking at the past. Recently, the availability of many reports and observations related to climate change, in addition to the increased availability of various data sets, gives climate change researchers new opportunities to use data mining techniques to provide new scientific insights.

The raw data used in this study is obtained from datasets uploaded to the server of the World Bank¹. The original dataset has seven variables or parameters, the response parameter CO2 emissions is dependent variable, and the remaining 6 parameters are independent or explanatory variables (year, Solid Fuel, Liquid Fuel, Gas Fuel, Cement, Per Capita, and Bunker fuels) (Table 1). The datasets contain 1248 observations.

Table 1. Dependent and Independent variables used in the study (WEKA)

Variable name	Minimum	Maximum	Mean	StdDev
Solid Fuel	8	17368	7488.994	4778.887
Liquid Fuel	0	75482	21840.31	29555.807
Gas Fuel	0	44127	7779.026	13100.996
Cement	0	6503	1812.413	2255.59
Per Capita	0	2.22	0.641	0.855
Bunker fuels	0	7664	1781	2383.877
Total	8	129253	38920.761	46801.968

The datasets present CO2 emissions in Italy for the period 1960 to 2014. The main characteristics of the datasets are available in electronic format like CSV files, it is a time series data, shows interesting patterns, and meeting the overall objectives of the study. A preprocessing of the datasets is required before using it with the WEKA tool. The raw data need to be formatted in a way compatible with WEKA [21].

3.3. Data Mining

Data mining approach enables us to make scientific discoveries, gain fundamental insights, discover new hidden knowledge, find new patterns or identify patterns of climatic relevance, associations, anomalies, and statistically significant structures or information in data [22]. Data mining is used for classification, clustering, and association of the different data types. It uses machine-learning algorithms. It consists of different phases for data manipulation starting with the data preparation model that includes the different processes for cleaning the data by removing noisy data and filling missing values, integrating data from the different digital resources by eliminating and reducing redundancy, then we select the interesting data for manipulation. Finally, we format and transform the data into other forms. The transformed data are the output of this data preparation model.

The second model is called Data Mining. This model includes different techniques and algorithms for classification, clustering, and association to search and identify interesting patterns in the data. The selected patterns then visualized in different ways to the user. Finally, data mining ends with an evaluation model where users can predict, validate, and interpret the results to confirm or prove some results or hypothesis.

4. RESULTS AND DISCUSSION

Data mining tool WEKA used for analysis and prediction based on the CO2 emission datasets. Data mining approaches can be used to generate new insights, to turn raw data into actionable insights, and find new trends. We used the Sequential Minimal Optimization regression (SMOreg) algorithm for time series for the analysis of

¹<https://data.worldbank.org/country/italy>

our data. The data were split into training and testing data to run WEKA on our data. In our study, 66% of the data used as training data and the remainder used as testing data. WEKA can manipulate data in different formats such as ARFF, CSV, and others. We used the CSV format for our data.

To analyze the various attributes associated with CO2 emissions datasets, we identify the main characteristics of these data types and classify related data. In this study, we used the SMOreg algorithm to recognize CO2 emissions patterns that can be used for prediction. Figure 1 shows CO2 Emissions in Italy in the period 1865-2014.

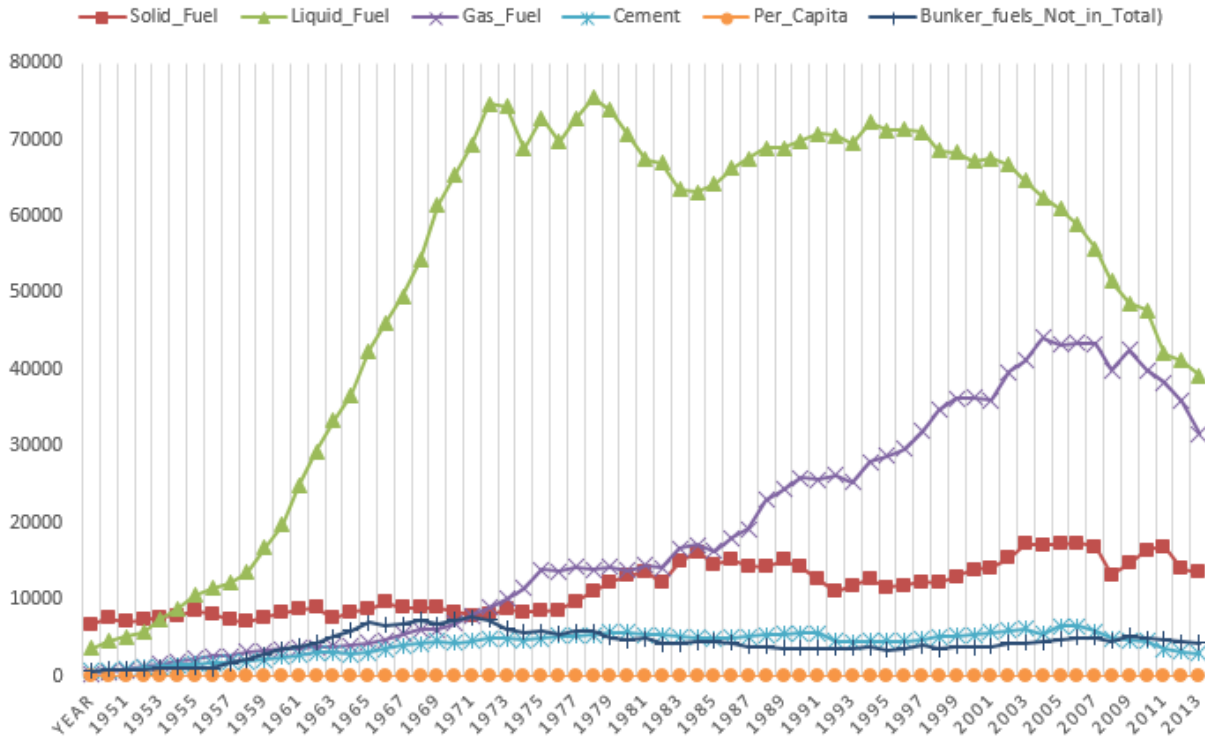


Figure 1.CO2 Emission by energy resources in Italy 1865-2014

In linear regression, the dependent variable is the predictor variable and the other independent variables help us in measuring and assessing the performance of the predictor variables. Linear regression is used to predict our dependent variable by evaluating and weighting all the independent variables in our datasets. As mentioned previously, the dependent variable is CO2 Emissions, and the independent variables are Liquid Fuel, Gas Fuel, Solid Fuel, Cement, Per Capita, and Bunker fuels.

Regression is different from classification in that classification places objects into pre-defined and clear classes while regression gives a numerical result. Therefore, we will analyze our data using WEKA to predict the weighting of the independent variables in the CO2 emission data set.

Based on the analysis and the results we got from WEKA, the average percentage contribution of CO2 emission is 56.8% for Liquid Fuel, 33.1% for Gas Fuel, 13.7% for Solid Fuel, 4.6% for Cement, 2.9% for Per Capita, 0.1% for Bunker fuels in Italy between 1860 -2014. The Liquid Fuel sector has had the highest rate of increase in CO2 emission starting from 1860 until 1995 with minor irregular in the percentage of emission of CO2 within these years. Then the percentage contribution of CO2 emission starts decreasing after 1995. This is true for Gas fuel compared to other resources. Gas fuel is the second highest source of the high percentage contribution of CO2 emission. In 2007, the percentage contribution of CO2 emission for Gas fuel and Cement starts decreasing too. The percentage contribution of CO2 emission for other resources is changing up and down. In the year 1978, Liquid Fuel surpassed the other resources as the leading contributor to CO2 emission and reached a peak of 56.8%. (Table 2) compares the result of running the SMOreg algorithm with a different value for the Complexity parameter C.

We run WEKA with different configurations for SMOreg based on the complexity parameter C. Table 2 shows the result of running SMOreg algorithm with two values for the parameter C=1 and 3. Referring to the results computed from WEKA (Table 3), we found that, the best accuracy of the prediction of CO2 emission by using SMOreg when C=3. We calculated Root Mean Squared Error measurement (RMSE) to choose the accurate configuration in the regression model.

Table 2. Results of SMOreg algorithm with different value for the complexity parameter C

Variable	Average Percentage Contribution Of CO2 Emission with	
	The complexity parameter C = 1	The complexity parameter C = 3
Liquid Fuel	56.8%	57.0%
Gas Fuel	33.1%	33.3%
Solid Fuel	13.7%	13.6%
Cement	4.6%	4.6%
Per Capita	2.9%	2.5%
Bunker fuels	0.1%	0.1%

Table 3. RMSE for SMOreg with C=1 and C=3

Shop floor	complexity parameter C = 1	complexity parameter C = 3
Correlation coefficient		1
Mean absolute error	99.9999	88.8143
Root mean squared error	124.9488	115.9269
Relative absolute error	0.2317 %	0.2058 %
Root relative squared error	0.2572 %	0.2386 %

Root Mean Square Error (RMSE) allow us to estimate how much error there is between two timeseries data sets. In other words, it compares a predicted value and an observed or known value (equation (1)). The smaller an RMSE value, the closer predicted and observed values are.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \tag{1}$$

Where

i = variable, N =number of non-missing data points, xi = actual observation time series and \hat{x} = estimated time series.

We used WEKA to evaluate the linear regression on our problem. We find that the linear regression for CO2 emissions can be predicted by the following formula equation (2):

$$Total = 1 * Solid_Fuel + 1 * Liquid_Fuel + 1 * Gas_Fuel + 1 * Cement + -0.0001 * Bunker_fuels) + 0.1126 \tag{2}$$

The performance measurements for linear regression is presented in Table 4

Table 4. The performance measurements for linear regression

Measurements	Value
Correlation coefficient	1
Mean absolute error	0.3668
Root mean squared error	0.576
Relative absolute error	0.0008 %
Root relative squared error	0.0012 %

Let us look at the regression model (equation 2); we have the coefficients, our independent variables and then the offset. Therefore, what we can see is that each one of our attributes was included. Sometimes some of these will be dropped out. Therefore, the Bunker_fuels has a minus or negative sign, because of that it reflects negatively. It does not have a huge affect on total emissions of CO2. They are independent variables and they do not have any correlation or covariance with each other.

The result computed by WEKA for the Root Mean Squared Error (RMSE) of equal to 0.0012 %, which gives us a good performance of the prediction for the CO2 emissions since the value for RMSE < 0.5.

The result for the second algorithm Simple Linear Regression as follows (equation 3):

$$Linear\ regression\ on\ Per_Capita = 54502.11 * Per_Capita + 3993.7 \tag{3}$$

Table 5 presents the different performance measurements for simple linear regression

Table 5. The performance measurements for simple linear regression

Measurements	Value
Correlation coefficient	0.9956
Mean absolute error	3642.8606
Root mean squared error	462.0385
Relative absolute error	8.4397 %
Root relative squared erro	9.1855 %
Total Number of Instances	53

By comparing the results, it is clear that Linear Regression algorithm performance is better than the second algorithm Simple Linear Regression. Linear Regression algorithm gives us an amore accurate prediction for the CO2 emissions than the second algorithm based on the value of Root Mean Squared Error (RMSE). Table 6 gives a comparison between the different algorithms used.

Table 6. A comparison between the different regression algorithms based on RMSE%

Algorithm	Root Mean Squared Error %
SMOreg with parameter C = 1	114.9488
SMOreg with parameter C = 3	115.9269
Linear Regression algorithm	0.576
Simple Linear Regression	462.0385

Figure 1 shows CO2 emissions in Italy for the period from 1865 to 2014. The chart presents the contribution of each fuel source to the CO2 emissions in Italy. This chart provides a visual representation of annual CO2 emissions for the following energy recourses: Solid Fuel, Liquid Fuel, Gas Fuel, Cement, Per Capita, or Bunker fuels. Italy's energy mix has a significant impact on the low emission of carbon dioxide in the atmosphere and changes as a country moves to or from a specific energy source. [23].

In 1997, Italy ratified and committed to the Kyoto Protocol for GHG emissions. Italy followed legal policies and conventions to decrease greenhouse gas (GHG) emissions over the period 2008-2012. By adopting Kyoto Protocol, and a national climate change policy focused on climate change in accordance with the United Nations Framework Convention on Climate Change (UNFCCC), Italy reduced GHG level emissions by 6.5% below 1990 levels. However, Italian CO2 emissions have increased by 12.3% since 1990, due to the advanced use of fossil fuels in the energy sector (figure 2) [1]. This increase came from transport and power generation sector. In these sectors, it is not easy to apply policies for CO2 emissions.CO2 emission directions are affected by different energy resources and non-energy activities such as agricultural and waste sector.

In Italy, mobility demand is growing constantly. Italy is distinguished by having the largest number of vehicles per capita, and transportation mainly depends on private vehicles and trucks that consume fossil fuels, especially petroleum products. Energy generation reports show that the residential sector constitutes an additional energy consumption in Italy, especially from intensive use of air conditioning. To take advantage of modern technology, flexibility and complexity in manufacturing processes, the industrial sector needs radical changes in production patterns, which require additional energy. These new methods and technologies require higher electricity consumption, which leads to a strong increase in total final consumption and the need for additional energy generation [1]. These may explain the increase of CO2 emission between 1990 and 2004.

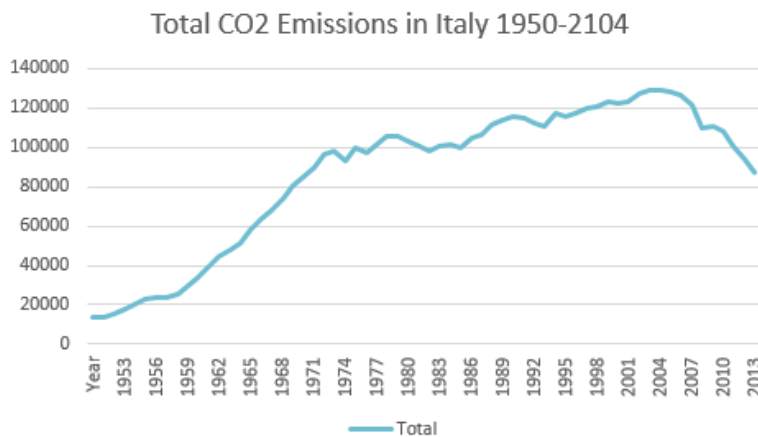


Figure 2. Total CO2 Emission in Italy 1865-2014

Furthermore, the chart shows the breakdown of annual CO₂ emissions in Italy and the change in CO₂ emissions in 2004 (figure 2). It is greatly affected by Italian energy mix, and changes as Italy government moves to or from a given source of energy.

With regard to air pollution and climate change, Italy is pressing for measures to reduce the emissions of black carbon, especially in the transport sector. Italy has co-funded approximately 190 initiatives with a cumulative cost of EUR 195 million through the Sustainable Mobility Fund. Italy has consistently reduced pollution in the last two decades [23, 24]. For example, in the transport sector, Italy adopted different policies and measures to reduce CO₂ emission as the promotion of hybrid vehicles or low or free emissions CO₂, new traffic rules and regulations, bikes, and motorcycle mobility [1].

According to the OECD report, Italy will look to integrate green initiatives into its economic policies [25, 26]. These initiatives may include “greening” the tax code, expanding environment-related markets and green trade policies, promoting eco-innovation and investing in green technology [24, 27].

Italy has supported higher targets for recycling and progress towards the full removal of EU landfills, green job development and tighter regulation of medium-sized combustion plant emissions [28].

Italy has made a big step toward promoting renewable energy technologies through its recent policies and financial incentives. Italy showed outstanding and remarkable evolution in the renewable energy sector. The green energy incentive scheme in Italy was highly successful and led to a 72 percent reduction in the cost of photovoltaic systems from 2008 to 2013 [24].

The COVID-19 pandemic has led to a major drop in air pollution (Figure 3). Satellite images shows that dense cluster of smoke from factories tends to be trapped against the Alps Mountains at the end of the Po Valley, making this one of the pollution hotspots of Western Europe. CO₂ emissions levels have decreased in Milan and other parts of northern Italy since the country lockdown on ninth of March 2020 [26, 29].

It is worth to mention here that, Italy is the first country to incorporate climate change and sustainable development into its national curriculum. Italian schools will spend almost one hour a week debating the problems of climate change. Therefore, Italy becomes first country to require climate change studies in schools [1].

Carbon dioxide emissions from the energy sector are severely affected by the level of energy consumption through energy supply, industry, manufacturing, transportation, higher education, residential, and families (households) and non-energy sector as agriculture [23]. Comparing the analysis of the actual data of CO₂ emissions in Italy with the predictive model of data mining using WEKA, the results of WEKA validate and prove the accuracy of the outcome of the prediction model.



Figure 3. The changes over northern Italy (Source Sentinel-5P, 1 Feb–17 Mar)

5. CONCLUSION

In this study, we used SMOREg data mining algorithm as a classification technique with different configurations for the Complexity parameter C , then measured the RMSE of each case to decide which one is better in the performance for the implementation. SMOREg used to predict the potential parameters that affect the percentage of CO₂ emissions in Italy.

We used two algorithms to find the prediction model. These algorithms are Linear Regression and Simple Linear Regression. The results showed that Linear Regression algorithm is the more accurate model for prediction based on the value of RMSE.

Based on historical data and the quantitative approach using WEKA, the study presented a qualitative analysis to discuss the trends in CO₂ emissions in Italy. The study shows the percentage distribution of CO₂ emissions for

energy resources is unchanged and CO₂ emissions are due to energy consumption. In terms of total emissions, the share of various energy resources remains roughly unchanged during the period 1951-2014 and so on. The liquid fuel is the largest contributor to national total GHG emissions 56.8%.

The identification of a consistent legislative framework that enhances policy collaboration at national and local levels is a crucial factor in assessing and determination of success of the effectiveness of potential efforts to reduce pollution.

This study focuses on carbon dioxide (CO₂) emissions in Italy, the most powerful greenhouse gas produce or come from burning fossil fuel, manufacturing, industry, transportation. Other emissions from agriculture, changes in land use, deforestation of agricultural land, and forestry change have a great impact too. It is clear that, CO₂ is not the only greenhouse gas that drives climate change globally.

In discussions on greenhouse gas (GHG) and climate change globally other recourses, such as methane (CH₄), nitrous oxide (N₂O), ozone (O₃), fluorinated gases, and trace gases have contributed greatly to climate change to date and maybe an interesting topic for future work.

Statement of Conflict of Interest: Author has declared no conflict of interest.

Author's Contributions: The contribution of the authors is equal

REFERENCES

- [1] Italian Report on Demonstrable Progress under Article 3.2 of The Kyoto Protocol (2018), United Nations Climate Change report, Feb. 2018. (61 pages). Access <https://unfccc.int/resource/docs/dpr/ital.pdf>
- [2] Proposal for a Council Decision concerning the approval, on behalf of the European Community (2002), of the Kyoto Protocol to the United Nations Framework Convention on Climate Change and the joint fulfilment of commitments thereunder /* COM/2001/0579 final - CNS 2001/0248 */, Official Journal 075 E , P. 0017 - 0032, March 2002 (10 pages).
- [3] K.I. Dervis (2007), Devastating for the world's poor: climate change threatens the development gains already, UN Chronicle, Volume: 44 Issue: 2 Page: 27, June 2007 (4 pages).
- [4] J. Blois, P. Zarnetske, M. Fitzpatrick and S. Finnegan (2013), "Climate Change and the Past, Present and Future of Biotic Interactions", Science 2ed issue, August 2013 (6 pages).
- [5] E. Massetti; P. Simona, Z. Davide, (2007). "National through to local climate policy in Italy". J. Integr. Environ. Sci., 4(3), 149–158. (11 pages)
- [6] J. Kwangbok; H. Taehoon; K. Jimin; L. Jaewook, (2020). A data-driven approach for establishing a CO₂ emission benchmark for a multi-family housing complex using data mining techniques. Renewable and Sustainable Energy Reviews, 110497–, 2020 (29 pages).
- [7] S. Jeslet and S Jeevanandham. (2015), CLIMATE CHANGE ANALYSIS USING DATA MINING TECHNIQUES, IJARSE, Vol. No.4, Special Issue (03), March 2015 (8 pages).
- [8] N. Somu, M. R. G.Raman, .&.K. Ramamritham (2021). A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 137, 110591, 2021 (21 pages).
- [9] CVV Farhate, ZMd Souza, O. SRdM, T. RLM, C. JLN (2018), Use of data mining techniques to classify soil CO₂ emission induced by crop management in sugarcane field. PLoS ONE 13(3): e0193537, 2018 (18 pages).
- [10] Report of the WHO–China Joint Mission on Coronavirus Disease 2019 (COVID-19) (2020), (40 pages) [https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))
- [11] .P. Friedlingstein, W. J. Matthew, O'S. Michael, M. A. Robbie, et al. (2019), Global Carbon Budget 2019 (2019). Earth Syst. Sci. Data 11, 1783–1838, 2019 (15 pages).
- [12] G.P. Peters, R.M. Andrew, J.G. Canadell, P. Friedlingstein, R. B. Jackson, J. I. Korsbakken, C. Le Quéré & A. Peregon (2020), Carbon dioxide emissions continue to grow amidst slowly emerging climate policies. Nat. Clim. Change 10, 3–6 (3 pages).
- [13] C. Le Quéré, R. Jackson, M. Jones, A. Smith, S. Abernethy, R. Andrew, A. De-Gol, Y. Shan, J. Canadell, , P. Friedlingstein, F. Creutzig, & G. Peters (2020). Supplementary data to: Temporary reduction in daily global CO₂ emissions during the COVID-19 forced confinement (Version 1.2). Global Carbon Project (8 pages). <https://doi.org/10.18160/RQDW-BTJU>
- [14] H. Hassani, X. Huang, and E. Silva (2019), Big Data and Climate Change, Big Data Cogn. Comput., 3, 12, 2019 (17 pages).
- [15] S. Deniz, H. Gokcen, and G. Nakhaeizadeh (2016), Application of Data Mining Methods For Analyzing of The Fuel Consumption And Emission Levels, International Journal Of Engineering Sciences & Research Technology (IJESRT), 5(10): October 2016 (13 pages).
- [16] B. Zhu, Y. Wei (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support

- vector machines methodology. *Omega*, 41(3), 517–524, 2013 (8 pages)..
- [17] S. Dai, D. Niu, & Y. Han (2018). Forecasting of energy-related CO2 emissions in China based on GM (1, 1) and least squares support vector machine optimized by modified shuffled frog leaping algorithm for sustainability. *Sustainability*, 10(4), 958, 2018 (17 pages).
- [18] F. Ye, X. Xie, L. Zhang, & X. Hu, (2018), “An improved grey model and scenario analysis for carbon intensity forecasting in the Pearl River delta region of china”, *Energies*, 11(1), 91, 2018 (17 pages).
- [19] Z. Zhang, and J. Li (2020), “Big Data mining For Climate Change”, Elsevier, 2020 (344 pages).
- [20] V. Lakshmanan, E. Gilleland, A. McGovern, & M. Tingley (2015), *Machine Learning and Data Mining Approaches to Climate Science*, Proceedings of the 4th International Workshop on Climate Informatics. Springer, 2015 (243 pages).
- [21] J. Brownlee (2016), "Machine Learning Mastery With Weka Analyze Data, Develop Models and Work Through Projects", 2016 Jason Brownlee, 2016 (248 pages).
- [22] Y. Abuzir, & S. Abuzir (2020), Data Mining Techniques for Prediction of Concrete Compressive Strength (CCS), *Palestinian Journal of Technology and Applied Sciences (PJTAS)*, No 3, 2020 (16 pages).
- [23] S. Agarwal, B. Bouille, Y. Briand et al. (2017), *Brown to Green: The G20 Transition to A Low-Carbon Economy, Italy Country Facts 2017, Climate Transparency 2017* (36 pages). <http://www.climate-transparency.org/g20-climate-performance/g20report2017>
- [24] S. Brett (2015), *How Clean is Your Country?*, Italy: Environmental Issues, Policies and Clean Technology (azocleantech.com), Jun 11 2015 (4 pages). <https://www.azocleantech.com/article.aspx?ArticleID=536>
- [25] M. Montini, *Italian Policies and Measures to Respond to Climate Change* (February 2000). SSRN Electronic Journal, 2000 (17 pages). Available at SSRN: <https://ssrn.com/abstract=235084>
- [26] J. Watts, and N. Kommenda (2020), *Coronavirus pandemic leading to huge drop in air pollution*, The Guardian, march 2020 (4 pages). <https://www.theguardian.com/environment/2020/mar/23/coronavirus-pandemic-leading-to-huge-drop-in-air-pollution>.
- [27] E. Massetti, P. Simona, D. Zanoni (2007). National through to local climate policy in Italy. *Environmental Sciences*, 4(3), 149–158, 2007 (9 pages).
- [28] *Italy's Action Plan On CO2 Emissions Reduction(2012)* , Ente Nazionale L'Aviazione Civile (ENAC), June, 2012. (73 pages), Access https://www.icao.int/environmental-protection/Documents/ActionPlan/Italy_AP_En.pdf
- [29] B. Rugani and D. Caro, (2020). Impact of COVID-19 outbreak measures of lockdown on the Italian Carbon Footprint. *Sci. Total Environ*, 737(), 1 October 2020, 139806 (14 pages).