ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/319166096

An Approach for Predicting CO2 Emissions using Data Mining Techniques

Article *in* International Journal of Computer Applications · August 2017

DOI: 10.5120/ijca2017915098

CITATIONS
4

READS
4,324

2 authors:

Douglas Kunda
Mulungushi University
44 PUBLICATIONS **429** CITATIONS

SEE PROFILE

Hazael Phiri
The Copperbelt University
5 PUBLICATIONS **31** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Design and development of a Grain Tracking and Warehouse Motion Detection Model Based on Internet of Things for Food Reserve Agency View project

STACE Framework View project

All content following this page was uploaded by Hazael Phiri on 12 November 2018.

The user has requested enhancement of the downloaded file.

An Approach for Predicting CO_2 Emissions using Data Mining Techniques

Douglas Kunda
Mulungushi University, School of Science, Engineering and Technology
P.O Box 80415, Kabwe, Zambia.

Hazael Phiri
Mulungushi University, School of Science, Engineering and Technology
P.O Box 80415, Kabwe, Zambia.

ABSTRACT
The effects of climate change such as drought, storms and extreme weather are affecting the earth. Greenhouse gases are the main cause of climate change and carbon dioxide (CO2) takes the larger proportion of the total greenhouse gases emitted. CO2 is projected to continuously increase based on simulation and data mining tools relying on historical data. Globally, 80% of CO2 emission is from combustion of fossil fuel mainly in manufacturing industry or transportation industry. Governments in both developed and developing countries have formulated policies to manage CO2 emission by orienting them towards consumers or manufacturers. In Zambia, CO2 emission control has mostly focused on transportation industry where carbon emission tax is charged on motor vehicles depending on combustion capacity. Transportation is an industry that has had a high increase of total CO2 emission currently standing at 35% but with average of 31.7 % from 1971 to 2014 in Zambia. The manufacturing sector, though being the highest emitter of CO2 has seen no policies formulated to regulate emissions. Predicated emission values for both transportation and manufacturing sector show a continuous domination of these two sectors regarding carbon emission. The task is then left with policy makers to introduce policies that will regulate emissions as the current carbon tax policy does not seem to be effective in reducing emissions. This paper brings out the trends in CO2 emission from fossil fuels in Zambia from 1964 to 2016. The paper also highlights the main industries contributing to CO2 emission, the policies implemented to control CO2 emission levels globally and in Zambia, provides a forecast for CO2 emission levels till 2021 and suggests future research directions based on the findings.

Keywords
Climate, Policy, Emission, Data Mining, WEKA, SMOreg, Carbon Tax

1. INTRODUCTION
We Climate change has become a prominent topic in modern human civilization as it poses a threat to future generations of human kind [6]. Researchers from different parts of the globe have conducted research concerning climate change and some of the effects predicted are being experienced such as extreme weather conditions [8]. Droughts, storms and extreme heat are leading headlines on news platforms. These conditions call upon researchers and policymakers to provide guidelines of sustaining the planet for the benefit of future generations. For solutions providing sustainability to be effective, the main factors contributing to climate change must be identified. Research has shown that the main contributing factors to climate change are greenhouse gases and carbon dioxide (CO2) is the main greenhouse gas contributing to climate change [4]. The Intergovernmental Panel on Climate Change IPCC has projected a continuous rise in CO2 emissions up to the year 2050. Further research, provides evidence that 80% of total CO2 emissions globally come from fossil fuel combustion [15]. The sectors of society were fossil fuels are used can be analysed to monitor trends among the various sectors and their total contribution to CO2 emissions. Effectiveness of existing policies in reducing the carbon footprint can be evaluated and new policies can be suggested.

To support formulation of policies regarding carbon emission data mining tools can be used as climate data is readily available in the public domain [11]. One of the data mining tools available is the Waikato Environment for Knowledge Analysis (WEKA). WEKA is a free open source software. WEKA is a collection of data processing tools and machine learning algorithms that can be used on data sets to facilitate a flexible knowledge discovery [17]. It has been in existence since 2000 and was developed by the University of Waikato in New Zealand [18]. WEKA supports the whole process of experimental data mining from data preparation, input, statistical analysis to visualizations. The simple graphical user interface make it easy to carry out exploratory data analysis.

The climate data available for analysis is mostly time series data which is useful for pattern analysis [13]. Time series analysis is performed on historical values of variables to determine the pattern in a historical data time lines and use the trend established for forecasting future conditions [19]. Sequential minimal optimization regression (SMOreg) is one of the forecasting algorithms that can be used in time series analysis and is based on based on support vector machine for the solution of regression problem [19]. SMOreg was used in analysis as it can support both linear and nonlinear regression models.

This paper begins by providing an overview of previous research on CO2 trends and projections for different parts of the world. The paper then looks at the policy provision in Zambia regarding carbon emissions. The paper further provides a time series analyses of CO2 emission trends from 1964 to 2016 and makes forecast to 2021 using WEKA as a data mining tool. The aim of the analysis was to find out what has been the trend in percentage contribution to carbon emission from different sectors based on total fossil fuel combustion annually. The second question was to find out which sector had the highest contribution of emissions and lastly make predictions for the next five years regarding the rate of carbon emission for each sector. The paper concludes with suggestions of policies that can be implemented to manage carbon emissions and future research directions.

2. CO2 EMISSIONS, POLICIES AND DATA MINING AT GLOBAL LEVEL
Combating the effects of climate change has been a main concern for countries at global level by applying various methods. In Europe, one of the main sectors contributing to CO2 emission was transport with 28% in 2008 [5]. Earlier studies showed that motor vehicle purchases were determined

by retail prices and not taxes incurred by owners over the lifespan of the vehicle but the research conducted by [5] proved otherwise. The research showed that circulation costs and costs on fuel determined the type of car with the large market share. Countries with motor vehicle manufacturing industries had strict standards set by the EU, targeting at reducing CO2 emissions on manufactured vehicles. Other countries that were non car manufacturing oriented their policies towards the consumer side. In Ireland, vehicle registration tax and annual motor tax were based on relative measured CO2 emission in g/km for a vehicle and not by engine size, as of July 2008 [10]. The higher the emission of the vehicle, the higher the tax rate charged. According to ex-post analysing conducted using data mining, this policy decreased the CO2 emissions by 8.6% in the first year of its introduction [10] concurring with [5] that car market share was based on taxes.

In Asia, China is projected to have the highest number of vehicles on the road by 2030 surpassing the United States of America as predicted using patterns of motor vehicle growth in Asia and Europe [12]. The estimated of CO2 is expected to be between 1.2 to 3.2 billion tonnes per annum by 2050. Policies on CO2 emission in China are targeted towards improving vehicle fuel economy.

In the United States of America, 25% to 30% of greenhouse gas emissions is from the transport sector and it has not fully been addressed [4] [11]. Researchers have suggested that with accurate measurements of CO2 emission mining traffic patterns, vehicle performance and driving behaviour can help to design solutions such as fleet maintenance policies, traffic signals and speed limits [4] [7]. A commercially available data mining tool called Minefleet has the ability to provide such types of analytics [11]. For the manufactures data mining can help designing low CO2 emission cars.

3. CO2 EMISSION POLICIES IN ZAMBIA

The need for regulating CO2 has been recognized as being important. In developing countries, CO2 emission is linked to Gross Domestic Product GDP. A rise in GDP leads to a rise in CO2 emission [15]. Most developing countries including Zambia have maintained a steady rise in GDP. The level of CO2 has been on the rise prompting government to implement carbon emission tax on motor vehicles based on engine combustion capacity [1]. The tax was introduced in 2006 to be collected at border points for imported vehicles but there was no follow up by Zambia Revenue Authority to collect the tax on an annual basis. Since 2010, the carbon emission tax is collected by Road Transport and Safety Agency on behalf over Zambia Revenue Authority when motor vehicle owners renew their road tax. The tax is viewed as a penalty for carbon emissions but the projects for which the funds realized are used, have not been clearly stated [16]. Other measures to control carbon emission have been through afforestation and practice of conservation farming [9].

4. DATA COLLECTION, PREPARATION AND WEKA TOOL

Data was obtained from http://data.worldbank.org/ country/zambia for transport, manufacturing and construction, electricity and heat production, residential, commercial and public services sectors for the period 1960 to 2016. The period 1960 to 1970 and 2015 to 2016 did not have any estimated values. The data collected was time series data as it is readily available and brings out interesting patterns or changing trends in evolution of objects under study [13]. Since data mining tools are required to work on clean and consistent data the raw data collected had to be formatted in a form compatible with WEKA[2].

```
@relation ZambiaCo2Emission

@attribute Year date "yyyy"
@attribute Transport numeric
@attribute other numeric
@attribute Manufacturing_and_Construction  numeric
@attribute Electricity_and_heat_production numeric
@attribute Residential_commercial_and_public_services numeric

@data
1960,?,?,?,?,?
1961,?,?,?,?,?
1962,?,?,?,?,?
1963,?,?,?,?,?
1964,?,?,?,?,?
1965,?,?,?,?,?
1966,?,?,?,?,?
1967,?,?,?,?,?
1968,?,?,?,?,?
1969,?,?,?,?,?
1970,?,?,?,?,?
1971,18.69436202,0.59347181,62.61127596,10.38575668,7.418397626
1972,18.25396825,0.793650794,64.81481481,8.994708995,7.142857143
1973,17.67554479,0.726392252,64.64891041,9.685230024,7.263922518
```

Figure 1: Data Sample for Analysis

5. ANALYSIS OF RESULTS AND INTERPRETATIONS

Data mining was used to analyse and forecast the data as simple data mining approaches can bring new scientific insights [14]. Trend analysis using SMOreg algorithm for time series was used for analysing the data in WEKA.

From the results, average percentage contribution of CO2 emission was 50.9% for manufacturing and construction, 31.7% for transport, 6.7% for electricity and heat production, 7.1% for residential, commercial and public services and 3.4% for other sectors in Zambia between 1971 -2014. The transport sector has had the highest rate of increase in CO2 emission starting from 1979 while manufacturing and construction has been decreasing. In the year 2000 the transport sector surpassed manufacturing and construction as the leading Contributor to CO2 emission and reached a peak of 51.5% in 2007. There was a sharp decline from 2007 to 2008 probably due to the recession and that can be considered to be a temporal trend [3]. The introduction of carbon emission tax has not drastically reduced the rate of emission. The manufacturing sector was leading in emission at 65% but has dropped to 45% to 50%. The other remaining sectors have been fluctuating below 10%.

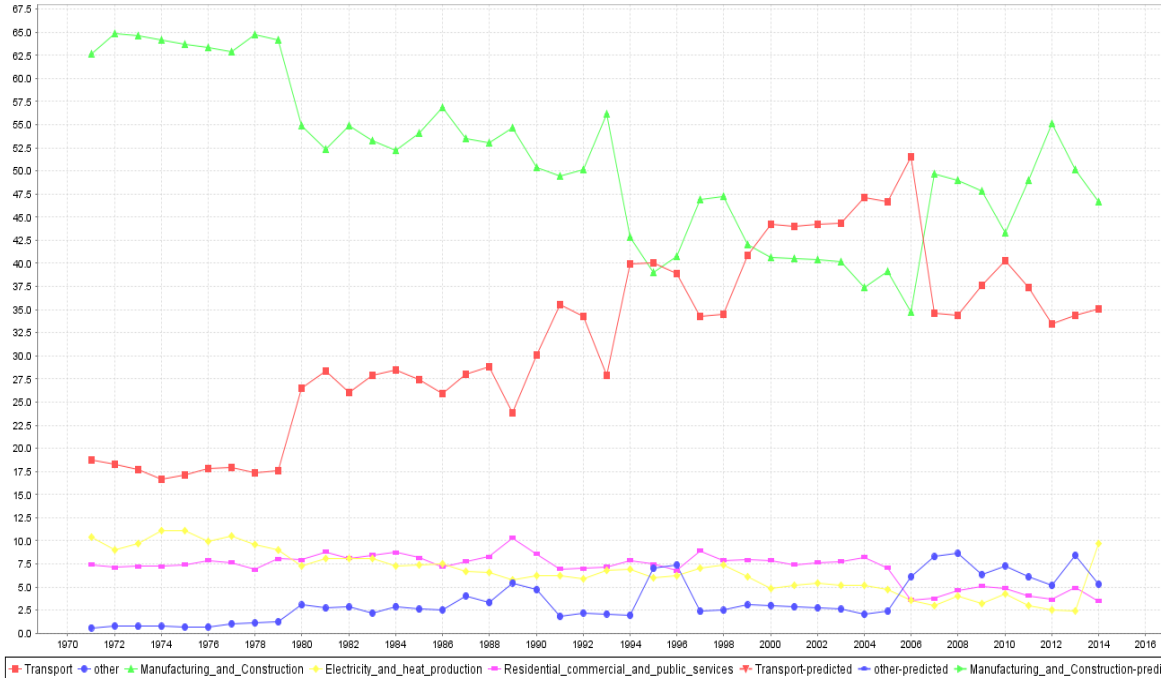

Figure 2: Zambia CO2 emission Trends 1971 to 2014

The results validate that there was a need to introduce carbon emission tax to control emissions in the transport sector but it has not controlled the emissions. Carbon tax revenue for 2011 to 2013 was 17.1, 22.1 and 26.1 million Kwacha but it has not been stated clearly were it was used [1]. Research also shows that the average age of a car on Zambia road has increased from 13 years in 2006 to 17 years in 2014 [16]. This is an indication that there could be an increase of used cars imported into the country there by increasing the age. This increase in age implies high CO2 emission levels and there is no existing policy regarding end life of motor vehicles.

Manufacturing and production has not received any policies regarding the control of carbon emissions though it remains as the other highest contributor to CO2 emission apart from the transport sector.

Regarding the forecast from 2017 to 2021, the prediction is an upward increase in proportion of carbon emission contribution in transport and manufacturing and construction sector. Residential, commercial and other sectors will have a slight decline but will remain fairly constant over the same period. Electricity and heat production will continue to decline at a slow rate.

Based on the results and forecasts, there is need to relook at policies for controlling carbon emission in transport and manufacturing sectors. Looking for alternative forms of transport such as trains, high speed buses and bicycles for short distances instead of individual cars might help in reducing CO2 emission levels. Implementing policies that ensure that all vehicles moving on high ways have maximum number of passengers allowed instead of individuals could be another alternative. Zambia can also consider restricting age for vehicles imported and charge carbon emission tax based on actual measured emission rate per vehicle.

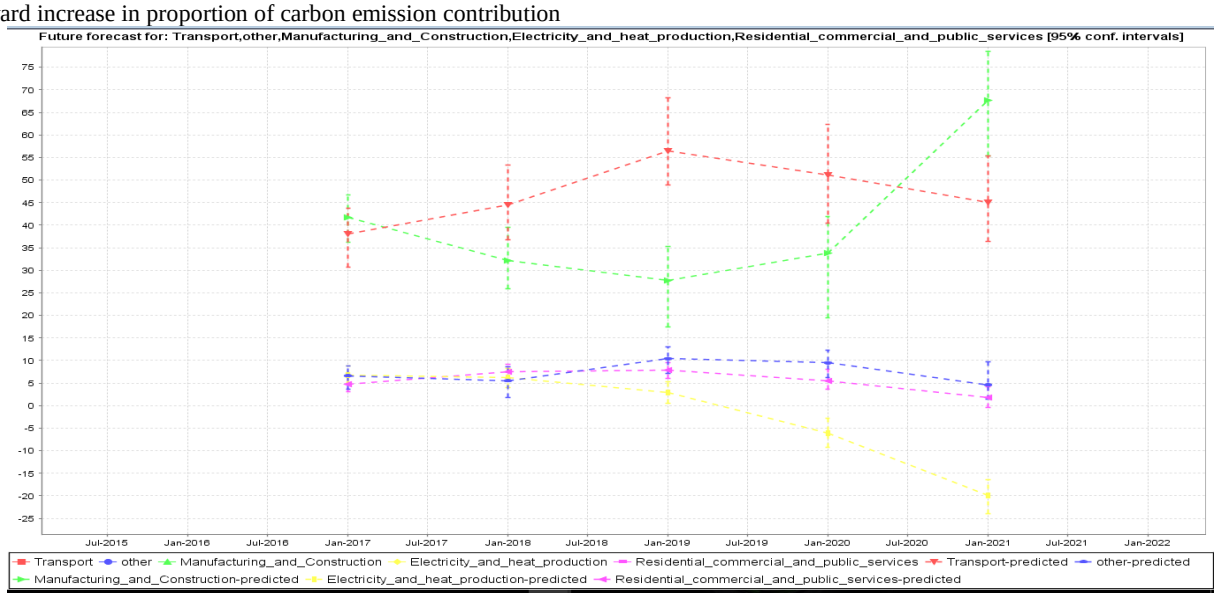

Figure 3: Zambia C02 Emission Forecast 2017 to 2021

6. CONCLUSION AND FUTURE

Climate change has mainly been affected by CO2 emissions. Analysing historic data using data mining can help to make future forecasts and formulate policies that can be used to manage carbon emissions. Across the globe, carbon emission policies have been targeted either towards the manufactures side or consumer's side. In Zambia, carbon emission has been increasing with the transport sector as the main contributor to the increase from 1971 to 2014. The forecast for the future predicts a continuous increase with transport and manufacturing. The increasing in transport is due to average age increase in motor vehicle age.

There is a need to formulate new policies to manage carbon emissions in Zambia. For future research, the work presented in this paper can be extended by breaking down the transport sector into various categories such as goods vehicles, public transport, personal vehicles and others to determine which category is responsible for high levels of CO2 emission. Another alternative for future research could be to determine the source of high CO2 by province or by type of road i.e highways or local urban and rural roads. Conducting such studies could help in formulating policies that are more specific than generic.

7. REFERENCES

[1] Nhekairo, W. M. "The Taxation System in Zambia, Jesuit Centre for Theological Reflections, Final Report".

[2] Han, J. 1997. "OLAP Mining: An Integration of OLAP with Data Mining,".

[3] Kriegel, H.-P. and Borgwardt, K. M. 2007. "Future Trends in Data Mining," Data Mining Knowledge Discovery, Springer.

[4] Kargupta, H., Gama, J,. and Fan, W. 2010. "The Next Generation of Transportation Systems,," ACM.

[5] Vance, C., and Mehlin, M. 2009. "Tax Policy and CO2 Emissions – An Econometric," Ruhr economic paper.

[6] Faghmous, J. H., and Kumar, V. 2014. "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science," Mary Ann Liberty, Inc,, vol. 3, p. 2.

[7] Chen, V., Han, J., and Yu, P. S. 1996. IEEE, Trandactions on Knowledge and data Engineering,, vol. 8, no. 8,.

[8] Savage, L. 2006. "An Oveview of Climate Change and Possible Consequences for Gisbone District".

[9] Stringer, L. C., Dougill, A. J., Mkwambisi, D. D., Dyer, J. C., and Kalaba, F. K. 2012. "Challenges and opportunities for carbon management in," Carbon Management, pp. 159-173.

[10] Rogan, F., Dennehy, D., Daly, H., Howley, M., and Gallachóir, B. P. Ó. 2011. "Impacts of an emission based private car taxation policy – First year," Transportation Research Part A, Elsevier.

[11] Morik, K., Bhaduri, K and Kargupta, H. 2012. "Introduction to data mining for sustainability," Data Mining and Knowledge Discovery, Springer,.

[12] Huo, H., Wang, M., Johnson, L., and He, D. "Projection of Chinese Motor Vehicle".

[13] Fu, Y. 1998. "Data Mining: Tasks, Techniques and Applications".

[14] Ganguly, A. R. and Steinhaeuser, K. 2008. "Data Mining for Climate Change and Impacts," in IEEE International Conference on Data Mining Workshops.

[15] Schmalensee, R., Stoker, T. M., and Judson, R. A. 1997. "World Carbon Dioxide Emissions 1950-2050".

[16] Kriegel, H.-P and Borgwardt, K. M. 2007. "Future Trends in Data Mining," Data Mining Knowledge Discovery, Springer.

[17] Witten, I. H., Eibe, F., Hall, M., Holmes, G., Kirkby, R and Pfahringe, B. "WEKA A Machine Learning Workbench for Data Mining," Hamilton.

[18] Witten, I. H., Eibe, F., Hall, M., Holmes, G., Reutemann, P and B. Pfahringe, "The WEKA Data Mining Software: An Update," SIGKDD Explorations, vol.11, no. 2

[19] Munir, A. Q and Winarko, E. 2015. "Classification Model Disease Risk Areas Endemicity Dengue Fever Outbreak based Prediction of Patients, Death, IR and CFR using Forecasting Techniques," International Journal of Computer Applications, vol. 114, no. 2, pp. 20-25.

View publication stats