TEACHING TO SOLVE GLOBAL WARMING WITH DATA SCIENCE

Cédric Mesnage
Solent University
Southampton, United Kingdom

Abstract

We report on the design and execution of a Data Science assignment in which students work on Global Warming issues and predict the outcome of solutions. The paper outlines how to solve Global Warming through the design of an assignment and gives elements of answers to research questions regarding research in the classroom, the benefits to students and how to influence policy makers.

Introduction

In this paper we report on conducting Data Science research in the classroom with two groups of Bachelor students. We are interested in knowing:

1. RQ1: Whether it is possible to do Data Science research in the classroom and how;
2. RQ2: how it benefits the students; and
3. RQ3: if the discoveries made are valuable.

To answer to RQ1, we describe our methodology in setting up the class, the various activities undertaken with the students and the artefacts produced. We set ourselves to research and solve Global Warming. RQ2 is answered by carrying out a survey with the students, focusing on how they benefited from the class, what could be improved and what they learned. To evaluate RQ3 we share the results of the students' research and their findings about global average temperature, gas emissions and potential solutions. We discuss how to improve this process, how to involve the students better, and future directions.

Solving Global Warming

Global warming is a major environmental issue of our time. The gas emissions due to industrial changes and the increase of population lead to a rise of the global average temperature and sea level as the ice melts. This results in dramatic disasters, floods, earthquakes, storms and species extinctions. Recently it has been shown the largest king penguins' colony has collapsed (Weimerskirch, Le Bouard, Ryan, & Bost, 2018), its population falling from 2 million individuals to 90,000. Another alarming fact is the change in the ocean microbiotic life; Morán et al. (2015) have shown the bacteria are smaller and there are more of them as a result of global warming. What is the effect of this on the rest of ocean life? The reader will certainly find many more examples of such disasters. So, what do we do about this? Even though the situation is dramatic, there are solutions, e.g., limiting the gas emissions has been shown to enable stopping the temperature rise (Smith et al., 2019) while keeping it

under the 1.5 degrees boundary. Planting trees would get the temperature back to a normal level: there are three trillion trees on the planet according to Crowther et al. (2015) and another 1.2 trillion would be necessary to absorb the excess of CO_2 in the atmosphere based on Goymer (2018). Identifying major industries and activities which contribute to gas emissions and convincing people to find alternatives is essential, as are reforestation efforts and planting trees in arid areas.

Data Science

Data Science is a recent research field of Computer Science evolving from Statistics, Data Mining, Data Visualisation and Machine Learning. Since much larger amounts of information are available, the diversity of the analyses which can be performed is vast. Every industrial sector needs data scientists to work on their data flow; therefore, the profession is growing quickly. The typical Data Science process consists of collecting data, cleaning and pre-processing the data, applying data mining or machine learning algorithms, performing analysis and producing visualisations to communicate the knowledge extracted. In previous occurrences of a Data Science class, students worked on the musicbrainz database for their assignment (Mesnage, 2017) and in a social software class (Mesnage & Jazayeri, 2008). How can Data Science help in solving Global Warming?

Design of a Data Science Assignment

The assignment proposed is composed of 11 tasks ranging from researching the topic of Climate Change, to finding relevant datasets, plotting data, making predictions and predicting the outcomes of solutions. The goal is to convince policy makers on changing behaviours and taking action. Figure 1 provides the text of this assignment.

Global warming is a problem clearly identified, the result of human industrialisation which recently led to disastrous gas emissions in the atmosphere. The climate is getting warmer every year due to the lack of ozone in the stratosphere and the presence of greenhouse gas. In this assignment you will work on climate change data to gain insights on global warming and motivate solutions.

Task 1. Research the topic of global warming and outline ideas on how Data Science can help. List potential solutions and how to influence policy makers.

Task 2. Based on your knowledge of Next Generation Databases discuss the advantages of the different NoSQL database categories to store climate change data.

Task 3. Find a relevant dataset to climate change (for instance global sea level or average global temperature, effect of planting trees/deforestation, gas emissions of cars, planes, ships, veganism...) on https://toolbox.google.com/datasetsearch and write a Python script to load the data and import it in a MongoDB database.

Task 4. Write 3 JavaScript queries to access your data from your MongoDB database.

Task 5. Based on your knowledge of Data Science and your readings, discuss the different data mining methods that can be applied to climate change data and what insights they can give you.

Task 6. Use https://toolbox.google.com/datasetsearch to find a relevant dataset, open it with Orange and plot the data with a Scatterplot.

Task 7. Recall how the Kmeans algorithm functions and the purpose of clustering. Apply the Kmeans algorithm to cluster data. You might need to process the data or link it with another dataset. For instance, you could cluster UK cities based on their gas emissions levels.

Task 8. Apply a classification tree and a classification tree viewer to the output of Kmeans. Produce the screenshot of the tree and interpret the results.

Task 9. Explain the APRIORI algorithm to compute frequent itemsets. What is the complexity of the problem of finding frequent itemsets and what is the technique used to improve its efficiency? If applicable, find association rules on your data or find another relevant dataset; for instance, you could find patterns of pollution levels and economic activities in regions.

Task 10. Use linear regression, polynomial regression and neural networks to predict the global average temperature in the coming 30 years; compare and discuss the results.

Task 11. Research prediction models of the global average temperature using gas emissions and predict the 5 year outcomes of solutions to global warming. Use the tool colab (https://colab.research.google.com) and your knowledge of Python for data visualisation to conduct this task.

Figure 1. Assignment 11 tasks.

The assignment was completed by 43 students to various levels of success. One student had an interesting take on task 11, which we will look at in the next section.

Elements of Answers

In this section we answer our research questions, by looking at the student work for some of the assignment tasks, the student feedback, success rate and discussions with policy makers.

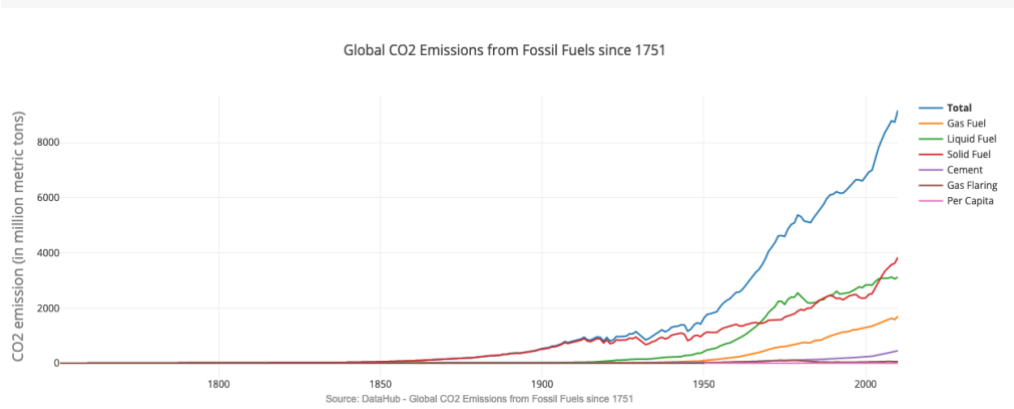

Figure 2. Line chart of global emissions.

Figure 2 shows the work of a student in building a visualisation showing the datasets of provenance of CO_2 emissions.

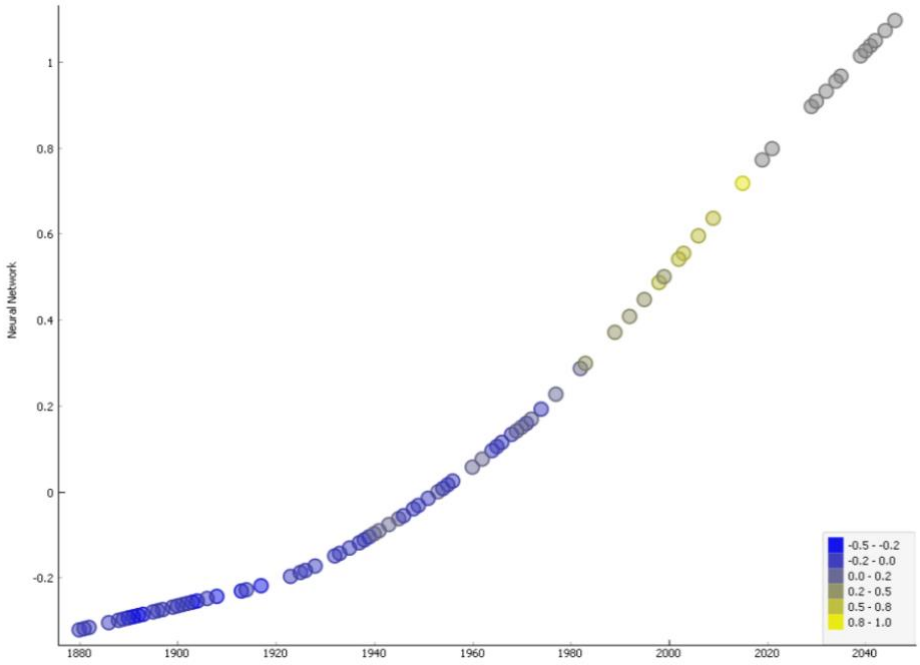

Figure 3. 30 year prediction of the global average temperature.

The neural network model used in Figure 3 to answer question 10 suggests a temperature change of 1.524 degrees celsius until 2046, which is consistent with climatologists' predictions and alarming.

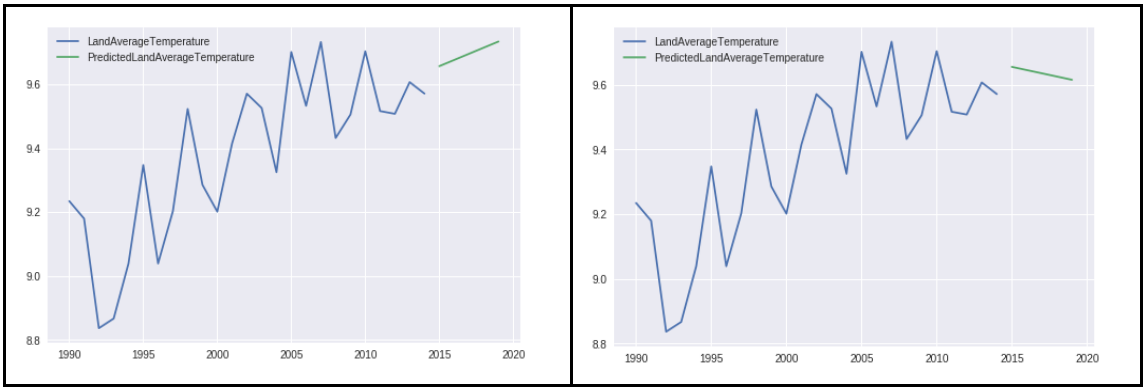

Figure 4. Predicted global average land temperature without (left graph) and with (right graph) increasing the number of planted trees.

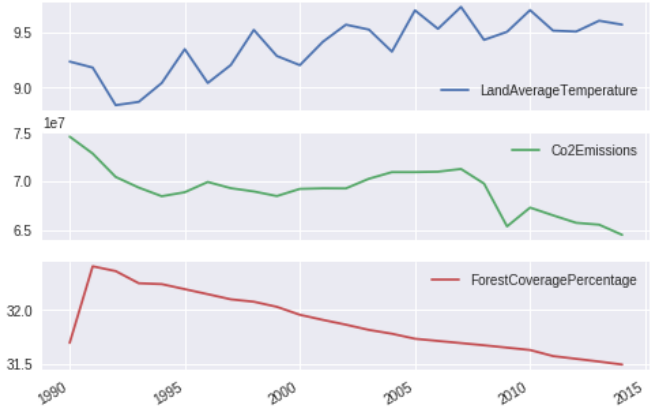

Figure 5. Datasets used for the prediction of the solution.

Figures 4 and 5 depict the work of a student in predicting the outcome of increasing the forest coverage of Earth, answering task 11. They show a potential regression of the global average land temperature in the case of an increase of the number of trees, which is an optimistic look on the problem and corroborates the work of Goymer (2018). By combining the land average temperature, the CO_2 emissions and the forest coverage, he trained a neural network model on the dataset shown in Figure 5 and used it to make the predictions in Figure 4, either with a continuing decrease of forest coverage or an increase of forest coverage.

This clearly answers our RQ1, as students have shown it is possible to do data science research in the classroom with the fantastic work they produced. As Little, Brookes, & Palmer (2008) discuss, research informs teaching, as well as teaching informs research.

Student Feedback

Ninety-seven and a half percent (97.5%) of students passed the class, which shows a great motivation to work on the assignment.

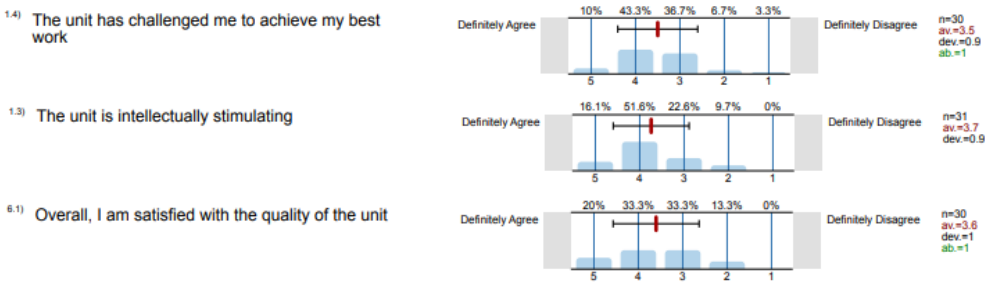

Figure 6. Student survey results.

Figure 6 shows some answers to the unit survey on student satisfaction, which are the questions most relevant to RQ2. Student found the assessment challenged them to achieve their best work and found the unit intellectually stimulating. These are good elements of answers to RQ2 on how the assessment benefitted the students. They were also overall satisfied with the unit and found the staff were good at explaining things. This does not account for another outcome: students learned through their research that it is possible to solve global warming. This is certainly the best outcome.

Sharing with policy makers

In terms of sharing with policy makers, I recently have been in touch with members of UK parliament and discussed with them the solution of planting trees and ways to reduce gas emissions. For example, cruise ships account for a large proportion of emissions for an industry which is far from being essential, and a large cruise ship pollutes as much as millions of cars (Vidal, 2016). Cruise ships could be replaced by other activities such as sailing with the wind. Planting a trillion trees is a huge human effort which we must undertake, reforesting forests but also planting in desert areas to use the space. There were once 6 trillion trees on Earth, we can certainly achieve planting a trillion. This article will be shared with policy makers and starts answering RQ3 and so far their reception is positive; they are willing to work on solutions to Climate Change.

Conclusion

In this paper we have shown it is possible to do research in class (RQ1) even on difficult topics as Data Science and Climate Change. We have proposed an assignment and given examples of students' answers. We looked at the unit survey which exhibits that students were challenged by the assignment and the success rate of 97.5% shows the interest in the topic (RQ2). We discussed next steps in informing policy makers (RQ3). This paper raises awareness on Global Warming issues and motivates the solution of planting trees in large numbers and reduce gas emissions. Initiatives such as the great green wall (Great Green Wall, n.d.) and the trillion trees projects (Trillion Trees, n.d.) encourage us to contribute.

Acknowledgements

I would like to particularly acknowledge the 45 students who took part in this year's Data Science class.

References

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., ... & Tuanmu, M. N. (2015). Mapping tree density at a global scale. *Nature*, *525*(7568), 201.

Goymer, P. (2018). A trillion trees. *Nature ecology & evolution*, *2*(2), 208.

Great Green Wall. (n.d.). The great green wall: Growing a world wonder. Retrieved from https://www.greatgreenwall.org/about-great-green-wall

Little, V., Brookes, R., & Palmer, R. (2008). "Informed teaching and teaching-informed research: The Contemporary Marketing Practices (CMP) living case study approach to understanding marketing practice." *Journal of Business & Industrial Marketing 23*.2: 124-134.

Mesnage, C. (2017). Using the Musicbrainz database in the classroom. In L. Morris, L. & C. Tsolakidis, (eds.), *The Proceedings of the International Conference on Information Communication Technologies in Education (ICICTE) 2017 (pp.* 338-344). Rhodos, Greece: ICICTE.

Mesnage, C. & Jazayeri, M. (2008). "Social thinking to design social software: A course experience report.", *ASE Workshops 2008. In 2008 23rd IEEE/ACM International Conference on Automated Software Engineering-Workshops* (pp. 19-24). IEEE.

Morán, X. A. G., Alonso-Sáez, L., Nogueira, E., Ducklow, H. W., González, N., López-Urrutia, Á. & Huete-Stauffer, T. M. (2015). More, smaller bacteria in response to ocean's warming? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1810), 20150371.

Smith, C. J., Forster, P. M., Allen, M., Fuglestvedt, J., Millar, R. J., Rogelj, J., & Zickfeld, K. (2019). Current fossil fuel infrastructure does not yet commit us to 1.5 °C warming. *Nature Communications 10*(1), 101.

Trillion Trees. (n.d.). A shared vision for more trees on our planet. Retrieved from https://www.trilliontrees.org/

Vidal, J. (2016). The world's largest cruise ship and its supersized pollution problem. *The Guardian*, *21*. Retrieved from https://www.theguardian.com/environment/2016/may/21/the-worlds-largest-cruise-ship-and-its-supersized-pollution-problem

Weimerskirch, H., Le Bouard, F., Ryan, P., & Bost, C. (2018). Massive decline of the world's largest king penguin colony at Ile aux Cochons, Crozet. *Antarctic Science*, *30*(4), 236-242. doi:10.1017/S0954102018000226

Author Details

Dr. Cédric Mesnage
cedric.mesnage@solent.ac.uk