Visualizing and Predicting CO_2 Emissions

Neha Gupta[1], Priscilla Lee[1] Sanjana Vasireddy[1], and Shyam Krishnan Ondanat Veetil[1]

[1]University of Southern California, Los Angeles CA 90007, USA

Abstract. CO_2 has been considered a major contributor to the increasing the global warming in Earth. According to the reports from NASA various human activities have contributed to around 50% increase in the amount of CO_2. In this project, we systematically analyze CO_2 output for a selected set of countries. Are there any relation between the economic indicators and Carbon dioxide emissions? The visualizations included would make it easier to derive conclusions and see correlations between various features and their impact in CO_2 output. The dashboard developed can be utilized by anyone who is interested in understanding how various features effect the CO2 output of countries

Keywords: CO2 Emissions · Gross National Income(GNI) · Population growth · Energy Expenditure.

1 Introduction

Carbon Dioxide (CO_2) emissions have adversely affected the environment, contributing to global warming, which is a gradual increase in the overall temperature of the Earth's atmosphere. One large factor contributing to global climate change is the increased levels of atmospheric carbon dioxide which is produced by the use of fossil fuels, such as exhaust from cars and trucks, burning trash, and factory emissions. According to the World Economic Forum's Global Risks Report 2021, the failure to mitigate and adapt to climate change is "the most impactful" risk facing communities worldwide.

In efforts to identify and highlight trends in CO_2 emissions over the years for selected countries across several continents, we developed a dashboard to visualize and predict this paradigm further. Though there may be other dashboards and infographics on the same topic, we have highlighted correlations for outcomes of interest and made the charts interactive and intuitive. Our visualizations focus on the features selected by our machine learning model and their correlation with CO_2 emissions. The features used are GNI per capita, Energy use per capita (kilograms of oil equivalent), Population in urban agglomerations, Nationally terrestrial protected areas (percentage of total land area) and Population growth (annual percentage).

As an outline for the rest of paper, we'll present related work on CO2 emissions (Section 2), the datasets we used for our visualizations (Section 3), explain how we designed the dashboard and the process we used to build the system

(Section 4), and finally showcase the main aspects of the dashboard we have built (Section 5).

2 Related Work

There has been a significant amount research on Global CO_2 Emissions and how to mitigate it. An article published in 'Our World in Data' depicted an interactive map showing the growth rate of CO2 emissions year after year from 1750 to 2021 with a slider to choose a particular year[6]. Also in this article, a tree map visualization and an interactive scatterplot was shown for the annual CO2 emissions by country, and aggregated by region.

Another example is the Emissions Explorer project[5], published to the site Devpost. It is an interactive application designed to illustrate the potential of data visualization to derive CO2 Emissions insights from data.

In addition, an IEEE paper[3] "Prediction Model: CO_2 Emission Using machine learning published in 2018 International Conference on Artificial Intelligence and Data Processing" created predictions of the 2018 CO2 Emissions based on a Random Forest machine learning model trained on almost all continents. Several interactive charts were created to show expected CO_2 Emissions for those continents in 2018.

This dashboard is inspired by prior works but improves upon them by combining different visualizations with KPIs and making them not only more detailed and interactive, but also easier to understand.

3 Data

The main dataset used in the project was the Climate Change dataset from The World Bank. The organization has an open data platform and they update their information according to the amount of data that is brought in.

This dataset has various indicators including GDP, GNI per capita (Atlas $), Population in urban agglomerations, and Methane (CH_4) emissions. We have used this dataset as a base for all of the visualizations that we have created. The main idea was to find the important indicators for CO_2 emissions and see their effect. The dataset had these indicators as a time series data starting from 1980 to 2011. The values for the years 2009, 2010, 2011 many missing values so we did not use them in our model.

For all map-based visualizations, we used a TopoJSON format of the data extracted. We also had to reduce the scope of our visualizations to select countries due to the sheer volume of data. We selected representative countries from each continent and saw the change in their indicators and the impact on the emissions.

We also added an additional chart for CO_2 emission caused by burning of diesel. The data for this has been procured from Observable[1]

4 Approach

The website was designed to guide the user through a cohesive story of how CO_2 emissions are impacted by various indicators. There are close to 58 indicators in our original dataset so it was important to identify which ones had the largest impact on the prediction of CO_2 emissions.

4.1 Machine Learning Model

We employed a machine learning model for feature selection. Before the data could be fed to the machine learning model, we performed various data cleaning and pre-processing methods including data imputation and normalization. Feature selection was performed using a Recursive Feature Elimination with Cross Validation(RFECV)[4]. We used a 5-fold cross validation[4] with R^2 as the scoring function. The model then selected the top 5 most important features to be *GNI per capita (Atlas \$), Energy use per capita (kilograms of oil equivalent), Urban population growth (annual %), Nationally terrestrial protected areas (% of total land area), Population growth (annual %).*

We also utilized data visualization to confirm that the effect of these indicators on CO_2 emissions are reflective of their impact. The Fig.1 shows a scatter plot of the chosen features.

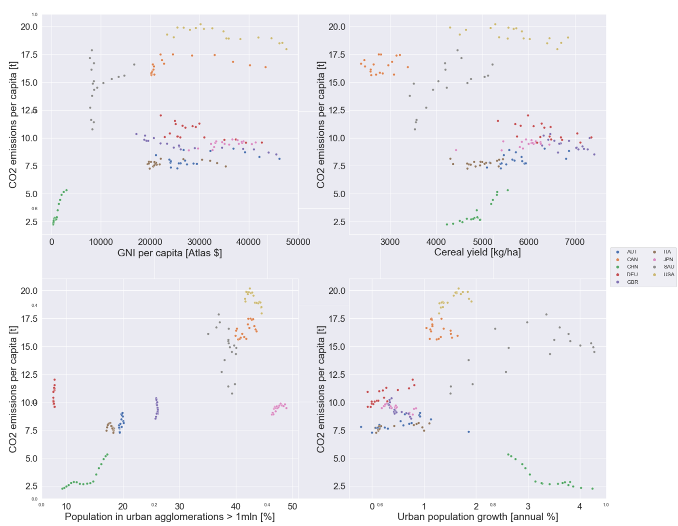

Fig. 1. Correlations of selected features

All four pieces of Fig. 1 above exhibit not only similar local tendencies for each country (in most cases), but also global trends for all data points. Note that these trends are nonlinear, with different structure in all plots – this may imply different kinds of theoretical contributions to the CO_2 emissions per capita and could be a valuable asset for the future prediction. When we created the dashboard, we kept in mind how these various factors could be explored and created a separate tab focused on viewing the changes in CO_2 emissions across the selected countries over time.

Once the data pre-processing and feature selection was completed, we implemented two models. Since this was a regression problem, we tried two different models the first was sklearn's Random Forest Regressor [4] and the next was XGBoost Regressor[2]. XGBoost showed more promising performance so we proceeded with this model. Once the algorithm was selected, we tuned hyperparameters using Randomized search[4]. We fit the model using the selected features and tuned hyper parameters to get a R^2 of 0.976. We then decided that the emissions predictions were best presented as an interactive mapbox chart for the user to peruse.

4.2 Visualization

In order to choose the most effective visualization types, we considered the type of data being displayed. For example, we decided that a scatter plot would be best to show the correlation of multiple indicators. When showing how countries perform on the selected features, we used a circle packing chart. For the map, we chose a dark theme to give a dramatic highlight when showing how the emissions increased year over year (Fig.2).

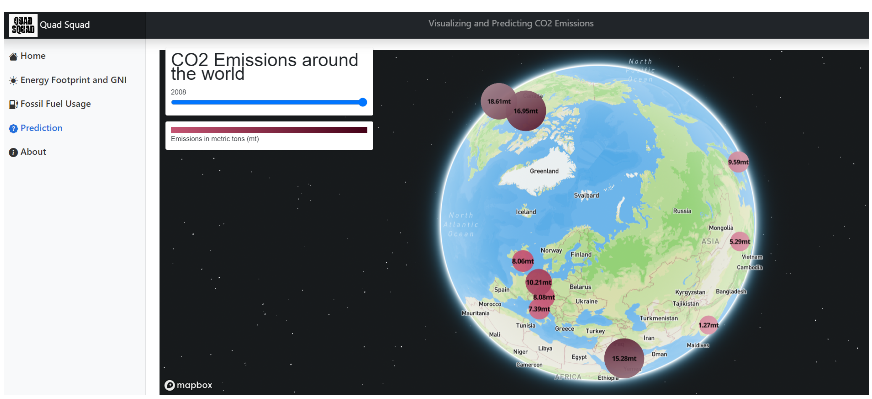

Fig. 2. Mapbox showing predicted CO2 emissions around the world

5 System

The base application for this infographic was built using Vue.js and Node.js. To format the application, Bootstrap-vue and Bootstrap templates were utilized. In order to create and implement the visualizations for the website we used D3.js.

The first tab of the dashboard focuses on how each country fares in the features selected, giving the user a good indication of the countries' relative positions. We also have an interactive scatter plot that plays out the effect of GDP on Emission Per Capita and Population Growth. These are some of the most important factors that can be explored further and are also the area of interest for the dashboard.

The next tab explores global CO_2 emissions from burning diesel from 1965 to 2021 for transportation and other purposes. The variation in colors indicates the consequent rise in global temperature. The visualization can also be viewed for continents separately.

The final tab is dedicated to a Mapbox visualization. The user is given the opportunity to interactively explore changes in CO_2 emission over the years. We have also included our emissions predictions in a JSON file so the user can view the predicted values as well. The slider available on the page allows the user to see how the values change across years.

6 Conclusion

In conclusion, as an effort to contribute to the mitigation of CO2 Emissions, we explored the climate change data, chose and implemented the most relevant models to generate CO2 emissions predictions, and transformed our findings into a cohesive sequence of several visualizations. In terms of future work, we could include more countries and with more data with can expand our predictions, visualizations, and more outcomes of interest. Our findings show that emissions continue to increase and the pace of global warming is accelerating – the urge for even more mitigation and action to be taken to minimize the impacts of climate change.

References

1. Ali Kayani, M.: co_2 emissions from burning of diesel: A disasterous journey from 1965 to 2021. https://observablehq.com/d/880a53d2f9888395
2. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. KDD '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939785, http://doi.acm.org/10.1145/2939672.2939785
3. Kadam, P.S., Vijayumar, S.: Prediction model: Co2 emission using machine learning. 2018 3rd International Conference for Convergence in Technology (I2CT) pp. 1–3 (2018)
4. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
5. Perry, L.: Carbon emission data visualization. https://devpost.com/software/carbon-emissions-calculator (2020)
6. Ritchie, H., Roser, M., Rosado, P.: co_2 and greenhouse gas emissions (2017)