Team 109 - Global Temperature Trends and Impact of CO2 and Deforestation on Temperature Change

Priyanka Bhardwaj, Gopi Chava, Imren Dinc, Phung Vy Luu, Sachin Sharma, Geneva Vezeau

1 INTRODUCTION

Climate change is one of the biggest threats in modern times, severely impacting Earth's ecosystem. The idea of climate change through global warming was first introduced in 1896 by Swedish scientist Svante Arrhenius [17]; however, it was not until 1988 that it became a worldwide topic of conversation as there was now proof that global warming was actively occurring [6]. Since then, it has been a "hot" topic for many researchers, with the majority of the work focused on high CO_2 emitting countries.

2 PROBLEM DEFINITION

After reviewing available literature, we identified that CO_2 and deforestation are two major factors impacting climate change. In this project, we will analyze temperature change for the last 250+ years and investigate effects of the two factors globally using machine learning (ML) and interactive visualization. We will use our analysis to predict the temperature change and impact of CO_2 emissions and deforestation for the next few decades based on historical datasets.

3 LITERATURE SURVEY

In this section, we summarize the studies done in climate change, deforestation, and CO_2 emissions. We started with review papers for each of our main topics. This helped us get more specific papers for our project's objective. These review papers gave us a broad idea of what has been studied in each of our main focuses so we could refine our study objectives.

Since the topic of climate change has been such a widely known issue, there have been many studies on its causes and effects and mitigation strategies. This helped us to get a starting point by building on what has been done in past studies. According to [11], the use of big data and ML methods enabled researchers to create specific solutions to climate change, opening our eyes to the use of big data and ML. [3] compared global climate change by using the latitudinal shift as a factor, predicting until 2050. This will be a good comparison for us as we are predicting until the same year, using a different method. [14]'s study shows an in depth climate shift for Europe; some of these ideas can be helpful when we are creating our models for European countries. Many recent studies were about ways to mitigate climate change, which involved impacts on the ecosystem [10] [15]; leading us to further research deforestation and its effects on climate change.

Several studies concluded that deforestation impacts climate change. Many of the papers mentioned the surface of the Earth cooling from the presence of forests or heating due to deforestation [4] [7] [13] [12]. [1]'s approach to predicting deforestation's impact on climate change used wet-bulb globe temperature estimate. However, this study was specifically for the Brazilian Amazon; while this worked for this region, it may not be useful for us when we expand to other countries. [16] investigates the impact of deforestation on climate change for different climates which will help us when we take a global approach.

Climate change has been linked to the greenhouse effect since its inception. CO_2 emissions are just one of the many gases that are associated with creating the greenhouse effect and causing climate change [19]. For CO_2 emissions studies, the majority of published papers are specific to a certain region; China [9], Iran [5], India [8], coastal areas [20] and developing ASEAN countries [2]. According to [19], CO_2 emission studies lack ML estimation and the global approach. In our project, we used several ML methods to evaluate performance. [18] applied the logarithmic mean Divisia index (LMDI). This identified major contributors to CO_2 emissions but does not relate to temperature change. [9]'s approach was to use demographic factors to predict CO_2 emissions while an innovative approach, this is difficult to use on a larger scale project like ours.

Through our research, we evaluated different ML approaches to model our data to reach our project objective.

4 PROPOSED METHOD

We predict the temperature change for the next few years and the impact of CO_2 emission and forest area on temperature change. We evaluated multiple models that can fit the available dataset and finalized two different approaches for two use-cases explained below.

4.1 Temperature Prediction Method

4.1.1 ARIMA. The first use-case is to predict the temperature based on the historical country specific dataset. The global temperature is a time series dataset, and we predict the temperature based on the historical time series temperature changes. ARMA and its variant models like ARIMA, SARIMA, SARIMAX are common methods for time series datasets. We evaluate which ARIMA model fits the best based on the datasets for different countries and use different models for different countries. Later, we fine tune the hyper-parameters to fit the country specific datasets. ARIMA or Autoregressive Integrated Moving Average model is a combination of 2 approaches - Auto Regression and Moving Averages. It also has a difference order (I) which is the number of transformations needed to make the data stationary. Below is the general forecasting equation:

$$\hat{y} = \mu + \phi_1 + y_{t-1} + \ldots + \phi p + y_{t-p} - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q}$$

ARIMA model takes 3 parameters : p: autoregressive d: integrated q: moving average

4.1.2 Artificial Neural Networks (ANN). The second use-case for our project is to predict the impact of CO_2 emission and forest area on temperature change. We use Neural Network for this use-case. Artificial neural networks (ANNs) are a type of machine learning algorithm that imitates the structure and function of the human brain. ANNs consist of interconnected nodes or "neurons" arranged in layers. Each neuron receives input signals, processes them, and produces an output signal as a result of an activation function. Finally, outputs are passed on to the next layer of neurons. ANNs are trained on a set of labeled data, adjusting the strength of connections between neurons to learn patterns and make predictions on new data. Specifically, in this study, we used Multilayer Perceptron (MLP), which is a type of feed-forward neural network, where the neurons are arranged in multiple layers, including an input layer, one or more hidden layers, and an output layer. We use ANN for regression. We use CO_2 emission and forest area data for different countries as features for the model and country average temperature are dependent variable.

4.2 Visualization Method

In our project, we use Tableau for interactive visualization of the temperature data predicted using the two approaches outlined in the previous section. The data used for interactive visualization is the outcome of the machine learning models used for temperature prediction.

To visualize the global temperatures forecast using ARIMA model, we used Tableau map layers feature along with an interactive date slider filter to change the year of interest. This visualization shows the global temperature change over time as the date slider is moved forward interactively. The data shown in this visualization covers both historical data used for model training as well as the temperature forecast by the model for future years.

The second visualization is based on the temperature prediction from ANN models that use CO_2 emissions and forest area as features. We will use Tableau map layers along with drop-down option for interactive selection of CO_2 emissions and forest area changes which allows visualizing the global temperature change based on chosen feature values.

4.3 Innovation

We identified during the literature survey that most of the studies are done for specific countries and regions since temperature show different patterns for different part of the earth. It is a challenging process to predict the temperature for all the countries in the world. In our project, we evaluate the impact of both factors on temperature trends in different countries globally with an interactive visualization tool. Our contributions can be summarized as follows:

(1) We predict the temperature for most of the countries around the globe based on the available historical city level temperature data using time series model.

(2) We create different models for different countries by tuning the hyper-parameters based on the country specific data, and then predict the temperature based on the country specific models. There are 100+ models for different countries.
(3) We predict the impact of CO_2 emission and forest area data on temperature for most of the countries using neural network.
(4) We train country specific models to evaluate the impact of CO_2 emission and forest area on temperature change as well.
(5) We create interactive visualizations to show country-specific temperature predictions and CO_2 emission and forest area impacts on temperature using a world map in Tableau.

Future Work: In this project, we only focus on predicting the temperature change for different countries across the globe to create awareness through interactive visualizations. We believe global view focusing on countries specific detail will create a lot of impact and awareness. The future work will include identifying the methods and processes to mitigate the impact of temperature change.

5 EXPERIMENTS & EVALUATION

This section explains the experiments collected and presents the proposed method results.

5.1 Dataset

We use the below 3 datasets in this project -

Global Land Temperatures By City : The dataset has the temperature details in Celsius starting for mid 1700s until 2013 for most major cities for 150+ countries across the world. It has approximately 8+ mm records. The granularity of this data is one record per month per city. With this dataset, we will calculate the mean of the monthly average temperature per year for each country. The Dataset is available at *Global Temperature Dataset by City*.

CO_2 **Emission :** This dataset has CO_2 emissions measured by metric tons per capita starting from 1990 until 2019 for all the countries across the world. It has one record per country with each column showing one recorded year. File link - *CO_2 Emission*

Forest Area: This dataset has the forest area details in hectares at regional and country level from 1990 until 2020. The granularity of this dataset is at country and year level. To prepare this dataset, we filter only country data and remove any record that does not have area data. The dataset is available at *Forest Area*

5.2 EDA, Data Preparation and Model Selection

Time Series Temperature Prediction - The global temperature dataset has temperature for major cities from November 1743 until September 2013, but a few countries do not have data from 1743 and some do not have data for all months. We dropped all the null temperature records and kept cities with data available for all 12 months and identified that there are a lot of differences in temperature ranges for different countries. We cannot create one model that can predict the temperature for all countries. Therefore, we decided to create country specific models. The dataset has monthly temperatures at city level. There are multiple records for the same time for different cities, which cannot be used in time series modeling. We calculated the yearly mean temperature at country level to prepare the time series data. We considered a 75-25% split for train and test dataset. We identified the best ARIMA model and hyper-parameters to fit the country specific data using *pmdarima* library.

CO_2 and Forest Area Impact - We used three datasets for this use-case. We need to transpose the CO_2 dataset to show years in rows for all countries to merge with other datasets on country and year. We dropped all the null values before merging the datasets. We only kept the CO_2 emission and forest area value in the dataset as features and temperature as the dependent variable for the ANN model. Figure 1 shows the summary of our ANN model.

Layer (type)	Output Shape	Param #
first_hidden (Dense)	(None, 128)	384
dropout_8 (Dropout)	(None, 128)	0
second_hidden (Dense)	(None, 64)	8256
dropout_9 (Dropout)	(None, 64)	0
output (Dense)	(None, 1)	65

Total params: 8,705
Trainable params: 8,705
Non-trainable params: 0

Figure 1: ANN Architecture

5.3 Results

5.3.1 SARIMAX. Temperature prediction using SARIMAX model based on historical time series temperature data for US only -

- MAE for Train dataset - 3.5 (approx.)
- MAE for Test dataset - 0.2 (approx.)

Figure 2 and Figure 3 show train and test dataset predictions:

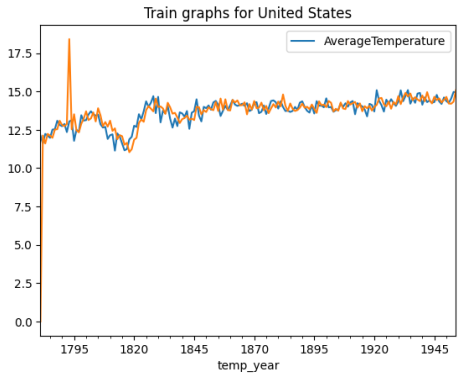

Figure 2: US Temperature Prediction for Training Dataset

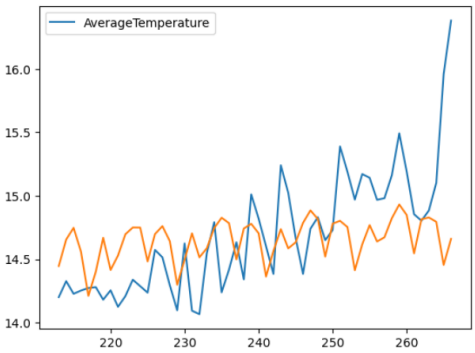

Figure 3: US Temperature Prediction for Test Dataset

5.3.2 ANN. Figure 4 shows the loss plot for ANN model with 150 epochs for impact of CO_2 emission and forest area on US temperature only.

5.3.3 Tableau Visualization. Figure 5 is an interactive global map to show the temperatures. We will use two similar interactive maps to show predicted results for both use-cases.

6 PLAN OF ACTIVITIES

The project progress is shown in Figure 6.

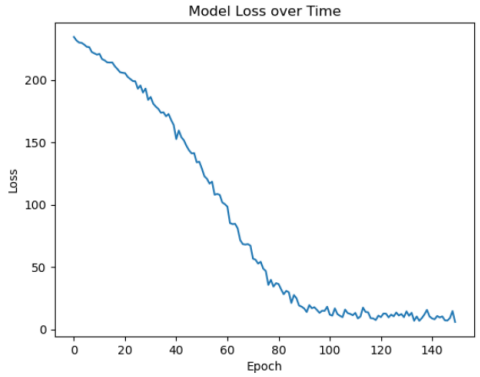

Figure 4: ANN Loss Plot

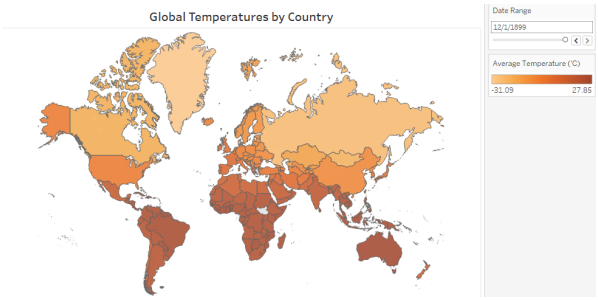

Figure 5: Country Specific Temperature Prediction

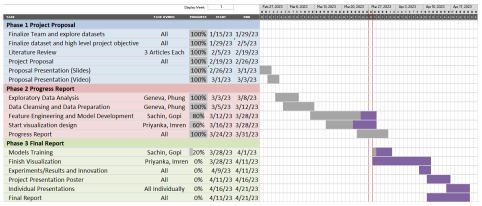

Figure 6: All team members have contributed a similar amount of effort

7 FINDINGS

We realized during EDA and model development that temperature is regional and cannot be predicted for all countries with just one model, so we pivoted our approach to build separate models for all countries. We also identified that temperature is not linearly co-related to CO_2 emission and forest area. CO_2 emission and forest area value are also completely different for different countries. So, We need to create country specific ANN models for second use-case also. We anticipate that it will take considerable amount of time to train so many country specific models.

REFERENCES

[1] Beatriz Fátima Alves de Oliveira, Marcus J Bottino, Paulo Nobre, and Carlos A Nobre. 2021. Deforestation and climate change are projected to increase heat stress risk in the Brazilian Amazon. *Communications Earth & Environment* 2, 1 (2021), 207.

[2] NAM Azmin, Z Ahmad, R Mahmood, ASM Zahari, and H Hendar. 2022. The Dynamic Linkages between CO2 Emissions, Energy Consumption and Economic Factors in ASEAN Countries. In *IOP Conference Series: Earth and Environmental Science*, Vol. 1102. IOP Publishing, 012038.

[3] Jean-Francois Bastin, Emily Clark, Thomas Elliott, Simon Hart, Johan Van Den Hoogen, Iris Hordijk, Haozhi Ma, Sabiha Majumder, Gabriele Manoli, Julia Maschler, et al. 2019. Understanding climate change from a global analysis of city analogues. *PloS one* 14, 7 (2019), e0217592.

[4] David Ellison, Cindy E Morris, Bruno Locatelli, Douglas Sheil, Jane Cohen, Daniel Murdiyarso, Victoria Gutierrez, Meine Van Noordwijk, Irena F Creed, Jan Pokorny, et al. 2017. Trees, forests and water: Cool insights for a hot world. *Global environmental change* 43 (2017), 51–61.

[5] Seyed Mohsen Hosseini, Amirali Saifoddin, Reza Shirmohammadi, and Alireza Aslani. 2019. Forecasting of CO2 emissions in Iran based on time series and regression analysis. *Energy Reports* 5 (2019), 619–631.

[6] Peter Jackson. 2007. From Stockholm to Kyoto: A Brief History of Climate Change. Retrieved March 31, 2023 from https://www.un.org/en/chronicle/article/stockholm-kyoto-brief-history-climate-change#:~:text=In%201988%2C%20global%20warming%20and,public%20debate%20and%20political%20agenda

[7] Alexander Koch and Jed O Kaplan. 2022. Tropical forest restoration under future climate change. *Nature Climate Change* 12, 3 (2022), 279–283.

[8] Surbhi Kumari and Sunil Kumar Singh. 2022. Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research* (2022), 1–16.

[9] Shijie Li and Chunshan Zhou. 2019. What are the impacts of demographic structure on CO2 emissions? A regional analysis in China via heterogeneous panel estimates. *Science of the Total Environment* 650 (2019), 2021–2031.

[10] Yadvinder Malhi, Janet Franklin, Nathalie Seddon, Martin Solan, Monica G Turner, Christopher B Field, and Nancy Knowlton. 2020. Climate change and ecosystems: Threats, opportunities and solutions. , 20190104 pages.

[11] Felix Creutzig Nikola Milojevic-Dupont. 2021. Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society* 64 (2021).

[12] Álvaro Salazar Oswaldo Maillard, Roberto Vides-Almonacid and Daniel M. Larrea-Alcazar. 2022. Effect of Deforestation on Land Surface Temperature in the Chiquitania Region, Bolivia. *Land* 12, 1 (2022), 2.

[13] Jayme A Prevedello, Gisele R Winck, Marcelo M Weber, Elizabeth Nichols, and Barry Sinervo. 2019. Impacts of forestation and deforestation on local temperature across the globe. *PloS one* 14, 3 (2019), e0213368.

[14] Guillaume Rohat, Stéphane Goyette, and Johannes Flacke. 2018. Characterization of European cities' climate shift–an exploratory study based on climate analogues. *International Journal of Climate Change Strategies and Management* 10, 3 (2018), 428–452.

[15] Haroon Sajjad, Pankaj Kumar, Md Masroor, Md Hibjur Rahaman, Sufia Rehman, Raihan Ahmed, and Mehebub Sahana. 2022. Forest vulnerability to climate change: A review for future research framework. *Forests* 13, 6 (2022), 917.

[16] Catherine E Scott, Sarah Anne Monks, DV Spracklen, SR Arnold, PM Forster, A Rap, M Äijälä, P Artaxo, KS Carslaw, MP Chipperfield, et al. 2018. Impact on short-lived climate forcers increases projected warming due to deforestation. *Nature communications* 9, 1 (2018), 157.

[17] Holly Shaftel. 2023. How Do We Know Climate Change is Real? Retrieved March 31, 2023 from https://climate.nasa.gov/evidence/#:~:text=In%201896%2C%20a%20seminal%20paper,Earth's%20atmosphere%20to%20global%20warming

[18] Saeed Solaymani. 2019. CO2 emissions patterns in 7 top carbon emitter economies: The case of transport sector. , 989–1001 pages.

[19] Lebunu Hewage Udara Willhelm Abeydeera, Jayantha Wadu Mesthrige, and Tharushi Imalka Samarasinghalage. 2019. Global research on carbon emissions: A scientometric review. *Sustainability* 11, 14 (2019), 3972.

[20] Kelvin O Yoro and Michael O Daramola. 2020. CO2 emission sources, greenhouse gases, and the global warming effect. In *Advances in carbon capture*. Elsevier, 3–28.