The Intersection of Residential CO2e and Climate

Aditya Kulkarni
COMS 6998: ML and Climate

December 2022

Contents

1 Abstract

The human race collectively emits approximately 51 billion tonnes of carbon dioxide equivalent (CO_2e) per year. In recent years, the COVID-19 pandemic has changed the trajectory and contributed to a record high in residential emissions. Every household has an impact on climate change, and the residential sector is responsible for a significant portion (20%) of total emissions. As the importance of addressing climate change and sustainability management becomes increasingly recognized, there is a growing need for solutions to combat global warming and its effects.

Through machine learning, this research project aims to uncover patterns and insights within residential households that can support climate change efforts in the current environment, where publicly available emission data is scarce. By examining the relationship between climate change policy and residential CO_2e emissions at a macro scale, our study provides a detailed and comprehensive understanding of the connection between these factors through various case studies and forecasting models. Overall, this project provides a valuable starting point for further studies in sparse data emission tracking models.

2 Introduction

2.1 Overview

In recent years, the growth of cryptocurrency mining and the COVID-19 pandemic have contributed to the rise in residential emissions, which now account for a hefty 20% of total emissions [1]. This means that households and related, through their energy use and consumption habits, are contributing significantly to the overall climate impact of the human race. In short term, time series prediction of emission data is stated to be inaccurate. Further, climate laws take a long time to foster merit during which pure emission data does not accurately represent the space. Through our research on the relationship among CO_2e, climate change policy, and residential emissions, we aim to provide a vivid picture of the connection between these factors and offer insights on how effective current policy is in tackling the increasing residential emissions.

The recent past has shown the usefulness in using technology in the field of climate change. The intersection of machine learning and artificial intelligence can help identify the most effective ways to reduce emissions and support climate change efforts through models with a focus on improving energy efficiency, promoting renewable energy sources, and engaging households in climate action. With machine learning, we can use our abundance of historical climate data and observations to improve predictions of Earth's future climate [2].

On the other hand, there is no universal climate law governing residential greenhouse gas (GHG) emissions. Each country has its own laws and policies in place to address this issue. These laws aim to reduce or limit the amount of GHG emissions produced by households and residential buildings, and can be implemented through a variety of measures such as setting emissions targets or limits, implementing carbon pricing schemes, promoting energy efficiency, and providing incentives for households to reduce their emissions. Some countries also have regulations that require residential buildings to meet certain energy efficiency standards. In fact, all countries targeted for zero-carbon-ready codes for new buildings by 2030 [3]. However, the validity of laws in climate policy is a space that is often not as often explored. Instead, machine learning engineers rely on instantaneous factual data in the form of accuracy- the effectiveness of laws over time is less investigated.

A main motivation for this project is the lack in publicly and easily accessible data on emissions and climate laws. We aim to address this by creating a simple model that could project future emissions and take climate laws into account. With less numeric data but a sufficient amount of public climate policy into account, this project aims to deliver two principal objectives:

1. Analyze multiple trends in various countries CO_2 emissions over time, and examine climate policies on these patterns. This will be done through case studies for certain groups of countries based on similar in emission data trends.

2. Inspect the relationship between these projections, its validity and governing climate laws to create a generalizable model for short term emissions.

Through these efforts, we hope to gain a better understanding of the complex relationship between CO_2e emissions, climate change, and the residential sector.

2.2 Data

As we intend to develop forecasting models as a part of this project, it was of paramount importance to obtain a balanced and comprehensive dataset.

All data for this project was collected from the Organization for Economic Cooperation and Development (OECD) website, specifically from the statistics section. The OECD offers a wide range of free data that is organized by theme, pollutant source, and variables, among other categories. We focus on residential and related sources of greenhouse gas (GHG) emissions, specifically in terms of carbon dioxide equivalent (CO_2e) measured in thousands of tonnes. We use residential CO_2e data for approximately 56 countries and country groupings, covering the period from 1990 to 2020.

RangeIndex: 75691 entries, 0 to 75690 Data columns (total 17 columns): #	Column	Non-Null Count	Dtype	
0	COU	75691 non-null	object	
1	Country	75691 non-null	object	
2	POL	75691 non-null	object	
3	Pollutant	75691 non-null	object	
4	VAR	75691 non-null	object	
5	Variable	75691 non-null	object	
6	YEA	75691 non-null	int64	
7	Year	75691 non-null	int64	
8	Unit Code	72758 non-null	object	
9	Unit	72758 non-null	object	
10	PowerCode Code	75691 non-null	int64	
11	PowerCode	75691 non-null	object	
12	Reference Period Code	0 non-null	float64	
13	Reference Period	0 non-null	float64	
14	Value	75691 non-null	float64	
15	Flag Codes	4 non-null	object	
16	Flags	4 non-null	object	
dtypes: float64(3), int64(3), object(11)				

Table 1. Table containing initial data with disposable information. We only focus on Value, Year and Country for this project.

Table 1 presents a detailed overview of our gathered dataset. Our dataset contains approximately $76,000$ rows of data, with each row containing information on the Country, Year, and CO2e Value. After removing all dispensable variables, our final dataset consists of $76,000$ entries of **univariate** CO_2e value data sorted by country and year. Table 2 presents a few selected entries for Ukraine from the years 1990 to 1994.

One significant limitation of the dataset used in this study is the maximum number of data points per country, which is only 31. This presents a challenge for accurate forecasting using most methods, as a small amount of data can limit the reliability and robustness of the resulting models. To address this issue, we leverage resources specifically designed for univariate time series data. Despite these efforts, the limited amount of data remained a potential limitation of the study.

Country	Value	Year
Ukraine	102010.596	1990-01-01
Ukraine	80058.191	1991-01-01
Ukraine	81660.872	1992-01-01
Ukraine	72460.327	1993-01-01
Ukraine	65156.889	1994-01-01

Table 2. Excerpt of DataFrame containing values of emission data of Ukraine from 1990 to 1994.

3 Case Studies

In this project, we analyzed a large number of laws and policies from the 56 countries and country groupings involved. Due to page limitations, we are unable to present most of the information we gathered. Out of the six case studies we explored, we have included **three** which have specific policy-emission relationships that were used to draw our conclusions. Some of the rest will be available on the Github.

3.1 Case Study I: Four countries resilient to climate change

Using the ND-GAIN Country Index, a tool which summarizes a country's vulnerability to climate change and ranks them accordingly, we selected certain countries which are stated to be the most resilient to climate change i.e. have great climate laws and policies. Their emission data is then compared to their policy making structure to observe law-emission relationships. See Figure 1.

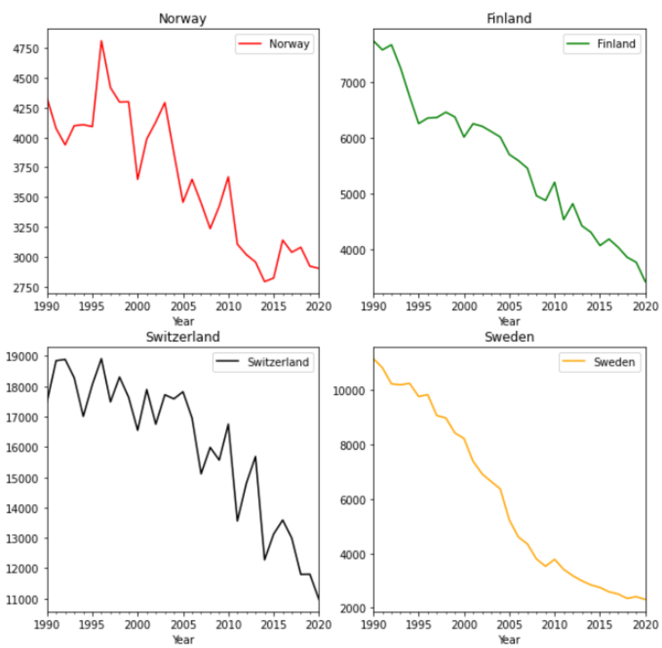

Figure 1. Emission Data of Four Resilient Countries- Norway, Finland, Switzerland, Sweden

The above countries take the top four spots on the ND-GAIN Country Index. Each of these countries have certain polices that give reason for their low household emission levels. Sweden (bottom-right) for instance, is stated to trend towards lower emissions from household consumption in all areas [4]. This is attributed to the country's increasing reliance on clean energy sources. Renewable sources of energy, hydroelectric and nuclear power in particular, now account for more than 66% percent of electricity generation [5]. Along similar lines, Norway (top-left) has kept climate into account for more than two decades. For instance, all plans under the Planning and Building Act of 2008 [6] act were to take the climate into account in energy supply solutions.

3.2 Case Study II: Countries most affected by Climate Change

Another case study done while preparing data for forecasting was done by looking for countries most affected by climate change without considering the dataset, their respective policies, and finally relating trends in the dataset. See Figure 2.

Figure 2. Emission Data of Four Vulnerable Countries- Japan, India, Canada, Russia

Japan is particularly vulnerable to the impacts of climate change due to its geography and dependency on fossil fuels. For instance, in October 2019, Japan was hit by Typhoon Hagibis, the most powerful typhoon in over 60 years [7]. While Japan's other sectors, such as industry and transportation, have seen a decrease in energy consumption, the household sectors have experienced a significant increase (16.9% from the 1990 level) [8]. The graph on the other hand does not display this imminent threat. Therefore, implementing various emission measures is essential for addressing this issue in Japan. On the other hand, One of the key components of India's climate policy is the National Action Plan on Climate Change (NAPCC) which outlines eight specific "national missions" that aim to address key areas related to climate change [9]. In addition, India has also pledged to reduce its emissions intensity, or the amount of greenhouse gases emitted per unit of GDP, by 33-35% by 2030 compared to 2005 levels. However, India (top-right) displays current emission levels implications with ever-increasing CO_2e values. In 2019, the yearly monsoon was extremely critical, and had 11.8 million people affected with the economic damage estimated to be $10 billion [10].

3.3 Case Study III: Four countries with most positively changed emission trend

For this case study, we parse through the data that had the most percentage decrease in emissions on average. Ukraine, Lithuania, Sweden, and Bulgaria seemed to trend the most downwards in the whole dataset.

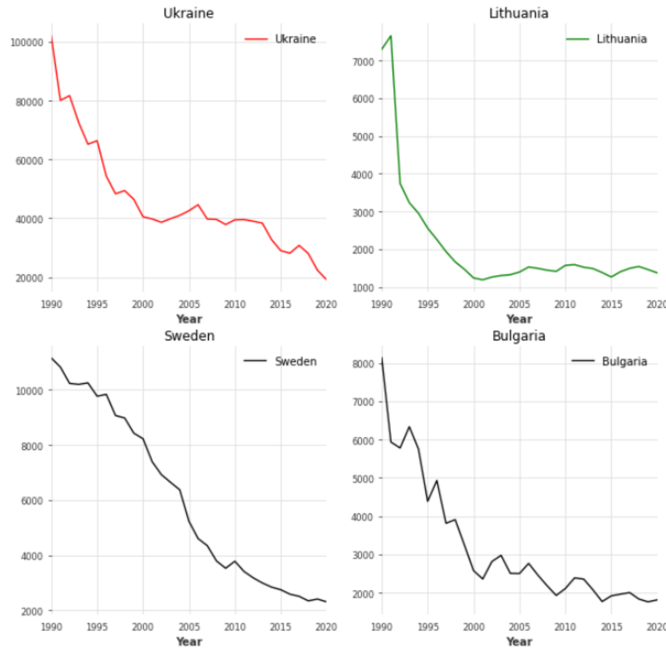

Figure 3. Emission Data of Four Countries Trending Downwards- Ukraine, Lithuania, Sweden, Bulgaria

Surprisingly, all four of the countries generated are a part of the European Union (EU). One could attribute it to the their emission laws and how they have improved in the last decade. In 2012, households and industry were responsible for half of the EU greenhouse gas emissions from fossil fuels.[11]. In 2019, Ukraine passed Law no 0875 on Principles of Monitoring, Reporting and Verification of GHG Emissions. This law aims to align the country's emission monitoring framework with the EU Emissions Trading System- the legislation which sets minimum energy efficiency standards for buildings [12]. In similar fashion, Sweden has a strong track record when it comes to climate policy and emissions laws; the country has been a leader in the transition to renewable energy and has some of the most ambitious emissions reduction targets in the world. One of the key pillars of Sweden's climate policy is its target to become a net-zero emissions country by 2045. To achieve this goal, Sweden has implemented a number of policies and programs, including a number of programs and incentives to promote energy efficiency and conservation, such as grants for energy-efficient buildings and appliances, and public education campaigns.

4 Emission Forecasting

This second half of this research project involves in making predictions about future emissions data until 2030 using data from 1990 to 2020. The goal is to understand the potential impact of climate change laws on these countries and to identify whether policies have strong influence on periodic emission data. Specifically, we are creating forecasting models using various machine learning models to gain future projections of CO_2 emissions and observing their relationship to climate change policy and current climate laws. Each section will provide the model used as well as information on certain predictions and their respective accuracies.

Data Preparation

For various time series models such as ARIMA, we have to make the data stationary for it to work. By becoming stationary, we remove the trends and seasonality from the information which makes it easier for ARIMA to make a prediction. To check if a time series is stationary, we apply a differencing technique known as Augmented Dickey-Fuller (ADF) test [13]. The ADF test uses hypothesis testing to check for stationarity and has an important parameter known as the 'p-value' that determines whether a time series is stationary. We ensure each set of data presented in our pipeline is made stationary. This is done by finding the difference of the $\sqrt(\log)$ of the values. We also make sure that all data is continuous, at the same scale, and is of the same units.

4.1 Neural Prophet

This machine learning algorithm uses deep learning techniques to predict time series data. Neural Prophet extends the Prophet algorithm with neural networks, which allows it to make more accurate predictions and handle a wider range of data types and patterns including univariate time series data. Using Neural Prophet, we predict emission data for several countries from the case studies, two of which is listed below:

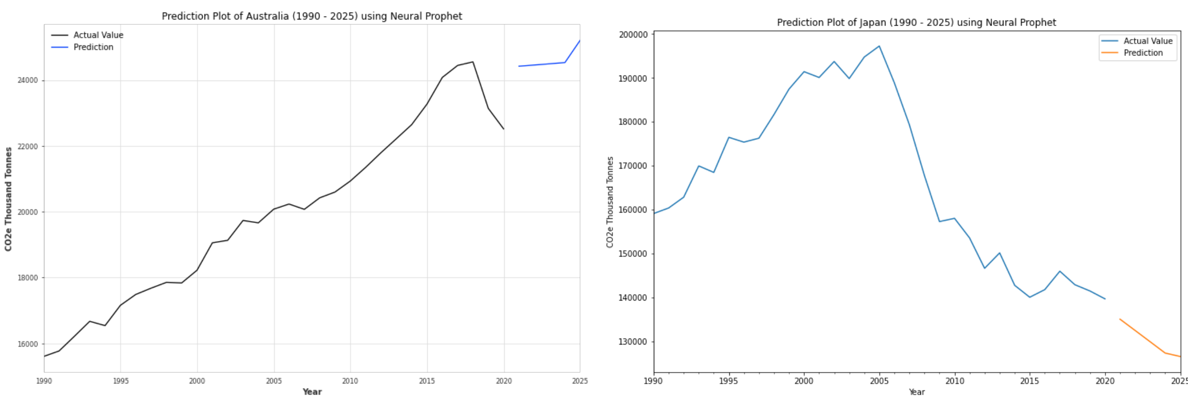

Figure 4. NeuralProphet. **Left:** Australia prediction. Clearly does not take climate policy into account. **Right:** Japan prediction. Interesting downward trend forecast.

- **Australia-** According to our analysis 4, Australia's residential emissions are expected to steadily increase until 2024, when they are projected to increase exponentially. Interestingly, this trend aligns with that of the predicted emissions that of the for 2021-2022 [14].
- **Japan-** We already went over a clear case study of Japan in the section before this. From the analysis, we admitted that Japan is prone to climate change in the near future. However, NeuralProphet does not have any knowledge of that and predicts lowering emission data from 2020- 2025.

In our analysis, we found that the use of the NeuralProphet model for predicting emissions in the case study countries was convenient but inadequate. This model lacks the ability to consider emission policies and other relevant factors, resulting in naive and inaccurate predictions. Although the model had good short term predictions, to improve the accuracy for a longer timespan, we would need to incorporate additional information or/and models, such as those related to climate policies.

4.2 ARIMA

ARIMA (AutoRegressive Integrated Moving Average) is a statistical technique used for forecasting time series data. It consists of three components: autoregressive (AR), integrated (I), and moving average (MA). Below, we see gathered data that is grouped based on proximity. Figure 5 displays three country groupings- Asia Oceania, America, and the Total. From the image, it is easily inferable that the decrease in CO_2e emissions in Asia Oceania is what contributed to the total decrease in emissions worldwide when compared to America.

When running the ARIMA model on the total emissions, we obtain the following results as displayed in 6. The *coef* column in the second row has a larger constant for *const* than the values (D.Diff). Furthermore, the values are not statistically significant as their p-values are above the 0.05 threshold. Overall, we can conclude that the ARIMA model produces undesirable results and is not suitable for modeling our type of data. This is likely due to the low number of data points for each country. We ran ARIMA for various hyperparmeter configurations on our data but were unable to produce accurate predictions.

Figure 5. Emissions by location. **Topleft:** We see that emissions in Asia Oceania have had a great decrease in the last 15 years. **Topright:** Emissions in America have had been oscillating since 1995 with an overall decrease. **Bottom:** Total emissions have been on a decrease ever since 1990.

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Diff   No. Observations:                   29
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -130.262
Method:                       css-mle   S.D. of innovations             20.180
Date:                Thu, 15 Dec 2022   AIC                            268.523
Time:                        20:28:40   BIC                            273.992
Sample:                             1   HQIC                           270.236

================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.2895      0.343     -0.845      0.398      -0.961       0.382
ar.L1.D.Diff    -0.2751      0.182     -1.508      0.132      -0.633       0.082
ma.L1.D.Diff    -1.0000      0.094    -10.609      0.000      -1.185      -0.815
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -3.6348           +0.0000j            3.6348            0.5000
MA.1            1.0000           +0.0000j            1.0000            0.0000
-----------------------------------------------------------------------------
```

Figure 6. ARIMA Model Output for OECD- Total.

4.3 PyAF (Python Automatic Forecasting)

A great open source prediction tool available is PyAF (Python Automatic Forecasting). It works as an automated process for predicting future values of a trend using a machine learning approach by providing a set of features that is comparable to some other forecasting methods mentioned in this paper. Two of the several countries we modeled with this method were as follows:

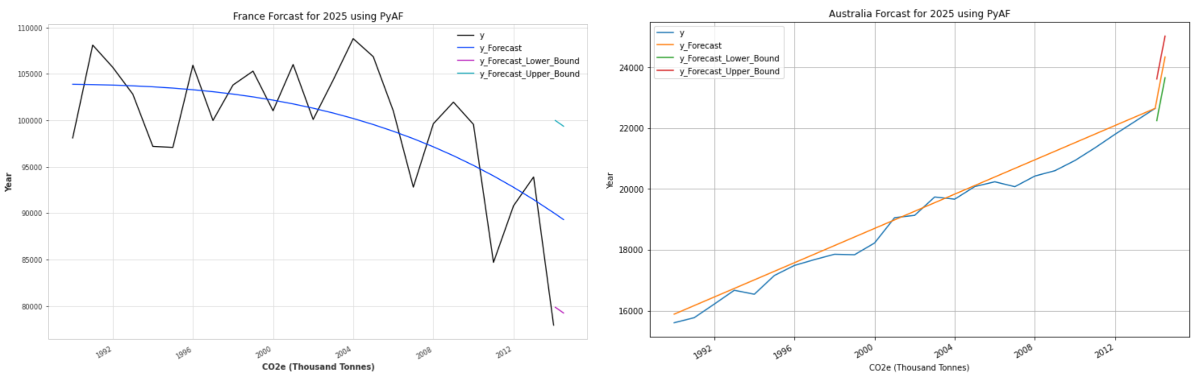

Figure 7. PyAF. Left: France Forecast **Right:** Australia Forecast. The addition of variance in univariate emission data seems to nullify the need to add climate policy for short term predictions.

- **France**: We observe an accurate forecast prediction a using 75-25 split on a decreasing slope with dedicated accurate bounds. A comparison with other sources such as this one from Statista validates it. Furthermore, our predictions align with the emission reduction goals set by France in 2021.
- **Australia**: Compared to France, the model prediction for Australia is very confident- the difference between bounds is very low. We wonder how the Climate Change Bill of 2022 will effect the confidence scores for the next few years. We cannot model this as we do not have data from 2021 onwards.

4.4 Exponential Smoothing

Holt-Winters Exponential Smoothing is a variant of exponential smoothing that uses a weighted moving average to forecast future values. Holt-Winters Exponential Smoothing extends this approach by incorporating additional components to model trends and seasonality in data. The addition of climate policy and governing laws is paramount to this project.

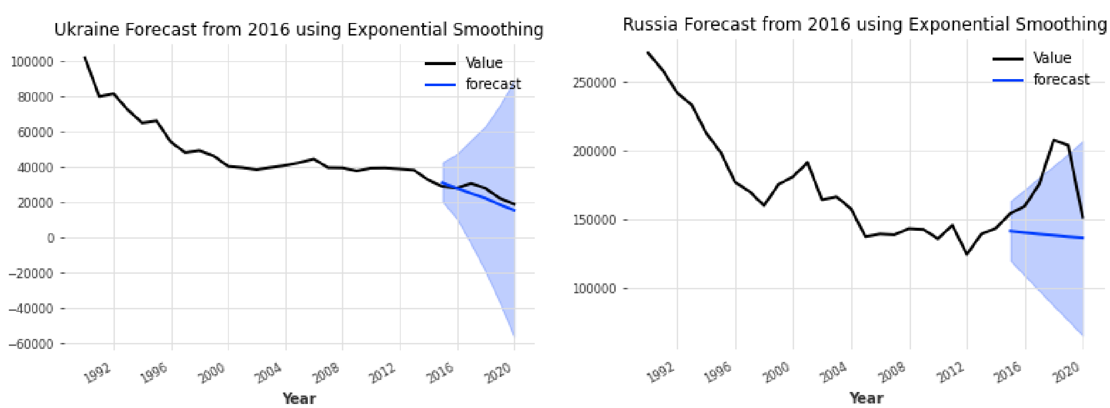

Figure 8. Exponential Smoothing. Left: Ukraine Emission Prediction **Right:** Russia Emission Prediction. A simpler model but greater training data seems to have the best predictions.

- **Ukraine:** Ukraine's emission data does not include seasonality and the initial trend does not have any major fluctuations. The forecast is extremely accurate as displayed in 8 but is worrysome and has a high area of variance.
- **Russia:** Figure 8 displays the issue with having fewer data points. The data from 2016 to 2020 had great deviation from the mean due to COVID-19. However, the forecast is a linear plot without any vertical scaling. Although the start and final emission value predicted are correct, our model accurately capture the need to follow the emission graph precisely. This is an important consideration for future forecasting. Additionally, we are interested in how Russia's first household emission law will effect future emissions from 2023.

5 Future Work

A potential direction for future work as stated earlier is to incorporate new data sources into our machine learning models. For instance, reinforcement learning can be used to improve the emission prediction tool. Here, the reward function can be made such that it gives positive rewards for every emission related policy issued *if* there has been a decrease in emissions. Doing so, we would not have the issue of data scarcity and could gather enough data for larger more complex models. Further, data augmentation could be used to improve the accuracy of our machine learning model. As the limited size and scope of this dataset limited the performance of our models, data augmentation could be a great method to scale the dataset while retaining the properties of the emission trend. We specifically attempted utilizing feature augmentation using deep generative models, but could not complete it within the time constraint.

6 Conclusion

In conclusion, this research project demonstrates the potential of using machine learning to understand the relationship between climate change policy and residential emissions. Our selection of case studies and forecasting models provide a generalizable method for making short-term emission predictions. In addition, it is evident that law and policy usually does not change the forecast in the span of five to ten years unlike a pandemic or interest surge. The findings of this project also highlight the importance in data abundancy and display the implication of inaccuracy due to it.

Clearly, climate change laws do work but they need to be done effectively right [15]. Further, it is not surprising that many laws that have been mandated do not, for the lack of a better word, *harvest* fast enough. For instance, France's first climate law in decades is said to have fallen short [16]. Machine learning models such as reinforcement learning models can be used to add in multiple forms of data such as climate policy.

Overall, the results of this research project demonstrate the potential of machine learning to support climate change efforts and provide valuable insights into the relationship between CO_2e and governing laws. Machine learning as displayed in this paper can be used an a great medium of interpreting emission policy and its effectiveness. With new modes of data interpretation, prediction models can be enhanced to discover uncovered patterns within emissions to support climate change efforts in the current environment, all while policy is being processed.

References

[1] uc davis. Analysis of greenhouse gas emissions from residential heating technologies in usa, Feb 2022. URL https://wcec.ucdavis.edu/analysis-of-greenhouse-gas-emissions-from-residential-heating-technologies-in-usa-2.

[2] The role of machine learning in helping to save the planet. URL https://www.weforum.org/agenda/2021/08/how-is-machine-learning-helping-us-to-create-more-sophisticated-climate-change-models.

[3] Iea. All countries targeted for zero-carbon-ready codes for new buildings by 2030 – analysis. URL https://www.iea.org/reports/all-countries-targeted-for-zero-carbon-ready-codes-for-new-buildings-by-2030-2.

[4] Greenhouse gas emissions from domestic demand remained unchanged. URL https://www.scb.se/en/finding-statistics/statistics-by-subject-area/environment/environmental-accounts-and-sustainable-development/system-of-environmental-and-economic-accounts/pong/statistical-news/environmental-accounts--environmental-pressure-from-consumption-2018/.

[5] Published by Statista Research Department and Jul 5. Electricity: Share of renewable energy 2007-2018, Jul 2021. URL https://www.statista.com/statistics/419480/sweden-share-of-electricity-from-renewable-sources/.

[6] Ministry of the Environment. Planning and building act (2008), Jun 2008. URL https://www.regjeringen.no/en/dokumenter/planning-building-act/id570450/.

[7] David Eckstein, Vera Künzel, and Laura Schäfer. Global climate risk index 2021. URL https://www.germanwatch.org/en/19777.

[8] URL https://iea.blob.core.windows.net/assets/d0031107-401d-4a2f-a48b-9eed19457335/GlobalEnergyReview2021.pdf.

[9] URL https://static.pib.gov.in/WriteReadData/specificdocs/documents/2021/dec/doc202112101.pdf.

[10] Climate risk index 2021: India is 7th most climate-affected country. URL https://uk.movies.yahoo.com/india-ranks-7th-on-the-global-climate-risk-index-065701509.html.

[11] Households and industry responsible for half of eu greenhouse gas emissions from fossil fuels, Jun 2016. URL https://www.eea.europa.eu/highlights/households-and-industry-responsible-for.

[12] Vizzuality. Planning and building act no. 71 of 2008. URL https://climate-laws.org/geographies/norway/laws/planning-and-building-act-no-71-of-2008.

[13] Augmented dickey–fuller test, Nov 2022. URL https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test.

[14] National greenhouse gas inventory quarterly update: December 2021. URL https://www.dcceew.gov.au/climate-change/publications/national-greenhouse-gas-inventory-quarterly-update-december-2021.

[15] David Vetter. How do you tell if climate laws really work? start by counting them, Aug 2020. URL https://www.forbes.com/sites/davidrvetter/2020/08/11/how-do-you-tell-if-climate-laws-really-work-start-by-counting-them/?sh=164df8ea4e11.

[16] France passes climate law, but critics say it falls short, July 2021. URL https://www.nytimes.com/2021/07/20/world/europe/france-climate-law.html.