

**Data Mining Analysis on Climate Change and Regional Agriculture Proposal**

02.23.2022

**─**

Jason Ortiz, Daniel Cohen, Rhodiam Arango

Computer Science Senior Project

CSCI 455 - M01

NYIT Manhattan Campus

# **Overview**

Climate Change has continued to change Earth’s environment, as a result of our actions over decades, and even now. We’re aware of the impacts that have already occurred as a result of Climate Change, but we need to look to the future, and Machine Learning can be helpful in predicting how Climate Change will affect our world in the future, if we remain on the same track.

# **Goals**

1. Use Machine Learning & Data Mining models and techniques to analyze and visualize very large datasets to find new patterns and reinforce expected patterns regarding Global Warming and Global agricultural data.
2. Use the patterns from the previous goal to predict which region’s agriculture will be affected the most, either for the country's benefit or detriment, and how this will impact the country as a whole.

# **Materials**

# **Datasets to be used:**

* Global CO2 Emissions Data
* Monthly CLIMAT reports from stations worldwide (DWD global climate data 2003 - 2020 from 4169 weather stations)
* Climate Change: Earth Surface Temperature Data

# **Literature Review**

***CO2 Emissions: Our World in Data***

This source provides detailed information and data on a country’s CO2 emissions, and presents detailed metrics in the data set. Some of these metrics include annual emissions, average emissions per person, how much CO2 has been emitted over time. The data even looks at CO2 emission due to coal, oil, gas, flaring, or cement production, etc. Overall, this source is very detailed with regards to CO2 emissions, and the source of those emissions, which is very positive for our project, and definitely provides a very useful data source. Something missing from this source specifically, which would be a huge help for our project, is a metric showing at the very least, an average temperature measurement every year. Looking at the change in temperature over a period of time is crucial for our machine learning project. Despite this however, this source still remains very beneficial because looking at the CO2 emissions for individual countries can give us an idea as to which countries will see a greater temperature change, due to the correlation between CO2 emissions and climate change, and overall temperature change. Aside from the importance of the data set and the different visualizations provided, there is also information provided on the importance of the sources behind the CO2 emission, like Greenhouse Gases, and coal, oil, gas, etc. This source overall provides useful information not only for the application of our project, but from an informational standpoint, so that we know more about the topic that we are focusing on.

***Predicting Weather Temperature Change using Machine Learning Models***

Some of the sources that we found demonstrated multiple use cases where Machine Learning was used to try and solve a problem/issue related to Climate Change, such as predicting average land/ocean temperatures, predicting extreme precipitation, and even predicting the impacts of climate change on crop yields, which relates very heavily to our own topic/use case, so that’ll be helpful guidance for us.

Other sources go in-depth relating the data that has been collected, such as global CO2 emissions. These sources can help us gain a better understanding of the topic at hand, and allow us to determine which features of our datasets will be important for our use case.

# **TechStack**

* **Python:** Many useful libraries to use for data analysis and visualization
  + Pandas
  + NumPy
  + Seaborn
  + Matplotlib
  + SKLearn
* **MySQL:** For integrating, storing, and querying

## Machine Learning and Data Mining Techniques

* Various visualization techniques and tools for data analysis and visualization.
* Avoiding the “Curse of Dimensionality” by using techniques such as feature selection and Support Vector Machines

# **Timeline**

## Past Week: 02/16/2022 - 02/23/2022

We accumulated more resources, looking into other cases where machine learning had been applied to the topic of Climate Change, to get a better idea of what has been done.

## Next Week: 02/23/2022 - 03/02/2022

We plan to analyze and visualize various weather and agriculture datasets, including both historical data as well as predictions made by previous researchers. We will take a closer look at which features will be important throughout the project through correlation analysis, as well as models used by previous researchers.

# **Work Division**

## Jason Ortiz

Taking into account my prior experience with Machine Learning Projects I will both be; making sure that Daniel is up to date with any tools/technologies we will be using throughout the project, as well as visualizations for the data we are using.

## Daniel Cohen

I will be supporting both Jason and Rhodiam with any of the coding/visualizations of our database, as well as analyzing the results from our coding, and what those results mean, and why it’s significant.

## Rhodiam Arango

Considering that I have prior experience with Machine Learning projects, I’ll be working on some of the coding and manipulation of our database through Python. I’ll work with Jason on some of the visualizations that may be needed for our data.

# **Dataset Sources**

<https://github.com/owid/co2-data>

<https://www.kaggle.com/christopherlemke/monthly-climat-reports-from-stations-worldwide>

<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>

<https://data.world/datasets/agriculture>

# **Research Sources**

<https://www.fao.org/3/i2490e/i2490e01c.pdf>

[CO2 emissions - Our World in Data](https://ourworldindata.org/co2-emissions)

<https://medium.com/swlh/predicting-weather-temperature-change-using-machine-learning-models-4f98c8983d08>

<https://vitalflux.com/machine-learning-use-cases-climate-change/>

<https://climate.nasa.gov/news/3124/global-climate-change-impact-on-crops-expected-within-10-years-nasa-study-finds/>