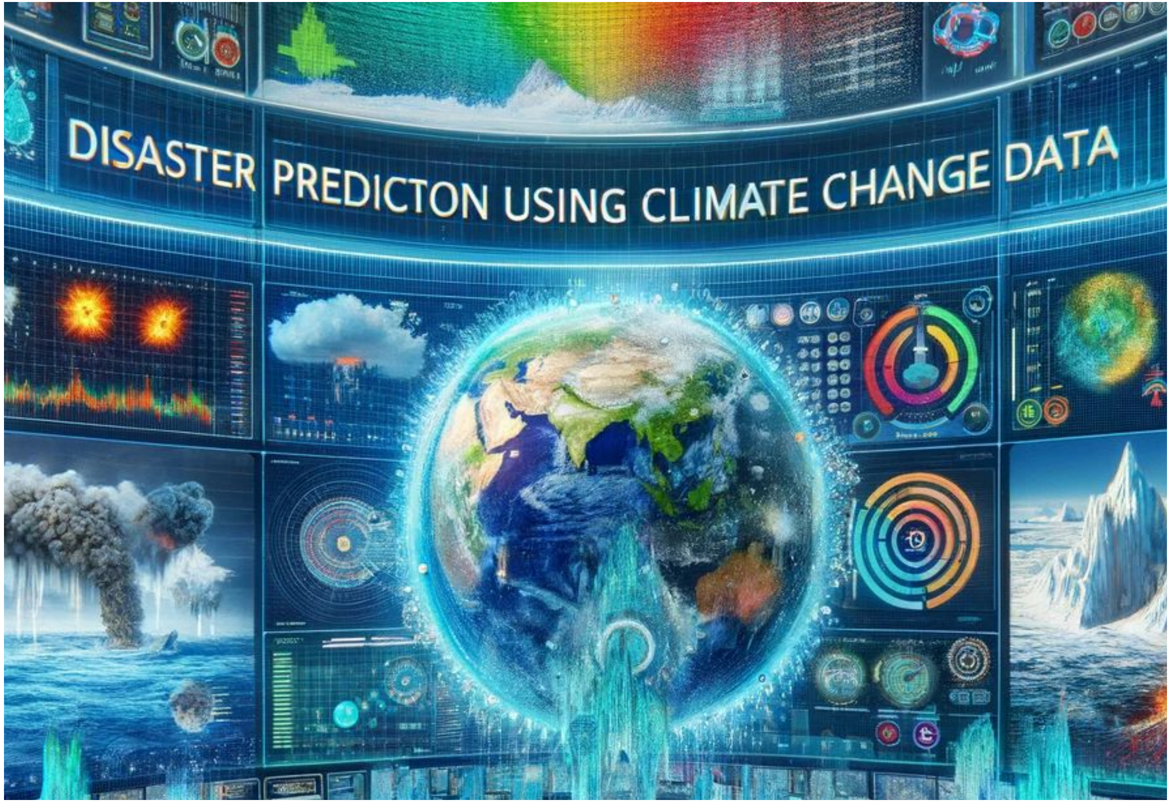


DS-5500

By: Sean Klein, Kevin Russell, Ataish Nehra

Problem Statement.

https://www.reuters.com/pictures/wildfires-devastate-british-columbia-interior-communities-2023-08-24/7D77Z5R4KZNMNOOIQASJ7DONK4/ (Chris Hengren)

https://en.wikipedia.org/wiki/Effects_of_Hurricane_Matthew_in_Haiti

https://www.weather.gov/bgm/pastFloodSeptember072011

- As the climate worsens, there has been an increase in the amount of natural disasters that occur.
- Our goal is to find trends for natural disasters in relation to the effects of climate change.
- We aim to be able to be able to predict natural disaster occurrences and discover which climate change affects are the best predictors for Natural Disasters
- If time permits we would like to be able to predict the locations and types of disasters. But firstly we are focusing on predicting the amount.
- For our analysis, we will be employing complex neural network models

Data Description.

- Our data set is a combination of 6 different data sets.
- It is a public dataset provided by the IMF in collaboration with many other trusted organizations.
- 5 of the data sets were used to collect climate change predictor measurements: Forest and Carbon, Land Cover, CO Concentrations, Sea Level, and Surface Temperature.
- The sixth data set contains the number of natural disasters that have occured by country as well as globally.
    - It also contains columns that specifically kept track of the types of natural disasters example: drought, flood.

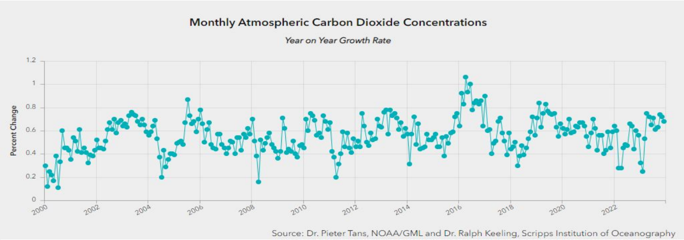
Dataset Structure

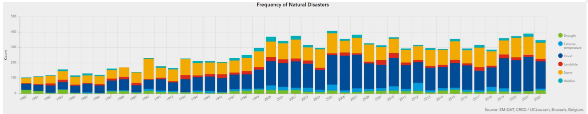
The dataset is comprised of 930 rows and 74 columns.
It was constructed from 6 different dataset. It was combined using the country and date/year data set.
25 of the columns are for the sea/ocean sea level changes 30 of the columns are for the surface temperatures. These columns will be condensed down to 1 column each. 1 will be an average for the closest sea level changes, and 1 will be the surface temperature for that specific country.

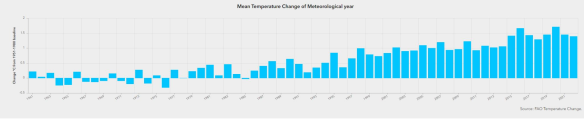
Data Cleaning.

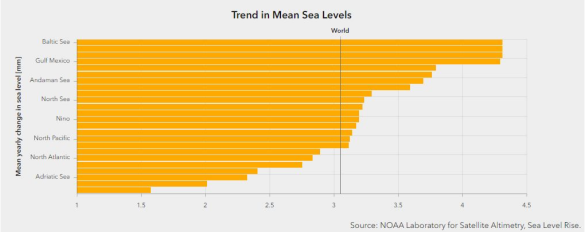
- We first had to cut down the data sets so that they only were the years that they shared (earliest 1963, common: 1992-2020).
- Filtering by ‘world’ column value.
- Pivot the data sets so that the rows were years and the columns represented the required metrics and for some datasets the cell values was the mean average for that year.
- Combined all 6 of the datasets to make one completed set.
- This left us with 30 row entries.
- Expanded the dataset by filtering for the top 30 countries by “Total Natural Disaster”.
- Will train on the countries and then try to estimate the global total per year.

EDA Visualizations.





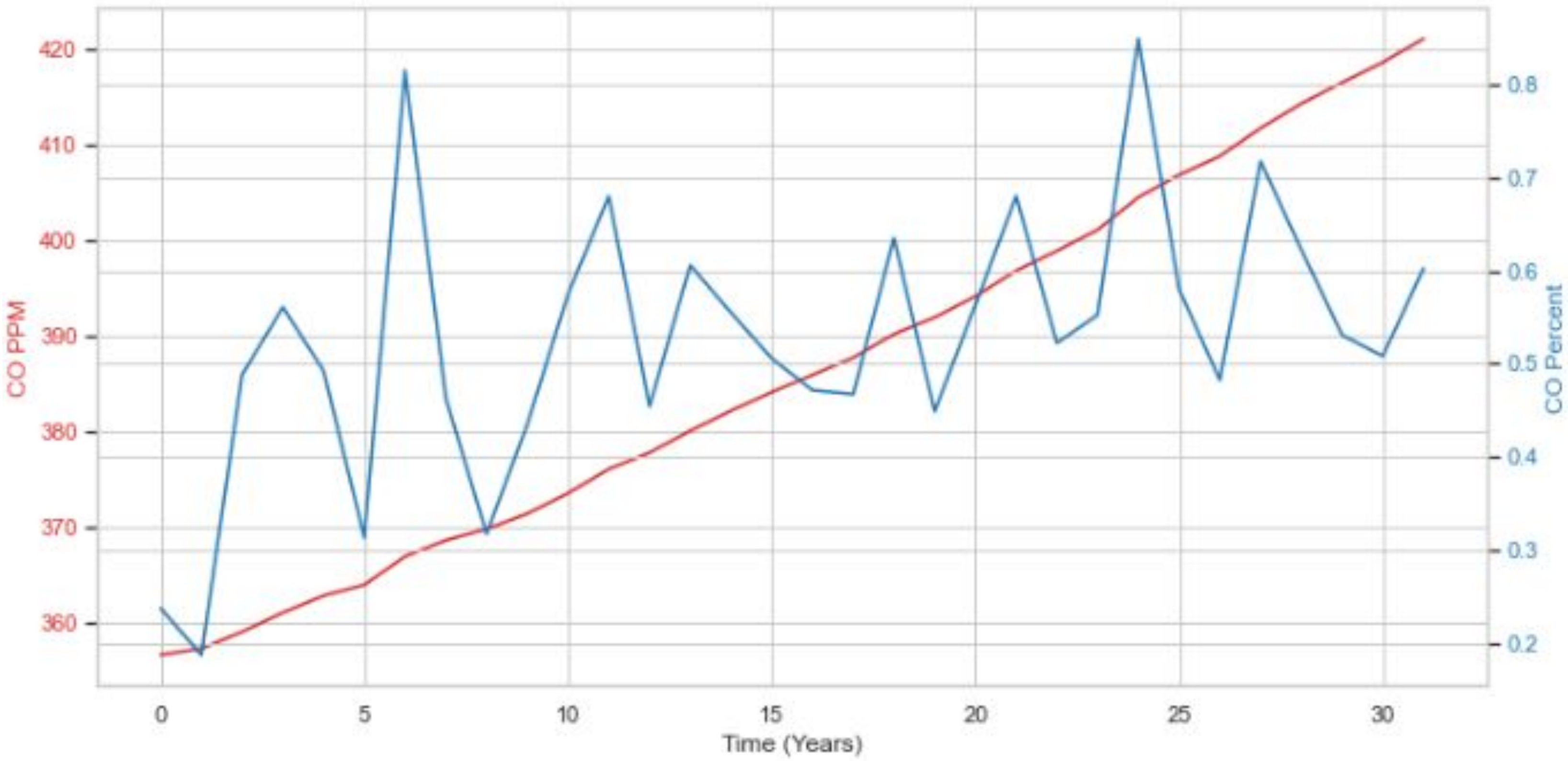




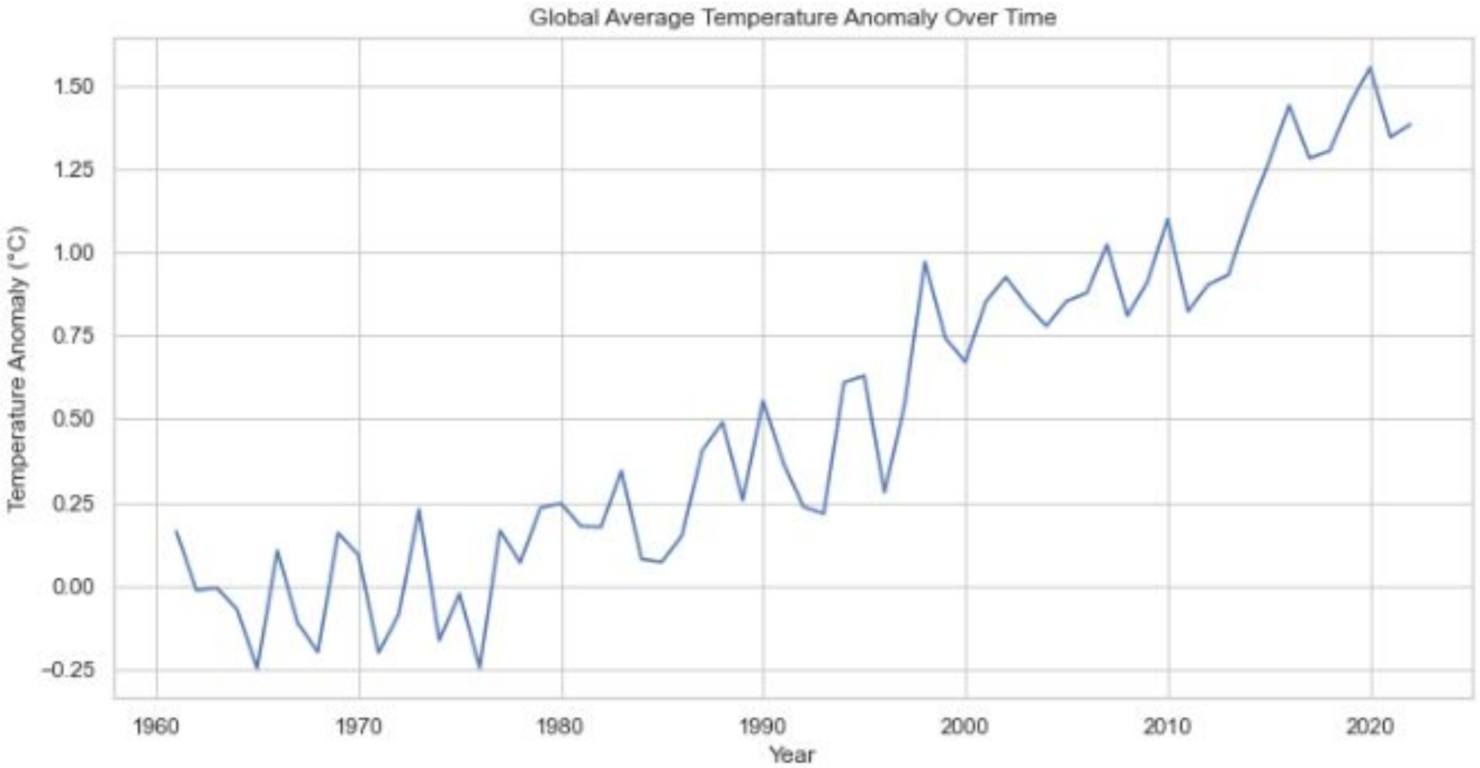
Note: It has been performed on jupyter notebook as well as R. Our end aim is to submit a RShiny web app for visualisations of all datasets (already developed, enhancement of visualisations is work in progress).

Graphs on this page are from: https://climatedata.imf.org/pages/climatechange-data

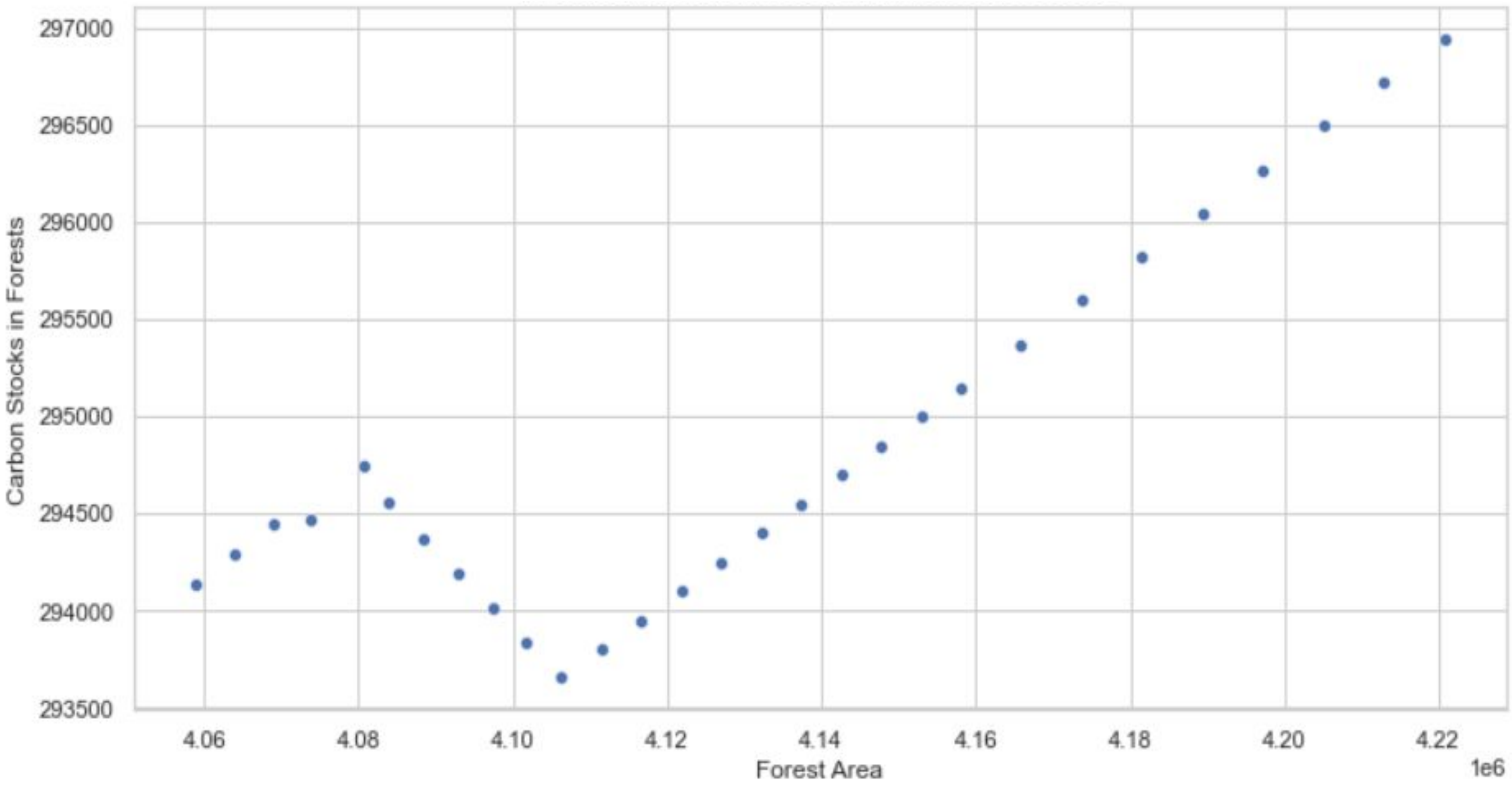
CO Concentrations.



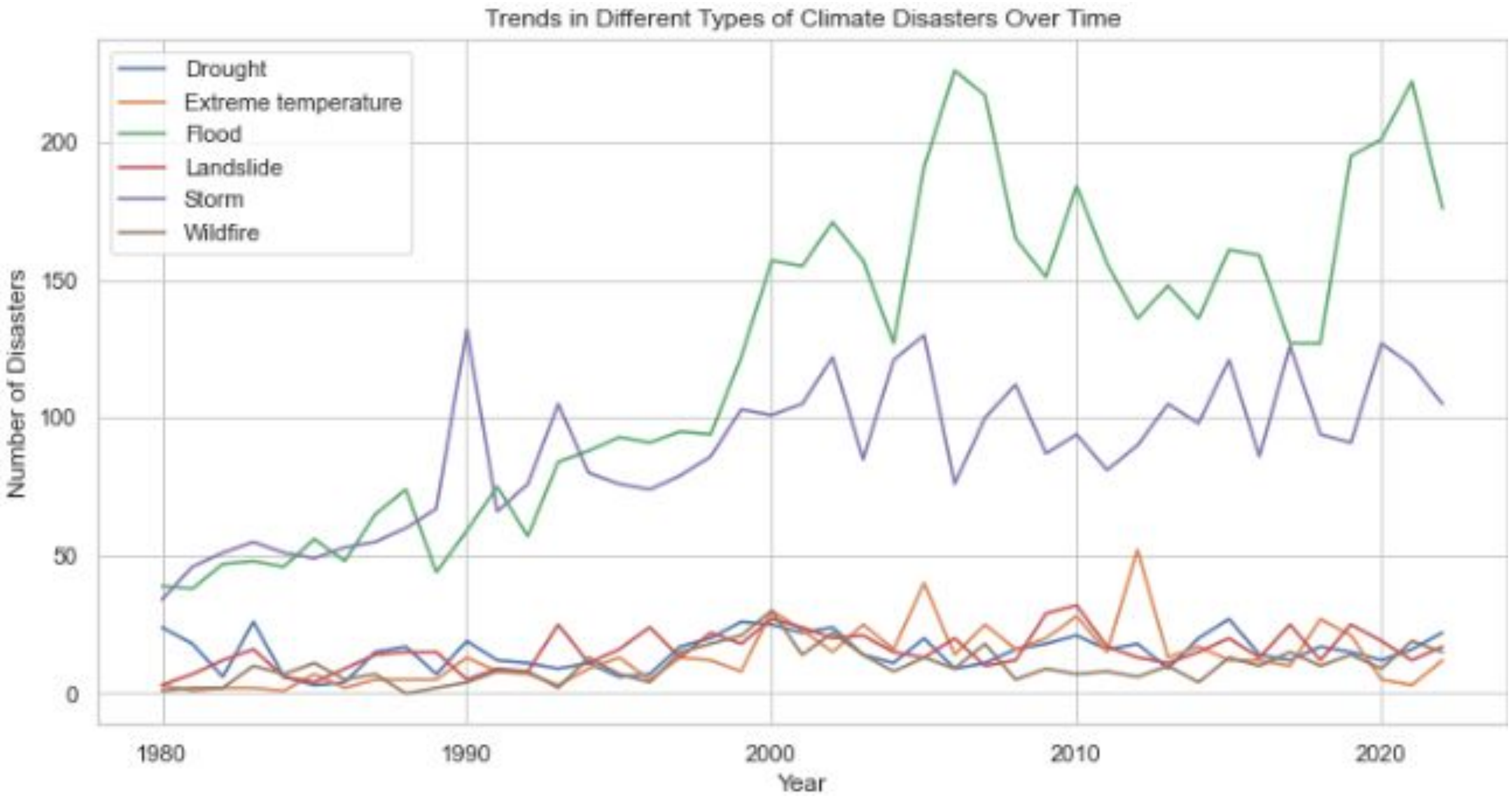
Surface Temperature Change.



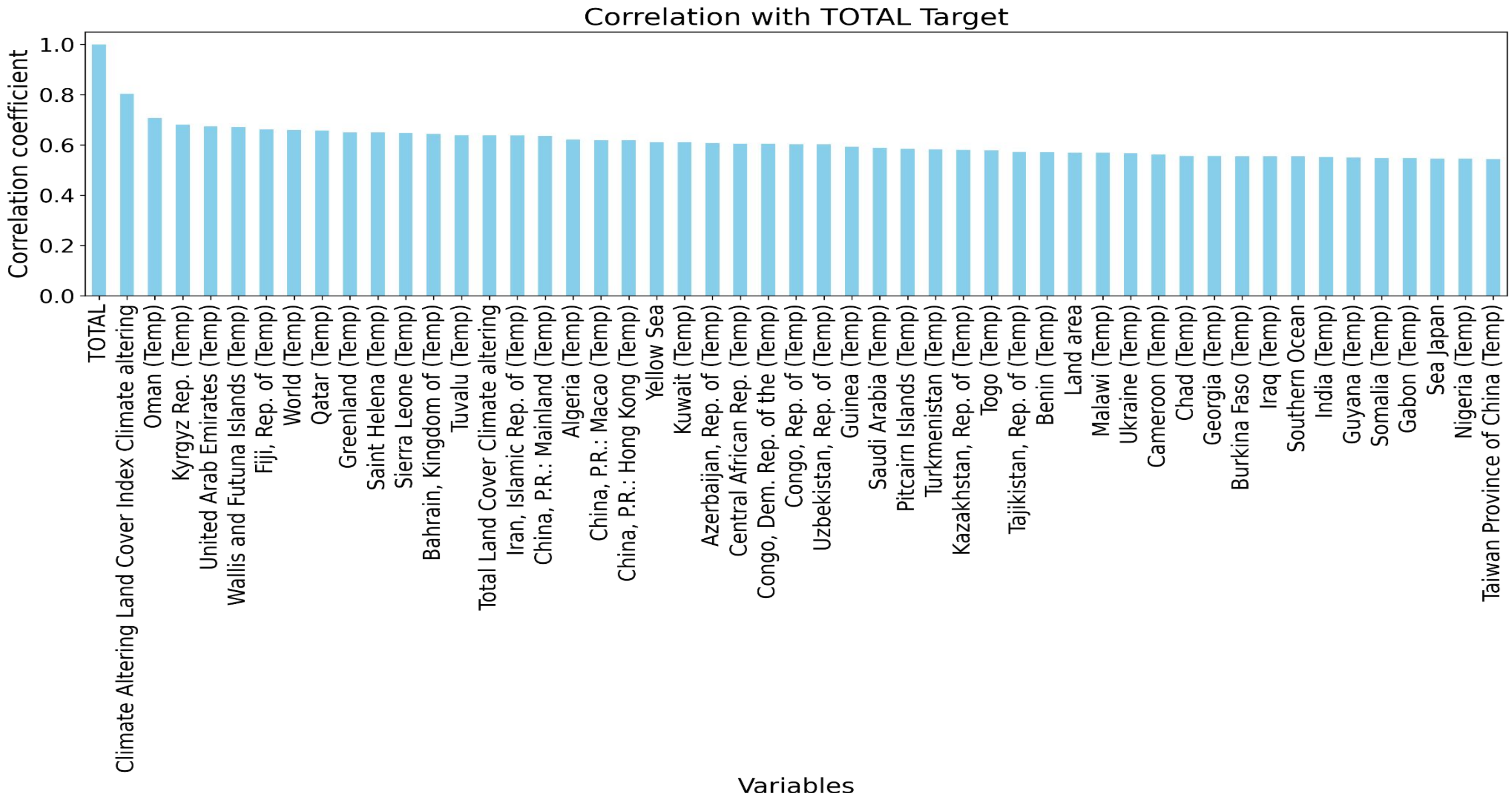
Relationship between Forest area and Carbon Stocks.



Natural Disasters Over Time.



Feature Importance

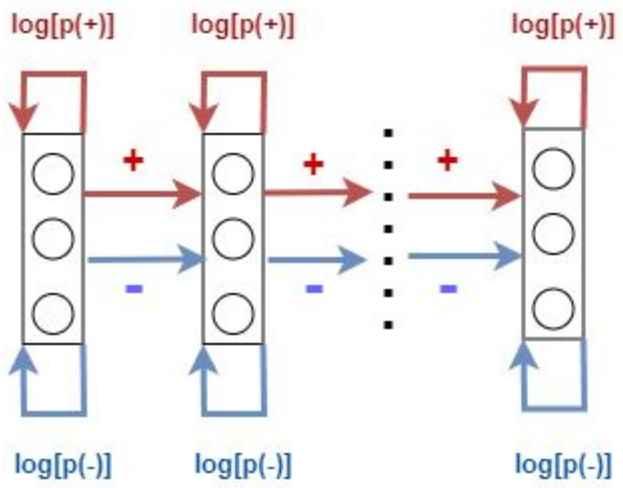


Neural Networks.

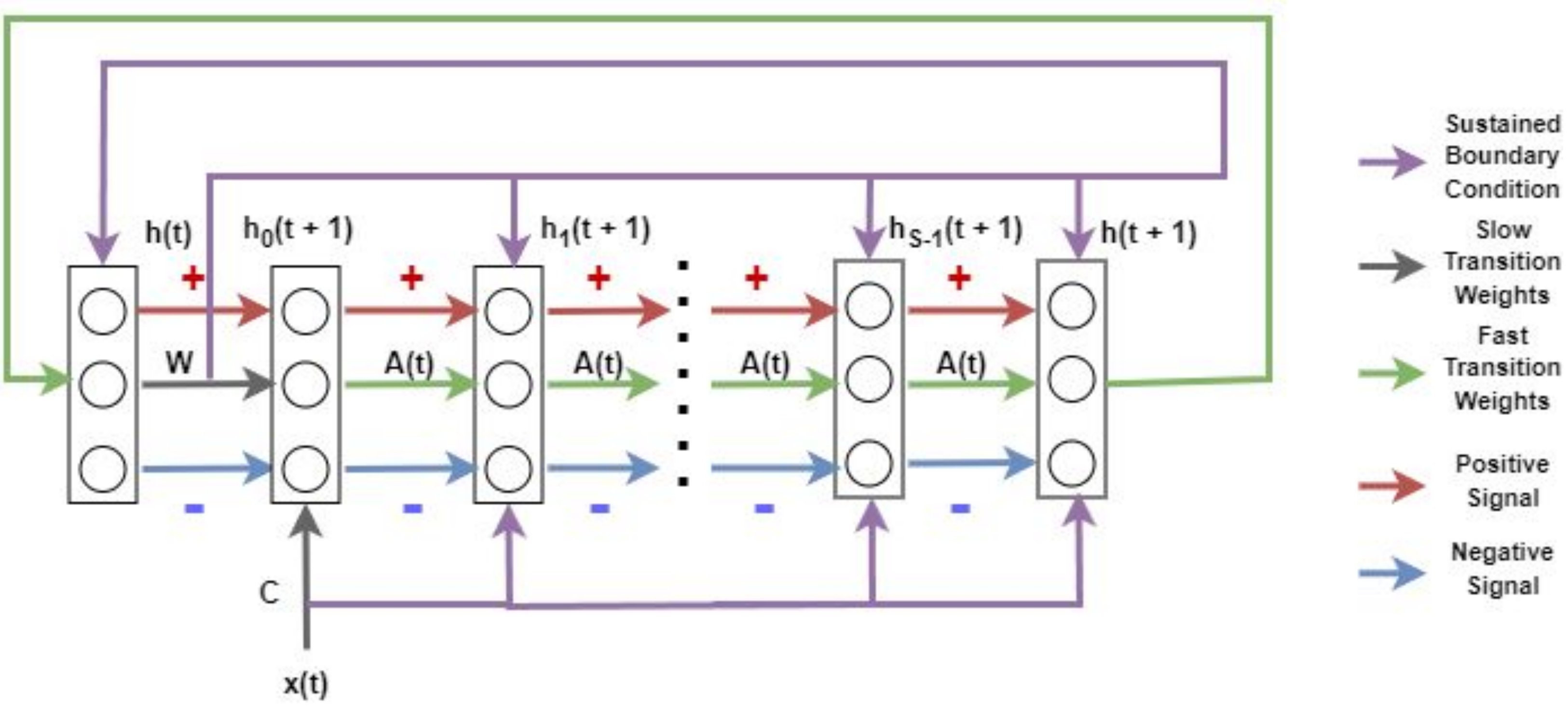
- We are going to be build an LSTM, Forward-Forward RNN, and Attention Neural Networks.
- Our Goal is to compare and contrast how these different architectures perform on the data set that we constructed.
- We will be performing a basic regression model to act as a base-case for our comparisons.
- LTSM and Attention Neural Networks are two of the leading models in Data Science.
- The Forward-Forward RNN is a more theoretical architecture in comparison to the others that doesn't use back propagation.

Supervised Forward-Forward Algorithm (Fall Semester)

- Uses two forward passes with positive and negative goodness to classify inputs
- Each layer is trained independently where positive goodness is maximized while negative goodness is minimized
- Hot encoding is used to corrupt or classify inputs for training
- Training on MNIST yielded a 98% accuracy with hand-tuning.
- Earlier layers doesn’t learn from later layers because there is no backward correction of errors. An RNN would need to be built to fix this issue.



Forward-Forward RNN Architecture (Part One)

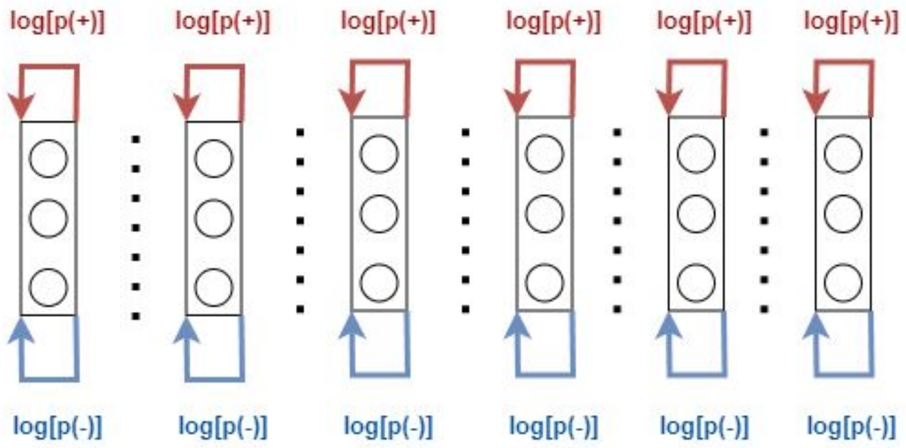


Forward-Forward RNN Architecture (Part Two)

- Extending the Forward-Forward Algorithm by applying Using Fast Weights to Attend the Past.
- RNN architecture will require tweaking whether fast weight is dropped after step interval or is retained.
- Not sure whether a secondary fast weight and h(t) is needed for the negative weight or if the positive and negative signal can share the weight and h(t)
- Initial experiments will use previous forward-forward (supervised) method paired with new architecture before adjusting architecture accordingly.

Forward-Forward RNN Architecture (Part Three)

- Fast learning will be tested and researched as a secondary extension.
- Fast learning will allow the goodness function to be trained independently from RNN circuit activities.
- Previous experiment confirmed hidden activities passed to proceeding layers is independent from goodness function being trained.



Conclusion/Questions/Future Plans.

- Decide the basic model and 1-2 extra advanced models so that we can compare their performance.
- Accordingly do data manipulations (if required).
- For now we have enough data, with a lot of features. Hence, we are sure of not lacking in the database required for building a model for our problem statement.
- We have also created a webapp using RShiny, that will display EDA on our datasets (will be included in GitHub Repo).
- Mapping Water bodies to nearest Land mass.