Impact of Climate Change on Global Food Supply

Aditya Gupta (E19CSE311), Anuj Agarwal (E19CSE327), Priyanka Mahawar (E19CSE183), Vishrut Sharma (E19CSE295)

***Abstract*: Droughts, fires, storms, and floods have become vigorous and more periodical, these repercussions of climate variability have become progressively visible. The global ecosystem is changing, including the environmental resources and cultivation on which we are heavily dependent. Hence climate change has become one of the most prominent challenges faced by humanity, and we, as machine learning enthusiasts, aim to play our part in it. Under our project, we will be exploring the role of machine learning in mitigating the adverse effects of climate change and helping society adapt to these changes. Our primary goal is to understand the changing climate and predict its long-term effect on the global food supply/production. In this study, we also saw that linear regression models are less accurate at capturing the variability of crop yield than non linear models.**

I. Introduction

We started collecting data on the temperature of the earth from the 1850s and the alarming fact is that the rate at which temperature is increasing is the fastest ever. It is expected that by the end of the 21st century the temperature rise would be somewhere between 1.5 to 5.7°C. The emission of greenhouse gases such as methane, water vapor, ozone, carbon dioxide, and chemicals like chlorofluorocarbons (CFCs) is known as Global Warming. Most of these gases are released in the atmosphere from airplane exhaust, fossil fuel extraction, car exhaust, and factory farming, these gases warm the earth by trapping the heat in the atmosphere as they have an insulating effect.

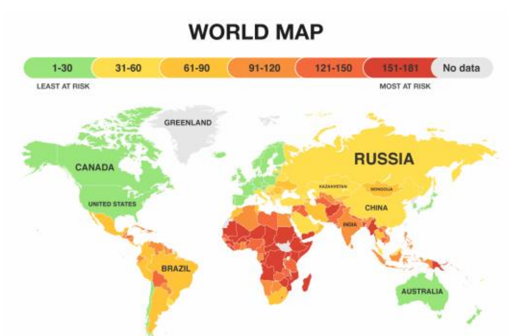


Figure shows 181 nations on their vulnerability and how ready they are to adapt to a warming planet.

Global warming at its pinnacle can be the cause that leads to an increase in the frequency of extreme meteorological events such as floods, droughts that affects food production around the world. Over a long period, the small differences in average temperature when compounded result in disproportionately huge changes in the recurrence of these extreme events.

Despite the prodigious advancements in science and crop yield potential, food production remains greatly dependent on climate, because temperature, solar radiation, and precipitation are the main drivers in the cultivation of crops. Pest infestations and plant diseases, as well as the demand for and supply of irrigation water, are governed by climate.

Hence, it is clear later or sooner the buildup of greenhouse gases in the atmosphere remains ignored and unchecked it will warm the surface of the earth. The biophysical processes of respiration and photosynthesis, insects, the infestations of weeds and, the whole hydrological and thermal dominion governing our agricultural systems are expected to be affected by such a warming trend.

Under our project, we use statistical methods to understand the effect of climate variability on global food production/supply in greater depth and propose solutions to tackle the adverse effects caused by climate change and extreme meteorological events.

II. Background Knowledge

Not any particular kind of technology can make the world a better place. Machine learning has given many contributions across different domain areas. ML has played a major role in weather forecasting to satellite imagery. Through remote sensing, ML can enable automatic monitoring (for example, detecting deforestation, gathering data on structures). It has the potential to speed up the process of scientific discovery and also be used to increase the efficiency of systems.

Using ML to combat changes in the climate has the potential to assist everyone while also furthering the discipline of ML. Problems mentioned in this paper like global warming, carbon dioxide emission, food wastage can be controlled or by prior analysis. Furthermore, taking real action on climate issues necessitates collaboration with fields both inside and outside of computer science, which can lead to multidisciplinary methodological advancements like enhanced physics-constrained machine learning approaches.

Climate change is a challenge and a threat to the food security of the whole world. As we all know the emission of greenhouse gases is increasing in the atmosphere every year and therefore due to this greenhouse effect, the temperature is also rising. The average global temperature is predicted to rise by 2 degrees Celsius by the end of the 21st century.

III. Related Work

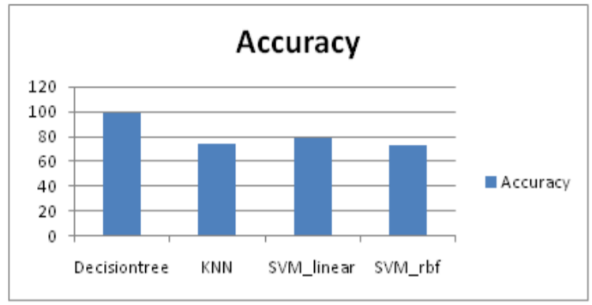
Botanists and agrarian researchers in Pakistan led a few measurable examinations and observed that there is a negative connection between pesticide use and harvest yield. Their examination contains how information mining joined cultivating data including pesticide use, meteorological information, and disturbance investigating can bring about productive pesticide use. Dealing with that, there are many proposed procedures to make a framework to attempt to control crop yield utilizing AI.

A yield forecast adaptation was proposed to utilize the digging methods for classification and expectation, in this model, they entered information and boundaries, for example, land area, soil type, crop name, bug data, environment, and so on, therefore to this, the model anticipated the plant infections, plant yield, and so on

Their examination proceeded to see designs in cultivating manifestations concerning the accessibility of data. K - means technique was utilized to pass on out measures of tainting to outside, the k – nearest neighbor associated with duplicate ordinary precipitation and other climate components, and other environment factors are taken apart utilizing SVM (support vector machine). They utilized a few procedures, for example, a GPS-based innovation for portioning soils, for horticulture data they utilized a 10-fold cross-approval, decision tree classifier strategy to test the informational indexes.

This statistical study uses Supervised Machine Learning Algorithms like **Decision Trees**, **K-Nearest Neighbor (KNN)**, and **Support Vector Machine (SVM)** to estimate soil fertility based on macronutrients and micronutrients levels found in the dataset. The implementation of supervised Machine Learning algorithms is tested with the test dataset and applied to the training dataset. The R Tool is used to create the algorithms.

The performance of these algorithms was evaluated a range of evaluation metrics like **Mean absolute error**, **cross-validation**, and correctness. The result of this analysis reveals that the best accuracy of 99 percent is achieved by using a decision tree with a very low mean square error (MSE).



Comparison graph for Result Analysis based on Accuracy

Using the above model as a base, there have been numerous studies on how to use and create a model using many structures such as decision tree, support vector machine, KNN, Random Forest, etc. to determine the impact of climate change on food supply globally.

IV. Proposed Methodology

A. About Dataset

Information for crop yield was assembled from FAOSTAT, which offers worldwide insights on food and agribusiness, and gives admittance to throughout 3 million time-series and cross-sectional information identifying with crop yield.

As rainfall dramatically affects food production, for this task precipitation, each year data was assembled from the World Data Bank notwithstanding the normal temperature for every country. The last information outline for normal precipitation incorporates; nation, year, and normal precipitation each year. The information outline begins from 1985 to 2017, on other hand, the normal temperature information outline incorporates nation, year, and normal recorded temperature. The temperature information outline begins at 1743 and closes in 2013. The variety in years will think twice about gathered information a piece was joining a year reach to exclude any invalid qualities. Information for pesticides was gathered from FAO, it is prominent that it begins in 1990 and closes in 2016. Consolidating these information outlines, it's normal that the year reach will begin from 1990 and close in 2013, which is 23 years of information.

For examining the conduct of the world's surface temperature, we have utilized The Berkeley Earth Surface Temperature Study which joins 1.6 billion temperature reports from 16 prior documents. It is well bundled and considers cutting into fascinating subsets (for instance by country). They distribute the source information and the code for the changes they applied. They likewise use strategies that permit climate perceptions from more limited-time series to be incorporated, which means fewer perceptions should be discarded.

B. *Data Preprocessing*

Data Preprocessing is a strategy that is utilized to convert the crude information into a perfect informational collection. The information is accumulated from various sources, it is gathered in a crude arrangement which isn't doable for the investigation. By applying unique methods like **supplanting missing qualities** and invalid qualities, we can change information into a reasonable organization. The last step on information preprocessing is the partitioning of training and testing information. The information normally will more often than not be parted inconsistent since preparing the model typically needs however many data points as could be expected. The preparation dataset is the underlying dataset used to prepare ML calculations to learn and create the right expectations.

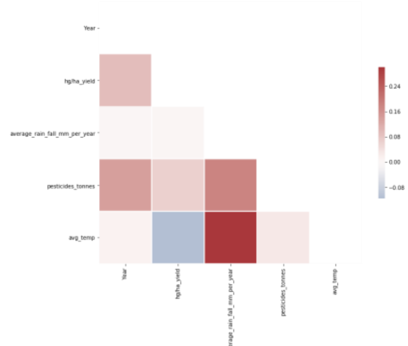
For setting up the harvest yield dataset for displaying, we dropped a couple of sections like Region Code, Area, Thing Code, and so on, which were of no utilization to the examination. Likewise, we renamed the Value attribute to hg/ha_yield to make it clear to perceive that this quality addresses crops yields creation esteem.

In the merged dataset, there are two categorical columns Many ML algorithms cannot operate on categorical data directly, therefore the labeled data must be converted to numeric format. For this, we use **One hot encoding**. Also, this dataset contains highlights with profoundly differing extents, units, and reaches. The highlights with high extents will make an appearance much more in the estimations than highlights with low sizes. To handle this impact, we want to carry all highlights to similar degree of extents. In our analysis, we have used **MinMaxScaler** to tackle this problem.

For the earth surface temperature dataset, we handled the missing data by performing **listwise deletion** because data was missing in chunks and the fact that it was time-series data. We also converted all the date columns to the proper format and created some new features to perform **exploratory analysis**.

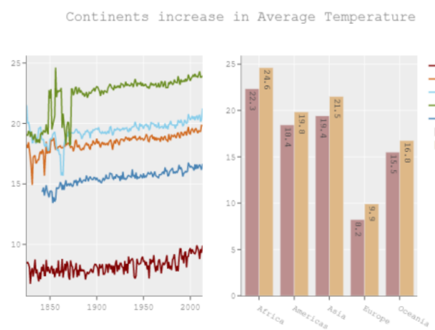
C. *Data Exploration & Feature extraction*

To explore the relationship between the attributes of the dataset we have visualized the **correlation matrix as a heatmap.**



It is clear from the heatmap above that the attributes have no correlations with each other. Therefore, we consider other elements that influence the yield of any crop and its creation. These are essentially the highlights that help in foreseeing the creation of any yield throughout the year. **In our research, we incorporate elements like rainfall, temperature, surface area, moistness, and wind speed.**

To understand the behavior of the earth's surface temperature, we grouped the data by continents and plotted the average land temperature from the year 1750 to 2010.



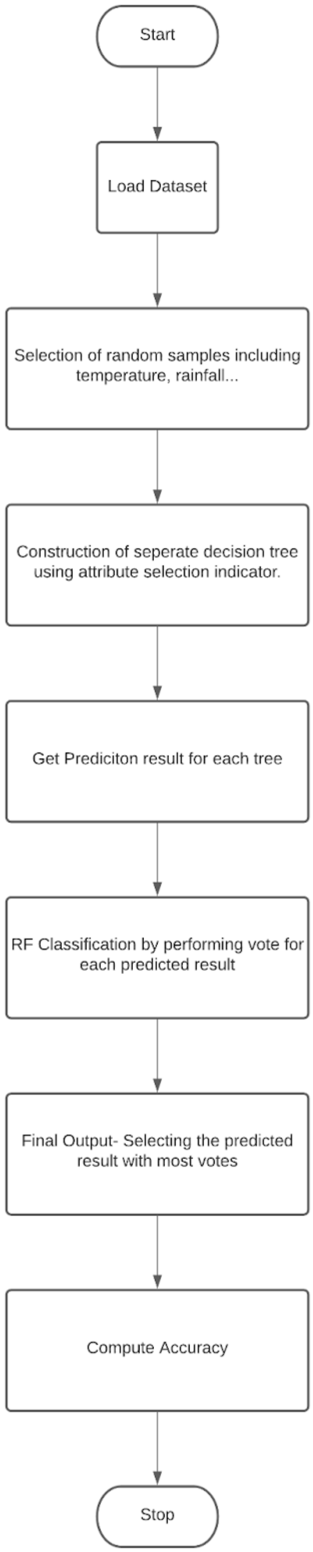
D. *Model comparison and selection*

The best technique to get answers for the problem of crop yield is ML. First evaluate and compare the model which will best compliment our data set post which we will decide the best model to use for our dataset.
There are a lot of machine learning algorithms available for predicting the yield of crop. In this paper we include the following machine learning algorithms for selection and accuracy comparison:

1) **Gradient Boosting** is an AI strategy utilized in relapse and order errands, among others. It gives a forecast model as an outfit of frail expectation models, which are normally choice trees. When a choice tree is a feeble student, the subsequent calculation is called gradient boosting tree. 89% accuracy is achieved when this algorithm is applied to our dataset.

2) **Random Forest** can break down crop development identified with the current climatic conditions and biophysical change. Random forest the calculation makes choice trees on various information tests and afterward foresee the information from every subset and afterward by casting a ballot gives the better answer for the Framework. Forest uses the stowing strategy to train the information which expands the precision of the Result. 68% accuracy is achieved when this algorithm is applied to our dataset.

3) **Decision Tree** is a unique gadget that uses a flowchart-like tree structure or is a model of decisions and all of their utility, expected results, input costs including results.96% accuracy is achieved when this algorithm is applied to our dataset



Flowchart for Random Forest Model

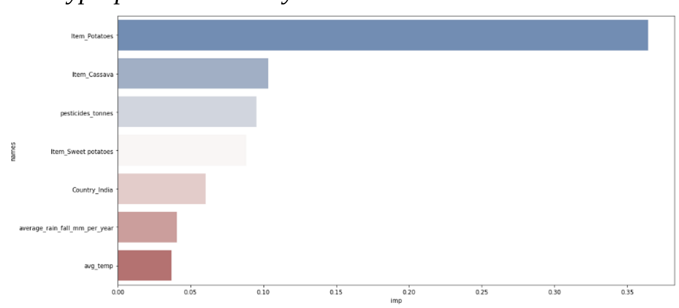
Decision Tree Algorithm is used in both Regression and classification related problems because of which it is used in many competitions and hence has become one of the most used algorithms in Machine Learning.

A decision tree normally begins with a solitary hub, which branches into potential results. Every one of those results prompts extra hubs, which branch off into different conceivable outcomes. A decision hub, addressed by a square, demonstrates a decision to be made, and an end hub shows the result of a decision way. A hub addresses a solitary information variable (X) and a split point on that factor, accepting the variable is numeric. The leaf hubs (additionally called terminal hubs) of the tree contain a yield variable (y) which is utilized to make an expectation.

The decision tree is showing up at a gauge by posing a progression of inquiries to the information, each question limiting our potential qualities until the model gets sufficiently sure to make a solitary forecast. The request for the inquiry just as their substance is being controlled by the model. Likewise, the inquiries posed are all in a Valid/Bogus structure. Decision trees relapse utilizes mean squared mistake (MSE) to choose to part a hub into at least two sub-hubs.

V. Result and Analysis

A. *Hyperparameter Analysis*

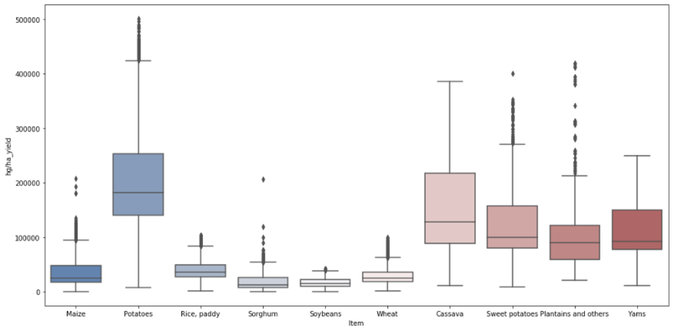


The yield of potatoes gives most elevated significance in the model for making the decision, it the dataset it is the most noteworthy.

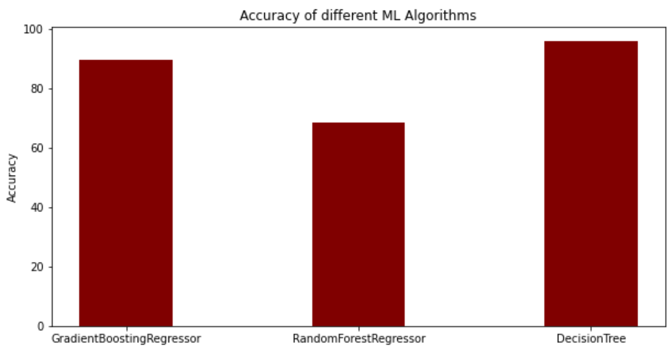
Cassava too, then, exactly as expected we see the effect of pesticides, where it's the third most critical part, and a short time later if the collect is sweet potatoes, we see likely the most raised yields in features importance in the dataset.

On the off chance that the collect is filled in India, it looks good since India has the greatest yields total in the dataset. Then, comes precipitation and temperature. The central assumption about these components was correct, where they all basically influence the yield of the ordinary collect in the model.

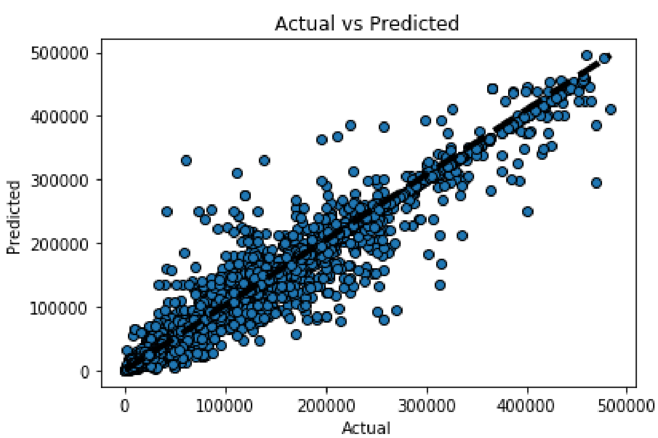
The boxplot beneath showcases yield for that is present in the dataset. Potatoes are the most noteworthy, Cassava, yams, and sweet potatoes.



B. *Classifiers Used*



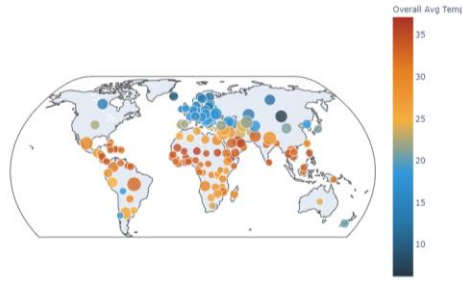
Gradient Boosting Regressor was the Machine learning classifier that was used for accuracy, comparison and prediction along with **Random Forest** and **Decision Tree**. These three classifiers were trained on the dataset and a comparison graph was plotted to showcase the performance of the models. Shown above in the figure showcases the performance. Of the three classifiers used, **Decision Tree Regressor resulted in high accuracy.**



The figure above shows the integrity of fit with the forecasts imagined as a line. It very well may be seen that the R^2 score is phenomenal. This implies that we have observed a decent fitting model to foresee the harvests yield an incentive for a specific country. Adding more elements, similar to environment information; wind and contamination, the monetary circumstance of a given nation, etc. will most likely upgrade the model's forecasts.

VI. Conclusion

Under this study, we were able to see that climate change has a significant impact on crop yield in just more ways than one.



In the image above, we can see the hotspots around the world where the current trend of increasing carbon pollution coincides with the rising median temperatures. The study also shows the trend of where this will lead if this issue is not addressed soon.

The impact of rising temperature on food production around the world is significant but it is auxiliary when compared to other environmental factors such as rainfall and soil fertility.

```
['GradientBoostingRegressor', 0.8965731164462923]
['RandomForestRegressor', 0.6842532317855172]
['SVR', -0.20353376480360752]
['DecisionTreeRegressor', 0.9600505886193001]
```

In this study, we also saw that linear regression models are less accurate at capturing the variability of crop yield than non-linear models. We saw that Decision Trees gave the best results, with an accuracy of 96%.

VII. References

1. David Rolnick1, Priya L. Donti2 and Lynn H. Kaack (2019). Tackling Climate Change with Machine Learning.

2. Sadrach Pierre, Ph.D. (2019). Climate Change and Food Scarcity in Developing Regions.

3. Cynthia Rosenzweig, Ana Iglesius, X.B.Yang, Paul R. Epstein and Eric Chivian (2001). Climate change and extreme weather events.

4. Aman Kharwal (2020). Predict Weather with Machine Learning.

5. Jayalakshmi and Savitha Devi (2021). Predictive Model construction for prediction of soil fertility using ML.