Observing agricultural trends and predicting yield change in crops affected by climate change over time

Aayush Kamath (ak4808)

December 16, 2022

Contents

1 Abstract

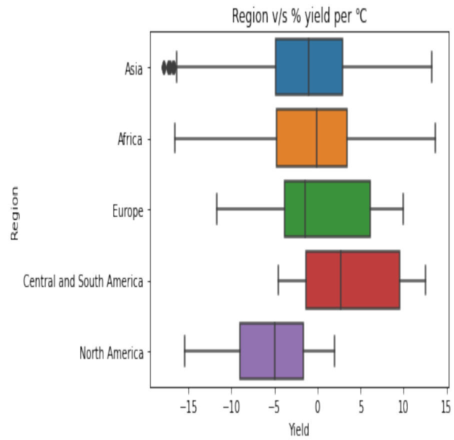
Climate Change is impacting all aspects of our lives. Even if majority indicators of changes in our life due to climate change might have been subtle so far, the impacts are going to be more pronounced in the upcoming decades in the form of a possible food crisis, if not tackled effectively. The term climate change is a broad blanket of various indicators and it is necessary to examine these specific indicators and their impacts on a global scale that is beyond just the surface level. Through this project, I have looked at climate change simulation data where related factors such as temperature, latitude, longitude, agricultural practices (Fertiliser,Irrigation,etc.) play a part in impacting the yield of crops across time. Analyzing past data and future data obtained through simulation, could help with determining certain trends which can then be used to come up with key steps to tackle the problem at hand. It is imperative for us to determine not just the factors but the extent to which these factors have impacted crop production over the years. Using ML models this project tries to look at the climate change impact on crop yield of four major crops and make inferences with regards to the findings of this study.

2 Introduction

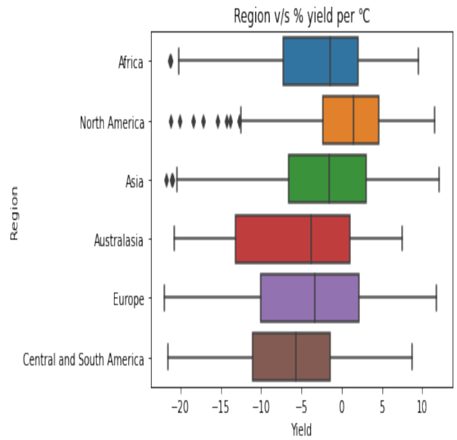
Numerous factors impact agriculture and it is necessary that we remain up to date with the negative impacts caused by climate change on crop yield. With increasing population, rising global temperatures and decreasing farming area and crop yield, we could be looking at a logistical nightmare with severe gaps in supply and demand of food in the near future. This makes it imperative for us to globally examine the detrimental aspects of climate change on crops and figure out ways to either reverse or significantly dampen the rapidly decreasing yield of crops. The challenge lies in having a data that gives an overview of global trends.

A significant chunk of climate related studies such as [3] Chen, X. et al. and [6] S. Jambekar et. al. focus on a specific region and its agricultural data. A more generalized approach is usually lacking where machine learning techniques can be applied on a global consolidated data. Hasegawa, T. et. al. [1] have created a dataset which consolidates decades worth of climate scenarios impacting crop yields in a given region and predicting the future trends for these scenarios. In this dataset, each row represents a climate scenario simulation generating future trends based on existing data for a particular crop in a given region. Applying ML techniques on this data set is the basis of this study.

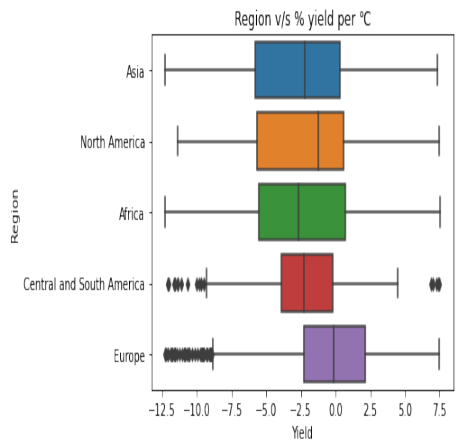
Existing literature suggests a variety of regression techniques for the given problem. [8] Khosla, E. et. al., [2] Rehana, S. et. al. and [11] Shah, A. et. al. had support vector regression as the model with the best performance. [11] Shah, A. et. al. had a ϵ value of 0.05 suggesting that the scope for error allowed with no penalty was minimal. Whereas, Random Forest performed significantly better in case of [5] Elavarasan, D. et. al. and [7] A. Nigam et. al. for crop yield prediction. Based on the approaches implemented in existing literature, this study implements five machine learning techniques namely Support Vector Regression, Random Forest Regression, XGBoost Regression, Linear Regression and Ridge Regression.



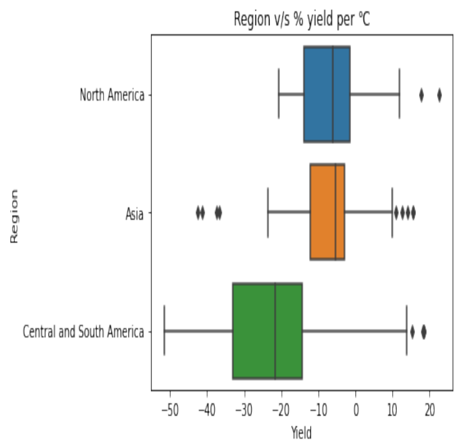
(a) Rice



(b) Wheat



(c) Maize



(d) Soybean

Figure 1: Region wise yield change per °C

3 Data Exploration

We explore data of each of the four crops namely Maize, Wheat, Rice and Soybean individually. Fig. 1. shows a plot for the region wise change in yields for all four crops with respect to per degree Celsius change in temperature. This helps in recognizing the varying behaviour of each crop across the globe. Most crops for most regions see a significant change (decrease) in median yield across thousands of climate scenario simulations. Take for example the median case of soybean in Central and South America suggesting 20% decrease in yield for temperature change of 1 degree Celsius. These plots were obtained after trimming certain outlier cases.

We further explored our data to look for yield changes as per the time slices available to us in our dataset as well as extrapolating all scenarios for recognizing agricultural trends in 2050's (mean future mid point value for all scenarios in our data). The 3 different time slices are 'MC', 'NF' and 'NC' which represent Mid-Century, Near-Future and End-Of-Century time periods respectively. Each simulation of the dataset belongs to either one of the aforementioned time slices. For the baseline point of 2005 to the future midpoint of 2050's, there was decrease in yield for wheat, soybean and maize across regions. The decrease was limited to a range of 0-10% for the most part for wheat and maize with the decrease in soybean yield suggesting severe impacts of climate change (range of 13-31% decrease across regions). Rice saw a concerning drop of 11% in simulations for North America while remaining fairly stable for the rest. Exploration with regards to respective time slices showed an increment in the magnitude of decrease of yield over time periods. The end of century projections especially remain concerning for all crops barring rice which is relatively better but lacks a positive trend in and of itself. The near future time slice (time frames from 2005 to 2030's) show a marginal decrease in yield but the impact of climate change starts to worsen mid-century onward.

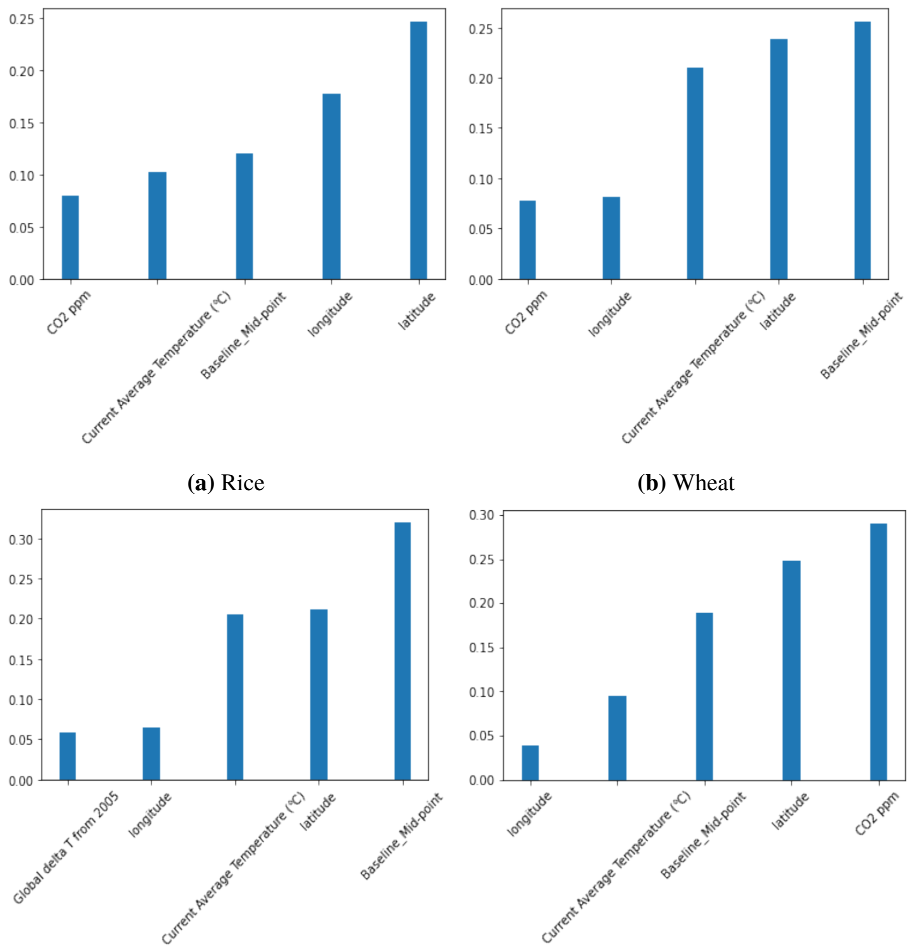
4 Methods

The dataset used has 8000 rows with each row representing a simulation of a given climate scenario. The columns range from the region, temperature, yield, prevalent farming practices to the chosen climate scenario and its time slice. The idea of this study was to apply regression techniques for predicting the %change in yield for per degree Celsius change in temperature for the given data. Each of the four crops were separately put through the modeling phase as data showed significant variation from crop to crop while exploration. We trained and tested the crop data on five regression techniques (Table 1). Linear Regression, Ridge Regression and Support Vector Regression perform poorly across all crops. SVR's performance can be improved by changing its ϵ value which represents the margin of tolerance of no penalty. In this implementation, we choose the default values to avoid overfitting and to have a fair comparison with rest of the regressor models. For ridge regression, changing the α barely caused a significant difference in the score of the model. Overall ridge regression performed similar to Linear Regression.

Random Forest and XGBoost performed way better than the rest of the regression techniques for all crops, which is in line with the existing literature. This is because they train on different sets of samples which results in decrease in variance. While maize, wheat and soybean didn't require parameter optimization for the two regressors in question, rice required it. The high variance in data relating to the rice crop causes low test accuracy, especially seen in the drop in test score of the other 3 regressors. To counter this we made changes to number of estimators and max depth of the tree in order to overcome overfitting on train data. There was a huge jump in the R2 score of Random Forests and XGBoost following this. For Maize and Wheat, Random Forest performed better whereas, for Rice and Soybean XGBoost performed better. Hence, these two techniques are significantly better as compared to other methods which is in line with findings of a lot of studies. The only negative outcome was that Support Vector Regression performed poorly which has shown to outperform random forest [11] Shah et. al. in certain cases.

Apart from the R2 score it was necessary to figure out the features that greatly impacted each of the four crops. Figure 2 shows the plots for the top 5 features of each crop. We calculated feature importance using random forest for the same. Maize and Wheat had Baseline Mid point as their most important feature suggesting that the time period when each of the scenarios were simulated held the maximum importance for their prediction of yield. Soybean depended the most on CO2 ppm concentration. Since soybean had the most significant decrease in yield across simulations, CO2 ppm data can be closely looked at for soybean production. Rice depended mostly on latitude and longitude which means there was a significant demarcation in yield based on its region. Adopting suitable farming practices prevailing in the region of better yield could be an approach for regions with low yield.

Across all 4 crops the most common important features were CO2 ppm concentration, Latitude and Longitude, Baseline Midpoint (Baseline midpoint year for simulations), Average Temperature and expected change in global temperature (global delta T).



(a) Rice

(b) Wheat

(c) Maize

(d) Soybean

Figure 2: Top 5 Most Important Features for each crop

5 Future Work

Hybrid Random Forest Regression techniques as well as neural networks can be implemented on the given data to identify better modeling methods. For preprocessing, a comparison of multiple feature engineering techniques can be made for each regression technique. In terms of data exploration, region wise trends of farming techniques such as fertiliser, irrigation, Tillage, Soil organic matter management, etc. can be looked at for their relation to yield changes. This can help with identifying key farming practices that can help in arresting the projected decrease in yield for the crops. This study as well as most of the existing literature focuses on 4 major crops where we saw that the data and trends vary significantly from crop to crop. Hence, we need to identify more crops that are being impacted due to climate change and examine their trends.

Models	Crops			
	Maize	Rice	Soybean	Wheat
Linear Regression	0.387	0.071	0.334	0.278
Ridge Regression	0.387	0.071	0.329	0.278
Support Vector Machines	0.396	0.099	0.082	0.272
Random Forest	0.926	0.909	0.923	0.919
XGBoost	0.874	0.922	0.945	0.877

Table 1: R2 Score of Crops on Test Data

6 References

[1] Hasegawa, T., Wakatsuki, H., Ju, H. et al. A global dataset for the projected impacts of climate change on four major crops. Sci Data 9, 58 (2022). https://doi.org/10.1038/s41597-022-01150-7

[2] Rehana, S. (2019). River Water Temperature Modelling Under Climate Change Using Support Vector Regression. In: Singh, S., Dhanya, C. (eds) Hydrology in a Changing World. Springer Water. Springer, Cham. https://doi.org/10.1007/978-3-030-02197-9_8

[3] Chen, X. et al. The effects of projected climate change and extreme climate on maize and rice in the Yangtze River Basin, China. Agric. For. Meteorol. 282–283, 107867 (2020).

[4] Liang, S. et al. Modeling crop yield and nitrogen use efficiency in wheat and maize production systems under future climate change. Nutr. Cycl. Agroecosystems 115, 117–136 (2019).

[5] Elavarasan, D., Vincent, P.M.D.R. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. J Ambient Intell Human Comput 12, 10009–10022 (2021). https://doi.org/10.1007/s12652-020-02752-y

[6] S. Jambekar, S. Nema and Z. Saquib, "Prediction of Crop Production in India Using Data Mining Techniques," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697446.

[7] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125-130, doi: 10.1109/ICIIP47207.2019.8985951.

[8] Khosla, E., Dharavath, R. Priya, R. Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environ Dev Sustain 22, 5687–5708 (2020). https://doi.org/10.1007/s10668-019-00445-x

[9] Rahimi-Moghaddam, S., Kambouzia, J. Deihimfard, R. Adaptation strategies to lessen negative impact of climate change on grain maize under hot climatic conditions: A model-based assessment. Agric. For. Meteorol. 253–254, 1–14 (2018).

[10] Iizumi, T. et al. Responses of crop yield growth to global temperature and socioeconomic changes. Sci. Rep. 7, 1–10 (2017).

[11] Shah, A., Dubey, A., Hemnani, V., Gala, D., Kalbande, D.R. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. In: Vasudevan, H., Deshmukh, A., Ray, K. (eds) Proceedings of International Conference on Wireless Communication . Lecture Notes on Data Engineering and Communications Technologies, vol 19. Springer, Singapore. https://doi.org/10.1007/978-981-10-8339-6$_6$