TECHNOLOGICAL ANALYSIS OF THE IMPACT OF CLIMATE CHANGE ON FOOD PRODUCTION

(A CASE STUDY OF UNITED STATES OF AMERICA)

By

TOLULOPE OLUSEGUN GBOKOYI

A Dissertation

Submitted to the Cardiff School of Technology

Cardiff Metropolitan University

In partial fulfilment of the Requirement for

The Degree of Master of Science in

Data Science

August 2023

Supervised by Dr. Wan Nor Ashikin

DECLARATION

I hereby declare that this dissertation titled " Technological Analysis of the Impact of Climate Change on food production (USA as a case study)" is entirely my own work and has never been submitted nor is it currently being submitted for any other degree.

…………………………………….. ……………………………….

Tolulope Olusegun Gbokoyi Dr Wan Nor Ashikin

(Student) (Supervisor)

ABSTRACT

The vital connection between climate change and food production in the US is investigated in this dissertation, with a particular emphasis on important crops including Maize, Rice, Wheat, and Potato. It is essential to understand how meteorological factors affect the yields of these main food crops from 1980 to 2021 since climate change poses enormous threats to agricultural sustainability.

The goal of the study is to address the three fundamental research questions. Firstly, it investigates the variation in meteorological factors (Solar radiation, Temperature, Precipitation, and Atmospheric CO_2 concentration) and the yields of the crops above. By analyzing comprehensive historical datasets, the research highlights the trends and patterns that have emerged over time. Secondly, in order to develop significant Correlations and comprehend how varying climatic trends interact with food production, the research explores the relationships between climatic factors and crop yields. Thirdly, the study explores the application of Machine Learning (ML) algorithms, specifically Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR), to predict crop yields. The comparison of these ML algorithms' performance metrics aims to identify the most suitable approach for yield prediction in the context of climate change and its impact on food production.

This is achieved through the collection of secondary datasets from reliable sources comprising historical crop yield and climatic data. Data science tools and techniques were used for rigorous cleaning, pre-processing and analysis. The ML models will be trained and optimized to produce reliable yield predictions. The findings would provide insight into how climate change is affecting food production in the US and deepen our understanding of the relationships between technology, climatic variability, and agricultural sustainability. The ML prediction model may help all agricultural stakeholders make informed decisions to promote food security and resilience in the face of a changing climate. In conclusion, this dissertation contributes to the field of Data Science and environmental studies by shedding light on the interplay between climate change and food production in the United States.

ACKNOWLEDGEMENT

I am extremely grateful to God and the numerous people who have contributed to the completion of my Master's dissertation because it has been an amazing journey.

First and foremost, I want to express my profound gratitude to my esteemed supervisor Dr Wan Nor Ashikin, whose constant support, insightful observations, and constructive criticism were crucial in shaping this research. Your guidance and support have not only improved the quality of my writing but have also deepened my understanding of the subject.

I would like to express my heartfelt appreciation to the Professors and Researchers at Cardiff School of Technology of Cardiff Metropolitan University whose love of learning and dedication to excellence inspired my intellectual curiosity. Your classrooms transformed into a sanctuary of knowledge where I developed as a student.

My heartfelt appreciation goes out to my beloved wife, and wonderful children for their unwavering support and confidence in me. You have been the pillars of strength that have helped me go forward during trying times with your unceasing support and understanding.

I am grateful to my friends and colleagues for their inspiring and intellectual discussions. I've been able to expand my horizons and think critically about my studies because of your diverse perspectives and varying ideas.

Furthermore, I want to extend my sincere gratitude to all the unsung heroes who ceaselessly labour behind the scenes to maintain the efficiency of the academic system and the university.

Thank you from the bottom of my heart to everyone who impacted my life in a lasting way and helped me reach this accomplishment.

Thank you

Tolulope Olusegun Gbokoyi

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Background Study

In the last 50 years, the globe has experienced a continuous change in weather patterns. The alteration in the weather pattern is a major concern for the entire world. Long-term variations in typical weather conditions have an effect on almost all of the ecosystem's main components. Climate change is the long-term alteration of the usual weather patterns that have come to define local, regional, and global climates on Earth according to the National Aeronautics and Space Administration (NASA) (2023). Climate change is among the most urgent problems our world is currently experiencing. It has significant impacts on many facets of human life, including agriculture and food security. The United Nations' 13th Sustainable Development Goal (SDG) for 2023 demonstrates the necessity for immediate intervention to counter its negative effects. Worldwide food production is impacted to some extent by these changes in weather patterns which include precipitation, humidity, evapotranspiration, and temperature. One cannot emphasize enough how much agriculture depends on the climate.

Global food production is closely tied to the United Nations' second sustainable development goal, which seeks to end hunger (Zero Hunger) by 2023. The world's main food crops are staple foods which include maize, wheat, rice, and potatoes since they are necessary for human nourishment. The first 3 crops are considered cereal crops while potato is a tuber crop. Maize is the major global staple food crop accounting for a global production of about 1.223 billion Metric tons as of July 2023 as projected by World Agricultural Production (WAP) (WAP 2023). It is cultivated basically for feeds for poultries, industrial purposes and majorly for food in sub-Saharan Africa. Global wheat production will contribute about 800.19 Metric tons to the global staple food crop production in July 2023 as projected by WAP (2023). When it comes to supplying nutritional calories, protein, and important nutrients including dietary fiber, vitamins, and minerals, wheat is largely used for human consumption (Shewry et al., 2009). Rice is another staple food crop that contributed about 520.77 Million Metric tons in July 2023 as projected by WAP

(2023). There are several ways to eat rice as a staple diet, including boiled rice, rice noodles, rice flour, and rice bran oil. Additionally, it is employed in the manufacture of a number of processed food items Food and Agriculture Organization (FAO) (FAO 2020). Potato is an excellent source of dietary carbohydrates and it is a tuber crop. Based on production and consumption, it ranks as the fourth-most significant food crop globally. The global production of potatoes is 368 million metric tons in 2020 (FAO 2020). According to the United Nations' SDG's (2023), Covid-19, violence, climate change, and rising inequality collaborate together to endanger global agricultural production and security. Climate change is affecting the production of various food crops, which is resulting in a decline in yield and quality.

Climatic factors such as elevated atmospheric Carbon dioxide, changes in precipitation trends and increasing temperature affects crop production biophysically. Global crop output has not been able to catch up with rising food demand, partially due to changes in weather patterns. These meteorological variables have varying impacts on climate change. The impacts of increasing temperatures, changing precipitation, and CO_2 fertilization vary depending on the crop, the area, and the degree of parameter change as stated by Malhi et al. (2021). In addition, the adaptation ability of the type of crops also influences productivity as seen in Iran Karimi et al., (2018). The production of staple food crops is impacted by weather changes both directly and indirectly. A change in agroecological conditions, which affects the growth and yield of these staple food crops, has a direct impact on food production. The four main staple food crops examined in this study, which are grown in various temperate and tropical parts of the world, have been greatly impacted by these extreme climatic changes.

The world of data analysis and prediction has been completely transformed by technical advancements in the analysis of large data using machine learning (ML) techniques and algorithms. It is feasible to effectively extract significant patterns and insights from large datasets with the aid of this ML algorithm, enabling researchers to locate previously unrecognized relationships and make accurate predictions. The ML analysis helps to identify patterns of the meteorological variables and predict the change in climate over the period based on the available dataset from over 30 years. This can help researchers to estimate the possible impact of these changes on crop production. Recent developments have made it possible for ML to progress the state-of-the-art in climate prediction (Rolnick et al., 2022). This offers priceless insight into the dynamics of the complex relationship of how climate change affects crop production. There are quite a number of ML algorithms with different applications in agriculture and crop production. The ML algorithm that would be used for this dissertation is a supervised algorithm and would work according to (Liakos et al., 2018) It predicts the missing outputs (labels) for the test data using the learned knowledge (trained model).

In this research, ML algorithms like Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forests Regression (RFR) are used to analyze the impact of climate change on Crop yield using the United States of America (USA) as a case study. The climatic factors that are considered are Solar Radiation, Temperature, Precipitation and Atmospheric CO_2 Concentration and the Crop yield of the four major staple food crops which are Maize (Corn), Wheat (spring wheat), Rice, and Potato. This analysis is focused on the USA and all the data used both meteorological and agricultural data are USA Data.

1.2 Problem Statement

Effective adaptation and agricultural planning require an understanding of the intricate connection between climate factors and crop productivity. Agricultural production is highly sensitive to weather and climate (Gowda et al., 2018) and the low performance of agriculture has been triggered by climate change (Kogo et. al., 2021). Agricultural yield is influenced by a range of factors including climate conditions, crop type etc. (Anderson et. al., 2020). Scholars have been looking for newer insights and methods to contain these impacts. In this pursuit, Data is considered a very powerful tool, the collection, collation and analysis of data can provide critical insights (Mangal et. al., 2020). Historical climate data and crop yield data are available. Therefore, to understand the type of complex connections and quantify them, an in-depth technological analysis is required, one that combines long-term trend research with exploratory data analysis for hidden connections.

Also, one of the key challenges of crop production is prediction performance (Shook et. al., 2021). Given how crucial it is to global food production, a precise and prompt crop output prediction is essential according to Bali et al. (2022) and Rashid et al. (2021). Consequently, the application of different machine learning algorithms to study the relationship and create a model suitable for predicting crop yields under changing climatic conditions with accurate performance metrics. The lack of knowledge about the intricate connections between these climate change and crop production, also, the lack of accurate crop yield prediction could result in a decline in crop production, hinder the implementation of sustainable agricultural strategies to mitigate these effects, and exacerbate global hunger.

1.3 Research Question

1. What variation in meteorological factors (solar radiation, temperature, precipitation, and Atmospheric CO_2 concentration) as well as crop yields (corn, rice, wheat, and potatoes) have taken place over the previous 41 years in the US?

2. What kind of connections exist between these climatic factors and the yields of these major food crops in the United States over this period (1980s to 2021)?

3. How can ML algorithms (SVR, DTR, and RFR) be used to develop a prediction model for the yield of these crops and compare the performance metrics of the ML to determine the most suitable algorithm for this analysis?

1.4 Aim and Objectives

The aim of this study is to conduct a technological analysis of how climate change is impacting food production, in the US, using different ML algorithms to develop a model for the yields of these crops.

The following Objectives are the focus of this dissertation:

1. To do a long-term analysis of the trend of the climatic variables (Solar Radiation, Temperature, Precipitation and Atmospheric CO_2 Conc.) and yearly crop yield variables (Maize, Rice, Wheat and Potato)
2. To analyze and quantify the strength of the hidden connection between each climatic variable and the yield of the crops using Exploratory Data Analysis (Output Correlation).
3. To apply 3 different ML Algorithms (Support Vector Regression, Decision Tree Regression and Random Forest Regression), develop models and compare the performance metrics of the model to know the most suitable.

1.5 Scope and Limitation

This research is focused on analyzing how specific climatic conditions affect the crop yield of some staple food crops using technological analysis (Machine Learning algorithms). The 4 climatic factors covered by this study are Solar Radiation, Temperature, Precipitation & Atmospheric CO_2 Concentration and the crop yield of 4 staple crops which are Maize, Rice, Wheat, & Potato in the United States of America. All data would be sourced from secondary sources. Given the short time frame of less than 3 months within which this analysis must be done and reported, this study is limited to these 4 meteorological conditions and these 4 staple crops. Also, this study is limited to the study of Spring Wheat and not Winter Wheat.

1.6 Research Contribution

The current research contributes significantly by aiming to close a number of gaps in the technological evaluation of how climate change impacts major food crops. This study, which focuses on maize, rice, wheat, and potato, adds to the scarce past studies about the effects of climatic change on food production. Data science has powerful tools and techniques to identify key Climatic variable that has the highest impact on the crop yield of these major food crops. This analysis can identify the key variables that should be the focus of mitigation and adaptation strategies by highlighting the climate elements that have the most impact on agricultural yields. Also, this study will improve the precision and prediction of models that evaluate how climate change will affect food production. This would make it easier to understand complex relationships and offer more precise projections of crop yields, output changes, and perhaps food security problems.

Furthermore, the non-linear correlations between climate variables and the results of food production can be captured by machine learning techniques. Although linear correlations may be assumed by conventional statistical models, machine learning techniques like Support Vector Machines may handle non-linear patterns and reveal intricate linkages that affect crop productivity. Overall, the use of data science techniques to analyze historical data on climate and food production can help create possible early warning systems for anticipated disruptions in food production caused by climate change. This early warning would aid in the decision-making process for the Agricultural Planning, Management Team, policymakers, farmers, and other stakeholders to take proactive measures in adaptive strategies, resource allocation, and policy intervention. This would lessen the negative impact of climate change on food production.

1.7 Research Structure

This research work is organized in the following pattern. Chapter one is focused on the introduction of the key terms, problem statement, research Aim and objectives, questions, scope and limitation and research contribution. The second chapter is about the review of all existing literature that is related to this study and Chapter three is about the methodology which breaks down the processes involved in the data collection, data pre-processing/cleaning and analysis of this research work. Chapter four focuses on the Result and Evaluation of the analysis which is about achieving the objectives in chapter one and chapter five is the conclusion and recommendation.

CHAPTER TWO

LITERATURE REVIEW

2.0 Literature Review

Globally, a number of things are affected by climate change in varied ways. The production of food and agriculture are two areas where the effects can be seen, particularly in light of the increasing human population and the need for food. Food security on a global scale and sustainable food production is crucial for humanity. Agricultural production is highly sensitive to weather and climate (Gowda et al. 2018). Nearly 60% of the variability in yield is accounted for by climatic factors according to Arya et al, (2020). Hence, there is a need for a critical analysis of the nature and extent of the impact. These impacts on global food security were quantified by Gbegbelegbe (2014). The productivity of important staple crops is impacted by irregularities in climatic fluctuations. Uncertainty exists over the current effects on a variety of sub-national crops and their consequences for food security (Deepak et. al 2019). Scientists are working very hard to come up with innovative solutions to deal with less predictable weather challenges because poor weather conditions in developing countries adversely affect agricultural yields (Raza et al. 2019).

2.1 Climatic Factors

The study of the effects of various climatic conditions on the world has become more popular among scholars as a result of global climate change. Intergovernmental Panel on Climate Change (IPCC), which is one of the leading climate change bodies in their 6th report found that since pre-Industrial times (beginning in 1750), human emissions of greenhouse gases have already warmed the climate by almost 2 degrees Fahrenheit (1.1 degrees Celsius). According to the 3rd and 4th National Climate Assessment Reports by NASA, the future effect of global climate change in the United States predicted that Climate change would continue through this century and beyond, Hurricanes will become stronger and more intense, etc. (NASA 2023). Temperature, precipitation, rainfall, evapotranspiration, extreme weather, carbon dioxide concentration, etc. are some of the main climatic elements which have an effect on agriculture as studied by various researchers.

a). ***Temperature***

The National Oceanic and Atmospheric Administration (NOAA) found out that the earth's average surface temperature has increased by roughly 1 degree Celsius ever since the pre-industrial era (1880-1900) (NOAA 2023). Due to regional causes and climatic variability, this global warming has a distinct impact on various regions of the planet. According to Karl et al (2015) the average annual temperature in the United States has increased by approximately 1.2 degrees Celsius from 1901 to 2014. This rise in the mean temperature is consistent with the patterns in global warming that have been accelerated by human activity as recorded by IPCC. Crop development and potential yield can be impacted by high temperatures because they can shorten crucial crop growth periods, encourage crop disease, and make crops more susceptible to insect pests (Jones & Yosef 2015). Along a similar line Zhao et al. (2017) worked on the analysis of the impact of temperature on the yields of Wheat, Rice, Maize and soybeans and discovered that globally, for every degree increase in mean global temperature, there is a loss of 7.4 ±4.5%, 6.0 ±2.9%, 3.2 ±3.7% and 3.1% of maize, wheat, rice and soybean respectively. These show the adverse effect the change in temperature can have on crop yields.

Wu et al (2021) used the feasible generalized least square (FGLS) model to analyze the impact of climate change on maize yield in China from 1979 to 2016, it was established that temperature has a negative impact on maize yield. The study of the analysis of the impact of temperature on the various rice growth stages was done by Abba et. al (2021) using a statistical approach from 1981 to 2017. It was discovered that at the replanting stage of the vegetative phase, the highest temperature has a detrimental effect on rice crops. Also, the increased maximum temperature at the time of replanting speeds up photosynthesis and shortens the rice crop's life cycle, which lowers rice yield. However, the rapid development of leaves suggests increased rice production, which is influenced favorably by the lowest temperature. This shows that temperature has varying impacts on different stages of rice growth stages.

Potato is a tuber crop and tuber growth is limited above 33°C (Jennings et. al 2020). There seems to be a compelling reason to suggest that extreme temperatures can have an adverse effect on potato yield. The effect of temperature on grains (wheat) is most felt with respect to the presence of irrigation. Rising temperatures of 1, 2, 3, and 4 °C reduced grain yield by 4.8%, 7.8%, 12.0%, and 17.2%, respectively, in comparison to baseline yield when averaged over-irrigation and compost conditions (Ding et al 2021).

b). ***Precipitation***

Precipitation is another significant climatic component that affects food crop output. The world has witnessed varying trends of precipitation over the years. Since the turn of the 20th century, global land precipitation has increased by around 2% (Hulme et. al 1998). Different metrological stations in the United States records this variation in trends which is subjected to regional climatic condition. The mean annual precipitation from 1981 to 2016 in Kentucky USA increased by quantities of 6.312 $mm \cdot a^{-1}$(Bia et. al 2019). The delayed arrival and early cessation of rainfall have caused changes in agricultural productivity patterns in several areas of Nigeria (Ani et al 2022). According to researchers, high levels of precipitation in the form of rainfall and floods, which have a negative impact on crop productivity, make sub-Saharan Africa and some regions of South Asia particularly susceptible to climatic risks (Chawdhery et. al. 2022 Ani et. al, 2022). In terms of precipitation and crop yield, many different models are simulated. It was investigated by Sidhu et al (2023) and the model inferred that there is a strong correlation between the total seasonal precipitation and the number of precipitation days (R2 values of 0.23, 0.68, 0.24 for rice, wheat, pearl millet, and respective values). Furthermore, Lobel et al (2017) established that the relationship between precipitation and crop yield is non-linear as supported by Wang et. al (2017) and the study by Agnolucci et. al (2020) who used a Statistical model along a similar line.

c). ***Atmospheric CO_2 Concentration***

Everyday human activity has contributed to the increase in atmospheric carbon dioxide concentration, which has an impact on crop yields. Research and data from different stations have shown that there has been an unprecedented increase in Atmospheric CO_2 concentration ever since the pre-industrial era. The US is not left out of the global increase in Atmospheric CO_2. In 2020, the Atmospheric CO_2 concentration at Mauna Loa, Hawaii, was 414 ppm, which is about 50% higher than the preindustrial level (Keeling et al 2021). The impact on crop yield seems to vary depending on the crop. A percentage rise in Atmospheric CO_2 concentration reduces rice crop production by 1.3% as studied by Hussain et al (2018) in Pakistan. According to past studies by Guo et al. (2010) which were confirmed by Saddique et al (2020) shows that elevated CO_2 somewhat favours maize crops by increasing maize yields which are done by offsetting the detrimental effects of temperature on maize yields. High levels of atmospheric carbon dioxide have been demonstrated to affect crop output in both favorable and unfavorable ways, according to research.

Some categories of crops according to findings thrive under this condition while others do not (Ani et al 2022). The increase in CO_2 concentration as it affects different categories of crops has been a case study for more than 30 years. Ainsworth et al. (2005), (2008) investigated the effects of increased CO_2 on various groups of crops in their multiple publications. In their 2020 publication, they concluded that there is no consensus about the degree of response of woody crops to elevated CO_2 due to the varying degree of response by the study of two commercial varieties of coffee, which were between 12% and 14.6% (Ghini et al. 2015) and 40% to 45% (Mauney et al., 1994) under the same condition of elevated CO_2 of 550μmol/mol. It is known that it does, however, boost the production of woody crops. This means that it has varying effects on different woody crops. According to studies, rice yield tends to rise by 16.2% on average at high CO_2 levels, with a 95% confidence interval (CI), as opposed to ambient CO_2 levels, which range from 14.5% to 18% (Hu et al 2022).

d). ***Solar Radiation***

The radiant energy from the sun is one of the major sources of renewable energy in the world which is converted into other forms of energy. The intensity of solar radiation that gets to the earth varies depending on different factors like aerosol, cloud cover, longitude and latitude, regional climatic variability etc. According to Zou et. al (2019) significant decreases in worldwide surface solar radiation (3.42 W m2 year-1) occurred from 1951 to 1992 and increases (4.75 W m2 year-1) from 1993 to 2005 were observed. Photosynthesis, the process through which plants make their food, depends heavily on solar energy. The rate of photosynthesis is impacted by changes in the amount of solar radiation that is accessible. Crop photosynthesis, the growth and development of plant organs, and crop yields are all influenced by solar radiation (Yang et al. 2019). At various phases of their growth, crops require differing intensities of solar radiation for photosynthesis which is suitable for that phase.

A key factor influencing prospective crop productivity is solar radiation as established by (Wilson et al 1995) and confirmed by further studies of Hou et al. (2014), Liu et al (2013) and Deng et al (2015). Reduced sun radiation has a major impact on maize growth at all phases of development, particularly during grain filling, which lowers grain production (Zhang 2005). Studies using shading strategies have demonstrated that maize yield rises in response to rising solar radiation and falls in response to falling solar radiation (Lui et al. 2017; Zhao et al. 1990; Cui et al. 2013; Early et al. 1967; Zhao et al. 1990). According to research by Gonocruz et al. (2021), solar radiation has a positive correlation with rice yield. Grain yields generally are positively correlated to solar radiation during reproductive and ripening stages (Evans et al. 1979).

Wheat has different types and lifecycles, there is winter wheat and spring wheat. In order to catch enough solar radiation for photosynthesis, wheat planted in a site with insufficient seasonal solar radiation needs a longer growth season (Yan et al., 2022). Since potatoes are tuber crop, they need a lot of moisture and sunlight. Studies conducted in Bangladesh by Janat (2021) indicate that sun radiation above a particular threshold has a negative impact on potato yield across all regions. Samanta et al. (2020) showed that the solar radiation barrier is (above 120-210 W m2 day1).

2.2 Critical Review of Existing Work

Every study in this body of knowledge employs a different technique, whether it is in terms of the climatic elements taken into account, the approach to evaluation and analysis, or the research work's main purpose. These analyses can be crop modelling or agricultural production model e.g., Hoogenboom et al, (2019) used the Decision Support System for Agrotechnology Transfer (DSSAT) to evaluate how climate change will affect the production of food, Bassau et al (2014) and Mairano et al (2017) used crop models to evaluate the impact of one climatic factor or the other on crop production. Dawson (2013) used the FEEDME Model (Food Estimation and Export for Diet and Malnutrition Evaluation) developed under the framework QUESTI-GSI Program under the United Nations Food and Agriculture Organization (FAO). This model uses the analysis of undernourishment in the population at the country level using the FAO measure of food deprivation, referred to as the prevalence of undernourishment (FAO 2014). Also, climate change scenarios used by Abd-Elmabod et al (2020) evaluated land suitability and yield reduction under climate change scenarios, ZIlli et al. (2020) used GLOBIOM-Brazil Scenario to evaluate the impact of climate change on Brazil Agriculture. Xie et al. (2020) used RCP Series Scenarios to analyze climate change's impact on China’s Agriculture.

Parry et al. (2004) simulated the agro-economic effects of these consequences and evaluated the crop yield to greenhouse gas-induced climate change. All of these originate from the Special Report on Emission Scenarios (SRES) of the Intergovernmental Panel on Climate Change (IPCC). It combined the crop model, the socioeconomic model, and the climatic model. Although it was intended to have more balanced work that took into account a wide range of issues, models like this are unfortunately predicated on a number of assumptions.

The statistical approach is another common way of analyzing the impacts. Gay et al. (2006), Li et al. (2021) used Statistical analysis to evaluate the impact of climate change on agriculture. Ray et al. (2019) used meta-analysis to synthesize their data which is a statistical approach. They used this statistical technique to synthesize results from different published articles and journals. This is more like a secondary literature review and if there is an element of bias in the analysis done by this paper and journals it would most likely be replicated. Haile et al (2017) study the Impact of Climate Change, Weather Extremes, and price risk on global food supply using a statistical approach in their analysis and statistical analysis is often time-based on assumptions which might not necessarily be true in real-life situations. Also, Fishman (2016) used a multi-variate regression analysis of spatially disaggregated crops with respect to the weather in India which is another statistical model.

Statistical models are limited because of the complexity of the intricacies between climatic factors. Agnlolucci et al. (2020), Wang et al. (2017) and Lobel et al..(2017) all used a statistical approach to analyze crop yield as a function of climatic factors. Shmudhuber and Tubiello (2007) conducted an in-depth review of works of literature based on several facets of food production, such as the impact of climate change on food security, the stability of the food supply, and the availability of food. This paper basically relies on the quality of work done in the paper that is being reviewed. The paper is descriptive in nature It does not include a thorough analysis of the size of these effects or the likelihood of various outcomes. Mohanty (2021) is also a review paper.

It can also be Remote Sensing technology as used by Xu (2021) for the evaluation of Agricultural climate change based on remote sensing images and Paraforos (2022) focused on integrating climate change with remote sensing for sustainable Agriculture. Statistical models and scenario simulations are the most typical traditional methods of analyzing agricultural output as a function of climate change. Ordinary least-squares linear regression (LR) is essentially used in this statistical model. A priori assumptions like the linear relationship between the variables, independent variables, etc. are necessary for this LR model (Sidhu et al., 2023) and these are the limitations of statistical Models. However, Sidhu et al. (2023) used technological analysis by using Boosted Regression Trees Machine Learning Algorithm to analyze crop yield as a function of Climatic factors.

Overall, all these methodological approaches as a fair share of their limitation and application but the technological approach gives more insight because of the flexibility of the algorithms and it has fewer assumptions. Also, it is capable of analyses non-linear relationships. The goal of this project is to employ machine learning algorithm as a technological method for a better predictive model with better performance metrics while using US Climatic Data and Crop yield.

2.3 Crop Yields

Food production has felt the brunt of climate change. In 2020, between 720 million and 811 million persons worldwide were suffering from hunger, roughly 161 million more than in 2019 United Nations (UN) (UN 2023). The UN's Sustainable Development Goal 2 aims to end hunger in the world by 2030 (UN 2023) by promoting sustainable agriculture and food production. Food accessibility, availability, and supply are all aspects of the large field of food production. Agriculture is the backbone of food production, and the four main staple crops of maize, rice, wheat, and potatoes make up a large portion of it globally. The amount of these 4 major staple food crops produced per unit land area (tones or hectare) is known as crop yield. The three cereal crops that makeup 94% of total human cereal consumption are rice, maize, and wheat. (Ranum 2014).

a). ***Maize***

It is a cereal crop that is grown all over the world and is the most widely grown source of staple foods. Its scientific name is zea mays. It is a highly versatile crop that can be cultivated in a variety of soil types and climates, making it a staple crop in many nations (Shiferaw et. al 2011). Variables like climate, soil quality, and agricultural techniques all affect maize production. Globally, maize is grown for dry grain on an estimated 197 million hectares of land per year (Erenstein et al. 2022). It was first grown in Mexico about 7000 years ago and has since spread quickly to become a major crop in South America Africa and Asia (Gul et al 2022). (Wu et al 2021) is prominent in the literature on the impact of climate change on Maize yield in China from 1979 to 2016 and it was discovered that precipitation had a marginally beneficial but generally insignificant impact on maize output and that temperature had a detrimental effect. Along similar line Guntukula et al. (2020) Studied Climate change's effects on maize yields and its variability in India using Panel study approach and found out that the yield of maize per hectare is substantially lower than in industrialized countries when factors like rainfall, maximum and minimum temperatures, and others are taken into account. The temperature standard deviation is also much lower than 1. Also, according to Zizinga et al. (2022)'s simulations of the effects of climate change on maize productivity in Uganda, CSA practices would boost grain yield by 14–37% in the RCP8.5 temperature scenario.

The review paper by Mulungu et al. (2019) revealed that sustainable maize production in Sub-saharan Africa is under threat by climate change and other literature has demonstrated the seriousness of this impact on maize production. If appropriate adaptation strategies to the adverse effect of climate change are not followed, the impacts will only worsen in the near future. The study by Jiang et al. (2021) focused on adaptation measures to lessen the effects and made the prediction that the large reduction in maize production under future climatic scenarios might be connected to the higher temperature. Additionally, seed genetic enhancement and agronomic adaption strategies could lessen the unfavourable effects. This shows that there is a varying effect of climate change on maize yield depending on local climatic factors and other variables but an effective mitigation strategy would help to truncate these impacts.

b**). *Rice***

Rice is also one of the world's leading staple food crops for large populations of the world most especially in Asia and Africa. It is grown in a variety of ecosystems, including upland, irrigated, and flooded ones. Rice yields are influenced by factors such as variety, water management, nutrient availability, and pest management. According to restudies, rice needs an ideal temperature of 22 to 33 °C and 5 to 6 hours of sunlight each day to flourish (Santos et. al 2017). There is a rapidly growing literature on the impact of climate change on Rice in different location which indicates the need to analyze these impacts. According to Pickson et al. (2021), who examined this impact across 30 different provinces in China, average temperature and average rainfall were taken into consideration. The results showed that rice cultivation will decline by 0.66% if the mean temperature increases by 1% over time, and the elasticity estimate (0.0455) of average rainfall demonstrates that a 10% increase in mean rainfall causes rice cultivation to increase by 0.46% over time. This demonstrates how heavily dependent rice yield is on rainfall. There are many studies that follow a similar path. For example, Attiaoui et al.

(2019) used the ARDL-PMG methodology to assess the influence of climate change on Tunisian farmers' production of cereal crops and found that it is more responsive to rainfall than temperature.

In addition, Ansari (2021) evaluated how Indonesian rice production will adapt to climate change using the DSSAT-CERES-Rice model while taking these factors (temperature, rainfall, and solar radiation) into account. It was discovered that during the rainy season, moisture stress brought on by high rain frequency and intensity is the main reason for decreased rice yield. It was discovered that the rising temperatures, which reduce rainfall frequency and intensity, as well as increasing sun radiation, are factors that contributed to the drop in future rice production during the first and second dry seasons. This demonstrates that in order to quantify an individual climatic variable's impact, it must also consider other variables. Furthermore, it was reaffirmed by the study of Pickson et al. (2020) and Warsame et al. (2021) that rainfall had a good impact on the cultivation of grain (Rice) crops. Globally, the effect on rice yield varies and is dependent on regional climate parameters.

c). ***Wheat***

Wheat is also one of the most popular staple crops which is consumed by so many countries around the world. Depending on the season in which it is grown, wheat is either categorized as a winter crop or a spring crop (Wang et al. 2018). Quite a number of literatures has been written about the response of wheat to climatic variables. When analyzing the impact of these climatic parameters (Max air temperatures, Min air temperatures, and monthly precipitation) on wheat production in the Iberian Peninsula, Bento et al. (2021) employed simulated regression models created by employing seven EURO-CORDEX regional climate models (RCM). The outcome demonstrated a favourable link between wheat yield and precipitation as well as warm temperatures. High temperatures negatively impact the wheat and barley yield in the second model.

The analysis by Adeshina et al. (2019), which forecast hotter and slightly drier weather in Cambridgeshire in the 2080s, came to a similar conclusion that extremely high temperatures will increase wheat loss in the future. The effect of climate conditions on wheat production varies depending on where in the world wheat is grown and harvested at different times of the year. Additionally, the various varieties of wheat have varying degrees of climatic adaptation. According to Sabella et al. (2020), a rise in global temperatures is predicted to result in a decline in wheat yield. Amiri et al. (2021) and Lv et al. (2013) both confirm that a rise in temperature has a detrimental effect on wheat grain output, which can be countered by the beneficial effects of CO_2 fertilization. These effects change depending on the local climate.

d). ***Potato***

Potato is a versatile and widely consumed staple food crops that belongs to the Solanaceae family. The growth and yield of potato can be affected by quite a number of variables but it is a very adaptable crop that can grow in various climates and soil types. Researchers have evaluated the response of potato yield to different climatic factors. According to Rana et al. (2020) used WOFOST simulation to evaluate the effect of climate change on potato production and found that Atmospheric CO_2 enhances potato yield, which may be related to an improvement in photosynthesis's ability to use radiation more effectively. This indicates that an increase in Atmospheric CO_2 along with a rise in solar radiation will enhance potato yield. According to studies, rising temperatures have a negative impact on potato yield.

According to an analysis by Patil et al. (2018) and confirmation by Pradel et al. (2019), a 3-degree Celsius increase in temperature can cause a yield reduction of up to 16-17%. In the same line, Tooley (2021) concluded that various types of potatoes respond to climate factors differently. According to many studies, rising temperatures and increased Atmospheric CO_2 have a detrimental and beneficial impact on potato yield, respectively. The necessity for variation in other climatic factors before the true impact of precipitation(rainfall) and solar radiation can be evaluated makes the impact appear to be small or not yet quantifiable.

2.4 Technological/Machine Learning Analysis Approach.

Understanding the intricate relationships between climatic data and crop productivity has proven to be challenging due to the complexity of historical climate data. Schlenker et al. (2009) research work indicates that there is a non-linear intricacy between climatic factors and crop yields as confirmed by Hsaing et al. (2013). Further analysis by using a polynomial regression model to improve the model performance was filled with errors because of the approximations as done by Fishman (2016). Due to these limitations, the linear and polynomial regression model is not adequate for estimating agricultural yield in relation to climate change. Fukuda et al. (2013) used Random Forest estimation of mango yield with respect to water and the model has high accuracy which shows the importance of water to the yield of mango.

Jeong et al. (2016) used a machine learning algorithm to analyze global crop yield and found out that for predicting agricultural yields at the national and international levels, RF regression is quite efficient. Also, Khaki et al. (2020) used a deep-learning approach for the prediction of crop yield and it accurately predicted corn maize and soybean yields across the entire corn belt of the United States. Technological analysis of crop yields and climate factors has proven to be much more accurate with predictions compared to other methods. For this study, the linear regression model would also be used. Machine learning algorithms are effective instruments for analyzing intricate complexity while making fewer assumptions. There are fewer assumptions made between the variables and none of the a priori LR assumptions are required (Sidhu et al., 2023). This aids in the development of models that perform better and have higher accuracy when projecting crop yields in the future in relation to climate variables.

Different machine learning algorithms can be used to analyze the intricate relationship between climatic factors and crop yield. Xu et al. (2019) employed machine learning algorithms like Support Vector Machine, Random Forest, and Kolmogorov-Smirnov (KS) to analyze the complex relationship between typical meteorological conditions and de-trended wheat yields at crucial stages of growth in Jiangsu Province, China and established that the de-trend wheat yield is closely related to the climatic factors.

After a detailed and critical analysis of this literature, based on the methodological techniques, the focus, and the climatic factors taken into account, it provides insight into the negative impacts of climate change on food production. However, some research has been conducted on the technological aspects of analyzing how climate change would affect agriculture and food production in various parts of the world. This research work is going to focus on using a Machine Learning algorithm to analyze the impact of climate change on food production while focusing on the top four staple food crops which are Maize, Rice, Wheat and Potatoes in the US. Different machine learning algorithms can be used to identify relationships and hidden patterns between climatic factors and crop yields and develop a model but some are more suitable than others depending on the kind of data being analyzed. The Machine learning Algorithms that are suitable and would be used for this analysis are Support Vector Regression (SVR), Decision Tree Regression (DT) and Random Forest Regression (RF).

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This study focuses on the technological application of using ML algorithms to uncover the intricate relationships between meteorological conditions and agricultural yields of the main staple food crops in the United States. Solar Radiation, Surface Temperature, Precipitation, and Atmospheric Carbon Dioxide Concentration are the climatic variables to be examined, and Maize, Rice, Wheat, and Potatoes are the staple crops. The underlying connection between the trends of these climatic elements and these food crops will be unravelled. A model would be developed using a variety of ML algorithms, and the determinants of these algorithms would be compared to find the one with the highest accuracy.

3.2 Research Framework

This research work is classical scientific research (i.e., positivism) which follows a deductive methodological approach. It is a quantitative analysis and the climatic conditions to be considered are for a period of 41 years (1980-2021). Climate data is the collection of records on daily variations in weather patterns over at least a 30-year time span. The crop yield data is for each calendar year that it is cultivated and harvested. The four main crops under investigation in this study are typically planted towards the end of February or mostly, the beginning of March, and they are harvested mostly in late October or very early November. For the purpose of objectives 2 and 3, March to October is considered the crop life cycle (cultivating to harvesting).

For a focused study, the climatic data components are taken into account from the first of March, which is considered to be the beginning of crop cultivation, to the last day of October, which is thought to be the end of crop harvesting. This would make it easier to conduct very focused research because climate information from the period (November to February) when the crops are not yet planted, growing, or harvested won't interfere with the study. Table 3.1 shows the breakdown of how this research would be conducted. It shows the step-by-step approach that would be taken in order to achieve each of the objectives of this research work.

3.2.1 PHASE1: DATA COLLECTION AND ANALYSIS

All the data for this research work are sourced from reliable secondary sources. All these secondary data are painstakingly stored over a long period by government parastatals and global organizations.

a) Climate Change Data Collection

The following steps were taken in the process of sourcing for Temperature, Precipitation, Solar Radiation and Atmospheric CO_2.

I. Solar Radiation

Step 1: go to "https://www.nrel.gov"

Step 2: Under “Research” Click on "Data & Tools"

Step 3: Click on "Alphabetical Listing"

Step 4: Under the listing, look for "BMS: Baseline Measurement System" under "B" and click on it.

Step 5: The (MIDC) page would open, search for "Historical *Monthly data* are available, this includes:" and click on the "Monthly data" in the ""Historical Monthly data are available, this includes:"

Step 6: On the new "SRRL BMS Monthly Data Sets" page, scroll down and under "Select Data type:" click on "Monthly Report"

Step 7: Under "Select Month" choose the year and the month of solar radiation data and click on "Submit"

NO	RESEARCH QUESTION	RESEARCH OBJECTIVES	METHODOLOGY		
			PHASE	*ACTIVITIES*	*OUTPUTS*
1	What changes in climatic factors (Solar radiation, Temperature, Precipitation, Atmospheric CO_2) as well as yield (maize, Rice, Wheat, Potato) have taken place over the period of 1980 to 2021 in the USA?	To do a longitudinal analysis of the trend of the Climatic Factors and the yield of the crops.	DATA COLLECTION AND ANALYSIS	Pre-processing (Data Cleaning) Descriptive/ Statistical Analysis Data Preparation for time series analysis.	Time series of the climatic factors and Time series of the Crop yield variables.
2	What are the intricate connections between Climatic variables from 1980 to 2021 and crop yields variables in the USA?	To analyze hidden trends and patterns between the climatic variables and crop yields using scattered plots	Further Data Analysis	Further Pre-processing for the analysis and visualization.	Scattered plot of Temp and crop yield d of each crop, Solar Rad and yield d of each crop, Atmospheric CO_2 and yield d of each crop, Precip and crop yield d of each crop.
3	How can ML algorithms be used to develop a prediction model and compare the performance metric?	To develop SVR, DTR and RFR models and compare their performance metrics.	APPLICATION OF MACHINE LEARNING ALGORITHM	Standardizing Splitting into Training and Test Model Fitting for SVR, DTR, RFR	R-square, Mean Square Error, Mean Absolute error and Root Mean Square Error of SVR, DTR and RFR.

Table 3.1: Research Methodology Break down

Step 8: A new window will open that contains the solar radiation text data

Appendix 1 shows the raw text data on the new page.

II. Temperature

Step 1: Open the website www.visualcrossing.com

Step 2: Scroll down and click on "Data Download"

Step 3: Sign up on the web page and log into the site after confirmation of email.

Step 4: Choose Location (USA)

Step 5: on the "Weather Query Builder" page, choose “manual explore”.

Step 6: Choose the date range of the data type needed (01-01-1980 to 31-12-2021).

Step 7: On the "Query option" click on the "weather element" and a pop out would be displayed.

Step 8: Click on “core weather elements” and chose the variables that are needed or leave the default selection and click "ok"

Step 9: Choose the format of the data type needed (grid or chart or json or csv. "CSV")

Step 10: Click on "Download" Data

Step 11: Under "Output Section" Select "Daily" and click on "Continue"

Step 12: Edit or write a preferred name for the dataset "USA" and click "Continue"

Step 13: Click on "Submit Query"

Step 14: Wait a few minutes and click “Download” to download your data.

Appendix 2 shows the raw data downloaded through the steps listed above.

III. Precipitation

Similar steps taken in a) were applied to get the data.

Appendix 2 shows the raw data downloaded through the steps.

IV. Atmospheric CO_2 Concentration.

Step 1: go to "https://www.gml.noaa.gov/"

Step 2: Click on "Products" and on the list click on "Trends in CO_2, CH_4, N_2O, SF_6"

Step 3: Make sure "Mauna Loa, Hawaii" is selected usually in blue when selected.

Step 4: Click "Data"

Step 5: Click on the CSV version of "Mauna Loa CO_2 Monthly mean data (Text) or (CSV)

Appendix 3 shows the raw data downloaded through the steps.

b) Crop Yield Data Collection

The following steps were taken in the process of sourcing for Maize, Rice, Wheat, and Potato.

I. Maize

Step 1: go to https://www.fao.org/home/en/

Step 2: Click on “Statistics” from the dropdown and click on “FAOSTAT”

Step 3: On the FAOSTAT home page, click on “Data

Step 4: Click on “Production” in the drop-down click on “Crops and Livestock Products”

Step 5: On the new page under “COUNTRY”, Select “United States of America”

Step 6: Under the “ELEMENT” Section, Click on “Yield” and “Area Harvested”

Step 7: Under the “ITEMS” section search for “Maize” and Select “Maize (Corn)”

Step 8: Under “YEARS” Select “1980,…, 2021”

Step 9: Under “Output Type”, Select “Table”

Step 10: Under “File Type” Select “CSV”

Step 11: Click on “Show Data” to have a preview of the dataset

Step 12: Click on “Download”

Appendix 4 shows a preview of the raw data downloaded through the steps above.

II. Rice

Similar Steps were taken in (I) above.

Step 7: Under the “ITEMS” Section search for “Rice” and Select “Rice”

Continue with the steps in (I) above

Appendix 4 shows a preview of the raw data downloaded through the steps above.

III. Wheat

Similar steps were taken in (I) above.

Step 7: Under the “ITEMS” Section search for “Wheat” and Select “Wheat”

Continue with the steps in (I) above

Appendix 4 shows a preview of the raw data downloaded through the steps above.

IV. Potato

Similar steps were taken in (I) above.

Step 7: Under the “ITEMS” section search for “Potato” and Select “Potato”

Continue with the steps in (I) above.

Appendix 4: Shows a preview of the raw data downloaded through the steps.

ACTIVITY 1: Data Pre-processing (Cleaning)

The following steps were taken for the pre-processing and cleaning of the raw dataset before being used for this analysis.

Climatic Data Pre-Processing (Cleaning)

The following steps were taken in cleaning the different climatic variable data.

I. Solar Radiation

Unfortunately, the solar radiation dataset from January 1980 to June 1981 was not available. The text data extraction from the website started from July 1981. After downloading the July to December 1981 Data.

Step 1: On the website, select the year "1981", Select the month "July" and click on "Submit" The text file would open in a new tab.

Step 2: Copy the Solar radiation data as the data for one month is in Appendix 1. The description is shown below:

Global Solar Radiation on a Horizontal Surface

Hourly Integrated and Daily Totals

Instrument: Eppley PSP

Watt-Hours per Square Meter

Step 3: “Paste” the copied one-month data in a "notepad"

Step 4: Repeat Step (1), (2) and (3) for August, September, October, November and December.

Step 5: In the notepad containing the monthly data from July to December 1981, click on "File" click on "Save as" and choose a "file name"(like "1981"), in the section "save as type" click the drop down and choose "All files". Then add ".csv" to the file name. The file name is now "1981.csv". This saves as a csv file.

Step 6: Repeat the process above for all the months in the year from 1982 to 2021. After this a total of 40 CSV file (each containing 12 monthly data set) corresponding to each year from 1981 to 2021

Step 7: "Open" the "1981" and "1982" file, copy the dataset in 1982.csv and "paste" under the dataset in 1981.csv.

Step 8: "Open" the 1983 and "copy" the dataset, "paste" under the first dataset 1981 which contains data from 1981 and 1982.

Step 9: "Repeat" Step 8 for "1984, 1985,.... 2021" file one after the other. The 1981 file now contains all the data set from July 1981 to December 2021.

Step 10: "Open" 1981 file. Click "copy" the dataset in the file, on the toolbar of the excel, "click" on 'Data", from the options available click on "Remove Duplicates".

Step 11: Open another excel file, on the first column, label it "Datetime" and type "01-january-1980", on the third row, type "02-January-1980". Select the two dates and put the cursor at the lower right end of the selection box, a "+" sign would appear.

Step 12: Drag the "+" Sign down the column till it's filled up to 31 December 2021.

Step 13: Open the 1981 csv file (that contains all the daily solar radiation datasets), go to column "AX" which is the cumulative sum of the daily solar radiation.

Step 14: Copy the Daily data input for each month and paste it on the second column of the new dataset that has the daily date from the first day in 1980 to the last day in 2021 as it corresponds to the date. The new dataset should be 14965 rows by 2 columns.

(N.B Some days of the month were skipped therefore, the need to check with the corresponding date. For the few missing values in a month, the mean of other input of the available days are used to replace them)

Step 15: Select the two columns in the new dataset, click on "Insert" on the toolbar, click on "Pivot Table" and select "From Table/Range", from the pop-up, click on "New Worksheet"

Step 16: On the right side select "Datetime" and "solar_rad". The cumulative sum of the daily solar radiation would be converted to Monthly solar radiation.

Step 17: “Copy” the monthly solar radiation to a new Excel file and save it as m_solar_rad.

To yearly solar radiation which is from March to October every year for the purpose of a focused crop yield analysis

Step 18: Click on the box in front of the October value in the third column, click on "Home" on the tool bar, click on "AutoSum" on the top right corner. The prompt to select a list of numbers starting with the October value of the temp would appear.

Step 19: Drag the cursor and select the values from October to March(upward) and click on "Enter" on the keyboard. The value for the AutoSum of all the selected data would appear in the column that was first selected, this is the year solar radiation value.

Step 20: Repeat step (18) and (19) for each year from 1982 to 2021.

Step 21: Copy the yearly solar radiation from step (20) to another excel file with the corresponding year on one column and the solar radiation on the second column.

N.B: From January 2012 the dataset for the instrument with the above description in Step (2) is not available, a similar dataset from the instrument with the description below was copied.

Global Solar Radiation on a Horizontal Surface

Hourly Integrated and Daily Totals

Instrument: Yankee TSP-1

Watt-Hours per Square Meter

Also, 1980 and 1981 yearly input was replaced with the mean value of 1982 to 1985.

NB: There were missing data from July 1983 to September 1984 and it was replaced by the mean of the previous 6month (January to July 1983)

II. Temperature

Step 1: The “**datetime**” column from 1980 to 2021 was copied to a new excel page.

Step 2: The corresponding “temp” for each day of the date was also copied to the page with the date.

Step 3: Select the “Datetime” and the “temp” column and click on “insert” on the toolbar.

Step 4: Click on “Pivot Table” and Select “From Table/Range”, from the pop-up, click on “New Worksheet”

Step 5: On the right side Select “Datetime” and “temp”. The cumulative sum of the daily temperature would be converted to monthly Temperature.

Step 6: Open another Excel file and create a column “Datetime” which is monthly Jan 1980, February 1980, …., December 2021.

Step 7: Copy the new Monthly data one after the other into the new Excel file with the corresponding monthly datetime and save it. Saveit as “M_Temp”. To “yearly temp” which is from March to October every year for the purpose of a focus crop yield analysis.

Step 8: Click on the box in front of the October value in the third column, click on "Home" on the toolbar, and click on "Autosum" in the top right corner. The prompt to select a list of numbers starting with the October value of the temp would appear.

Step 9: "Drag the cursor and select the values from October to March" and click on "Enter" on the keyboard. The value for the auto sum of all the selected data would appear in the column that was first selected. This is the "yearly temp" value.

Step 10: "Repeat" Steps (8) and (9) for each year from 1981 to 2021

Step 11: Copy the "yearly temp" from Step (10) to another Excel file with the corresponding year on one column and the "temp" on the second column.

III. Precipitation

Similar steps are taken in (II) above replacing "temp" with "precip"

IV. Atmospheric CO_2 Concentration

The Atmospheric C02 Concentration is a monthly Data The following steps were taken during Data pre-processing.

Step 1: "Open" a new Excel file and also "Open" the downloaded Atmospheric CO_2 Conc file.

Step 2: "Scroll down" on the downloaded file and "Copy" the "year" column from 1980 to 2021 and "Paste' in the new Excel file.

Step 3: on the "average" column, copy the values from the row on the first month in 1980 down to 2021 to the new Excel file that contains the year and the month. This new dataset contains "year' and "average"

Step 4: On the "year" column the first 1980 represents Jan 1980, change it to "January 1980", and change the next 1980 to "February 1980". Select the two rows containing the new date and move the cursor to the lower right end of the selection box, a "+" sign would appear.

Step 5: Drag the cursor down all the way through the last "2021" which represents "December 2021" and all the year changes to the format of the first and second. Save it. M_CO_2

Step 6: Click on the box in front of the October value in the third column, click on "Home" on the toolbar, and click on "AutoSum" in the top right corner. The prompt to select a list of numbers starting with the October value of the temp would appear.

Step 7: Drag the cursor and select the values from October to March(upward) and click on "Enter" on the keyboard. The value for the auto sum of all the selected data would appear in the column that was first selected. This is the yearly value for the Atmospheric CO_2 concentration.

Step 8: Repeat Steps (6) and (7) for each year from 1981 to 2021.

Step 9: Copy the yearly Atmospheric CO_2 Value from step (8) to another Excel file with the corresponding year on one column and the Atmospheric CO_2 on the second column.

Overall, combine all climatic datasets into three different Excel file, One containing all the monthly data from January 1980 to December 2021, the other containing all the Yearly data which is the summation of the values from March to October and the one containing all the yearly data from January to December.

Appendix 5: Shows the data generated by combining all the monthly climatic data.

Crop Yield Data Pre-processing (Cleaning)

The following steps were taken during the process of cleaning the different yields of the staple crops data.

I. Maize

The maize yield dataset is a yearly dataset and the processes involved are as follows:

Step 1: “Open” a new Excel file and create a column called “year”, under the column type “1980” and next “1981”.

Step 2: “Select” the two years and “move” the cursor to the low right corner and it will change to a "+" sign, “drag” it down till the column is filled up with years following that pattern until it is “2021”.

Step 3: In the downloaded data, under the column Element, “scroll down” and search for "yield", and "Select" the row with the first yield Check the next column “Item code” input must be “112” and the column “Item” must have an input “Maize” then “Scroll up” and search for the column with the name "value". Then “scroll down” to check the selected row.

Step 4: The input under the "Value" column on the selected row represents the value of the “Maize yield” for the year "1990".

Step 5: “Select all” the inputs under the column "value" from the first "yield" under the column "Element" to the last "yield" and “copy” it to the new Excel file which corresponds to the years “1990 to 2021”.

II. Rice

The rice yield dataset is a yearly dataset and the processes involved are as follows:

Step 1: Repeat Steps (1) and (2) under Maize above.

Step 2: In the downloaded data, under the column Element, “scroll down” and search for “yield”. Check the next column “item code” input must be 113” and the column “Item” must have an input “Rice” and “Select” the row with the first yield. “Scroll up” and search for the column with the name “value”. Then “scroll down” to check the selected row.

Step 3: The input under the "Value" column on the selected row represents the value of the “Rice yield” for the year "1990".

Step 4: “Select all” the inputs under the column "value" from the first "yield" as long as it satisfies all the conditions in Step (2) and “copy” it to the new Excel file which corresponds to the years “1990 to 2021”.

III. Wheat

The Wheat yield dataset is a yearly dataset and the processes involved are as follows:

Step 1: Repeat Steps (1) and (2) under Maize above.

Step 2: In the downloaded data, under the column Element, “scroll down” and search for “yield”. Check the next column “item code” input must be 111” and the column “Item” must have an input “Wheat” and “Select” the row with the first yield. “Scroll up” and search for the column with the name “value”. Then “scroll down” to check the selected row.

Step 3: The input under the "Value" column on the selected row represents the value of the “Wheat yield” for the year "1990".

Step 4: “Select all” the inputs under the column "value" from the first "yield" as long as it satisfies all the conditions in Step (2) and “copy” it to the new Excel file which corresponds to the years “1990 to 2021”.

IV. Potato

The Potato yield dataset is a yearly dataset and the processes involved are as follows:

Step 1: Repeat Steps (1) and (2) under Maize above.

Step 2: In the downloaded data, under the column Element, “scroll down” and search for “yield”. Check the next column “item code” input must be 1510” and the column “Item” must have an input “Potato” and “Select” the row with the first yield. “Scroll up” and search for the column with the name “value”. Then “scroll down” to check the selected row.

Step 3: The input under the "Value" column on the selected row represents the value of the “Potato” for the year "1990".

Step 4: “Select all” the inputs under the column "value" from the first "yield" as long as it satisfies all the conditions in Step (2) and “copy” it to the new Excel file which corresponds to the years “1990 to 2021”.

Overall, the crop yield data was downloaded twice from the site from 1973 to 1989 and from 1990 to 2021. The data from 1973 was extracted in a similar way while paying attention to the dates in the dataset. Combine all the Crop yield data into one CSV file.

Activity II: Descriptive/Statistical Analysis

All the new yearly Excel files created during the data preprocessing (Cleaning) are combined into one particular dataset with the following columns:

“year”, “country”, “maize_yield”, “pot_yield”, “wheat_yield”, “rice_yield”, “solar_rad”, “Co2”, “temp” and “precip”

Appendix 6: Shows the data generated after combining all the processed data.

The descriptive analysis of the data involves reading the new CSV file into Python 3 on jupyter notebook on Anaconda and examining the basic descriptive properties of the dataset. The following steps were taken.

Step 1: Importation of necessary Library as shown in Appendix 7

Step 2: Reading the CSV file into the Python Kernel as “df”.

Step 3: Checking the descriptive properties of the dataset by using the following to check

Activity III: Data Preparation for Time Series of Climatic Factors

This involves exploring the trends and patterns of the different variables in the dataset from 1980 to 2021. This can be done by visualization.

Climatic Factors: The following steps were taken:

Step 1: Plot the graph showing the Monthly and Annual Trend of Solar Radiation from 1980 to 2021.

Step 2: Plot the graph showing the Monthly and Annual trends of Temperature from 1980 to 2021.

Step 4: Plot the graph showing the Monthly and Annual trends of Precipitation from 1980 to 2021.

Step 4: Plot the graph showing the Monthly and Annual trends of Atmospheric CO_2 Conc. from 1980 to 2021.

Step 5: Plot a graph showing the Annual Percentage change in all 4 climatic factors.

The graphs generated by Solar Radiation, Temperature, Precipitation and Atmospheric CO_2 can be visualized in Figures 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, and 4.9 in Chapter 4 respectively.

Activity IV: Data Preparation for Time Series of the Yield of Crops.

Crop yields: The following steps were taken:

Step 1: Plot the graph showing the trends of Maize from 1980 to 2021.

Step 2: Plot the graph showing the trends of Rice from 1980 to 2021.

Step 3: Plot the graph showing the trends of Wheat from 1980 to 2021

Step 4: Plot the graph showing the trends of Potato from 1980 to 2021

Step 5: Plot a graph showing the percentage change in the yield of all the crops.

The graphs generated from the yields of Maize, Rice, Wheat and Potato can be visualized in Figures 4.10, 4.11, 4.12, 4.13 and 4.14 in Chapter 4 respectively.

3.2.2 PHASE II:

Activity 1: Further pre-processing for Correlation Analysis.

All the major pre-processing has been done in ACTIVITY I. The CSV file containing all the Monthly Climatic Factors would be used. Also, the Csv file containing all the Crop yield Data and Climatic variables would be used.

Activity II

The following visualization was done.

a). Solar Radiation.

The correlation between Solar radiation on each of the crop yields is given below i.e solar_rad as y variable and use matplotlib.pyplot library for all the scattered plots.

Step 1: X variable is “maize_yield” for the Correlation between Maize yield and Solar Radiation and plot.

Step 2: X variable is “rice_yield” for the Correlation between Rice yield and Solar Radiation and plot

Step 3: X variable is “wheat_yield” for the Correlation between Wheat Yield and Solar Radiation.

Step 4: X variable is “pot_yield” for the Correlation between Potato Yield and Solar Radiation and plot.

The graphs of the correlation between Solar radiation and Maize, Rice, Wheat and Potato are in Figures 4.15, 4.16, 4.17, and 4.18 in Chapter 4.

b). Temperature.

The correlation between Temperature on each of the crop yields is given below i.e. temp as y-variable and use matplotlip.pyplot library for the scattered plots.

Step 1: X variable is “maize_yield” For the Correlation between Maize yield and Temperature and plot.

Step 2: X variable is “rice_yield” For the Correlation between Rice Yield and Temperature and plot.

Step 3: X variable is “wheat_yield” For the Correlation between wheat Yield and Temperature and plot.

Step 4 X variable is “pot_yield” For the Correlation between potato Yield and Temperature and plot.

The Scattered plot of the correlation between Temperature and Maize, Rice, Wheat and Potato are in Figures 4.19, 4.20, 4.21, and 4.22 in Chapter 4.

c). Precipitation

The correlation between Precipitation on each of the crop yields is given below i.e. “precip” as y-variable and use matplotlib.pyplot library for the scattered plots.

Step 1: X variable is “maize_yield for the correlation between maize yield and Precipitation and plot.

Step 2 X variable is “rice_yield for the correlation between Rice yield and Precipitation and plot.

Step 3: X variable is “wheat_yield for the correlation between Wheat yield and Precipitation and plot.

Step 4: X variable is “pot_yield for the correlation between Wheat yield and Precipitation and plot..

The Scattered plot for the correlation between Precipitation and Maize, Rice, Wheat and potato are in Figures 4.23, 4.24, 4.25, and 4.26 in Chapter 4.

d). Atmosphere CO_2

The correlation between Atmospheric CO_2 concentration and each of the crop yields is given below i.e ."CO_2" as y-variable and use matplotlib.pyplot librar for the scattered plots.

Step 1: X variable is "maize_yield for the correlation between maize yield and Atmospheric CO_2 and plot.

Step 2: X variable is "rice_yield for the correlation between Rice yield and Atmospheric CO_2 and plot.

Step 6: X variable is "wheat_yield for the correlation between Wheat yield and Atmospheric CO_2 and plot.

Step 8: X variable is "pot_yield for the correlation between Potato yield and Atmospheric CO_2 and plot.

The Scattered plot for the correlation between Precipitation and Maize, Rice, Wheat and potato are in Figures 4.27, 4.28, 4.29, and 4.30 in Chapter 4.

Overall, The SNS heatmap shows the correlations of all the variables with each other in the figures. The heat map showing the correlation between all these variables is shown in Figure 4.31 in Chapter 4.

3.2.3 PHASE III: Further Pre-processing.

As shown in PHASE (I) Activity2

APPLICATION OF MACHINE LEARNING ALGORITHM

The following ML algorithms would be applied to pre-processed data from Phase 1. They are SVR, DTR and RFR.

a). Support Vector Regression (SVR).

The following steps were taken in using SVR for analyzing the impact of all the meteorological variables (“solar rad”, “temp”, “precip”, and “CO_2”) which are the independent variable and the Crop Yields (“Maize_yield”, “rice_yield”, “wheat_yield” and “pot_yield”) which are the dependent variables.

I. Maize_yield

Step 1: on a new kernel, Import the necessary libraries as shown in Appendix 8.

Step 2: Reading the processed CSV file into the python karnel as “df”

Step 3: Selecting the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “CO_2” and the dependent variable “y” which is “maize_yield”.

Step 4: Scaling the Features “X” and “y” using Standardization. Importing the “StandardScaler()” from sklearn.preprocessing and assigning it to a variable “scaledX” and “scaledy”.

STANDARDIZATION

Step 5: Transforming “X” and “y” by fitting the scaled and scaled on “X” and “y” using “scaled.fit_transform(X)” and “scaled.fit_transform(y)”

SPLITTING INTO TRAIN AND TEST

Step 6: Splitting the dataset into training and test data set using 75% to 25% respectively. Importing “train_test_split” from “sklearn.model_selection.

Step 7: Applying the splitting function on our data which divides it into X_train, X_test, y_train and y_test. The function train_test_split is used specifying the test_size as 0.25, train_size as 0.75 and the random state as 10 for it not to keep changing each time we re-run it.

CHOOSING HYPERPARAMETERS

Step 8: Choose “rbf” as the “kernel function” since the relationship between the “X” and “y” is not a linear relationship.

Step 9: Train the SVR model with the training Dataset “X_train” and “y_train” by fitting the SVR model function svr_model.fit which was imported from sklearn.svm.

Step 10: Evaluate the model by applying the model on the testing data “X_test” to predict the result which y_predict which should be similar to y_test.

Step 11: Evaluating the performance of the model by calculating regression metrics which is done by comparing the testing output “y_test” and the predicted output by the model “y_predict”

Step 12: R-Squared: importing r2_score from sklearn.metrics and applying r2_score on the testing output “y_test” and the output predicted by the SVR model which is “y_predict”

Step 13: Mean Absolute Error (MAE): Importing “mean_absolute_error” from sklearn.metrics and applying it on “y_test” and y_predict”.

Step 14: Mean Square Error (MSE): Importing “mean_square_error” from sklearn.metrics and applying it on “y_test” and y_predict”.

Step 15: Root Mean Squared Error (RMSE): by using the function rmse = float (format(np.sqrt(mean_squared_error(y_test, y_predict)), '.3f'))

II. Rice_yield

Similar Steps from Steps (1) and (2) in I above.

Step 3: Selecting the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “rice_yield”.

Similar steps Step (4) to (15) in I above.

III. Wheat_yield

Similar Steps from Steps (1) and (2) in I above.

Step 3: Select the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “wheat_yield”.

Similar steps Step (4) to (15) in I above

IV. Pot_yield

Similar Steps from Steps (1) and (2) in I above.

Step 3: Selecting the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “pot_yield”.

Similar steps Step (4) to (15) in I above

The computation of the metrics of SVR is given in Table 4.2 in Chapter 4.

b) Decision Tree Regression (DTR)

The following steps were taken in using DTR for analyzing the impact of all the meteorological variables (“solar rad”, “temp”, “precip”, and “CO_2”) which are the independent variable and the crops Yields (“Maize_yield”, “rice_yield”, “wheat_yield” and “pot_yield”) which are the dependent variables.

I. maize_yield

Similar steps in a) (I) from Step (1) to (7) above.

CHOOSING HYPER-PARAMETERS

Step 8: Import DecisionTreeRegressor from Sklearn.tree.

Step 9: Then train the data with the DecisionTreeRegression model while choosing a (max_depth=3, min_samples_split=5, random_state = 0)as hyperparameters. Fit the model on the training data which is “X_train” and “y_train”.

Similar steps in a) (I) from Step (10) to (15)

II. Rice_yield

Similar Steps from b) (I) Steps (1) and (2) in I above.

Step 3: Selecting the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “rice_yield”.

Similar steps b) (I) from Steps (4) to (15) in above.

III. Wheat_yield

Similar Steps from b) (I) Steps (1) and (2) in I above.

Step 3: Select the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “wheat_yield”.

Similar steps b) (I) from Step (4) to (15) in above.

IV. Pot_yield

Similar Steps from b) (I) from Steps (1) and (2) in I above.

Step 3: Select the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “wheat_yield”.

Similar steps b) (I) from Step (4) to (15) in above.

The computation of the metrics of SVR is given in table 4.2 in Chapter 4.

c) Random Forest Regression (RFR)

The following steps were taken in using RFR for analyzing the impact of all the meteorological variables (“solar rad”, “temp”, “precip”, and “CO_2”) which are the independent variable and the Crop Yields (“maize_yield”, “rice_yield”, “wheat_yield” and “pot_yield”) which are the dependent variables.

I. maize_yield

Similar steps in a) (I) from Step (1) to (7) above.

CHOOSING HYPER_PARAMETERS

Step 8: Import RandomForestRegressor from Sklearn.ensemble.

Step 9: Then train the data with the RandomForestRegressor model while choosing the number of decision trees (n_estimators = 100, random_state = 5) and a “random_state” of “5” for it not to keep changing and n_estimator is the number of trees. Fit the model on the training data which is “X_train” and “y_train”.

Similar steps in b) (I) from Step (10) to (15)

II. Rice_yield

Similar Steps from a) (I) Steps (1) and (2) in I above.

Step 3: Selecting the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “rice_yield”.

Similar steps c) (I) from Step (4) to (15) in above.

III. Wheat_yield

Similar Steps from c) (I) Steps (1) and (2) in I above.

Step 3: Select the features which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “wheat_yield”.

Similar steps c) (I) from Step (4) to (15) in above.

IV. Pot_yield

Similar Steps from b) (I) from Steps (1) and (2) in I above.

Step 3: Feature Selection which are the independent variables “X” Which are “solar_rad”, “temp”, “precip”, “co2” and the dependent variable “y” which is “wheat_yield”.

Similar Steps c) (I) from Step (4) to (15) in the above.

The computation of the metrics of RFR is given in Table 4.2 in Chapter 4.

CHAPTER FOUR

RESULT AND EVALUATION

4.1 Introduction

The focus of this study is on the use of data science tools and techniques for the analysis of the variation of climatic factors like solar radiation, temperature, precipitation, and Atmospheric CO_2 concentration as it affects crop yield of the main staple food crops of the world, which are maize (corn), rice, wheat, and potato. Using the USA as a case study, this is aimed to simplify the intricate relationships between climatic elements and how various crops respond from cultivating to harvesting time. Over a 41-year period from 1980 to 2021, trends in both the agricultural yield and these climate conditions are analyzed. Additionally, through correlation research, the unspoken links between these climatic parameters and crop yield are revealed. These data are also subjected to the application of three different machine learning algorithms (Support Vector Regression, Decision Tree, and Random Forest), which result in the development of several models. These various ML algorithms' performance metrics are evaluated.

All the data used for this analysis are secondary data and highly reliable sources (majorly the official government website of the USA). The raw data required quite a number of data cleaning and pre-processing before they can be used for this analysis. The Food and Agriculture Organization of the United Nations is the source of all crop yield data, which is based on the annual yield of the crops since the majority of crops are grown between March and harvested by the end of October. The Meteorological variables are sourced from different places. Solar Radiation was hourly data but was converted to Daily, monthly and finally yearly Data, Temperature and Precipitation Data are daily data and Atmospheric CO_2 is monthly data. For the purposes of a focused study in Output 2 and 3, the months of March through October are regarded as one calendar year for the meteorological data because they coincide with the crop life cycle.

This chapter primarily discusses the analysis's outcomes, including visualizations of those outcomes, as well as their interpretation and the outcomes of all the processes outlined in Chapter 3 overall. An explicit interpretation of the achievement of the research aims and objectives.

4.2 OUTPUT 1: Time Series Visualization

Over the 41-year period that is being addressed in this study, both the climatic factors and agricultural production have undergone changes. Analyses are done on the overall fluctuation throughout this time. Below are the trends based on the data.

a). Climatic Variables Trends

The study would consider the climate factors below and look at their trend over a 41-year period. The climatic data visualization is of two types which are Monthly and yearly visualization for this period. This is broken into two visualizations for the purpose of better visualization and easy interpretation.

I. Solar Radiation

The variation in the recorded solar radiation on a horizontal surface measured in Kilo Watts-Hours Per square meter from January 1980 through to December 2021 is given in Figures 4.1 and 4.2

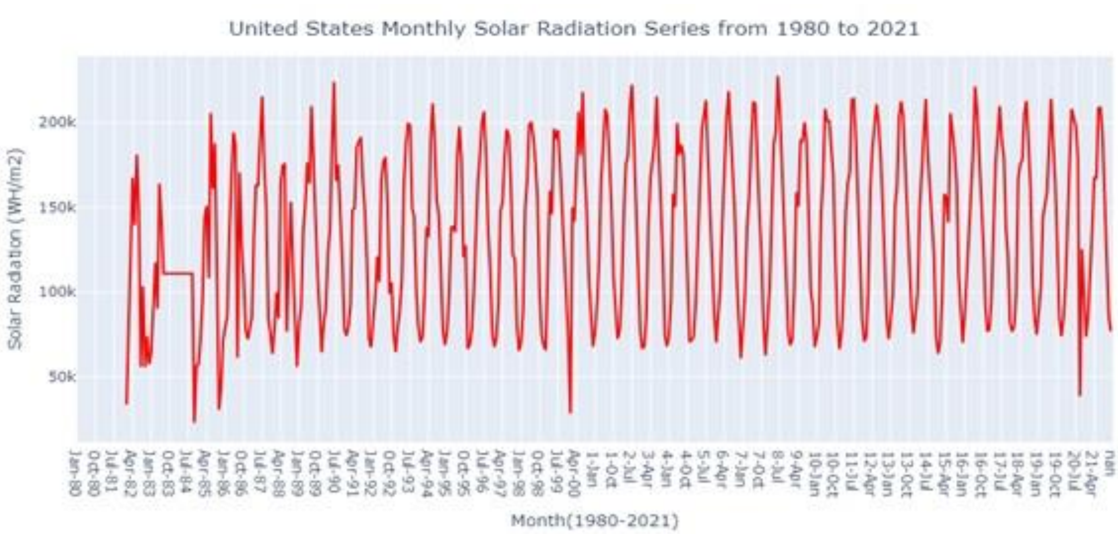

Figure 4.1: Monthly solar radiation trend from 1980 to 2021.

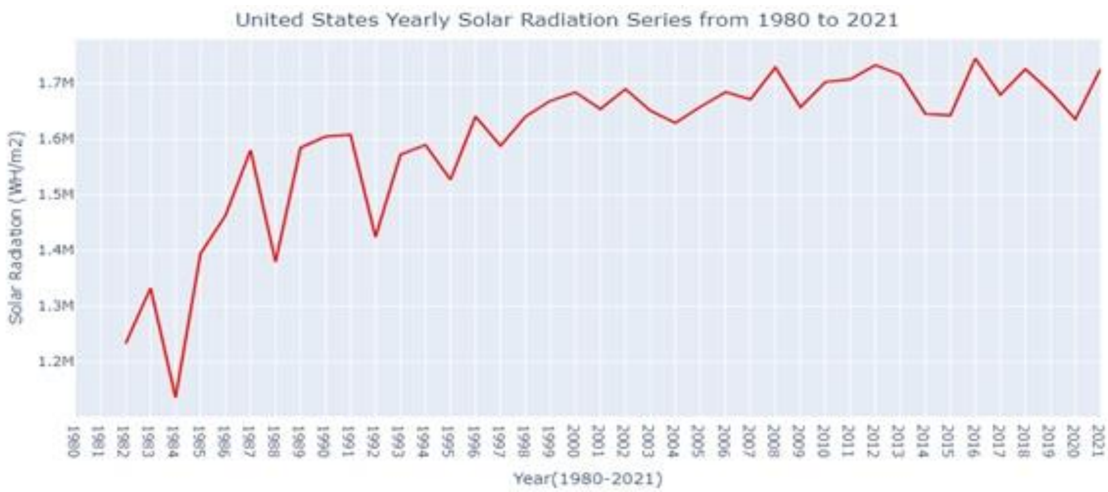

Figure 4.2: Annual Solar Radiation trend from 1980 to 2021.

According to Figures 4.1 and 4.2, the average annual solar radiation measured in Kilo-Watts hour per square meter increased significantly in 1986, and from 1987 until 2021, it increased steadily on average. However, the total solar radiation slightly decreased from 1996 to 1997 until 1998, after which it began to rise and kept doing so with an average annual rise. Also, the sharp drop from 1983 to 1984 is because of the missing value from July 1983 to July 1984 which was filled by the mean and it is more obvious in Figure 4.1.

II. Temperature

Figures 4.3 and 4.4 show the fluctuation of the Total temperature measured in degrees Celsius over the period of 41 years for both monthly and yearly data. The second graph helps with betterer visualization and it suggests that there was a significant drop in the temperature from 1991 to 1992 before the steady increase going forward. Also, a little decrease was noticed from 2002 to 2003 and an increase from 2009 to 2010 which was like a significant decrease from 1991 to 1992.

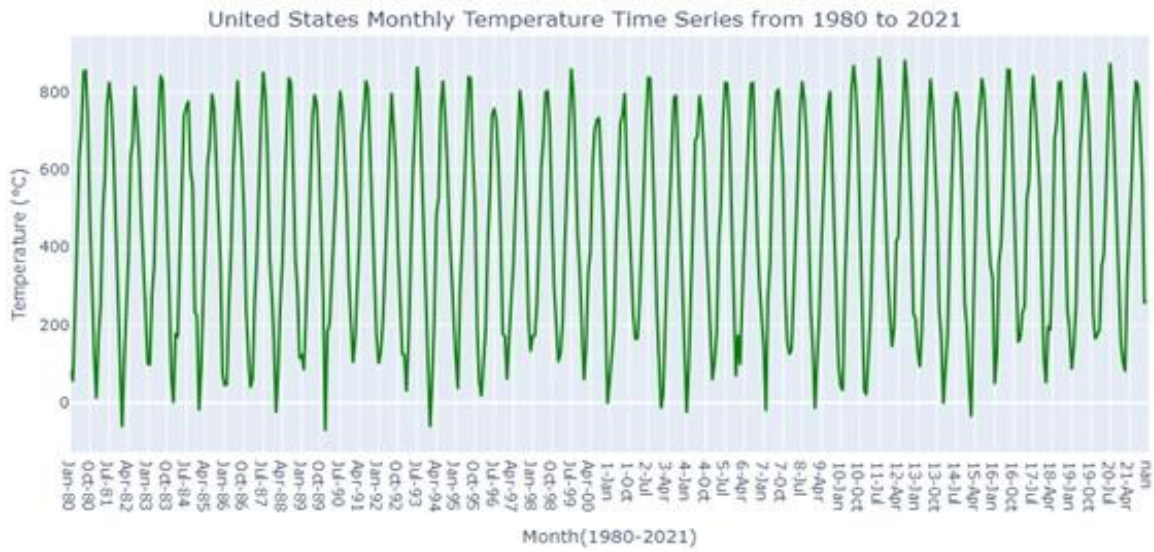

Figure 4.3: Monthly Temperature trend from 1980 to 2021.

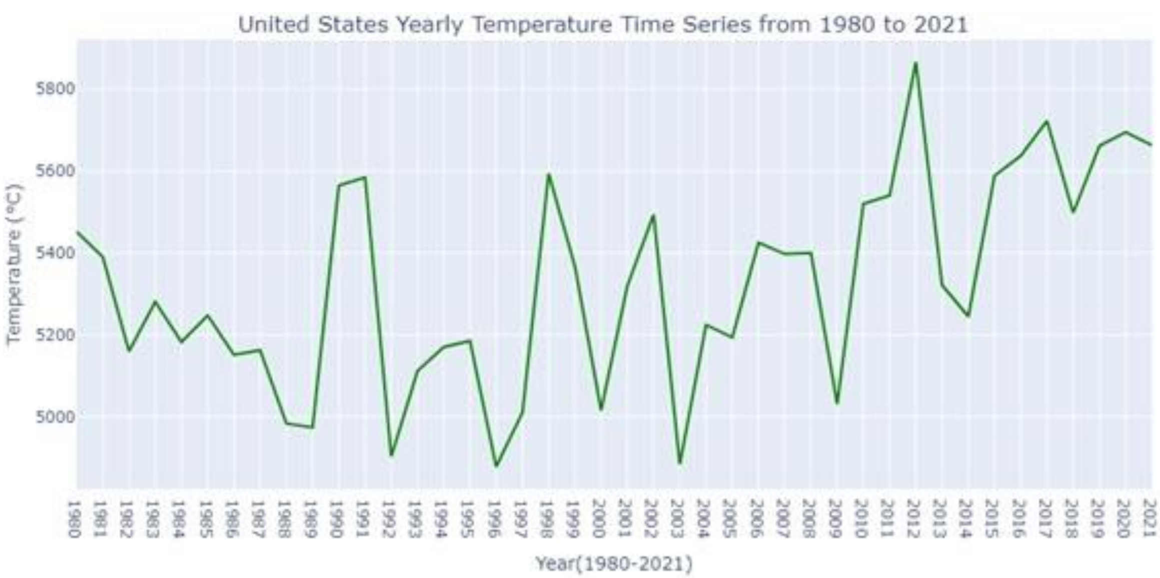

Figure 4.4: Annual Temperature trend from 1980 to 2021.

III. Precipitation

Figure 4.5 and 4.6 above shows the variation in total precipitation measured in millimetres of data over a 41-year period. Figure 4.6 appears to indicate that the average annual precipitation for this time period varied only slightly over 10-year intervals, with some years experiencing little precipitation, such as 1986, 1991, 1998,

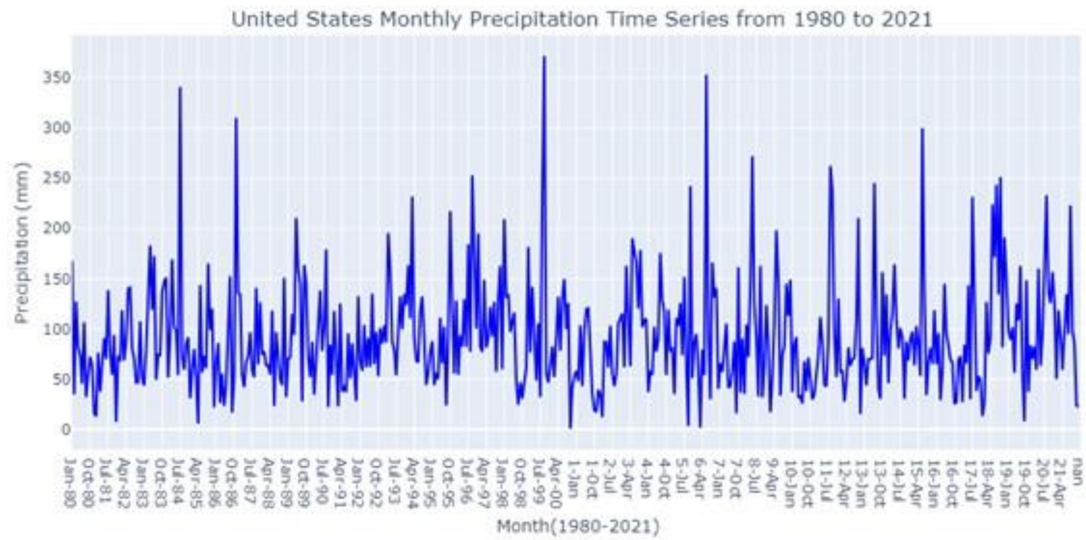

Figure 4.5: Monthly Precipitation trend from 1980 to 2021.

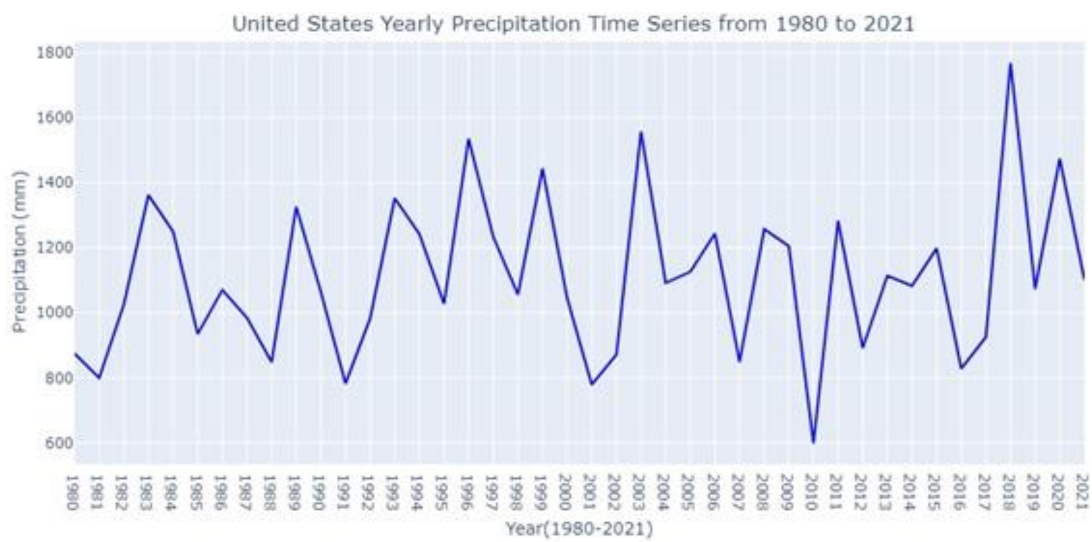

Figure 4.6: Total Annual Precipitation trend from 1980 to 2021.

2002, 2007, and 2016, and others experienced a lot, such as 1984, 1989, 1996, 1999, and 2003. The amount of precipitation in 2010 was exceptionally low however, 2018 was extraordinarily high.

IV. Atmospheric CO_2 Concentration.

Since 1980, the annual Atmospheric CO_2 measured in parts per million level has steadily increased, as shown in Figures 4.7 and 4.8. There have been discussions concerning global warming concerning consistent increase, which is occurring at an unprecedented rate for a variety of reasons. One of the main gases that trap heat and raise the average temperature of our planet is Atmospheric CO_2 which also has an effect on Crop yield. Figures 4.7 and 4.8 shows the unprecedented rate at which it has been increasing in the USA since 1980 to 2021.

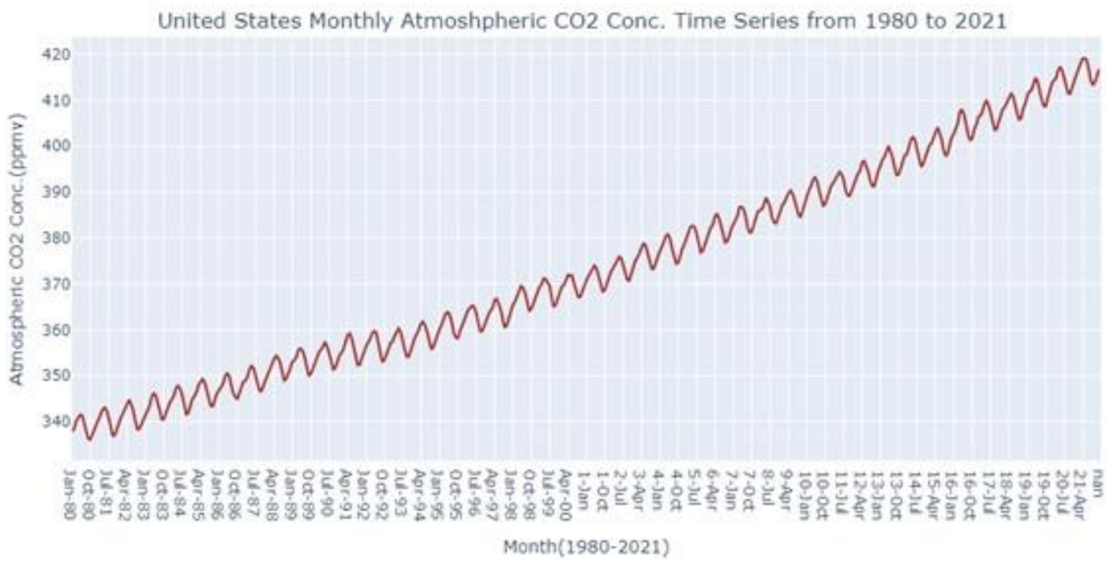

Figure 4.7: Monthly Atmospheric CO_2 Conc. trend from 1980 to 2021.

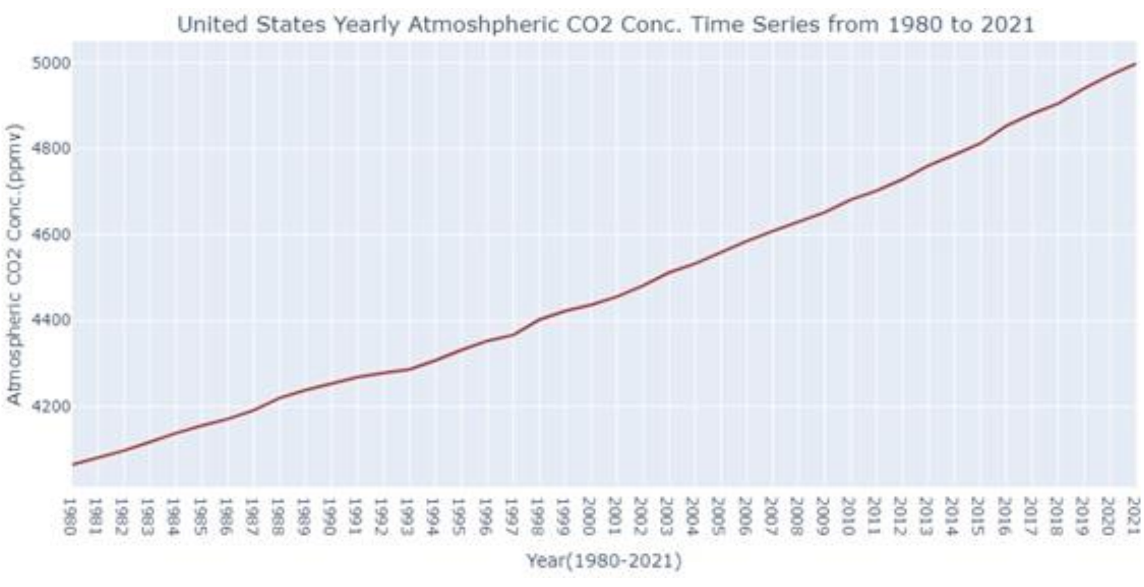

Figure 4.8: Annual Atmospheric CO_2 Conc. trend from 1980 to 2021.

V. Annual Percentage Change in Climatic Variables

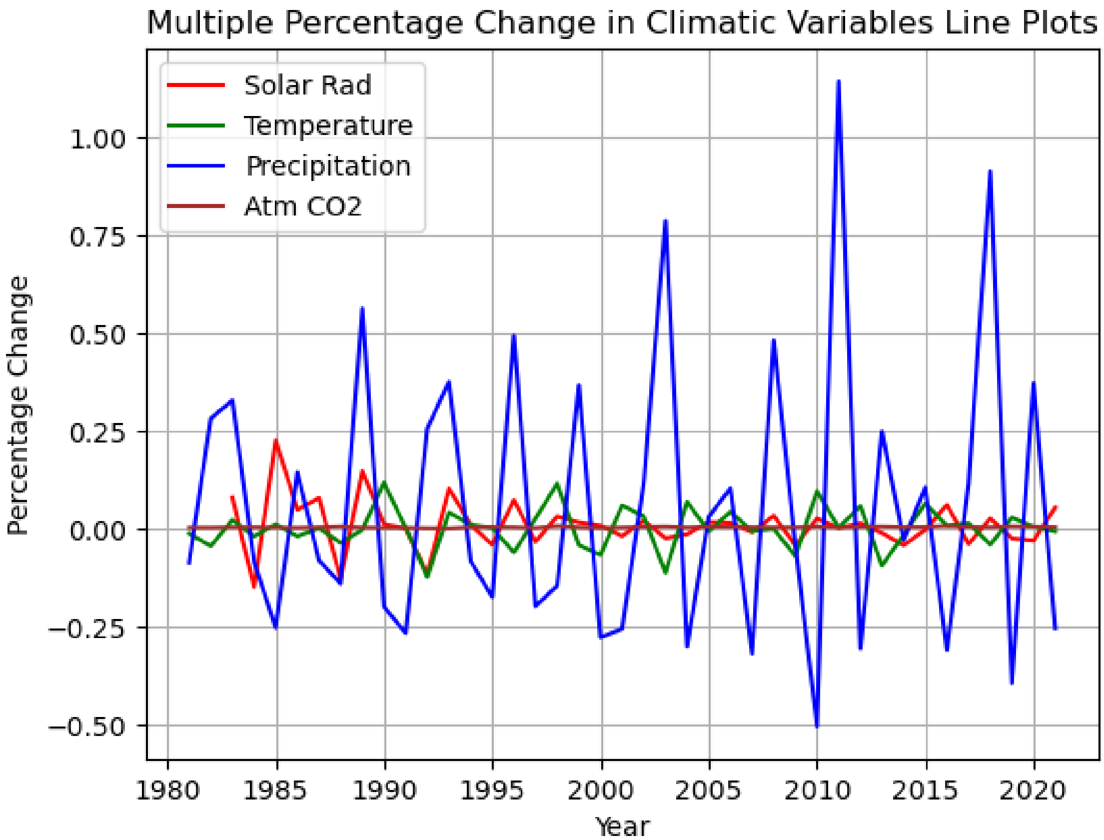

Figure 4.9: Annual Percentage Change in Climatic Variables.

The meteorological variables exhibit significantly different annual percentage fluctuations in their trends. The percentage change in these climatic variables over the period of 41 years is shown in Figure 4.9. A deeper examination of the graph shows that precipitation, particularly between the years 2010 and 2012, has the greatest percentage variance in magnitude. From 2009 to 2010, it experienced a percentage decrease afterwards from 2010 to 2011, it increased by almost 127%. That wasn't the first time the annual precipitation pattern had gone through an increase or decrease of close to or greater than 100% over the years. This indicates that droughts are frequent in the US and are usually followed by a year with high precipitation. There hasn't been a noticeable change in the rates of change over time since the Atmospheric CO_2 concentration has largely increased at the same rate throughout the study's time frame. Prior to being mostly stable with little volatility, solar radiation experienced some percentage variation between

1983 and roughly 1985. There hasn't actually been a substantial percentage variation in the temperature.

b) Crop yield Trends

The yield of these 4 major food crops which supply the world population with the major daily nutritional value is compiled over the period of 41 years (1980 to 2021). The trends of the yield of these individual crops are given below.

I. Maize

The change or fluctuation in the pattern of the maize yield(hg/ha) over a 41-year period is shown in Figure 4.10. The graph shows that the overall yields have increased on average steadily over this time period. However, some years, like 1983, 1998, and 1993, had a poor yield, while others, like 1982, 1987, 1992, 1994, 2004, and 2016, had a good yield. There was a significant drop in yield in 2012. Note that poor or good in this context refers to the year being considered as well as the year before and after.

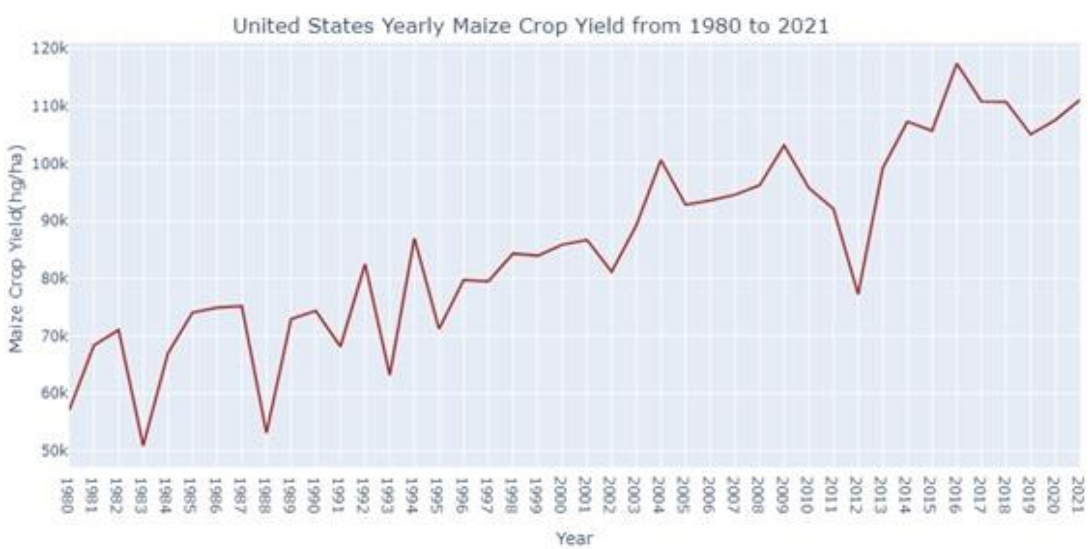

Figure 4.10: Maize yield (hg/ha) trend from 1980 to 2021

II. Rice

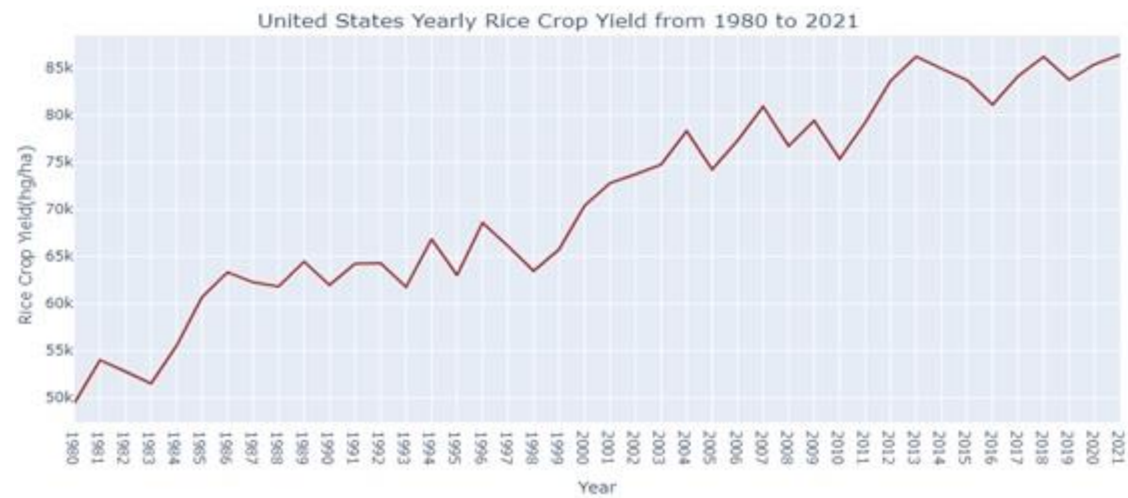

Figure 4.11: Rice Yield Trend from 1980 to 2021

The variation in the pattern of rice yield(hg/ha) over the 41-year period taken into account in this study is shown in Figure 4.11 above. The yield has been increasing steadily throughout this time period, and the graph reveals a specific trend. A substantial increase over a period of roughly 4 to 6 years, followed by a small decline in yield before another regular fluctuation, and then significant growth.

III. Wheat

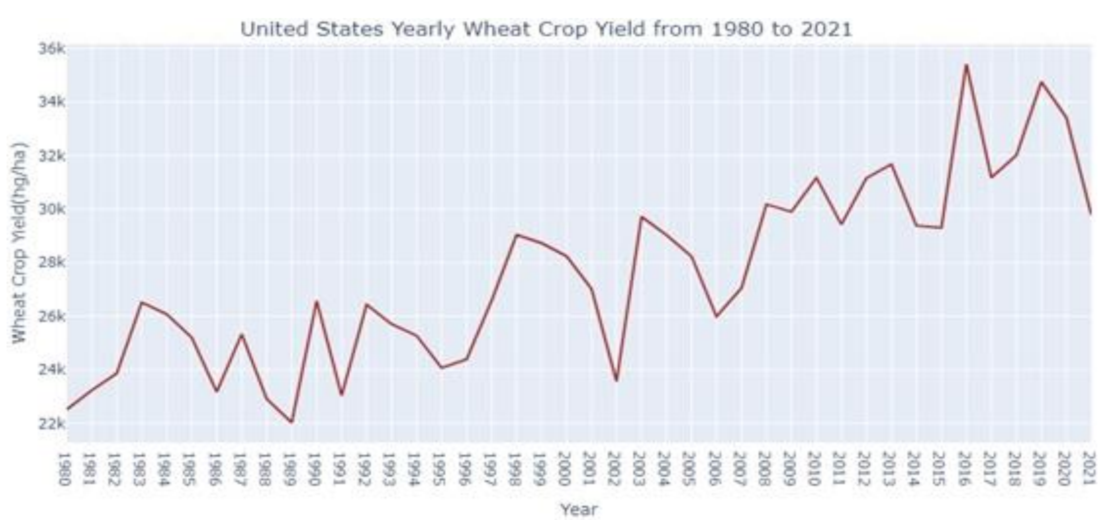

Figure 4.12: Wheat Yield trend from 1980 to 2021

Figure 4.12 shows the variations in wheat yield in the USA, from 1980 to 2021. Wheat yield has fluctuated over this time with highs and lows, but overall, the average yearly output is gradually rising. Some years, like 1989, 1991, 2002, 2006, and 2015, had low yields, while others, like 1990, 1992, 1998, 2003, and 2016, had high yields.

IV. Potato

The trend of the potato yield(hg/ha) during a 41-year period in the USA is shown in Figure 4.13. Over this period, the yield has steadily increased with just a little amount of low output, but even in that case, it still indicates an increase over the previous year when there was a terrible yield. This is a notable increase from this time period.

V. Percentage Change in Crop Yield Variables.

These important food crops experience variable rates of change in yields as a percentage. A closer look at the graph in Figure 4.14 reveals that maize has had a significant percentage change throughout the years from 1980 to about 1996 before it started to stabilize for about 12 years, and it then experienced another significant variation between 2012 and 2015. The percentage change for wheat yield varies slightly as well. The percentage changes in yield for potatoes and rice were not all that substantial over this period of time because the rate of their growth was consistent.

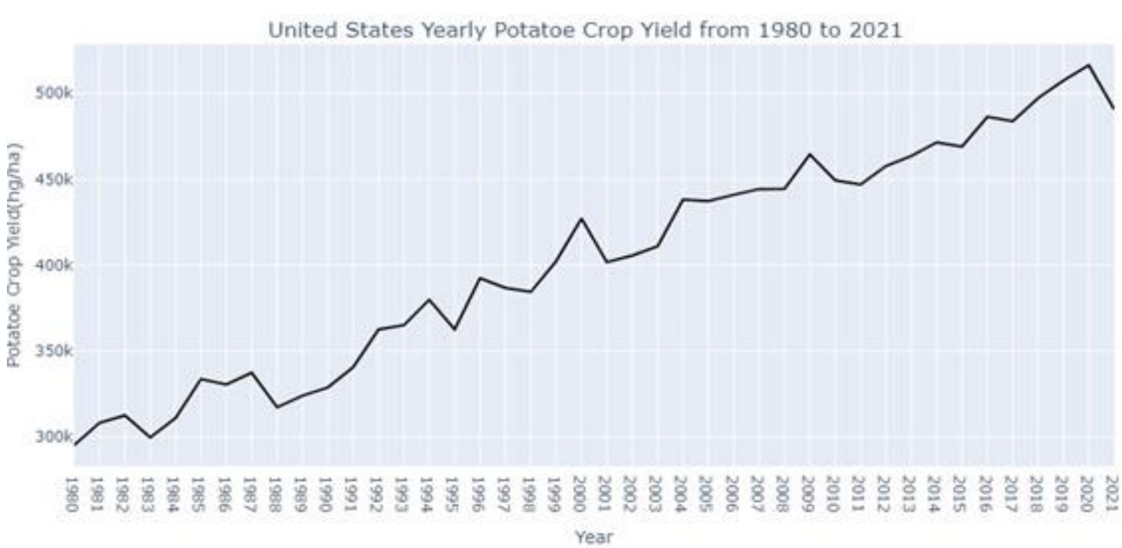

Figure 4.13: Total Potato yield trend from 1980-2021

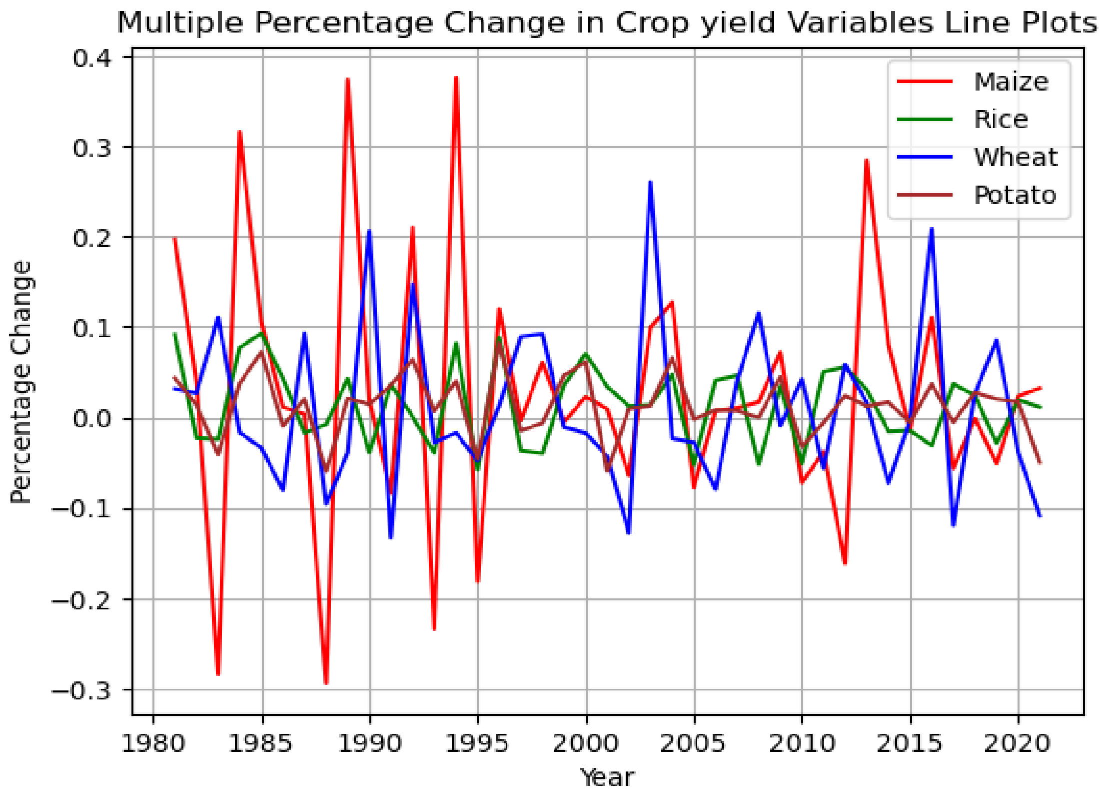

Figure 4.14: Percentage Change in Crop Yield Variables.

4.3. OUTPUT 2: Correlation Analysis

This analysis evaluates the intricate relationships between each meteorological factor and each staple food crop's crop yield. It is useful for understanding how variations in one climatic element affect the crop yield of these staple crops. The degree and direction of the linear link between these variables are quantified by the correlation coefficient. The scale runs from -1 to 1, with -1 denoting a fully negative linear relationship, 0 denoting no relationship, and 1 denoting a perfectly positive linear relationship. The correlation between each climatic factor as it relates to the crop yield is given below.

a) Solar Radiation

I. Figure 4.15 shows the scattered plot of the correlation between Solar radiation and Maize Yield. The correlation coefficient is 0.65 which means Solar Radiation has a fairly strong positive correlation with the yield of maize. An increase in solar radiation will lead to a fairly strong positive response in the yield of Maize.

II. Figure 4.16 shows the scattered plot of the correlation between solar radiation and Rice yield. The correlation coefficient is 0.77 which shows a strong positive correlation. An increase in Solar radiation will lead to a corresponding positively strong response in the yield of Rice.

III. Figure 4.17 shows the scattered plot of the correlation between solar radiation and Wheat yield. The correlation coefficient is 0.53 which shows a positive and moderate correlation. An increase in Solar radiation will lead to a corresponding moderate response in the yield of wheat.

IV. Figure 4.18 shows the scattered plot of the correlation between solar radiation and Potato yield. The correlation coefficient is 0.76 which shows a positive and strong correlation. An increase in Solar radiation will lead to a corresponding strongly positive response in the yield of Potato.

b) Temperature

I. Figure 4.19 shows the scattered plot of the correlation between Temperature and Maize Yield. The correlation coefficient is 0.46 which means Temperature has a positive and moderate correlation with maize yield. An increase in temperature would lead to a corresponding moderate positive response in the yield of Maize.

II. Figure 4.20 shows the scattered plot of the correlation between Temperature and Rice Yield. The correlation coefficient is 0.54 which means the temperature has a positive and moderate correlation with Rice yield. An increase in Temperature would lead to a corresponding positive and moderate response in the yield of Rice.

III. Figure 4.21 shows the scattered plot of the correlation between Temperature and Wheat Yield. The correlation coefficient is 0.52 which means the temperature is moderately correlated with Wheat yield. An increase in Temperature would lead to a corresponding positive moderate response in the yield of Wheat.

IV. Figure 4.22 shows the scattered plot of the correlation between Temperature and Potato Yield. The correlation coefficient is 0.53 which means the temperature is positive and moderately correlated with Potato yield. An increase in Temperature would lead to a corresponding positive moderate response in the yield of Potato.

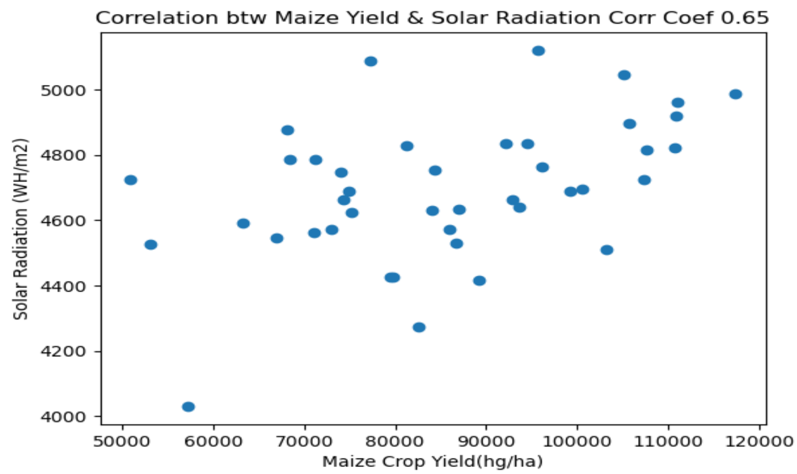

Figure 4.15: Correlation of Maize Crop yield and Solar Radiation.

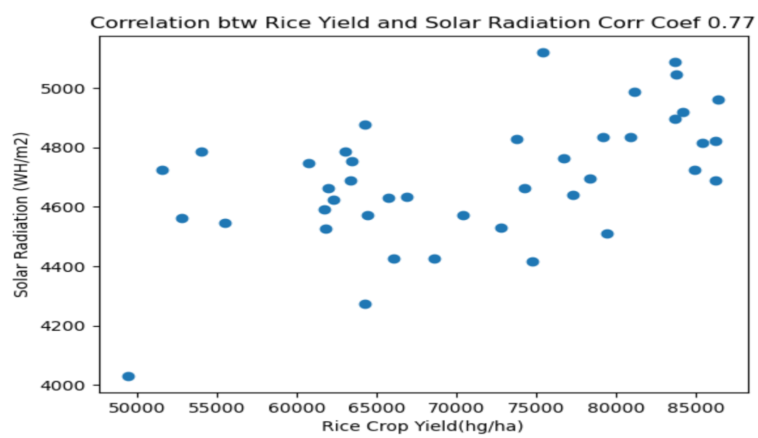

Figure 4.16: Correlation of Rice Yield and Solar Radiation

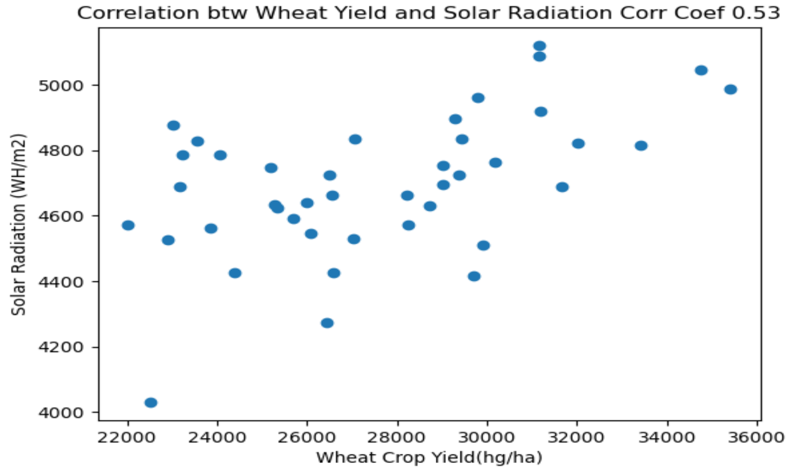

Figure 4.17: Correlation of Wheat Yield and Solar Radiation.

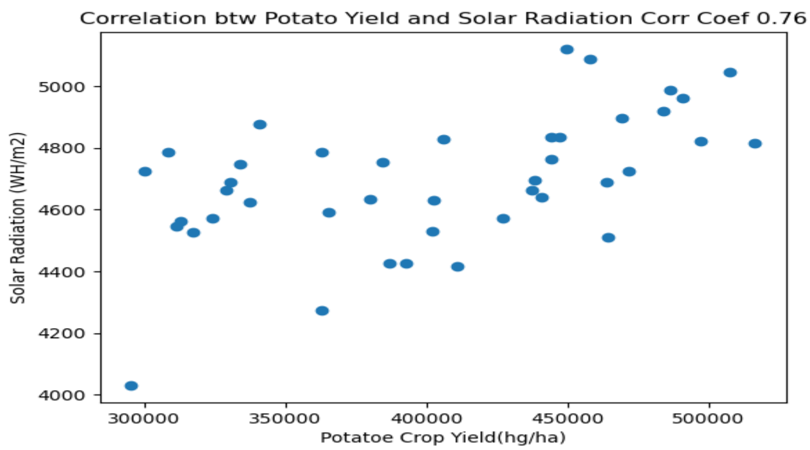

Figure 4.18: Correlation of Potato and Solar Radiation.

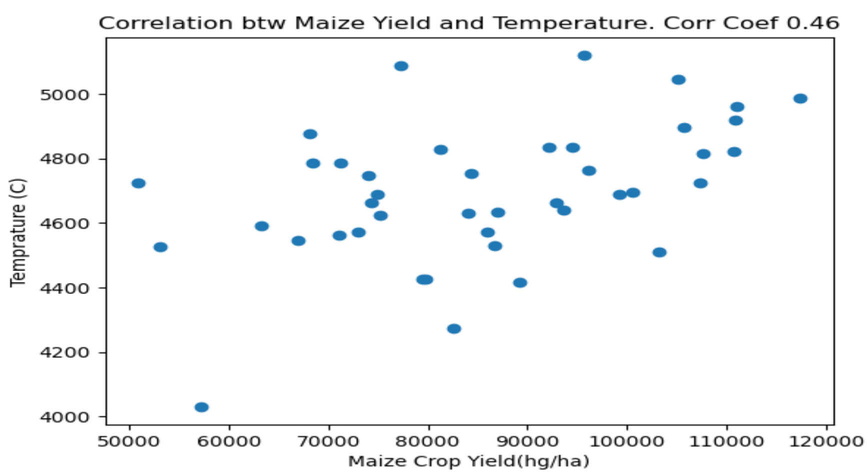

Figure 4.19: Correlation of Maize yield and Temperature.

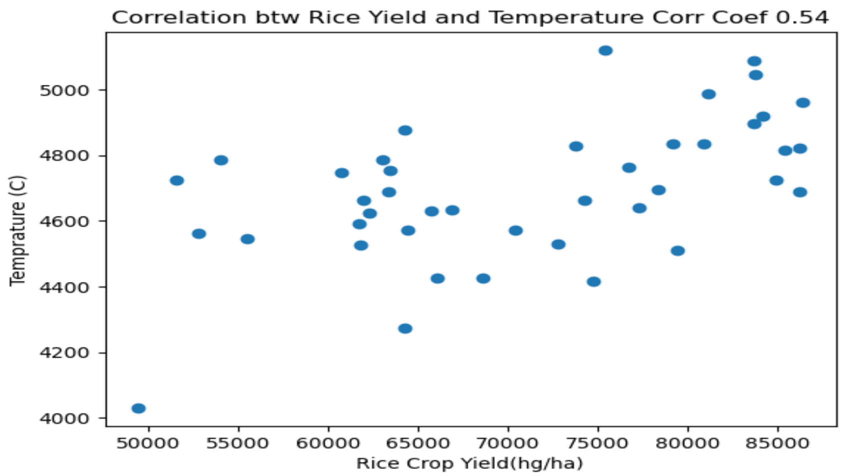

Figure 4.20 Correlation of Temperature and Rice Yield.

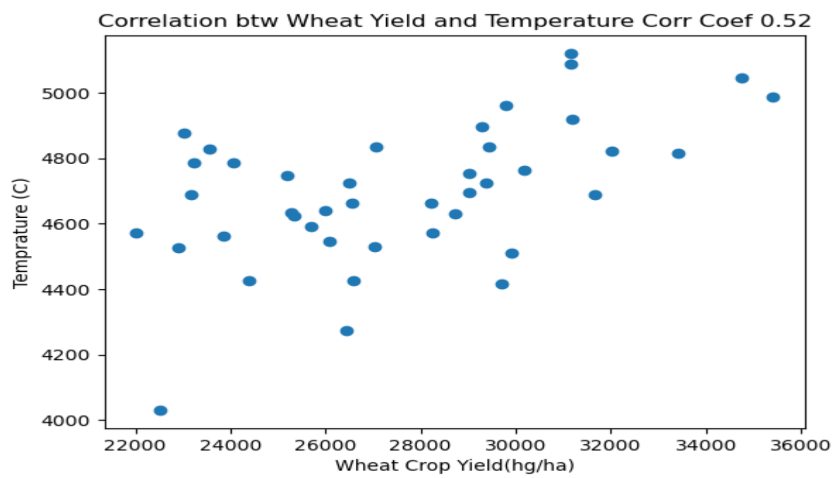

Figure 4.21: Correlation of Wheat yield and Temperature.

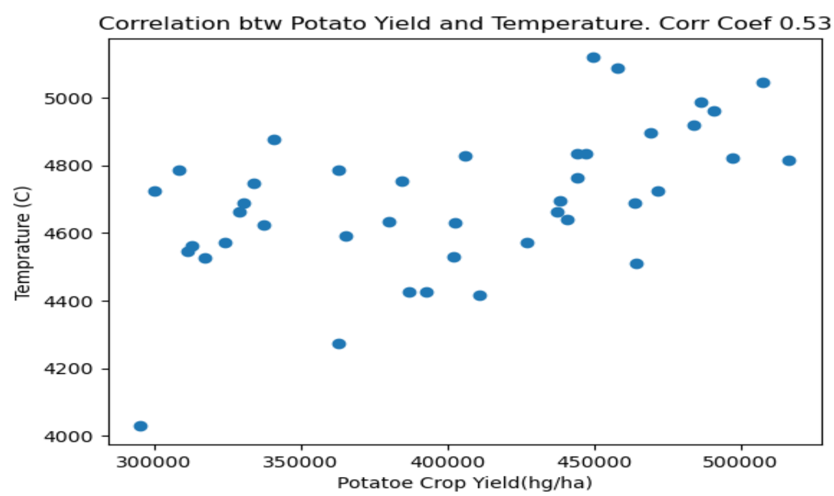

Figure 4.22 Correlation of Temperature and Potato yield.

c) Precipitation

I. Figure 4.23 shows the scattered plot of the correlation between Precipitation and Maize Yield. The correlation coefficient is 0.23 which means the Precipitation has a positive but weak correlation with Maize yield. An increase in Precipitation would lead to a corresponding positive but weak response in the yield of Maize.

II. Figure 4.24 shows the scattered plot of the correlation between Precipitation and Rice Yield. The correlation coefficient is 0.23 which means the Precipitation has a positive but weak correlation with Rice yield. An increase in Precipitation would lead to a corresponding positive but weak response in the yield of Rice.

III. Figure 4.25 shows the scattered plot of the correlation between Precipitation and Wheat Yield. The correlation coefficient is 0.24 which means the Precipitation has a positive but weak correlation with Wheat yield. An increase in Precipitation would lead to a corresponding positive but weak response in the yield of Wheat.

IV. Figure 4.26 shows the scattered plot of the correlation between Precipitation and Potato Yield. The correlation coefficient is 0.27 which means the Precipitation has a positive but weak correlation with Potato yield. An increase in Precipitation would lead to a corresponding positive but weak response in the yield of Potato

d) Atmospheric CO_2

I. Figure 4.27 shows the scattered plot of the correlation between Atmospheric CO_2 and Maize Yield. The correlation coefficient is 0.9 which means the Atmospheric CO_2 has a positive and very strong correlation with Maize yield. An increase in Atmospheric CO_2 would lead to a corresponding positive and very strong response in the yield of Maize.

II. Figure 4.28 shows the scattered plot of the correlation between Atmospheric CO_2 and Rice Yield. The correlation coefficient is 0.96 which means the Atmospheric CO_2 has a positive very strong correlation with Rice yield. An increase in Atmospheric CO_2 would lead to a corresponding positive and very strong response in the yield of Rice.

III. Figure 4.29 below shows the scattered plot of the correlation between Atmospheric CO_2 and Wheat Yield. The correlation coefficient is 0.86 which means the Atmospheric CO_2 has a positive and very strong correlation with Wheat yield. An increase in Atmospheric CO_2 would lead to a corresponding positive and very strong response in the yield of Wheat.

IV. Figure 4.30 shows the scattered plot of the correlation between Atmospheric CO_2 and Potato Yield. The correlation coefficient is 0.27 which means the Atmospheric CO_2 has a positive but weak correlation with Potato yield. An increase in Atmospheric CO_2 would lead to a corresponding positive but weak response in the yield of Potato.

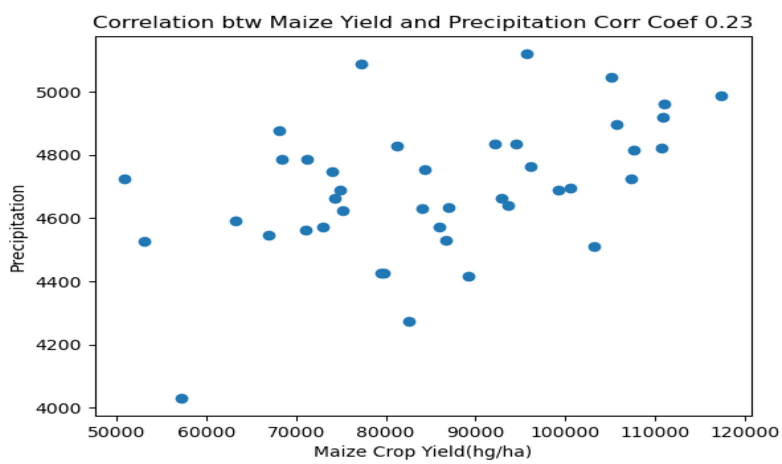

Figure 4.23: Correlation of Precipitation and Maize yield.

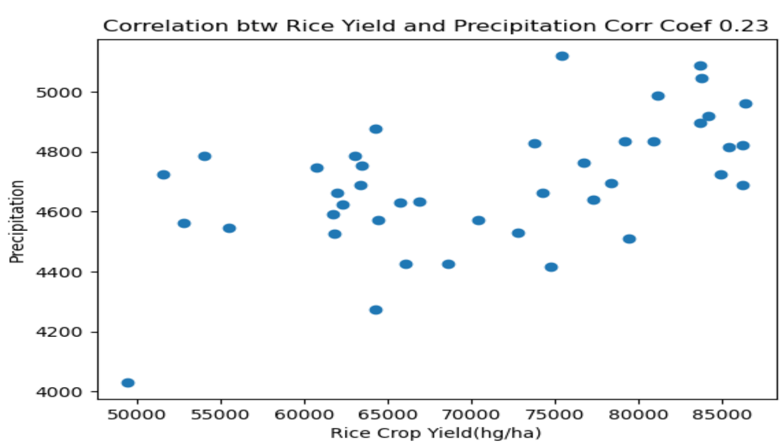

Figure 4.24: Correlation of Precipitation and Rice Yield.

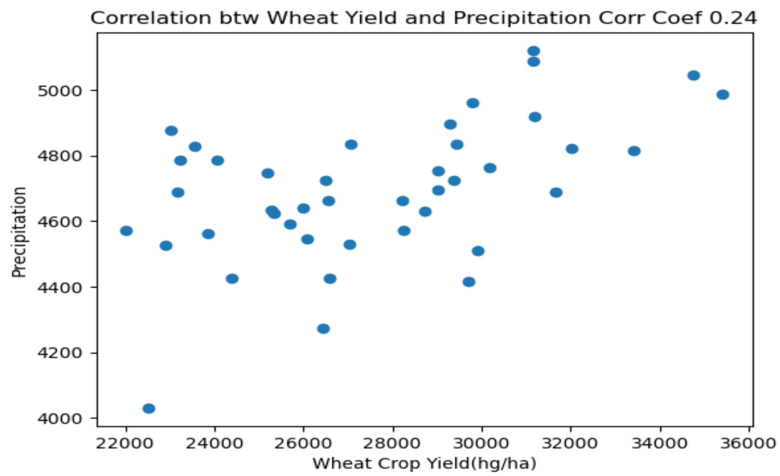

Figure 4.25 Correlation of Precipitation and Wheat Yield.

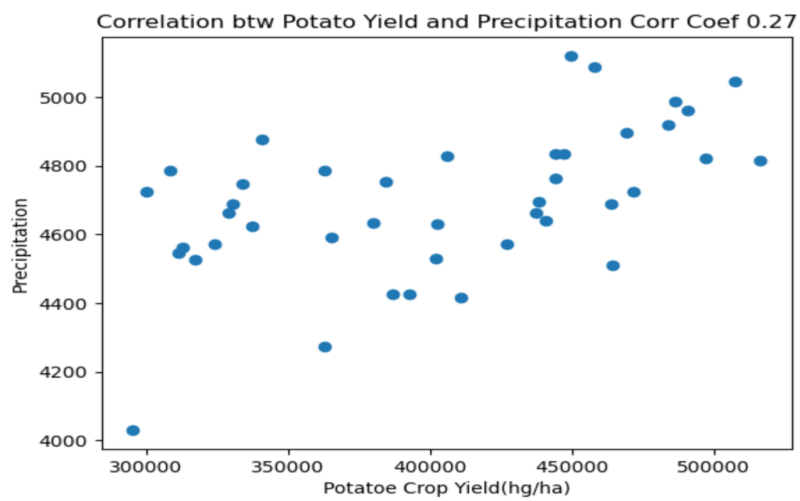

Figure 4.26: Correlation of Precipitation and Potato Yield.

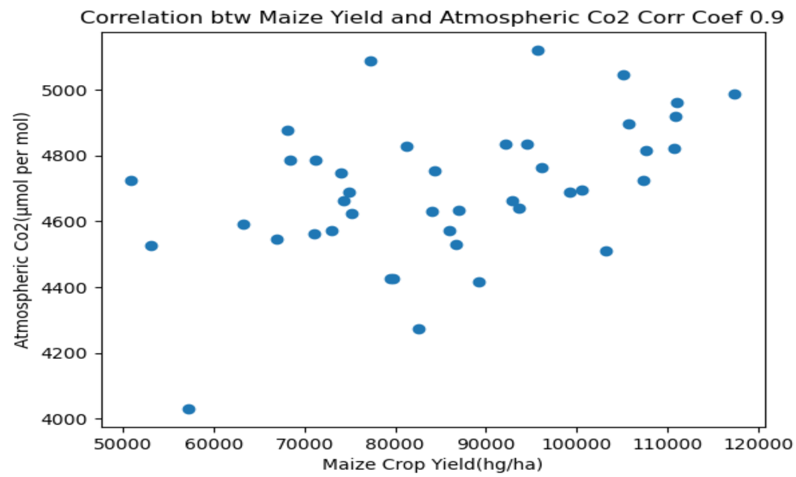

Figure 4.27: Correlation of Atmospheric CO_2 and Maize Yield

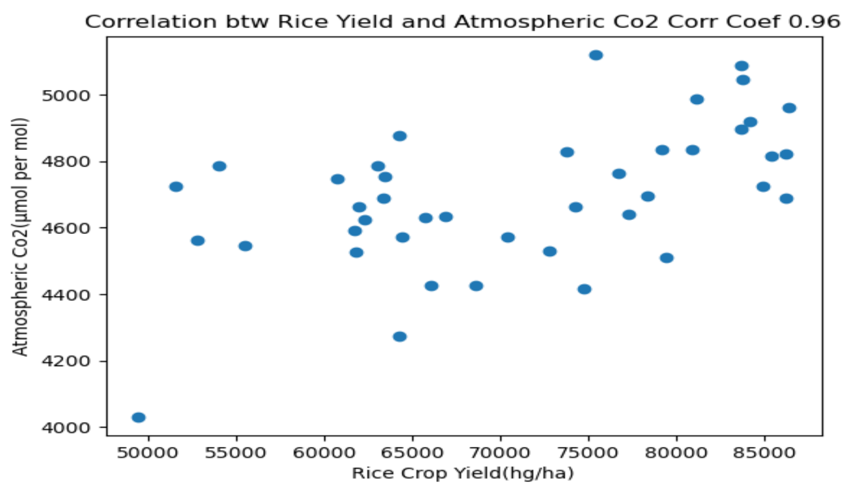

Figure 4.28: Correlation of Atmospheric CO_2 and Rice Yield

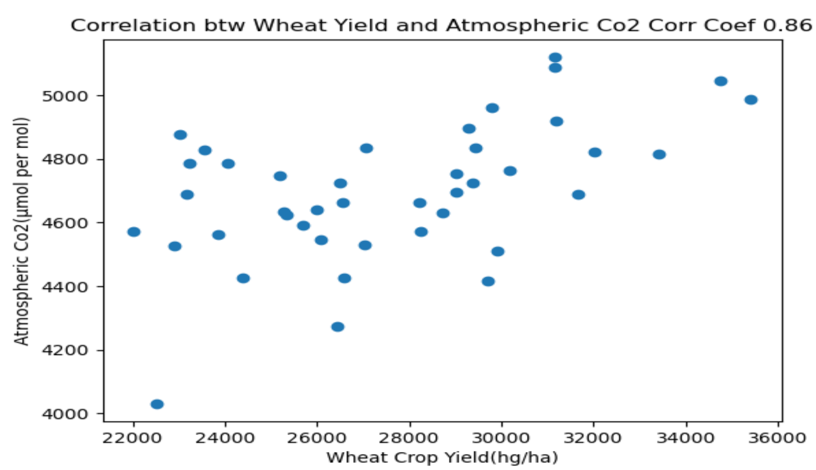

Figure 4.29: Correlation of Atmospheric CO_2 and Wheat Yield

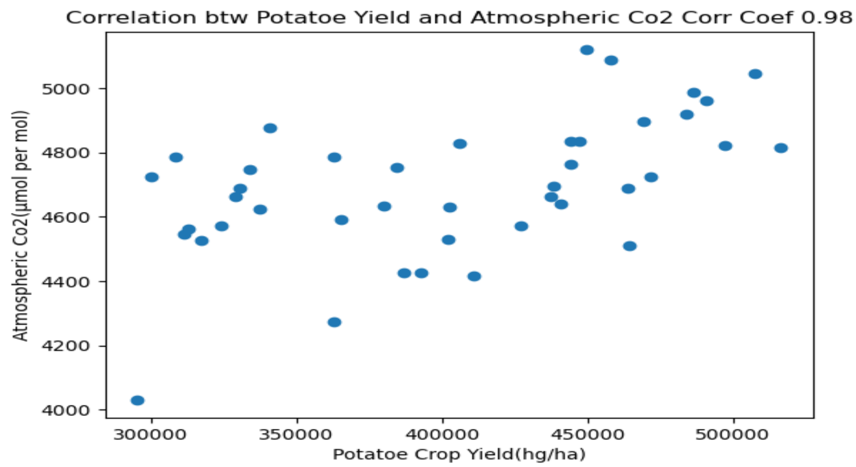

Figure 4.30: Correlation of Atmospheric CO_2 and Potato Yield

Overall, the heatmap showing the correlation between all the climatic variables and crop yield is given below.

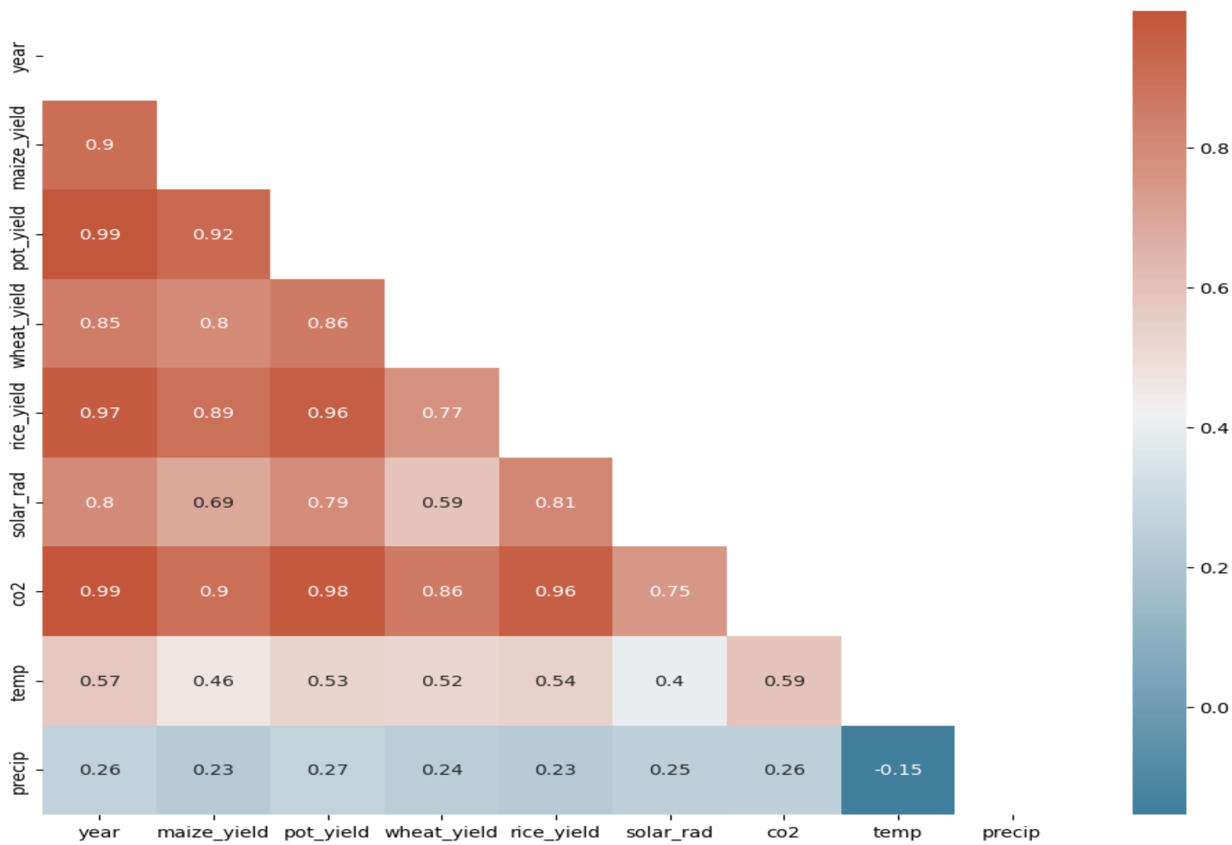

Figure 4.31: Heatmaps of the correlation between Climatic Variables and Crop Yields.

4.4. OUTPUT 3: Application of Machine Learning Algorithm.

The following Machine learning Algorithms were used to analyze the Yields of these crops as the *Dependent Variables* and all the climatic variables as the *Independent Variables*. For this investigation, three machine learning algorithms—Support Vector Regression (SVR), Decision Tree Regression (DTR), and Random Forest Regression (RFR) will be used. These three have different hyperparameters that are used to improve their performance model. The hyperparameters employed in this analysis using Python on a Jupyter Notebook are displayed in Table 4.3.2.

Model	Parameters	Description	**values**
Support Vector Regression	Gamma Kernel	This control trade-off between model flexibility and generalization. The kernel function needed for a non-linear relationship	0.01 rbf
Decision Tree Regression	Max_depth Minimum Sample_Split	The maximum depth of the tree. The minimum number of samples required to split an internal node.	3 5
Random Forest Regression	n_estimators Random State	Number of trees used in the random forest The seed used by the random number generator reproducibility	100 5

Table 4.1: Hyperparameters used in the application of the ML algorithms.

Furthermore, the performance metrics of each of the machine learning algorithm used in this analysis is given in Table 4.2 The outcome shows that random forest regression has the best performance metrics and the least error when it was applied to create a model for the yields of Maize, Rice, Wheat Potato.

Crop Yields	Support Vector Regression				Decision Tree Regression				Random Forest Regression			
	R2	MAE	MSE	RMSE	R2	MAE	MSE	RMSE	R2	MAE	MSE	RMSE
maize	0.74	0.17	0.05	0.22	0.69	0.49	0.45	0.67	0.77	0.48	0.34	0.58
Rice	0.80	0.28	0.10	0.32	0.93	0.25	0.09	0.31	0.89	0.30	0.15	0.38
Wheat	0.49	0.35	0.19	0.44	0.28	0.67	0.74	0.86	0.52	0.61	0.50	0.70
Potato	0.91	0.14	0.04	0.20	0.96	0.19	0.06	0.24	0.97	0.16	0.04	0.20

Table 4.2: Performance Metrics of SVR, DTR and RFR analysis.

R2 R-Squared Sore

MAE Mean Absolute Error

MSE Mean Square Error

RMSE Root Mean Square Error

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study examines the impact of climate change on agricultural yields of four staple food crops in the USA over a 41-year period. The research utilizes data science tools, trend analysis, correlation analysis, and machine learning algorithms to analyze four meteorological variables (Solar Radiation, Temperature, Precipitation, and Atmospheric CO_2 concentration) in relation to the yields of Maize, Rice, Wheat, and Potato crops. The technological approach aims to build a predictive model to assess the influence of these climatic factors on crop yields.

Over the years 1980 to 2021, a longitudinal analysis was conducted to examine the trends and patterns of each of the meteorological variables.

a). The study examined climatic variables using time series visualization, but yearly visualization was found to be more interpretable. Total solar radiation increased over the years with fluctuations, while temperature showed more fluctuations but had an overall increasing trend. Precipitation exhibited the greatest percentage variance, with significant fluctuations, including a sharp decrease in 2010 and a remarkable increase of almost 127% in 2011. Atmospheric CO_2 showed a noticeable uptick since 1980. Solar radiation had some percentage variation in the early 1980s, but the temperature remained relatively stable throughout the study period as shown in Chapter 4 section 4.2(a).

b). The yearly trends of crop yields were visualized, showing a general increase in all crops' yields. Maize had six new record yields, but there were five years with poor yields, including a significant drop in 2012 compared to yields from 2006 to 2021. Rice yield increased steadily with some fluctuations, while wheat yield showed varying trends. Potato yield was similar to rice yield, relatively steady with minor fluctuations. The percentage change in maize yield exhibited high variance from 1980 to 1996, stabilized for about 12 years, and had another significant variation between 2012 and 2015. Wheat yield showed slight variations in percentage change, while potatoes and rice had relatively consistent growth rates with fewer fluctuations as shown in Chapter 4 section 4.2(b).

Also, the second objective is the correlation analysis between each of the climatic variables against the yield of the 4 crops. According to the results, the yields of these crops have a positive correlation with solar radiation, with rice and potato yields showing a particularly strong positive correlation (0.77). Additionally, with an average coefficient of 0.93, they are all strongly positively correlated with Atmospheric CO_2 concentration. However, these yields have a weak correlation to precipitation with the highest correlation coefficient being 0.27 and that's for potatoes. Furthermore, their yields typically show a mildly beneficial correlation with temperature. On this basis, it would seem reasonable to conclude that solar radiation and Atmospheric CO_2 have a significant beneficial effect on the yield of maize, rice, wheat, and potatoes. However, there is a limit to the amount of temperature and precipitation needed before it starts having an adverse effect on them as shown in all the scattered plots in Chapter 4 section 4.3.

Finally, the third objective is the application of machine learning algorithms (Support Vector Regression (SVR), Decision Tree Regression (DTR) and Random Forest Regression (RFR)) and the evaluation of their performance Metrics. The 3 Machine learning algorithms seem to be good for the Potato and Rice yield model because of their high R2-score which is an average of 95% and 87% accuracy respectively. DTR performed poorly for Wheat yield with an R2-score of 28% while SVR and RFR performed a bit better but still poor with an R2 Score of 49% and 52% which is not a good model. RFR regression performed best for the Maize yield Model with an R2 score of 77%, SVR followed up with 74% and DTR had a 69% R2 score. Overall, Random Forest Regression has the best model for all the yields of the crops and is more suitable for this analysis as shown in the tables in Chapter 4 Section 4.4.

5.2. **Recommendation**

This study provides insight into the technological analysis impact of climate change on food production and also developed a predictive model for the different staple food crops considered. The following recommendations are made to improve this study in light of the conclusions and analyses provided in previous chapters.

1. Climatic variables: Additional climatic factors can be taken into account, such as humidity, which affects crop water requirements and can influence the prevalence of disease in crops. Temperature can also be broken down into maximum and minimum daily temperatures, which are related to crop yield. Also, wind speed can be considered.
2. Extreme Weather Events: Data on extreme weather events, such as heatwaves, drought, hurricanes, and flooding in addition to this analysis would improve this study.
3. Agricultural Factors: The type of soil and fertilizer used to cultivate the soil can have a significant impact on the production of the crops. The yields are also impacted by farming practices and insect management.

The results of this study would be much enhanced if all the pertinent information and facts about these recommendations were available.

REFERENCES

Abbas, S. and Mayo, Z.A., 2021. Impact of temperature and rainfall on rice production in Punjab, Pakistan. *Environment, Development and Sustainability*, *23*(2), pp.1706-1728.

Abd-Elmabod, S.K., Muñoz-Rojas, M., Jordán, A., Anaya-Romero, M., Phillips, J.D., Jones, L., Zhang, Z., Pereira, P., Fleskens, L., van Der Ploeg, M. and de la Rosa, D., 2020. Climate change impacts on agricultural suitability and yield reduction in a Mediterranean region. *Geoderma*, *374*, p.114453.

Agnolucci, P., Rapti, C., Alexander, P., De Lipsis, V., Holland, R.A., Eigenbrod, F. and Ekins, P., 2020. Impacts of rising temperatures and farm management practices on global yields of 18 crops. Nature Food, 1(9), pp.562-571.

Ainsworth, E.A. and Long, S.P., 2021. 30 years of free-air carbon dioxide enrichment (FACE): what have we learned about future crop productivity and its potential for adaptation?. Global Change Biology, 27(1), pp.27-49.

Ainsworth, E.A. and Long, S.P., 2005. What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. New phytologist, 165(2), pp.351-372.

Ainsworth, E.A., 2008. Rice production in a changing climate: a meta-analysis of responses to elevated carbon dioxide and elevated ozone concentration. Global Change Biology, 14(7), pp.1642-1650.

Amiri, S., Eyni-Nargeseh, H., Rahimi-Moghaddam, S. and Azizi, K., 2021. Water use efficiency of chickpea agro-ecosystems will be boosted by positive effects of CO_2 and using suitable genotype× environment× management under climate change conditions. *Agricultural Water Management*, *252*, p.106928.

Anderson, R., Bayer, P.E. and Edwards, D., 2020. Climate change and the need for agricultural adaptation. Current opinion in plant biology, 56, pp.197-202.

Andreas, A.; Stoffel, T.; (1981). NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data); NREL Report No. DA-

5500-56488. Accessed [June 3rd 2023] http://dx.doi.org/10.5439/1052221 https://midcdmz.nrel.gov/apps/html.pl?site=BMS;page=historical

Ani, K.J., Anyika, V.O. and Mutambara, E., 2022. The impact of climate change on food and human security in Nigeria. International Journal of Climate Change Strategies and Management, 14(2), pp.148-167.

Aryal, J.P., Sapkota, T.B., Khurana, R., Khatri-Chhetri, A., Rahut, D.B. and Jat, M.L., 2020. Climate change and agriculture in South Asia: Adaptation options in smallholder production systems. *Environment, Development and Sustainability*, *22*(6), pp.5045-5075.

Bali, N. and Singla, A., 2022. Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. Archives of computational methods in engineering, pp.1-18.

Bassu, S., Brisson, N., Durand, J.L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C. and Basso, B., 2014. How do various maize crop models vary in their responses to climate change factors?. *Global change biology*, *20*(7), pp.2301-2320.

Bai, Y., Ochuodho, T.O. and Yang, J., 2019. Impact of land use and climate change on water-related ecosystem services in Kentucky, USA. Ecological Indicators, 102, pp.51-64.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R. and Chhetri, N., 2014. A meta-analysis of crop yield under climate change and adaptation. *Nature climate change*, *4*(4), pp.287-291.

Chawdhery, M.R.A., Al-Mueed, M., Wazed, M.A., Emran, S.A., Chowdhury, M.A.H. and Hussain, S.G., 2022. Climate change impacts assessment using crop simulation model intercomparison approach in northern indo-gangetic basin of Bangladesh. International Journal of Environmental Research and Public Health, 19(23), p.15829.

Cui, H.Y., Jin, L.B., Li, B., Dong, S.T., Liu, P., Zhao, B. and Zhang, J.W., 2013. Effects of shading on dry matter accumulation and nutrient absorption of summer maize. Ying Yong Sheng tai xue bao= The Journal of Applied Ecology, 24(11), pp.3099-3105.

Ding, Z., Ali, E.F., Elmahdy, A.M., Ragab, K.E., Seleiman, M.F. and Kheir, A.M., 2021. Modeling the combined impacts of deficit irrigation, rising temperature and compost application on wheat yield and water productivity. *Agricultural Water Management*, *244*, p.106626.

Deng, N., Ling, X., Sun, Y., Zhang, C., Fahad, S., Peng, S., Cui, K., Nie, L. and Huang, J., 2015. Influence of temperature and solar radiation on grain yield and quality in irrigated rice system. European Journal of Agronomy, 64, pp.37-46.

Evans, L.T. and De Datta, S.K., 1979. The relation between irradiance and grain yield of irrigated rice in the tropics, as influenced by cultivar, nitrogen fertilizer application and month of planting. Field Crops Research, 2, pp.1-17.

Early, E.B., McIlrath, W.O., Seif, R.D. and Hageman, R.H., 1967. Effects of Shade Applied at Different Stages of Plant Development on Corn (Zea mays L.) Production 1. Crop Science, 7(2), pp.151-156.

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K. and Prasanna, B.M., 2022. Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, *14*(5), pp.1295-1319.

Fishman, R., 2016. More uneven distributions overturn benefits of higher precipitation for crop yields. Environmental Research Letters, 11(2), p.024004.

Fukuda, S., Spreer, W., Yasunaga, E., Yuge, K., Sardsud, V. and Müller, J., 2013. Random Forests modelling for the estimation of mango (Mangifera indica L. cv. Chok Anan) fruit yields under different irrigation regimes. *Agricultural water management*, *116*, pp.142-150.

Food and Agriculture Organization of the United Nations. [FAOSTAT]. [Crops and livestock products]. Latest update: [March 24, 2023]. Accessed: [June 6th 2023]. [https://www.fao.org/faostat/en/#data/QCL]

Food and Agriculture organization of the United Nations 2020. "The state of food and Agriculture" Accessed July 9th 2023 https://www.fao.org/3/cb1447en/cb1447en.pdf

Gay, C., Estrada, F., Conde, C., Eakin, H. and Villers, L., 2006. Potential impacts of climate change on agriculture: a case of study of coffee production in Veracruz, Mexico. *Climatic Change*, *79*(3-4), pp.259-288.

Ghini, R., Torre-Neto, A., Dentzien, A.F., Guerreiro-Filho, O., Iost, R., Patrício, F.R., Prado, J.S., Thomaziello, R.A., Bettiol, W. and DaMatta, F.M., 2015. Coffee growth, pest and yield responses to free-air CO 2 enrichment. Climatic Change, 132, pp.307-320.

Gonocruz, R.A., Nakamura, R., Yoshino, K., Homma, M., Doi, T., Yoshida, Y. and Tani, A., 2021. Analysis of the rice yield under an Agrivoltaic system: A case study in Japan. Environments, 8(7), p.65.

Gul, A., Chandio, A.A., Siyal, S.A., Rehman, A. and Xiumin, W., 2022. How climate change is impacting the major yield crops of Pakistan? an exploration from long-and short-run estimation. *Environmental Science and Pollution Research*, *29*(18), pp.26660-26674.

Guntukula, R. and Goyari, P., 2020. The impact of climate change on maize yields and its variability in Telangana, India: A panel approach study. *Journal of Public Affairs*, *20*(3), p.e2088.

Hoogenboom, G., Porter, C.H., Boote, K.J., Shelia, V., Wilkens, P.W., Singh, U., White, J.W., Asseng, S., Lizaso, J.I., Moreno, L.P. and Pavan, W., 2019. The DSSAT crop modeling ecosystem. In *Advances in crop modelling for a sustainable agriculture* (pp. 173-216). Burleigh Dodds Science Publishing.

Hou, P., Liu, Y., Xie, R., Ming, B., Ma, D., Li, S. and Mei, X., 2014. Temporal and spatial variation in accumulated temperature requirements of maize. Field Crops Research, 158, pp.55-64.

Hsiang, S., Lobell, D., Roberts, M. and Schlenker, W., 2013. Climate and crop yields in Australia. Brazil, China, Europe and the United States

Hu, S., Chen, W., Tong, K., Wang, Y., Jing, L., Wang, Y. and Yang, L., 2022. Response of rice growth and leaf physiology to elevated CO_2 concentrations: A meta-analysis of 20-year FACE studies. *Science of the Total Environment*, *807*, p.151017.

Hu, S., Wang, Y. and Yang, L., 2021. Response of rice yield traits to elevated Atmospheric CO_2 concentration and its interaction with cultivar, nitrogen application rate and temperature: a meta-analysis of 20 years FACE studies. Science of the Total Environment, 764, p.142797.

Hulme, M., Osborn, T.J. and Johns, T.C., 1998. Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simulations. Geophysical research letters, 25(17), pp.3379-3382.

Jennings, S.A., Koehler, A.K., Nicklin, K.J., Deva, C., Sait, S.M. and Challinor, A.J., 2020. Global potato yields increase under climate change with adaptation and CO_2 fertilisation. *Frontiers in Sustainable Food Systems*, *4*, p.519324.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R. and Kim, S.H., 2016. Random forests for global and regional crop yield predictions. *PloS one*, *11*(6), p.e0156571.

Jiang, R., He, W., He, L., Yang, J.Y., Qian, B., Zhou, W. and He, P., 2021. Modelling adaptation strategies to reduce adverse impacts of climate change on maize cropping system in Northeast China. *Scientific Reports*, *11*(1), p.810.

Karimi, V., Karami, E. and Keshavarz, M., 2018. Climate change and agriculture: Impacts and adaptive responses in Iran. *Journal of Integrative Agriculture*, *17*(1), pp.1-15

Karl, T.R., Arguez, A., Huang, B., Lawrimore, J.H., McMahon, J.R., Menne, M.J., Peterson, T.C., Vose, R.S. and Zhang, H.M., 2015. Possible artifacts of data biases in the recent global surface warming hiatus. Science, 348(6242), pp.1469-1472.

Katny, M.A.C., Hoffmann-Thoma, G., Schrier, A.A., Fangmeier, A., Jäger, H.J. and van Bel, A.J., 2005. Increase of photosynthesis and starch in potato under elevated CO_2 is dependent on leaf age. *Journal of plant physiology*, *162*(4), pp.429-438.

Keeling, R.F. and Graven, H.D., 2021. Insights from time series of atmospheric carbon dioxide and related tracers. Annual Review of Environment and Resources, 46, pp.85-110.

Khaki, S., Wang, L. and Archontoulis, S.V., 2020. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, *10*, p.1750.

Kogo, B.K., Kumar, L. and Koech, R., 2021. Climate change and variability in Kenya: a review of impacts on agriculture and food security. Environment, Development and Sustainability, 23, pp.23-43.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors*, *18*(8), p.2674.

Liu, Y., Hou, P., Xie, R., Li, S., Zhang, H., Ming, B., Ma, D. and Liang, S., 2013. Spatial adaptabilities of spring maize to variation of climatic conditions. Crop Science, 53(4), pp.1693-1703.

Lobell, D.B. and Asseng, S., 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environmental Research Letters, 12(1), p.015001.

Lv, Z., Liu, X., Cao, W. and Zhu, Y., 2013. Climate change impacts on regional winter wheat production in main wheat production regions of China. *Agricultural and Forest Meteorology*, *171*, pp.234-248.

Maiorano, A., Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R.P., Ruane, A.C., Semenov, M.A., Wallach, D., Wang, E. and Alderman, P.D., 2017. Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field crops research*, *202*, pp.5-20.

Malhi, G.S., Kaur, M. and Kaushik, P., 2021. Impact of climate change on agriculture and its mitigation strategies: A review. *Sustainability*, *13*(3), p.1318.

Mangal, P., Rajesh, A. and Misra, R., 2020, June. Big data in climate change research: Opportunities and challenges. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) (pp. 321-326). IEEE.

Mauney, J.R., Kimball, B.A., Pinter Jr, P.J., LaMorte, R.L., Lewin, K.F., Nagy, J. and Hendrey, G.R., 1994. Growth and yield of cotton in response to a free-air carbon dioxide enrichment (FACE) environment. Agricultural and Forest Meteorology, 70(1-4), pp.49-67.

Mulungu, K. and Ng'ombe, J.N., 2019. Climate change impacts on sustainable maize production in Sub-Saharan Africa: A review. *Maize Prod. Use*, pp.47-58.

NOAA National Centers for Environmental Information (2023). Climate Change: Global Temperature Accessed July 9, 2023 from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

Patil, D.D., Pandey, V., Acharya, R.R. and Baraiya, L.N., 2018. Effect of intra-seasonal variation in temperature on tuber yield of potato in middle Gujarat using SUBSTOR model. *Journal of Agrometeorology*, *20*(1), pp.22-27.

Paraforos, D., Muzirafuti, A., Randazzo, G. and Lanza, S., 2022. Sustainable Agriculture and Advances of Remote Sensing (Volume 1).

Pickson, R.B., He, G. and Boateng, E., 2021. Impacts of climate change on rice production: evidence from 30 Chinese provinces. *Environment, Development and Sustainability*, pp.1-19.

Pradel, W., Gatto, M., Hareau, G., Pandey, S.K. and Bhardway, V., 2019. Adoption of potato varieties and their role for climate change adaptation in India. *Climate Risk Management*, *23*, pp.114-123.

Rana, A., Dua, V.K., Chauhan, S. and Sharma, J., 2020. Climate change and potato productivity in Punjab—Impacts and adaptation. *Potato Research*, *63*, pp.597-613.

Ranum, P., Peña-Rosas, J.P. and Garcia-Casal, M.N., 2014. Global maize production, utilization, and consumption. *Annals of the new York academy of sciences*, *1312*(1), pp.105-112.

Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A. and Khan, N., 2021. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, pp.63406-63439.

Ray, D.K., West, P.C., Clark, M., Gerber, J.S., Prishchepov, A.V. and Chatterjee, S., 2019. Climate change has likely already affected global food production. *PloS one*, *14*(5), p.e0217148.

Raza, A., Razzaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y. and Xu, J., 2019. Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. *Plants*, *8*(2), p.34.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A. and Luccioni, A.S., 2022. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, *55*(2), pp.1-96.

Saddique, Q., Khan, M.I., Habib ur Rahman, M., Jiatun, X., Waseem, M., Gaiser, T., Mohsin Waqas, M., Ahmad, I., Chong, L. and Cai, H., 2020. Effects of elevated air temperature and CO_2 on maize production and water use efficiency under future climate change scenarios in Shaanxi Province, China. *Atmosphere*, *11*(8), p.843

Samanta, S., Banerjee, S., Mukherjee, A., Patra, P.K. and Chakraborty, P.K., 2020. Determining the radiation use efficiency of potato using sunshine hour data: A simple and costless approach. Spanish Journal of Agricultural Research, 18(2), pp.e0801-e0801.

Santos, E.S., Abreu, M.M., Magalhães, M.C., Viegas, W., Amâncio, S. and Cordovil, C., 2017, April. Nutrients levels in paddy soils and flood waters from Tagus-Sado basin: The impact of farming system. In *EGU General Assembly Conference Abstracts* (p. 17129).

Shiferaw, B., Prasanna, B.M., Hellin, J. and Bänziger, M., 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food security*, *3*, pp.307-327.

Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S. and Singh, A.K., 2021. Crop yield prediction integrating genotype and weather variables using deep learning. Plos one, 16(6), p.e0252402.

Sidhu, B.S., Mehrabi, Z., Ramankutty, N. and Kandlikar, M., 2023. How can machine learning help in understanding the impact of climate change on crop yields?. *Environmental Research Letters*, *18*(2), p.024008.

Sossou, S., Igue, C.B. and Diallo, M., 2019. Impact of climate change on cereal yield and production in the Sahel: case of Burkina Faso. *Asian Journal of Agricultural Extension, Economics & Sociology*, *37*(4), pp.1-11.

Shewry, P.R., 2009. Wheat. *Journal of experimental botany*, *60*(6), pp.1537-1553. Available at https://academic.oup.com/jxb/article/60/6/1537/517393 [Accessed 17th July 2023}

Thoning, K.W., Crotwell, A.M. and Mund, J.W., 1973. Atmospheric carbon dioxide dry air mole fractions from continuous measurements at Mauna Loa. Hawaii, Barrow, Alaska, American Samoa and South Pole, 2019, pp.2021-02. https://doi.org/10.15138/yaf1-bk21

https://www.gml.noaa.gov/ccgg/trends/data.html

Visual Crossing Corporation. (2020). Visual Crossing Weather (2017-2019). [Weather Query Builder]. Accessed (June 2nd 2023). Retrieved from https://www.visualcrossing.com/weather/weather-data-services

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W. and Reynolds, M.P., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nature plants, 3(8), pp.1-13.

Wang, J., Vanga, S.K., Saxena, R., Orsat, V. and Raghavan, V., 2018. Effect of climate change on the yield of cereal crops: A review. *Climate*, *6*(2), p.41.

Wu, J.Z., Zhang, J., Ge, Z.M., Xing, L.W., Han, S.Q., Chen, S.H.E.N. and Kong, F.T., 2021. Impact of climate change on maize yield in China from 1979 to 2016. *Journal of Integrative Agriculture*, *20*(1), pp.289-299.

Warsame, A.A., Sheik-Ali, I.A., Ali, A.O. and Sarkodie, S.A., 2021. Climate change and crop production nexus in Somalia: an empirical evidence from ARDL technique. *Environmental Science and Pollution Research*, *28*(16), pp.19838-19850.

Wilson, D.R., Muchow, R.C. and Murgatroyd, C.J., 1995. Model analysis of temperature and solar radiation limitations to maize potential productivity in a cool climate. Field crops research, 43(1), pp.1-18.

World Health Organization 2022 "UN Report: Global hunger numbers rose to as many as 828 million in 2021" [Accessed July 14, 2023] https://www.who.int/news/item/06-07-2022-un-report--global-hunger-numbers-rose-to-as-many-as-828-million-in-2021

World Agricutural Production 2023, USDA Available at https://apps.fas.usda.gov/psdonline/circulars/production.pdf [Accessed 17th July 2023]

Xie, W., Huang, J., Wang, J., Cui, Q., Robertson, R. and Chen, K., 2020. Climate change impacts on China's agriculture: The responses from market and trade. *China Economic Review*, *62*, p.101256.

Xu, D., 2021. Agricultural climate change based on remote sensing image and emergency material supply management of agriculture, rural areas and farmers. *Arabian Journal of Geosciences*, *14*, pp.1-18.

Yan, H., Harrison, M.T., Liu, K., Wang, B., Feng, P., Fahad, S., Meinke, H., Yang, R., Li Liu, D., Archontoulis, S. and Huber, I., 2022. Crop traits enabling yield gains under more frequent extreme climatic events. Science of the Total Environment, 808, p.152170.

Yang, Y., Xu, W., Hou, P., Liu, G., Liu, W., Wang, Y., Zhao, R., Ming, B., Xie, R., Wang, K. and Li, S., 2019. Improving maize grain yield by matching maize growth and solar radiation. Scientific reports, 9(1), p.3635.

Zhang, J. Effects of light & temperature stress on physiological characteristics of yield and quality in maize (Zea Mays L.). Shandong Agric Univ, Shandong, China (2005).

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P. and Durand, J.L., 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of sciences*, *114*(35), pp.9326-9331.

Zhao, J.R. and Chen, G.P., 1990. Effect of shading treatment at different stages of plant development on grain production of corn (Zea Mays L.) and observations of tip kernel abortion. Scientia Agricultura Sinica, 23(4), pp.28-34.

Zou, L., Wang, L., Li, J., Lu, Y., Gong, W. and Niu, Y., 2019. Global surface solar radiation and photovoltaic power from Coupled Model Intercomparison Project Phase 5 climate models. Journal of Cleaner Production, 224, pp.304-324.

APPENDIX

```
Solar Radiation Research Laboratory (SRRL) Baseline Measurement System (BMS)
39.74 N Latitude     105.18 W Longitude     1829 Meters AMSL     Time Zone -7

November 1981

Global Solar Radiation on a Horizontal Surface
Hourly Integrated and Daily Totals
Instrument: Eppley PSP
Watt-Hours per Square Meter

Day,1,,2,,3,,4,,5,,6,,7,,8,,9,,10,,11,,12,,13,,14,,15,,16,,17,,18,,19,,20,,21,,22,,23,,24,,Daily

1,0,,0,,0,,0,,0,,0,,11,,137,,312,,459,,558,,595,,512,,509,,379,,215,,0,,0,,0,,0,,0,,0,,0,,0,,3687,
2,0,,0,,0,,0,,0,,0,,4,,56,,301,,406,,545,,592,,581,,436,,376,,232,,0,,0,,0,,0,,0,,0,,0,,0,,3529,
3,0,,0,,0,,0,,0,,0,,6,,130,,298,,427,,517,,564,,544,,470,,346,,190,,0,,0,,0,,0,,0,,0,,0,,0,,3492,
4,0,,0,,0,,0,,0,,0,,9,,131,,,*,,*,,*,,*,,*,,*,,*,189,,0,,0,,0,,0,,0,,0,,0,,0,,329,*
5,0,,0,,0,,0,,0,,0,,7,,116,,292,,437,,368,,293,,369,,302,,248,,211,,0,,0,,0,,0,,0,,0,,0,,0,,2643,
6,0,,0,,0,,0,,0,,0,,7,,114,,278,,285,,522,,556,,536,,451,,314,,90,,0,,0,,0,,0,,0,,0,,0,,0,,3153,
7,0,,0,,0,,0,,0,,0,,2,,47,,107,,310,,423,,454,,341,,320,,174,,77,,0,,0,,0,,0,,0,,0,,0,,0,,2255,
8,0,,0,,0,,0,,0,,0,,6,,97,,180,,207,,255,,331,,402,,286,,324,,146,,0,,0,,0,,0,,0,,0,,0,,0,,2234,
9,0,,0,,0,,0,,0,,0,,5,,107,,268,,414,,518,,562,,545,,469,,341,,175,,0,,0,,0,,0,,0,,0,,0,,0,,3404,
10,0,,0,,0,,0,,0,,0,,4,,106,,266,,411,,512,,560,,536,,452,,344,,173,,0,,0,,0,,0,,0,,0,,0,,0,,3364,
11,0,,0,,0,,0,,0,,0,,3,,101,,262,,404,,500,,546,,529,,454,,324,,160,,0,,0,,0,,0,,0,,0,,0,,0,,3283,
12,0,,0,,0,,0,,0,,0,,2,,99,,257,,402,,503,,549,,531,,454,,339,,156,,0,,0,,0,,0,,0,,0,,0,,0,,3292,
13,0,,0,,0,,0,,0,,0,,2,,27,,115,,195,,391,,532,,487,,457,,339,,169,,0,,0,,0,,0,,0,,0,,0,,0,,2714,
14,0,,0,,0,,0,,0,,0,,2,,45,,133,,217,,349,,487,,511,,294,,316,,142,,0,,0,,0,,0,,0,,0,,0,,0,,2496,
15,0,,0,,0,,0,,0,,0,,2,,115,,217,,389,,513,,537,,492,,422,,123,,61,,0,,0,,0,,0,,0,,0,,0,,0,,2871,
16,0,,0,,0,,0,,0,,0,,1,,61,,177,,306,,473,,505,,494,,422,,309,,169,,0,,0,,0,,0,,0,,0,,0,,0,,2917,
17,0,,0,,0,,0,,0,,0,,0,,85,,195,,306,,343,,417,,436,,359,,248,,116,,0,,0,,0,,0,,0,,0,,0,,0,,2505,
18,0,,0,,0,,0,,0,,0,,2,,32,,205,,397,,490,,537,,519,,454,,331,,156,,0,,0,,0,,0,,0,,0,,0,,0,,3123,
19,0,,0,,0,,0,,0,,0,,1,,57,,208,,372,,504,,463,,447,,341,,180,,143,,0,,0,,0,,0,,0,,0,,0,,0,,2716,
20,0,,0,,0,,0,,0,,0,,0,,52,,168,,394,,424,,532,,592,,449,,155,,106,,0,,0,,0,,0,,0,,0,,0,,0,,2872,
21,0,,0,,0,,0,,0,,0,,0,,88,,169,,280,,223,,314,,502,,303,,263,,124,,0,,0,,0,,0,,0,,0,,0,,0,,2266,
22,0,,0,,0,,0,,0,,0,,0,,60,,224,,352,,252,,440,,414,,355,,,?,178,,0,,0,,0,,0,,0,,0,,0,,0,,2275,
23,0,,0,,0,,0,,0,,0,,0,,63,,215,,356,,327,,382,,420,,386,,219,,79,,0,,0,,0,,0,,0,,0,,0,,0,,2447,
24,0,,0,,0,,0,,0,,0,,0,,38,,193,,334,,466,,432,,481,,413,,297,,109,,0,,0,,0,,0,,0,,0,,0,,0,,2763,
25,0,,0,,0,,0,,0,,0,,0,,23,,115,,325,,452,,494,,408,,331,,154,,52,,0,,0,,0,,0,,0,,0,,0,,0,,2354,
26,0,,0,,0,,0,,0,,0,,0,,63,,,?,404,,511,,529,,506,,428,,294,,134,,0,,0,,0,,0,,0,,0,,0,,0,,2869,
27,0,,0,,0,,0,,0,,0,,0,,50,,168,,289,,413,,443,,470,,386,,239,,87,,0,,0,,0,,0,,0,,0,,0,,0,,2545,
28,0,,0,,0,,0,,0,,0,,0,,53,,208,,351,,425,,486,,473,,387,,190,,128,,0,,0,,0,,0,,0,,0,,0,,0,,2701,
29,0,,0,,0,,0,,0,,0,,0,,26,,107,,239,,220,,210,,373,,168,,78,,49,,0,,0,,0,,0,,0,,0,,0,,0,,1470,
30,0,,0,,0,,0,,0,,0,,0,,58,,204,,346,,451,,462,,342,,336,,199,,72,,0,,0,,0,,0,,0,,0,,0,,0,,2470,


Avg,0,,0,,0,,0,,0,,0,,3,,75,,209,,345,,429,,476,,476,,389,,266,,136,,0,,0,,0,,0,,0,,0,,0,,0,,2783,
S D,0,,0,,0,,0,,0,,0,,3,,34,,62,,72,,99,,95,,68,,76,,82,,50,,0,,0,,0,,0,,0,,0,,0,,0,,496,
S/A,0,,0,,0,,0,,0,,0,,1.21,,0.46,,0.30,,0.21,,0.23,,0.20,,0.14,,0.19,,0.31,,0.37,,0,,0,,0,,0,,0,,0,,0,,0,,0.18,
Min,0,,0,,0,,0,,0,,0,,0,,23,,107,,195,,220,,210,,341,,168,,78,,49,,0,,0,,0,,0,,0,,0,,0,,0,,1470,
Max,0,,0,,0,,0,,0,,0,,11,,137,,312,,459,,558,,595,,592,,509,,379,,232,,0,,0,,0,,0,,0,,0,,0,,0,,3687,
n,0,,0,,0,,0,,0,,0,,26,,30,,28,,29,,29,,29,,29,,29,,28,,30,,0,,0,,0,,0,,0,,0,,0,,0,,284,


Daylight Data Recovery: 2.4% missing; 0.7% exceed the 15% QC threshold -- 296 total daylight measurements.

Instrument: Eppley PSP     Serial Number: 18039F3            Calibration Factor:  8.43 µV/W/m²     Calibration Date: Unknown
```

Appendix 1: Solar Radiation Raw text data.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	datetime	tempmax	tempmin	temp	feelsliken	feelsliken	feelslike	dew	humidity	precip	precippro	precipcov	preciptyp	snow	snowdept
2	1/1/1980	8.3	-0.7	3.1	6.6	-3.9	0.9	-3.5	64	0	0	0			
3	1/2/1980	5.5	-0.1	2.7	1.2	-4.6	-2	-3.7	63.4	0	0	0			
4	1/3/1980	7.1	1.1	3.7	4.5	-3.3	0.6	-3.4	60.5	0	0	0			
5	1/4/1980	2.8	-1	0.4	2.8	-6.7	-4.1	-2.3	82.8	4.398	100	54.17	rain,snow		2.1
6	1/5/1980	0	-1.1	-0.8	-4.4	-7.6	-6.4	-2.7	86.9	10.012	100	79.17	rain,snow		11.7
7	1/6/1980	1.6	-5.5	-1.9	0.2	-11.4	-6	-8.5	61.7	0	0	0			13.6
8	1/7/1980	5.5	-1.7	2.2	3.2	-7.6	-2.7	-2.2	72.7	0.553	100	12.5	rain		10
9	1/8/1980	6	1.6	3.6	4.1	-2.7	0.9	-5	54.2	0	0	0			6.1
10	1/9/1980	3.7	-2.8	2	3.6	-3.4	0.3	-2.1	74.7	0	0	0			5
11	1/10/1980	3.3	-2.9	0.4	-0.5	-6.9	-3.1	-3.9	74.1	0	0	0			2.2
12	1/11/1980	11.2	2.1	5.9	11.2	-1.3	4.7	4.3	89.7	11.569	100	41.67	rain		0.7
13	1/12/1980	11.1	0	4.2	11.1	-4.6	0.4	-7.3	44.5	0.327	100	4.17	rain		0
14	1/13/1980	2.1	-1.6	0.5	-0.5	-5.4	-2.9	-7.7	54.9	0	0	0			
15	1/14/1980	7.6	2.1	4.8	4.8	-1.9	1.4	3	87.8	0.706	100	16.67	rain		0
16	1/15/1980	11.1	3.9	7.7	11.1	0.8	4.9	4.6	81.2	0.054	100	8.33	rain		0
17	1/16/1980	13.7	2.8	7.4	13.7	-0.4	5.8	-2.9	53.1	0	0	0			
18	1/17/1980	7.1	2.5	5.4	5.4	-0.5	3.4	1.7	77.6	0.005	100	4.17	rain		0
19	1/18/1980	7.2	5.4	6.5	7.2	3.9	5.1	5.5	93.8	30.518	100	70.83	rain		0
20	1/19/1980	8.4	2.8	5.5	4.9	-1.2	2.1	-0.4	66.7	0.81	100	8.33	rain		0
21	1/20/1980	8.8	1.5	5.1	6	-1.9	2	-3.2	56.3	0	0	0			
22	1/21/1980	8.8	-1.6	3.5	6	-3.5	0.7	-6.6	49.8	0	0	0			
23	1/22/1980	8.3	1.8	5.6	8.3	-0.6	3.8	1.9	78.2	8.561	100	37.5	rain		0
24	1/23/1980	8.2	-2.2	4.4	6.5	-9.1	0.3	-2.9	60.8	0.331	100	8.33	rain		0
25	1/24/1980	0.6	-6.7	-2.5	0.4	-13.8	-7.3	-10.9	54.5	0.018	100	8.33	rain,snow		0

Appendix 2: The raw Temperature and Precipitation Data

	A	B	C	D	E	F	G	H
55	# at the M	approximately 21 miles north of the Mauna Loa Observatory.						
56	#							
57	year	month	decimal da	average	deseasona	ndays	sdev	unc
58	1958	3	1958.203	315.7	314.43	-1	-9.99	-0.99
59	1958	4	1958.288	317.45	315.16	-1	-9.99	-0.99
60	1958	5	1958.37	317.51	314.71	-1	-9.99	-0.99
61	1958	6	1958.455	317.24	315.14	-1	-9.99	-0.99
62	1958	7	1958.537	315.86	315.18	-1	-9.99	-0.99
63	1958	8	1958.622	314.93	316.18	-1	-9.99	-0.99
64	1958	9	1958.707	313.2	316.08	-1	-9.99	-0.99
65	1958	10	1958.789	312.43	315.41	-1	-9.99	-0.99
66	1958	11	1958.874	313.33	315.2	-1	-9.99	-0.99
67	1958	12	1958.956	314.67	315.43	-1	-9.99	-0.99
68	1959	1	1959.041	315.58	315.55	-1	-9.99	-0.99
69	1959	2	1959.126	316.48	315.86	-1	-9.99	-0.99
70	1959	3	1959.203	316.65	315.38	-1	-9.99	-0.99
71	1959	4	1959.288	317.72	315.41	-1	-9.99	-0.99
72	1959	5	1959.37	318.29	315.49	-1	-9.99	-0.99
73	1959	6	1959.455	318.15	316.03	-1	-9.99	-0.99
74	1959	7	1959.537	316.54	315.86	-1	-9.99	-0.99
75	1959	8	1959.622	314.8	316.06	-1	-9.99	-0.99
76	1959	9	1959.707	313.84	316.73	-1	-9.99	-0.99
77	1959	10	1959.789	313.33	316.33	-1	-9.99	-0.99

Appendix 3: The raw text data of Atmospheric CO_2 concentration

FAOSTAT_USA CROP PRODUCTION_6-1-2023 - Excel

A1 | Domain Code

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Domain C	Domain	Area Code	Area	Element C	Element	Item Code	Item	Year Code	Year	Unit	Value	Flag	Flag Description	
2	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1990	1990	ha	2.7E+07	A	Official figure	
3	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1991	1991	ha	2.8E+07	A	Official figure	
4	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1992	1992	ha	2.9E+07	A	Official figure	
5	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1993	1993	ha	2.5E+07	A	Official figure	
6	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1994	1994	ha	2.9E+07	A	Official figure	
7	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1995	1995	ha	2.6E+07	A	Official figure	
8	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1996	1996	ha	2.9E+07	A	Official figure	
9	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1997	1997	ha	2.9E+07	A	Official figure	
10	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1998	1998	ha	2.9E+07	A	Official figure	
11	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	1999	1999	ha	2.9E+07	A	Official figure	
12	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2000	2000	ha	2.9E+07	A	Official figure	
13	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2001	2001	ha	2.8E+07	A	Official figure	
14	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2002	2002	ha	2.8E+07	A	Official figure	
15	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2003	2003	ha	2.9E+07	A	Official figure	
16	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2004	2004	ha	3E+07	A	Official figure	
17	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2005	2005	ha	3E+07	A	Official figure	
18	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2006	2006	ha	2.9E+07	A	Official figure	
19	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2007	2007	ha	3.5E+07	A	Official figure	
20	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2008	2008	ha	3.2E+07	A	Official figure	
21	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2009	2009	ha	3.2E+07	A	Official figure	
22	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2010	2010	ha	3.3E+07	A	Official figure	
23	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2011	2011	ha	3.4E+07	A	Official figure	
24	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2012	2012	ha	3.5E+07	A	Official figure	
25	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2013	2013	ha	3.5E+07	A	Official figure	
26	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2014	2014	ha	3.4E+07	A	Official figure	
27	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2015	2015	ha	3.3E+07	A	Official figure	
28	QCL	Crops and	840	United Sta	5312	Area harv	112	Maize (co	2016	2016	ha	3.5E+07	A	Official figure	

Appendix 4: The raw Crop yield data.

E26 | 33624

	A	B	C	D	E	F
1	Datettime	m_temp	m_precip	CO2	solar_rad	
2	Jan-80	84.3	167.865	337.9		
3	Feb-80	54.2	34.67	338.34		
4	Mar-80	233	127.139	340.07		
5	Apr-80	455.9	81.479	340.93		
6	May-80	635.3	74.154	341.45		
7	Jun-80	702.3	45.359	341.36		
8	Jul-80	852.7	106.727	339.45		
9	Aug-80	854.8	32.174	337.67		
10	Sep-80	734	52.814	336.25		
11	Oct-80	462.5	72.349	336.14		
12	Nov-80	261.8	63.803	337.3		
13	Dec-80	121.6	15.88	338.29		
14	Jan-81	10.8	12.835	339.29		
15	Feb-81	173.3	76.416	340.55		
16	Mar-81	252.3	37.451	341.63		
17	Apr-81	492	67.436	342.6		
18	May-81	584.4	91.071	343.04		
19	Jun-81	769.2	69.88	342.54		
20	Jul-81	826.2	138.861	340.82		
21	Aug-81	767.5	76.185	338.48		
22	Sep-81	645.6	54.346	336.95		
23	Oct-81	447.3	94.076	337.05		
24	Nov-81	315.1	7.384	338.58		
25	Dec-81	106.7	73.434	339.91		
26	Jan-82	-62.7	67.761	340.93	33624	
27	Feb-82	98.3	118.865	341.76	79230	
28	Mar-82	223.6	68.387	342.77	138324	
29	Apr-82	371.4	86.612	343.96	167723	
30	May-82	634.4	140.232	344.77	139497	
31	Jun-82	671.7	141.241	343.88	181128	
32	Jul-82	815.8	80.512	342.42	147562	
33	Aug-82	737.8	70.939	340.24	55798	
34	Sep-82	636	46.449	338.38	103493	
35	Oct-82	472.3	46.59	338.41	55799	

Appendix 5: The combination of the monthly climate change variables.

	A	B	C	D	E	F	G	H	I	J
1	year	country	maize_yiel	pot_yield	wheat_yie	rice_yield	solar_rad	co2	temp	precip
2	1980	United Sta	57118	295245	22513	49461	894031	2713.57	4030.5	592.2
3	1981	United Sta	68380	308199	23229	54015	894031	2723.25	4784.5	629.3
4	1982	United Sta	71082	312544	23859	52791	851000	2734.78	4563	681
5	1983	United Sta	50893	299754	26509	51532	957440	2748.52	4722.5	941
6	1984	United Sta	66981	311082	26073	55523	800268	2761.06	4546.8	982
7	1985	United Sta	74074	333689	25197	60704	967417	2773.18	4747.4	654.4
8	1986	United Sta	74930	330540	23167	63340	1014520	2783.23	4689.8	470.7
9	1987	United Sta	75225	337394	25330	62271	1277547	2797.01	4622.6	594.8
10	1988	United Sta	53109	317265	22913	61801	1063344	2815.86	4525.9	523.1
11	1989	United Sta	72978	324007	22024	64466	1260174	2827.57	4573.1	1060.1
12	1990	United Sta	74380	328809	26567	61975	1279369	2836.93	4663.6	770
13	1991	United Sta	68171	340579	23038	64231	1147922	2848.46	4877.4	525.4
14	1992	United Sta	82526	362533	26433	64294	1106325	2854.98	4273	669.4
15	1993	United Sta	63210	365091	25699	61763	1252046	2859.6	4590.5	904.4
16	1994	United Sta	86996	379868	25270	66851	1273503	2873.38	4634.5	835.1

Appendix 6: The combination of pre-processed Data.

```
In [19]:
import matplotlib.pyplot as plt
import matplotlib.dates
import datetime
import pandas as pd
import pylab as pl
import seaborn as sns
import numpy as np
from sklearn.model_selection import train_test_split
%matplotlib inline
from sklearn.linear_model import LinearRegression
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from scipy import stats
import plotly.express as px
import plotly.graph_objs as go
from sklearn.metrics import mean_squared_error, r2_score
```

Appendix 7: The needed Python Libraries.

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn import svm
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVR
```

In [ ]:
```
df = pd.read_csv(r"C:\Users\TOLS\OneDrive - Cardiff Metropolitan University\Dissertation\DISSERTATION DATA\DISSERTATION.csv"
df.head(10)
```

In [ ]:
```
# Filtering the independent variables (X) and the dependent variable (y)
X = df[["solar_rad", "co2", "temp", "precip"]].values.astype(float)
y = df[["maize_yield"]].values.astype(float)
X
```

In [ ]:
```
from sklearn.preprocessing import StandardScaler
```

In [ ]:
```
#Scaling or Standardizing the two variables for better analysis by calling the standardscaler() function from sklearn
scaledX = StandardScaler()
scaledy = StandardScaler()
```

In [ ]:
```
#Transforming the variables into the standadized form.
X = scaledX.fit_transform(X)
y = scaledy.fit_transform(y)
X
```

In [ ]:
```
y = y.ravel()
y
```

In [ ]:
```
from sklearn.model_selection import train_test_split
```

In [ ]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, train_size=0.75, random_state = 10)
```

In [ ]:
```
gamma_values = [0.01]  # You can adjust these values according to your needs
```

In [ ]:
```
from sklearn.svm import SVR
```

In [ ]:
```
for gamma in gamma_values:
    svr_model = SVR(kernel='rbf', gamma=gamma)  # Using 'rbf' kernel with specified gamma
    svr_model.fit(X_train, y_train)
```

In [ ]:
```
#for gamma in gamma_values:
#svr_model = SVR(kernel='rbf' gamma=gamma)
#svr_model.fit(X_train, y_train)
```

In [ ]:
```
y_predict = svr_model.predict(X_test)
y_predict
```

In [ ]:
```
y_predict= y_predict.reshape(-1,1)
y_predict
```

In [ ]:
```
y_pred=scaledy.inverse_transform(y_predict)
y_pred
```

In [ ]:
```
from sklearn.metrics import r2_score
r2 = r2_score(y_test, y_predict)
r2
```

In [ ]:
```
intercept = svr_model.intercept_
intercept
```

In [ ]:
```
from sklearn.metrics import mean_absolute_error as mae
```

Appendix 8. Support Vector Regression Python codes.

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [ ]:
```
import numpy as np # for array operations
import pandas as pd # for working with DataFrames
import matplotlib.pyplot as plt # for data visualization
%matplotlib inline

from sklearn.model_selection import train_test_split # for splitting the data
from sklearn.metrics import mean_squared_error # for calculating the cost function
from sklearn.ensemble import RandomForestRegressor # Import the model we are using

```

In [ ]:
```
df=pd.read_csv(r"C:\Users\TOLS\OneDrive - Cardiff Metropolitan University\Dissertation\DISSERTATION DATA\DISSERTATION.csv")
df.head(5)
```

In [ ]:
```
df.columns
```

In [ ]:
```
df=df[['maize_yield', 'solar_rad', 'co2', 'temp', 'precip']]
df.head()
```

In [ ]:
```
#Separating the features and the target variable
X = df.drop('maize_yield', axis = 1).values.astype(float) # Features
y = df[["maize_yield"]].values.astype(float) # Target
X
```

In [ ]:
```
### Sandardization of data ###
from sklearn.preprocessing import StandardScaler
scaledX = StandardScaler()
scaledy = StandardScaler()
```

In [ ]:
```
# Storing the fit object for later reference
X = scaledX.fit_transform(X)
y = scaledX.fit_transform(y)
```

In [ ]:
```
y = y.ravel()
```

In [ ]:
```
# Splitting the dataset into training and testing set (80/20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 5)
```

In [ ]:
```
# Import the model we are using
from sklearn.ensemble import RandomForestRegressor
# Instantiate model with 100 decision trees
rf = RandomForestRegressor(n_estimators = 100, random_state = 5)
# Train the model on training data
rf.fit(X_train, y_train)
```

In [ ]:
```
# Predicting the target values of the test set
y_pred = rf.predict(X_test)
```

In [ ]:
```
from sklearn.metrics import mean_absolute_error as mae
error = mae(y_test, y_pred)
print("Mean absolute error : " + str(error))
```

In [ ]:
```
# RMSE (Root Mean Square Error)
rmse = float(format(np.sqrt(mean_squared_error(y_test, y_pred)), '.3f'))
print("\nRMSE: ", rmse)
```

In [ ]:
```
from sklearn.metrics import r2_score
r2 = r2_score(y_test, y_pred)
print( "R-Squared :", r2)
```

In [ ]:
```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
print( "Mean Square Error :", mse)
```

Appendix 9: Random Forest Regression Python codes.

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [ ]:
```
import numpy as np # for array operations
import pandas as pd # for working with DataFrames
import requests, io # for HTTP requests and I/O commands
import matplotlib.pyplot as plt # for data visualization
%matplotlib inline

from sklearn.model_selection import train_test_split # for splitting the data
from sklearn.metrics import mean_squared_error # for calculating the cost function
from sklearn.tree import DecisionTreeRegressor # for building the model
```

In [ ]:
```
df=pd.read_csv(r"C:\Users\TOLS\OneDrive - Cardiff Metropolitan University\Dissertation\DISSERTATION DATA\DISSERTATION.csv")
df.head(5)
```

In [ ]:
```
df.columns
```

In [ ]:
```
df=df[['maize_yield', 'solar_rad', 'co2', 'temp', 'precip']]
df.head()
```

In [ ]:
```
#Separating the features and the target variable
X = df.drop('maize_yield', axis = 1).values.astype(float) # Features
y = df[["maize_yield"]].values.astype(float) # Target
X
```

In [ ]:
```
### Sandardization of data ###
from sklearn.preprocessing import StandardScaler
scaledX = StandardScaler()
scaledy = StandardScaler()
```

In [ ]:
```
# Storing the fit object for later reference
X = scaledX.fit_transform(X)
y = scaledy.fit_transform(y)
```

In [ ]:
```
# Splitting the dataset into training and testing set (80/20)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 5)
```

In [ ]:
```
# Initializing the Decision Tree Regression model
model = DecisionTreeRegressor(max_depth=3, min_samples_split=5, random_state = 0)

# Fitting the Decision Tree Regression model to the data
model.fit(X_train, y_train)
```

In [ ]:
```
# Predicting the target values of the test set
y_pred = model.predict(X_test)

# RMSE (Root Mean Square Error)
rmse = float(format(np.sqrt(mean_squared_error(y_test, y_pred)), '.3f'))
print("\nRMSE: ", rmse)
```

In [ ]:
```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
print( "Mean Square Error :", mse)
```

In [ ]:
```
from sklearn.metrics import mean_absolute_error as mae
error = mae(y_test, y_pred)
print("Mean absolute error : " + str(error))
```

In [ ]:
```
from sklearn.metrics import r2_score
r2 = r2_score(y_test, y_pred)
r2
```

Appendix 10 Decision Tree Regression Python Codes.

File Edit View Insert Cell Kernel Widgets Help Not Trusted | Python 3 (ipykernel)

```
from sklearn.linear_model import LinearRegression
from sklearn import preprocessing
from sklearn.model_selection import train_test_split

from scipy import stats
from pandas import to_datetime
import datetime as dt
import plotly.express as px
import plotly.graph_objs as go
```

In [ ]:

```
df=pd.read_csv(r"C:\Users\TOLS\OneDrive - Cardiff Metropolitan University\Dissertation\DISSERTATION DATA\monthly_climateDat
df
```

In [ ]:

```
df.isna
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=df['Datetime'].astype(dtype=str),
                                y=df['solar_rad'],
                                marker_color='red', text="solar_rad"))
fig.update_layout({"title": 'United States Monthly Solar Radiation Series from 1980 to 2021',
                   "xaxis": {"title":"Month(1980-2021)"},
                   "yaxis": {"title":"Solar Radiation (WH/m2)"},
                   "showlegend": False})
fig.show()
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=df['Datetime'].astype(dtype=str),
                                y=df['m_temp'],
                                marker_color='green', text="temperature"))
fig.update_layout({"title": 'United States Monthly Temperature Time Series from 1980 to 2021',
                   "xaxis": {"title":"Month(1980-2021)"},
                   "yaxis": {"title":"Temperature (°C)"},
                   "showlegend": False})
fig.show()
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=df['Datetime'].astype(dtype=str),
                                y=df['m_precip'],
                                marker_color='blue', text="Precipitation"))
fig.update_layout({"title": 'United States Monthly Precipitation Time Series from 1980 to 2021',
                   "xaxis": {"title":"Month(1980-2021)"},
                   "yaxis": {"title":"Precipitation (mm)"},
                   "showlegend": False})
fig.show()
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=df['Datetime'].astype(dtype=str),
                                y=df['CO2'],
                                marker_color='brown', text="Atmospheric CO2"))
fig.update_layout({"title": 'United States Monthly Atmoshpheric CO2 Conc. Time Series from 1980 to 2021',
                   "xaxis": {"title":"Month(1980-2021)"},
                   "yaxis": {"title":"Atmospheric CO2 Conc.(ppmv)"},
                   "showlegend": False})
fig.show()
```

In [ ]:

```
dfy=pd.read_csv(r"C:\Users\TOLS\OneDrive - Cardiff Metropolitan University\Dissertation\DISSERTATION DATA\y_climate.csv")
dfy
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=dfy['Year'].astype(dtype=str),
                                y=dfy['y_solar_rad'],
                                marker_color='red', text="solar_rad"))
fig.update_layout({"title": 'United States Yearly Solar Radiation Series from 1980 to 2021',
                   "xaxis": {"title":"Year(1980-2021)"},
                   "yaxis": {"title":"Solar Radiation (WH/m2)"},
                   "showlegend": False})
fig.show()
```

In [ ]:

```
fig = go.Figure(data=go.Scatter(x=dfy['Year'].astype(dtype=str),
                                y=dfy['y_temp'],
                                marker_color='green', text="temperature"))
fig.update_layout({"title": 'United States Yearly Temperature Time Series from 1980 to 2021',
                   "xaxis": {"title":"Year(1980-2021)"},
                   "yaxis": {"title":"Temperature (°C)"},
                   "showlegend": False})
fig.show()
```

Appendix 11: Objective 1 Python codes.

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

In [ ]:
```python
fig = go.Figure(data=go.Scatter(x=dfy['Year'].astype(dtype=str),
                                y=dfy['y_precip'],
                                marker_color='blue', text="Precipitation"))
fig.update_layout({"title": 'United States Yearly Precipitation Time Series from 1980 to 2021',
                   "xaxis": {"title":"Year(1980-2021)"},
                   "yaxis": {"title":"Precipitation (mm)"},
                   "showlegend": False})
fig.show()
```

In [ ]:
```python
fig = go.Figure(data=go.Scatter(x=dfy['Year'].astype(dtype=str),
                                y=dfy['y_CO2'],
                                marker_color='brown', text="Atmospheric CO2"))
fig.update_layout({"title": 'United States Yearly Atmoshpheric CO2 Conc. Time Series from 1980 to 2021',
                   "xaxis": {"title":"Year(1980-2021)"},
                   "yaxis": {"title":"Atmospheric CO2 Conc.(ppmv)"},
                   "showlegend": False})
fig.show()
```

In [ ]:
```python
dfy
```

In [ ]:
```python
dfy.columns
```

In [ ]:
```python
dfy=dfy.drop(['Year'], axis=1)
dfy
```

In [ ]:
```python
dfy=dfy.T
dfy
```

In [ ]:
```python
p_dfy=dfy.pct_change(axis='columns')
p_dfy
```

In [ ]:
```python
p_dfy=p_dfy.T
p_dfy
```

In [ ]:
```python
p_dfy['year']= [1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997,
p_dfy
```

In [ ]:
```python
# Example data for two line plots
#x_values = [1, 2, 3, 4, 5]
#y_values1 = [2, 4, 6, 8, 10]
#y_values2 = [1, 3, 5, 7, 9]

x= p_dfy['year']
y1= p_dfy['y_solar_rad']
y2= p_dfy['y_temp']
y3= p_dfy['y_precip']
y4= p_dfy['y_CO2']


# Create the first line plot
plt.plot(x, y1, label='Solar Rad', color='red', linestyle='-')

# Create the second line plot
plt.plot(x, y2, label='Temperature', color='green', linestyle='-')

# Create the second line plot
plt.plot(x, y3, label='Precipitation', color='blue', linestyle='-')
# Create the second line plot
plt.plot(x, y4, label='Atm CO2', color='brown', linestyle='-')

# Add labels and title
plt.xlabel('Year')
plt.ylabel('Percentage Change')
plt.title('Multiple Percentage Change in Climatic Variables Line Plots')
plt.grid(True)

plt.legend()

# Show the plot
plt.show()
```

Appendix 12: Objective 1 Python codes continuation.