A Model for Forecasting Land Productivity Decline

By

Seth Nyawacha

171231

Submitted in Partial Fulfilment of the Requirements for the Master of Science in Computing Information System at Strathmore University

School of Computing & Engineering Sciences

Strathmore University

Nairobi, Kenya

March 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Student's Name: Seth Odhiambo Nyawacha

Sign: ________________________ Date: ______19/03/2025__________________

Approval

The dissertation of Seth Odhiambo Nyawacha was reviewed and approved for examination by the following:

Sign: B.K.S Date: 27/03/2025

Prof. Bernard Shibwabo
Associate Professor
School of Computing & Engineering Sciences,
Strathmore University

Abstract

Sub-Saharan Africa (SSA) faces significant challenges in agricultural sustainability due to its vulnerability to climate change, which directly threatens crop productivity and economic stability. While existing crop simulation models have demonstrated potential in optimizing water and soil resources, their application within SSA's unique agro-climatic conditions has remained limited. This study sought to bridge this gap by developing a robust farm productivity decline simulation model tailored specifically to SSA's agricultural landscape. By leveraging predictive analytics, the model forecasted productivity trends and provided actionable insights to mitigate yield gaps, reduce food insecurity, and enhance land-use strategies. Through scenario modelling based on climate variability and resource availability, this study provides practical information on yield decline, fostering resilience against climate-induced shocks. The trained model demonstrated strong predictive performance, achieving a Train R^2 of 0.86 and Test R^2 of 0.84**,** indicating a high explanatory power in capturing the relationships between agro-climatic variables and farm productivity decline. The Cross-validation R^2 of 0.83 and an Out-of-Bag (OOB) Score of 0.83 further validated the model's robustness, ensuring it generalized well across different data partitions without overfitting. Additionally, Pearson correlation coefficients of 0.93 for training and 0.92 for testing confirmed a strong linear relationship between the observed and predicted productivity values, reinforcing the model's reliability in capturing real-world agricultural trends. Collectively, these evaluation metrics highlight the model's effectiveness in forecasting productivity decline while maintaining high predictive accuracy and generalization across unseen data. The forecasting system provided 20 years of future projections, revealing that Kenya's food basket regions are expected to experience a significant productivity decline starting around 2026, with the most substantial reductions projected to occur by 2040. The model's Mean Squared Error (MSE) ranged between 0.001 and 0.002, further confirming its ability to generalize well on unseen data. To facilitate accessibility and usability, the model was deployed through a Next.js-based web interface, with a Python-powered backend, leveraging containerized architecture via Docker for enhanced scalability and efficiency. This study contributes to the integration of advanced agricultural modeling into real-world farming practices**,** supporting scalable climate adaptation strategies aimed at safeguarding food security and economic resilience across SSA.

Table of Contents

List of Figures

Abbreviations and Acronyms

ANN	Artificial Neural Network
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station Data
CNN	Convolution Neural Network
FAO	Food and Agriculture Organization
GGCM	Global Gridded Crop Models
ML	Machine Learning
NASA	National Aeronautics and Space Administration
RS	Remote Sensing
SHAP	Shapley Additive Explanations
SSA	Sub-Saharan Africa
RF	Random Forest

Definition of Terms

APSIM-Maize: A crop simulation model within the Agricultural Production Systems Simulator (APSIM) framework, designed to simulate the growth, development, and yield of maize. It incorporates factors like soil, climate, crop management om a daily basis and interactions between them to assess productivity and environmental impact (Archontoulis et al., 2014).

AquaCrop: According to Heng et al, (2009), the FAO AquaCrop model predicts crop productivity, water requirement, and water use efficiency (WUE) under water-limiting conditions.

CropSyst: Stockle et al, 1994 describes CropSyst as a multi-year, multi-crop, daily time step cropping systems simulation model developed to serve as an analytical tool to study the effect of climate, soils, and management on cropping systems productivity and the environment. CropSyst simulates the soil water and nitrogen budgets, crop growth and development, crop yield, residue production and decomposition, soil erosion by water, and salinity.

CropWat: A software tool developed by the FAO to calculate crop water requirements and irrigation needs using weather, crop, and soil data. It supports irrigation planning and water management (Archontoulis et al., 2014).

DSSAT CREES: A module or variation in the (Decision Support System for Agrotechnology Transfer) framework where typically is used in simulating other crop production other than maize using climate, soil and farm management practices (Liu et al., 2022).

DSSAT-CRES-Maize: A component of the Decision Support System for Agrotechnology Transfer (DSSAT) that simulates maize growth, development, and yield based on weather, soil, and management practices (Liu et al., 2022).

EPIC-Maize: A variation of the EPIC (Environmental Policy Integrated Climate) model that focuses on maize production in terms of its adaptability to soil and management practices around. (Mearns et al., 1999).

InfoCrop: Crop simulation model developed in India that predicts the effects of weather, pests, soil, and management practices on the growth and yield of crops. It also assesses the impact of climate change (Aggarwal et al., 2005).

Chapter 1: Introduction

1.1 Background to the Study

Farmland productivity is defined as the capability of a piece of land in supporting agricultural production based on biophysical conditions such as climate, elevation, soil conditions and farm practices (Wang et al., 2020). Farmland acts as the center for agricultural production. On the other hand, agriculture plays a pivotal role in combatting food insecurities brought about by the gradual effects of climate change on agricultural ecosystems (Dasgupta et al., 2021). The devastating climate change effects has not only affected productivity, but also farmland quality, which currently has a multi-dimensional view with inclusion of crop suitability, production potential, soil environmental quality and sustainability (Wang et al., 2021).

Fuglie (2011) describes agricultural productivity in Africa as a dwindling endeavor with farmers receiving less and less yield with variation in time. Although agriculture picked up in mid 1990s in Africa, the inhabitants of the continent are yet to realize a full potential of their major economic activity. It is estimated that Sub-Saharan Africa's rural population is economically active in agriculture, with between 45 – 50 % of East Africa's population being actively engaged in agricultural practices. Agriculture continues to play a pivotal role in influencing the economy of Africa, and its dwindling fortunes, this means that a larger population of Sub-Saharan Africa, is not able to contribute regularly and constantly to economic growth (Edwards et al., 2009).

Error! Reference source not found. highlights some of the direct and indirect effects of climate change to the environment and the socio-economic aspects of farmers.

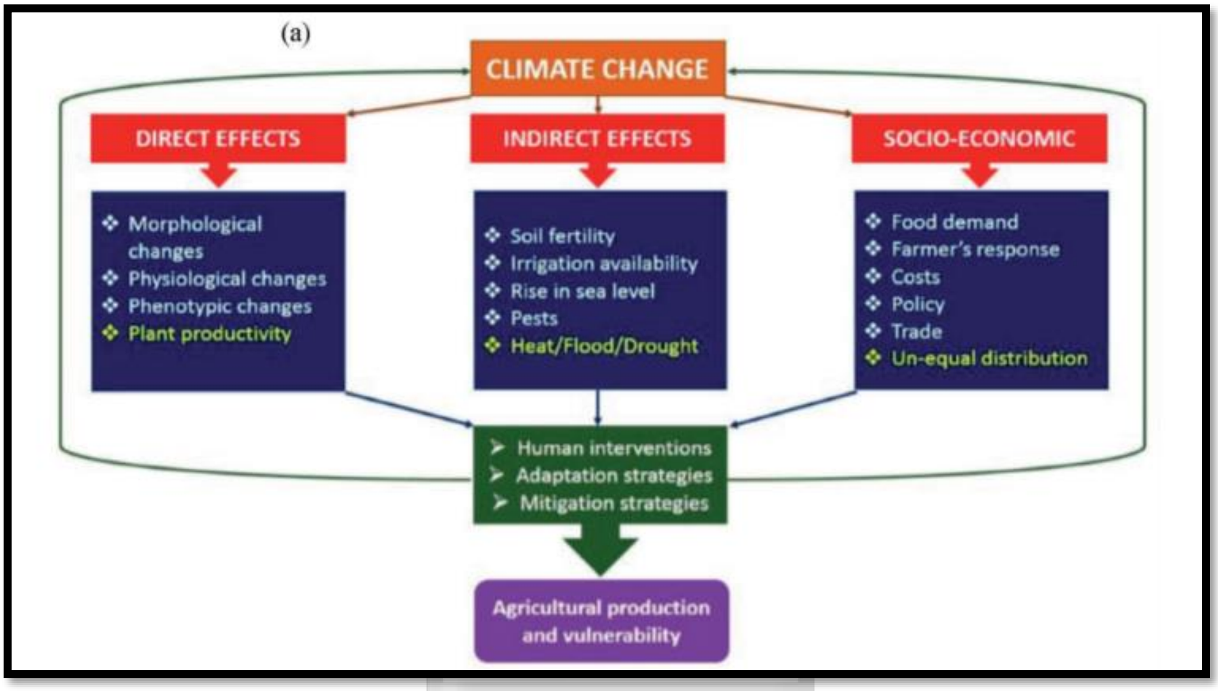

Figure 1.1: Effects of Climate Change on Agricultural Ecosystems

According to Lilian et al. (2020), climate change has been the main driver to farm productivity. Climate change impacts have manifested itself with extreme weather events such as drought, floods, high and very low temperatures. On the flip side, climate change has also driven high production of crops on especially warmer climates, improved crop suitability for some specific regions, longer growing seasons for high yielding crops, however, its negative effects have been felt across the world.

Daou et al. (2021) argues that there are direct, indirect and socio-economic factors driving dwindling crop production for farmers in Sub-Saharan Africa. Among the direct effects are the morphological, physiological, phenotypical and plant productivity, while the indirect causes include soil fertility, irrigation, sea level rise, pests and extreme weather events. The socio-economic aspects may include the varying food demands, costs of farm inputs and policy that drives agricultural sector and supply of Agro-inputs as illustrated in Figure 1.1.

According to Dasgupta et al. (2021), to monitor impact of drivers to crop production at scale, there is need to do analysis covering an expansive geographic area, to have a holistic view of the extent of such effects at a regional scale. Over the years, various technologies have arisen to try and model the drivers of farmland productivity at scale, and at field level. Coupled with machine learning models, researchers have attained considerable fair results that could assist them in understanding more the variability of farm productivity. Jones et al. (2022) highlights the importance that machine learning brought towards modelling the non-linear drivers of crop production at scale. Such a model has been cited to provide management insights into optimizing the allocation of agronomic resources.

Despite the successes of modern methods such as Machine Learning methods to assess the variability of crop production across time, sharp criticism has arisen due to their inability to provide comprehensive information on parameter importance as the major driver to influencing a certain direction of output. Jones et al. (2022) argues that the use of Shapley Additive explanations (SHAP), which integrates efficient computation using sample approximation while conserving the properties of influencing parameters, would solve the black box effect brought by machine learning models in modelling farm productivity.

Nurcahyo et al. (2023) hhighlights the needs of capturing the main drivers of productivity variability within a farm set up by using explainable methods and models. Although factors influencing crop production are multi-faceted, climate and soil variables play a critical role in the historical impact of crop production. Such critical factors have been picked out by novel machine learning methods, such as the use of interpretable machine learning models, as well as crop simulation models, based on varying crop husbandry coupled with climatic variations as input parameters.

In the past, researchers relied on the use of traditional methods of having boots on the ground to assess crop productivity at farm level. This data was then fed into a statistical method or a processed based model to identify patterns at a regional scale. With increasing need of having relatively accurate data at scale, machine learning and regression models have been applied, these new methods, although most are marred with black box effects, are having a relatively accurate prediction at scale (Leng & Hall, 2020).

On a global scale, Africa has since improved periodically, and the use of modern scientific methods, to explain different phenomena in the agricultural space has improved tremendously. However, this growth rate is more in the horizontal expansion, in terms of quantity, however, the quality of agriculture in terms of knowledge, and continuous application of novel methods to retain the high yields have been curtailed. The major driver to this effect has been the rise of poverty index across the African rural population, and the increasingly expensive technology for a small holder farmer to apply (Fugies et al., 2020). Adamopoulos and Restuccia, (2021) argues that to reduce the productivity disparity and close the gap on lack of financial access for small holder farmers, there needs to have additional aggregate productivity gains through spatial reallocation and changes in crop production systems. Their problem has not been in unfavorable geographic endowments, but rather in the efficient existing agricultural systems.

1.2 Problem Statement

Research has continuously shown that the Sub-Saharan Africa (SSA) agricultural eco-system is highly vulnerable to the impact of climate change with its crop productivity being highly susceptible to climate variations (Omotoso et al., 2023). Florent Baarsch et al. (2020) argues that climate change on agricultural lands continuously impart detrimental effects on the African economic development increasing intense effect of economic disparity.

According to Branca et al (2021), most undernourished population are found in the sub-Saharan Africa, with a projected population of 40 million being food insecure despite the (FAO et al., 2020) estimation that the population continue to increase progressively. For Africa to be food sufficient, the region needs to increase its crop production by up to 260% by 2050 (UN, 2019a).

With the increasing demand for food, minimal efforts have been made to diversify and disseminate information from existing crop simulation models, such as the AquaCrop model, which emphasize the optimization of biophysical resources like water, soil fertility, and nutrients to maximize yields (Mansour et al., 2020) that can assist farmers to cope with changing climate in minimizing productivity decline.

Although Silva and Giller (2020), highlights the importance of crop modelling to quantify the magnitude of crop yield gaps, the gaps between food demand and availability, and the land area needed to feed the population in the present and in the future, as well as evaluating adaptation capability for biodiversity, the outputs of these models have not been applied directly on small holder farms in Sub Saharan Africa (SSA), while accessing that information has been very complicated for farmers. Existing models have also largely been tested in other continents and not in Africa due to budget constraints on collecting ground data and very complex agricultural scenarios that normally comes into play.

This research therefore intends to develop a farm productivity decline simulation model that forecasts on the productivity capability of SSA lands, based on key parameters that affect crop production directly. Kephe et al. (2021) explains that there is very little research done on minimal uptake, development and testing of crop simulation models in SSA. This research development, therefore, provides productivity trends for farmers in SSA, with a forecasting capability for contextual relevance, while making it accessible for SSA farmers towards building resilience against climate and productivity shocks.

1.3 Aim of the Study

The aim of this research is to develop a model for forecasting land productivity decline

1.4 Research Objectives

i. To evaluate existing models and methodologies that have been used in previous studies to model land productivity decline.
ii. To establish input covariates for farm productivity decline model.
iii. To develop a model for predicting land productivity decline
iv. To evaluate the accuracy and reliability of the developed model in predicting land productivity decline.

1.5 Research Questions

i. What models and methodologies have been used in previous studies to assess land productivity decline?
ii. What are the key input covariates required for developing a farm productivity decline model?
iii. What is the most suitable model for predicting land productivity decline?
iv. How can the accuracy of the developed model be evaluated?

1.6 Justification

Jägermeyr et al. (2021) discusses the impacts of climate change on crop yield as a major societal concern. One of the proposals to tackle the challenge is the ability to identify and quantify the scope of associated uncertainties, which are deemed to be substantial as a result of climate change. Benami et al., (2021) assessed the intersection of remote sensing and crop modelling with economic growth. The inability of small holder farmers to assess crop production losses leads to more uncertainties, as a result losing their financial ability to build resilience in cases of climate shocks. Therefore, this research is geared towards enhancing the ability of small holder farmers to gauge the level of uncertainties of their land productivity in time with its forecasting capability.

1.7 Scope of Research

This study developed a prototype model to simulate land productivity decline using agro-climatic data and multi-temporal satellite imagery. The model integrated biophysical land properties and climate characteristics to predict productivity trends, providing a tool for assessing land productivity dynamics and enhancing resilience against potential productivity shocks.

1.8 Limitation of Research

The primary complexity in developing this model lies in the need for extensive ground truthing and validation to ensure its accuracy and applicability in real-world scenarios. Due to time constraints during its development, large-scale ground truthing was not fully conducted, which may impact the model's reliability across diverse agro-climatic conditions.

Chapter 2: Literature Review

2.1 Overview

Farm productivity research and study has attracted interest from various scientists and research groups across the world. Leng and Hall (2020) used a combination of process-based crop models and machine learning models coupled with a traditional multiple linear regression model to simulate maize yield production. Methodologies such as random forest, a nonparametric algorithm uses two randomization states and relies on an ensemble of decision trees to minimize overfitting of predictor variables. Notably, one advantage of machine learning models is that they do not assume a certain shape of response functions e.g. linear or polynomial.

Leng (2017b) evaluated productivity across time with two climate variables as input parameters in a machine learning models to evaluate farm productivity. The study also determined the advantages and disadvantages of each approach. It is noteworthy that while process-based crop models, even when fed with harmonized parameters, exhibit significant bias, regression and machine learning models can accurately replicate the observed pattern of yield averages with minimal bias and margins of errors.

From the overview, there are currently two methodologies that are outstanding in simulating land and crop productivity decline, notably, they include the following

i. Use of Machine learning and statistical methods for land productivity decline forecasting
ii. Use of crop simulation models/processed models for land productivity decline forecasting.

They are discussed in the following sections.

2.2 Use of Machine Learning Models and Statistical Methods for Land Productivity Forecasting

Wolanin et al., (2019) estimated gross primary productivity using the combination of satellite imageries originating from sentinel 2 constellation as input parameter in a machine learning model. The model estimated productivity without the need for any local information or ground truth information, with a considerably high accuracy. Some machine learning models considered for this study included neural network model with flux tower-based gross primary productivity data.

Global gridded crop models (GGCMs) were used for estimating agricultural crop yields and externalities at large scales, typically at coarse resolution. However, machine learning models have been used with GGCMs output data as covariates of high spatial resolution on models such as extreme gradient boosting and random forest while developing a meta model for prediction of crop model output with fine resolution (Folberth et al., 2019).

Artificial Neural Network (ANN) uses a network of artificial neutrons, a supervised learning. However, ANN is a data greedy model which becomes more accurate with increasing data or parameter inputs, besides being a computational heavy model. This model has been used extensively in forecasting crop yields especially the major crops including maize, potato, cassava where there was data availability (Mishra et al., 2016).

Pant et al., (2021) applies a set of comparison machine learning models to assess crop productivity on agricultural lands including, decision trees model, random forest, support vector machine model and gradient boosting method. The availability of open datasets that can be used to train the model eased the process of data acquisition, however, with the history of datasets not assured, there may be introduced biases on the modelling. According to this research, decision trees tended to perform much better contrary to other studies such as Wolanin et al., (2019) and (C. Folberth et al., 2019) which had random forest model performing better when trained with minimal dataset. **Error! Reference source not found.** highlights the general concept of machine learning as cited in the study.

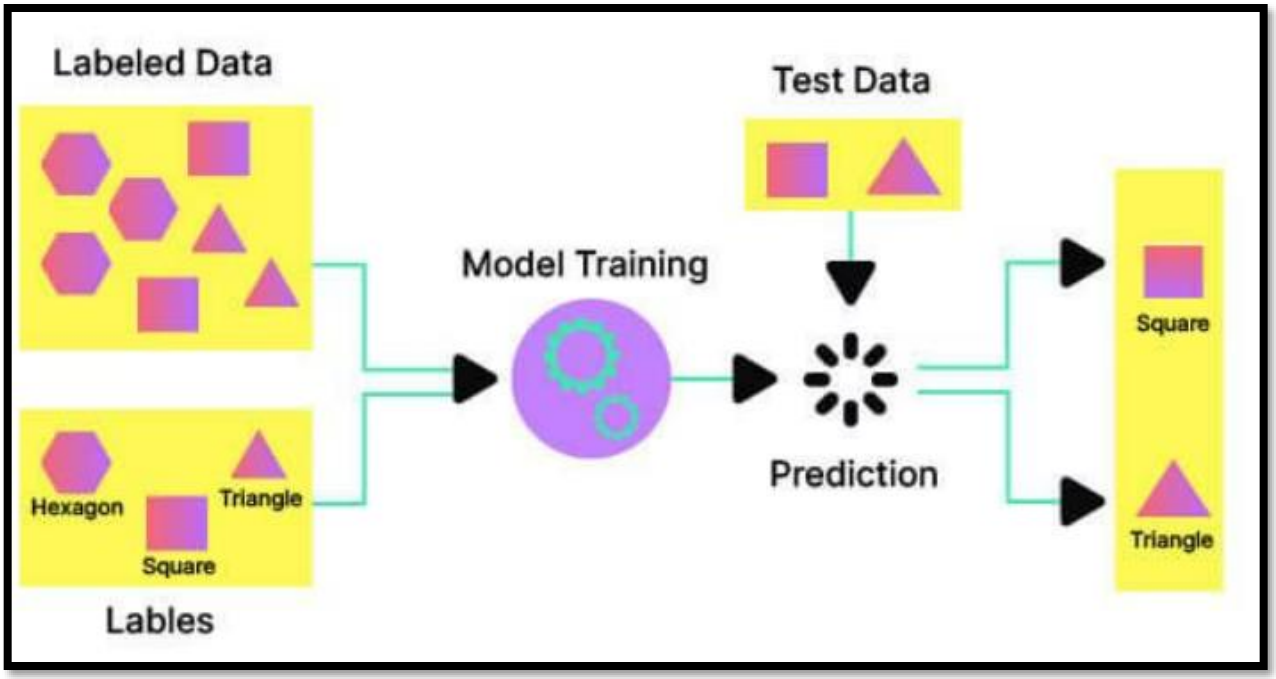

Figure 2.1: Machine Learning Implementation Plan

2.3 Process Based Crop Models for Crop Productivity Simulation

Crop models are simulators that allow the description of how crop genetics, crop management practices and environment influence crop growth. Crop models have been used for purposes such as crop production or yield prediction based on bio-physical characteristics, management decisions and adaptations policies (Ntuchu et al., 2021).

Gohain et al., (2021) applies Decision Support System for Agrotechnology Transfer (DSSAT)—crop simulation model (CSM) model to assess the crop water stress during crop growing period on drought resistant crops such as cassava. Notably, the short comings of crop simulation models are the needs for input parameters which are demanding and may not directly be available at the time of modelling (Ntuchu et al., 2021).

Crop simulation models are popular as they quantify impact of changes to climate, or crop management for different Agro-ecological zones however, great uncertainties always present itself as the models requires localized farmer inputs.

Where the input parameters are not present, the use of crop models tends to provide very uncertain outputs. Feleke et al., (2021) takes an ensemble approach of process-based model including APSIM-Maize, DSSAT CREES -Maize and Aqua Crop models to simulate maize growth, however the scale of the models was localised as there was need to collect data for the area of interest. José Luis Noriega-Navarrete et al., (2021) compares the future scenario maize production using different climate scenario datasets on dynamic and stochastic models, such as DSSAT-CERES-Maize, APSIM-Maize, CropSyst, AquaCrop, EPIC-Maize, CropWat InfoCrop, and WOFOST where the model showed a varying level of production with Africa being hit the most. However, it is unclear why Africa as a continent, with the most favourable climate and largest area with natural vegetation, still under performs continents such as Europe on climate change impacts.

These crop model requirement makes it difficult to scale across an extensive geographical area, making it un-applicable for context that require huge spatial variation and ground data is not available.

2.4 Comparative Review Between Process-based Model and Machine Learning Model Performance on Farm Productivity Forecasting

Leng and Hall, (2020) compared the use of different methods on assessing the impact of climate change on crop production with a focus on computing yield disparities overtime. The assessment show that both regression and machine learning models can be used to explain yield pattern variability, while large bias is found for process-based crop models even fed with harmonized parameters with less predictive capability of below 50%. Batool et al., (2022) had similar agreement with those of (Leng & Hall,2020) where machine learning regression algorithm performed better in yield prediction with fewer data than the simulation model, with a general error margin of less than 30% for the machine learning models. When assessed closely, although simulation model takes time to develop and understand the factors that influence productivity, the black box effect of machine learning model tends to identify patterns better than simulation models, and its ability to be used with less data input than the processed based model is promising.

2.5 Periodic Polynomial Regression to assess Farm Productivity

Periodic Polynomial Regression (PPR) has been used to simulate farm productivity decline. PPR is a type of regression analysis where the relationship between the independent variable (Weighted Regularized Extreme Learning Machine (SCA-WRELM) is used to estimate a dependent variable.

It statistically computes the relationship between input variables using the second order equations, that contains a cross-product interaction of independent variables. PPR is used to assess the relationship between the dependent and independent variables with no linear relationship similar to the input parameters considered in this study. The polynomial regression takes the shape as highlighted in **Error! Reference source not found.**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_h X^h + \epsilon, \qquad (1)$$

Where *h* is called the **degree** of the polynomial defined by the number of independent regression coefficients variables. The second-degree coefficients are described using equation **Error! Reference source not found.**

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{50} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{50} & x_{50}^2 \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{50} \end{pmatrix} \qquad (2)$$

PPR estimates better when subjected to large data sample size, where big spatial data is downloaded and engineered for the model. Statistical overflows is also curtailed where the computing power being applied is huge especially when working with higher degree terms.

When PPR is subjected to periodic time series data, the pattern of the regression model repeats itself based on the time interval of the data. The repeating polynomial regression model is then termed as the periodic regression model which is the linear combination of polynomial of n degree and periodic components (Čobanović et al., n.d.). The cyclic polynomial regression model maybe explained by n number of degree of polynomial trends, depending on the degree that fits the data distribution better. Figure 2.2 presents a sample cyclic polynomial regression model.

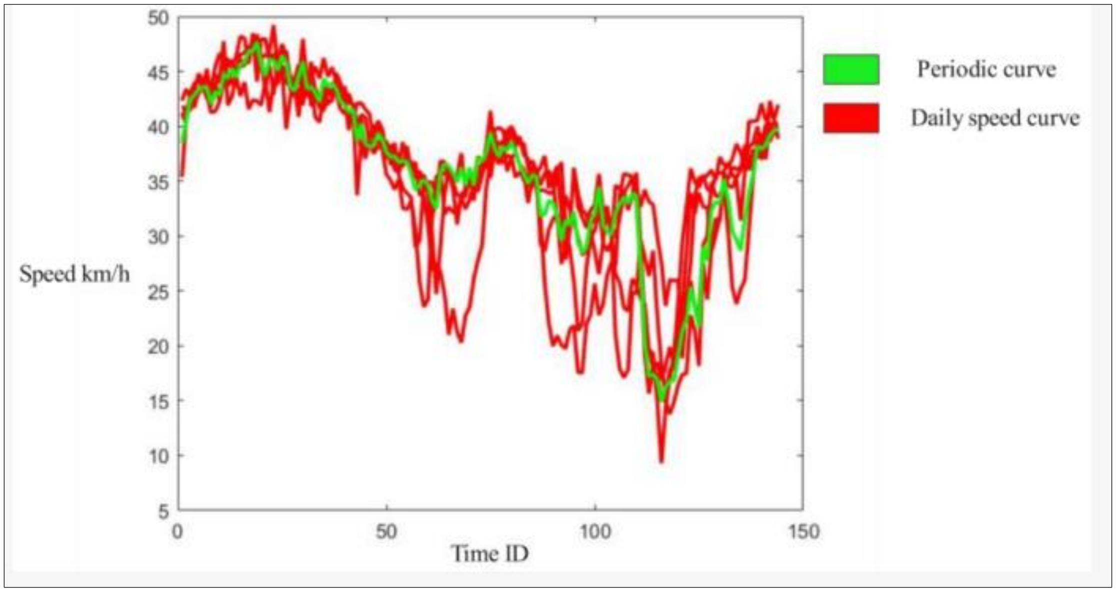

Figure 2.2: Example of a Periodic/Cyclic Polynomial Regression Model

The graph in Figure 2.2, shows an example of cyclic polynomial output. The output captures the motorist speed against time. The green curve represents the fitted periodic curve, while the red curve represents the daily speed curve for the motorist.

2.6 Conceptual Framework

The conceptual framework outlines the interaction between the dependent and independent variables. The independent variables included climate and soil-based factors, while the dependent variable was computed from satellite-based products and assessed against the independent variables. These relationships were evaluated using machine learning and cyclic polynomial regression models to identify trends over time. The research's capability has been demonstrated through the development of a forecasting model for farm productivity decline, validating the interaction between these variables. Figure 2.3 presents the conceptual framework for this research

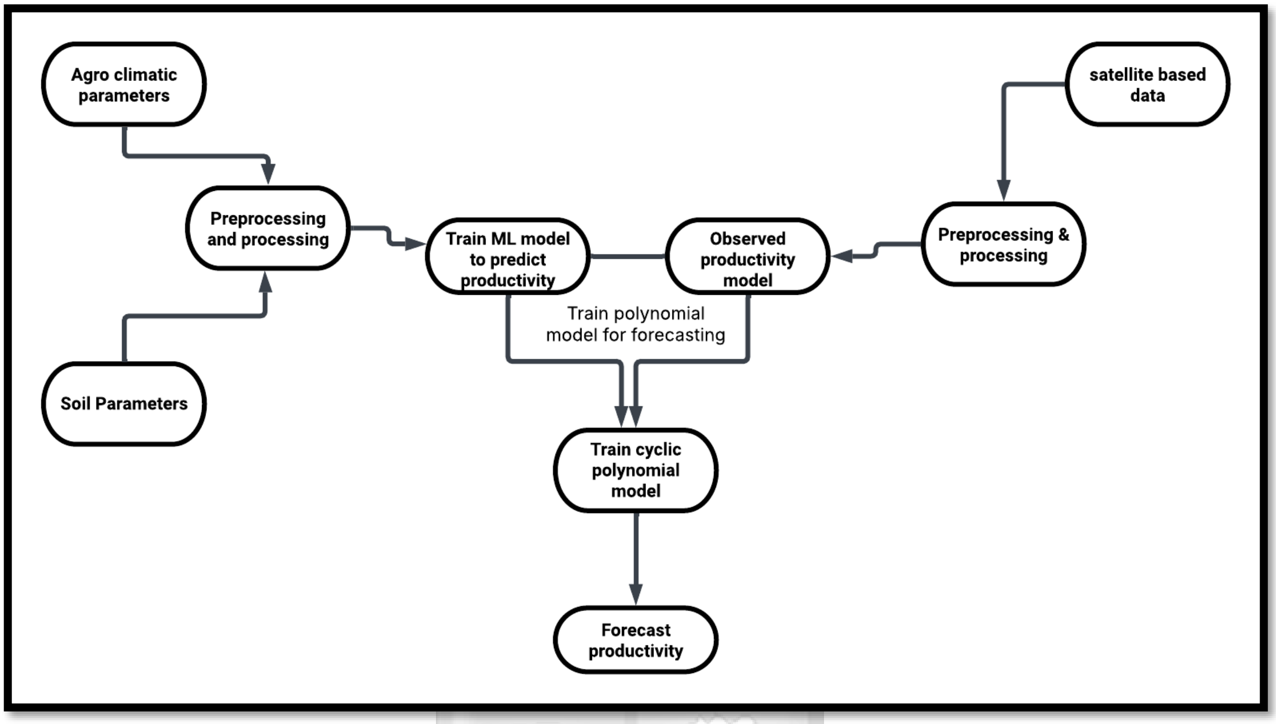

Figure 2.3: Conceptual framework

2.7 Operationalization of variables

In this study, operationalization is vital for understanding the interplay between climate factors, soil health, and agricultural productivity. The independent variables included climate-based datasets, soil-based datasets, and satellite imagery datasets, which were operationalized using independent variables as follows:

i. **Climate-Based Datasets**

Minimum and maximum daily temperatures (°C) were sourced from open databases referenced in this study and integrated with other independent variables to assess farm productivity decline. Similarly, total monthly precipitation (mm) was obtained as part of the independent variables used to describe variations in productivity. These climate variables played a crucial role in evaluating the impact of weather patterns on land productivity trends over time.

ii. **Soil-Based Datasets**

Soil organic carbon layers and soil moisture levels were obtained from remotely sensed data collected by various satellite sensors orbiting the Earth. These satellites provide high-resolution soil property estimates, enabling the assessment of soil organic carbon distribution and moisture variability across different landscapes. These datasets were acquired to form part of the independent variables for farm productivity model.

iii. **Satellite Imagery Datasets**

These datasets were used to assess vegetation health, with indices such as the Normalized Difference Vegetation Index (NDVI) calculated to evaluate changes in crop vitality over time as a proxy to farm productivity. NDVI was derived from satellite imagery by analysing the difference between near-infrared (NIR) and red-light reflectance, providing insights into crop vigor, stress levels, and overall crop condition throughout the study period.

These variables were combined using regression modeling to estimate changes and trends in farm productivity over time, allowing for the identification of productivity declines. Existing research often overlooks scalability, making it challenging to apply findings to a larger number of farmers. Addressing this gap is essential for developing effective interventions that enhance the resilience of agricultural practices, particularly for smallholder farmers. The extensive methodology applied is discussed in the next chapter.

Chapter 3: Methodology

3.1 Introduction

This chapter outlines the methodology used to achieve the research objectives. The study primarily relied on remote sensing data, integrating machine learning and polynomial regression to estimate farm productivity decline and forecast productivity trends.

The conceptual framework centres on modelling farm productivity decline based on key influencing factors, including soil and climate-related parameters. The process began with the collection of input data, which included climate variables (e.g., temperature and rainfall), soil characteristics, and historical farm productivity records derived from multi-temporal satellite imagery. These datasets underwent a pre-processing phase, followed by data engineering to ensure compatibility with machine learning algorithms.

To address data gaps, particularly in soil parameters, a machine learning model was applied for imputation. Subsequently, cyclic polynomial regression was used to analyse periodic trends in the data, focusing on how climate variability such as seasonal fluctuations and long-term climate cycles affects productivity. The final step involved employing the cyclic polynomial model to forecast future productivity trends, providing insights into potential declines and aiding in proactive decision-making for sustainable farm management.

3.2 Research Design

This research employed a quantitative design, utilizing machine learning methods and mathematical modelling. The quantitative approach enabled the forecasting of productivity decline. The study focused on analysing how productivity and climate factors interacted over time, incorporating aspects of a longitudinal study design. The research aimed to develop a model that predicted future productivity levels based on past and current data. Numerical variables related to farm productivity, such as soil health parameters and climate factors including rainfall and temperature influx were used across the study area

This research employed a quantitative design, integrating machine learning methods and mathematical modelling to analyse and forecast farm productivity decline. The study follows a structured data-driven approach, aligning with the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which consists of six key phases: problem identification, data understanding, data preparation, modelling, evaluation, and deployment, as outlined in section 1.4 Research Objectives.

The study adopts a longitudinal design, focusing on how productivity and climate factors interact over time. By utilizing historical and current numerical variable including soil health parameters, rainfall, and temperature fluctuations, the research develops predictive model for farm productivity decline. The ultimate output being understanding the trend of farm-land productivity and identify regions that are performing low or experiencing productivity decline.

The structured CRISP-DM framework ensures systematic data handling, from preprocessing and feature selection to model training, validation, and forecasting. The final model is evaluated for accuracy and reliability before being used for predictive analysis, contributing to informed farm management and climate adaptation strategies. The process followed for this research is shared in Figure 3.1.

Figure 3.1: CRISP-DM Framework (Adapted from Schröer et al, 2021)

3.3 Study Area

This research focused on implementing the land productivity decline model in the counties that mostly practice agriculture coloured in green, well-known as Kenya's food basket as indicated in Figure 3.2.

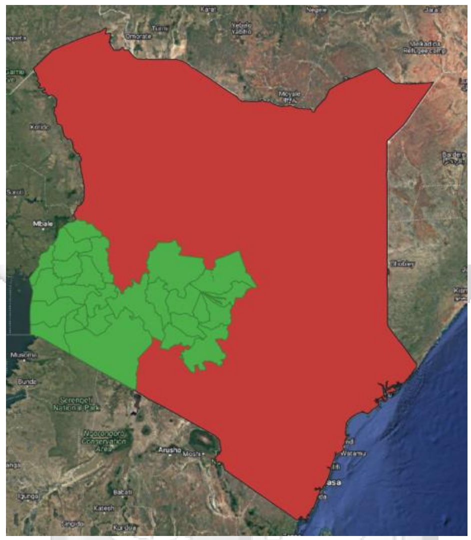

Figure 3.2: Location of the Study Area relative to the Map of Kenya

3.4 Data Collection

Error! Reference source not found. gathered and applied in analysis aspects for this research.

Table 3.1: Data Sources and their Temporal Resolution

PARAMETER	TEMPORAL RESOLUTION	DATA SOURCE
Soil Moisture	Daily (2013 – 2024)	NASA Global Land Data Assimilation System Version 2 (GLDAS-2)
Rainfall	Daily (2013 – 2024)	Climate Hazards Centre InfraRed Precipitation (CHIRPS)
Radiation photosynthetic data	Daily (2013 – 2024)	MODIS
Soil organic carbon	Annual (2013 – 2017)	Innovative Solutions for Decision Agriculture Ltd. (iSDA)
Temperature	Daily (2013 – 2024)	Climate Hazards Centre InfraRed Temperature (CHIRTS)
Satellite imagery	Weekly (2013 – 2024)	Sentinel 2

3.5 Data Pre-Processing

Data pre-processing steps were applied to the raw datasets to ensure they were ready for analysis. Since this research acquired data from multiple sources, various pre-processing techniques were required. The datasets obtained were largely categorized into three main groups, each undergoing specific pre-processing steps to ensure consistency and suitability for analysis

i. Climate Based datasets: Temperature, Rainfall, Radiation photosynthetic data
ii. Soil based datasets: Soil Organic Carbon layers
iii. Satellite imagery datasets: Sentinel 2 datasets

3.5.1 Pre-processing of Climate Based Datasets

Climate based datasets are often acquired as analysis ready data. Occasionally they come with data gaps especially around areas that satellite sensors malfunctioned, or there was a heavy cloud cover at the time of capture. To handle missing data and data gaps, this research applied interpolation methods such as kriging to fill the data gaps and estimate the missing values. Kriging is based on semi variogram which quantifies autocorrelation based on Tobler's first law of geography, that states that closer things are more related. However, at a certain distance, autocorrelation becomes independent (sill), where there is not any spatial autocorrelation between the data points.

Figure 3.3: The Logic of Kriging Interpolation

outline the aspects of sill where relationship between the predicted point, and the known point flattens out.

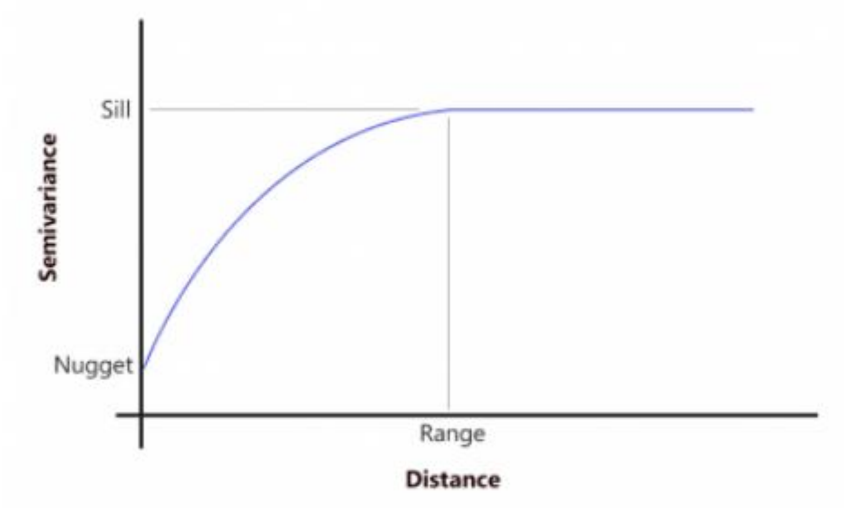

Figure 3.3: The Logic of Kriging Interpolation

3.5.2 Soil Based Datasets

Soil-based datasets, as highlighted in, did not cover all years as indicated. The datasets that did not span the entire research period were estimated using the Random Forest (RF) machine learning model. The RF regressor, a non-parametric ensemble classifier, generated multiple decision trees by randomly selecting subsets of training samples and variables. The RF model incorporated measured soil organic carbon data points across Africa to predict the missing layers from 2017 to 2022 (Belgiu et al., 2016).

The major advantage of using the RF model to predict soil organic carbon content was its ability to internally determine regression errors using unused datasets, known as out-of-the-bag samples (Mayesse et al., 2021). Figure 3.4 illustrates the role of Random Forest regression in utilizing decision trees to predict the dependent variable.

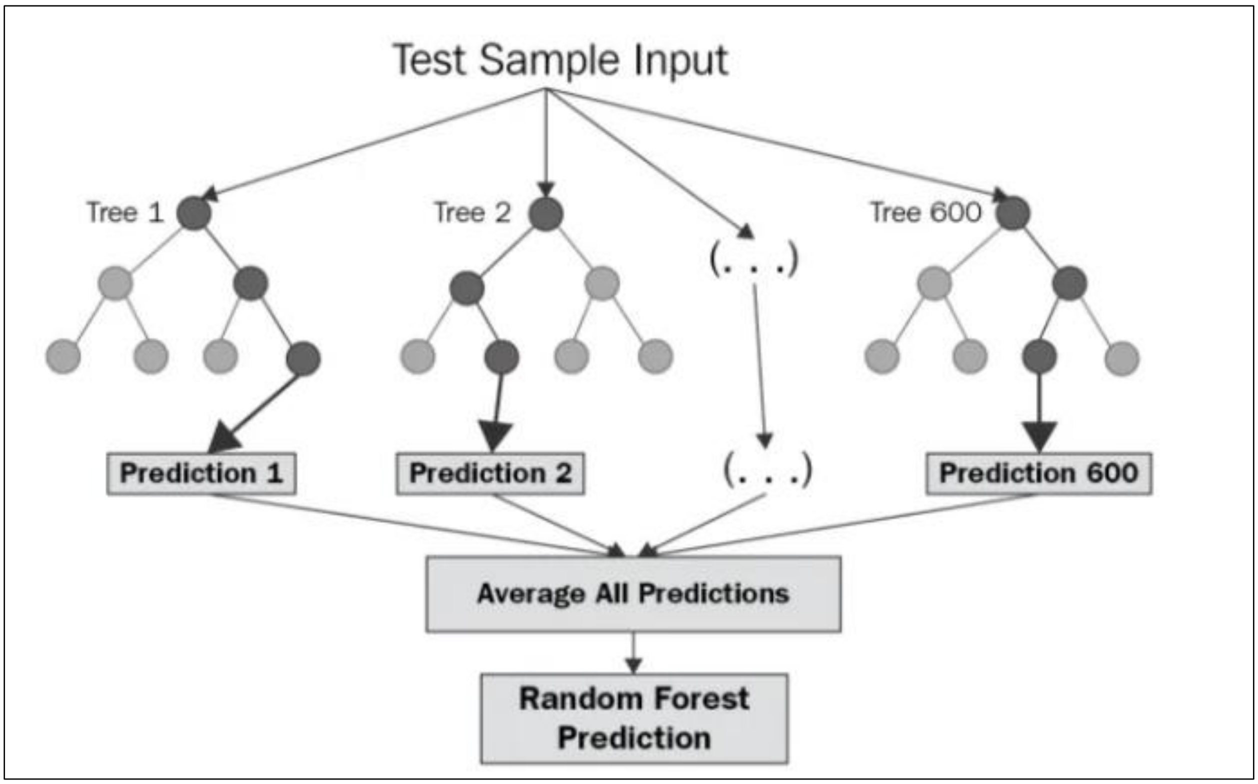

Figure 3.4: Random Forest Ensemble Decision Tree Logic

3.5.2 Satellite imagery datasets

A digital remotely sensed image is composed of Digital Number (DN) or Brightness Value (BV) located at the intersection of each row i and column j in band k in imagery. Raw images contain distortions, where the sources may range from variation in the altitude and velocity of sensor to the Earth rotation and curvature. Pre-processing steps aim to correct distortions and perform image rectification. These processes preceded processing of the data as highlighted in Figure 3.5

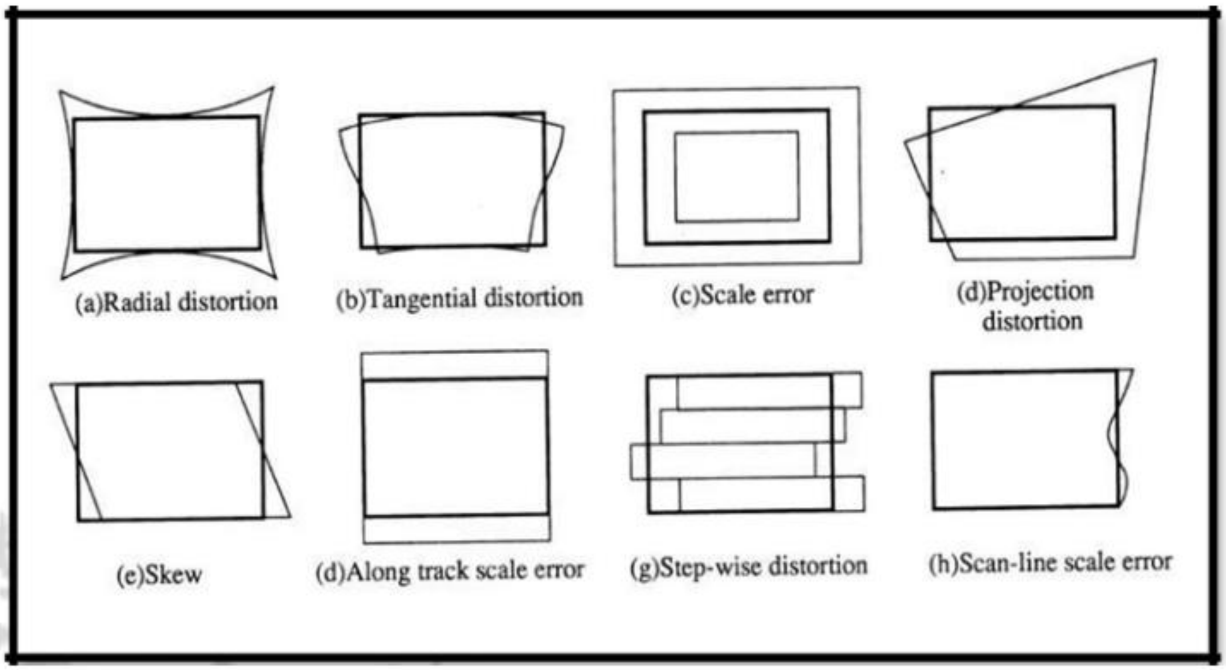

Figure 3.5: Types of Satellite Imagery Errors

To correct the errors as outlined in Figure 3.5, the following methodologies were applied:

I. Relative radiometric correction using histogram adjustments, used to normalize multi-temporal data taken on different dates to a selected reference data at specific time.

II. Single-image normalization using histogram adjustment: This method is based on examination of spectral characteristics of objects of known or assumed brightness recorded by multispectral imagery also used in correcting spectral errors in the dataset.

III. Solar elevation correction: Sun elevation correction accounts for the seasonal position of the sun relative to the earth. Image data acquired can be normalized by assuming that sun was at zenith at each date of sensing.

3.6 Modelling Farm Productivity

Various indices have been applied to model drought or flood conditions with indices such as Standard Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI), Palmer Drought Severity Index (PDSI), Soil Water Deficit Index (SWDI), Normalized Difference Vegetation Index (NDVI), Vegetation Health Index (VHI), vegetation Drought Response Index (VegDRI), and Scaled Drought Condition Index (SDCI) being applied. They incorporate factors such as rainfall, land surface temperature (LST), potential evapotranspiration (PET), soil moisture content (SM) (Lee and Kim, 2021).

To model multi temporal farm productivity before applying the periodic/cyclic polynomial regression model, the following indices were computed from the acquired datasets. The productivity was a composite index of parameters computed using the equation (3) to equation (6).

$$\frac{\boldsymbol{Rain} - \boldsymbol{Rain}_{\min})}{\boldsymbol{Rain}_{\max} - \boldsymbol{Rain}_{\min})} = \boldsymbol{PPT} \tag{3}$$

$$\frac{\boldsymbol{Temp}_{\max} - \boldsymbol{Temp}}{\boldsymbol{Temp}_{\max} - \boldsymbol{Temp}_{\min})} = \boldsymbol{Temp} \tag{4}$$

$$\frac{\boldsymbol{Soil\ Moisture} - \boldsymbol{Soil\ Moisture}_{\min})}{\boldsymbol{Soil\ Moisture}_{\max} - \boldsymbol{Soil\ Moisture}_{\min})} = \boldsymbol{SM} \tag{5}$$

$$\frac{\boldsymbol{EVI} - \boldsymbol{EVI}_{\min})}{\boldsymbol{EVI}_{\max} - \boldsymbol{EVI}_{\min})} = \boldsymbol{VH} \tag{6}$$

Farm productivity output was modelled using soil moisture, rainfall, radiation photosynthetic data, soil organic carbon, and temperature parameters. The soil moisture and radiation photosynthetic data were rescaled and incorporated into the farm productivity model. The rescaling process utilized a large membership algorithm, which transformed an input raster into a fuzzified raster, assigning values between 0 and 1. A value of 0 indicated no membership within the defined fuzzy set, while a value of 1 represented full membership. Intermediate values followed the large membership function to determine the degree of membership of raster values.

The large function was constructed using two user-defined input raster values, which set the point of half membership (midpoint, resulting in 0.5), along with a predefined function spread that controlled the function's uptake. The rescaled soil organic carbon and radiation photosynthetic data were represented by the symbols SO and RP, respectively. Figure 3.6 illustrated the shape and character of the transition zone for large membership function.

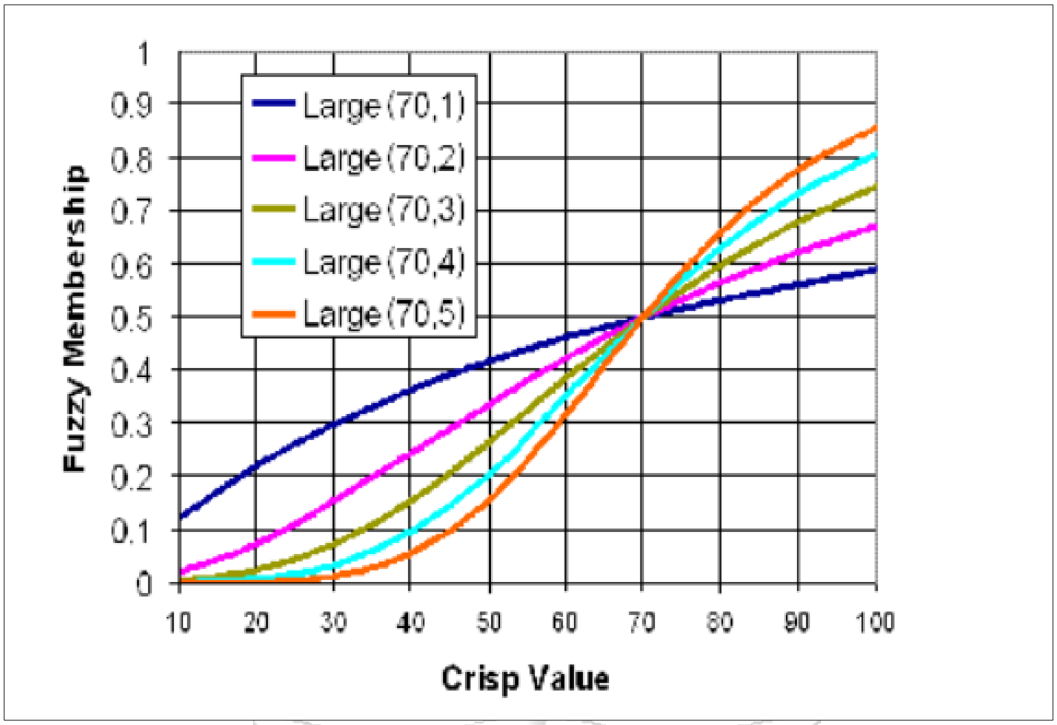

Figure 3.6: Variations of the Fuzzy Large Membership Function

Equation (7) illustrates how these parameters interact with each other to generate a multi-temporal productivity output.

$$\boldsymbol{Productivity} = (\boldsymbol{PPT} + \boldsymbol{Temp} + \boldsymbol{SM} + \boldsymbol{VH} + \boldsymbol{SO} + \boldsymbol{RP})/\mathbf{5} \qquad (1)$$

From equation (7), farm productivity is modelled as an average of the input parameters. The modelled farm productivity was then subjected to cyclic polynomial regression model to assess trends in farm productivity and consequently productivity decline.

Figure 3.7 illustrates the general implementation workflow of the research before model evaluation

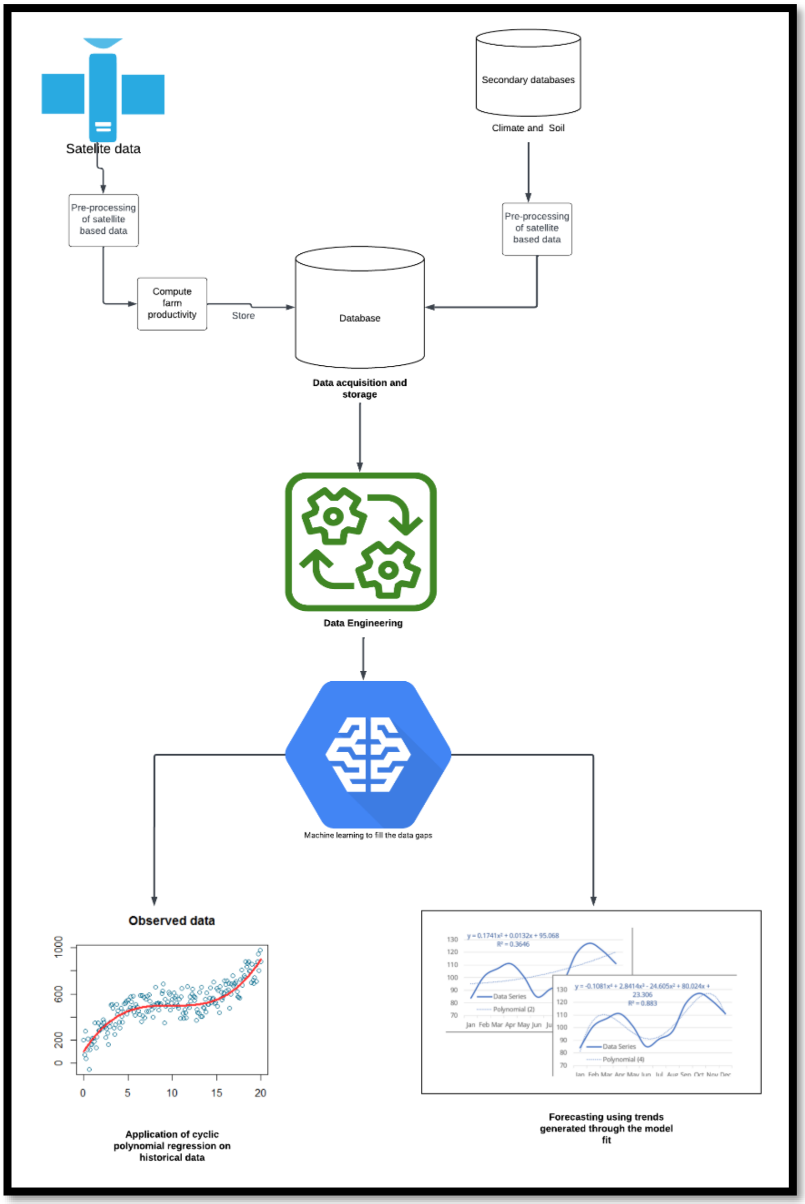

Figure 3.7: General implementation workflow

3.7 Model Evaluation

3.7.1 Random Forest Evaluation

The Out-of-Bag (OOB) assessment method was used in this study to provide a robust evaluation of the Random Forest model. OOB evaluation offered a quick and efficient technique to assess the model's quality without requiring a separate validation dataset. The lower the OOB estimate, the higher the accuracy of the model.

3.7.2 Polynomial regression evaluation

The polynomial regression model evaluation involved assessing how well the model fit the data and how effectively it generalized to new data. Several metrics were tested, including Mean Absolute Error (MAE), which evaluated the average magnitude of errors in predictions and provided an estimate of deviation from the actual values. Mean Squared Error (MSE) was also applied to estimate the average of the squared differences between actual and predicted values. MSE penalized larger errors more than smaller ones due to squaring the differences, making it particularly useful when large errors were undesirable as for this research.

Root Mean Squared Error (RMSE) served as an important metric, providing an error estimate in the same units as the target variable, making it more interpretable while still emphasizing larger errors. Additionally, R^2 (Coefficient of Determination) was used to indicate the proportion of variance in the dependent variable that could be explained by the independent variables. A higher R^2 value indicated a better model fit (Sun et al., 2014). For productivity decline tool, the MSE metric was however choses as the default value to assess accuracy for the end users

3.8 Tools and Software

This research utilized both R and Python scripting languages to develop and test the models. The implementation was scaled up using Fast-API, a web-based framework that serves application with data in form of application interface (API), allowing for an automated approach to understanding productivity decline for any farm. Additionally, Docker and gunicorn were used to deploy and manage the application efficiently.

For visualization, libraries including Seaborn, Matplotlib, and ggplot2 were employed to evaluate the graphical outputs generated by the models. Furthermore, QGIS, an open-source software, was used to analyse some of the prediction layers generated through the Random Forest processes. The front-end interface was developed using Next.js with TypeScript to facilitate user interaction with the model outputs

3.9 Dissemination of Results

The results of this study is disseminated through a web platform, allowing users to access the farm productivity decline model at scale. This platform enabled multiple users to simultaneously view insights and access relevant information pertaining to their productivity trend.

3.10 Utilization of Results

The results could be utilized by smallholder farmers, agricultural extension officers, and policymakers. Farmers can use the forecasts to assess potential declines in productivity and take proactive measures to mitigate risks. Agricultural advisors and policymakers can leverage the insights to develop strategies and support systems aimed at enhancing resilience and preparedness across farming communities.

3.11 Ethical Consideration

This research used data sets that are available online and are published under open licensing agreement that allows unlimited use. They are in openly accessible databases as highlighted in **Error! Reference source not found.**. Some of the licensing agreement applicable include the following:

i. Creative Commons Licenses (CC) which requires attribution but allows users to share, use, and adapt the data freely and without any restriction.
ii. Open data commons (ODC) Licenses allows free use, sharing and adaptation of the data if attribution is provided.
iii. General Public Licensing (GPL) for data which allows free use and modification, provided that any derived works are also shared with the same freedom.

Although this research was based on the above licenses, it still underwent an ethical review to ascertain the research methodology and design. The ethical approval is attached in the Appendices sections.

Chapter 4: System Analysis and Design

4.1 Introduction

This chapter delves into the system development process, aiming to transform the defined requirements into a structured conceptual framework or architecture. It outlines user interactions with various system components and details the responses generated based on user requests.

4.2 System Requirement Analysis

The success of system development and implementation largely depends on effectively gathering requirements from both stakeholders and end users. Requirement elicitation, a fundamental aspect of requirement engineering, focuses on identifying the necessary system functionalities and constraints. As described by (Fauzan et al., 2025), requirement engineering involves not just elicitation but also analysis, validation, documentation, and ongoing management of requirements.

A well-structured requirement-gathering process guarantees that system specifications are clear, verifiable, traceable, consistent, and complete. Given the nature of the system developed in this study, a rapid design approach was necessary. As a result, the requirements gathering heavily relied on an extensive literature review, complemented by the researcher's expertise in the agricultural domain as supported by Suarez et al, (2011). This approach ensured a deep understanding of the problem, as outlined in the problem statement, allowing for the development of a system that effectively addresses the identified challenges. The system developed relied on the categorization of functional and non-functional system needs for inclusion in implementation.

Functional Requirements includes:

i. Access platform through a web address
ii. Upload farm boundary as a ***geojson*** format data
iii. View predicted and forecasted outputs

Non-Functional Requirements includes:

i. Functionality to achieve the intended prediction goals
ii. Simplicity and ease of interpretation
iii. High precision in prediction-making

4.3 System Architecture

The system architecture includes a middleware layer that serves as the core for prediction and forecasting logic. This middleware was implemented using the Fast-API framework in Python, ensuring efficient request handling and seamless integration with other system components, as outlined in the system design. Figure 4.1outlines the system architectural design.

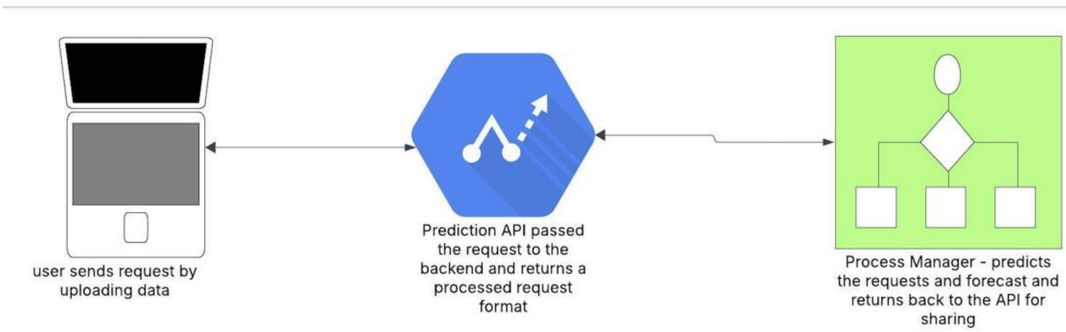

Figure 4.1: System Architectural Design

4.3.1 Use-case Diagram

Use Case Diagram was essential for outlining the key operations of the forecasting model system, including data input, model training, prediction generation, and visualization. It helps identify user interactions by illustrating how the user interacts with components of the system. Moreover, it served as a visual tool that aids communication, allowing stakeholders to easily understand how the model functions. The use-case diagram for the system is outlined in Figure 4.2

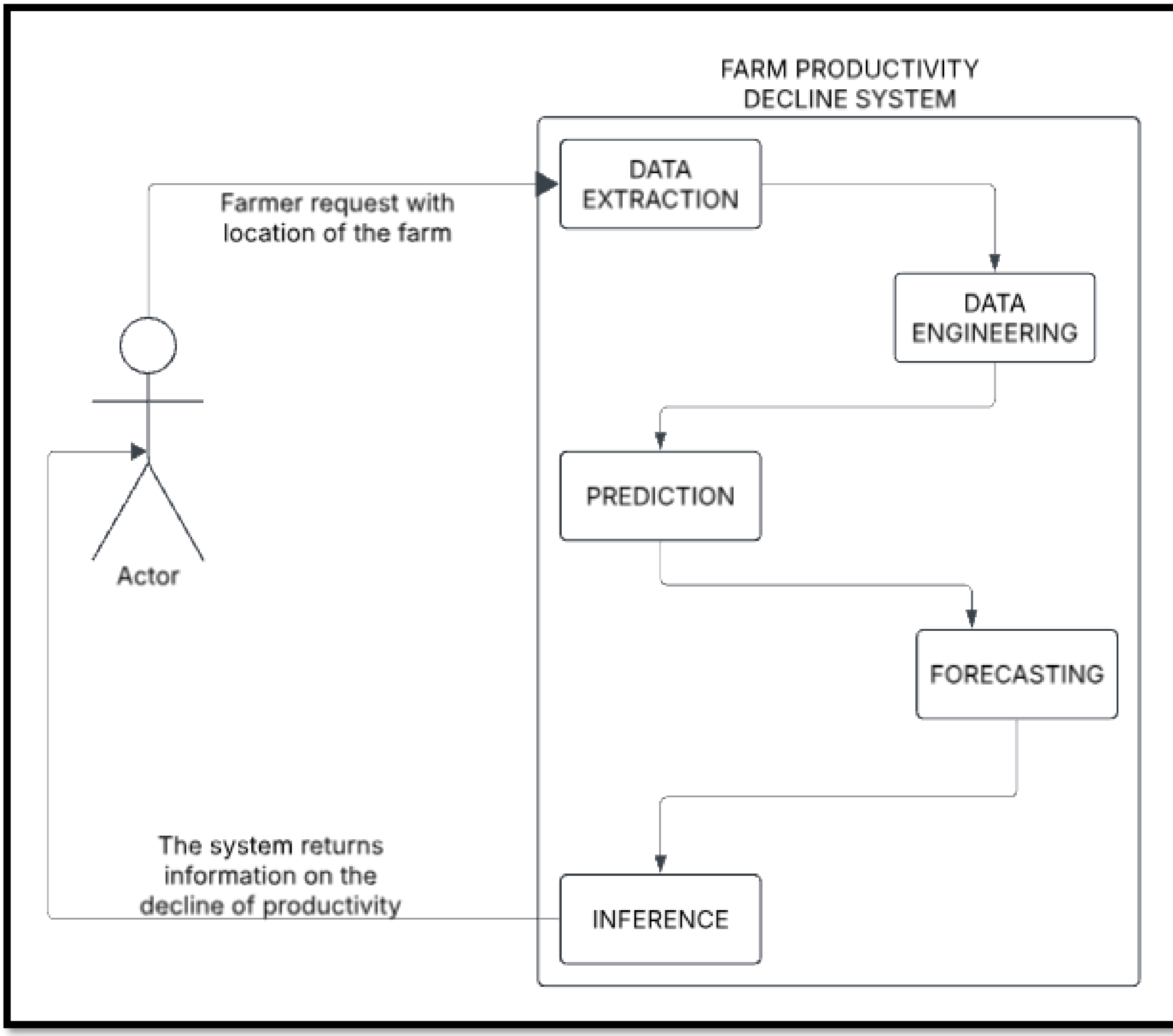

Figure 4.2: Use Case Diagram of the System

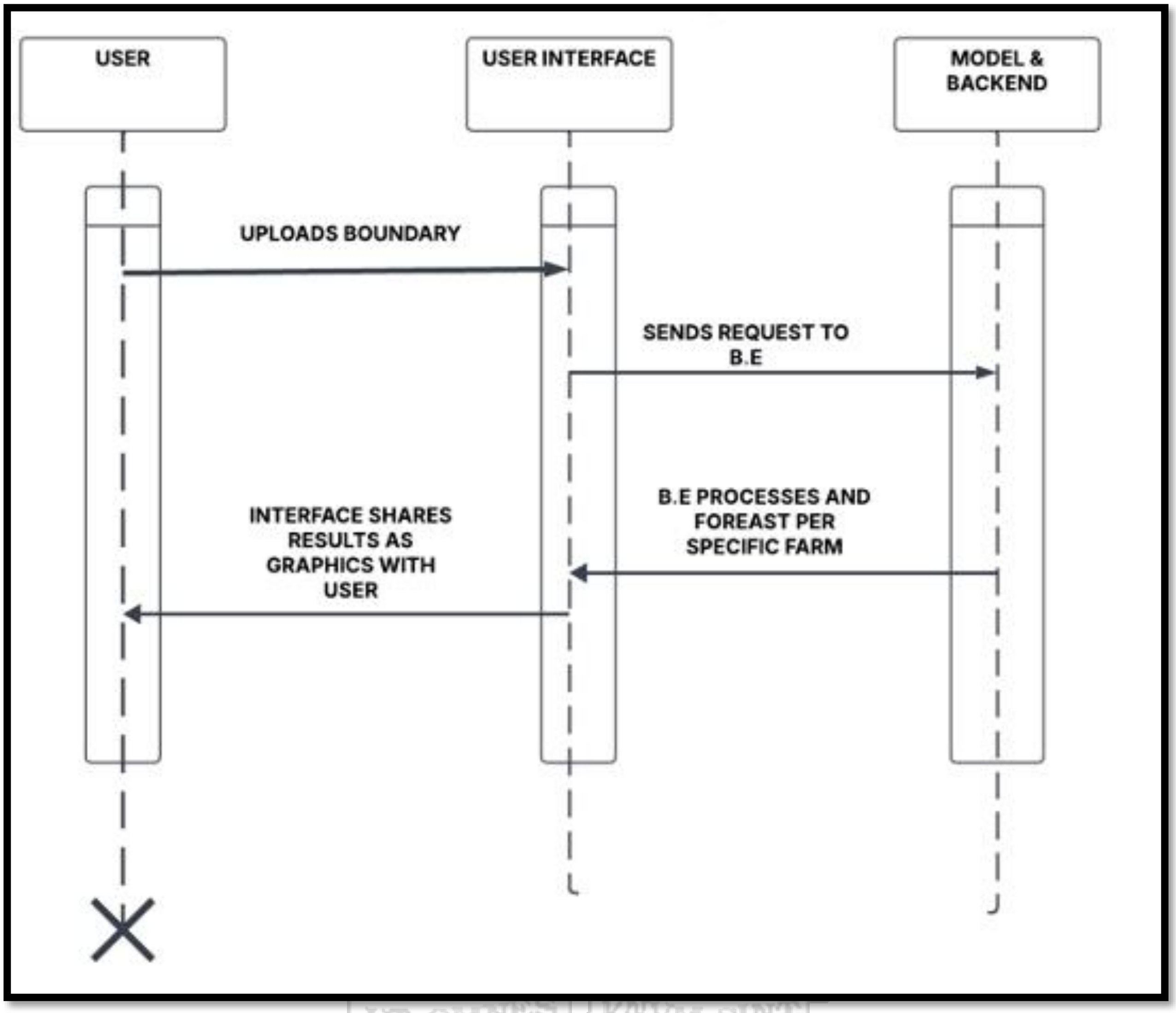

Figure 4.3: Sequence Diagram Illustrating System Interaction

Chapter 5: System Implementation & Testing

5.1 Background

The study aims not only to forecast productivity trends and identify points of decline but also to predict the productivity of specific farmland within the farming counties in Kenya. To achieve this, various machine learning models were explored and evaluated for their effectiveness in predicting land productivity. The models tested include:

i. Convolutional Neural Networks (CNNs) – Leveraging deep learning to capture spatial patterns and complex relationships in productivity data.
ii. Random Forest (RF) – A robust ensemble learning method that enhances predictive accuracy by averaging multiple decision trees.
iii. Support Vector Machine (SVM) – A powerful algorithm known for its ability to handle high-dimensional data and identify optimal decision boundaries.

Each model was trained using historical productivity data, with the dependent variable representing land productivity. The goal was to assess the prediction capability of each model and determine which provides the most reliable results for forecasting productivity trends and identifying potential areas of decline.

The findings and comparative analysis of these models are presented in the next section

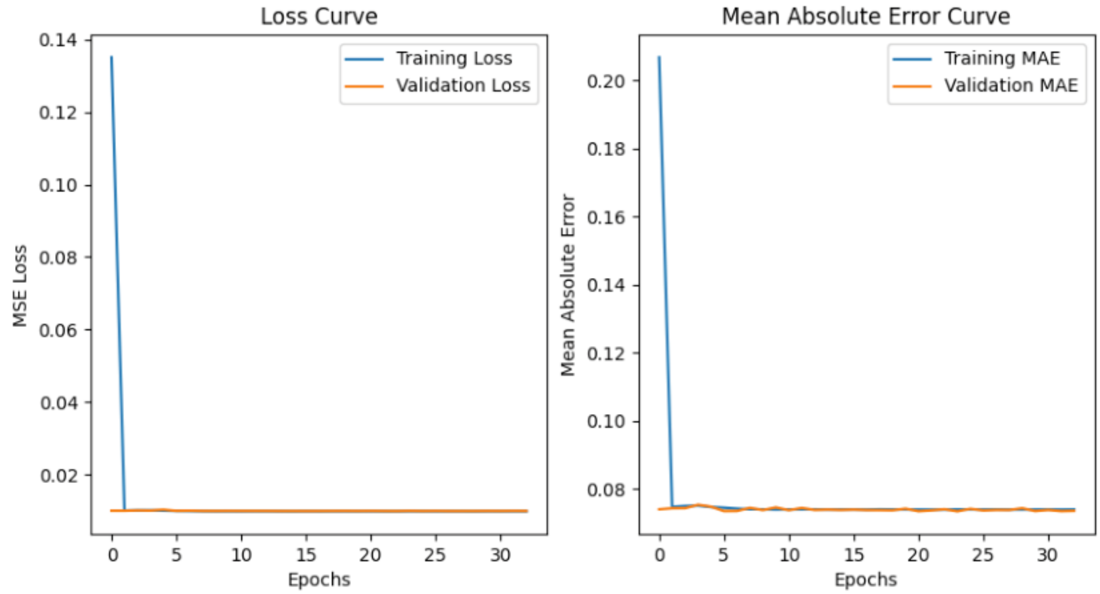

Figure 5.1: CNN Training Curves

Figure 5.1 shows the Loss Curve (Left) and Mean Absolute Error (MAE) Curve (Right) for Convolutional Neural Network (CNN) model. The CNN model demonstrates fast convergence, learning quickly within the first few epochs and stabilizing efficiently. Additionally, there is no major overfitting, as indicated by the stable validation loss and Mean Absolute Error (MAE), ensuring that the model generalizes well to unseen data. However, there is a possibility of underfitting, as the loss is very low, suggesting that the model might not be capturing deeper relationships in the data. Although (Teja Kattenborn et al., 2021) outlines the usability of CNN model to predict different vegetation scenarios, the model tends to underfit. This prompted the testing of random forest model in predicting land productivity. Figure 5.2 shows random forest training curve.

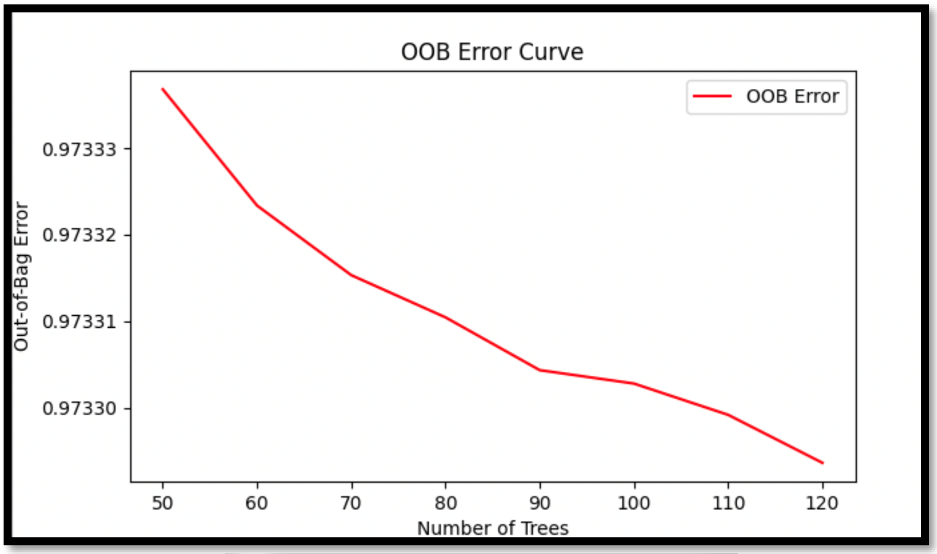

Figure 5.2: Out of the bag (OOB) Random Forest Training Curve

The OOB error curve shows how the out-of-bag error decreases as the number of trees in the Random Forest model increases from 50 to 120. Initially, the OOB error starts high and gradually decreases with more trees, indicating improved model performance. However, the decline in OOB error slows down as more trees are added, suggesting diminishing returns in further increasing the number of trees. This trend indicates that the model stabilizes around 100-120 trees, beyond which adding more trees does not significantly improve generalization.

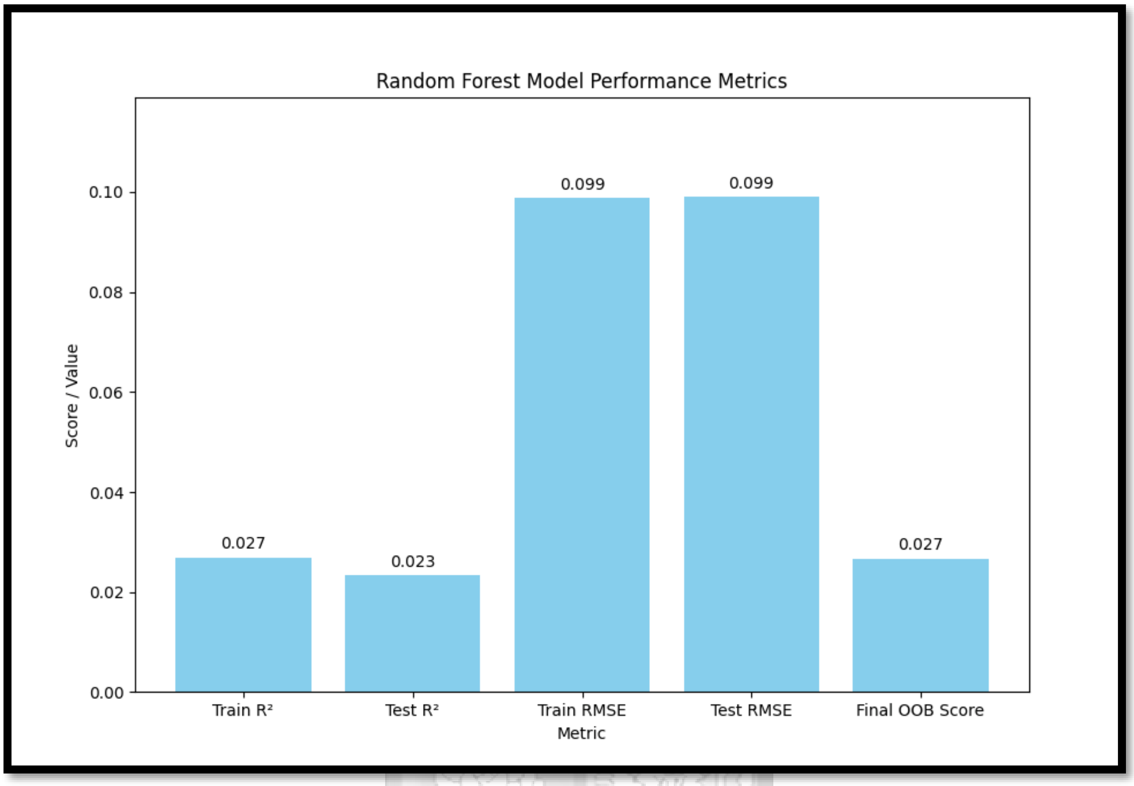

Figure 5.3: Random Forest Performance Metrics

The bar chart presents key performance metrics of the Random Forest model, highlighting its ability to generalize but also revealing its limited predictive power. The Training R^2 (0.027) and Testing R^2 (0.023) values are both low, indicating that the model explains very little variance in the data. However, the similarity between these values suggests that the model generalizes well without overfitting. The Training RMSE (0.099) and Testing RMSE (0.099) are identical, demonstrating that the model maintains a consistent level of error across both datasets. This stability further supports the conclusion that the model is not overfitting but may not be capturing the underlying patterns effectively. Additionally, the Final OOB Score (0.027) is quite low, reinforcing the observation that the model struggles to explain variance in the dataset. The output shows lack of depth with random forest model in predicting the farm productivity in **Error! Reference source not found.** in comparison with support vector machine model in Figure 5.5.

Figure 5.4: Random Forest Prediction

Figure 5.5: SVM Prediction

Comparatively, the SVM model demonstrated a greater ability to capture the spatial dynamics of the area of interest, providing more detailed and nuanced predictions. In contrast, the Random Forest model tends to generalize the predictions across the area, treating it as a uniform region rather than accounting for local variations. This over-generalization in the Random Forest model results in a loss of finer spatial details, which may reduce its accuracy in capturing subtle differences within the dataset. Consequently, the SVM model proves to be more effective in modelling the underlying complexity of the region.

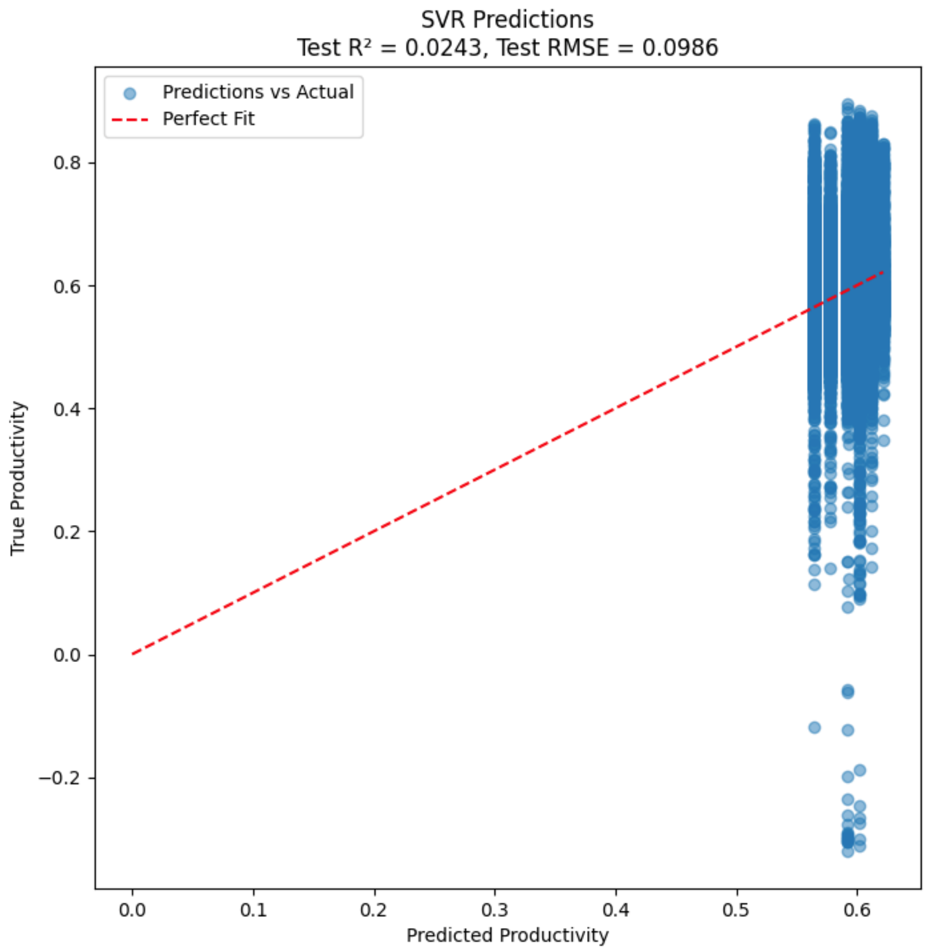

Figure 5.6: SVM Model Performance Graph

Figure 5.6 indicates that the SVM regression model is not capturing much of the variance in the data. However, it outperformed the other models under consideration. Although the test $R^2 \approx 0.0243$ suggests that the model explains only about 2% of the variability in the target variable (productivity), training on a larger dataset covering a broader area could improve its ability to capture data dynamics and better explain the complexities of the target variable. Some predicted outputs are presented in Figure 5.7 to Figure 5.9.

5.2 Assessment of Model Prediction for Farm Productivity

The following figures illustrate the predictive capacity of the trained model across different years. The model is utilized to forecast the decline in farm productivity within the study area, providing insights into long-term trends and potential future outcomes.

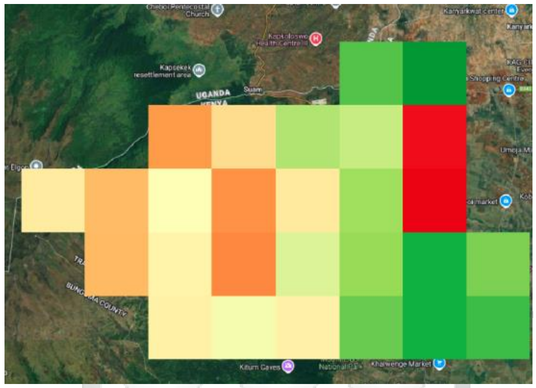

Figure 5.7: 2022 Predicted Productivity

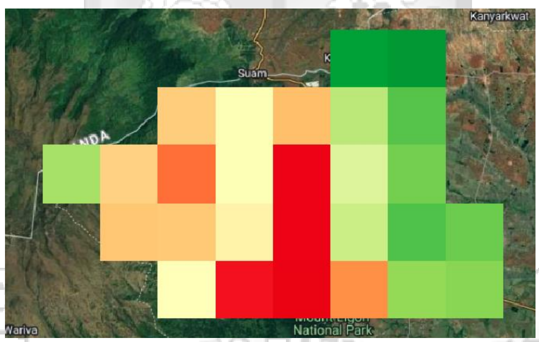

Figure 5.8: Predicted Productivity for 2021

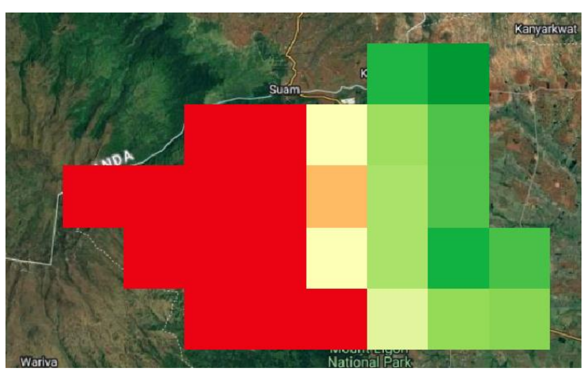

Figure 5.9: Predicted Productivity for 2020

It is evident that the trained model is able to pick the dynamism in the data which is a pre-requisite for a regression-based model to be used in this study.

With more retraining of the random forest model, the model improved and provided stronger metrics as compared to other models initially performing better, including SVM model. This prompted to using the random forest model to generate the prediction going forward. The metrics are highlighted in Figure 5.17.

5.3 Comparative Assessment and forecasting

To forecast farm productivity trend, independent variables were prepared for the area of interest spanning 23 years i.e. from 2000 to 2023.The trained model predicted a productivity as shown in Figure 5.10 in comparison with observed productivity outputs in. Figure 5.11.

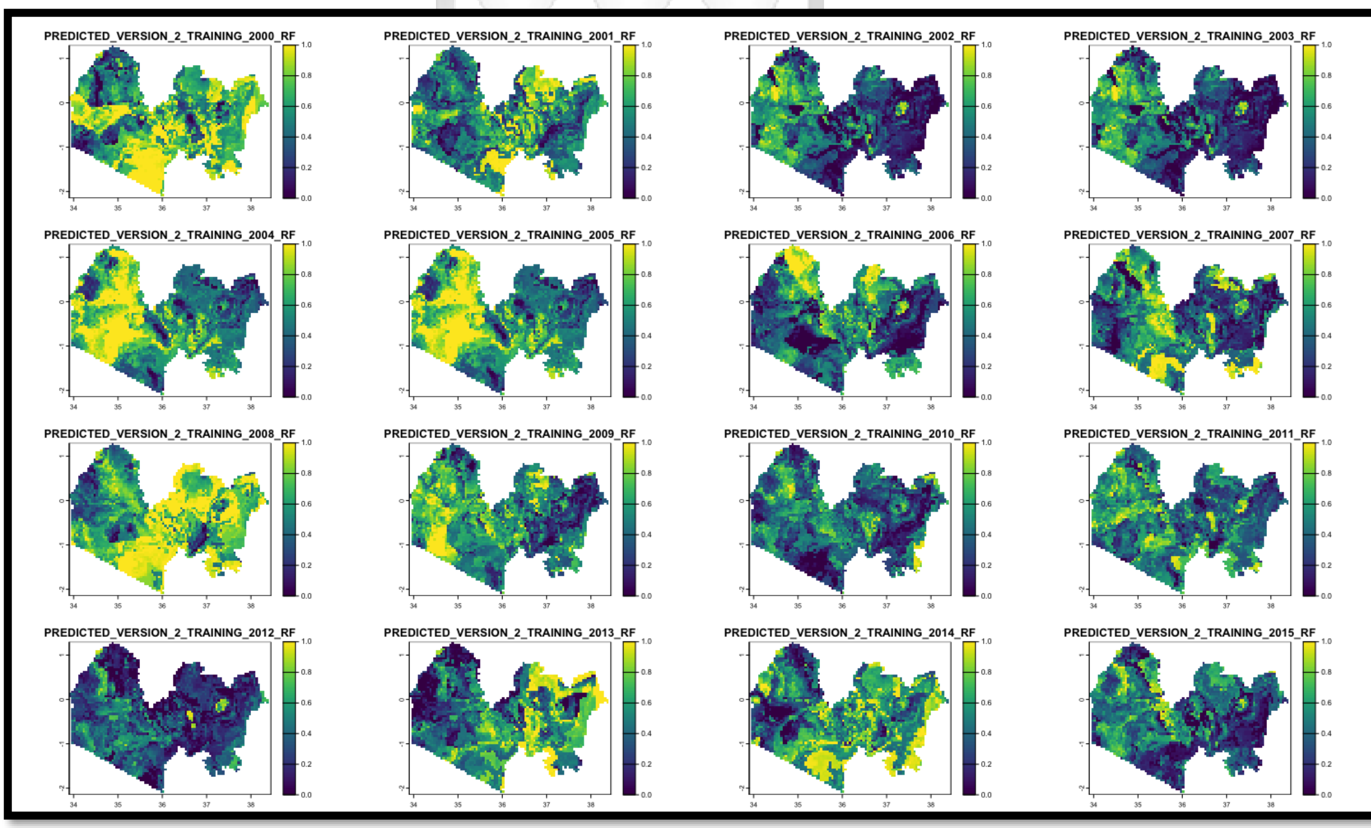

Figure 5.10: Predicted Productivity for the Food Basket Areas Of Kenya 2000 to 2015

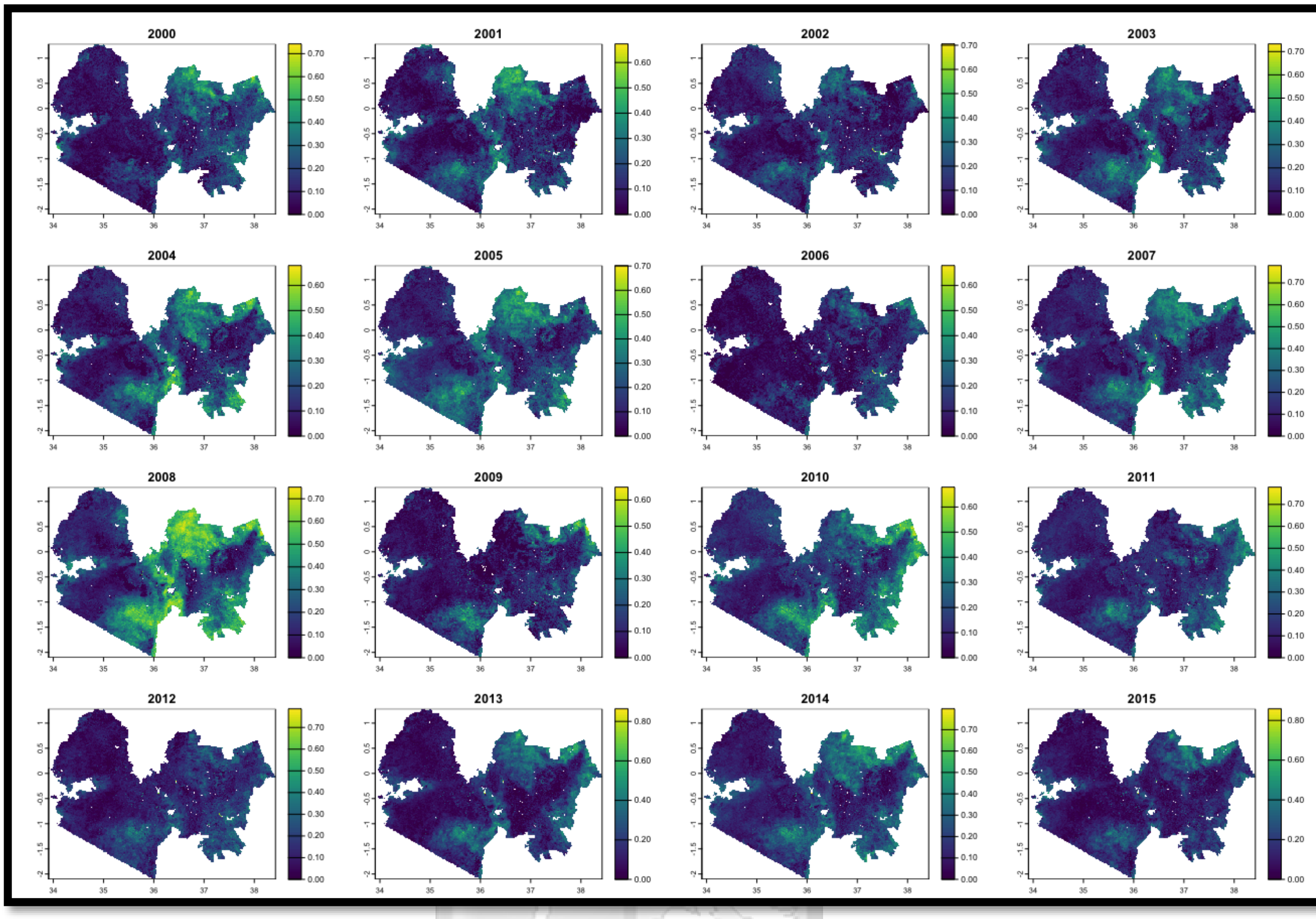

Figure 5.11: Observed Productivity Values for Years 2000 – 2015

The code used to extract observed productivity is shown in Figure 5.12.

```
// Function to calculate annual mean GPP
var annualGPP = years.map(function(year) {
  var start = ee.Date.fromYMD(year, 1, 1);
  var end = start.advance(1, 'year');

  var yearlyGPP = dataset
    .filterDate(start, end)
    .select('Gpp')
    .mean()
    .clip(table);

  // Compute mean value over the region
  var meanValue = yearlyGPP.reduceRegion({
    reducer: ee.Reducer.mean(),
    geometry: table,
    scale: 500,
    maxPixels: 1e10
  });

  return ee.Feature(null, {
    'year': year,
    'mean_GPP': meanValue.get('Gpp')
  });
});

// Convert to a FeatureCollection
var gppCollection = ee.FeatureCollection(annualGPP);


// Export the results as a CSV file
Export.table.toDrive({
  collection: gppCollection,
  description: 'Annual_GPP_2002_2023',
  fileNamePrefix: 'Annual_GPP_2002_2023',
  fileFormat: 'CSV'
});
```

Figure 5.12: Code Snippet to Extract Productivity Datasets

Figure 5.13 and Figure 5.14 shows the density plots comparing the differences between observed productivity and predicted productivity.

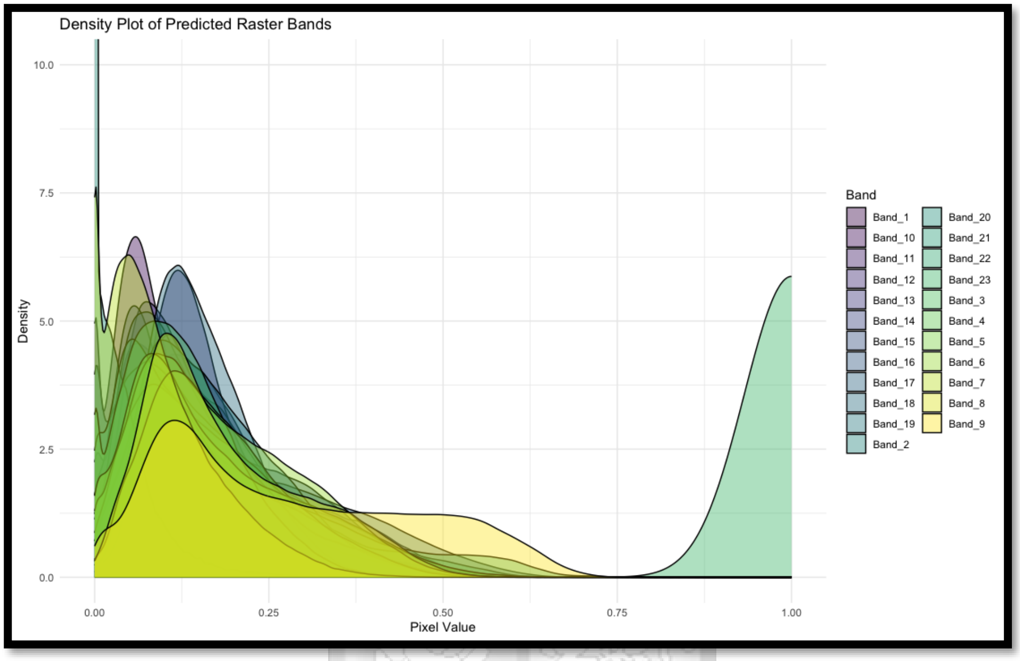

Figure 5.13: Density Plot of Predicted Productivity per Year

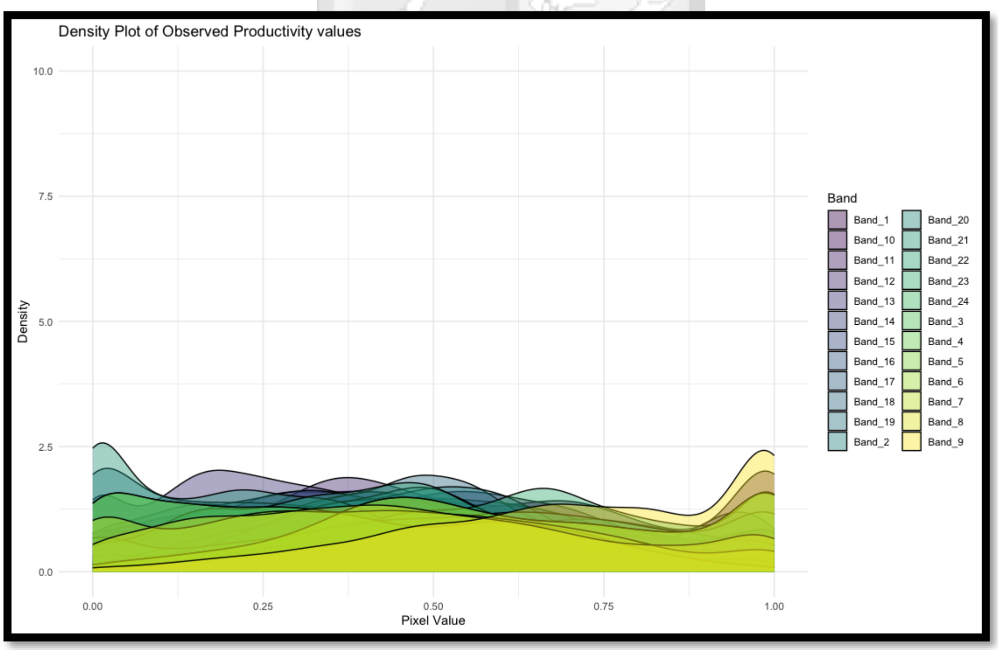

Figure 5.14: Density Plot of the Observed Productivity per Year

Although the predicted layers exhibited some outliers, the majority of the predicted pixels remained within the expected range, not exceeding three times the standard deviation. The post-processing phase also included the extraction of outlier values before forecasting. Notably, the prediction model generalized well for values below 0.75 but performed worse for very high values. This outcome suggests that the model was optimized for farmland and cultivated regions, while areas with exceptionally high productivity likely represent forested zones and other non-agricultural areas, which have very little changes across the years, limiting the forecasting ability of the prediction model.

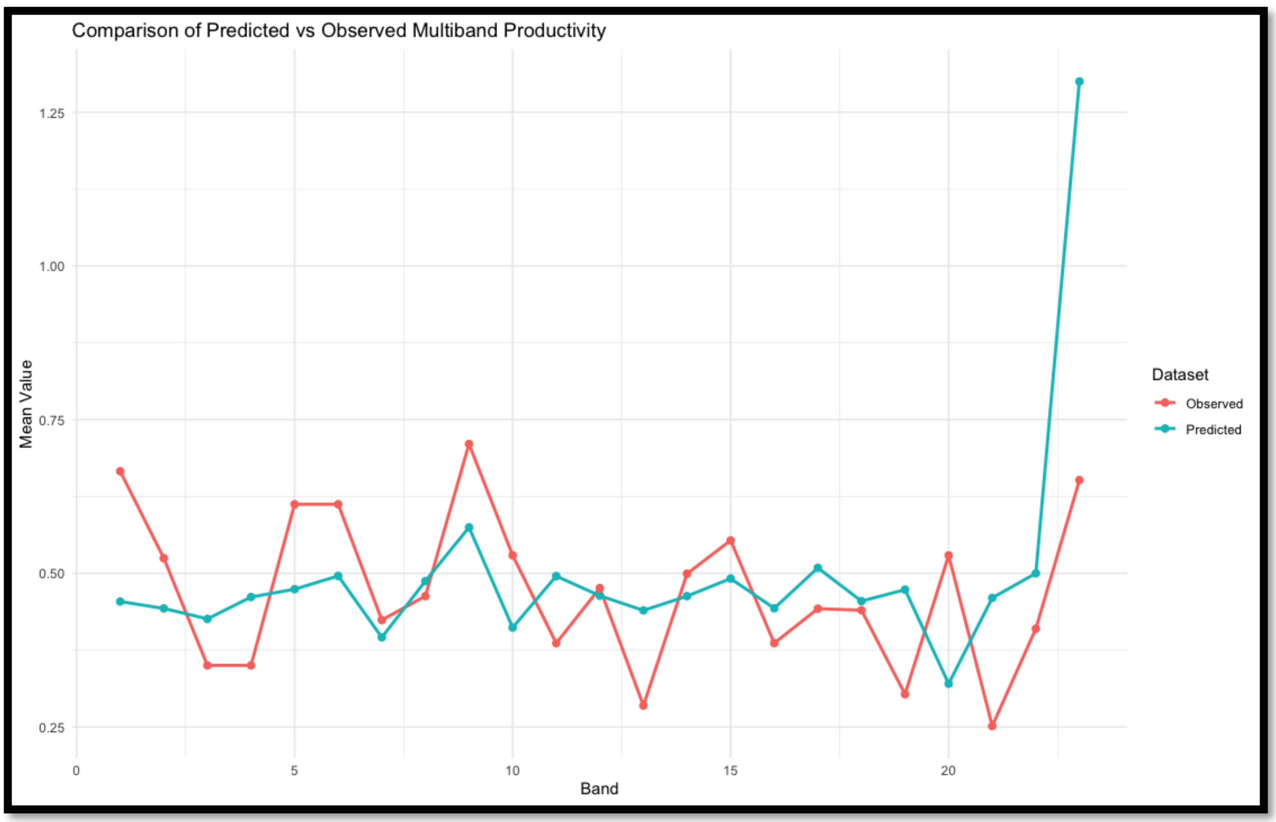

Figure 5.15: Values of Predicted versus Observed Productivity

From Figure 5.15 the predicted values closely follow the observed values, demonstrating the model's strong ability to generalize well within the standard deviation of real productivity data. The overall trend remains consistent, accurately mapping productivity variations across multiple bands. The model effectively captures the natural fluctuations present in the observed data while maintaining a stable correlation.

To forecast generate forecasted data, a polynomial regression model was implemented on predicted datapoints as indicated in section 0. Some of the outputs for forecasted model is shown in Figure 5.16.

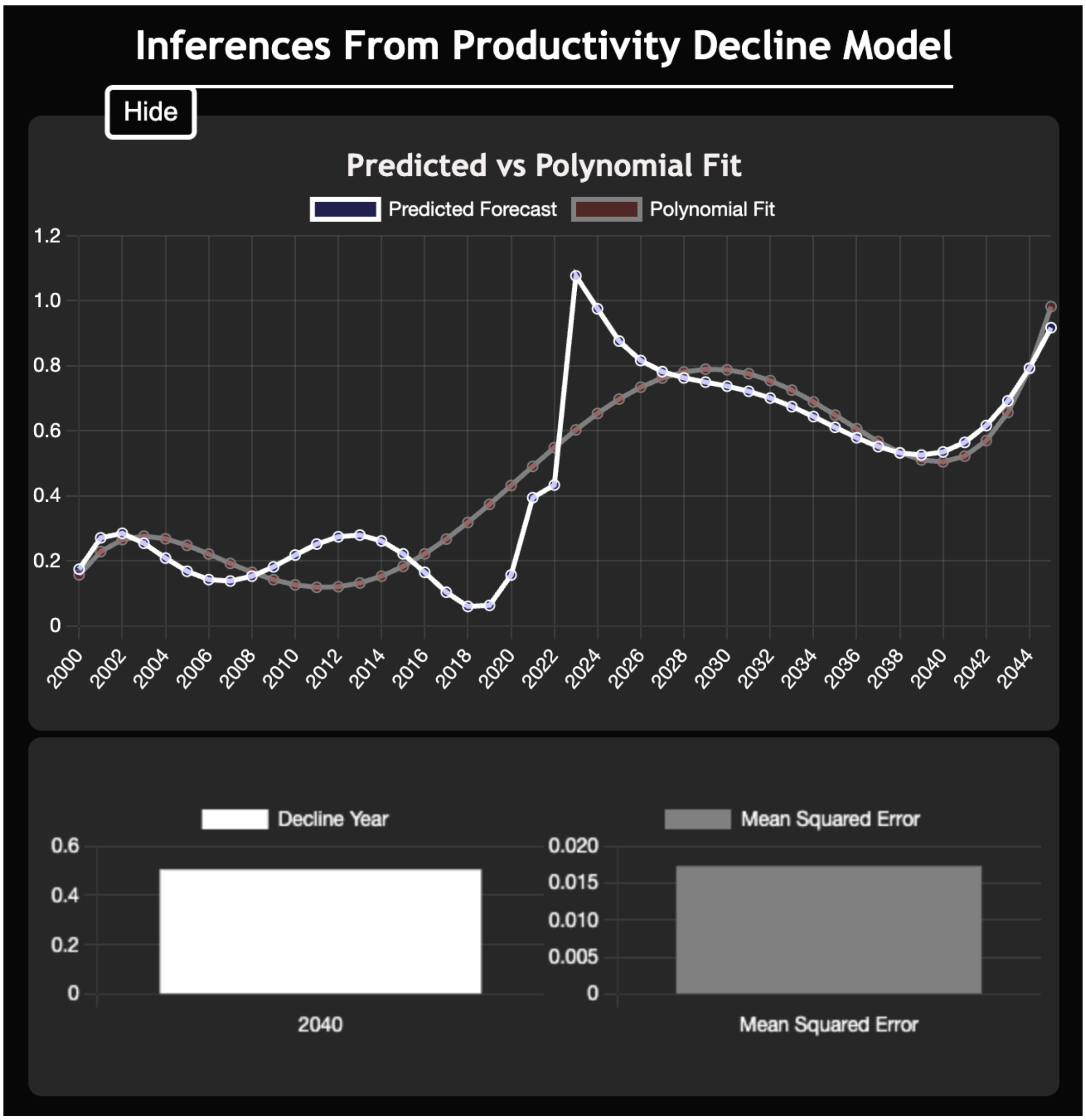

Figure 5.16: Forecasted Output and Inference Generated from User Interface

5.4 Model Evaluation

Figure 5.17 illustrate the model performance metrics after training

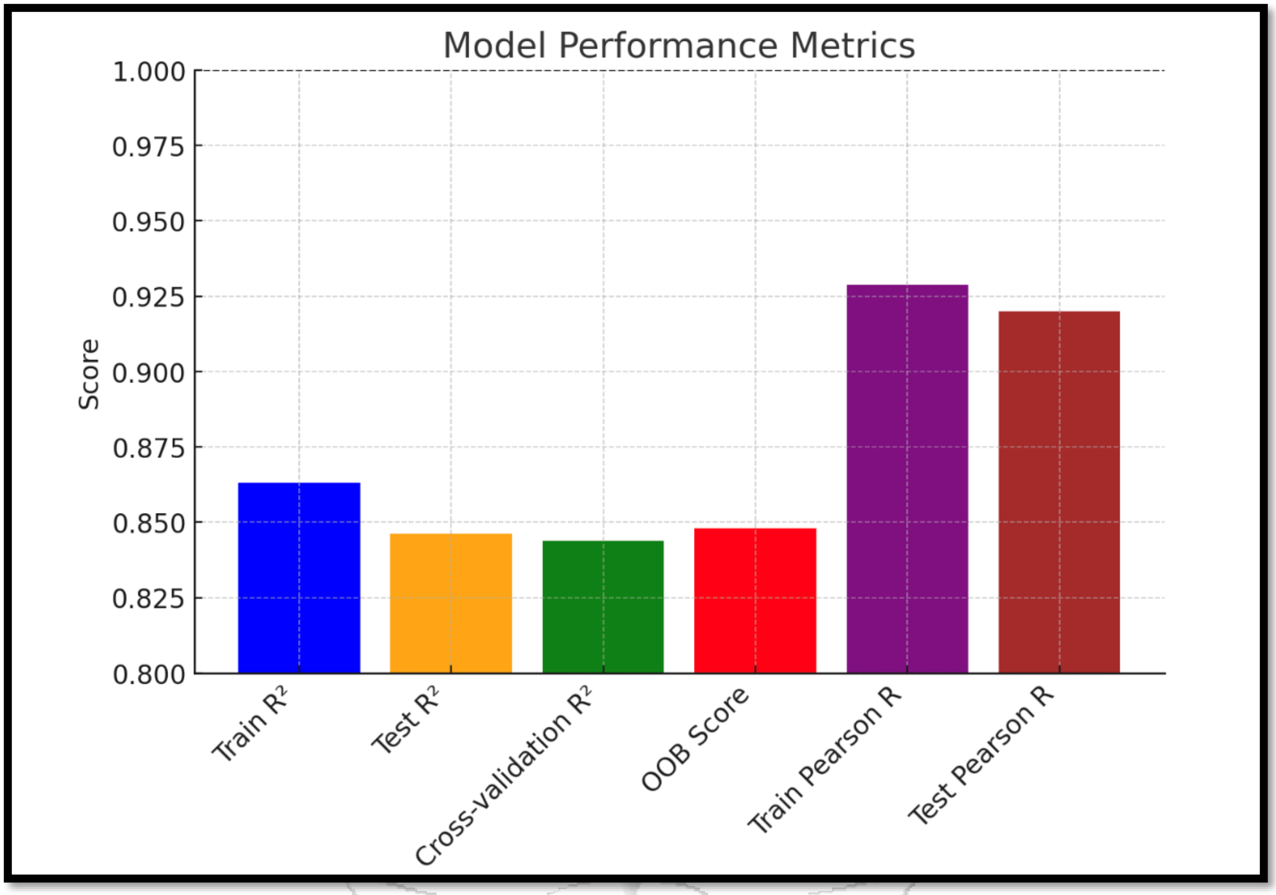

Figure 5.17: Model Performance Metrics

The bar chart presents a comprehensive comparison of the model's performance metrics. The coefficient of determination R^2 for the training dataset (0.8630) and the test dataset (0.8463) are closely aligned, indicating that the model generalizes well without signs of overfitting. Furthermore, the cross-validation R^2 (0.8440) and the out-of-bag (OOB) score (0.8479) provide additional confirmation of the model's performance (McAvaney et al., 2001) . R^2 measures how well the model's predictions align with the actual values. It quantifies the proportion of variance in the dependent variable that is explained by the independent variables while OOB error is an internal validation metric in Random Forest, estimated using samples that were not included in each bootstrap iteration for testing of the model performance. Notably, the metrics were above 0.8 (80% in terms of accuracy. The training loss function is shown in Figure 5.18.

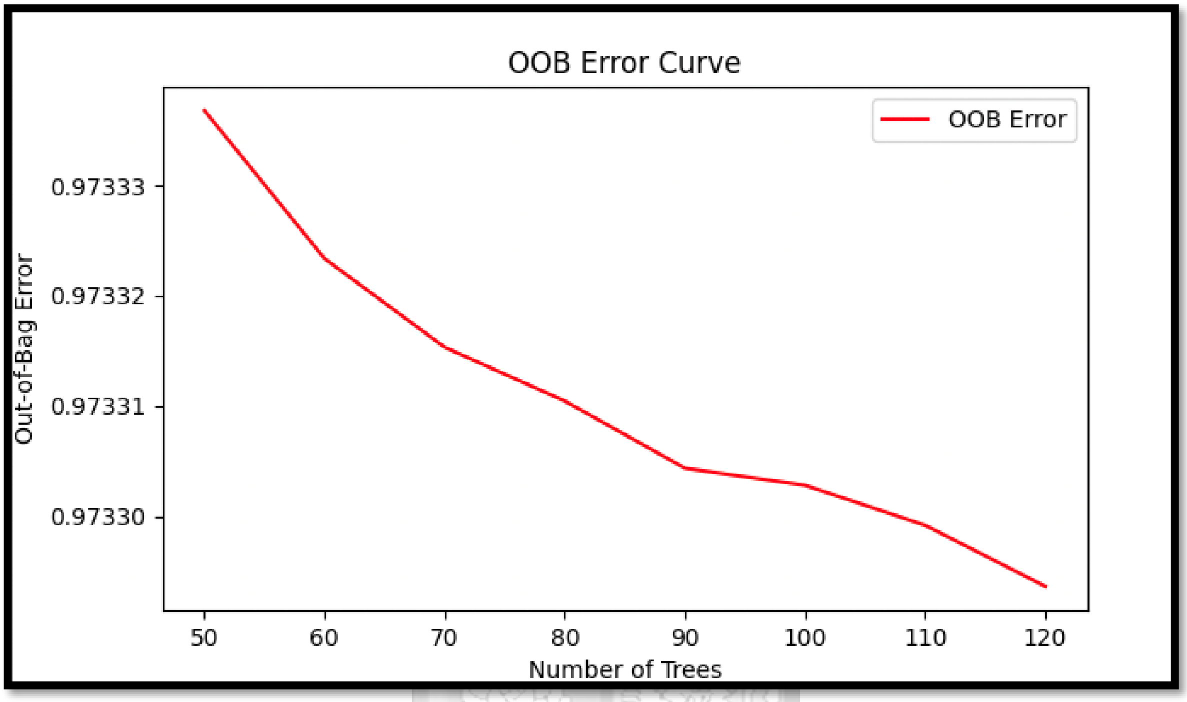

Figure 5.18: Training Loss Function for Random Forest Model Training for High Accuracy Prediction Level

5.5 Productivity Decline System Components

5.5.1 Backend

The implementation utilized Python-based frameworks, where multiple iterations of the model were trained using Python scripts. Various model types were explored, and the best-performing model, **Random Forest**, was selected for result generation and predictions.

To enable scalability and accessibility, the implementation was built using the [FAST API] framework (Peralta et al., 2024) for microservices, [pandas library], [scikit-learn], [pyspatialml]. This allows the creation of API endpoints, ensuring seamless model deployment and web-based access. Figure 5.19 shows the system implementation flow.

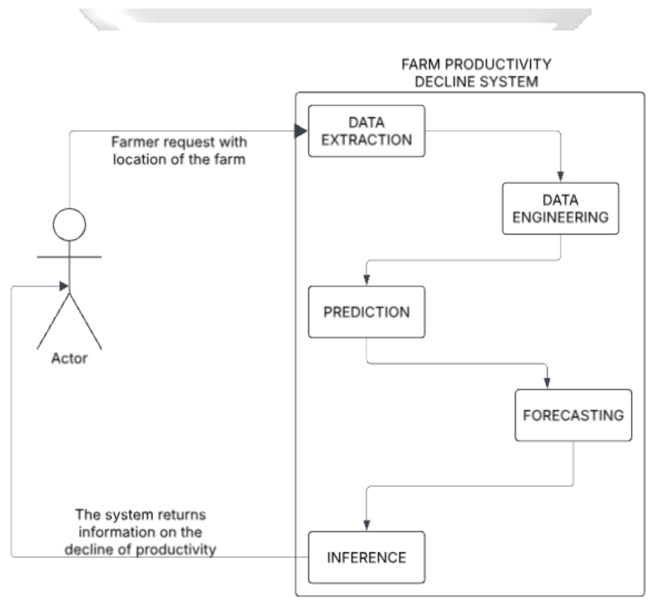

Figure 5.19: UML Flow for System Components

UML diagram in Figure 5.19 represents the Farm Productivity Decline System, which processes farm location requests from users to analyse and predict productivity decline. The Farm Productivity Decline System is designed to provide a data-driven approach to analysing and forecasting agricultural productivity trends. The system operates through a structured pipeline, beginning with user interaction, where farmers submit farm location data for analysis. Upon receiving the request, the system initiates the data extraction process, sourcing diverse datasets, including climate, soil, and satellite imagery, to ensure a comprehensive evaluation of farm productivity. These datasets undergo a data engineering phase, where they are pre-processed through techniques such as cleaning, normalization, and feature selection, ensuring data quality and relevance for subsequent analysis.

The prediction phase utilizes machine learning methodologies, with the Random Forest model demonstrating superior performance in estimating farm yield and soil health based on historical and real-time data. This is followed by the forecasting stage, where advanced time-series modelling techniques are employed to anticipate future productivity trends, enabling proactive decision-making for farmers and agricultural stakeholders. The final inference stage synthesizes these outputs into actionable insights, providing a detailed assessment of productivity trends and potential declines. By integrating predictive analytics with a scalable Fast API-based deployment framework, the system ensures seamless accessibility and adaptability, offering a reliable tool for data-driven agricultural decision-making and sustainability planning.

To facilitate seamless deployment and ensure accessibility via web-based platforms, the system was containerized using Docker, enabling a highly portable and efficient deployment process. Docker provides a consistent runtime environment, ensuring that the application runs identically across different server environments, eliminating dependency conflicts that often arise due to variations in system configurations (Grambow et al., 2018). By encapsulating the application along with its dependencies, Docker guarantees isolated and self-sufficient environments, preventing conflicts with server-specific environment variables and configurations. This approach significantly enhances scalability, as containers can be easily replicated or orchestrated using container management tools such as Kubernetes, enabling efficient load balancing and resource allocation. Furthermore, Docker's lightweight nature ensures minimal overhead, allowing for faster deployment times and optimal performance. Figure 5.20 shows sections of code that prompts docker to start running in the background, with Figure 5.21 showing the configuration file for docker, and Figure 5.22 showing docker running the productivity decline model on my local machine.

docker-compose.yml

```yaml
services:
  app:
    container_name: dissertation_app
    build:
      context: .
      dockerfile: Dockerfile
    ports:
      - "8080:8080"
    volumes:
      - .:/app
    environment:
      - PYTHONUNBUFFERED=1
    command: >
      uvicorn main:app
      --host 0.0.0.0
      --port 8080
      --reload
      --reload-dir /app
      --reload-include "*.py"
      --reload-include "*.html"
      --reload-include "*.css"
      --reload-include "*.js"
      --reload-include "*.json"
      --reload-include "*.yaml"
      --reload-include "*.yml"
    restart: unless-stopped


```

Figure 5.20: Docker Compose file for Docker

```
FROM python:3

RUN apt-get update \
    && apt-get install -y binutils libproj-dev gdal-bin python3-gdal python3-pip python3-fiona

RUN apt-get update && apt-get install -y gdal-bin \
     apt-utils libgdal-dev

RUN apt-get update && \
    apt-get install -y gcc python3-dev musl-dev \
    gdal-bin \
    python3-gdal \
    libgeos-dev \
    libffi-dev \
    proj-bin \
    libopenblas-dev \
    python3-scipy python3-numpy python3-pandas \
    netcat-traditional \
    nano

RUN apt-get update && apt-get install -y gdal-bin \
     apt-utils libgdal-dev

RUN pip install pygdal=="`gdal-config --version`.*"

RUN pip install -I GDAL=="`gdal-config --version`.*"

COPY requirements.txt requirements.txt

# RUN pip install -r requirements.txt
RUN pip install --no-cache-dir -r requirements.txt

RUN pip install pystac_client

RUN pip install odc-stac

RUN pip install rioxarray

RUN pip install geoserver-rest

RUN pip install boto3

RUN pip install rasterstats


WORKDIR /app

COPY . .

EXPOSE 8080

CMD ["uvicorn", "main:app", "--host", "0.0.0.0", "--port", "8080"]
```

Figure 5.21: Dockerfile

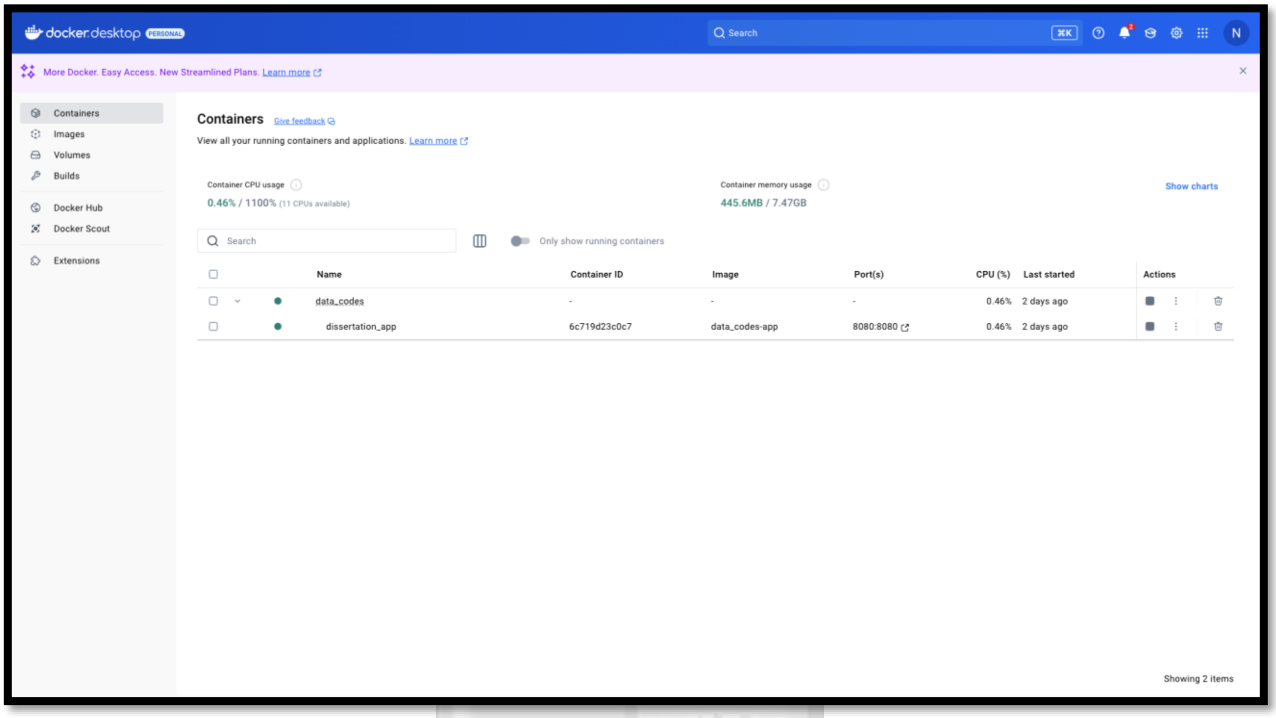

Figure 5.22: Farm productivity Decline Model Running on Docker

Figure 5.23 shows the entry point of my application before the application invokes other background processes.

```python
main.py > ...
from rf_test import random_predictor_2
load_dotenv()

app = FastAPI(
    title="MODEL FOR FORECASTING FARM PRODUCTIVITY DECLINE",
    description="API forecasts farm productivity decline",
    version="1.0.0",
    contact={
        "name": "Strathmore University",
        "email": "seth.nyawacha@strathmore.edu",
    },
)
origins = [
    "http://localhost",
    "http://localhost:8085",
    "http://localhost:8080",
    "*",
]

app.add_middleware(
    CORSMiddleware,
    allow_origins=origins,
    allow_credentials=True,
    allow_methods=["*"],
    allow_headers=["*"],
)


class Annual(BaseModel):
    geojson: dict
    farmername: str


@app.post("/fetch_data", summary="Fetch annual monthly mean bands")
async def fetch_annual(farm: Annual):
    farmername, geojson = (
        farm.farmername,
        farm.geojson,
    )

    bbox_data = process_data_bounding_box(geojson)

    print("moving to download satellite imagery")

    data_dwonload = aws_download_annual(bbox_data, farmername)

    print("moving to climate risk analysis")

    climate_risk_analysis(bbox_data, farmername)

    random_predictor_2()

    return {"message": "Data downloaded, processed and predicted successfully"}
```

Figure 5.23: Sections of the Main File as Application Entry Point

5.5.2 Front End/Client Side

The dissemination of this tool required developing a client-side interface, or front end, to share the model-generated data with platform users. This front end was developed using React, a widely adopted framework for client-side applications. React enhances performance through its use of the Virtual Document Object Model(DOM), a lightweight, in-memory representation of the actual DOM. When changes occur, React updates the Virtual DOM first and then employs a process called reconciliation to compare it with the previous version. Only the necessary changes are then applied to the real DOM, resulting in faster and more efficient rendering. This approach minimizes the performance drawbacks associated with direct DOM manipulation, as the Virtual DOM allows React to manage updates more effectively by reducing the need for direct manipulation of the real DOM, this hugely borrows from Sprague, (1980) where the framework for web development was outlined. Figure 5.24 shows the interface of the productivity decline model. Figure 5.25 shows code snippet used to plot the output on the user interface, Figure 5.26 shows section of cascading styling sheet used to give the interface graphical feel and Figure 5.27 shows section of code used to generate statistics.

Figure 5.24: User Interface Implementation for the Model

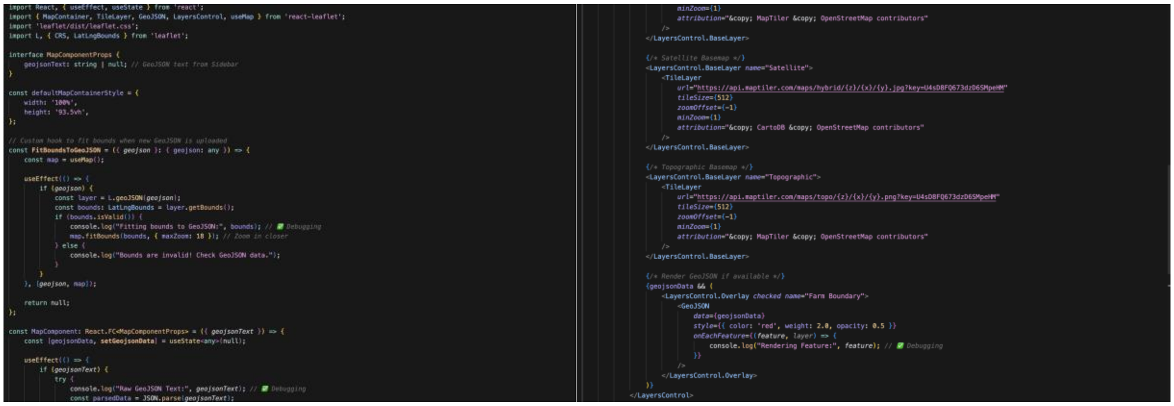

Figure 5.25: Code Snippet Boundary Plot

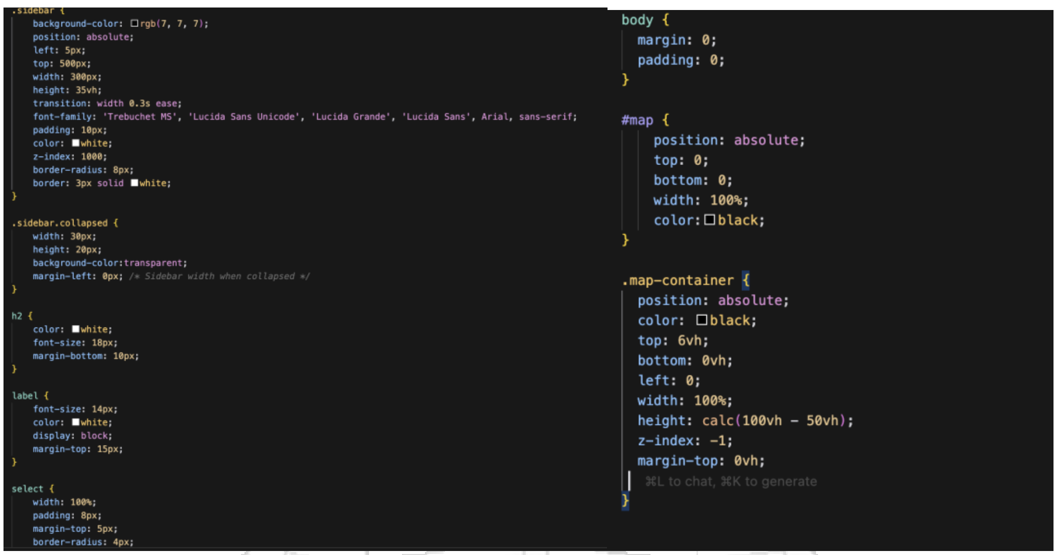

Figure 5.26: Cascading Styling Sheet To Style The Front End

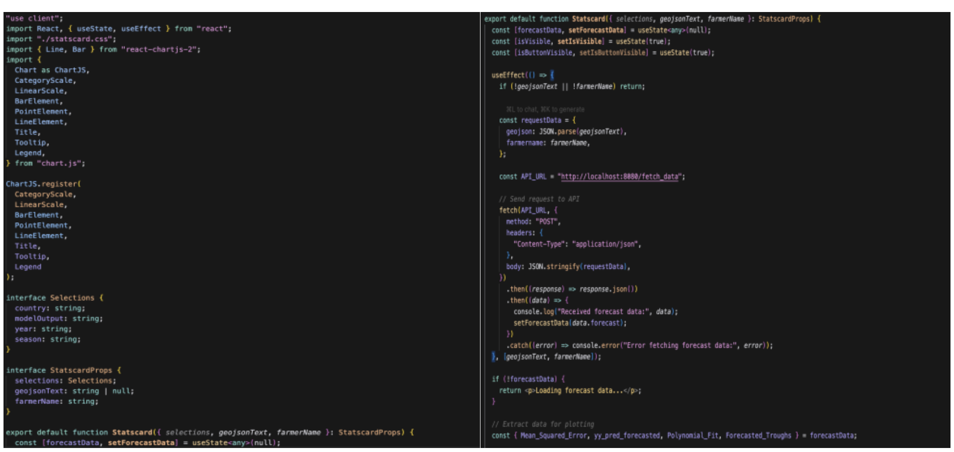

Figure 5.27: Statistics Code Snippet

5.5.3 Model Design and API access

For a user to access the system and generate the intended inferences from the trained model, the system was designed to accept farm boundary in form of *.geojson* format data type. The boundary of the farm where prediction and forecasting needs to happen, is then returned with *json* type feature through the system containing the intended details from the backend including

- The prediction trend from 2000 to 2024
- The forecasted trend using polynomial regression model
- The associated metrics including the mean square error and the forecasted decline date as shown in Figure 5.29. Figure 5.28 shows the request structure made to productivity decline API.


```
{
    "geojson": {
    "type": "FeatureCollection",
    "features": [
      {
        "type": "Feature",
        "properties": {},
        "geometry": {
          "coordinates": [
            [
              [
                34.93999828610123,
                -0.2471347034667417
              ],
              [
                34.93999828610123,
                -0.2488224389046536
              ],
              [
                34.941222369421695,
                -0.2488224389046536
              ],
              [
                34.941222369421695,
                -0.2471347034667417
              ],
              [
                34.93999828610123,
                -0.2471347034667417
              ]
            ]
          ],
          "type": "Polygon"
        }
      }
    ]
  },
    "farmername": "string"
}
```


Figure 5.28: Payload Structure for the Model

```json
{
    "forecast": {
      "Mean_Squared_Error": 0.024682,
      "yy_pred_forecasted": [
        0.07843017578125,
        1.1435546875,
        1.1463623046875
      ],
      "Polynomial_Fit": {
        "2000": 0.1156,
        "2001": 0.2031,
        "2002": 0.2625,
        "2043": 0.8624,
        "2044": 1.0581,
        "2045": 1.3557
      },
      "Forecasted_Troughs": {
        "2040": 0.6841
      }
    },
    "geojson": {
      "type": "FeatureCollection",
      "features": [
        {
          "type": "Feature",
          "properties": {},
          "geometry": {
            "coordinates": [
              [
                [
                  34.93999828610123,
                  -0.2471347034667417
                ],
                [
                  34.93999828610123,
                  -0.2488224389046536
                ],
                [
                  34.941222369421695,
                  -0.2488224389046536
                ],
                [
                  34.941222369421695,
                  -0.2471347034667417
                ],
                [
```


Figure 5.29: Response from the Model’s Application Processing Interface (API)

The response is them plotted on the front end in different components to showcase the information as a graphical output for easier readability to the user as indicated in Figure 5.30.

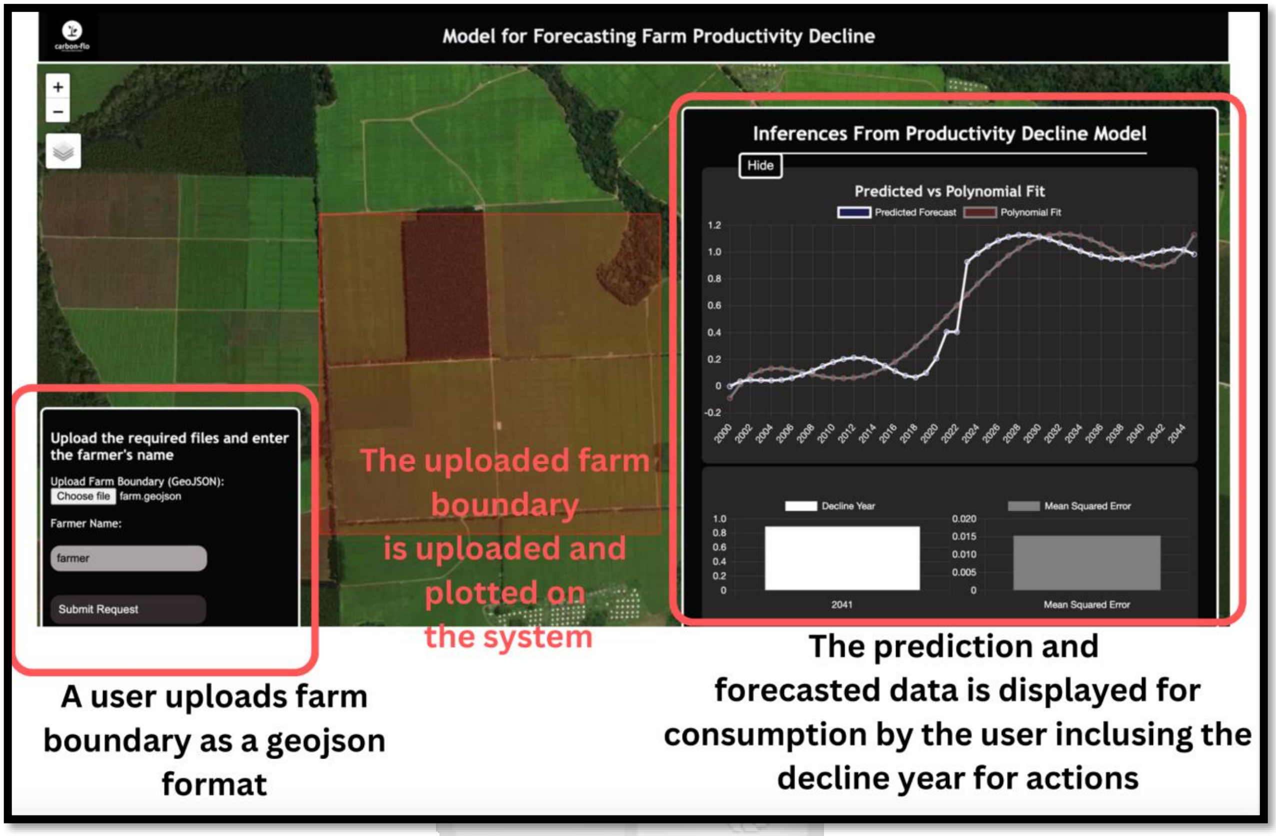

Figure 5.30: Farm Productivity System Components

Chapter 6: Discussions

6.1 Achievements of Research Objectives

This section outlines how the study captured the objectives as outlined in the study, which focused on leveraging machine learning models to predict and forecast farm productivity.

6.1.1 Evaluate Existing Models and Methodologies

A comprehensive study was conducted to assess the existing methodologies employed in forecasting and predicting farm productivity. This evaluation involved analysing various models, techniques, and frameworks used in agricultural forecasting, considering their accuracy, reliability, and applicability to different farming conditions. The study also explored the integration of machine learning algorithms, statistical models, and remote sensing data in enhancing prediction accuracy (Mark, 2006). Additionally, the strengths and limitations of these methodologies were examined to identify gaps that could be addressed in the proposed model. The detailed findings and insights from this assessment are thoroughly outlined in Chapter 2:, providing a foundation for selecting the most suitable approach for developing a robust farm productivity decline prediction system.

6.1.2 Compile Input Covariates for Farm Productivity Decline Model

The input covariates were carefully selected and compiled to develop the farm productivity prediction model, as outlined in the Chapter 3:. These covariates were chosen based on their direct influence on agricultural productivity and their ability to provide meaningful insights into crop performance over time. The key input variables included daily rainfall patterns, maximum and minimum temperatures, soil moisture components, and satellite imagery for productivity modelling as outlined in Chapter 3:

Daily rainfall data was incorporated to assess the availability of water, which is crucial for plant growth, while temperature extremes were analysed to determine their impact on crop stress and overall yield. Soil moisture components played a vital role in evaluating soil water retention capacity and its effect on plant health. Additionally, satellite imagery provided remote sensing data, enabling the monitoring of vegetation health and identifying patterns related to crop growth. These covariates were used to build the dependent variable productivity, ensuring that the model could effectively capture the relationships between environmental conditions and farm output. By integrating these diverse data sources, the prediction model aimed to enhance the accuracy and reliability of farm productivity forecasts.

6.1.3 Develop a Forecasting Land Productivity Decline Model

Several models were evaluated to determine their accuracy in predicting farm productivity, including Random Forest, Linear Regression, Polynomial Regression, Neural Networks, and XGBoost. Each model was tested to assess its performance, and the results were compared to identify the most reliable approach. Among the tested models, Random Forest emerged as the best-performing model, demonstrating higher accuracy than the others. Due to its ability to handle complex relationships and minimize overfitting and a lower mean square error (Nhambiu & Chichango, 2024), it was selected as the model of choice for the final implementation. A detailed analysis of the model performance and selection process is provided in Chapter 5:

6.1.4 Assess the Accuracy of Land Productivity Decline Model

The model's accuracy was evaluated using Mean Squared Error (MSE), as detailed in Chapter 5:. Additionally, the platform provides a real-time preview of the MSE for each user request, allowing for an assessment of prediction reliability. A lower MSE indicates a higher prediction accuracy, reinforcing the system's capability to generate precise farm productivity forecasts.

Chapter 7: Conclusions and Recommendations

7.1 Conclusions

The farm productivity decline model was developed by integrating key environmental and climatic covariates into two predictive models: a Random Forest model for prediction and a Polynomial Regression model for forecasting. The Random Forest model was utilized to predict farm productivity based on historical and real-time data, leveraging its robustness in handling complex relationships among variables. Meanwhile, the Polynomial Regression model was implemented to provide a forecasting capability, enabling users to anticipate potential declines in productivity over time.

This model was embedded into a Decision Support System (DSS)**,** designed to deliver actionable insights on farm productivity trends (Sprague, 1980), (Liu et al., 2009). The system identifies potential periods of productivity decline, allowing users to take proactive measures. Additionally, it presents both fitted and predicted values for a specific farm, offering a comparative view of past performance and expected outcomes.

To ensure transparency and reliability, the platform also provides an error rate associated with each prediction. This error rate serves as an indicator of model confidence, helping users assess the reliability of the insights before making any decision. The integration of these predictive and forecasting models within the DSS enhances farm management strategies, enabling informed decision-making to mitigate productivity decline.

7.2 Recommendations

For a more realistic and accurate forecasting, future research should explore the use of more advanced machine learning models and incorporate higher-resolution temporal datasets, particularly for the dependent variable (farm productivity). Enhancing the model with more granular data can improve its predictive capability, making it more responsive to subtle changes in environmental and climatic conditions (Arkabaev et al., 2025).

This research holds significant potential as a pioneering study in the industry. Its insights can be leveraged not only by farmers to optimize agricultural planning but also by banks, financial institutions, and lenders seeking to assess farm productivity risks before issuing loans or financial support. By providing a data-driven approach to monitoring farm productivity decline, the model can contribute to better risk management, improved decision-making, and enhanced sustainability in the agricultural sector

7.3 Future Works

As outlined in section 7.2 Recommendations, this research could benefit from refinement of the dependent variable, and exploring more robust models that can be used for forecasting in the near real time that could even span within seasons.

References

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Abeeb Babatunde Omotoso, Letsoalo, S., Kehinde Oluseyi Olagunju, Christopher Seleke Tshwene, & Abiodun Olusola Omotayo. (2023). Climate change and variability in sub-Saharan Africa: A systematic review of trends and impacts on agriculture. *Journal of Cleaner Production*, *414*, 137487–137487. https://doi.org/10.1016/j.jclepro.2023.137487

Aggarwal, P. K., Kalra, N., Chander, S., & Pathak, H. (2005). InfoCrop: A dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. I. Model description. *Agricultural Systems*, *89*(1), 1–25. https://doi.org/10.1016/j.agsy.2005.08.001

Aldian Nurcahyo, Yaya Heryadi, Lukas, N., Wayan Suparta, & Ilvico Sonata. (2023). Interpretable Machine Learning for Multi-Class Crop Yield Prediction. https://doi.org/10.1109/icicyta60173.2023.10428914.

Archontoulis, S. V., Miguez, F. E., & Moore, K. J. (2014). Evaluating APSIM Maize, Soil Water, Soil Nitrogen, Manure, and Soil Temperature Modules in the Midwestern United States. *Agronomy Journal*, *106*(3), 1025–1040. https://doi.org/10.2134/agronj2013.0421

Arkabaev, N., Rahimov, E., Abdullaev, A., Padmanaban, H., & Salmanov, V. (2025). Modelling and analysis of optimization algorithms. Jurnal Ilmiah Ilmu Terapan Universitas Jambi, 9(1), 161-177.

Batool, D., Shahbaz, M., Hafiz Shahzad Asif, Shaukat, K., Talha Mahboob Alam, Hameed, I. A., Ramzan, Z., Waheed, A., Hanan Aljuaid, & Luo, S. (2022). A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. *Plants*, *11*(15), 1925–1925. https://doi.org/10.3390/plants11151925.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

Benami, E., Jin, Z., Carter, M. R., Ghosh, A., Hijmans, R. J., Hobbs, A., Benson Kenduiywo, & Lobell, D. B. (2021). Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews Earth & Environment*, *2*(2), 140–159. https://doi.org/10.1038/s43017-020-00122-y

Branca, G., Arslan, A., Paolantonio, A., Uwe Grewer, Cattaneo, A., Romina Cavatassi, Lipper, L., Hillier, J., & Vetter, S. (2021). Assessing the economic and mitigation benefits of climate-smart agriculture and its implications for political economy: A case study in Southern Africa. Journal of Cleaner Production, 285, 125161–125161. https://doi.org/10.1016/j.jclepro.2020.125161

C. Folberth, A. Baklanov, J. Balkovič, R. Skalský, N. Khabarov, & M. Obersteiner. (2019). Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. *Agricultural and Forest Meteorology*, *264*, 1–15. https://doi.org/10.1016/j.agrformet.2018.09.021

Čobanović, K., Lozanov-Crvenković, Z., & Nikolić-Đorić, E. (n.d.). *ICOTS-7, 2006: Cobanovic, Lozanov-Crvenkovic and Nikolic-Doric (Refereed) 1 PERIODIC REGRESSION*. Retrieved September 9, 2024, from https://iase-web.org/documents/papers/icots7/C207.pdf

Daou, R., Joubrane, K., Maroun, R. G., Lydia Rabbaa Khabbaz, Ismail, A., & André El Khoury. (2021). Mycotoxins: Factors influencing production and control strategies. AIMS Agriculture and Food, 6(1), 416–447. https://doi.org/10.3934/agrfood.2021025.

Dasgupta, S., Nicole van Maanen, Gosling, S. N., Piontek, F., Otto, C., & Carl-Friedrich Schleussner. (2021). Effects of climate change on combined labour productivity and supply: an empirical, multi-model study. *The Lancet Planetary Health*, *5*(7), e455–e465. https://doi.org/10.1016/s2542-5196(21)00170-4

Edwards, S., Johnson, S., & Block, S. (2009). Title: The Decline and Rise of Agricultural Productivity in Sub-Saharan Africa since 1961. Publisher. https://www.nber.org/system/files/chapters/c13435/c13435.pdf

Edwards, S., Johnson, S., & Block, S. (2009). Title: The Decline and Rise of Agricultural Productivity in Sub-Saharan Africa since 1961. Publisher. https://www.nber.org/system/files/chapters/c13435/c13435.pdf

Encyclopedia of Agriculture and Food Systems. (2014). ScienceDirect. https://www.sciencedirect.com/referencework/9780080931395/encyclopedia-of-agriculture-and-food-systems

Fauzan, R., Hamidi, M. Z., Safitri, W. A., Siahaan, D. O., & Karimi, M. I. (2025). Ontology in Requirements Software Development Method: A Systematic Literature Review. *Journal of Information Technology and Cyber Security*

Feleke, H. G., Savage, M. J., & Tesfaye, K. (2021). Calibration and validation of APSIM–Maize, DSSAT CERES–Maize and AquaCrop models for Ethiopian tropical environments. South African Journal of Plant and Soil, 38(1), 36-51.

Florent Baarsch, Granadillos, J. R., Hare, W., Knaus, M., Krapp, M., Schaeffer, M., & Hermann Lotze-Campen. (2020). The impact of climate change on incomes and convergence in Africa. *World Development*

Fuglie, K., Gautam, M., Goyal, A., & Maloney, W. F. (2019). Harvesting prosperity: Technology and productivity growth in agriculture. World Bank Publications.

Gohain, G. B., Singh, K. K., Singh, R. S., Dakhore, K. K., & Ghosh, K. (2021). Application of CERES-sorghum crop simulation model DSSAT v4.7 for determining crop water stress in crop phenological stages. *Modeling Earth Systems and Environment*, *8*(2), 1963–1975. https://doi.org/10.1007/s40808-021-01194-5.

Grambow, M., Hasenburg, J., Pfandzelter, T., & Bermbach, D. (2018). *Dockerization Impacts in Database Performance Benchmarking*. ArXiv.org. https://arxiv.org/abs/1812.04362

Heng, L. K., Hsiao, T., Evett, S., Howell, T., & Steduto, P. (2009). Validating the FAO AquaCrop Model for Irrigated and Water Deficient Field Maize. *Agronomy Journal*, *101*(3), 488–498. https://doi.org/10.2134/agronj2008.0029xs

Hengl, T., Miller, M.A.E., Križan, J., et al. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci Rep 11, 6130 (2021). doi:10.1038/s41598-021-85639-y

Jägermeyr, J., Christoph Müller, Ruane, A. C., Elliott, J., Juraj Balkovic, Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Toshichika Iizumi, Jain, A. K., Kelly, D., Nikolay Khabarov, Lange, S., & Lin, T.-S. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food*, *2*(11), 873–885. https://doi.org/10.1038/s43016-021-00400-y

João Vasco Silva, & Giller, K. E. (2020). Grand challenges for the 21st century: what crop models can and can't (yet) do. *The Journal of Agricultural Science*, *158*(10), 794–805. https://doi.org/10.1017/s0021859621000150

Jones, E. J., Thomas F.A. Bishop, Malone, B. P., Hulme, P. J., Whelan, B. M., & Filippi, P. (2022). Identifying causes of crop yield variability with interpretive machine learning. Computers and Electronics in Agriculture, 192, 106632–106632. https://doi.org/10.1016/j.compag.2021.106632

José Luis Noriega-Navarrete, Salazar-Moreno, R., & Irineo Lorenzo López-Cruz. (2021). Revisión: modelos de crecimiento y rendimiento de maíz en escenarios de cambio climático. *Revista Mexicana de Ciencias Agrícolas*, *12*(1), 127–140. https://doi.org/10.29312/remexca.v12i1.2552

Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. Agriculture & Food Security, 10, 1-24.

Klawonn, F., Kruse, R., & Winkler, R. (2015). Fuzzy clustering: More than just fuzzification. Fuzzy sets and systems, 281, 272-279.

Lee, S.-J., Kim, N., & Lee, Y. (2021). Development of Integrated Crop Drought Index by Combining Rainfall, Land Surface Temperature, Evapotranspiration, Soil Moisture, and Vegetation Index for Agricultural Drought Monitoring. Remote Sensing, 13(9), 1778–1778. https://doi.org/10.3390/rs13091778

Leng, G., & Hall, J. W. (2020). Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models. *Environmental Research Letters*, *15*(4), 044027–044027. https://doi.org/10.1088/1748-9326/ab7b24

Leng, G., & Hall, J. W. (2020). Predicting spatial and temporal variability in crop yields: aninter comparison of machine learning, regression and process-based models. Environmental Research Letters, 15(4), 044027–044027. https://doi.org/10.1088/1748-9326/ab7b24

Liliane, T. N., & Charles, M. S. (2020). Factors affecting yield of crops. Agronomy-climate change & food security, 9.

Liu, S., Alex, Whitfield, R. I., & Boyle, I. M. (2009). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261–286. https://doi.org/10.1007/s10115-009-0192-4

Liu, Y., Zhang, H., Li, G., Sun, X., & Wang, M. (2022). A comprehensive method to increase yield and narrow the yield gap of winter wheat for sustainable intensification. *Journal of the Science of Food and Agriculture*, *102*(10), 4238–4249. https://doi.org/10.1002/jsfa.11775

Mansour, H., Abd El-Hady, M., Eldardiry, E., Saad, S., Abd Elmabod, S., Aziz, A., Essa, S., & Eldwainy, C. (2020). SUSTAINABLE AGRICULTURE AND FOOD -CHALLENGES AND SOLUTIONS: A REVIEW. 20, 3218–3223. https://www.plantarchives.org/SPECIAL%20ISSUE%2020-1/215__3218-3223_.pdf

Mark. (2006). Statistical Inference for Variable Importance. *The International Journal of Biostatistics*, *2*(1). https://doi.org/10.2202/1557-4679.1008

McAvaney, B. J., Covey, C., Joussaume, S., Kattsov, V., Kitoh, A., Ogana, W., ... & Zhao, Z. C. (2001). Model evaluation. In Climate Change 2001: The scientific basis. Contribution of WG1 to the Third Assessment Report of the IPCC (TAR) (pp. 471-523). Cambridge University Press.

Mearns, L. O., Mavromatis, T., E. Tsvetsinskaya, Hays, C., & Easterling, W. (1999). Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. *Journal of Geophysical Research Atmospheres*, *104*(D6), 6623–6646. https://doi.org/10.1029/1998jd200061

Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: a review paper. *Indian J. Sci. Technol*, *9*(38), 1-14.

Nhambiu, J., & Chichango, F. (2024). Forecasting energy consumption in Mozambique: A comparative analysis of advanced machine learning models from 2025 to 2045. *Research Society and Development*, *13*(9), e3613946830–e3613946830. https://doi.org/10.33448/rsd-v13i9.46830

Peralta, J. H. (2023). Microservice APIs: Using Python, Flask, FastAPI, OpenAPI and More. Simon and Schuster.

Priscilla Ntuchu Kephe, Kingsley Kwabena Ayisi, & Brilliant Mareme Petja. (2021). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. Agriculture & Food Security, 10(1). https://doi.org/10.1186/s40066-020-00283-5.

Rainford, S., Martín-López, J. M., & Mayesse Da Silva. (2021). Approximating Soil Organic Carbon Stock in the Eastern Plains of Colombia. *Frontiers in Environmental Science*, *9*. https://doi.org/10.3389/fenvs.2021.685819.

Sprague, R. H. (1980). A Framework for the Development of Decision Support Systems. *MIS Quarterly*, *4*(4), 1–1. https://doi.org/10.2307/248957

Sprague, R. H. (1980). A Framework for the Development of Decision Support Systems. *MIS Quarterly*, *4*(4), 1–1. https://doi.org/10.2307/248957

Stockle, C. O., Martin, S. A., & Campbell, G. S. (1994). CropSyst, a cropping systems simulation model: Water/nitrogen budgets and crop yield. *Agricultural Systems*, *46*(3), 335–359. https://doi.org/10.1016/0308-521x(94)90006-2

Suárez-Figueroa, M. C., & Asunción Gómez-Pérez. (2011). Ontology Requirements Specification. *Springer EBooks*, 93–106. https://doi.org/10.1007/978-3-642-24794-1_5

Sun, B., Liu, H., Zhou, S., & Li, W. (2014). Evaluating the Performance of Polynomial Regression Method with Different Parameters during Color Characterization. Mathematical Problems in Engineering, 2014, 1–7. https://doi.org/10.1155/2014/418651.

Teja Kattenborn, Jens Leitloff, Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Wang, L., Zhou, Y., Li, Q., Xu, T., Wu, Z., & Liu, J. (2021). Application of Three Deep Machine-Learning Algorithms in a Construction Assessment Model of Farmland Quality at the County Scale: Case Study of Xiangzhou, Hubei Province, China. Agriculture, 11(1), 72–72. https://doi.org/10.3390/agriculture11010072.

Wolanin, A., Gustau Camps-Valls, Gómez-Chova, L., Mateo-García, G., van, Zhang, Y., & Guanter, L. (2019). Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. *Remote Sensing of Environment*, *225*, 441–457. https://doi.org/10.1016/j.rse.2019.03.002

Yang, P., Zhao, Q., & Cai, X. (2020). Machine learning based estimation of land productivity in the contiguous US using biophysical predictors. Environmental Research Letters, 15(7), 074013–074013. https://doi.org/10.1088/1748-9326/ab865f.

Appendices

Appendix A: Similarity Report

A Model for Forecasting Farm Productivity Decline.docx

ORIGINALITY REPORT

18%	18%	15%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ouci.dntb.gov.ua Internet Source	2%
2	ebooks.inflibnet.ac.in Internet Source	1%
3	Submitted to Strathmore University Student Paper	1%
4	su-plus.strathmore.edu Internet Source	1%
5	iopscience.iop.org Internet Source	1%
6	Submitted to La Trobe University Student Paper	1%
7	researchspace.ukzn.ac.za Internet Source	<1%
8	krishikosh.egranth.ac.in Internet Source	<1%
9	d-nb.info Internet Source	<1%
10	journals.ametsoc.org Internet Source	<1%
11	www.mdpi.com Internet Source	<1%
12	fastercapital.com Internet Source	<1%

Appendix B: Ethical Review Approval

Strathmore
UNIVERSITY

4th December 2024

Mr Nyawacha Seth,
seth.nyawacha@strathmore.edu

Dear Mr Nyawacha,

RE: A Model for Forecasting Farm Productivity Decline

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2435/24.** The approval period is from **4th December 2024 to 3rd December 2025.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu