AGRICARE

A report on Project Phase – II submitted

in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology
in
Computer Science and Engineering

by

Rahul Shah
Registration No.: 2020105251

Satyam Raj
Registration No.: 2020105253

Under the supervision of
Dr. Shouvik Dey
Associate Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY NAGALAND

MAY 2024

राष्ट्रीय प्रौद्योगिकी संस्थान नागालैंड
NATIONAL INSTITUTE OF TECHNOLOGY NAGALAND
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
Chumukedima, Dimapur – 797 103, Nagaland

BONAFIDE CERTIFICATE

Certified that this Project titled **"AGRICARE"** is the bonafide work of **Rahul Shah (2020105251) & Satyam Raj (2020105253)** who carried out the work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other student.

Dr. ARAMBAM NEELIMA
Project Coordinator
Assistant Professor
Department of Computer
Science & Engineering
NIT Nagaland
Dimapur – 797 103

Dr. SHOUVIK DEY
Supervisor
Associate Professor
Department of Computer
Science & Engineering
NIT Nagaland
Dimapur – 797 103

Dr. LITHUNGO MURRY
Professor & Head of Department
Assistant Professor
Department of Computer
Science & Engineering
NIT Nagaland
Dimapur – 797 103

Dr. H. ROHEN SINGH
External Examiner
Assistant Professor
Department of IT School of
Engineering & Technology
Nagaland University
Dimapur – 797 103

ABSTRACT

Agricare, an advanced agricultural software, revolutionizes precision farming through its three core components: a sophisticated Crop Recommendation Model and an efficient Crop Water Model, and a Crop Yield Prediction Model by state-of-the-art machine learning models.

The Crop Recommendation Model is a standout feature, utilizing machine learning models. It integrates soil fertility data and weather forecasts for immediate and contextually relevant crop suggestions. This approach ensures comprehensive and dynamic crop recommendations, considering soil chemical compositions and weather conditions. The Crop Water Model calculates optimal irrigation amounts at various stages of the crop based on crop type, weather conditions, and soil characteristics, promoting water conservation and efficient resource utilization.

Complementing this is the Crop Yield Prediction Model, which estimates crop yields based on various factors such as crop type, seasonal conditions, and agricultural inputs. The model's predictive insights empower farmers to make informed decisions about crop management practices and optimize agricultural productivity.

Agricare's development relies on a robust foundation of materials and libraries for scalability and maintainability. The framework and libraries contribute diverse features, and the design and implementation phases showcase the software's effectiveness in providing accurate crop recommendations, irrigation guidance, and yield predictions.

In conclusion, Agricare is a comprehensive precision farming solution, empowering farmers with data-driven insights for optimizing practices and promoting sustainable and efficient farming. The report suggests potential avenues for future research and development in the field.

ACKNOWLEDGEMENT

We wish to place on record our deep sense of gratitude to our honorific Guide **Dr. Shouvik Dey**, Designation, Institute for his / her supervision, valuable guidance and moral support leading to the successful completion of the work. Without his / her continuous encouragement and involvement, this project would not have been a reality.

We would like to extent my sincere thanks to **Dr J Arul Valan** for their valuable suggestions. We wish to thank **Dr. Lithungo Murry**, Department of Computer Science and Engineering, NIT Nagaland for continuous support. We would also like to thank all our friends who have developed us to gain a sense of dutifulness, perfection and sincerity in the effort (if Desired).

We wish to dedicate this work to **our parents**, for they are the pillars of support giving us confidence in whatever we do. We would like to thank **Dr. Shouvik Dey** who has motivated us to work harder and do our best. Last but not least, we would like to owe our sincere and incessant gratitude to the almighty God for the immense blessing on us.

RAHUL SHAH

(2020105251)

&

SATYAM RAJ

(2020105253)

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	III
	LIST OF TABLES	VIII
	LIST OF FIGURES	IX
	LIST OF ABBREVIATIONS	X
1	**INTRODUCTION**	
	1.1 GENERAL	1
	1.2 STATE OF THE ART	2
	1.3 OBJECTIVES	3
	1.4 ORGANIZATION OF THE REPORT	4
2	**AGRICARE A SOFTEARE SOLUTION**	**5**
	2.1 INTRODUCTION	5
	2.2 TECHNOLOGY USED	
	2.2.1 PYTHON	6
	2.2.2 ML ALGORITHMS FOR CLASSIFICATION	6
	2.2.3 ML ALGORITHMS FOR REGRESSION	6
	2.2.4 PYTHON LIBRARIES	7
	2.2.5 TECHNOLOGY STACK FOR WEB APP	8
	2.3 MODELS OVERVIEW	
	2.3.1 CROP RECOMMENDATION MODEL	8
	2.3.2 CROP WATER MODEL	9
	2.3.4 CROP YIELD PREDICTION MODEL	9
	2.4 INTEGRATING ML MODEL FOR WEB APP	10
	2.5 CONCLUSION	10

CHAPTER NO.	TITLE	PAGE NO.

3	**PREVIOUS WORKS**		**11**
	3.1	INTRODUCTION	11
	3.2	PROPOSED DESIGN	11
	3.3	IMPLEMENTATION	12
		3.3.1 DATA COLLECTION	12
		3.3.2 DATA PREPROCESSING	13
		3.3.3 CROP RECOMMENDATION MODEL	13
		3.3.4 CROP WATER MODEL	14
	3.4	RESULT	16
		3.4.1 CROP RECOMMENDATION MODEL	16
		3.4.2 CROP WATER MODEL	16
	3.5	CONCLUSION	16
4	**DESIGN AND IMPLEMENTATION**		**17**
	4.1	INTRODUCTION	17
	4.2	DESIGN	17
	4.3	IMPLEMENTATION	18
		4.3.1 DATA PREPROCESSING	18
		4.3.2 CROP YIELD PREDICTION MODEL	19
	4.4	CONCLUSION	21
5	**RESULTS AND DISCUSSION**		**20**
	5.1	INTRODUCTION	22
	5.2	CROP RECOMMENDATION MODEL	21
	5.4	CROP WATER MODEL	25
	5.5	CROP YIELD PREDICTION MODEL	26
	5.5	ANALYSIS	29
	5.6	CONCLUSION	29
6	**CONCLUSION**		**30**

CHAPTER NO.	TITLE	PAGENO.

6.1 INTRODUCTION 30
6.2 HIGHLIGHT OF WORK DONE 30
6.2.1 CROP RECOMMENDATION MODEL 30
6.2.2 CROP WATER MODEL 31
6.2.3 CROP YIELD PREDICTION MODEL 31
6.3 FUTURE SCOPE 31

REFERENCES **32**

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
3.1	Ccrop Soil Database	12
3.2	Crop Stages and Crop Coefficient Database	13
3.3	Comparison of accuracy of different algorithms	16
4.1	Crop Yield Database	19
4.2	Accuracy Comparison of different regression models	21
5.1	Comparative study of regression model	27

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
3.1	Proposed design of the project in phase-1	12
3.2	Working of Random Forest Algorithm	14
3.3	Flowchart for calculating water WR for irrigation	15
4.1	Proposed design of our project	18
4.2	XGBoost and Random Forest algorithm for Regression	19
4.3	ANN for Regression	20
5.1	Scatter plot for XGBoost Classifier	22
5.2	Comparison Matrix for Random Forest Classifier	23
5.3	Accuracy comparison of various classification models	23
5.4	Crop stage wise variation in actual, FAO-PM and pan-derived evapotranspiration (mm day^{-1}).	25
5.5	Calculated weekly ET_0 from historical data	25
5.6	Calculated average monthly ET_0 from historical data	26
5.7	Actual vs Predicted yield by XGBoost regression	27
5.8	Actual vs Predicted yield by Random Forest regression	28
5.8	Actual vs Predicted yield by ANN for regression	28
5.8	Comparison of R2 scores of various regression models	28

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Networks
ET	-	Evapotranspiration
ET_0	-	Reference Evapotranspiration
ET_c	-	Crop Evapotranspiration
FAO	-	Food and Agricultural Organization United Nations
GBDT	-	Gradient Boost Decision Tree
GPS	-	Global Positioning System
IDE	-	Integrated Development Environment
K_c	-	Crop Coefficient
ML	-	Machine Learning
MAE	-	Mean Absolute Error
MSE	-	Mean Square Error
OOPs	-	Object Oriented Programming
R2 score	-	R^2 Score or Coefficient of Determination
RH	-	Relative Humidity
SVM	-	Support Vector Machine
WR	-	Water Requirement
XGBoost	-	Extreme Gradient Boost

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

India is a large agricultural country with a huge population approximately 1.4 billion. To meet the food demand of its large population India is largely dependent on its agricultural sector. A total of 40% population of rural household is involved in agriculture. Despite being a large population involved in agriculture it contributes only 17% of GDP and it is declining every year. People which are involved in agriculture have low income and incur losses in it. This is due to the various reason such as Lack of infrastructure, production risks, small and fragmented land holdings, lack of opportunities. These include unfavorable weather conditions, drought, flood, poor irrigation facilities, and lack of timely advisory and data driven decision. They usually take historical parameters and ancestral farming patterns into consideration without knowing that crop depends on weather, present-day, and soil conditions.

Precision farming, revolutionizes traditional agricultural methods by leveraging technology and data analysis to optimize crop production. It empowers farmers to make precise decisions based on factors like weather patterns, soil, historical data analysis ultimately enhancing efficiency and sustainability in farming practices.

However, modern agriculture faces several challenges, including the difficulty in selecting the most suitable crops for varying soil and environmental conditions. Additionally, water scarcity and inefficient irrigation practices pose significant hurdles, leading to resource wastage and reduced crop yields. Furthermore, accurately predicting crop yields remains a persistent challenge, impacting farmers' ability to plan and manage their crops effectively.

In response to these challenges, Agricare emerges as a solution-driven project that harnesses technology and data analysis to revolutionize farming practices. By providing data-driven insights, Agricare empowers farmers to make informed decisions about crop selection and management, addressing challenges related to crop suitability. Moreover, Agricare optimizes irrigation practices using advanced methods like the FAO-Penman equation, promoting water conservation and sustainability in agriculture. Additionally, by accurately predicting crop yields through historical data analysis, Agricare enables farmers to enhance productivity and profitability, ultimately striving towards sustainable success in today's complex agricultural landscape.

1.2 STATE OF THE ART

In this section, a literature survey of previous related works in the field of Crop Recommendation System, Crop Water Requirement and Crop Yield Prediction is presented below:

In the proposed system by Kumar et al [1]," crop yield prediction is facilitated through historical data analysis. Factors such as temperature, humidity, pH, rainfall, and crop type are considered. The system aims to cover a wide variety of crops across different districts of India and predicts the best crop based on field weather conditions. Random Forest algorithm and decision trees are employed for crop prediction, with the Random Forest algorithm yielding highly accurate results, thereby potentially increasing crop yield profitability.

A soil and weather parameter-based recommendation Satish Babu (2013) et al [2], which outlines the development of a precision farming software model for small farms in Kerala State, emphasizing the adaptation of Precision Agriculture principles. The objective is to offer direct advisory services to even the smallest farmers at the crop level. Specifically designed for Kerala's agricultural landscape, the model demonstrates adaptability for potential implementation in various regions across India.

A case study by Mr. Omkar Kulkarni [3] which proposes a machine learning-driven crop recommendation system to aid farmers in making informed decisions. The system integrates historical and scientific data on soil, weather patterns to predict suitable crops. It details the approach including dataset gathering, pre-processing, feature selection, and the application of the Random Forest algorithm.

A case study by U. Surendran et al [8] discusses the serious water shortages in India, particularly in the state of Kerala, where agriculture is the largest consumer of water. The study emphasizes the importance of irrigation in enhancing crop yields but notes the low efficiency due to a lack of location-specific information on irrigation scheduling. The goal is to assess the current status, future demand, and water resource planning for sustainable agriculture in the region.

Research by M. Smith et al [14] outlines the updated approach for estimating crop-water requirements, developed in partnership with FAO and IAEA. It would summarize previous methodologies, detail enhancements made to the FAO method, and discuss the integration of IAEA expertise in soil and water management. Comparisons with older methods and implications for agricultural water management may also be included.

The study proposed by Aymen E Khedr et al [15] addresses food insecurity in Egypt by developing a framework to predict production and imports for a given year. They utilize Artificial Neural Networks, specifically Multi-layer Perceptron implemented in WEKA, for prediction. The framework aims to visualize production, import, need, and availability of food, facilitating decision-making regarding further imports. This work is detailed in the International Journal of Agriculture.

The paper at et al [16] discusses the use of the Random Forest algorithm to predict crop yields based on existing data. Real data from Tamil Nadu were utilized to construct and test the models. The study concludes that the Random Forest Algorithm can be effectively employed for accurate prediction of crop yields.

In et al [17] generated output demonstrates the effectiveness of Random Forests (RF) as a machine-learning method for predicting crop yields at both regional and global scales due to its high accuracy. The paper also explores the implementation of k-nearest neighbor and Support Vector Regression (SVR) techniques.

1.3 OBJECTIVE

The objective of this project is to develop Agricare, a web-based software solution aimed at delivering precision farming solutions. With a focus on enhancing crop production efficiency, Agricare employs Machine Learning algorithms to construct a Crop Recommendation Model, providing farmers with personalized recommendations for optimized yields. Additionally, the project will implement a Water Requirement Model utilizing the FAO-Penman Monteith equation to calculate precise irrigation strategies. Leveraging historical data, Agricare will forecast crop yields, informing farmers for future planning and decisions. By integrating precision farming techniques, Agricare will offer tailored recommendations, facilitating optimal crop management and promoting agricultural sustainability.

1.4 ORGANIZATION OF THE REPORT

The report is organized as follows:

Chapter 1: This chapter gives the introduction of the report, the related works of the project and objective of the report.

Chapter 2: This chapter describes about our project main idea, which is the Agricare, and technical base of it.

Chapter 3: This chapter gives a recap of the previous work of this project.

Chapter 4: This chapter discuss the design and implementation of our models.

Chapter 5: This chapter is based on the result and analysis.

Chapter 6: This includes conclusion, work done, future scope and references.

CHAPTER 2

AGRICARE – A SOGTWARE SOLUTION

2.1 INTRODUCTION

Agricare stands as an innovative software solution at the forefront of precision agriculture, aiming to revolutionize farming practices through advanced technological interventions. At its core, Agricare comprises three distinct models meticulously designed to address crucial aspects of agricultural decision-making.

The Crop Recommender Model harnesses the power of classification machine learning algorithms, including Random Forest, Naive Bayes, Decision Trees, SVM, and XGBoost, to analyze soil composition, weather conditions, and other pertinent factors, thereby offering tailored crop recommendations to farmers.

Complementing this, the Crop Water Requirement Model utilizes the FAO-Penman equation and meteorological data to accurately calculate crop water requirements, aiding in optimizing irrigation strategies and water usage efficiency.

Furthermore, the Crop Yield Prediction Model, leveraging techniques such as Random Forest regression, XGBoost regression, and Artificial Neural Networks (ANN), predicts crop yields based on historical data and various influencing factors such as geographical location, crop type, and environmental variables.

Additionally, Agricare features a user-friendly web application developed using Next.js for frontend and Python Flask for backend, facilitating seamless interaction and integration of these models into farming operations.

This chapter provides a comprehensive overview of Agricare's architecture, functionality, and its pivotal role in enhancing agricultural productivity and sustainability.

2.2 TECHNOLOGIES USED

VsCode and Python Jupyter Notebook IDE is used for the development of the project. Technologies included in the project are

2.2.1 Python 3 – Programming Language

Python is a high-level, interpreted programming language known for its readability and versatility. With a clean syntax and extensive libraries, it facilitates rapid development and supports various paradigms, making it a popular choice for web development and machine learning.

2.2.2 Machine learning algorithms for classification

Gradient Boosting Framework: XGBoost is a high-performing classification algorithm known for its accuracy and efficiency. It belongs to the boosting family of algorithms, utilizing a series of decision trees to progressively improve predictions. Unique features like regularized objective functions help prevent overfitting and enhance generalization. XGBoost is widely used for its ability to handle complex datasets and is favored across various industries, including precision agriculture applications like the Crop Recommender Model in Agricare.

Random Forest Classifier: Random Forest is a versatile ensemble learning algorithm used for classification and regression tasks. It builds multiple decision trees during training and combines their predictions to improve accuracy and reduce overfitting. Widely favored for its robustness, scalability, and interpretability, Random Forest is extensively employed in various fields, including agriculture, for tasks like crop yield prediction in systems like Agricare.

Other Classifiers: Decision Trees, SVM, and Naive Bayes Classifier each bring unique strengths to machine learning tasks. Decision Trees offer interpretability, SVM excels in high-dimensional spaces, and Naive Bayes is simple yet efficient. These algorithms, including Agricare's Crop Recommender Model, play vital roles in various machine learning applications.

2.2.3 Machine Learning algorithm for regression

Artificial Neural Network: ANNs, or Artificial Neural Networks, are effective regression models for crop yield prediction. They utilize interconnected nodes to process input data and capture complex relationships between various factors like location, crop type, and weather conditions. By training on historical data, ANNs offer accurate predictions, aiding farmers in optimizing agricultural practices and maximizing yields, as seen in systems like Agricare.

Random Forest: Random Forest Regression is a robust technique for predicting crop yields. It constructs multiple decision trees during training, with each tree independently forecasting yield based on factors like location, crop type, weather, and farming practices. By combining predictions from these trees, it mitigates overfitting and boosts accuracy. Known for handling large, complex datasets well, it's ideal for agricultural systems like Agricare. Here, it offers farmers valuable insights to refine crop management and enhance overall yield.

Extreme Gradient Boost (XGBoost): XGBoost Regression uses decision trees to predict crop yields by capturing complex relationships among location, weather, and agricultural factors. Its regularized function prevents overfitting, ensuring reliable performance. In applications like Agricare, it provides accurate predictions, helping farmers optimize crop management and improve yields.

2.2.4 Python Libraries

Scikit-Learn, often abbreviated as SKlearn, is a robust and user-friendly machine learning library for Python, providing various machine learning algorithms and workflows.

Pandas: Pandas is a powerful data manipulation library for Python, offering high-performance data structures and tools for cleaning, transforming, and analyzing data.

Matplotlib: Matplotlib is a versatile and customizable plotting library, essential for Python developers, providing a flexible framework to create publication-quality visualizations across a variety of plot types.

Seaborn: Seaborn, built on Matplotlib, specializes in statistical data visualization, offering aesthetically pleasing and informative plots with minimal code, making it a go-to choose for professionals seeking efficient data exploration.

NumPy: NumPy is a foundational library for numerical operations in Python, empowering developers with efficient array manipulations, and mathematical functions critical for scientific computing and data analysis workflows.

2.2.5 Technology Stack for Web Application

NextJs Framework: React framework for server-side rendering and static site generation. Simplifies development of React applications with features like automatic code splitting, hot module replacement, and server-side rendering. Ideal for building fast, scalable, and SEO-friendly web applications.

Python Flask: Lightweight web framework for Python. Simple, flexible, ideal for small to medium projects. Provides essential tools for web servers, HTTP requests, and responses. Widely used for APIs, web services, and dynamic websites. Allows developers to design applications according to specific needs.

2.3 MODELS OVERVIEW

2.3.1 Crop Recommendation Model

The Crop Recommendation Model within Agricare utilizes a classification algorithm to predict the optimal crop based on soil attributes such as nitrogen (N), phosphorus (P), potassium (K), pH of the soil, and weather parameters including humidity, rainfall, and temperature. Leveraging the best suitable machine learning algorithms, the model offers accurate predictions by analyzing the intricate relationships between these factors. By integrating these advanced technologies into the software solution, Agricare ensures precise crop recommendations, empowering farmers to make informed decisions about crop selection. The effectiveness of the model lies in its ability to consider multiple variables simultaneously, resulting in tailored recommendations that maximize crop yield and optimize agricultural practices, thereby enhancing overall farm productivity.

2.3.2 Crop Water Model

Evapotranspiration (ET) is a key component in the hydrological cycle, representing the combined process of water evaporation from the soil surface and transpiration from plant leaves. It plays a crucial role in determining the water requirements for crops, as it quantifies the amount of water that needs to be supplied to maintain optimal plant growth. The FAO-PM (Food and Agriculture Organization - Penman-Monteith) equation is a widely accepted method for estimating potential evapotranspiration, and it forms the basis for many crop-water models.

The FAO-PM equation is an empirical approach that estimates reference evapotranspiration based on meteorological parameters. This equation includes additional factors related to the specific crop, soil characteristics, and management practices. To calculate water requirements for a specific crop, the crop coefficient **Kc** is introduced into the equation, adjusting for the crop's stage of development.

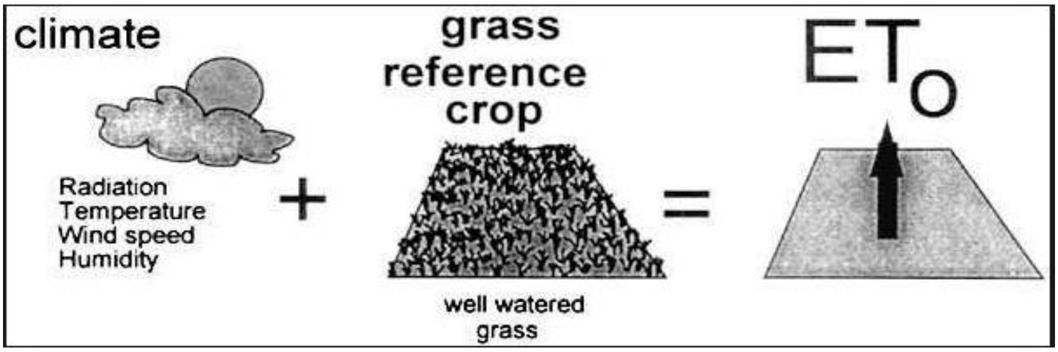

Figure 2.1 Evapotranspiration - simple analysis

2.3.3 Crop Yield Prediction Model

The Crop Yield Prediction Model, which harnesses regression algorithms including Artificial Neural Networks (ANN), XGBoost, and Random Forest (RF) to forecast crop yields. This model operates on a dataset encompassing key features such as state, season, crop type, agricultural area, rainfall, and fertilizer usage. By employing these advanced algorithms, the model effectively analyzes the complex interplay between various factors influencing crop production. Its functionality revolves around processing historical data to generate accurate yield predictions, enabling farmers to anticipate crop output and make informed decisions. It provides farmers with actionable insights, allowing for optimized crop management strategies and improved yield outcomes.

2.4 INTEGRATING ML MODELS WITH WEB APP

Python ML models are seamlessly integrated into Agricare's web application via APIs, allowing users to input data and receive real-time predictions or recommendations. This integration empowers farmers to make informed decisions regarding crop selection, water management, and yield optimization, enhancing the utility and effectiveness of the platform.

2.5 CONCLUSION

The Agricare project report introduces a comprehensive precision agriculture solution, which integrates three vital machine learning models: Crop Recommender, Crop Water Requirement, and Crop Yield Prediction. Leveraging advanced algorithms Agricare delivers customized crop recommendations, water needs, and yield forecasts to farmers. This technological synergy revolutionizes agricultural practices, offering farmers valuable insights to optimize productivity and make informed decisions.

CHAPTER 3

PREVIOUS WORK

3.1 INTRODUCTION

In this chapter, we explore the details of Phase 1 of the Agricare project During Phase 1, we developed two key models: the Crop Recommendation Model and the Crop Water Model. The Crop Recommendation Model suggests suitable crops by analyzing soil and weather data, while the Crop Water Model calculates water needs using the FAO-Penman equation. In this chapter, we discuss the design, implementation, and results of Phase 1, highlighting our methodologies and achievements.

3.2 PROPOSED DESIGN

The proposed design of Phase-1 which includes two key models the Crop Recommender Model and Crop Water Model is illustrated below through a pictorial representation. The illustration shows the workflow of each model providing a clear overview of their functionalities. The design includes a flowchart showing models, dataset used by mentioned models, processing steps, procedures and decision making. Through visual representation it gives insight into the structural composition and operational flow of Agricare's Phase 1 models, facilitating a comprehensive understanding of their implementation.

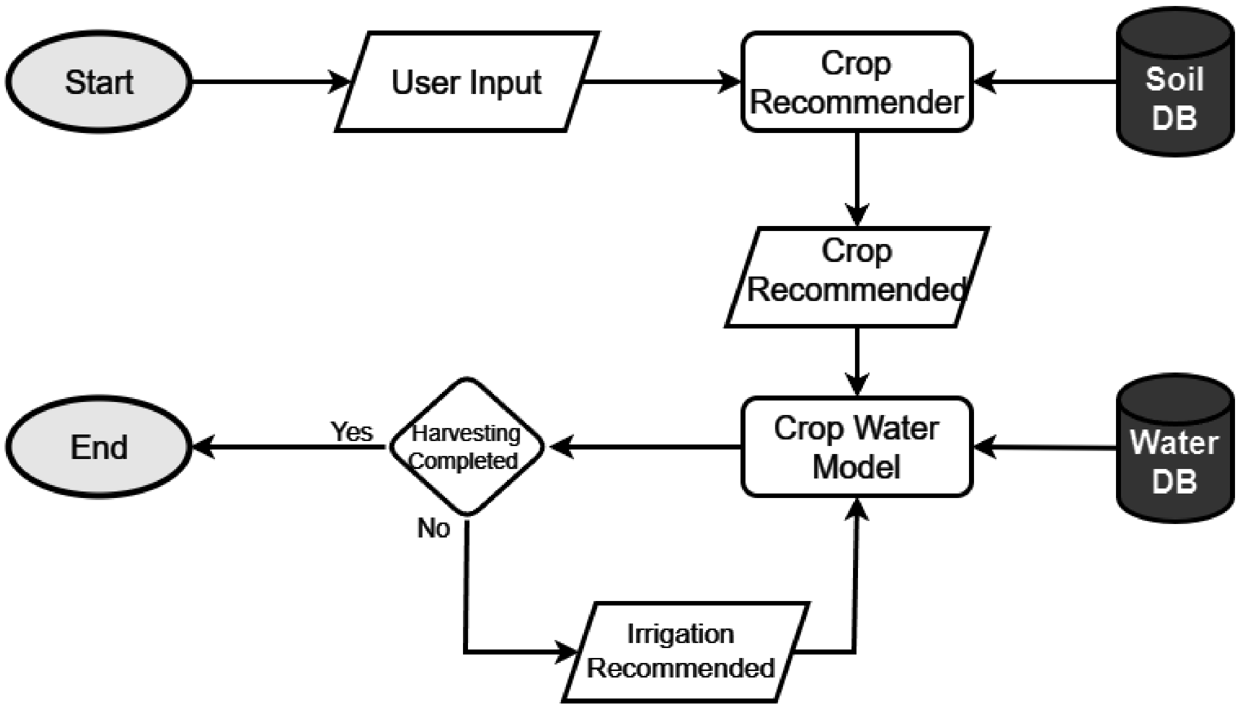

Figure 3.1 Proposed design of the project in Phase-1

3.3 IMPLEMENTATION

The implementation steps for both the Crop Recommendation model and the Crop Water Model involves the following:

3.3.1 Data Collection

Crop recommendation model: This mainly constitutes the past 20 years of soil, and crop data set of India. This data consists of over 2000 observations and over 20 different crops. Soil texture and nutrient values like N, P, K and pH determine the fertility of the soil. This data set is used for training the crop recommendation model.

Table - 3.1 Crop soil database

N	P	K	temperature	humidity	pH	rainfall	label
90	42	43	20.87974	82.00274	6.502985	202.9355	rice
85	58	41	21.77046	80.31964	7.038096	226.6555	rice
60	55	44	23.00446	82.32076	7.840207	263.9642	rice

Crop water model: For the crop water model we have gathered a data set for different crops growing period and crop coefficient for different stages. This database is used by the Crop water model.

Table – 3.2 Crop Stages and Crop Coefficient database

Crop	Total	Lini	Ldev	Lmid	Llate	kc_ini	kc_dev	kc_mid	kc_late
Barley/Oats/Wheat	120	15	25	50	30	0.35	0.75	1.15	0.45
Bean/green	75	15	25	25	10	0.35	0.7	1.1	0.9
Bean/dry	95	15	25	35	20	0.35	0.7	1.1	0.3
Cabbage	120	20	25	60	15	0.45	0.75	1.05	0.9

3.3.2 Data Preprocessing

The dataset in preprocessed for cleaning and transforming data to ensure quality and compatibility. It involves handling missing values, removing duplicates, and standardizing formats. Also perform normalization, scaling and label encoding on dataset for better algorithm performance. Extracting meaningful features from the data, such as calculating derived variables or creating new attributes that enhance predictive capabilities.

3.3.3 Crop Recommendation Model

Dataset is split into 4:1 for training and testing respectively. Machine learning classification algorithms such as Random Forest, Naive Bayes, Decision Trees, SVM, XGBoost are selected for model training. After analysis of result by each model, best model is selected to use by our model.

Random Forest: Random Forest is a versatile machine learning algorithm for classification tasks. It operates by creating an ensemble of decision trees. Each tree is trained on a random subset of the training data, selected with replacement (bootstrap sampling). This diversification helps prevent overfitting and ensures robustness. Moreover, at each node of the tree, only a random subset of features is considered for splitting, further enhancing diversity and generalization. Once the trees are built, they collectively vote on the class of unseen data points. The class with the most votes across all trees is then predicted as the final output. This approach tends to yield accurate and reliable predictions while being relatively resistant to noise and outliers in the data.

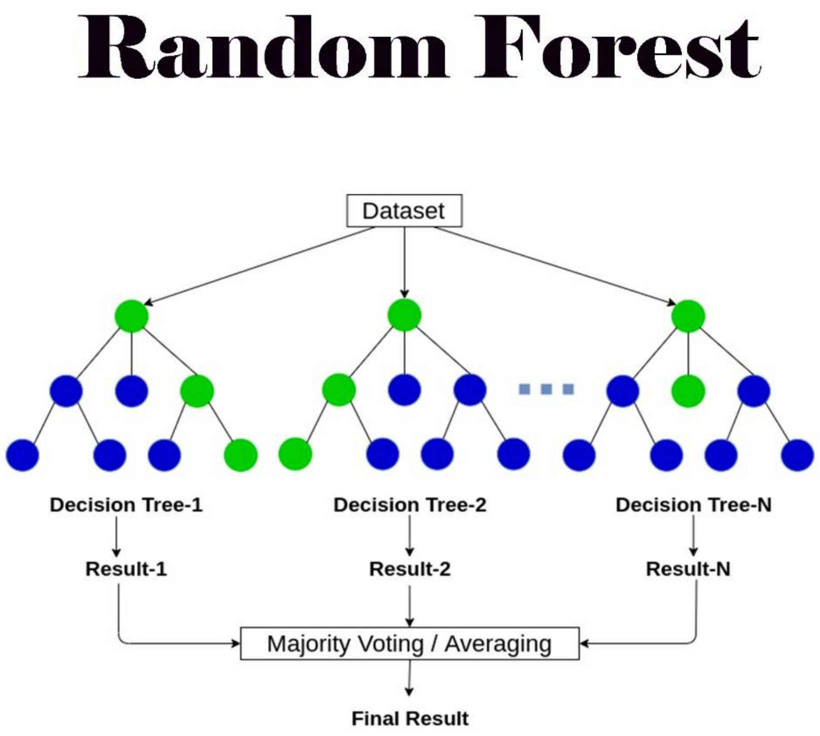

Figure 3.2 Working of Random Forest Algorithm

3.3.4 Crop Water Model

The FAO-PM (Food and Agriculture Organization - Penman-Monteith) equation is a widely accepted method for estimating potential evapotranspiration, and it forms the basis for many crop-water models. It that estimates reference evapotranspiration based on meteorological parameters.

The equation is expressed as follows:

$$\boldsymbol{ET_0} = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad \text{.........} \quad \text{Eq 3.1}$$

Where:

ET_o reference evapotranspiration [mm day^{-1}],

R_n net radiation at the crop surface [MJ $m^{-2}day^{-1}$],

G soil heat flux density [MJ $m^{-2}day^{-1}$],

T mean daily air temperature at 2 m height [°C],

u_2 wind speed at 2 m height [m s^{-1}],

e_s saturation vapor pressure [kPa],

e_a actual vapor pressure [kPa],

$e_s - e_a$ saturation vapor pressure deficit [kPa],

Δ slope vapor pressure curve [kPa C^{-1}],

γ psychrometric constant [kPa C^{-1}].

To calculate water requirements for a specific crop, the crop coefficient **Kc** is introduced into the equation, adjusting for the crop's stage of development. The actual evapotranspiration **ET_c** can be calculated using the following relationship:

$$ET_c = K_c\, ET_o \qquad \text{......... Eq 3.2}$$

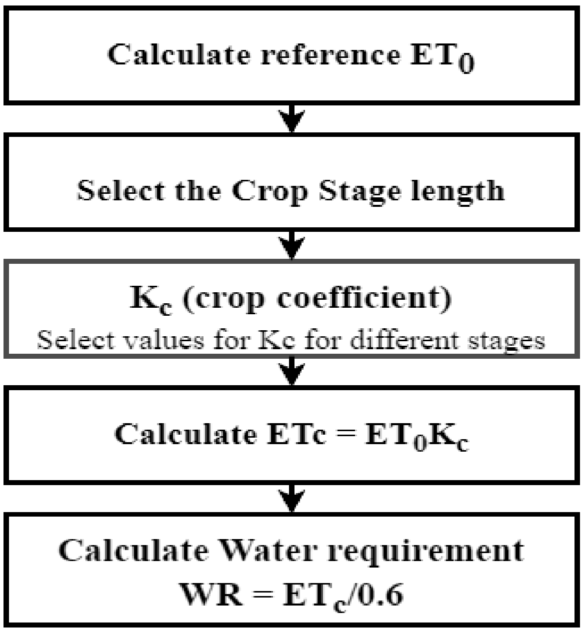

Figure 3.3 flowchart for calculating water requirement for irrigation

3.4 RESULT

3.4.1 Crop Recommendation Model

Crop recommendation models, trained with various classification algorithms on different datasets, showcase diverse effectiveness. By analyzing succinct evaluation metrics, we have selected the Random Forest algorithm for its highest accuracy among all classification algorithm for our model which helps in optimal crop selection across agricultural contexts, ultimately enhancing productivity and sustainability.

Table 3.3 Comparison of Accuracy of different algorithm

Classification Algorithm	Accuracy
Decision Tree	93.41
SVM	8.86
Gaussian Naïve Bayes	98.86
XGBoost	98.86
Random Forest	99.09

3.4.2 Crop Water Model

The crop water model calculates the total water requirement for irrigation by estimating daily evapotranspiration for the crop. This value is then multiplied by the crop coefficient, which accounts for the specific crop's water needs, and adjusted by irrigation efficiency to determine the precise water needed for the crop.

Further results are mentioned in Result and Discussion chapter.

3.5 CONCLUSION

In conclusion, Phase 1 of the Agricare project marked the successful development and implementation of the Crop Recommender Model and the Crop Water Requirement Model. These models, utilizing machine learning algorithms and meteorological data, offer valuable insights for optimal crop selection and water management in precision agriculture. With the completion of Phase 1, the groundwork has been laid for further advancements, including the integration of the crop yield prediction model and the deployment of the Agricare software solution.

CHAPTER 4

DESIGN AND IMPLEMENTATION

4.1 INTRODUCTION

The foundation of any successful project lies in its design and implementation. In this chapter, we embark on a journey through the intricate details of Agricare's structure and the meticulous steps taken for its practical application. From the conceptual design to the tangible results, each aspect contributes to the effectiveness of this agricultural software. Agricare is put together and the careful steps we have taken to make it work in the real world.

We'll start from the big ideas in the design phase down to the actual, tangible outcomes. It's like unfolding a map that guides us through the details of Agricare's structure, and each part of this journey is essential to making sure this agricultural software does its job effectively. Understanding the synergy between design and implementation, Agricare's implementation becomes a harmonious blend of planning and execution. From the conceptual blueprint to the actual results, each aspect is explored holistically, contributing to Agricare's effectiveness in the agricultural landscape.

4.2 DESIGN

The design phase is the architectural blueprint of Agricare. A conceptual design acts as the guiding start, outlining the overall structure and functionality. Every feature, including the algorithms, is carefully conceptualized to ensure accuracy and efficiency. This section unravels the design thinking behind Agricare, showcasing how every element aligns with the goal of empowering farmers.

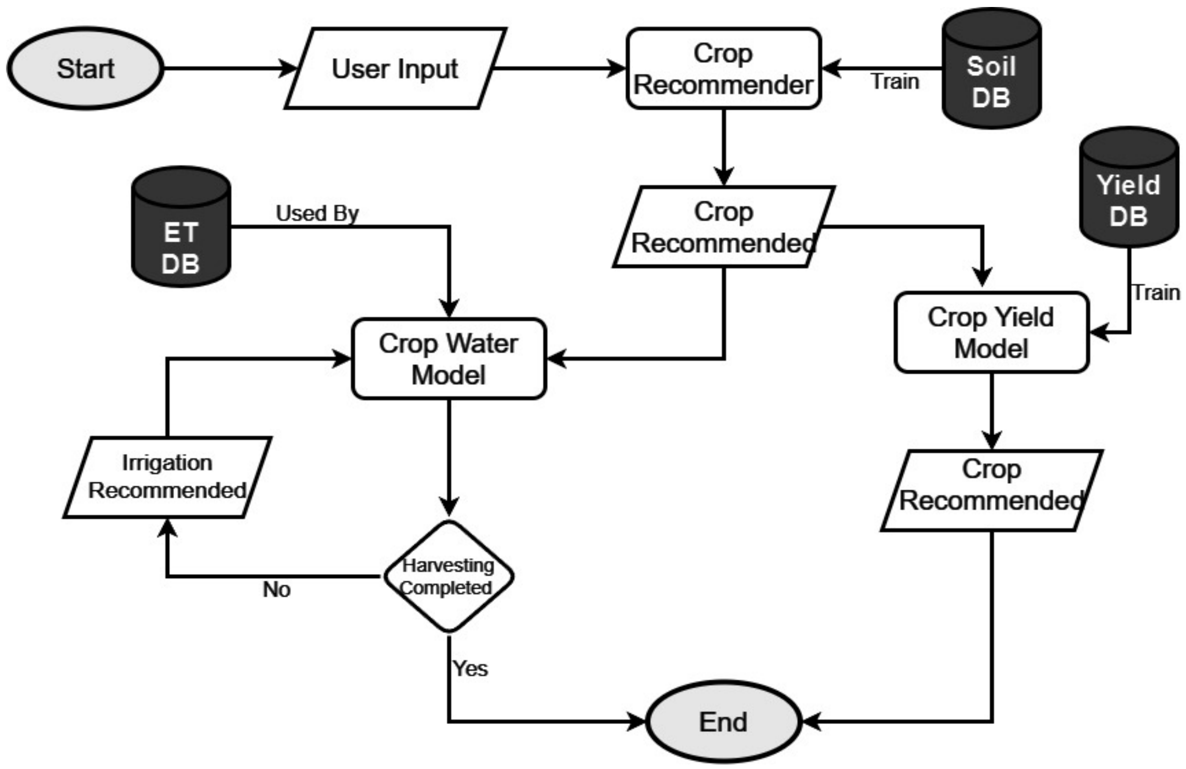

Figure 4.1 proposed design of our project

The three models used in Agricare allows it to be suitable for any farm size and any soil type. The models include the Crop Recommendation Model, the Crop Water Model and the Crop Yield Prediction Model. The later two models work in parallel to each other.

The input which includes Soil Chemical compositions, Soil moisture content, weather conditions, and land location etc. is collected from user, then processed through the crop recommendation model. The crop selected is passed on along with soil data to the crop water model, and crop yield prediction model. The former two models are discussed in detail in previous work. In current chapter we will focus on yield model.

4.3 IMPLEMENTATION

4.3.1 Data Preprocessing

This step includes removal redundancies and outlier for clear processing of the data set.

Crop yield database This database consists of yield of crop from the various state of India for years 1997-2014. It consists of 19000 records. This database is used for crop yield model.

Table - 4.1 Crop yield database

Crop	Season	State	Area	Annual Rainfall	Fertilizer	Yield
Arecanut	Whole Year	Assam	73814	2051.4	7024878	0.796087
Arhar/Tur	Kharif	Assam	6637	2051.4	631643.3	0.710435
Castor seed	Kharif	Assam	796	2051.4	75755.32	0.238333
Coconut	Whole Year	Assam	19656	2051.4	1870662	5238.052

4.3.2 Crop Yield Prediction Model

Data set for training: This model takes gets trained on the Crop soil database.

Machine Learning Algorithms: XG Boost, Random Forest and Artificial Neural Network for regression in parallel.

XG Boost: Extreme Gradient Boosting, is a scalable, distributed gradient-boost decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems. Gradient Boosting is based on "**boosting**" or improving a single weak model by combining it with a number of other weak models in order to generate a collectively strong model

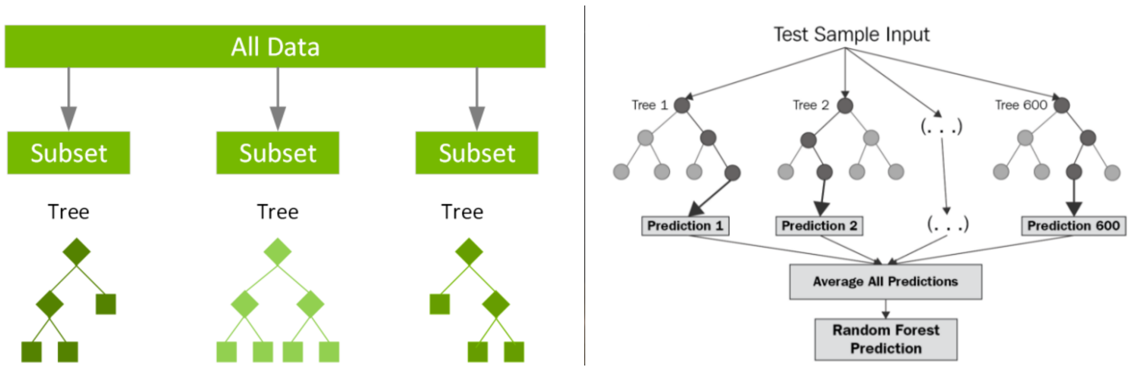

Figure 4.2 XGBoost and Random Forest

Random forest uses a technique called "bagging" to build full decision trees in parallel from random bootstrap samples of the data set. The final prediction is an average of all of the decision tree predictions. The crop, Indian state, season, area, average rainfall, and fertilizer acts as a major role in estimating the crop yield.

Artificial Neural Network for Regression: Artificial Neural Networks (ANNs) for regression tasks are a powerful class of machine learning models that can learn complex patterns and relationships in data. Comprising interconnected nodes organized in layers, ANNs can capture non-linearities and interactions between features, making them well-suited for tasks where the relationship between input and output variables is intricate.

In the context of regression, ANNs learn to map input features to continuous output values, making them capable for predicting numerical value of crop yields. By using multiple hidden layers, ANNs can extract hierarchical representations of data, allowing them to learn from raw input features. Below is the diagram of ANN implementation for crop yield, consisting 6 input nodes, and 1 output node.

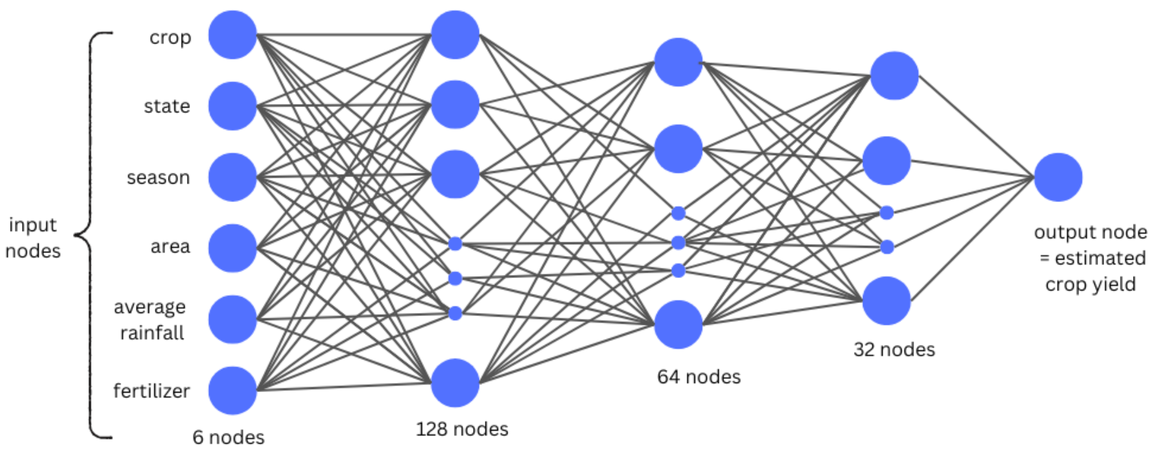

Figure 4.3 Artificial Neural Network for regression

During training, ANNs use a process known as **back-propagation** to update the model parameters iteratively, minimizing the discrepancy between predicted and actual values. This involves propagating the error backward through the network and adjusting the weights and biases of each node to reduce the loss function.

<u>Expected Input</u>: Crop, state (of India where farm is located), season of agriculture, farm area, average rainfall for land, fertilizer quantity used on the land area.

<u>Output</u>: Estimated yield (in Tonnes/Hectare) for the land.

Table 4.2 Accuracy Comparison of different regression models

	XG Boost Regression	Random Forest Regression	ANN for Regression
MSE	0.02509159522926599	0.0187574264511114	0.08796776256550387
MAE	0.01431275297249475	0.0099971727739125	0.05166765806641393
R^2 Sore	0.9754514818316115	0.9816485552305481	0.9139361926637085

4.4 CONCLUSION

In the culmination of the design and implementation chapter, it becomes apparent that energy between thoughtful design and effective implementation is the bedrock of Agricare's success. The meticulous planning, coding endeavors, and real-world results all underscore the commitment to creating a tool that not only meets but exceeds the expectations of precision farming.

The comparison of XGBoost, Random Forest Regression, and Artificial Neural Network (ANN) models reveals their distinct strengths and applications in predictive modeling. XGBoost stands out for its robustness, efficiency in handling large datasets, and ability to capture complex relationships with boosted decision trees. Random Forest Regression excels in handling noisy and correlated data, providing reliable predictions through ensemble learning and feature importance analysis. On the other hand, ANN offers flexibility in modeling nonlinear relationships and extracting intricate patterns from data, making it suitable for tasks with complex input-output mappings. Each model has its unique characteristics and trade-offs, and the choice depends on the specific requirements and nature of the datasets.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 INTRODUCTION

Through a rigorous examination of data outputs, performance metrics, and real-world applications, we scrutinize the efficacy of the precision farming solutions provided by Agricare. The discussion dissects the implications of these results, elucidating their significance in the context of agricultural decision-making. This chapter serves as the empirical cornerstone, substantiating the practical impact and viability of Agricare in enhancing agricultural practices through data-driven insights.

5.2 CROP RECOMMENDATION MODEL

The accuracy of the model can be measured using confusion matrix, accuracy score or by graphical visualization. However, accuracy score would not be a great measure for this, since this model has more than 2 classes. Each crop with holds one class.

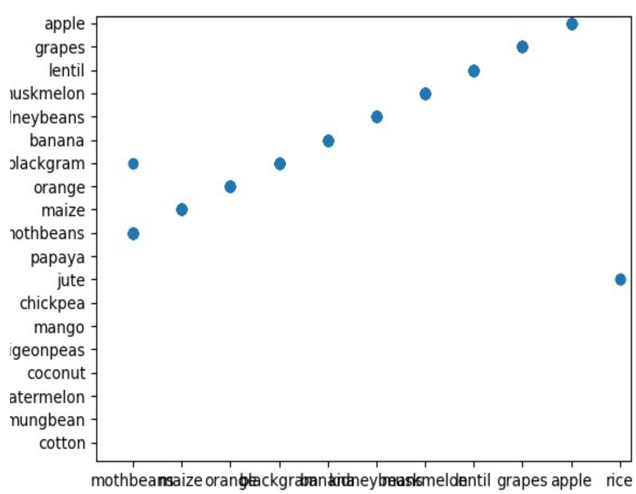

Figure 5.1 Scatter plot for XGBoost classifier

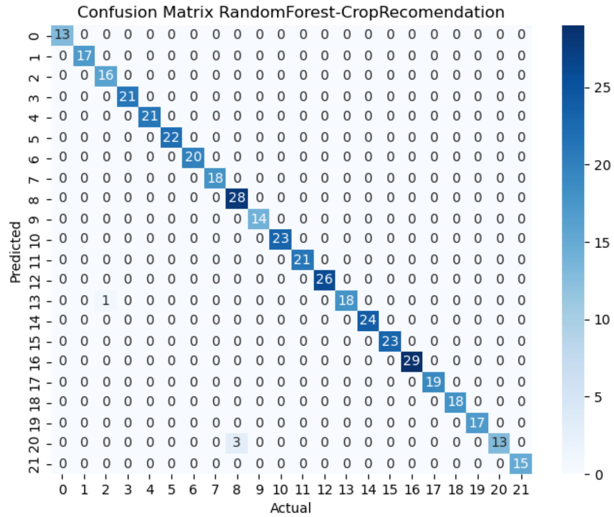

Figure 5.2 Confusion matrix for Random Forest classifier

Number of classes = Number of distinct crops is the database.

Label each class with integer values we get a confusion matrix of the form:

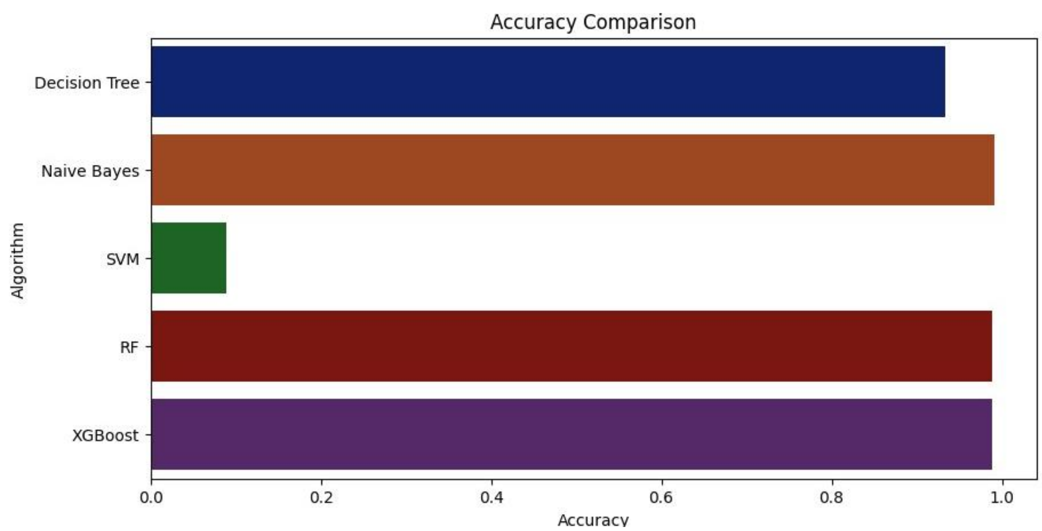

Figure 5.3 Accuracy comparison of the various classification models

INPUT:

```
data = np.array([[104, 18, 30, 23.603016, 60.3, 6.7, 140.91]])
best_crop = crop_model.predict_best_crop_for_input(data)
print(best_crop)
```

OUTPUT:

```
---Predicting user input values---
coffee
```

Classification Report: Random Forest Classifier

	precision	recall	f1-score	support
0	1.00	1.00	1.00	20
1	1.00	1.00	1.00	22
2	1.00	1.00	1.00	22
3	1.00	1.00	1.00	21
4	1.00	1.00	1.00	21
5	1.00	1.00	1.00	26
6	1.00	1.00	1.00	21
7	1.00	1.00	1.00	23
8	0.86	0.90	0.88	21
9	1.00	1.00	1.00	12
10	1.00	1.00	1.00	21
11	1.00	1.00	1.00	20
12	1.00	1.00	1.00	21
13	1.00	1.00	1.00	16
14	1.00	1.00	1.00	23
15	1.00	1.00	1.00	26
16	1.00	1.00	1.00	23
17	1.00	1.00	1.00	20
18	1.00	1.00	1.00	19
19	1.00	1.00	1.00	12
20	0.86	0.80	0.83	15
21	1.00	1.00	1.00	15
Accuracy			**0.99**	**440**
macro avg	**0.99**	**0.99**	**0.99**	**440**
weighted avg	**0.99**	**0.99**	**0.99**	**440**

5.3 CROP WATER MODEL

In the provided Python implementation using the Penman library to calculate reference evapotranspiration (ETo) and adjust crop water requirements based on growth stages and irrigation efficiency, the results can be analyzed and summarized in a results section. Additionally, a brief conclusion can be drawn based on the outcomes.

The comparison of crop stage wise actual evapotranspiration with FAO Penman-Monteith estimates reveals insights into the accuracy and effectiveness of the model, aiding in understanding variations and optimizing irrigation strategies for different growth stages

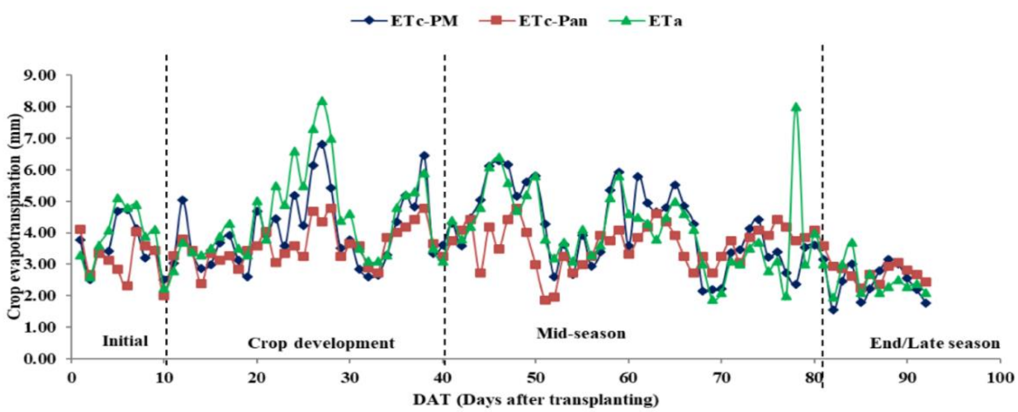

Figure 5.4 Crop stage wise variation in actual, FAO-PM and pan-derived evapotranspiration (mm day^{-1})

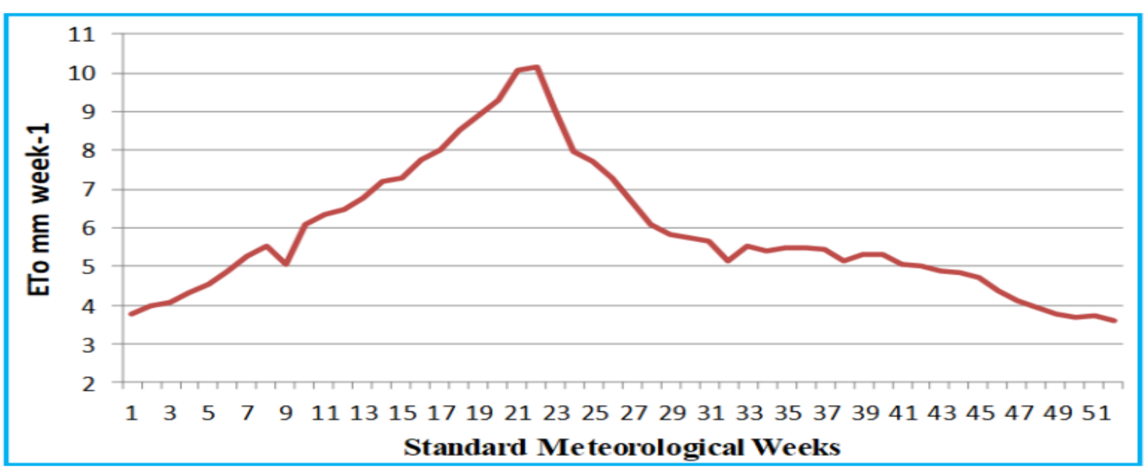

Figure 5.5 Calculated weekly ET_0 from historical data

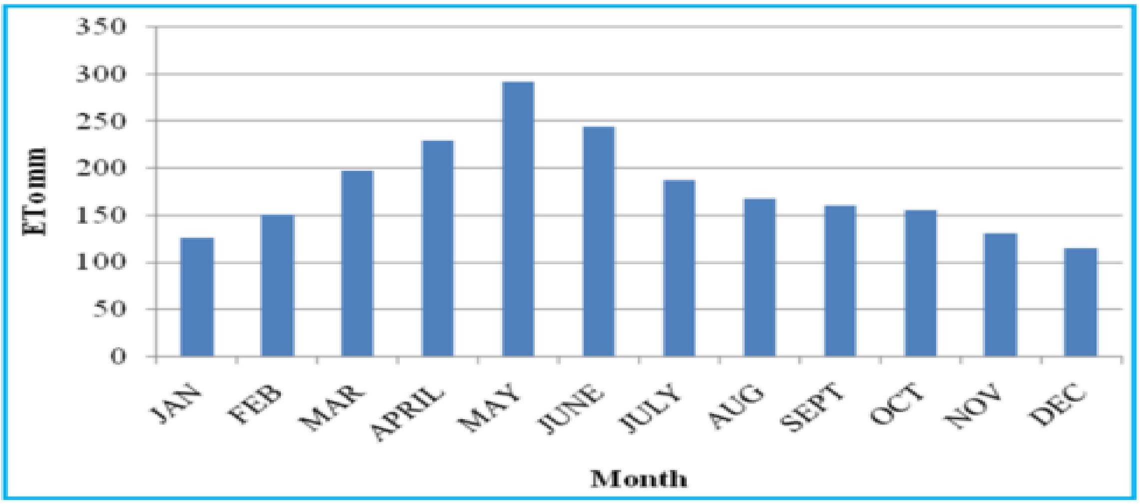

Figure 5.6 Calculated average monthly ET_0 from historical data

1. Daily ETo:

The calculated daily reference evapotranspiration using the Penman library is (X) mm/day.

2. Crop Water Requirement for Each Growth Stage:

Crop Water Requirement (Ini): Y mm/day

Crop Water Requirement (Dev): Z mm/day

Crop Water Requirement (Mid): W mm/day

Crop Water Requirement (Late): V mm/day

3. Total Crop Water Requirement:

The total water requirement for the entire crop growth cycle is X mm/day.

4. Adjusted Crop Water Requirement with Irrigation Efficiency:

The adjusted crop water requirement, considering an irrigation efficiency factor c, is A mm/day.

5.4 CROP YIELD PREDICTION MODEL

The crop yield prediction model is built on top of the XG Boost Regression, the Random Forest Regression, and the Artificial Neural Network with Regression algorithms. Their accuracy can be measured in terms of:

Mean Squared Error (MSE): MSE measures the average squared difference between the predicted values and the actual values. Lower values of MSE indicate better model performance.

Mean Absolute Error (MAE): MAE measures the average absolute difference between the predicted values and the actual values. Lower values of MAE indicate better model performance

R-Squared value (R2): An R-Squared value shows how well the model predicts the outcome of the dependent variable. R-Squared values range from 0 to 1.

Table 5.1 Comparative study of Regression Models for crop prediction

	XG Boost Regression	Random Forest Regression	ANN for Regression
MSE	0.02509159522926599	0.0187574264511114	0.08796776256550387
MAE	0.01431275297249475	0.0099971727739125	0.05166765806641393
R^2 Sore	0.97545148183161151	0.9816485552305481	0.9139361926637085

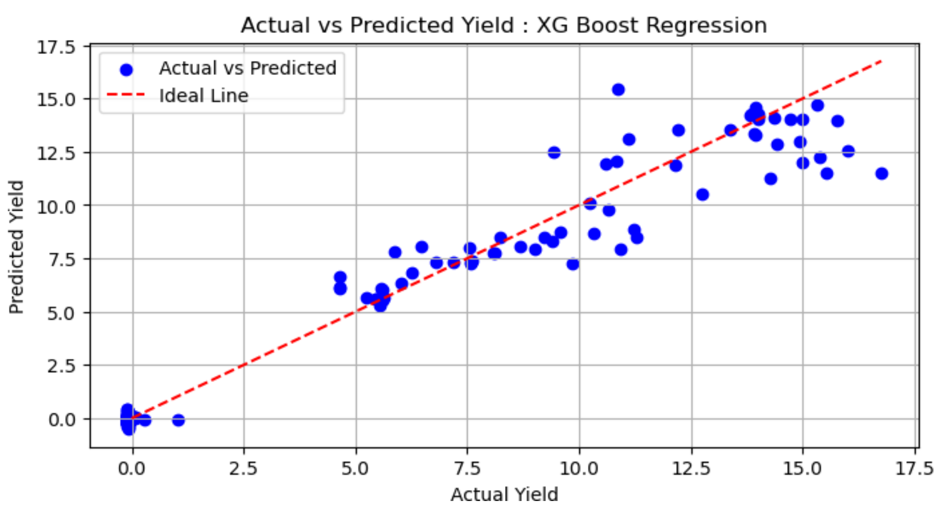

Figure 5.7 Actual vs Predicted yield by XGBoost regression

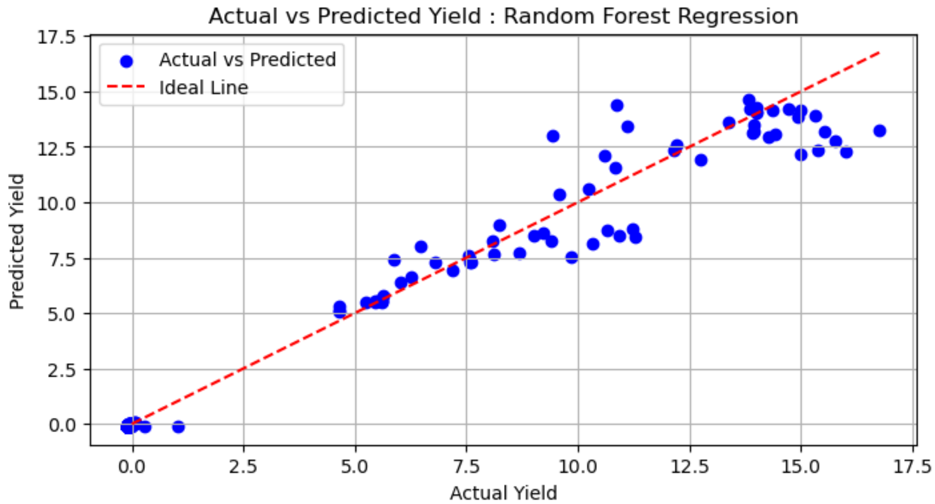

Figure 5.7 Actual vs Predicted yield by Random Forest regression

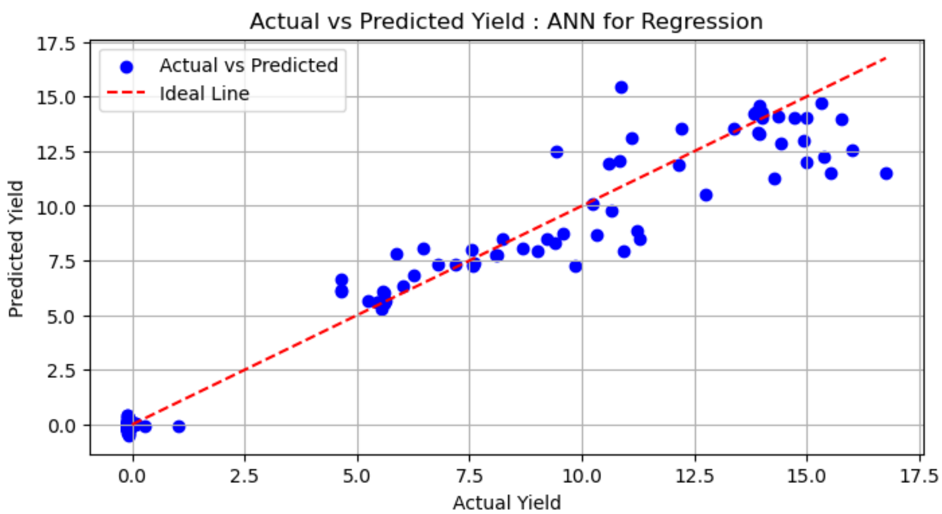

Figure 5.8 Actual vs Predicted yield by ANN for regression

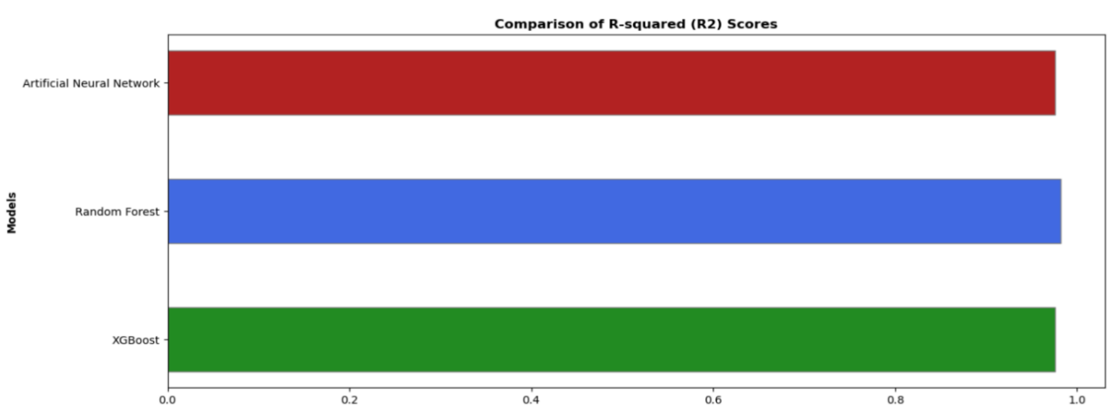

Figure 5.9 Comparison of R^2 values for various regression models.

5.5 ANALYSIS

In reviewing the graphs, it's evident that Random Forest outperformed XGBoost and ANN with a higher R-squared score, indicating better capture of data variability and more accurate predictions. However, while Random Forest excelled in this aspect, factors like computational efficiency and model interpretability should also be considered when determining the most suitable model.

Despite Random Forest's superiority in R-squared, both XGBoost and ANN demonstrated valuable performance, each with unique strengths in handling complex relationships and nonlinear patterns, respectively. Therefore, a comprehensive evaluation considering various factors beyond R-squared is crucial for making informed decisions about model selection and deployment.

5.6 CONCLUSION

In conclusion, for the case of Crop Prediction Model, the XGBoost outperforms the Random Forest classification. These models help us to determine the best crop for the farmer. The utilization of the FAO Penman equation in our crop water model provides a robust framework for estimating reference evapotranspiration and determining crop water requirements. For the case of Crop Yield Prediction, the Random Forest outperforms the XGBoost and Artificial Neural Networks. They help us to estimate the yield for a farmer's land.

CHAPTER 6

CONCLUSION

6.1 INTRODUCTION

Our project is designed to support farmers. By inputting soil data and weather information, the system recommends the optimal crop for production, considering the specific soil features. Additionally, it calculates the precise water requirement for efficient crop irrigation using the FAO-Penman Monteith equation, which relies on the evapotranspiration process. This technology aims to provide farmers with tailored guidance for crop selection and optimal water management, contributing to enhanced agricultural productivity.

6.2 HIGHLIGHT OF THE WORK DONE

6.2.1 Crop Recommendation Model: Tailored Insights

The Crop Recommendation Model is the result of intricate work with Machine Learning. It's not just a tool; it's your personalized farming advisor. By analyzing soil fertility and historical crop data, it gives tailored recommendations. This model acts as a dependable guide for farmers, aiding in optimal crop choices. The integration of machine learning isn't just about suggesting crops—it's about arming farmers with data for better decision-making, ultimately enhancing agricultural productivity.

6.2.2 Water Requirement Model: Precision in Irrigation

The Water Requirement Model exemplifies our dedication to precise irrigation. Anchored in the FAO-Penman Monteith equation, it serves as a strategic tool, not just conserving water but contributing significantly to crop health. By intricately considering the evapotranspiration process, the model calculates precise water requirements. This ensures efficient irrigation tailored to the specific needs of each crop, emphasizing its crucial role in promoting sustainability in agriculture.

6.2.3 Crop Yield Prediction Model: Estimating the yield

In reviewing the graphs, it's evident that Random Forest outperformed XGBoost and ANN with a higher R-squared score, indicating better capture of data variability and more accurate predictions. Despite Random Forest's superiority in R-squared, both XGBoost and ANN demonstrated valuable performance, each with unique strengths in handling complex relationships and nonlinear patterns, respectively. Therefore, a comprehensive evaluation considering various factors beyond R-squared is crucial for making informed decisions about model selection and deployment.

6.3 FUTURE SCOPE

Future work includes refining the Crop Recommendation Model by adapting it to diverse agricultural landscapes and integrating real-time data sources. The Water Requirement Model will evolve by incorporating a broader range of environmental factors and extending its applicability to various crops.

The performance and accuracy of the Crop Yield Prediction Model can be increased by applying it over a larger data set, which can be generated with the help of IoT devices. This makes it real-time operable as well. Further, the website can be enhanced to support farmers on areas other than crop recommendation, irrigation, and yield estimation, such as fertilizers quantity, weather analysis etc.

REFERENCES

[1] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741. IEEE, 2020.

[2] Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open-Source Software (ICFOSS) Trivandrum, India.

[3] Mr. Omkar Kulkarni, Mr. Sushil Buragute, Mr. Ashwin Khode, Mr. Melvin Fernando, Prof. Nilesh Korade (2022) "Crop Recommendation System Using Machine Learning Algorithm", International Journal of Creative Research Thoughts (IJCRT)

[4] Pudumalar S, Ramanujam E, Rajashree RH, Kavya C, Kiruthika T, Nisha J. Crop recommendation system for precision agriculture. In: 2016 8th International Conference on Advanced Computing. 2016; p. 32–36.

[5] Dhruv Piyush Parikh, Jugal Jain, Tanishq Gupta and Rishit Hemant Dabhade, "Machine Learning Based Crop Recommendation System" in IJARSCT 2021

[6] Lakshmi.N, Priya. M, Sahana Shetty, and Manjunath C. R (2018), Crop Recommendation System for Precision Agriculture, IND: International Journal for Research in Applied Science & Engineering Technology

[7] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. (2019) "Crop Recommendation System to Maximize Crop Yield in Ramtek Region using Machine Learning." International Journal of Scientific Research in Science and Technology

[8] U. Surendran, C. M. Sushanth, George Mammen, and E. J. Joseph (2015), 'Modelling the Crop Water Requirement using FAO-CROPWAT and Assessment of Water Resources

for Sustainable Water Resource Management: A case study in Palakkad district of humid tropical Kerala, India', International Conferences on Water Resources, Coastal and Ocean Engineering.

[9] Rohitashw Kumar, Vijay Shankar and Mahesh Kumar (2011), 'Modelling of Crop Reference Evapotranspiration: A Review', Universal Journal of Environmental Research and Technology.

[10] Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998): Crop evapotranspiration – Guidelines for computing crop water requirements – FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations, Rome.

[11] Bailey, B.J., Montero, J.I., Biel, C., Wilkinson, D.J., Anton, A., Jolliet, O. (1993): Transpiration of Ficus Benjamin: comparison of measurements with predictions of the Penman–Monteith model and a simplified version. Agric. For. Meteorology. 65, 229-243.

[12] Christiansen, J. E. (1968): Pan Evaporation and evapotranspiration from climatic data.'' ASCE J. of Irrigation and Drainage, Vol. 94: 243–265.

[13] Doorenbos, J. and Pruitt, W.O. (1977): Guidelines for predicting crop water requirements. Irrigation and Drainage Paper 24 (revised) Food and Agricultural Organization of the United Nations, Rome.

[14] M. Smith, R. Allen, L. Pereira 1998 "Revised FAO Methodology for Crop-Water Requirements" International Atomic Energy Agency (IAEA).

[15] Aymen E Khedr, Mona Kadry, Ghada Walid (2015), "Proposed Framework for implementing Data Mining Techniques to Enhance Decisions in Agriculture Sector Applied Case on Food Security Information Center Ministry of Agriculture, Egypt, International.

[16] P. Priya, U. Muthaiah M. Balamurugan. Predicting Yield of the Crop Using Machine Learning Algorithm. International Journal of Engineering Science Research Technology.

[17] Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, United States of America.

[18] Aditya Shastry, H. A Sanjay and E. Bhanushree, (2017) "Prediction of crop yield using Regression Technique", International Journal of Computing.

[19] Djodiltachoumy, (2017) "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics.

[20] J. Jeong, J. Resop, N. Mueller and team. Random forests for global and regional crop yield prediction. PLoS ONE Journal.

[21] Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod (2019) "Crop Yield Prediction Using Machine Learning", International Journal of Science and Research (IJSR).