Abstract

India being an agricultural country, its economy mainly depends on agriculture yield growth and allied agroindustry products. In India agriculture is largely influenced by rain water which is highly unpredictable. Agriculture growth depends on diverse soil parameters like nitrogen, phosphorous, potassium, crop rotation, soil moisture, surface temperature. It also depends on weather aspects which include temperature, rainfall etc. Agriculture is one of the major fields in our country and also plays a major role in our country's economy. India is the second –largest producer of agriculture crops and agriculture is one of the major and least paid occupation in India. Variability in seasonal climate conditions can have harmful effects, with incidents of drought reducing production. Developing better techniques to predict crop productivity in various climatic conditions can help farmer and other stakeholders in their decision making in terms of agronomy and crop choice.

CHAPTER 1

INTRODUCTION

Agriculture is the most important sector of Indian Economy. Indian agriculture sector accounts for 18 percent of India's GDP and provides employment to 50% of the country's workforce. But latest studies have shown a steady decline in the contribution made by agriculture to the Indian economy although it is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India.

OVERVIEW

The Data Analysis is process of inspecting cleansing, modelling data with the goal of discovering useful information and conclusions. It is a process of analyzing, extracting and predicting the meaningful information from huge data to extract some pattern. This process is used by companies to turn the raw data of their customer to useful information. This analysis can also be used in the field of Agriculture. Most farmers were relied on their long-terms experiences in the field on particular crops to expect a higher yield in the next harvesting period but still the they don't get worth price of the crops. It is mostly happening due to improper irrigation or inappropriate crops selection or also sometimes the crop yield is less than that of expected. Agricultural researchers insist on the need for an efficient mechanism to predict and improve the crop growth and Majority of research works in agriculture focus on biological mechanisms to identify crop growth and improve its yield. The outcome of crop yield primarily depends on parameters such as variety of crop, seed type and environmental parameters such as sunlight (Temperature), soil (ph.), water (ph.), rainfall and humidity. By analyzing the soil and atmosphere at particular region best crop in order to have more crop yield and the net crop yield can be predicted. This prediction will help the farmers. To choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months. Crop yield estimation is a difficult task since it is affected by various factors such as genetic potential of crop cultivar, soil, weather, cultivation practices (date of sowing, amount of irrigation and fertilizer, etc.) and biotic stress. Several methods of crop yield estimation have been developed such as statistical, agro-meteorological, empirical, biophysical, mechanistic.

India is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation. The number of studies Information and Communication Technology (ICT) can be applied for prediction of crop yield. By the use of Data Mining, we can also predict the crop yield. By fully analyze the previous data we can suggest the farmer for a better crop for the better yield.

Smart agriculture is the way of conveying information from traditional farmers to the educated farmers. To obtain estimates of aggregate physical production functions for the yields of various crops in specified states, considering various technological factors and a newly developed weather index as inputs. Regression and coefficient of determination analysis along with Average Error rate were carried out to make a decent comparison between our actual result which is called target and prediction model that is friendly interface for farmers, which gives the analysis of rice production based on available data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity Accurate and timely monitoring of agricultural crop conditions and estimating potential crop yields are essential processes for operational programs Because of the importance of predicting crop yield, the purpose of this study is to apply several forecasting methods for evaluating crop yield estimates in Ghana. Crop yield forecasting, which provides information for decision Makers.

PROBLEM STATEMENT

The production of agriculture is affected by several climate factors. Like as metrological parameters (Humidity, wind speed, temperature, and moisture), precipitation parameters (rainfall, region wise rainfall, irrigation etc.), and soil parameters (PH, organic carbon, phosphorus, fiber etc.). And due to continuously change in climate condition everything is messed.

In India farmers still follow the traditional technology which they adopted from their ancestor. But the problem is that in earliest time climate was very healthy everything was happened on time. But now most of the things have been changed due to global warming and many other factors. The main problem with agriculture in India is lack of rainfall in seasonal time. Humidity is also necessary for crops, but it has been excessive, it also converts as drawback. Winter season is been affected so Rabi crops are widely affected. Since few years the rainfall in winter season was high as expected.

To overcome these above issues, we need to develop a system which will able to find the hidden facts or results, patterns and insights. The farmer can predict which crop should sow so that he/she can get more benefit. In proposed system we are applying data analytics techniques on agriculture production-based datasets and find the insights so that it can help to the farmers and their decision making.

EXISTING SYSTEMS

This section provides the proposed methodology used for crop yield prediction. The purpose of crop yield prediction is to estimate production in agriculture sector for better crop management and make strategic decisions for improving crop yield in future. The Existing model can be incorporated with a decision support system (DSS) that can be used in precision agriculture which aims at complete farm management.

DEMERITS

- Factors like climate and location of market and planting area is not taken into consideration
- The system doesn't take area of land being cultivated and the sowing date. The market price of the cultivated crops after harvesting is not considered.

PROPOSED SYSTEM

Prediction of the crop yield using the efficient algorithm and suggest how much quantity of fertilizer should be used to get the proper yield for the crop using naïve Bayesian algorithm. The data mining techniques on historical climate and crop production data several predictions are made which increase the crop productivity.

The decision support system must be implemented for the farmers to take proper decisions about soil and crop to be cultivated. They have collected the dataset with attributes of the crop season, Area and production in hectares and analyzed with various algorithms in WEKA.

MERITS

- Factors like climate and location of market and planting area is taken into consideration
- . The market price of the cultivated crops after harvesting is considered.

OBJECTIVES AND METHODOLOGY

OBJECTIVES

- Provide the farmer with the yield of a crop based on land area, rainfall, temperature and district using machine learning.
- Predict the future market price of crops by taking previous crop price and predicted yield data into consideration.
- To predict the crop price for six major districts and eight crops using multiple linear regression and random forest.
- To compare the predicted result of the different algorithms and determine which approach is more suitable.

METHODOLOGY

❖ CROP YEILD PREDICTION

The outcome of crop yield primarily depends on parameters such as variety of crop, seed type and environmental parameters such as sunlight (Temperature), soil (ph), water (ph), rainfall and humidity. By analyzing the soil and atmosphere at particular region best crop in order to have more crop yield and the net crop yield can be predicted. This prediction will help the farmers. To choose appropriate crops for their farm according to the soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer and months.

❖ FERTILIZER PREDICTION

India is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation.

Based on soil type and soil PH we suggest what kind of fertilizer should be used for particular crop.

❖ RANDOM FOREST

Random Forest algorithm is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the result. But one thing to note is that creating the forest is not the same as constructing the decision with information gain or gain index approach.

How Random Forest Algorithm Works?

There are two stages in Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage. The whole process is shown below, and it's easy to understand using the figure.

firstly, shows the Random Forest creation pseudocode:

1. Randomly select "**K**" features from total "**m**" features where **k <<m**
2. Among the "**K**" features, calculate the node "**d**" using the best split point
3. Split the node into **daughter nodes** using the **best split**
4. Repeat the **a to c** steps until "l" number of nodes has been reached
5. Build forest by repeating steps **a to d** for "n" number times to create **"n" number of trees**

F11	**F12**	**F13**	**F14**	**F15**	**T1**
F21	**F22**	**F23**	**F24**	**F25**	**T2**
:	**:**	**:**	**:**	**:**	**:**
:	**:**	**:**	**:**	**:**	**:**
Fm1	**Fm2**	**Fm3**	**Fm4**	**Fm5**	**Tm**

F11	**F12**	**F13**	**F14**	**F15**	**T1**
F81	**F82**	**F83**	**F84**	**F85**	**T8**
:	**:**	**:**	**:**	**:**	**:**
:	**:**	**:**	**:**	**:**	**:**
Fj1	**Fj2**	**Fj3**	**Fj4**	**Fj5**	**Tj**

F21	**F22**	**F23**	**F24**	**F25**	**T2**
F51	**F52**	**F53**	**F54**	**F55**	**T5**
:	**:**	**:**	**:**	**:**	**:**
:	**:**	**:**	**:**	**:**	**:**
Fm1	**Fm2**	**Fm3**	**Fm4**	**Fm5**	**Tm**

F31	**F32**	**F33**	**F34**	**F35**	**T3**
F61	**F62**	**F63**	**F64**	**F65**	**T6**
:	**:**	**:**	**:**	**:**	**:**
:	**:**	**:**	**:**	**:**	**:**
Fk1	**Fk2**	**Fk3**	**Fk4**	**Fk5**	**Tk**

Fig: 1.1 Different random forest trees dataset

In the next stage, with the random forest classifier created, we will make the prediction. The random forest prediction pseudocode is shown below:

1. Takes the **test features** and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target)
2. Calculate the **votes** for each predicted target
3. Consider the **high voted** predicted target as the **final prediction** from the random forest algorithm

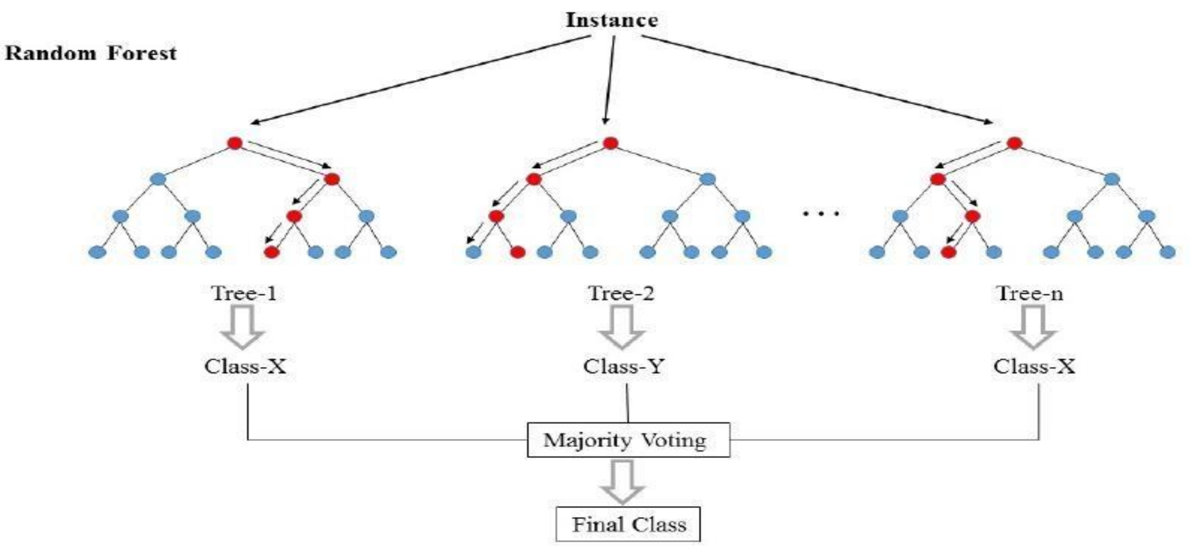


Fig: 1.2 Random Forest Tree

Applications of Random Forest

- **Banking** Random forest algorithm is used to find loyal customers, which means customers who can take out plenty of loans and pay interest to the bank properly, and fraud customers, which means customers who have bad records like failure to pay back a loan on time or have dangerous actions.
- **Medicine** Random forest algorithm can be used to both identify the correct combination of components in medicine, and to identify diseases by analyzing the patient's medical records.
- **Stock Market** Random forest algorithm can be used to identify a stock's behaviour and the expected loss or profit.
- **E-Commerce** Random forest algorithm can be used for predicting whether the customer will like the recommend products, based on the experience of similar customers.

Advantages of Random Forest Algorithm

1. For applications in classification problems, Random Forest algorithm will avoid the overfitting problem
2. For both classification and regression task, the same random forest algorithm can be used
3. The Random Forest algorithm can be used for identifying the most important features from the training dataset, in other words, feature engineering.

CHAPTER 2

LITERATURE SURVEY

A literature survey or a literature review in a project report shows the various analyses and research made in the field of interest and the results already published, taking into account the various parameters of the project and the extent of the project.

A literature survey includes the following

• Existing theories about the topic which are accepted universally.

• Books written on the topic, both generic and specific.

• Research done in the field usually in the order of oldest to latest.

• Challenges being faced and on-going work, if available.

Literature survey describes about the existing work on the given project .It deals with the problem associated with the existing system and also gives user a clear knowledge on how to deal with the existing problems and how to provide solution to the existing problems .different thing.

2.1 OBJECTIVES OF LITERATURE SURVEY

• Learning the definitions of the concepts.

• Access to latest approaches, methods and theories.

• Discovering research topics based on the existing research

• Concentrate on your own field of expertise– Even if another field uses the same words, they usually mean completely

• It improves the quality of the literature survey to exclude sidetracks– Remember to explicate what is excluded.

Data Security in Cloud Computing Using RC and AES

Before building our android application, the following system is taken into consideration

1. Rice yield estimation at pixel scale using relative vegetation indices from unmanned aerial systems

Timely and accurate prediction of rice yield information is closely related to the people's livelihood, which has been attached great importance by all levels of government. Satellite remote sensing provides the possibility for large-scale crop yield estimation, but they are usually limited by spatial and spectral resolution. Unmanned Aerial Vehicles (UAV) remote sensing with hyperspectral sensors can obtain high spatial temporal resolution and hyperspectral images on demand. Generally, time-series Vegetation Indices (VIs) are used for estimating grain yield. But multi-day vegetation indices may be affected by different background and illumination condition, so the differences between vegetation indices may include the effects induced from external condition, which will pose a negative effect on the accuracy of crop yield estimation. Therefore, in this study, the relative vegetation index and relative yield were proposed and used to estimate rice yield at pixel scale. And the optimal growth stages for crop yield estimation would also be determined. Hyperspectral images of critical rice growth stages at tillering stage, jointing stage, booting stage, heading stage, filling stage, ripening stage were obtained from July 28 to November 24 in 2017.Firstly, all possible two-band combinations of discrete channels from 500nm to 900nm was used to create Relative Normalized Difference Vegetation Index (RNDVI). Then the best RNDVI at different growth stages were determined for rice yield estimation. Finally, different combinations of growth stages were tested to obtain the optimal combinations for yield estimation. These models were validated at pixel scale using the measured yields. The result shows that four-growth-stage model with RNDVI[635, 784] at tillering stage, RNDVI[744, 807] at jointing stage, RNDVI[712,784] a booting stage, RNDVI[736, 816] at heading stage with the multiple linear regression function gain a higher R2 (0.74) and lower RMSE(248.97kg/ha). The mean absolute percentage error of estimated rice yield of 4.31%. Results shows that the yield estimations at pixel scale with relative vegetation indices were acceptable. In the study, a yield estimation method with relative vegetation indices is proposed and the optimal growth stage combinations for rice yield estimation were determined. This study explores the possibility of yield estimation at pixel scale using hyper spectral images fromUAV platform, which will further improve the method system for remote sensing of yield estimation.

2. Use of Deep Neural Networks for Crop Yield Prediction: A Case Study of Soybean Yield in Lauderdale County, Alabama, USA

World population is constantly increasing, and it is necessary to have sufficient crop production. Monitoring crop growth and yield estimation are very important for the economic development of a nation. The prediction of crop yield has direct impact on national and international economies and play important role in the food management and food security. Deep learning gains importance on crop monitoring, crop type classification and crop yield estimation applications with the recent advances in image classification using deep Convolutional Neural Networks. Traditional crop yield prediction approaches based on remote sensing consist of classical Machine Learning methods such as Support Vector Machines and Decision Trees. Convolutional Neural Network (CNN] and Long-Short Term Memory Network (LSTM] are deep neural network models that are proposed for crop yield prediction recently. This study focused on soybean yield prediction of Lauderdale County, Alabama, USA using 3D CNN model that leverages the spatiotemporal features. The yield is provided from USDA NASS Quick Stat tool for years 2003-2016.The satellite data used is collected from NASA's MODIS land products surface reflectance, land surface temperature and land surface temperature via Google Earth Engine. The root mean squared error (RMSE] is used as the evaluation metric in order to be able to compare the results with other methods that

generally, uses RMSE as the evaluation metric.

3.Crop Yield Prediction and Efficient use of Fertilizers.

India being an agriculture country, its economy predominantly depends on agriculture yield growth and agroindustry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen(N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region,

amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which

enhance to increase the crop yield and increase farmer revenue.

4.Prediction of major crop yields of Tamilnadu using K-means and Modified KNN

Agriculture is the principal source of livelihood for more than 40 percent of the population of this state. According to Food and Agricultural Organization (FAO)researchers, between 2010 and 2050 the world population will increase by one third. The demand for crop production will increase by 60% higher than the current production. Hence prediction plays a major role to find out the demand of crop production for maximizing the yield. For that in this paper we propose a prediction method for the major crops of Tamilnadu using K-means and Modified K Nearest Neighbor (KNN). MATLAB and WEKA are used as the tool for clustering and classification

respectively. The number result shows that our method is better than traditional data mining approach.

5.Rice crop yield prediction in India using support vector machines

Food production in India is largely dependent on cereal crops including rice, wheat and various pulses. The sustainability and productivity of rice growing areas is dependent on suitable climatic conditions. Variability in seasonal climate conditions can have detrimental effect, with incidents of drought reducing production. Developing better techniques to predict crop productivity in different climatic conditions can assist farmer and other stakeholders in better decision making in terms of agronomy and crop choice. Machine learning techniques can be used to improve prediction of crop yield under different climatic scenarios. This paper presents the review on use of such machine learning technique for Indian rice cropping areas. This paper discusses the experimental results obtained by applying SMO classifier using the WEKA tool on the dataset of 27 districts of Maharashtra state, India. The dataset considered for the rice crop yield prediction was sourced from publicly available Indian Government records. The parameters considered for the study were precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield for the Kharif season (June to November) for the years 1998 to 2002. For the present study the mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE) were calculated. The experimental results showed that the performance of other techniques on the same dataset was much better compared to SMO.

6. Regional Winter Wheat Maturity Date Prediction Using Remote Sensing-Crop Model Data Assimilation and Numerical Weather Prediction

Optimizing harvesting schedules requires a method for maturity date prediction, to avoid the influence of adverse weather and prevent the decline of crop yield or quality due to inappropriate harvest schedule. However, most prediction models are statistical-based thus are not suitable for regional application, and remote sensing-based models lacked predictability. We presented a framework that assimilated leaf area index (LAI) derived from Moderate Resolution Imaging Spectro radiometer (MODIS) into World Food Studies (WOFOST) crop growth model, and forecast meteorological data from THORPEX Interactive Grand Global Ensemble (TIGGE) was used as weather data input for the future periods. We selected the winter wheat planting area in Henan Province as study area and recalibrated WOFOST model based on observation data from agrometeorological sites. A cost function based on normalization was constructed to quantify the difference between simulated LAI and MODIS LAI products. First the MODIS LAI profile was smoothed by Savitzky-Golay (S-G) filter, and then these two LAI profiles were normalized to keep their trend information. Then we selected parameters in WOFOST model that are sensitive to maturity date as optimization parameters, such as emergence date (IDEM), effective temperature sum from emergence to anthesis (TSUM1) and effective temperature sum from anthesis to maturity (TSUM2). These parameters have significant differences between years and no obvious spatial and temporal patterns. By means of Shuffled Complex Evolution method developed at the University of Arizona (SCE-UA) algorithm, we simulated in each pixel in the study area and retrieved the optimal parameters set of this pixel. Then we run WOFOST by this optimal parameter set to simulate the growth and development of winter wheat. Moreover, we transformed TIGGE data into the CABO-format weather file to drive WOFOST simulating winter wheat growth in the next 16 d and obtained a spatial distribution of winter wheat. Moreover, we transformed TIGGE data into the CABO-format weather file to drive WOFOST simulating winter wheat growth in the next 16 d and obtained a spatial distribution of winter wheat maturity date in the study area. Comparing the forecasting date with the observed date from agrometeorological sites, it demonstrated that this method had substantial accuracy in predicting regional maturity date with correlation coefficient (R2) of 0.90 and the root mean square error (RMSE) was 1.93 d. Besides that, the distribution map of maturity prediction showed obvious spatial variability. This method can remedy the shortages of poor predictability and lacking regional differences in most previous methods, and it provides a reference for the future study of crop maturity prediction at a regional scale with longer forecast period.

7.Multiple Crop Yield prediction using dual-polarimetricTerraSAR-X stripmap imagery

This paper presents the results of an experiment carried out to relate the yield from various crops to TerraSAR-X dual polarimetric imagery. X-band wavelength has higher sensitivity to smaller crop structures, especially stem and head density making it suitable for relating yield to backscatter. The coherent dual-polarimetric mode of TerraSAR-X was also used to emphasize the volume scattering through dual-polarimetric entropy/alpha decomposition. Good correlations to yield data as gathered by harvester telemetry were obtained.

8.Remote sensing indicators for crop growth monitoring at different scales

Crop growth monitoring is critical in yield estimation and prediction. In this paper, the authors investigated several indicators for crop growth monitoring by remote sensing at different scales. The experiments were conducted in a study area in Hebei province in North China Plain. The target crop in this research is winter wheat, which is one of the important grain crops in China. The study at canopy scale is based on field experiment with different fertilizer supply treatments to winter wheat to get different crop growth gradients. The canopy spectra and LAI data of different crop growth situation were collected and analyzed. The correlation coefficients between vegetation indices and LAI were calculated. The best growth monitoring indicators at canopy scale for each phenological stage were determined by the correlation coefficients. The best indicator in early elongation stage is SAVI (L=0.3). While in jointing stage, the best indicator is SAVI (L=0.2). In the heading and milk stages, the best indicator is NDVI. At the regional scales, the multi-spectral remote sensing data from HJ-IA satellite were used to study the crop growth monitoring indicators. The performances of the vegetation indices at different spatial scales for crop growth monitoring were studied in different phenological stages by comparing the correlation coefficients of vegetation indices and LAI data. The result shows that the vegetation indices which have a function to decrease soil-effect are suitable for the region in different crop coverage in early elongation stage and for the region in low crop coverage, but they have limitations in dense crop coverage while in booting and heading stages.

9.Assessing relevant climate data for agricultural applications

Climate change is expected to substantially reduce agricultural yields, as reported in the by the Intergovernmental Panel on Climate Change (IPCC). In Sub-Saharan Africa and (to a lesser extent) in South Asia, limited data availability and institutional networking constrain agricultural research and development. Here we performed a review of relevant aspects in relation to coupling agriculture–climate predictions, and a three-step analysis of the importance of climate data for agricultural impact assessment. First, using meta-data from the scientific literature we examined trends in the use of climate and weather data in agricultural research, and we found that despite agricultural researchers' preference for field-scale weather data (50.4% of cases in the assembled literature), large-scale datasets coupled with weather generators can be useful in the agricultural context. Using well-known interpolation techniques, we then assessed the sensitivities of the weather station network to the lack of data and found high sensitivities to data loss only over mountainous areas in Nepal and Ethiopia (random removal of data impacted precipitation estimates by ±1300 mm/year and temperature estimates by ±3 ◦C). Finally, we numerically compared IPCC Fourth Assessment Report (4AR) climate models' representation of mean climates and interannual variability with different observational datasets. Climate models were found inadequate for field-scale agricultural studies in West Africa and South Asia, as their ability to represent mean climates and climate variability was limited: more than 50% of the country-model combinations showed.

10.Crop Syst , a Cropping Systems Simulation Model: Water/Nitrogen Budgets and Crop Yield

In agriculture, water and nitrogen are two critical resources for growing a crop. However, their management cannot be analyzed independently of weather, soil characteristics, field hydrology, crop characteristics, crop rotation, and management factors. This paper describes the water, nitrogen, and crop growth components of CropSyst, a comprehensive cropping systems simulation model, and provides preliminary verification of these components. The water budget of the model properly describes crop water use. Predicted nitrogen contents throughout the soil profile did not exactly match the measured values from leaching experiments, but they did follow the general trends of the data. The agreement between simulated and observed biomass and yield of corn, winter wheat and spring wheat grown in two locations with a total of 77 data points was good as shown by several statistical indicators. Based on this preliminary validation, CropSyst appears promising as a tool to analyze management practices for water and nitrogen. Additional validation of model components, including a wide range of crops and conditions, should be conducted in the future

CHAPTER 3

SYSTEM REQUIREMENTS SPECIFICATION

A software requirements specification (SRS) – a requirements specification for a software system – is a complete description of the behavior of a system to be developed. In addition to a description of the software functions, the SRS also contains non-functional requirements. Software requirements are a sub-field of software engineering that deals with the elicitation, analysis, specification, and validation of requirements for software.

INTRODUCTION

Software requirements are a sub-field of software engineering that deals with the elicitation, analysis, specification, and validation of requirements for software.

GENERAL DESCRIPTION

PRODUCT PERSPECTIVE

- **Python** is an interpreted, high-level, general-purpose programming language.
- Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.
- Python is dynamically typed and garbage collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.
- **Multi-platform**: Python is available for all major operating systems, Windows, Linux/Unix, MacOS X, most likely your mobile phone OS, etc.
- **The Jupyter Notebook** is an **interactive computing environment** that enables users to author notebook documents that include: - Live code - Interactive widgets - Plots - Narrative text - Equations - Images – Video
- These documents provide a **complete and self-contained record of a computation** that can be converted to various formats and shared with others using email, Dropbox, version control systems

REQUIREMENTS

FUNCTIONAL REQUIREMENTS

- Identify the effect of various factors such as land area, rainfall and temperature on crop production and yield.
- Predict the future market price of crops by taking previous crop price and predicted yield data into consideration.
- To compare the accuracy of multiple linear regression and random forest and determine which algorithm is accurate.
- To integrate this with a user interface and for real-time data.

NON-FUNCTIONAL REQUIREMENTS

- **Usability:** The GUI is required to be easily usable.
- **Reliability:** The system is expected to make reliable prediction of crop yield and crop price.
- **Compatibility:** The system is required to be compatible on various platforms.
- **Performance:** The model is required to make prediction with minimum response time and minimum percentage error.
- **Scalability:** The system is required to be flexible and adaptive to growing and changing datasets.

SOFTWARE REQUIREMENTS

Operating System	Windows XP or above
Programming language	Python
Development environment	Anaconda-jupyter

HARDWARE REQUIREMENTS

Processor	Intel Core i5 or AMD FX 8 core series with clock speed of 2.4 GHz or above
RAM	2GB or above
Hard disk	40 GB or above
Input device	Keyboard, mouse
Sensors	Moisture, humidity, ph sensor

CHAPTER 4

SYSTEM ANALYSIS AND DESIGN

Analysis is the process of breaking a complex topic or substance into smaller parts to gain a better understanding of it. Analysts in the field of engineering look at requirements, structures, mechanisms, and systems dimensions. Analysis is an exploratory activity. The Analysis Phase is where the project lifecycle begins. The Analysis Phase is where you break down the deliverables in the high-level Project Charter into the more detailed business requirements. The Analysis Phase is also the part of the project where you identify the overall direction that the project will take through the creation of the project strategy documents.

INTRODUCTION

Gathering requirements is the main attraction of the Analysis Phase. The process of gathering requirements is usually more than simply asking the users what they need and writing their answers down. Depending on the complexity of the application, the process for gathering requirements has a clearly defined process of its own. This process consists of a group of repeatable processes that utilize certain techniques to capture, document, communicate, and manage requirements.

Systems design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could see it as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

If the broader topic of product development "blends the perspective of marketing, design, and manufacturing into a single approach to product development," then design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user.

Until the 1990s systems design had a crucial and respected role in the data processing industry. In the 1990s standardization of hardware and software resulted in the ability to build modular systems. The increasing importance of software running on generic platforms has enhanced the discipline of software engineering.

Object-oriented analysis and design methods are becoming the most widely used methods for computer systems design. The UML has become the standard language in object-oriented analysis and design. It is widely used for modelling software systems and is increasingly used for high designing non-software systems and organizations.

System design is one of the most important phases of software development process. The purpose of the design is to plan the solution of a problem specified by the requirement documentation. In other words, the first step in the solution to the problem is the design of the project.

The design of the system is perhaps the most critical factor affecting the quality of the software. The objective of the design phase is to produce overall design of the software. It aims to figure out the modules that should be in the system to fulfil all the system requirements in an efficient manner.

The design will contain the specification of all these modules, their interaction with other modules and the desired output from each module. The output of the design process is a description of the software architecture.

The design phase is followed by two sub phases

- High Level Design
- Low Level Design.

SYSTEM ARCHITECTURE

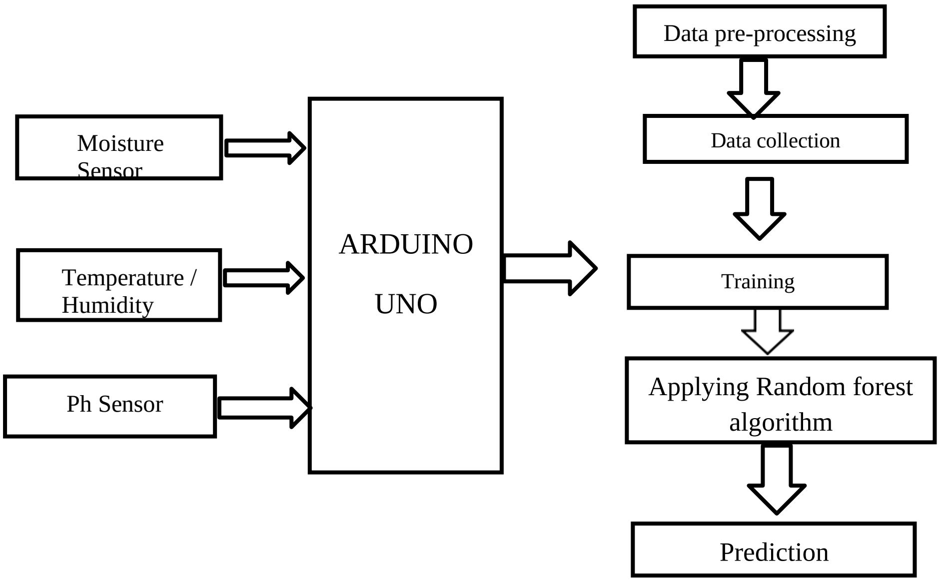


Fig: 4.1 General Block Diagram Describing the Activities

HIGH LEVEL DESIGN

In the high-level design, the proposed functional and non-functional requirements of the software are depicted. Overall solution to the architecture is developed which can handle those needs. This chapter involves the following consideration.

- Design consideration
- Data flow diagram

4.3.1 DESIGN CONSIDERATION

There are several design considerations issues that need to be addressed or resolved before getting down designing a complete solution for the system.

DATA FLOW DIAGRAM

A data flow diagram is the graphical representation of the flow of data through an information system. DFD is very useful in understanding a system and can be efficiently used during analysis. A DFD shows the flow of data through a system. It views a system as a function that transforms the inputs into desired outputs. Any complex systems will not perform this transformation in a single step and a data will typically undergo a series of transformations before it becomes the output.

With a data flow diagram, users are able to visualize how the system will operate that the system will accomplish and how the system will be implemented, old system data flow diagrams can be drawn up and compared with a new systems data flow diagram to draw comparisons to implement a more efficient system.

Data flow diagrams can be used to provide the end user with a physical idea of where the data they input, ultimately as an effect upon the structure of the whole system.

Below section explains the module wise data flow diagram.

DATA COLLECTION

In this module datasets are collected from various sources like GKVK, DES, agmarknet.gov.in and krishimaratavahini. The datasets include information like temperature, rainfall, price, area, production and yield of the previous 5 years (2013- 2018)

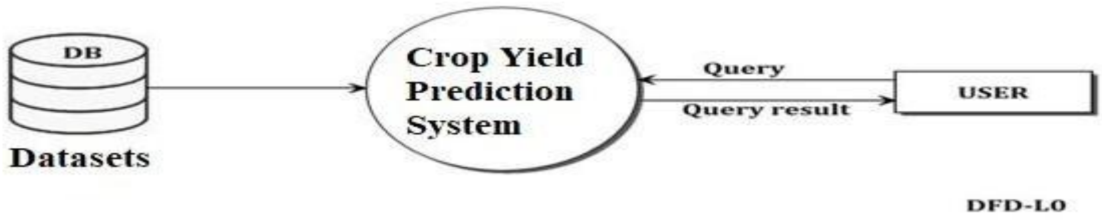


Fig: 4.2 Data Flow Diagram of Data Collection

DATA PRE-PROCESSING: Data pre-processing includes removing of the unwanted attributes from our datasets. Feature extraction is done in order to extract only the attributes that affect the price and yield of a crop like rainfall, temperature, location, area, production and yield.

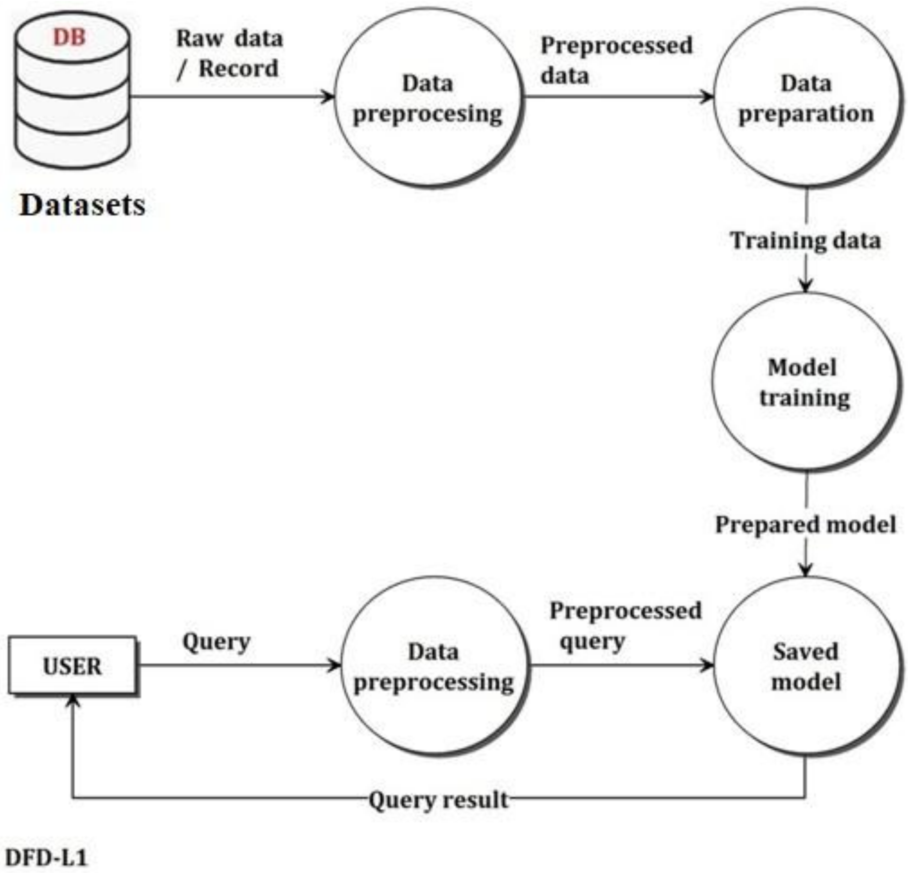


Fig: 4.3 Dataflow Diagram of Pre-Processing

ANALYSIS AND PREDICTION

In this module, patterns in data is recognized, percentage correlation between various factors affecting crop yield and price are determined. Various data visualization techniques are used to study the patterns in data and factors causing change. Algorithms like Multiple Linear Regression and Random Forest are used to predict crop yield and price. The accuracy of these algorithms are compared using mean absolute percentage error thus helping us determine the most suitable approach for prediction.

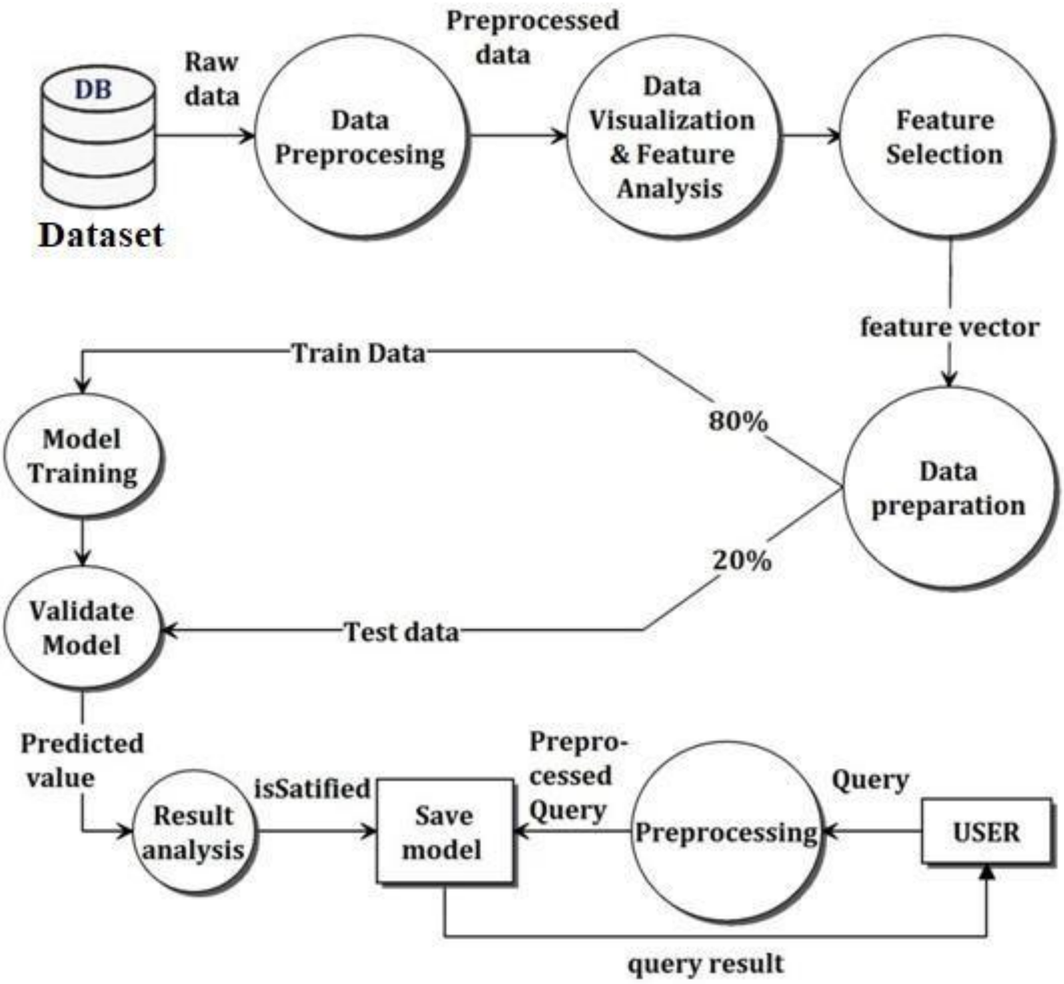


Fig: 4.4 Dataflow Diagram Of Analysis and Prediction

LOW LEVEL DESIGN

During the detailed phase, the view of the application developed during the high-level design is broken down into modules and programs.

INTRODUCTION

During the detailed phase, the view of the application developed during the high-level design is broken down into modules and programs. Logic design is done for every program and then documented as program specifications. For every program, a unit test plan is created. The entry criteria for this will be the HLD document. And the exit criteria will the program specification and unit test plan (LLD).

USE CASE DIAGRAM

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. While a use case itself might drill into a lot of detail about every possibility, a use case diagram can help provide a higher-level view of the system. It has been said before that "Use case diagrams are the blueprints for your system". They provide the simplified and graphical representation of what the system must do.

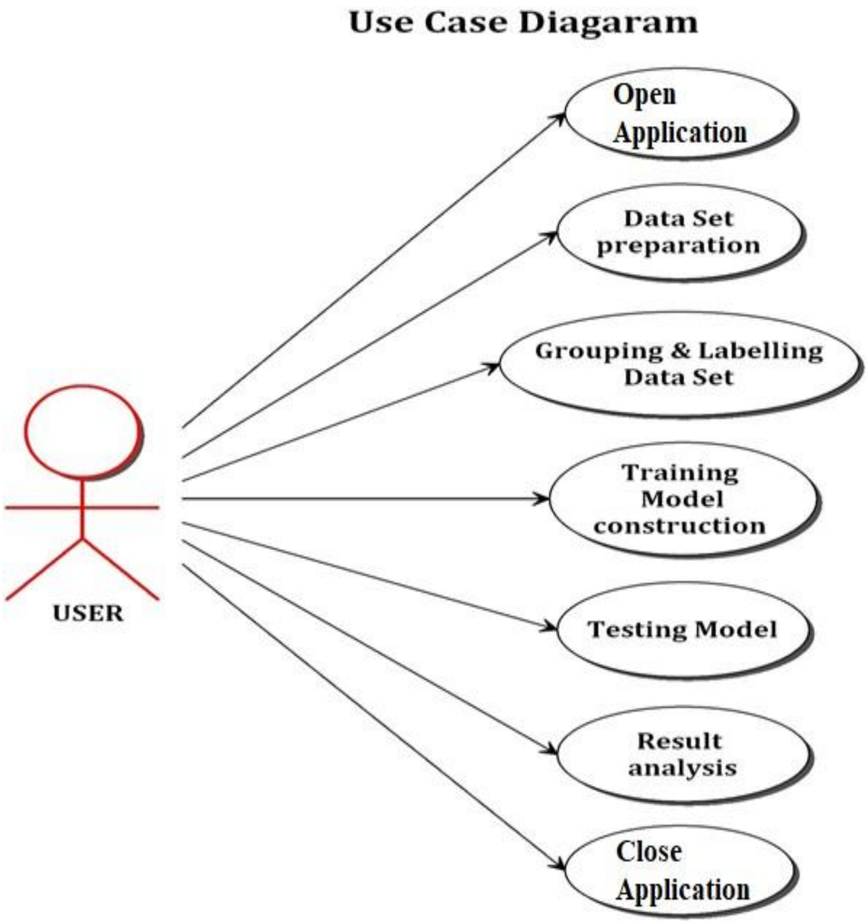


Fig: 4.5 Use Case Diagram

SEQUENCE DIAGRAM

A sequence diagram in a Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It shows the participants in an interaction and the sequence of messages among them; each participant is assigned a column in a table.

Below section shows the sequence diagram in this application.

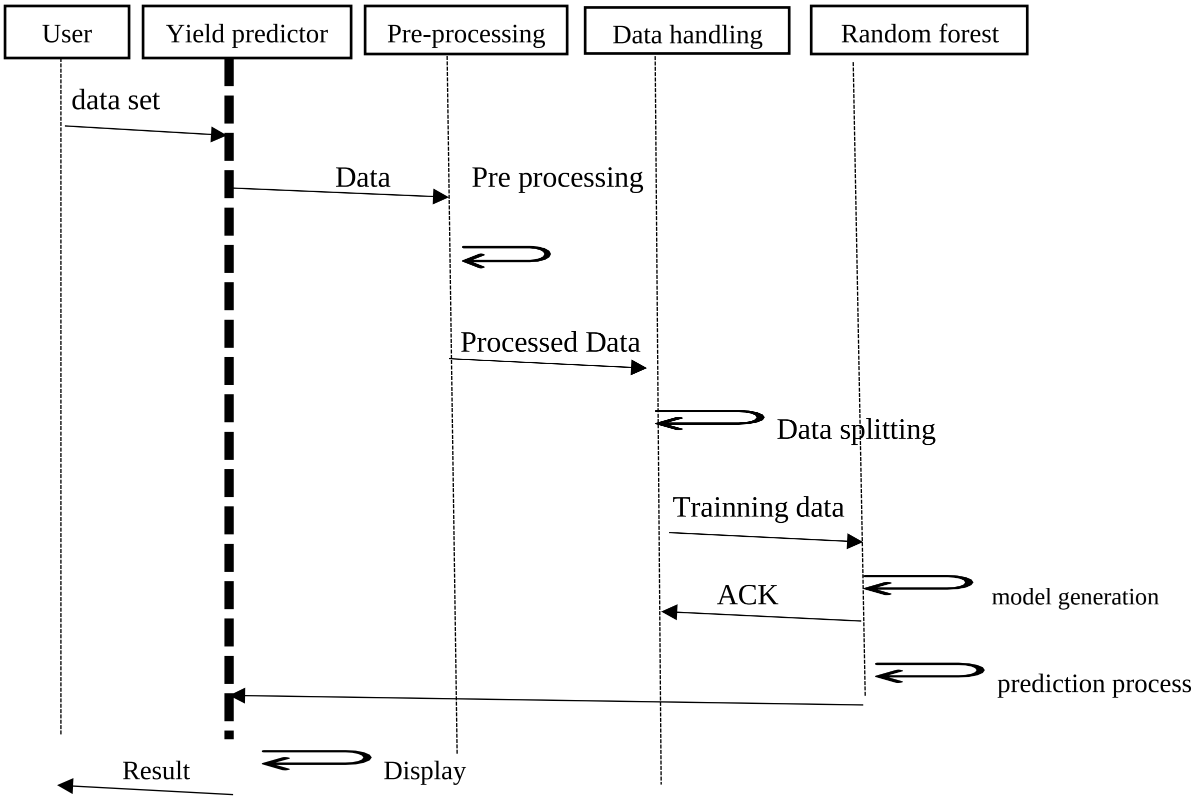


Fig: 4.6 Sequence Diagram

CHAPTER 5

IMPLEMENTATION

Implementation is the carrying out, execution or practice of a plan, a method, or any design, idea, model, specification, standard or policy for doing something. As such, implementation is the action that must follow any preliminary thinking in the order for something to actually happen. Implementations allow the users to take over its operation for use and evaluation. It involves training the users to handles the system and plan for a smooth conversion.

Implementation is a process of ensuring that the information system is

- Constructing a new system from scratch.
- Constructing a new system from the existing system.

HARDWARE IMPLEMENTATION

ARDUINO

The Arduino Uno is a microcontroller board based on the ATmega328. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header, and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started.

The Uno differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it features the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB- to-serial converter.

pinout: added SDA and SCL pins that are near to the AREF pin and two other new pins placed near to the RESET pin, the IOREF that allow the shields to adapt to the voltage provided from the board. In future, shields will be compatible both with the board that use the AVR, which operate with 5V and with the Arduino Due that operate with 3.3V. The second one is a not connected pin, that is reserved for future purposes. At mega 16U2 replace the 8U2. "Uno" means one in Italian and is named to mark the upcoming release of Arduino 1.0.

The Uno and version 1.0 will be the reference versions of Arduino, moving forward. The Uno is the latest in a series of USB Arduino boards, and the reference model for the Arduino platform; for a comparison with previous version

Arduino Uno R3 Front   Arduino Uno R3 Back

Fig: 5.1 ATmega328 Microcontroller

The Arduino Uno can be powered via the USB connection or with an external power supply. The power source is selected automatically.

External (non-USB) power can come either from an AC-to-DC adapter (wall-wart) or battery. The adapter can be connected by plugging a 2.1mm center-positive plug into the board's power jack. Leads from a battery can be inserted in the Gnd and Vin pin headers of the POWER connector.

The board can operate on an external supply of 6 to 20 volts. If supplied with less than 7V, however, the 5V pin may supply less than five volts and the board may be unstable. If using more than 12V, the voltage regulator may overheat and damage the board. The recommended range is 7 to 12 volts.

Features of ATmega328 Microcontroller

Microcontroller	:	ATmega328
Operating Voltage	:	5V
Input voltage	:	12v
Input Voltage (limits)	:	6-20v
Digital I/O Pins	:	14 (of which 6 provide PWM output)

Analog Input Pins	:	6
DC Current per I/O Pin	:	40 Ma
DC Current for 3.3V Pin	:	50 mA
Flash Memory	:	32 KB of which 0.5 KB used by bootloader
SRAM	:	2 KB (ATmega328)
EEPROM	:	1 KB (ATmega328)
Clock Speed	:	16 MHz

Moisture Sensor SA SM01

In this sensor we are using 2 Probes to be dipped into the Soil As per Moisture We will get Analogue Output variations from 0.60volts - 12volts. Soil moisture sensors measure the water content in soil. A soil moisture probe is made up of multiple soil moisture sensors. One common type of soil moisture sensors in commercial use is a Frequency domain sensor such as a capacitance sensor. Another sensor, the neutron moisture gauge, utilizes the moderator properties of water for neutrons.

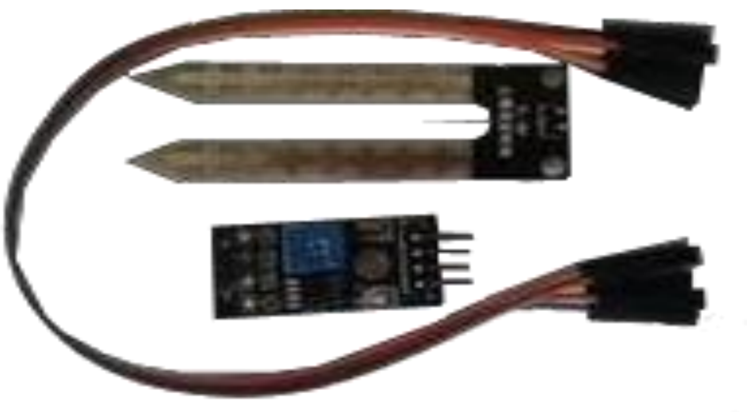


Fig: 5.2 Moisture Sensor SA SM01

Advantages

- Very Sensitive
- Very efficient and accurate
- Analogue and Digital Output

Applications

- Irrigation Applications
- Small- and Large-Scale Industries
- Green House

Humidity sensor

A humidity sensor (or hygrometer) senses, measures and reports both moisture and air temperature. The ratio of moisture in the air to the highest amount of moisture at a particular air temperature is called relative humidity. Relative humidity becomes an important factor when looking for comfort.

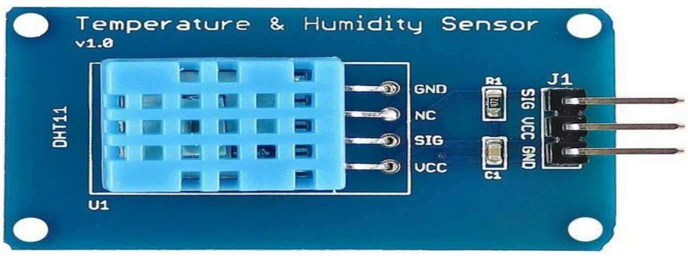


Fig: 5.3 Humidity Sensor

PH / NPK Sensor:

With Three-Way Soil Meter For Moisture, Light Intensity and pH Testing Meter, you can easily check on the condition of your plants. The moisture meter will quickly tell whether your plants are doing well. Firstly, let you know that dry and needs water or wet and could use a day to dry out, you will never over/under water your plants again. Secondly, it helps you to control pH level in soil, acidic or alkaline is suitable for your plants. Thirdly, testing whether plants getting adequate sunlight. Above all, equipped with this meter, you can give the best care to your plants and keep the soil and lawn healthy and happy. Note: Please remove probes from soil and wipe clean after each use.

Soil pH is one of the most important factors that can be overlooked in the garden. pH has impacts on the availability of nutrients and of the plants ability to take them up. If the pH of your garden soil is not in the optimal range for the plants you are trying to grow you may end up having issues. Often plants grown in a soil that does not have the optimal pH dont produce or if they do their harvests are low while the plant may look stressed. On todays joint episode between the Testing Garden Assumptions and Urban Garden Series, I am going to take a look at soil pH, how to easily measure it and how you can adjust the pH over time if need be. pH is measured on a 14 point scale with 0 being the most acidic 7 neutral and 14 the most basic.

Three Way Soil Meter With PH Detector is a Perfect tool to test the soil conditions (Moisture/Light/ PH) of fruits, flowers, vegetables, shrubs, etc. Its an Ideal and necessary tool for gardeners, vegetable and fruit growers, and all those who plant trees and grass. The meter can be Simply inserted into the soil and switch to the setting you want to measure and read the scale.

Features:

1. Three different soil test meters in one device; Measures moisture, pH/Acidity, and light.
2. 100% Accuracy; Easy to read moisture, pH and light levels; Perfect monitor for growing healthy plants.
3. Takes the guesswork out of gardening; Know exactly when to water, amend your soil or adjust lighting.
4. Compact, portable & easy to use, just plug and use; Compact soil meter works indoor/outdoors.
5. Save water, energy and keep your plants, lawn, flower in top condition.
6. Measures moisture at root level.
7. No battery required, simple and convenient to use.
8. Multifunctional : Can test Moisture / Light / pH.
9. Scientifically accurate

SOFTWARE IMPLEMENTATION

The Arduino Uno has a number of facilities for communicating with a computer, another Arduino, or other microcontrollers. The ATmega328 provides UART TTL (5V) serial communication, which is available on digital pins 0 (RX) and 1 (TX). An ATmega16U2 on the board channels this serial communication over USB and appears as a virtual com port to software on the computer. The '16U2 firmware uses the standard USB COM drivers, and no external driver is needed. However, on Windows, a .inf file is required. The Arduino software includes a serial monitor which allows simple textual data to be sent to and from the Arduino board. The RX and TX LEDs on the board will flash when data is being transmitted via the USB-to-serial chip and USB connection to the computer (but not for serial communication on pins 0 and 1).

A Software Serial library allows for serial communication on any of the Uno's digital pins.

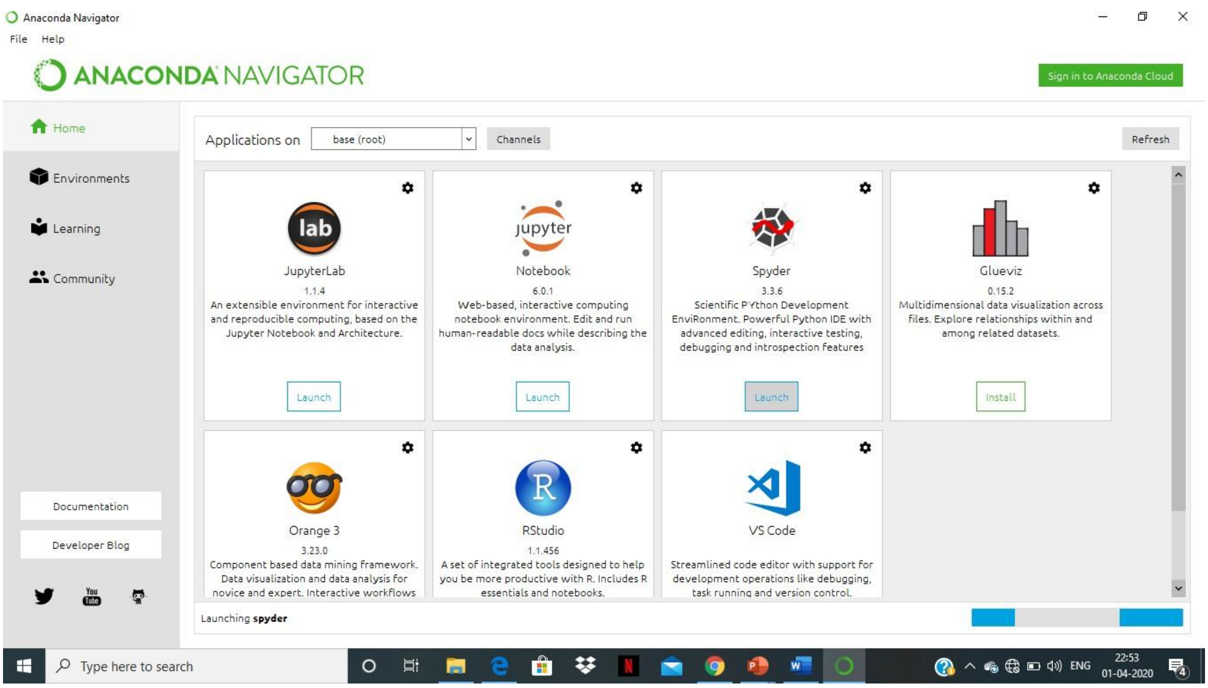
The ATmega328 also supports I2C (TWI) and SPI communication. The Arduino software includes a Wire library to simplify use of the I2C bus; see the documentation for details. For SPI communication, use the SPI library.

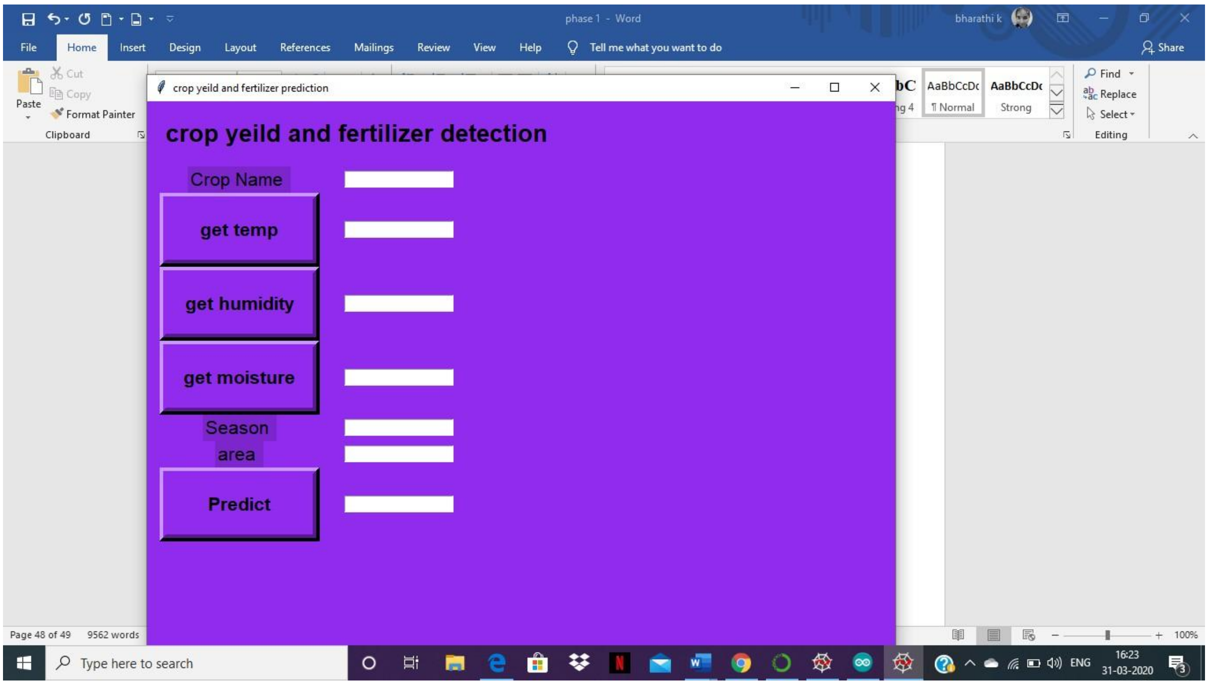
The Arduino Uno can be programmed with the Arduino software. The ATmega328 on the Arduino Uno comes pre burned with a bootloader that allows you to upload new code to it without the use of an external hardware programmer. It communicates using the original STK500 protocol (reference, C header files).

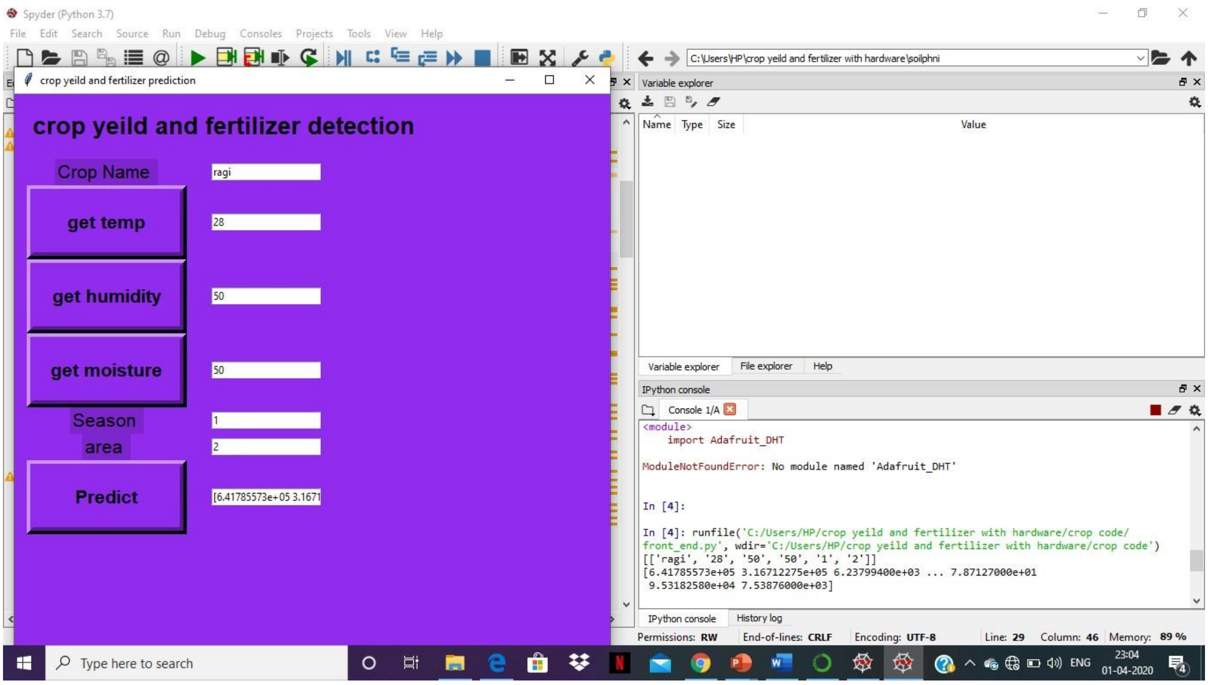
CHAPTER 6

PERFORMANCE ANALYSIS

Analysis is the process of breaking a complex topic or substance into smaller parts to gain a better understanding of it. Analysts in the field of engineering look at requirements, structures, mechanisms, and systems dimensions. Analysis is an exploratory activity. The Analysis Phase is where the project lifecycle begins. The Analysis Phase is where you break down the deliverables in the high-level Project Charter into the more detailed business requirements. The Analysis Phase is also the part of the project where you identify the overall direction that the project will take through the creation of the project strategy documents. Gathering requirements is the main attraction of the Analysis Phase. The process of gathering requirements is usually more than simply asking the users what they need and writing their answers down. Depending on the complexity of the application, the process for gathering requirements has a clearly defined process of its own. This process consists of a group of repeatable processes that utilize certain techniques to capture, document, communicate, and manage requirements




crop yeild and fertilizer prediction
crop yeild and fertilizer detection
Crop Name
get temp
get humidity
get moisture
Season
area
Predict


Spyder (Python 3.7)
crop yeild and fertilizer prediction
crop yeild and fertilizer detection
Crop Name
ragi
get temp
28
get humidity
50
get moisture
50
Season
1
area
2
Predict
[6.41785573e+05 3.1671
Variable explorer
IPython console
import Adafruit_DHT
ModuleNotFoundError: No module named 'Adafruit_DHT'
In [4]: runfile('C:/Users/HP/crop yeild and fertilizer with hardware/crop code/front_end.py', wdir='C:/Users/HP/crop yeild and fertilizer with hardware/crop code')
[['ragi', '28', '50', '50', '1', '2']]
[6.41785573e+05 3.16712275e+05 6.23799400e+03 ... 7.87127000e+01
9.53182580e+04 7.53876000e+03]

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

Agriculture is the backbone of counties like India. However, the usage of technology towards agriculture is to be given paramount importance towards preclusion agriculture. This paper proposes a system which will help farmers to have an idea of yield estimates based on weather parameters and area under cultivation Using this farmer can make decisions on whether to grow that particular crop or go for alternate crop in case yield predictions are unfavorable. This research work can be enhancing to the next level. We can build a recommender system of agriculture production and distribution for farmer. By which farmers can make decision in which season which crop should sow so that they can get more benefit. This system is work for structured dataset. In future we can implement data independent system also. It means format of data whatever, our system should work with same efficiency.

In future developing the web application based on this ideology and make the user use this easily and help the user to understand the yield of the crop, he is going to crop in that season.

REFERENCES

[1] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,DominguezLimaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo, P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).

[2] Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output. 2018 27th Wireless and Optical Communication Conference (WOCC).

[3] Manjunatha, M., Parkavi, A. (2018). Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. 2018 International Conference on Current Trends Towards Con- verging Technologies (ICCTCT).

[4] Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques.2017 1st International Conference on Next Generation Computing Applications (NextComp).

[5] Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C.,Basurto,C. (2015). Crop-planning, making smarter agriculture with climate data. 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics).

[6] Shah, P., Hiremath, D., Chaudhary, S. (2017). Towards development of spark based agricultural information system including geo-spatial data. 2017 IEEE International Conference on Big Data (Big Data).

[7] Afrin, S., Khan, A. T., Mahia, M., Ahsan, R., Mishal, M. R., Ahmed, W.,Rahman, R. M. (2018). Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Di_erent Agricultural Regions of Bangladesh.2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).

[8] Sekhar, C. C., Sekhar, C. (2017). Productivity improvement in agriculture sector using big data tools. 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

[9] Sahu, S., Chawla, M., Khare, N. (2017). An ancient analysis of crop yield prediction using Hadoop framework based on random forest approach. 2017 International Conference on Computing, Communication and Automation (ICCCA).

[10] Garg, A., Garg, B. (2017). A robust and novel regression based fuzzy time series algorithm for prediction of rice yield. 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT).

[11] Raja, S. K. S., Rishi, R., Sundaresan, E., Srijit, V. (2017). Demand based crop recommender system for farmers. 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR).

[12] Dey, U. K., Masud, A. H., Uddin, M. N. (2017). Rice yield prediction model using data

APPENDIX A

IMPLEMENTATION CODE

```
import numpy as np

import pandas as pd

import Adafruit_DHT

import time

import spidev

import os

import busio

import digitalio

import board

import adafruit_mcp3xxx.mcp3008 as MCP

from adafruit_mcp3xxx.analog_in import AnalogIn"""

import pandas as pd

from tkinter import *

from sklearn.preprocessing import LabelEncoder

from sklearn.model_selection import train_test_split

#from sklearn.cross_validation import train_test_split

root = Tk()

root.title('crop yeild and fertilizer prediction')

root.geometry('850x650')

root.configure(background="purple2")
```

```
var = StringVar()

label = Label( root, textvariable = var,font=('arial',20,'bold'),bd=20,background="purple2")

var.set("crop yeild and fertilizer prediction")

label.grid(row=0,columnspan=6)

# In[207]:

data = pd.read_excel('crop_csv_file.xlsx')

data = data[:10000]

data = data.drop(['State_Name','District_Name','Crop_Year'],axis=1)

#handling missing data

data = data.dropna()

le = LabelEncoder()

crop = le.fit_transform(data.Crop)

Season = le.fit_transform(data.Season)

data['Crop'] = crop

data['Season'] = Season

X = data.iloc[:,:-1]

y = data.iloc[:,-1]

X_train,X_test,Y_train,Y_test = train_test_split(X,y,test_size=0.2,random_state=100)

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import roc_auc_score , classification_report, mean_squared_error,
r2_score

forest = RandomForestRegressor(n_estimators=1000,

criterion='mse',
```

random_state=1,

n_jobs=-1)

forest.fit(X_train, Y_train)

y_train_pred = forest.predict(X_train)

y_test_pred = forest.predict(X_test)

def predict():

global forest , X_test

forest.predict(X_test)

y_pred = forest.predict(X_test)

X_data = [[Entry_5.get(),Entry_1.get(),Entry_2.get(),Entry_3.get(),Entry_4.get(),Entry_6.get()]]

print(X_data)

print(y_pred)

Entry_7.delete(0,END)

Entry_7.insert(0,y_pred)def th():

humidity, temperature = Adafruit_DHT.read_retry(11, 4)

h = print ('Temp: {0:0.1f} C Humidity: {1:0.1f} %'.format(temperature, humidity))

time.sleep(1)

h = print ("temp : {0:0.1f} c".format(temperature))

#label_10 = Label(bottomframe,text=h,height=1,bg='violet',font="Helvetica 15 bold").grid(row=10,column=2,sticky=E)

label_pred = Label(bottomframe,text=("Temperature = ",temperature),bg='violet',font="Helvetica 15 bold").grid(row=2, column=6)

```
label_pred = Label(bottomframe,text=("Humidity =",humidity),bg='violet',font="Helvetica 15 bold").grid(row=3, column=6)
entry_2.delete(0,END)
entry_2.insert(0,temperature)
def soil():
spi = spidev.SpiDev()
spi.open(0,0)
def readChannel(channel):
val = spi.xfer2([1,(8+channel)<<4,0])
data = ((val[1]&3) << 8) + val[2]
return data
delay = 0.2
if__name__== "_main_":
try:
val = readChannel(0)
if (val != 0):
print(val)
#print(dir(entry_9))
entry_9.delete(0,END)
entry_9.insert(0,val)
#label_pred = Label(bottomframe,text=("Soil_moisture =",val),bg='violet',font="Helvetica 15 bold").grid(row=6, column=6)
#time.sleep(delay)
```

```
except KeyboardInterrupt:

print ("Cancel.")

def hum():

humidity, temperature = Adafruit_DHT.read_retry(11, 4)

h = print ('Temp: {0:0.1f} C Humidity: {1:0.1f} %'.format(temperature, humidity)

time.sleep(1)

h = print ("temp : {0:0.1f} c".format(temperature)

#label_10 = Label(bottomframe,text=h,height=1,bg='violet',font="Helvetica 15 bold").grid(row=10,column=2,sticky=E)

#label_pred = Label(bottomframe,text=("Temperature = ",temperature),bg='violet',font="Helvetica 15 bold").grid(row=2, column=6)

#label_pred = Label(bottomframe,text=("Humidity =",humidity),bg='violet',font="Helvetica 15 bold").grid(row=3, column=6)

Entry_3.delete(0,END)

Entry_3.insert(0,humidity)

label_5 = Label(root, text ='Crop Name ',font=("Helvetica", 16),background="Purple3")

label_5.grid(row=1,column=0)

Entry_5 = Entry(root)

Entry_5.grid(row=1,column=1)

label_1 = Label(root, text ='rainfall ',font=("Helvetica", 16),background="Purple3")

#label_1.grid(row=1,column=0)

Entry_1 = Entry(root)

Entry_1.grid(row=2,column=1)
```

```
B1 = Button(root, text = "get temp",height=1,padx=16,pady=16,bd=8,font=('arial',16,'bold'),width=10,bg="purple2",command=th)

B1.grid(row=2,column=0)

label_2 = Label(root, text ='water level',font=("Helvetica", 16),background="Purple3")

label_2.grid(row=2,column=0)

Entry_2 = Entry(root)

Entry_2.grid(row=3,column=1)

B2 = Button(root, text = "get humidity",height=1,padx=16,pady=16,bd=8,font=('arial',16,'bold'),width=10,bg="purple2",command=hum)

B2.grid(row=3,column=0)

label_3 = Label(root, text ='water flow',font=("Helvetica", 16,),background="Purple3")

label_3.grid(row=3,column=0)

Entry_3 = Entry(root)

Entry_3.grid(row=4,column=1)

B3 = Button(root, text = "get moisture",height=1,padx=16,pady=16,bd=8,font=('arial',16,'bold'),width=10,bg="purple2",command=soil)

B3.grid(row=4,column=0)

label_4 = Label(root, text ='Season ',font=("Helvetica", 16),background="Purple3")

label_4.grid(row=5,column=0

Entry_4 = Entry(root)

Entry_4.grid(row=5,column=1)
```

```
label_6 = Label(root, text ='area ',font=("Helvetica", 16),background="Purple3")

label_6.grid(row=6,column=0)

Entry_6 = Entry(root)

Entry_6.grid(row=6,column=1)

Entry_7 = Entry(root)

Entry_7.grid(row=7,column=1)

B4 = Button(root, text = "Predict",height=1,padx=16,pady=16,bd=8,font=('arial',16,'bold'),width=10,bg="purple2",command = predict)

B4.grid(row=7,column=0)

root.mainloop()
```

hardware code

```
#include <Wire.h>

#include <dht.h>

dht DHT;

#define DHT11_PIN 5

int moist=A0;

void setup() {

Serial.begin(9600);

pinMode(moist,INPUT); }

void loop() {

humidity();

moist_read();
```

```
}
void humidity()
{
int chk = DHT.read11(DHT11_PIN);
Serial.println("Temp: ");
Serial.println(DHT.temperature);
Serial.println((char)223);
Serial.println("C");
Serial.print("Temp:");
Serial.println(DHT.temperature);
Serial.println("Humidity: ");
Serial.println(DHT.humidity);
//  Serial.print("%");
//  Serial.print("RH:");
// Serial.println(DHT.humidity);
delay(2000);
}
void moist_read()
{
//lcd.setCursor(0,0);
int val = 0,M;
char buff[50];
```

```
val = analogRead(A0);

// M=(val * 3.3)/1024.0;

M=1023-val;

sprintf(buff,"%d",M);

Serial.println("Moisture:");

Serial.print(buff);

delay(2000);

}
```

Crop Yield and Fertilizer Prediction Using Machine Learning

RAJASHREE INGALE [1], ASHWINI A NARAYAN [2], ANUSHA R[3], BHARATHI K [4], NAGASHREE S RAO[5]

[1] *Assistant Professor,* [2,3,4,5] *UG students*

Department of Information Science and Engineering, AMC Engineering College, Bangalore-82

email- [1] *di.rajashree@gmail.com,* [2] *ashwini.shrivalli@gmail.com,* [3] *anushargowda31@gmail.com,* [4] *sarasammabharathi626@gmail.com ,* [5] *nagashreerao29@gmail.com*

Abstract: India is an agricultural based business nation, its financial health mainly depends upon agriculture yield growth extension and agro industry goods. Data mining is a blooming field in crop yield inspection. Yield presumption is a notable issue in agricultural area. Any farmer is willing to know how much yield he is going to get. Dividing the different dependent properties like area, pH esteem from which alkalinity of the dirt is predicted. Percentage of additives like Nitrogen (N), Phosphorous (P), and Potassium (K) Geographic location is used along with the utilization of external applications like APIs for weather and temperature, variety of soil, nutrient value of the soil in that region, aggregate rainfall measure in the region, soil composition can be determined. Each information will be divided separately and this information will be trained using suitable machine learning algorithms to construct a model. System has a model that accurately determines the crop yield and provide the client with proper recommendations about the amount of the fertilizers required dependent on the kind of soil, crop and other atmospheric conditions which will help the farmers to reap healthy crops and increases his income as well as country's economy.

Keywords: IoT, humidity and moisture sensors, microcontroller etc...

INTRODUCTION

India is a populated nation and unusual arbitrary changes in the atmospheric conditions as to protect the world natural resources. Randers face major problems during the dry season. Sort of soil performs important role in crop yield.

Recommending the proper use of manure will help the farmers to take the correct decision during that particular period.

The number of studies Information and Communication Technology (ICT) can be utilized for harvest yield prediction. By using Data Mining approach, crop yield can be determined. By examining the past information, we can suggest the farmer the best crop to grow in during particular season for the better outcome.

for the better yield we need to inspect the kind of soil and soil fertility, rainfall which is one of the major element and groundwater obtainability suppose it is bare land it is good to go for money crops and if suppose it is moist land it is smart to go for wheat and sugarcane. there are 15 agro climatic areas in India these areas are split on the kind of land.

Each agro climatic areas can develop some particular crops. Depending on that we have to recommend the farmer that which crop is best suited to plant in his farm to that particular area during that current season. The main aim of the project is to get more crop at less yield. Timely observation of the mistakes will help to gain more crop yield. Crop yield prediction is the important sector which helps to protect food resources. for the perceiving the further information about the crop yield, we have to learn bulk data with the help of machine learning algorithm so that will give the precise yield for that crop and advise the farmer for the best crop. Increasing the crop yield is the ultimate goal of precision agriculture using information technology methods. Yield profit and sustainability are major goal of precision agriculture.

Agriculture is the backbone of Indian economy since ages. In recent times, due to rapidly fluctuating atmospheric conditions it has become difficult to decide which crop has to be grown. By using some artificial technologies, we have to find the crop information and suggest the farmers to grow best crop accordingly and in addition to this fertilizer is one of the most important part to grow crop. Fertilizer must be used in the right quantity otherwise the soil loses its productivity, and we may not get the expected crop yield. Agriculture is most important for Indian economy because it has the largest contribution in the India's GDP so crop yield prediction is required for India.

Usually, supervised learning algorithm approach will determine the efficient crop yield for that farm field. In the ancient methods crop yield was determined on the farmers past experiences but now due to rapid changes in the weather conditions crop yield cannot be predicted. Technology will help the farmers to decide whether to grow that crop or not if in case yield predictions are unfavourable. Machine learning model will understand the crop pattern and yield based on certain conditions and determines the crop yield. The aim is to develop the systematic model to determine the better model to determine the output therefore we will try with the various algorithms and compare them and the model with minimum faults will be selected to determine the particular crop for that region.

2. RELATED WORK

Niketa et al stated that crop yield depends on the climate. Climatic conditions in India fluctuate very rapidly. Due to scarcity of water, Randers come across critical problems. By seeing all this machine learning algorithms were used by them in order to help the farmers to recommend the crop for their farm for the better yield. Different kind of data will be taken to predict the upcoming data. SMO classifiers in WEKA was used by them to arrange the results. Minimum temperature, maximum temperature, average temperature and past year's crop and yield data are the important elements that was taken in to account. Past data was divided into high yield and low yield by using SMO tool. The determined result for the crop yield prediction using SMO classifier gives less incorrect when compared to naïve Bayes, multilayer perceptron and Bayesian network.

Eswari et al 2018 have point out that main factors for crop yield prediction are perception, average, minimum and maximum temperature. Besides all this one more factor is taken into account named crop evapotranspiration. It refers to both evaporation from soil and vegetative surfaces and transpiration from plants. Dataset with this attribute was collected and given as an input to the Bayesian network and it is classified into two classes namely, true and false and output is compared with the classifications in the model using confusion matrix and determine the correctness. At last, they commenced that crop yield prediction with Naive Bayes and Bayesian network give high exactness than SMO classifier and foretelling the crop yield prediction in various kind of atmosphere and cropping situations will be advantageous.

Shruthi Mishra et al 2018 have shown that by executing the data mining techniques on previous climate and crop yield data few anticipations are made which maximises the crop yield. Decision support system is made use to help the farmers to get an idea about the suitable crop and its yield. Dataset with the attributes of the crop season, region, and production in hectares is collected and various algorithms in WEKA is used to analyse them. Data was analysed with four methods J48, IBK, LAD tree, LWL in WEKA and spotted their correctness and compared with each other.

Chlingaryana et al 2017 shown that nitrogen quantity in the soil is the major element in the soil in the crop yield prediction. In recent years in decision making remote sensing systems are used generally. These data which will help the farmers to have idea of yield estimates and increase the yield. To maximize the crop yield nitrogen supplements are the important factor which also improves the soil richness. Machine learning algorithms are used to make the choice. Nitrogen, kind of soil, yield prediction of past data are the important factors taken into account to predict the crop yield. In recent days precision agriculture is made use to increase the yield. Information technology is used to certify the soil and crop. How to improve the manufacturing and condition of the soil is told by the information technology. The secured output is back propagation neural network is made use to obtain various flora events. The conventional neural network of long-term memory is used to determine the future information.

Dakshayini Patil at all 2017 shown thatrice crop plays a significant role in the economy. They employed different information mining technique to foresee the rice crop yield. Rice crop is the long-lasting security of India. It offers 40% to the general yield. Maximum return of the yield depends on the particular area atmospheric conditions. Learning well informed master plan as per the climatic conditions can improve the harvest. Reports use different mining technique based on past information of crop yield and dissimilar climatic areas. Data of 27 areas of Maharashtra are utilized to determine the crop yield.

Weighted commitment to its yield approving. One sort of framework reviews the hub to be as "fake neurons". These are called neural frameworks. The back-provoking computation (Rumelhart and McClelland, 1986) is made use in layered feed-forward ANNs. This suggests the artificial neurons are arranged through in layers and send their signs "forward", and later from that point onward, the bungles are outspread backward. The framework gains commitments by neurons in the data layer, and organised yield is given by the neurons on a yield layer. There may be at least one widely appealing covered layers. This neural arrange building is utmost popular, since it is associated with a broad scope of attempts. The superior term, "feed forward" represents how this neural arrange strategies and evaluate plan. In a feed-forward neural framework, neurons are simply connected with forward. Each layer of the neural framework consists link with backing layer (for example, from the commitment to the concealed layer), yet there is no connection back. The expression back spread characterize how this sort of neural framework is developed. The articulation back spread is a kind of control developing. While using control developing procedure, the framework should be provided with both model input data and predicted yields. The real yield is compared with predicted yield. Using the predicted yields, the back-expansion formulating computation at that tip grabs a anticipated error and modify heaps of the different layers backward from the harvesting layer to the information layer.

3. PROPOSED SYSTEM METHODOLOGY

Determine the crop yield by utilizing the effective algorithms and recommend how much amount of fertilizer must be used to gain the maximum yield of the crop.

I. **Data Set Description:** This is the training example collection utilized in this work. The information in Table I is information utilized to anticipate crop yield conditional on 7 components. These 7 components are state, region, crop, block, season, formation using this data we can design an AI model and train the model and we can foretell the formation and from Table II we can foresee the measure of manure to be utilized to gain the right yield information elements are the amount of nitrogen, phosphorus, and the result is the quantity of the manure to be utilized. Here in the info parameters 1, 2, 3, 4, 5, 6 constitutes to the very high, high, better than expected, beneath expected, low and extremely low amount present in the soil individually.

TABLE I SAMPLE DATA SET OF CROP DATA

STATE	DISTRICT	CROP YEAR	SEASON	CROP	AREA
Bihar	Gaya	2013	Rabi	rapeseed	1086

Bihar	Gaya	2013	Rabi	sunflower	1394
Bihar	Gaya	2013	Rabi	wheat	58783
Bihar	Gaya	2013	summer	maize	1165
Bihar	Gaya	2013	summer	moong	3856
Bihar	Gaya	2013	Whole year	sugarcane	459

TABLE II SAMPLE FERTILIZER DATA

N	P	K	Amt of n	Amt of p	Amt of k
3	5	6	64	50	60
1	4	2	40	46	30
5	1	5	93	16	32
3	1	3	63	20	39
4	6	6	87	37	39
2	1	1	65	19	32

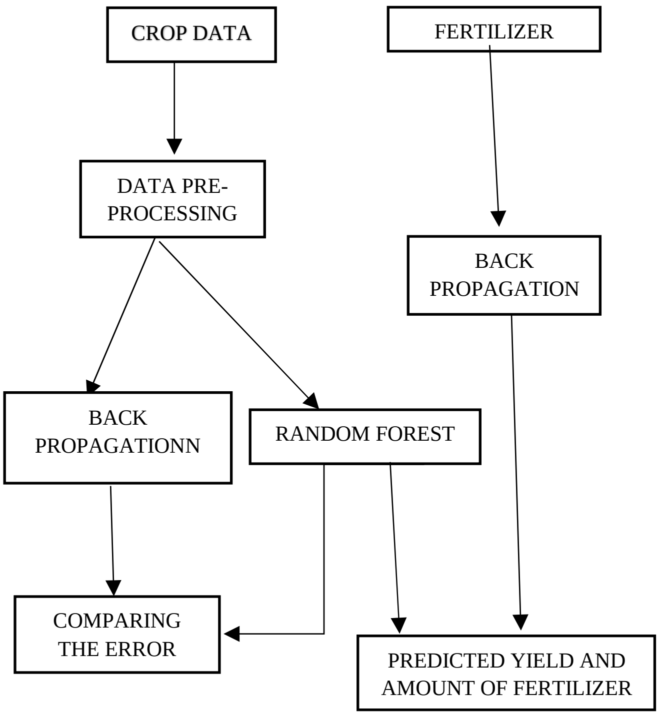
II. Necessary Packages

- x Numpy
- x Pandas
- x Matplotlib.pyplot
- x Scikit-learn
- x TensorFlow
- x Jupyter

Stock the information associated with the yield and manure in the csv which contains of the state, district, crop year, season, crop, area, production , and another training data include phosphorous level, level of potassium, level of nitrogen in the soil, how much quantity of phosphorous, potassium, nitrogen should be utilized to maximise soil richness.

III. Architecture

The Fig. 1 shown below represents the architecture diagram.



METADATA: All the major information utilized in the training information are set up with the number to use in the computation it is similar like setting up all the information. Here, we will initialize all the yield names with the numbers. This information makes us utilize the information productively in the computation. Here the metadata of the total number of yields is given with fixed number. This number isn't equivalent that is one number is unique to one yield, identical number isn't given to the other harvest. This metadata contains surplus of a hundred harvests that spread all over India.

IV. **Data Pre-processing:** Here the natural information in the harvest information is polished and the metadata is attached to it by clearing the things which are modified to the whole number. So, the data can be effortlessly trained. Right now, we first deliver the metadata into this and after that this metadata will be joined to the information and substitute the transformed information with metadata. Then this information will be moved further and clear the unfavourable information in the list and it will partition the information into the train and the test data. For this dividing of the information into train and test we should import train_test_split which in the Scikit-learn this will support the pre-operated data to divide the information into train and test as per the offered weight given in the code. The partition of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent individually.

V. **Random Forest Algorithm:** This calculation fits for both gigantic and tiny information to give a systemized expectation. Conditional on the offered information to the calculation it creates different choice trees and examine for what number of trees give a similar anticipation. It is determined on the votes it will enumerate and which trees give a similar result after that the yield given by the more trees it will seem as yield as described in Fig. 2. The given information in the task go to the random forest calculation and here it will build ten trees and move information to it. Each tree is arranged depending on the different circumstances and it will instruct the model as per to it and will sum up the number of trees which gives the similar results and which tree has maximum number it will be concluded as the output.

Fertilizer utilization using Back-propagation: The Fertilizer information set which is accessible in the CSV arrangement is pre-operated and made ready to develop the model with that dataset. Firstly, the data collection is partitioned into 80% for composing information and 20% for the test information. The calculation used to produce the dataset into a model is the Backpropagation calculation. The backpropagation algorithm is the scheme from the several layer perceptron in the artificial neural system. The backpropagation calculation is utilized for gigantic datasets which have no right connections between the credits of the dataset to shape a system model via developing the dataset and determining the yield. This algorithm contains three layers in the system model, they are the information layer, concealed layer, and yield layer. Info layer in the model is mainly subject for giving the input data to the model, after that the secret layer which is in the middle of the information layer and yield layer and this is capable to gain the result from the information layer as input and compute as per the loads residing on the contribution to invisible layer and gives the favorable yield output and the final layer is yield layer which gives the yield determined from the framework model. Backpropagation algorithm is a directed learning computation. To develop a dataset in a back-propagation algorithm it must have the favorable yield attribute in the dataset. Back propagation algorithm is instructed in such a manner we attach the result value to the dataset that is to be prepared, at first repetition result is calculated and notice the dissimilarity between the anticipated output and observed output depending on that inspection it back spreads the mistake and increment the load in middle of nodes and bias. The network is hence prepared with many repetitions until it gains the favourable result. After coaching the network model then it's verified and stimulated by the test dataset whether the determination is error-free or not. After verification of the network model, we can determine by giving the unspecified data and anticipate the result to the unambiguous data given to the model.

Method:

1. First, to coach the network we set up the loads and bias.
2. Reviewing all the credits x from the dataset d calculate the result for each block in the network.
3. Then it back spreads the fault in the web

   For each result from the web as k, compute its fault term ∂k

   $$\partial k = Ok\ (1-Ok)\ (t-Ok)$$

   For each hidden unit h, calculate its error term ∂h

   $$\partial h = Oh\ (1-Oh) \sum Wkh\partial k$$

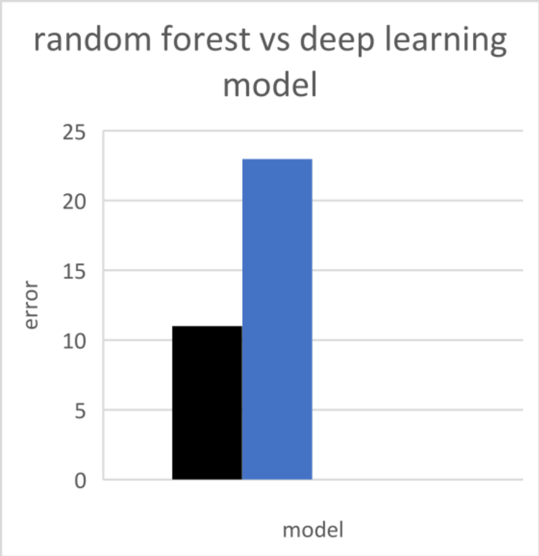
   Update each network weight

   $$Wji = Wji+\Delta Wji$$

   Where, $\Delta Wji = \eta\ \partial j\ xji$

IV. RESULTS

In this paper, attempt is made orderly to know the yield manufacturing inspection and is operated by applying both the Random Forest calculation and Backpropagation calculation. These models were investigated with various kind of crops in different regions across India to anticipate the harvest. Indeed, even manure information was developed utilizing the back-propagation calculation and estimated to acquire the output of how much nitrogen, phosphorus is necessary for the sector of land. Both the models for the crop creation were examined in expecting the yield and by different parameter concerning the fault rate. We inspected the mistake rate gained while focusing at the random forest calculation and backpropagation where we obtained the misconception rate lesser to the random forest than back propagation while determining the yield for both of the models and the solicitude is plotted in the chart Fig. 3. For prediction of the yield, the client will enter the data as indicated in Fig. 4. The client must have to enter the particulars consecutively. The result of harvest anticipation is indicated in Fig 5. The input of the compost information is given as shown in fig.6 and output of the fertiliser information is shown in Fig. 7.



-Random forest

-Deep learning

Fig. 3. Comparison between two models respective to their error rate

Enter state west Bengal

Enter district:

Fig.:4 User input-1

Enter state: west Bengal

Enter district: PURILIA

Enter year: 2014

Enter season: rabi

Enter crop: wheat

Enter area: 320

Production: 7065.322

Fig: 5 Output for the given user input-1

Enter nitrogen:

Enter phosphorus:

Fig: 6 User input-2

Enter nitrogen:

Enter phosphorus

Enter potassium

Amount of nitrogen fertilizer:45.02

Amount of phosphorus:28.09

Amount of potassium:35.10

Fig: 7 Output for the above user input-2

V. CONCLUSION AND FUTURE ENHANCEMENT

Crop yield prediction and efficient utilization of fertilisers is victoriously anticipated and effective algorithm is also established from both the algorithm and acquired the great efficient result of the harvest. In future enhancement the web based applications build on this principles will make the client utilise this effortlessly and help the client to get an idea of the yield of the crop that he is going to plant in his farm during that season.

ACKNOWLEDGEMENT

The acquired result will be beneficial for the ranchers to know the Yield of the harvest so, he can go for the better harvest which gives high return and also guide them the accomplished utilization of compost with the aim that he can utilize just the necessary quantity of manures for that farm. By all these we can support the ranchers to sow the best crop which offer them greater yield.

REFERENCES

[1] Niketa Gandhi et al," Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques", IEEE International Conference on Advances in Computer Applications (ICACA), 2016.

[2] K.E. Eswari. L.Vinitha. "Crop Yield Prediction in Tamil Nadu Using Bayesian Network ", International Journal of Intellectual Advancements and Research in Engineering Computations, Volume-6, Issue-2, ISSN: 2348-2079.

[3] Shruti Mishra, Priyanka Paygude, Snehal Chaudhary, Sonali Idate "Use of Data Mining in Crop Yield Prediction" IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number: CFP18J06DVD , ISBN:978-1-5386-0806-7.

[4] Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb — Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture 151 (2018) 61–69, Elisver, 2018.

[5] Dakshayini Patil et al,"Rice Crop Yield Prediction using Data Mining Techniques:An Overview", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 7, Issue 5, May 2017. [6] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015.

[7] K. Menaka, N.Yuvaraj,Indian Journal of Innovations and Developments Vol 5(12),

October, 2016 ISSN (online) : 2277-5390 ISSN (Print): 22775382 [8] Machine Learning by Tom M. Mitchell, ISBN: 0070428077

[9] Snehal S.Dahikar, Dr.Sandeep V.Rode ,"Agricultural Crop Yield Prediction Using Artificial Neural Network Approach ", International journal of innovative and research in electrical, instrumentation and control engineering, volume 2,Issue 2,2014.

[10] D Ramesh, B Vishnu Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of research in engineering and technology, eISSN:2319-1163.

[11] Ramesh A. Medar, "A survey on data mining techniques for crop yield prediction", International Journal of advance in computer science and management studies, ISSN: 2231-7782, volume 2, Issue 9, 2014.

[12] Pooja M C et al, ―Implementation of Crop Yield Forecasting Using Data Mining , International Research Journal of Engineering and Technology (IRJET),2018,p-ISSN: 2395-0072

[13] Jig Han Jeong et al , “Random Forests for Global and Regional Crop Yield Predictions ”,PLOS-ONE DOI:10.1371/journal.pone.0156571 June 3, 2016.

[14] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, Analysis of agriculture data using data mining techniques: application of big data, Springer journal,2017.

[15] S.kanaga Subba Raju et al, Demand based crop recommender system for farmers, International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2017.

[16] Shriya Sahu et al," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach, International Conference on Computing, Communication and Automation , 2017.

[17] E Manjula and S. Djodkitachoumy International journal of computation Intelligence and informatics,Vol 6:No 4 March 2017.