Estimate the Crop Yield using Data Analytics

CHAPTER 1

INTRODUCTION

Data Analytics

Data Analytics refers to the techniques used to analyze data to enhance productivity and business gain. Data is extracted from various sources and is cleaned and categorized to analyze various behavioral patterns. The techniques and the tools used vary according to the organization or individual.So,if you understand your Business Administration and have the capability to perform Exploratory Data Analysis, to gather the required information, then you are good to go with a career in Data Analytics.Data analytics is important because it helps businesses optimize their performances. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services.

Once the data that's needed is in place, the next step is to find and fix data quality problems that could affect the accuracy of analytics applications. That includes running data profiling and data cleansing tasks to ensure the information in a data set is consistents, errors and duplicate entries are eliminated. Additional data preparation work is done to manipulate and organize the data for the planned analytics use. Data governance policies are then applied to ensure that the data follows corporate standards and is being used properly

What are the tools used in Data Analytics?

With the increasing demand for Data Analytics in the market, many tools have emerged with various functionalities for this purpose. Either open-source or user-friendly, the top tools in the data analytics market are as follows.

- **R-programing**– This tool is the leading analytics tool used for statistics and data modeling. R compiles and runs on various platforms such as UNIX, Windows, and Mac OS. It also provides tools to automatically install all packages as per user-requirement.
- **Python**– Python is an open-source, object oriented programing language that is easy to read, write, and maintain. It provides various machine learning and visualization libraries such as TensorFlow,Pandas, Keras, etc. It also can be assembled on any platform like SQL server, a MongoDB database or JSON
- **SAS**– A programming language and environment for data manipulation and analytics, this tool is easily accessible and can analyze data from different sources.
- **Microsoft Excel** – This tool is one of the most widely used tools for data analytics. Mostly used for clients' internal data, this tool analyzes the tasks that summarize the data with a preview of pivot tables.

Types of Data Analytics

Data analytics is broken down into four basic types.

1. **Descriptive analytics:** This describes what has happened over a given period of time. Have the number of views gone up? Are sales stronger this month than last?

2. **Diagnostic analytics:** This focuses more on why something happened. This involves more diverse data inputs and a bit of hypothesizing. Did the weather affect beer sales? Did that latest marketing campaign impact sales?
3. **Predictive analytics:** This moves to what is likely going to happen in the near term. What happened to sales the last time we had a hot summer? How many weather models predict a hot summer this year?
4. **Prescriptive analytics:** This suggests a course of action. If the likelihood of a hot summer is measured as an average of these five weather models is above 58%, we should add an evening shift to the brewery and rent an additional tank to increase output.

Agriculture

Agriculture forms the basis for food security and hence it is important. In India, the majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance.

In India, wheat and rice are the major crops grown along with sugarcane, potatoes, oil seeds etc.Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture.This domain provides employment to more than 60% of the total population and has a contribution to GDPalso (about 17%).In the farm output, India ranks second considering the worldwide scenario.

This is the widest economic sector and has an important role regarding the framework of the socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature,irrigation,cultivation, soil, rainfall, pesticide and fertilizers. Historical information regarding crop yield provides major input for companies engaged in this domain. These companies

make use of agriculture products as raw materials, animal feed, paper production and so on. The estimation of production of crops helps these companies in planning supply chain decisions like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield.

Areas in Agriculture that uses data analytics

1. Receiving Useful Data to Help Fight Food Scarcity and Empower Small Farmers

Since data scientists have tools to process and analyze gigantic amounts of data efficiently, projects are underway to determine how that information might help small-scale farmers join in the battle to solve worldwide food shortages.

In September 2018, a coalition launched a project that will run through 2030 and look at data from approximately 500 million farmers in impoverished areas from 50 countries.The people behind the project hope the data will show whether agricultural investments in various countries are paying off and help develop policies for the farmers. On a larger scale, this project aligns with the United Nations' Sustainable Development Goals to double the agricultural productivity and incomes of farmers in developing nations and help them reduce world hunger.

2. Managing Crop Diseases and Pests

Agricultural pests can quickly cut into a farmer's profits. But, misusing pesticides can have adverse effects on people, plants and other living things. Fortunately, some companies recruit data scientists to help them develop user-facing platforms that analyze when to apply pesticides and how much to use.

One of them is a Brazilian company called Agrosmart. Its technology relies on Internet of Things (IoT) sensors and artificial intelligence to determine the kind of insects on a crop and the quantity present. Farmers then get an associated report and can use it to plan their pest management approaches. The goal is to help farmers cost-effectively control pests with a minimized environmental impact.

In another case, Saillog, an Israeli startup, developed a smartphone app called Agrio that informs farmers of the diseases currently affecting their crops or ones found on surrounding farms.

3. Investigating Agricultural Niches

Data scientists know how to use tools that identify patterns and relationships that may otherwise remain hidden. As such, they can draw conclusions that push agricultural science forward through the examination of specific factors. For example, researchers know trace minerals positively affect the metabolic functions of livestock and poultry, while carotenoids play a role in increasing egg yolk quality and nutrition. The findings brought about by sifting through databases and studies to conclude things like these show how seemingly small factors in agricultural processes can bring about substantial changes.

4. To Cope With Climate Change

Climate change is a looming concern that has already affected the agriculture sector. However, data scientists are hard at work figuring out ways to compensate for the shift.

One project involves giving IoT sensors to Taiwanese rice farmers so they

can collect crucial information about their crops. It'll all go into a database used to help farmers optimize their production cycles, even when climate change makes that task exceptionally challenging. Following the traditional farming calendar is no longer sufficient because of climate change. But, data analysis could forever change the future of farming.

Scientists are also scrutinizing agricultural soil data to improve their understanding of how soil contributes to climate change by releasing greenhouse gasses, as well as how soil data might aid in adapting to climate change. Collecting this kind of information is tricky, but scientists believe it could fill in knowledge gaps associated with the relationship between soil and climate change.

5. To Make Yield Predictions

A poor yield can result in a devastating season for farmers, as well as all the entities that depend on the crops. IBM has a platform that estimates corn yields 2 to 3 months in advance, reducing unpleasant surprises for agricultural professionals.

CHAPTER 2

LITERATURE SURVEY

1] Crop Monitoring and Crop Yield Prediction Computerized Tools in Mexico

Crop yield prediction is an information service provided by the National Institute Research for Forestry Agriculture and Livestock (INIFAP) to the Ministry of Agriculture in Mexico where it is used as a decision making aid. A comprehensive set of state of the art tools has been developed which is currently providing accurate results. The tools consist mainly of prediction models with input data taken on site at sample plots as well as remotely sensed data. The different levels of government agencies –federal, state and municipal– are already benefiting from timely information about the condition of agricultural crops, including the relationship with climatic effects and other potentially adverse factors and the impact on crop production.

The general aims of this information service project are:

- The timely provision of information from the production plots as it relates to the general agricultural regional productivity
- To develop a decision support system of agricultural related activities to support government farmer aid programs.

The benefits of this Project are:

- Timely response and reliability of the information,
- Reliable decision support,
- More accurate and complete information
- Production risk reductions
- Savings of government funds due to a better definition of required import volumes for crops.
- increased confidence concerning the requirements of import volumes

- Increased efficiency in crop management and production
- Availability of production volumes ahead of harvest.

2] Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen

Agriculture is the backbone of India which indirectly contributes to the Indian economy. Farmers who are the drivers of agriculture are facing a lot of problems for proper identification of the crops that can be cultivated for the specific soil conditions and to maximize the crops yield. All these problems are due to lack of technology and scientific techniques being used in agriculture. Crop yield varies as a result of variations in atmospheric and soil conditions. Data mining mainly focuses on methods to elicit useful knowledge from the dataset. There are several data mining approaches that can be used for the purpose of predicting crops yield and finding association among attributes contributing for the crops yield. This paper mainly intensifies on various association algorithms, namely Apriori, Eclat, and AprioriTid to find the association among temperature, rainfall, soil pH, soil nitrogen, and paddy yield.

Methodology

- Data Collection
- Data Preprocessing
- Applying Association Rule Mining Algorithm
- Apriori Approach
- Eclat Approach
- AprioriTid Algorithm
- Analyzing Association Rules

This work can be extended for predicting the type of crop that can be cultivated in a particular plot for high yield. Also this work can be enhanced to predict the better association rule by considering additional attributes like Phosphorus,

Carbon, and Zinc. It can be applied to more agricultural crops of various locations. More number of association algorithms can be implemented on the agricultural dataset.

3] Soil Analysis and Crop Prediction

Soil analysis is an important process to determine the available plant nutrients in the soil. Plants absorb the major nutrients through soil. In addition to soil, there are various major factors like rainfall, precipitation, fertilizer, etc that affect plant growth. Our aim is to create a prediction engine for the most suitable crop for a particular soil. As an initial step, we have focused on predicting the accurate crop yield to the user by analyzing the soil fertility and rainfall in the region entered by the user as an input.

Methods

All the data analyzed is continuously monitored,displayed and uploaded on the IoT cloud. Thingspeak provides a precise and accurate display of temperature and moisture data.The sensor accuracy and range is also taken care of, will collect respective data.Naive Bayes and K-NearestNeighbour (KNN) is performed .

In this project analysis of soil based on Temperature and Soil Moisture has been proposed using Arduino, Cloud Computing. The project has high efficiency and accuracy in fetching the live data of temperature and soil moisture. The project will assist the farmers in increasing the agriculture yield and take efficient care of food production as the stick will always provide a helping hand to farmers for getting accurate live feed of environmental temperature and soil moisture with more than 99% accurate results.

The project proposes a wise agricultural model in integration with IoT. IoT has always mattered in the Agriculture domain.It is a really challenging task

because of the highly localized nature of agriculture information specifically distinct conditions.Also only use soil to predict the crop yield which may not be applicable in many situations.

4]A Smart Agricultural Model by Integrating IoT, Mobile and Cloud-based Big Data Analytics

-S. Rajeswari

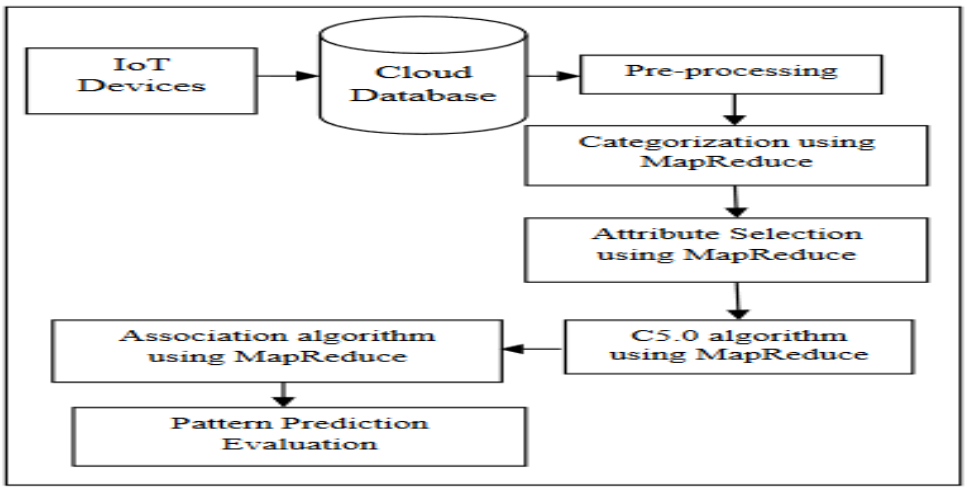
The traditional database paradigm does not have enough storage for the data produced by Internet of Things (IoT) devices lead to the need of cloud storage. These data's are analyzed with the help of Big Data mining techniques.Cloud based big data analytics and the IoT technology performs an important role in the feasibility study of smart agriculture.Smart or precision agricultural systems are estimated to play an essential role in improving agriculture activities. Mobile device usage is very common by everyone, including the farmers. In that, in the daily life of farmers the Information and Communication Technologies (ICT) play a vital role to get the agricultural Information. The IoT has various applications in digital Agriculture domain like monitoring the crop growth, selection of the fertilizer, irrigation decision support system, etc. In this paper, IoT device is used to sense the agricultural data and it is stored into the Cloud database. Cloud based Big data analysis is used to analyze the data viz. fertilizer requirements, analysis the crops, market and stock requirements for the crop. Then the prediction is performed based on data mining technique which information reaches the farmer via mobile app. Our ultimate aim is to increase the crop production and control the agricultural cost of the products using this predicted information.

Methods:

- Cloud database is used to store and share the crop information's, prices of the fertilizers and crop prices.
- In an agriculture sector, the cloud computing gives the smartness with flexibility, predictability, scalability, and optimization.
- It gives the information for farmer in an economical and reasonable cost.
- In the field of agriculture, IoT plays a very important role in collecting data.

MapReduce concept is easy to handle the data and process the multiple nodes. In this model, the process can be divided into maps and reduced. Map function is used to perform filtering and sorting and reduce function is used to perform a summary 2017 International Conference on Intelligent Computing and Control (I2C2) operation. So, the MapReduce technique is used in the predictive analysis concept to predict the data.

They propose a new prototype model for providing the sensing data as a service on the cloud.Wireless Sensor Network increasingly enables applications and services to interact with the physical world. It enables the farmer to have an effective and smart solution to improve the crop yield with less cost.



Our future work will be focussing on interfacing different soil nutrient sensors with IoT tools then collect the data with the sensor tools and store the data into cloud database to analyzing and predict with the data mining algorithms suitable for agricultural Big Data analysis for getting the desired outcome.

5] Crop yield prediction in agriculture using data mining predictive analytics techniques

- P.Surya

Data Mining is an emerging research field in Agriculture especially in crop yield analysis and prediction. As early into the growing season as possible, a farmer is focused on perceiving how much yield they are about to expect. As with many other sectors the amount of agriculture data is increasing on a daily source. In our proposed work, collected agriculture dataset will be used to get crop yield prediction models using various regression techniques. Regression analysis was tested for the effective prediction or forecast of the agriculture yield for various crops in Tamilnadu state particularly in the North Western zone of Tamilnadu. North western zone of Tamil Nadu state data consists of four districts. The North western zone of Tamilnadu districts are Dharmapuri, Salem, Namakkal, Krishnagiri. The analysis depends on the results of the predictor model, in the north western zone, under the area having more cultivated crops are Tapioca, Sugar cane, Ragi, Maize, Groundnut.

Methods

- **Linear Regression**

Linear Regression is one of the commonly used well-known modeling techniques in data mining concept, in which the dependent variable is to be taken as continuous, in other independent variables will be continuous or discrete, and the regression line is linear.

- **Logistic Regression**

Logistic regression technique is used to find the probability of an event of Success and event of Failure. Logistic regression will be used when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature.

- **Polynomial Regression**

A regression of y on x may be a polynomial regression of y on x if the ability of independent variable power is greater than one. Such an equation is known as polynomial regression equation.

- **Ridge regression**

Ridge Regression could be a technique used once the data suffers from multicollinearity (independent variables are extremely correlated).This paper dealt with various regression techniques for agriculture crop yield prediction. This mainly focused on getting a predictor model by using regression techniques. Predictor formula is most useful in the crop prediction of Agriculture crop Production in Tons.

6] Agricultural Crop Yield Prediction Using Artificial Neural Network Approach

-Miss.Snehal S.Dahikar , Dr.Sandeep V.Rode

By considering various situations of climatologically phenomena affecting local weather conditions in various parts of the world. These weather conditions

have a direct effect on crop yield. Various researches have been done exploring the connections between large-scale climatological phenomena and crop yield. Artificial neural networks have been demonstrated to be powerful tools for modeling and prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameters of soil and also parameters related to atmosphere. Parameters like type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulfur, manganese, copper, iron, depth, temperature, rainfall, humidity. For that purpose we are using an artificial neural network (ANN).

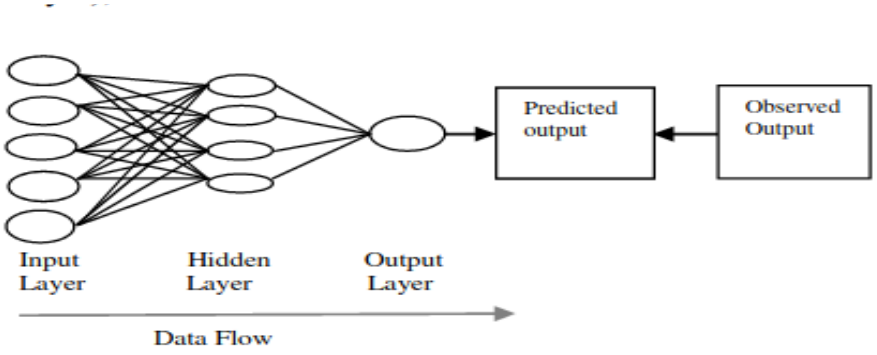
Methods:

Artificial Neural Network (ANN).

The word network in the term 'artificial neural network' refers to the inter–connections between the neurons in the different layers of each system. An example system has three layers. The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations.

An ANN is typically defined by three types of parameters:

1. The interconnection pattern between different layers of neurons
2. The learning process for updating the weights of the interconnections
3. The activation function that converts a neuron's weighted input to its output activation.One type of network sees the nodes as „artificial neurons“. These are called artificial neural networks (ANNs).



Layer and connection of a feed-forward back-propagation ANN

In this way we concluded that ANN is a beneficial tool for crop prediction. This paper includes the parameters of their regional soil parameters. Then it is analyzed by using feed forward back propagation ANN. Analyze in matlab ANN approach to make it more efficient. ANN Sharing Difficult and ANN Training Takes a Long Time.

7]Rice Crop Yield Prediction in India using Support Vector Machines.

-Niketa Gandhi; Leisa J. Armstrong;

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.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. Various climatic factors which are known to affect the rice crop yield, such as precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration**,**were considered with the rice yield production for the Kharif season.

Methods:

- Acquiring each parameter (precipitation, minimum, average, maximum temperature and reference crop evapotranspiration) monthly mean records of each district is considered.
- Calculating the total precipitation, average temperature for the minimum, average and maximum temperature for each year for each district during the Kharif season.Acquiring each districts area, production and rice crop yield details of the year 1998 to 2002 from the publicly available Indian Government records.
- For preparing the data set for applying data mining techniques, unrequired columns were omitted. They were sr. no,name of the district and year.The data set was then sorted on the basis of area,which is less than 100 hectares can be omitted.
- The dataset was then sorted on the basis of yield to classify the records into low, moderate and high. The yield has been calculated on the basis of area and production hence these two columns were omitted.
- This data set was then saved in .csv format for further applying data mining techniques. This file had following columns: precipitation, minimum temperature, average temperature, maximum temperature, reference crop

evapotranspiration, crop yield and class.

Table 1 A confusion matrix

Predicted		Observed	
		True	**False**
	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

WEKA was used to construct the algorithm.The algorithm achieved the accuracy of 78.76%, sensitivity of 68.17% and specificity of 83.97%.The F1 score was computed to measure the test's accuracy and achieved a score of 0.69. Correlation Coefficient was used to measure the quality of classification which resulted in 0.54.In terms of test's accuracy and quality also BayesNet and Multilayer Perceptron showed the highest accuracy and best quality and SMO showed the lowest accuracy and worst quality.

8]Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors.

-Preeti Tiwari and Piyush Shukla

Agriculture is one of the major revenue producing sectors of India and a source of survival. The number of biological, financial, and environmental factors affects the crop yield. So, unidentifiable changes in these factors lead to failure in agriculture. This paper focuses on prediction of crop yield where different geospatial features were utilized, such as normalized difference vegetation index, standard precipitation index, vegetation condition index. In order to learn from previous weather conditions, a standard error back propagation neural network was used. Here, the training was done in such a way that all sets of features were

utilized in pair with their yield value as output. To increase the reliability of the work, the whole experiment was done on a real geo-spatial dataset from the Madhya Pradesh region of India. The result shows that the proposed model has overcome various evaluation parameters on different scales as compared to previous approaches adopted by researchers.

Methods:

Normalized Difference Vegetation Index (NDVI)

For crop yield prediction, NDVI is ordinarily utilized as an input feature vector. This feature measures vegetation cover in the land. NDVI was first recommended as a record of vegetation well-being and thickness.

$$N = (b_{NIR} - b_{RED} b_{NIR} + b_{RED})$$

where variable N represents as the normalized difference vegetation index (NDVI),

while bRED are the reflectance of red bands and bNIR stands for near infrared.

Standardized Precipitation Index (SPI)

In 1993, this index was introduced by Mckee and his team members. SPI helps in quantifying the precipitation shortage occurs in different time scales for reflecting the effect of available water resources. So, the temporal pattern was identified by estimating the standard precipitation value for different month periods. In order to estimate SPI value, the difference of the precipitation from the average of the selected time scale was taken and finally divided by the standard deviation.

Vegetation Condition Index (VCI)

It is suggested which represents correlation between the current NDVI value with the minimum NDVI value from a huge set of periods of time. It was calculated where V j stands or the jth position VCI value in the vector. Similarly,

N j stands for the jth position NDVI value in the vector. However, Nmin and Nmax are minimum and maximum values of the NDVI vector.

$$V_j = (N_j - Nmin\ Nmax - Nmin) \times 100$$

As digital data on libraries and servers are drastically increased with every new Second, researchers are attracted to work on it. So, this work focuses on geo-spatial data for crop yield prediction by utilizing various features vectors of SPI, NDVI, and VCI. Classification based on the working of neural networks has also been performed by many researchers. In some of the works, document learning is based on single information, but through this proposed work it will overcome this dependency, as well as the learning can be performed on all the available data.

Here, the result shows that proposed work has improved the prediction accuracy by reducing the RMSE, relative error value. Also through proper training and rich input vector resultant, the neural network is a less time-consuming model. It was obtained that proposed work has reduced the relative error by 33% as compared to SNN, while RMSE also got reduced by 31.8%. In this work, average testing time for prediction of crop yield was decreased by 38.2% and overall accuracy was improved.Temporal aggregates might still be contaminated by cloud cover; the procedure will be biased by a single false value.

9]Crop Prediction Using Predictive Analytics

-P. S. Vijayabaskar

This work is to construct a model for testing the soil fertility. It also suggests the crop which has to be planted depending upon the value obtained from the sensor. It also provides the regional wise information about the crop in the form of a graph. We have farmer chat where the farmers can share and get ideas from the expert by registering in this application. It also suggests the fertilizer which has to be added to the soil in order to increase the crop productivity. It helps the farmer to analyze the fertility of their yard and plant the better crop to increase their productivity and profit. It also provides the information about the fertilizer to be added in the soil and also provides the information about the nearby fertilizer shop.

Predictive Analytics

Predictive analytics include various statistical techniques such as predictive modeling, machine learning and data mining which will analyze the current and historical data to make predictions about the unknown event.Predictive analytics has been classified into three types. They are

- Predictive model

  The main objective of this model is to obtain the similarity in different sample will exhibit the specific performance

- Descriptive model

  A descriptive model is used to establish the different relationship between the customer or products.

- Decision model

  A decision model is used to establish the relationship between all the elements of known data, the decision, and forecasting the result of the decision.

Performance Evaluation

The applications have the following modules like Admin, user, GSM, NPK sensor, raspberry pi. By using the NPK sensors we can get the soil results whenever it's needed. This will reduce the farmer's work.

Various soil samples taken from different places can be tested. Portable , Time consumption is low. Helps the farmer to plant the right crop,High accuracy and Speed.Agro algorithm helps farmers to select a particular crop .Laborious manual statistical analysis.

The NPK sensor used will detect only the nitrogen, potassium and phosphate present in the soil. Parameters like humidity, pressure, ph are not identified. The system has to be compared at rotational level which helps us to compare the farming system with different crop composition. Climatic conditions may change hence accurate results cannot be produced.The prediction based on atmosphere is not accurate.

10]Analysis Of Crop Yield Prediction Using Data Mining Techniques

-D Ramesh , B Vishnu Vardhan

Data Mining is emerging research field in Agriculture especially in crop yield analysis and prediction. As early into the growing season as possible, a farmer is focused in perceptive how much yield they about to expect. As with many other sectors the amount of agriculture data are increasing on a daily

source. The agrarian sector in India is facing a rigorous problem to maximize crop productivity. More than 60 percent of the crop still depends on monsoon rainfall. Recent developments in Information Technology for the agriculture field has become an interesting research area to predict the crop yield. The problem of yield prediction is a major problem that remains to be solved based on available data. Data Mining techniques are the better choices for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) technique and Density based clustering technique for the selected region i.e.East Godavari district of Andhra Pradesh in India.

Methods:

In this paper the statistical method namely Multiple Linear Regression technique and Data Mining method namely Density-based clustering technique were take up for the estimation of crop yield analysis.Multiple Linear Regression A regression model that involves more than one predictor variable is called the Multiple Regression Model. Multiple Linear Regression (MLR) is the method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes termed as a predictor and independent variables are called predictors. Multiple Linear Regression (MLR) technique is based on least squares and probably the most widely used method in climatology for developing models to reconstruct climate variables from tree ring services. This crop yield prediction model is presented with the use of Multiple Linear Regression (MLR) technique where the predictor is the Production and there are seven predictors namely Year,Rainfall, Area of Sowing, Yield and Fertilizers (Nitrogen,Phosphorus and Potassium).

Density-based Clustering Technique

The primary idea of Density-based clustering techniques is that, for each point of a cluster, the neighborhood of a given unit distance contains at least a minimum number of points. In other words the density in the neighborhood should reach some threshold. However, this idea is based on the assumption that the clusters are in the spherical or regular shapes.

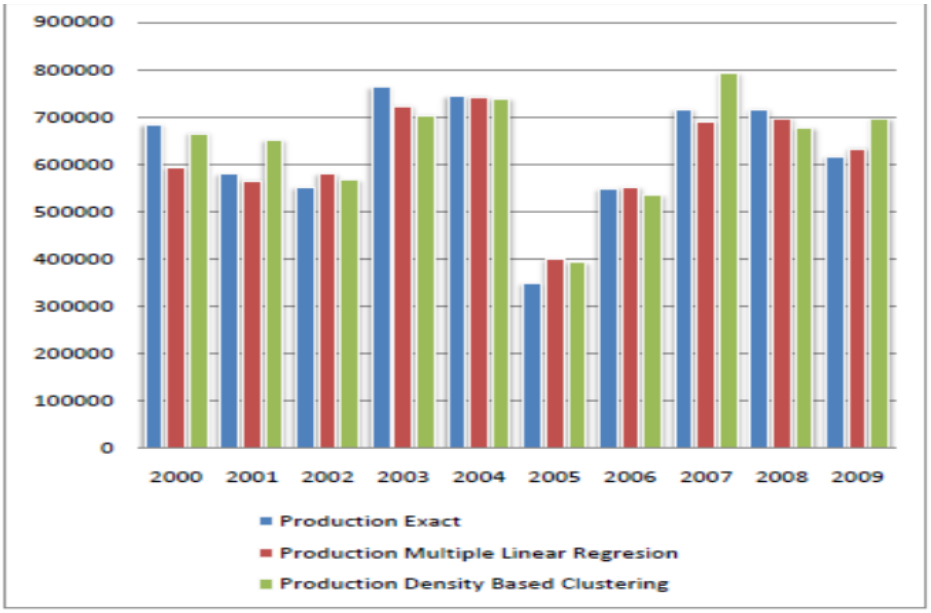
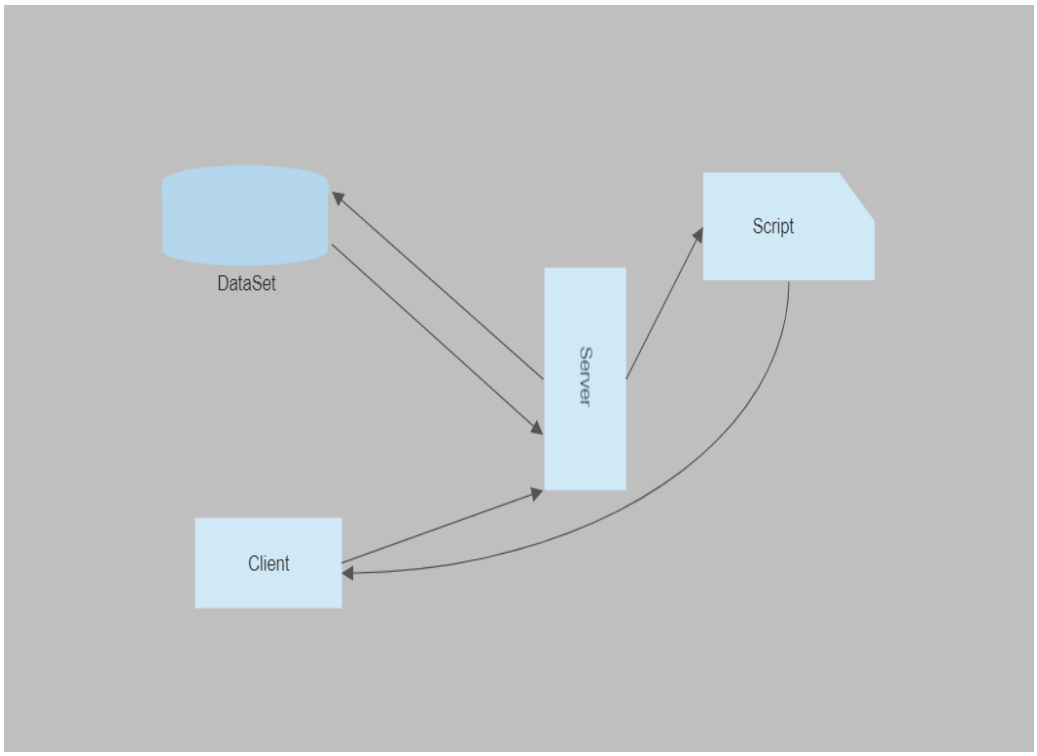


Fig-1: Comparison between Multiple Linear Regression technique and Density-based clustering technique

In this procedure the results of two methods were compared according to the specific region i.e. East Godavari district of Andhra Pradesh in India. Similar process was adopted for all the districts of Andhra Pradesh to improve and authenticate the validity of yield prediction which are useful for the farmers of Andhra Pradesh for the prediction of a specific crop.In the subsequent work a comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efficiency of the proposed technique.


DataSet
Script
Server
Client