A Mobile Application Through Crop Yield is Predicted using Data Mining Techniques

Thanushree C
Department of Computer Science and Engineering
Jain University
Bangalore, India

Harshitha .P
Department of Computer Science and Engineering
Jain University
Bangalore, India

Manjunath C R
Department of Computer Science and Engineering Associate professor
Jain University
Bangalore, India

Abstract:- The economy of India is mainly dependent on agriculture. Most of the people in India take up agriculture as their occupation. India also suffers from natural calamities like flood, drought are the main reasons for the crop damage and huge loss to the farmers. So the crop yield prediction in advance to harvest the farmers will be benefited to make the planning appropriately with the predicted results. The yield of the crop depends on the major factors like weather, soil, water, nitrogen, soil moisture, humidity, crop rotation, surface temperature rain water etc. Predicting the crop yield deals with the huge datasets along with lots of data analytics with are applied on those datasets. Prediction of agriculture crop yield is a job that requires the unification of knowledge from several areas such as data mining. Using data mining we extract knowledge of huge data using various algorithms/techniques.

A huge data coming from different parameters like weather, water, pH, temperature, soil quality etc are required for crop yield prediction .Prediction of crop yield is important for managing and storage of crops and data mining also focuses upon methodologies used for extracting the useful knowledge from data and there various tools used for extracting knowledge from data. But it is sad that farmers do not use any knowledge discovery process approach for their crop yield data. Data mining is also helpful in agriculture for decision making. Our system aims at making farmers aware of the crop yield prediction by using various data mining techniques and the result of the prediction as a message to his phone.

Keywords:- data mining, agriculture, prediction, mobile application.

I. INTRODUCTION

India is the seventh largest country in the world and it's an agriculture based country where most of its population derive their living and the source of income from the agriculture itself. India is practicing agriculture since many hundred years and it is the oldest countries to follow agriculture even today. Due to globalization, the trends in the agriculture are keep on changing .India is being ranked second in the farm output worldwide. Most of the economy in India is dependent on agriculture. Agriculture also play an important role in the development of socio-economic fabrication of India and which is demographically a broad sector of income. Agriculture supports the livelihood of about 58 per cent of India's population. Agriculture is the major contributor to Indian economy as it is a main source of income as large population of India depends on agriculture. Agriculture deals with the business of crops and also depends on some factors. Some of the factors on which the agriculture is dependent upon are the type of soil, the climatic conditions, cultivation of crop, temperature, fertilizers used, harvesting, irrigation, pesticides which are used and other factors. Sometimes there would also be a huge loss in agriculture due to the natural calamities like drought, flood which would result in the loss of crop production and loss to farmers. The large numbers of farmers are committing suicide due to this reason and sometimes there would also be a pest attack which will again result in the loss. Annually, farmers face a loss of 40% in total that is caused by either the damage of crops or the farmers deprived of being paid the right price for their yield. Therefore, this results in low productivity. The crop yield in agriculture is measured of the degree of yield per unit area that has been cultivated with the seeds that are produced from the same cultivated crop. Predicting the crop yield immediately after the cultivation will be helping the farmers and the government bodies to decide for storage, the yield profit, selling of crops, importing/exporting the crop. Government has taken few steps and is providing the messages and the information of the crop that is known as the agro vendor's information to farmers; also providing data related to soil quality and the utility of the soil at each region. The yield that is derived from the crop will be impacting directly on the National and International economic sectors of the country on annual basis and also plays an important and significant role in the food management. There is need for agricultural institutions that educate the farmers with short-term courses and workshops, update them with latest technologies and hoping to obtain better and effective results. Forecasting the information of the crop before harvest is very useful for the farmers for future activities. There are many researchers who have developed statistical models for the yield of crop that would suggest the pre-harvesting process and the crop yield after harvesting. These industries use agriculture product as chemicals, seeds, poultry, fertilizer, livestock, food animal feed, raw material, pesticide, paper. The accurate estimation of

crop production helps these companies to plan supply chain decision in production. Till today by keeping only the experience of the farmers was used, this included the random count of number of seeds, buds that plants have. Based on these the farmer would predict the crop yield. Now days there are number of different applications emerging and the technologies for the farmers to make their lives easier. The information technology has become more and more part of one's daily life, which is helpful especially in the field of agriculture. The Data mining is one of the processes of discovering the previously unknown and the likely impressive patterns in the large datasets. Therefore the Data Mining is the process used for extracting information from the given set of data and transforms it into a readable structure for further use of data. Data Mining is a process used for discovering the patterns in the large data sets which are involving methods at which include the statistics , machine learning, artificial intelligence and database systems. Finally the ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications. Data mining is the technique which plays a vital role in the analysis of data. Data mining is also computing process which is used for discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system. Unsupervised (clustering) and supervised (classifications) are two different types of learning methods which are implemented in the data mining. Clustering is the process of examining a collection of "data points," and grouping the data points into "clusters" according to some distance measure.

II. RELEATED WORK

The survey consists of study of various methodologies that were adopted to predict the crop yield.

In [1] using crop yield prediction system for enhancing the yield for better decisions and planning which can be planned out. They follow some step.

- Selecting the agriculture field: First for the crop yield prediction we need to select the agriculture field.
- Selecting the crop: The crop which needs to be sown should be selected.
- Input: Consider the climatic parameters like rainfall, temperature, humidity and also some agronomical variables like pesticides, nutrient contents N, P, K, and soil etc. The parameters values will be taken as input.
- Pre-processing: It is nothing but noise removing process. The data mining to be a successful application it requires dataset which is huge. From the various resources the data which is collected will be in raw form and that will cotain some redundant, inconsistent, incomplete data. So here

the redundant data will be removed. Finally that redundant data will be normalized.

Feature Extraction: Here selection of attribute will happen. It identifies and uses the relevant attribute from set of data. The classifiers application processes the irrelevant and redundant data and then those data will be removed.

In [2] it follows some methodology

- Dataset Collection: The dataset is gathered from Madurai district and in soil testing lab it has been tested for some specific attribute of soil. In addition to that some general crop data is collected from online sources which are similar. The crops considered may include cotton, groundnut, millet, pulses, sorghum, sugarcane, coriander, vegetables, banana, and paddy in this model. Texture, Depth, Drainage, Permeability, soil colour, Ph, Water holding and Erosion are the parameters considered. These above mentioned soil parameters play an important role in extraction of soil nutrients and water of crop's ability. For the satisfactory environment soil is must. The absorbed nutrients determine the water holding capacity. To hold onto the nutrients soil texture will affect the soil's ability. The availability of soil nutrients is affected by the acidity or alkalinity level which is an important variable. The penetration of roots is determined by water holding and drainage. For choosing a crop the following reasons are considered for the stated parameters.

- Ensemble technique for crop prediction: Ensemble is one of the models of data mining to gather greater efficiency and prediction this model combines the multiple models power than other model could alone achieve. This is also called as the Method of Committee or Combiners Model. In this system they use the ensemble technique called Majority voting technique. Base learners can be at least two.

➢ *Learners Used in the Model*
- *Random Tree*
  It is same as decision tree. But the difference is for every split their will be random attribute subset is available. Nominal and numerical data both can be built for random trees. Randomly chosen attributes can be considered for each node. The size of the subset can be specified by subset ratio parameter.

- Chaid
  It stands for Chi-squared Automatic Interaction Detection. It is based upon adjusted significance testing which is also a decision tree technique type. Chi-squared based criterion is used in place of gain or information ratio. It has many advantages which uses multi way splits that include highly interpretable and visual results.

- *K-Nearest Neighbor*

For both classification and regression this can be used. The cases available will be stored and new cases will be classified according to similarity measure.

III. METHODOLOGY

A. Working Principle

- Gather the data related to the field.
- Compute the collected data with the database
- Communicate the data to the server along with the necessary parameters like soil, climate etc.
- Get the result.
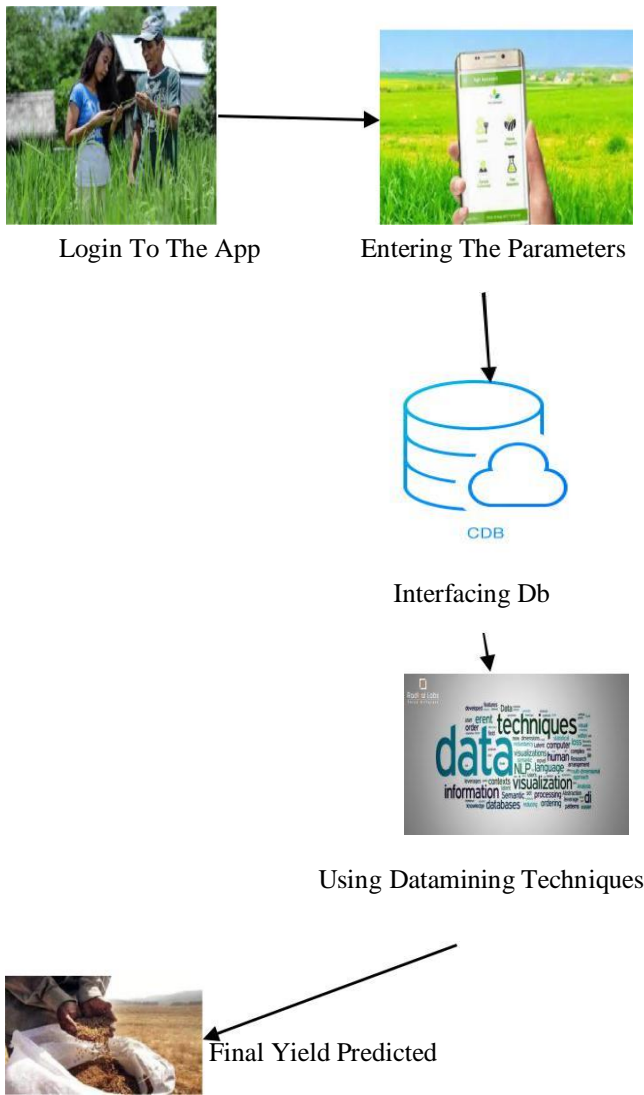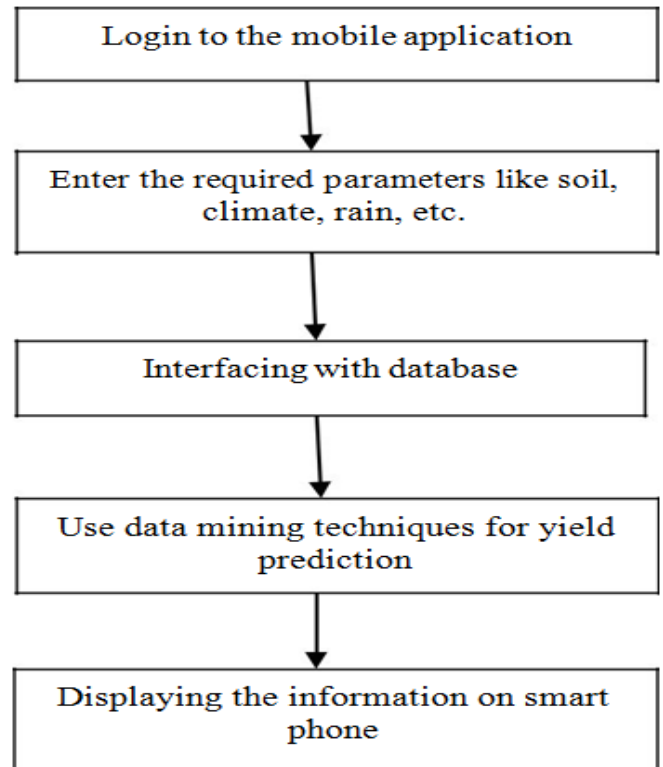- Notification to the farmer about the yield prediction.

B. Work Flow

Login To The App   Entering The Parameters

Interfacing Db

Using Datamining Techniques

Final Yield Predicted

Fig 1:- The system work flow

C. System Design

D. Exixting System

In existing systems there is no much technology used for the detection or the prediction of the yield, based on the previously collected data. This previously collected data will include the several factors like soil, climate, rainfall and other important features in that particular location. At present only the type of the crop that can be grown in that particular area or region is been told to the farmers and most the farmers fails to go to the government bodies to take suggestions about the land, soil, and other climatic conditions that will be effecting the growth of their crops in spite of all these facilities the farmers fail to go because they find difficult to travel to far place and no proper segments to support them therefore the farmers will not have much knowledge about the crop yield in advance.

E. Proposed System

The system that we are proposing here is the mobile application where the farmers need to download the app on their smart phone and enter the required details about their field and the data related to the crop they grow and for how much area along with the climatic conditions ,rainfall and other necessary factors and all these data will be compared with the previously noted data and are compared using various data mining techniques then finally the compared results of crop yield would be displayed on their smart phone.

REFERENCES

[1]. Yogesh Gandge, Sandhya, study in various data mining techniques for crop yield prediction, 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)

[2]. S.Pudumalar, E.Ramanujam,R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha. Crop Recommendation System for Precision Agriculture 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC)978-1-5090-5888-4/16/$31.(2016) IEEE.

[3]. S.Nagini, Dr. T. V. Rajini Kanth, B.V.Kiranmayee, Agriculture Yield Prediction Using Predictive Analytic Techniques, 978-1-5090-5256-1/16/$31.00(2016)IEEE.

[4]. Monali Paul, Santosh K. Vishwakarma, Ashok Verma, Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach,2015 International Conference on Computational Intelligence and Communication Networks.

[5]. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki Analysis of agriculture data using data mining techniques: application of big data Majumdar et al. J Big Data (2017).

[6]. A predictive modeling approach for improving paddy crop productivity using Data Mining Approach, Turk J Elec Eng & Comp Science (2017).

[7]. Effects of temperature and rainfall on paddy yield using data mining .kuljit kaur,kanwalpreet singh.987-1-5090-3519-9/17/$31.00 (2017)IEEE.

[8]. Niteta Gandhi, Liesa Armstrong, Applying data mining technology to predict yield of rice in humid subtropical climatic zone of india,978-9-3805-4421-2/16/$31.00(2016)IEEE