SYSTEM TO PREDICT THE PADDY YIELD DEPENDING ON AGRICULTURAL DATA USING DATA MINING

Lavanya N[1] , A C LOKESH KUMAR[2]
[1]P.G.Student, Dept of CSE ,P.E.S.C.E Mandya,Karnataka,India.
[2] Associate Professor , Dept of CSE, P.E.S.C.EMandya,Karnataka,India.

Abstract : Agriculture is the main occupation and backbone of our country. Agriculture is a necessary sector because it is one of the main income of India. Agricultural yield depends on the natural parameters like temperature, rainfall, humidity etc. Due to unpredictable climatic changes and unavailable of mandatory resources of water farmers are unable to attain the expected yield. The climatic changes plays a vital role in the expected production of agricultural yield. Predicting the crop yield is one of the challenging task and it requires unification of knowledge from statistics and agricultural data. Prediction of crop yield helps in managing the storage of crops as well as it directs the transportation decisions, and risk management issues related to crops. In this project we are using data mining techniques to predict the crop yield. Data mining focuses on the discovering of new pattern or knowledge in a large set of information. We use methodologies of data mining to extract the knowledge which results in an efficient output which can be deduced easily and rapidly from the dataset. Here we are predicting the agricultural yield of paddy crop. Farmers do not use any knowledge discovery process approach on paddy yield data so that Data mining can be used in agriculture for decision making. In the proposed system, we collect data from different government organizations, after preprocessing the available data ,we apply Eclat algorithm for analysis of temperature, rainfall , nitrate content, PH and history of the paddy yield to predict the paddy yield and to analyze the effect of all the above natural parameters on the paddy yield.

IndexTerms **- Data Mining , predicting crop yield, natural parameters, association rules.**

I. INTRODUCTION

Now a days, agricultural departments are able to generate and aggregate a bulk of data.Crop production rely on various environmental parameters such as temperature, soil moisture, rainfall, evaporation, radiation and other practices. The main objective is to discover the associated relationship between the environmental factors and paddy yield at Nanjangud district of Karnataka. Here , we are predicting the paddy yield by taking temperature ,rainfall,soil moisture and ph into consideration using Data Mining techniques which will be helpful for the agricultural enterprises[6].Data Mining is a field where a large amount of data is evaluated in distinct ways and classified in order to get useful information. Agriculture is a novel research field in data mining. Data Mining is the process of identifying hidden and frequent patterns in the large databases[1]. The figure 1 shows the various stages of the paddy plant and they are vegetative stage, reproductive stage, maturation stage and grain filling stage. System discovers the relationships between temperature, rainfall ,soil moisture, pH and paddy plant during all these stages using data mining technique called "Association Rules Mining". We make use of the data collected from government sectors for analysis and to predict the patterns between temperature, rainfall and paddy plant.

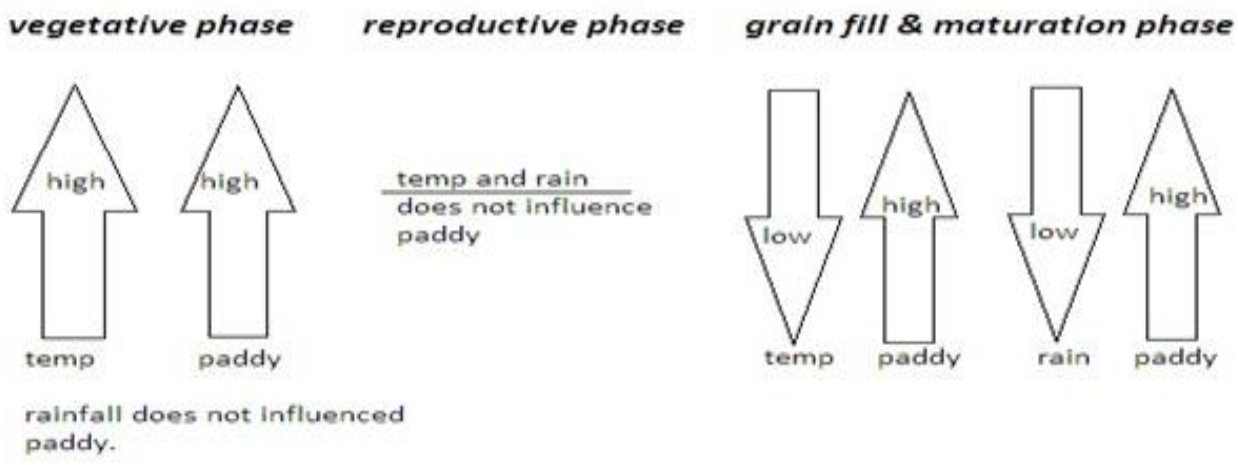

Fig 1 : Temperature Vs Rainfall Vs Paddy

The existing system is manual where we compare the previous with the present. Based on the previous experiences and results we came to know how much crop yield will be produced. There is no automation to predict the relationship between rainfall, temperature, soil moisture, pH and paddy plant. The existing system has the limitations and they are manual process, time consuming, less reliable, less efficient and less user satisfaction.

The proposed system is an agriculture application which analyzes the previous data related to rainfall, temperature and paddy crop yield. Proposed system makes use of data mining in agriculture for decision making. The research is conducted taking under consideration the various stages of the paddy plant that are vegetative stage, reproductive stage, maturation stage and grain filling stage. System discovers the relationships between temperature, rainfall, soil moisture, pH and paddy plant during all these stages using data mining technique "Association Rules Mining".

II. RELATED WORK

Agricultural productivity depends on natural factors like climate, economy and geography. The objective of this paper is to enhance the productivity of prediction of yield which is useful for the farmers. Several aspects influences the productivity of agriculture and it can solved using data mining techniques and other statistical methodologies. The authors Ami Mistry , Vinita Shah used some of the classification models like Linear Regression ,Artificial neural network, k-Nearest Neighbor(KNN),Regression Tree, Support Vector Machine and clustering models like K-means clustering , Density-based Clustering and Weight based clustering[4]. The y compared the data mining techniques mentioned above and concluded as Root Mean Square Error (RMSE) describes the machine learning algorithm performed on dataset. Different models of data mining like classification and clustering techniques provide the better result for the different crops.

This study concentrates mainly on rice crop yield. The authors proposed an rice prediction model to predict the yield of rice. They used multiple linear regression, partial square linear regression data mining techniques. The authors Umid Kumar Dey , Abdullah Hasan Masud , Mohammed Nazim took rainfall, temperature , humidity, yield and area as attributes to predict the crop yield. Then usage of Multiple Linear Regression, AdaBoost (Adaptive Boosting), Support Vector Machine Regression and the Modified Nonlinear Regression equation is implemented on the training set to find out parameter values[2]. Adaptive Boosting adapts the data to build a complex relationship to predict the yield. Modified Nonlinear Regression is a non linear regression .They concluded that modified non linear regression is more efficient comparing with other three predefined models. This paper shows the vital role of weather in prediction of crop yield.

This study deals with the prediction of crop using data mining approach by considering temperature, rainfall and soil behavior. The authors Monali Paul, Santosh K. Vishwakarma, Ashok Verma used K- Nearest Neighbor algorithm , Naïve Bayes algorithm which are well known classification algorithms applied to dataset of soil taken from the Madhya Pradesh state. The experiments are performed using RapidMiner.The algorithms are performed on the dataset to get accuracy on yield of the crop. RapidMiner 5.3 is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics[3].

Prediction of yield plays a vital role in agricultural problem which can be solved using the available data. By employing the data mining techniques and classification algorithms we can get the solution for yield prediction problem. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities. The authors D Ramesh, B Vishnu Vardhan used one of the clustering technique called k means algorithm. This study concentrates mainly on rainfall and temperature attributes effect on agricultural data[5]. They compared K means algorithm and MLR technique to predict the agricultural productivity. They used K means algorithm one of the basic algorithm. Biclustering techniques are used to identify agricultural related dataset.K means algorithm partitions into cluster and it provides the agricultural related information.

III. METHODOLOGY

A. Data Mining Techniques

The data mining techniques are used to describe the data recovery operation and type of mining Some of the data mining techniques are:

a. Association Rule
b. Classification
c. Clustering
d. Sequential Patterns and Predictions

a. **Association Rule** - As we know association is a correlation that exists between two or more items or patterns to discover new pattern. It is one of the most familiar and forthright data mining technique. For example, Market-basket analysis, where we analyse the data on buying habits of customer probability of identifying a consumer regularly purchases bread

when they purchase cream hence conveying that a consumer next time consumer purchase cream might also purchase bread.

b. **Classification -** Classification is based on multiple attributes to identify a particular class , it builds a concept on the type of item or object based on attributes. For example, we can easily classify people on their age, locality etc.and a university on departments, deptid , staff etc.

c. **Clustering -** Clustering is a process of identifying a group or cluster by relating one or more described attributes. Cluster is a group of objects or items having similarities which is useful to identify different patterns.

d. Sequential Patterns and Prediction - Sequential patterns is a approach to identify regular patterns and similar patterns. Consider the scenario of shopping basket , we take advantage of the purchasing history of the consumer and based on the available information we can automatically predict certain items.

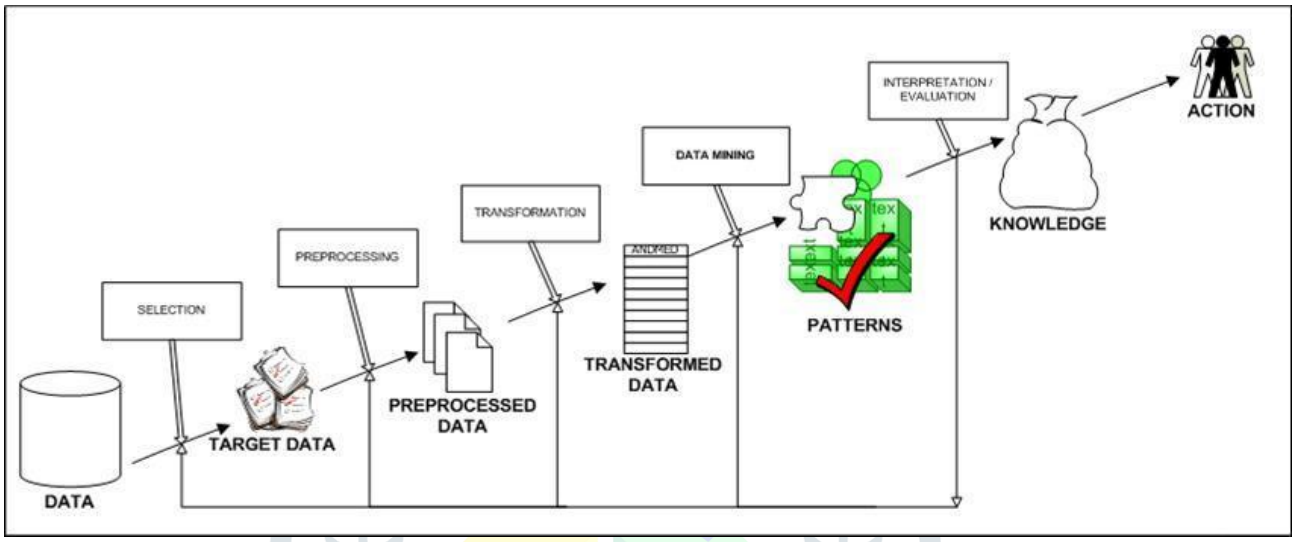

Fig 2: Phases of Data Mining

The figure 2 depicts the data mining phases and are as follows

a. **Data cleaning** :It removes or eliminates noisy and unnecessary data. Quality of data is one of the issue in data mining.

b. **Data integration** : In this phase the data got from the previous phase are combined. The data that are stored using different technologies ,are combined from different data sources in this phase.

c. **Data selection** : In this phase data is retrieved from data bases to perform analysis.

d. **Data transformation** : The selected data are transformed in the form which is appropriate for mining.

e. **Data mining** : In this phase algorithms are applied to the transformed data to extract hidden patterns .

f. **Pattern evaluation** – This phase identifies the interesting patterns which represents new knowledge.

g. **Knowledge presentation** – The various techniques of visualization are implemented to represent the knowledge which is mined to the user. We know that the information extracted using the algorithms of data mining are useful .

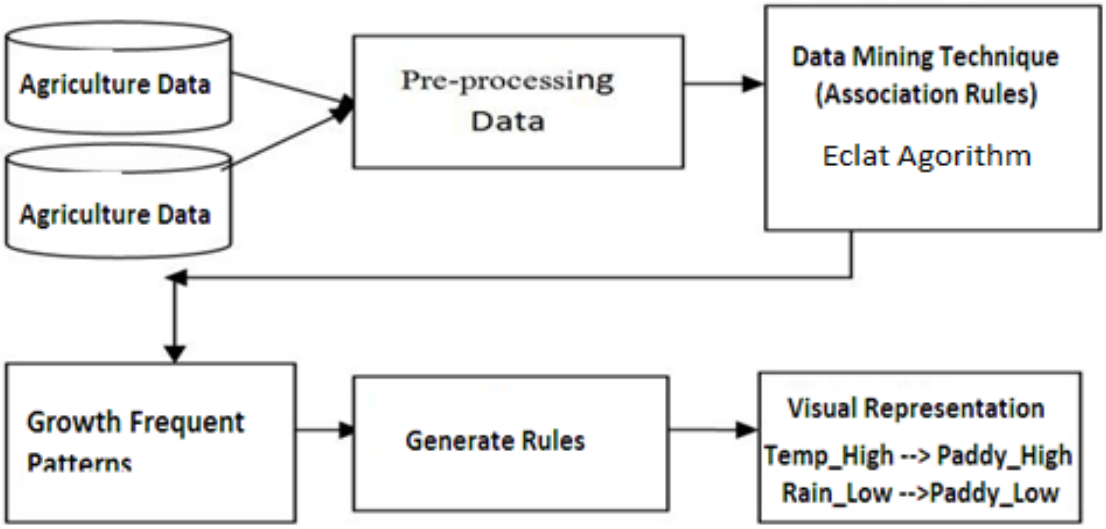

Fig 3:System Architecture Using Eclat Algorithm

The figure 3 depicts the architecture of the proposed system. Initially we collect the agricultural data such as temperature , rainfall , Ph , nitrate content and paddy yield data relevant to our project. Then we pre process the data i.e, the data which are not necessary for of crop yield are removed like farmer name , village etc. Using association rules and éclat algorithm we identify new frequent patterns and we generate rules .Finally the growth frequent patterns are represented to client.

B Algorithm

We are using Association Rules [Eclat Algorithm] for the following attributes:

1. Temperature
2. rainfall
3. Ph
4. nitrogen content
5. crop (paddy)

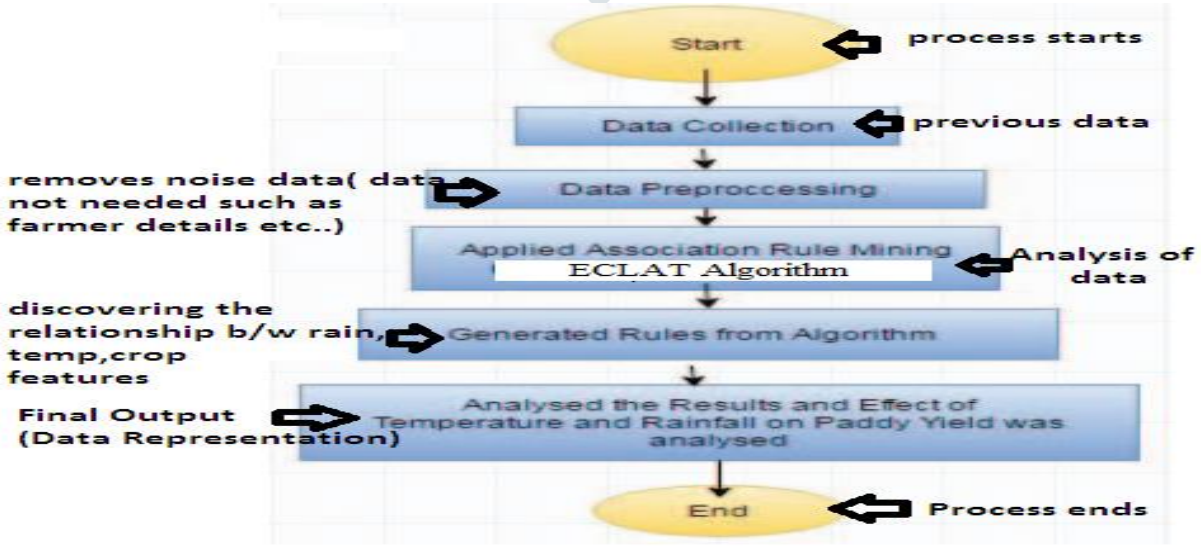

Fig 4:Methodology of a System Using Eclat Algorithm

The figure 4 above shows the methodology that we have adopted to develop the system. Initially data are collected from various government organisations and then noisy data are removed. By applying the éclat algorithm for the preprocessed data we'll get the paddy prediction pattern. The below pseudocode describes the steps of éclat algorithm.

1. Get tidlist for each item (DB scan)
2. Tidlist of {a} is exactly the list of transactions
   containing {a}
3. Intersect tidlist of {a} with the tidlists of all other
   items, resulting in tidlists of {a,b}, {a,c}, {a,d}, ...
   = {a}-conditional database (if {a} removed)
4. Repeat from 1 on {a}-conditional database
5. Repeat for all other items

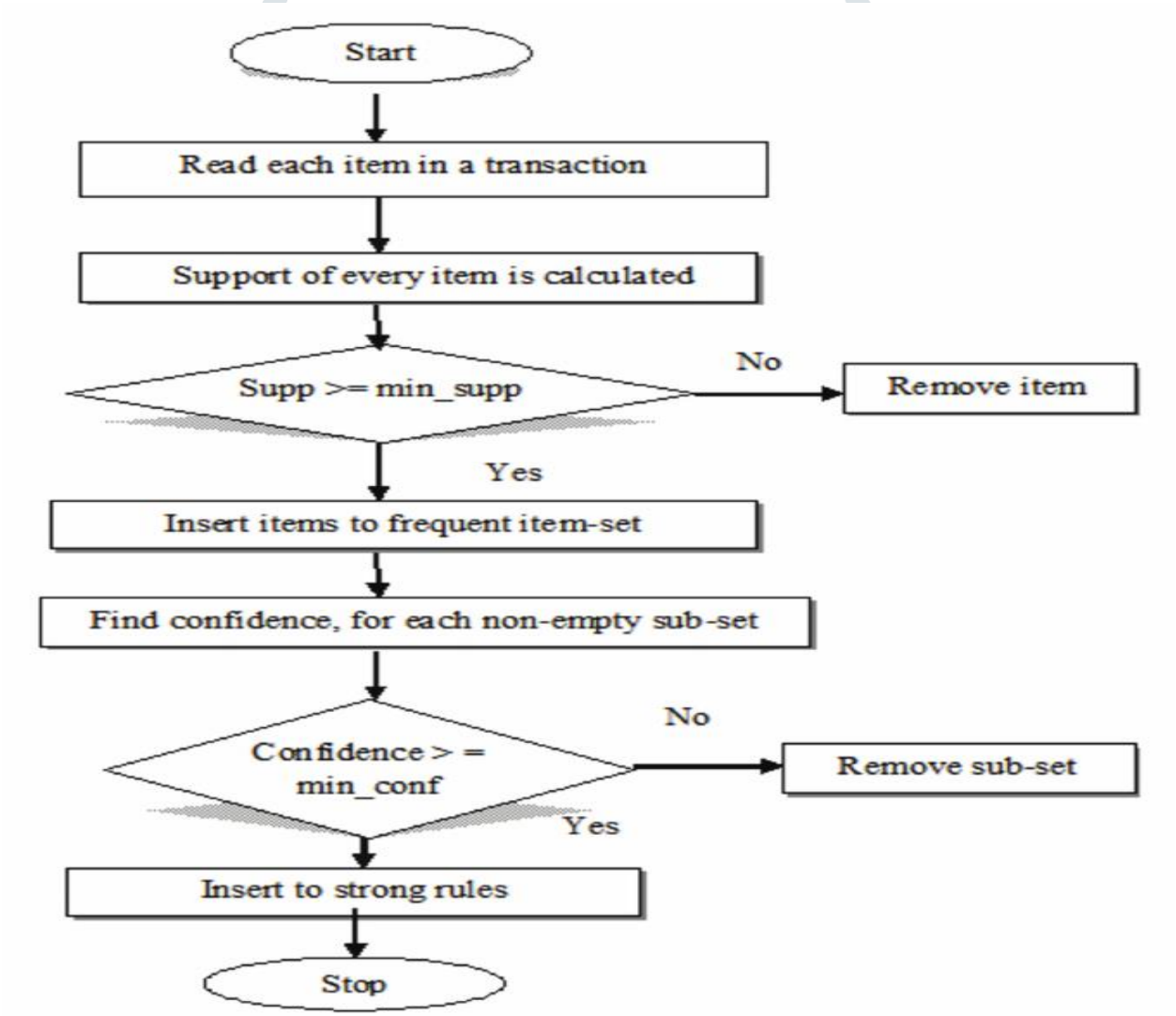

Fig 5:Flow of Algorithm

1. Initially it scans the data of agricultreand determine the support of each item i.e, temperature , rainfall , Ph and nitrogen content.
2. Then we'll fenerate the frequent item set 1 i.e, Lk-1 by eleminating the items having less value than support.

3. Join the Lk-1 to get candidate k-item set.
4. Generate Candidate item set C k item set by combibing the Lk-1 set.
5. Frequent item set are added till C=null set.
6. The items in the frequent item generates non empty subsets.
7. For every non empty subset we determine the confidence and we add the items confidence having greater or equal to the specified confidence to the strong association rule.

C. Actors and their modules

1. Agricultural Head/Administrator

The admin has the following modules;

a. Login Module - This module is for admin or agricultural head login to the web application by entering the credentials like id ,password.

b. Create Agriculture Department/regions - In this module administrator creates the agricultural departments for the respective regions.

c. Create Location Incharger/Staff - In this module administrator adds up the staffs/location incharger for the future use of the application. Staffs basically created for location. Administrator sets a unique staff Id and password for each staff, using these credentials staff can get login.

d. **Set Id and Password for region Incharger -** In this module administrator sets Id and password for location Incharger or staff.

e. **Import Excel Sheet Data (Manage Dataset) -** In this module administrator manages the existing dataset for the crop yield prediction. Here admin adds the existing dataset [sample soil attributes and temp, rainfall and crop yield details]

f. Crop Pattern Prediction Module (Core Module) - In this module the system generates the output for the input given by the application user, the output is to determine the relationship between daily temperature , rainfall and actual crop yield. Yield predicted using data mining technqiue “association rules mining”.

g. Change Password - In this module , the admin can update his profile like changing password etc.

2. Location Incharger/Staff

The Location Incharger/Staff has the following modules.

a. Login Module - This module is for staff or location incharger login to the web application by entering the credentials like id ,password.

b. **Import Excel Sheet Data (Manage Dataset) -** In this module location incharger manages the existing dataset for the crop yield prediction. Here staff adds the existing dataset [sample soil attributes and temp, rainfall and paddy yield destails]based on daily, monthy and yearly.

c. **Crop Pattern Prediction Module (Core Module) -** In this module the system generates the output for the input given by the application user, the output is to determine the relationship between daily temperature and actual crop yield; daily rainfall and actual crop yield. System predicts yield based on daily, monthy and yearly bases.

d. **Change Password** - In this module, the location incharger can change his password whenever he wants.

3. Visitor or Farmers

The Farmers has the following modules:

a. Home
b. About us
c. Contact us
d. Basic Agricultural info

IV. EXPERIMENTAL RESULTS

The outputs that we will get after step by step execution of all the modules of the system are defined by the following snapshots .The figure above depicts the login mmodule of admin to create the staff ,upload dataset and to get the pattern of paddy yield.

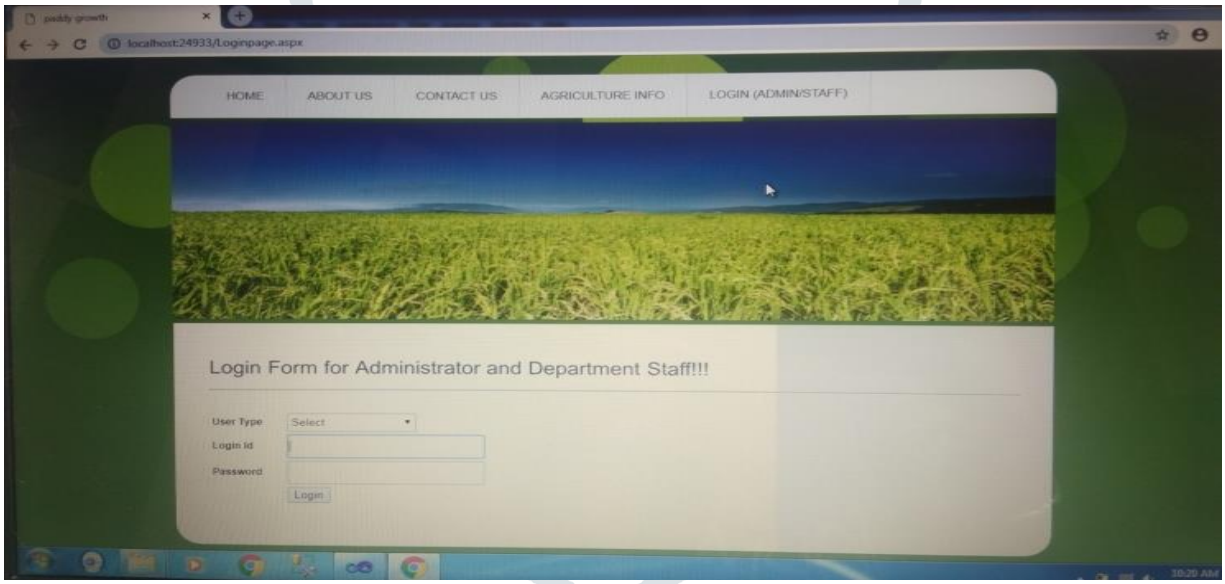

Fig 6 :Login Module

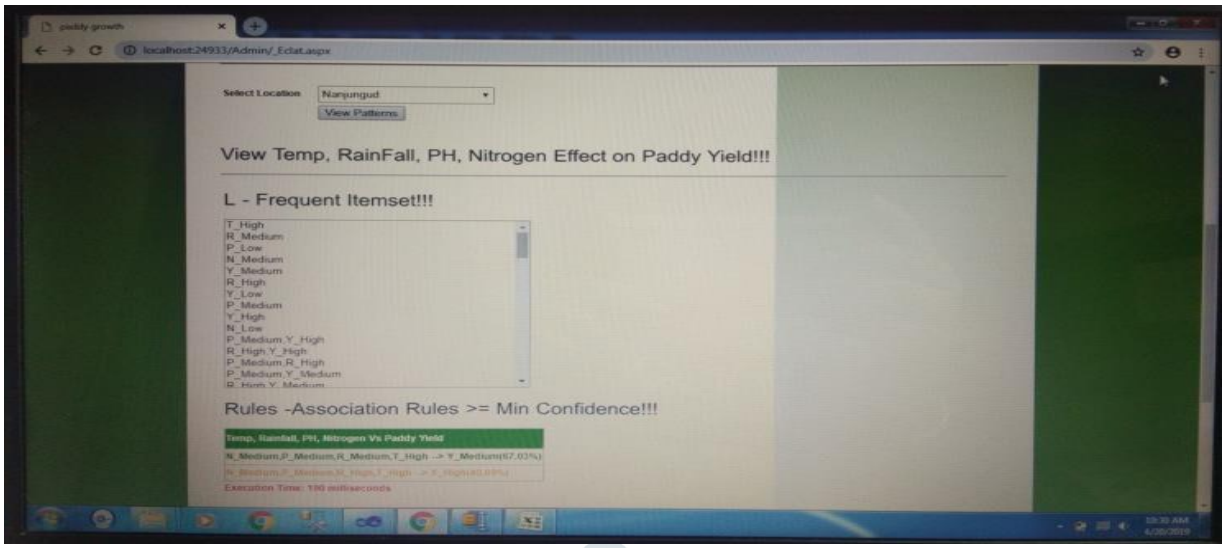

Fig 7 : Paddy Pattern Prediction Module

The above snapshot shows login module and the paddy prediction module with respect to temperature, rainfall, ph, nitrogen content.

CONCLUSION AND FUTURE SCOPE

The system to "predict the growth of a paddy using data mining techniques" has been flourished as favorable to the clients i.e, agricultural departments of various regions. The proposed system has achieved the goals and they are it reduces the manual work, efficient compared to the existing system and we can store large volume of data. We can add npk values of soil as attributes to the proposed system from soil testing lab. If any queries raises, the staff can communicate with the administrator directly with the addition of the new module.

REFERENCES

[1] Umid Kumar Dey, Abdullah Hasan Masud "Rice Yield Prediction Model Using Data Mining",IEEEE, 2017.
[2] Ami Mistry1, Vinita Shah Brief "Survey of data mining Techniques Applied to applications of Agriculture" , 2018.
[3] Monali Paul, Santosh K. Vishwakarma, Ashok Verma "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" , IEEE ,2018.
[4] Md. TahmidShakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty "Agricultural Production Output Prediction UsingSupervised Machine Learning Techniques", IEEE, 2018.
[5] D Ramesh , B Vishnu Vardhan "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering , Vol. 2, Issue 9, September 2018.
[6] Mucherino, Petraq Papajorgji, P. M. Pardalos ,"A surveyof data mining techniques applied to agriculture", 25 May 2017 Springer .
[7] Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider "Data Mining with neural networks for wheat yield prediction" ,Petra Perner, editor, Advances in Data Mining (Proc ICDM2018), 47–56, Berlin, Heidelberg, July 2018. Springer Verlag.