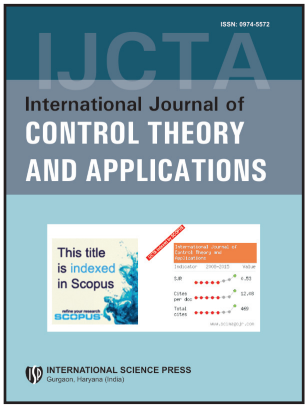

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 43 • 2016

Agriculture Crop Prediction System Based on Meteorological Information

Rageena P.M.[a], Aby Abahai T.[b] and Surekha Mariam Varghese[c]

[a-c]Department of Computer Science and Engineering, M.A College of Engineering, Kothamangalam, Kerala, India. Email: [a]rageenasalim@gmail.com; [b]abytom@gmail.com; [c]surekh.var@gmail.com

Abstract: Agriculture is the backbone of our economy, but Kerala is a consumer state which depends on other states for their needs and resources. The resources should be efficiently and properly utilized inorder to become a producer state. This work includes about the research and the building of an effective agricultural yield forecasting system based on real-time monthly weather information. This system provides a platform for right information at right time and also gives a clear idea about agriculture developments such as information on crop, fertilizer, soil. Due to the abnormal weather that happens every year and rapid regional climate change it is difficult to predict the agricultural crop production. Real-time weather information is urgently required for the development of agricultural yield forecasting system. The proposed system establishes the prediction system for soil, crop, and fertilizer prediction by process the weather data and datasets. Here this system collect weather data such as humidity, amount of rainfall, temperature, sunshine for each location entered by the user and establishes two types of prediction model inorder to predict the crop suitable for that location. The Prediction model consists of suitable classifier (Naive Bayesian). Proposed system implements two crop prediction model i.e., Naive Bayes crop prediction model and Fuzzy Naive Bayes prediction model. Also proposed system compares the accuracy of two model using confusion matrix.

Keywords: Naive Bayesian Classifier, Prediction System, Fuzzy logic, Fertilizer and Soil Prediction, Document Annotation.

1. INTRODUCTION

In recent years, different emerging technology and applications are introduced for information technology. Information technology has become more and more a part of our day to day life, especially for agriculture. Yield prediction is very popular among farmers these days, which may contribute to the proper selection of crops for their agriculture developments activities. Based on farmer's experience on a particular field and crop we perform the yield prediction in the earlier stage. Proposed system uses data mining techniques in order to predict the crop name by, analyzed the crop, fertilizer, soil datasets. The proposed contains about the research and the building of an effective agricultural yield forecasting system based on real-time monthly weather information. This system provides a platform for right information at right time and also gives the clear idea about our agriculture developments such as information on crop, fertilizer, soil. Proposed system collect weather data such as humidity, amount of rainfall, temperature, sunshine and establishes a Prediction model [6]. The agricultural observation induces farmers to adjust their agricultural production and shipment. Inorder to do business in the country, such as agriculture observations; there have been a lot of research and system building. Agricultural forecast is becoming more complex and more difficult due to global warming and abnormal weather.

Climate change is a major problem that affects to the agriculture development. The causes of climate change depends on many factors such as natural causes like changes in earth revolution, variations in solar system and human activities, concentrations of carbon dioxide and certain other harmful atmospheric gases have been increasing. Crop production is highly dependent on variation in weather and therefore any change in climate condition will effects on crop yields and productivity [1]. Elevated temperature and carbon dioxide affects the biological processes like respiration, photosynthesis, plant growth, reproduction, water use etc. Bayesian network classifiers are often used for classification problems. The Naive Bayes model is a simple Bayesian network classifier [5] that assumes the predictors are independent in given class value. Proposed system implements two crop prediction model i.e., Naive Bayes crop prediction model and Fuzzy Naive Bayes prediction model. Also proposed system compares the accuracy of two model using confusion matrix.

2. RELATED WORK

Some of the related papers that currently exist are given below i.e., 'Machine Learning Perspective for Prediction Agricultural Droughts [5]' this paper addresses the supervised learning mechanism and its one of the method Naive Bayesian Classification. Mainly Naive Bayesian classification is to evaluate the parameters and it often performs better in many complex real world applications. Due to drought it may cause more losses to society compared to other natural disasters [5] and it affects large number of people. Due to the frequent occurrence of drought poses an increasingly severe threat to the agricultural production and developments. Figure 1 shows the architecture of the proposed system it explains the working of the proposed system. Here monitoring of agricultural drought accurately done by integrating multi-source information. The Bayesian model attempts to characterize drought in a more accurate and comprehensive way. This model application using a variety of methods and data, there is still some work to be done in the future research because of the complex spatial and temporal characteristics of drought.

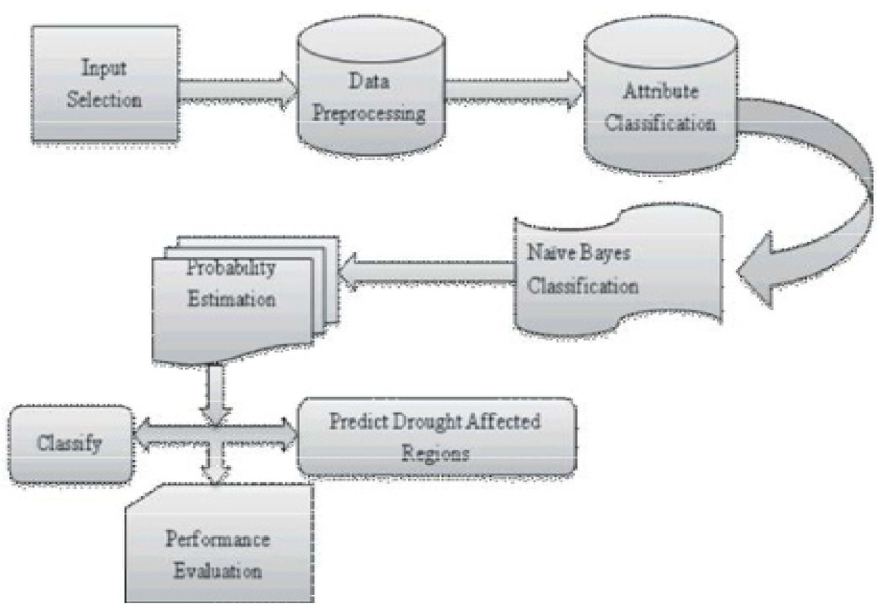

Figure 1: Architecture of Proposed System

Another existing system is the 'Tomato Crop Expert System Based on Artificial Intelligence [2]'. Nowadays Tomato is the main grown vegetable crop in world. This Expert System contains of two parts i.e., Tomato Information System and Tomato Crop Expert System where in Information system, the user can get all the relevant information about the tomato crop i.e., different diseases, symptoms, preventions, pests, Virus of Tomato fruits and plants. User having an interaction with the expert system through the Advisory System [2], the user has to answer the questions asked by the Expert System. Depends on the response given by the user trough online the expert system decides the disease and displays its control measure of disease.

Another method is 'Artificial Neural Networks for Rice Yield Prediction'. Here neural networks [6] have been gaining a great deal of importance and they are used in the different areas such as prediction and classification; the areas where regression and other statistical models are traditionally being used. This study presents a useful system consists of neural networks and their statistical counter parts used in the area of prediction of crop yield. Recently, Researchers have developed several forecasting and prediction models of various crop yields in relation to different parameters of artificial neural networks and by combining ANN and statistical techniques such as linear regression technique [6]. Agricultural management developers need simple and accurate prediction techniques to predict rice yields in the planning process.

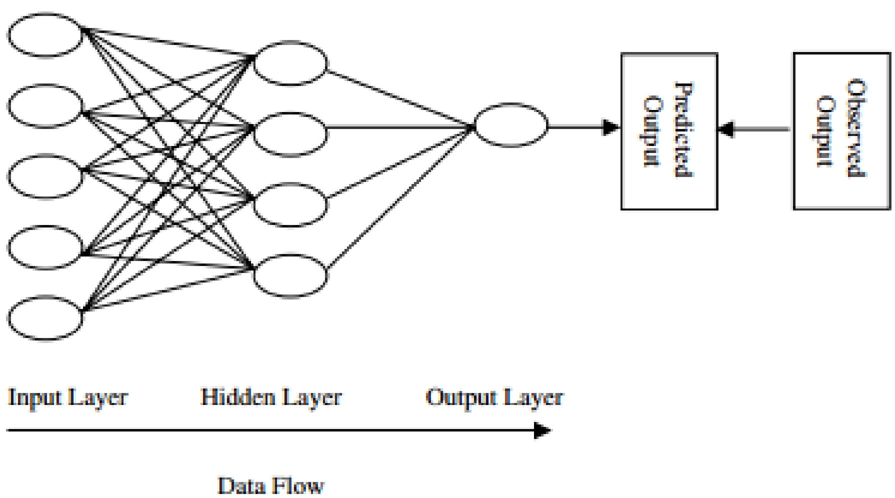

Figure 2: Neural Network

The input and output variables are the nodes correspond to input and output layers. Data move between layers and each data have weighted connections. A node accepts data from the previous layer and after that they calculate a weighted sum of all its inputs. ANN [6] may become over-trained, causing it to memorize the training data and resulting in poor predictions. Learning rate means the amount of weights change during a series of iterations to bring the predicted value within an acceptable range of the observed value. The proposed system contains about the research and the building of an effective agricultural yield forecasting system based on real-time monthly weather information. This system provides a platform for right information at right time and also gives the clear idea about our agriculture developments such as information on crop, fertilizer, soil. Proposed system collect weather data such as humidity, amount of rainfall, temperature, sunshine and establishes a prediction model [6]. The prediction model consists of suitable classifier (Naive Bayesian).

3. PROPOSED WORK

In this section the data set used, as well as the methods for the prediction is discussed.

A. Architecture

The proposed system consist of five modules sensor application, crop prediction, fertilizer prediction, soil prediction, score calculation and document annotation. Proposed system Architecture consist of Input screen, Prediction system, and Output stream each elements have its own functionalities. Through input screen the inputs are given to the prediction system also through output screen the outputs are generated .The prediction system consists of knowledge base, rule based system and Naive Bayesian classifier [3]. Knowledge base consists of the dataset and other related information about the crops. Rule based system consists of different rule used for fuzzy prediction system. Architecture of the proposed system is given below

1. *Sensor Application:* This application collects the corresponding weather parameters such as humidity, sunshine, rainfall, temperature for each location entered by the user and sends back to the server. Server stores this data in their database. The prediction model collects the data from the server and performs prediction with the help of Naive Bayesian. Here the Sensor Application collects real time weather information by using Yahoo Api from Yahoo weather web service. This real time weather information is used for the prediction.

2. *Crop Prediction:* The most common goal of cropping is to produce a greater crop yield [4] on a given piece of land by making use of resources that would otherwise not be utilized by a single crop. Careful planning is required, taking into account the soil, climate, crops, and varieties. For the prediction proposed system use a crop dataset consist of humidity, temperature, sunshine, rainfall. For each location proposed system finds the weather information and according to that information and crop dataset this system predicts the crop for that location. Admin can add the crop details and keep it up to date. Based on weather history, system predicts the suitable crops.

3. *Fertilizer Prediction:* Here proposed system predicts the fertilizer for each disease based on the given symptoms that enter by the user by using naïve bayesian classifier. For the prediction proposed system give a fertilizer dataset consist of the different symptoms. Admin can add the fertilizer details and keep it up to date. Identify each fertilizer based on symptoms.

4. *Soil Prediction:* Here proposed system predict the soil and crop suitable for a given location based on the given nitrogen, phosphorus content by using Naïve Bayesian classifier. For the prediction proposed system give a soil dataset [3] consist of the different elements in the soil i.e. phosphorus, potassium, calcium, ph level. Admin can add the soil details and keep it up to date. Predict the soil suitable for the crop.

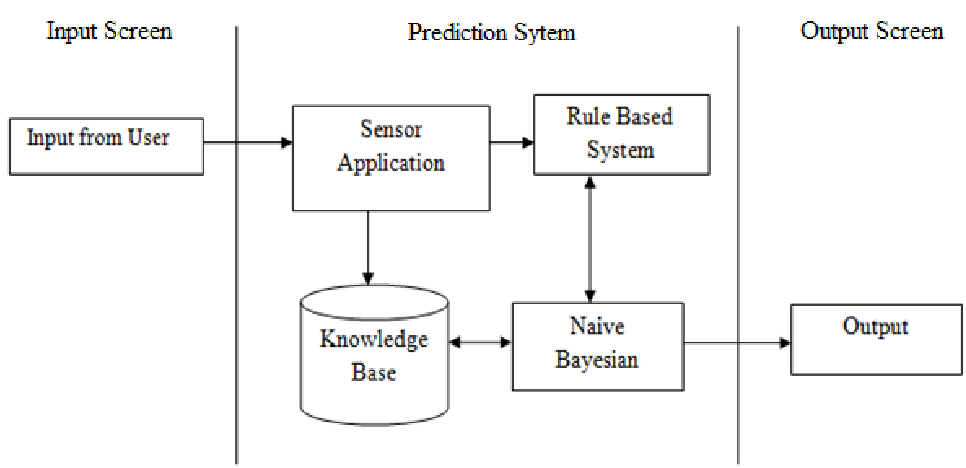

Figure 3: System Architecture

5. *Score Calculation:* Score calculation module is used to know the score of each location that search by the user. By knowing the score of each location thereby proposed system can findout which location having maximum score. Thus proposed system can concentrate more on the crop, fertilizer, soil of the maximum score location. Score of each location can be seen at the database. It can view by the admin and he maintains this module.

6. *Document Annotation:* Document Annotation [10] means sharing of data i.e., useful for effective searching and to support advanced applications. Here by using this searching mechanism (annotation) the user can search any data through this information system related to agriculture. User may need to give only the keyword in the search box. Then proposed system gives the txt files to the users that contain the corresponding keyword. The system has three databases: Document collection, Query collection, Annotation database [10]. Proposed system split the query terms and then searches them within the annotation database. Score of an attribute in a document is found by using a probabilistic approach. Searching can be done by using the formula given below. Total score of an attribute is the product of content score and query score. Attributes with high score is more relevant to that document.

Score of an attribute is given by:

Score[Attribute] = ContentScore[Attribute] × QueryScore[Attribute]

Where Query Score can be defined as:

QueryScore[Attribute] = P(Attribute in Query Collection) ÷ 1-P(Attribute in Query Collection)

P(Attribute in Query Collection) = COUNT(Queries in Query Collection with Attributes) + 1÷COUNT(Queries in Query Collection

Where Content Score can be defined as:

ContentScore[Attribute] = P(Attribute in Document Collection)/1-P(Attribute in Document Collection)

P(Attribute in Document Collection) = COUNT(Documents in Document Collection with Attribute)+1/ COUNT(Documents in Document Collection)+1

C. Prediction System

A prediction or forecast, is a statement about an uncertain event. It is often, but not always, based upon experience or knowledge. Predicting [4] any event requires knowledge about past performance. In machine learning, Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all Naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For some types of probability models, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. Proposed system provides a platform for right information at right time. It gives the clear idea about our agriculture developments such as information on crop, fertilizer, soil. Here proposed system collect weather data from sensor application such as humidity, amount of rainfall, temperature and we establish two types of prediction model.

1. Crop prediction system with Simple Naïve Bayesian Classifier

A naive bayes classifier is a simple probabilistic classifier based on applying bayes' theorem with strong (naive) independence assumptions. Here in this prediction mechanism we use the Naïve Bayesian [3] classification algorithm steps. Here we collect the weather data from yahoo weather site for each location that enter by the user as the input to the Naïve Bayesian classifier and the classifier predict the crop that suitable for that location. So here the input to naïve Bayesian is the weather parameter of the location such as, humidity, rainfall, sunshine, rainfall. The prediction model process the input using naïve bayesian classifier and crop dataset then gives the results. Using simple Naïve Bayesian [5] we perform prediction for fertilizer and soil. Crop dataset contains the different crops and their weather conditions required for the prediction, the dataset is prepared with the help of the agriculture information given from the Agriculture Office near Perumbavoor, Kerala. Naïve Bayesian classification algorithm is based on Bayes' theorem i.e.,

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Where X be a data tuple and belongs to class C, P(C|X) is the posterior probability of C conditioned on X, P(C) is the prior probability of C, P(X) is the prior probability of X

Naïve Bayesian algorithm steps

1. Set the Dataset
2. Set the Class label Ci (for i= 1,2……n)
3. Find out the count of Class label P(Ci) from dataset
4. Specify the Tuple X (contain n attribute)
5. Calculate the Class label of Tuple X

    5.1. Find P(X|Ci) for each attribute in Tuple X

    Repeat step 5.1 for each attribute in Tuple X

    5.2. Multiply each P(X|Ci) i.e.,

    P(X1|Ci) *P(X2|Ci) *…….P(Xn|Ci) for i= 1,2,…..n

    5.3. Maximize the Equation P(X|Ci)*P(Ci) for i = 1,2,…n

    5.4. From the results select the class label that have maximum P(X|Ci)P(Ci) value

Here the tuple X is the input given to the Naive Bayesian classifier i.e; in proposed system weather parameter is the input to the naive bayes classifier.

Example: Here data tuples described by the attributes humidity, sunshine, rainfall, temperature. Class label attribute status (crop growth) has two distinct values ({yes, no}). C1 correspond to the class status = yes and C2 correspond to status = no. Tuple that wish to classify is X = (humidity= low, temperature = 20, rainfall = medium, sunshine = warm). Inorder to Perform Naivebayes algorithm first they need to set a suitable dataset. Assume the dataset and perform prediction. Let us assume,

- P(status= yes) = 9/14 = 0.643, P(status = no) = 5/14 = 0.357

To compute P(X|Ci), for i = 1, 2 compute the following conditional probabilities

- P(humidity = low | status = yes) = 2/9 = 0.222
- P(humidity = low | status = no) = 3/5 = 0.600
- P(temperature = 20 | status = yes) = 4/9 = 0.444

- P(temperature = 20 | status = no) = 2/5 = 0.400
- P(rainfall = medium | status = yes) = 6/9 = 0.667
- P(rainfall = medium | status = no) = 1/5 = 0.200
- P(sunshine = warm | status = yes) = 6/9 = 0.667
- P(sunshine = warm | status = no) = 2/5 = 0.400

Using these probabilities, proposed system obtain

– P(X|status = yes) = P(humidity = low | status = yes × P(temperature = 20 | status = yes) × P(rainfall = medium | status = yes) × P(sunshine = warm | status= yes) = 0.222 × 0.444 × 0.667 × 0.667 = 0.044

Similarly,

– P(X|status = no) = 0.600 × 0.400 × 0.200 × 0.400 = 0.019

To find the class Ci, maximizes P(X|Ci)P(Ci)

– P(X|status= yes)P(status= yes) = 0.044 × 0.643 = 0.028

– P(X|status= no)P(status = no) = 0.019 × 0.357 = 0.007

Here the value of status = yes is greater than status = no. Therefore the naive Bayesian classifier predicts status = yes for tuple X.

2. Fuzzy Naive Bayes Crop Prediction System

Our attempt is to forecast yield prediction with the help of fuzzy logic based approximate reasoning. This process uses the concept of a pure fuzzy logic system where the fuzzy rule base consists of a collection of fuzzy IF–THEN rules. In order to build our models proposed system defined the fuzzy sets consist of four parameters: relative humidity, rainfall, temperature and sunshine are the input variables for our model. A fuzzy inference system consists of inputs and their membership functions, output and its membership functions and rules for the memberships. Fuzzy inference [7] is the process of formulating the mapping from a given input to an output using fuzzy logic. The project is going to have two input variables and one output variable. The inputs variables are the rainfall and temperature at a particular time and the output will be the crop growth. The input parameters for the fuzzy logic are collected from the yahoo weather webservice and are used for crop yield evaluation. Using Fuzzy logic [7] membership function proposed system generates fuzzy rules and are used to generate output. Rainfall and temperature are chosen as input because it was proved that they are the factors that influence more the occurrence of crop yield. The other input parameter humidity and sunshine is given directly to the dataset. The output of the fuzzy logic is used to generate a dataset and that is given to the input of naive bayes. The inputs variables are (rainfall and temperature) are grouped into five using the linguistic variable terms which are:

NL:- Negative large: Very Low

NS:- Negative small: Low

ZE:- Zero: Normal

PS:- Positive small: High

PL:- Positive large: Very High

For a particular input variable the variable will either belong to one of these or at most two of it with a membership value.

(a) *Temperature Equation (Input 1):* The average temperature ranges for the crops are from 10°*C* - 40°*C*.The equations given below produce the values of membership value for any temperature value in between 10 &40. Inorder to construct the equation proposed system use triangular member function [8] of fuzzy logic is used. After the equation we need to build a graph showing the membership value, we have eight lines in the graph therefore we must formulate eight equations to implement the graph. Some of the linguistic variables [9] (very low, low, normal, high and very high) might have the same temperature range but they will surely have different equations. A membership function for a fuzzy set A on the universe of discourse X is defined as μ_A:X → [0,1], where each element of X is mapped to a value between 0 and 1. This value, called membership value or degree of membership. Triangular function: defined by a lower limit a, an upper limit b, and a value m, where a < m < b.

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{m-a}, & a < x \le m \\ \dfrac{b-x}{b-m}, & m < x < b \\ 0, & x \ge b \end{cases}$$

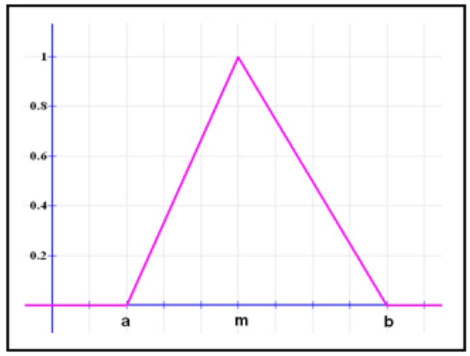

Figure 4: Triangular Membership Function

(Very Low): (10 <= Temprature <= 25) = Member (m1) = -.2T + 5

Low): (20 <= Temprature <= 25) = Member (m2) = .2T - 4

(Low): (25 <= Temprature <= 30) = Member (m1) = -.2T + 6

(Normal): (25 <= Temprature <= 30) = Member (m2) = .2T - 5

(Normal): (30 <= Temprature <= 35) = Member (m1) = -.2T + 7

(High): (30<=Temprature <=35) = Member (m2) = .2T-6

(High):(35 <=Temprature <=40) = Member (m1) = -.2T+8

(Very high)=(35 <=Temprature <=40) = Member (m1) = -.2T-7

Where *T* denotes Temperature and *m* denotes membership value respectively.

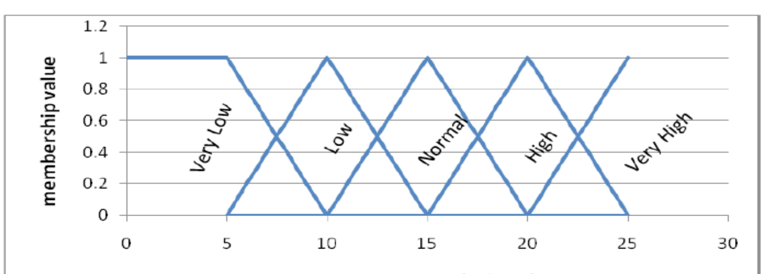

Figure 5:Membership function of temperature

(b) *Rainfall Equation (Input 2):* Rainfall range 0 – 12 cm. Here also proposed system find equation for calculating membership function [8] by using triangular membership function.

(Very Low): (0 < Rainfall <= 4) = Member (m1) = -.25R + 1

(Low): (0 < Rainfall <= 4) = Member (m2) = .25R

(Normal): (4 < Rainfall < 8) = Member (m1) = -.25R + 2

(Normal): (4 <= Rainfall <= 8) = Member (m2) = .25R - 1

(High): (8 <= Rainfall < 12) = Member (m1) = -.25R + 3

(Very high): (8<=Rainfall <=12) = Member (m2) = .25R-2

(c) *Output Parameter:* By using the membership degree of the inputs proposed system derive the fuzzy rules. Proposed systems give the output of the fuzzy logic as the input to the Naive Bayesian. Using the Bayesian network predict the result by using the naive algorithm. Output of the fuzzy logic is fuzzy rules that are given to the naïve bayes as input. For the prediction proposed system construct a dataset by using these fuzzy rules.

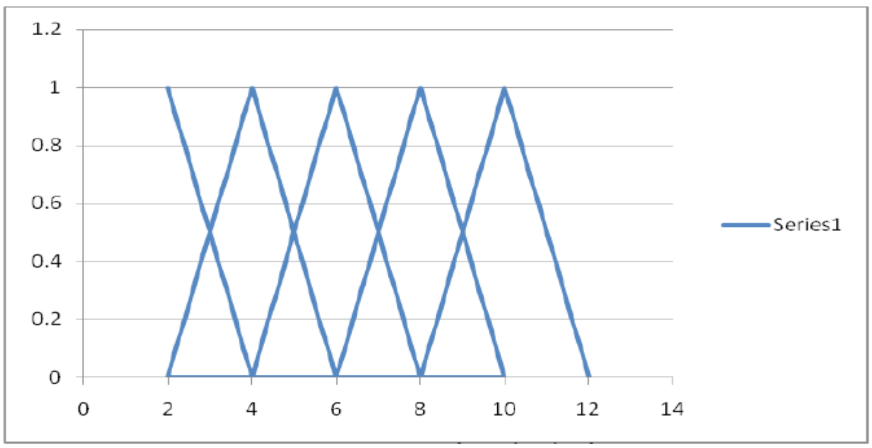

Figure 6: Membership function of rainfall

Example: Assuming that the Temperature and the Rainfall (which are factor that determines the crop growth) are 12.5°*C* and 5 respectively, the procedure or step that will be taken to know the crop growth is stated bellow:

1. Proposed system is going to determine the group that each input falls in and their membership value for that group:
   - For 12.5°C temperature proposed system can deduce from the graph or by using one of the equations. i.e., the temperature of 12.5°C falls between 10 and 25, therefore the equation to calculate the membership will be -0.2T + 5 for very low and 0.2T - 4 for low. By putting the value of the temperature in both equations will get
     - -.2 * 12.5 + 5 = 2.5(Very low) and .2 * 12.5 - 4 = -1.5(low)
   - For 5 cm rainfall it falls between 4 and 8, therefore the equation to calculate the membership will be -0.25R+2 for normal and 0.25R-1 for normal
     - -.25 * 5 + 2 = .75(normal) and .25 * 5 - 1 = .25(normal)
2. The next step is to determine how a particular value of temperature and rainfall will produce a particular value of Crop growth. The table Table 1 given below shows how the two input variable and the output value are related. The table was constructed using a data derive from United States department of Agriculture (USDA).

Table 1
A table showing the relationship between temperature, rainfall and cropgrowth

TP \ RF	NL	NS	ZE	PS	PL
NL	NL	NL	NS	NS	ZE
NS	NL	NL	NS	ZE	ZE
ZE	NS	NS	ZE	ZE	PS
PS	NS	ZE	ZE	PS	PS
PL	ZE	ZE	PS	PS	PL

The value of our temperature ($12°C$) shows that they can categorize the temperature into very low and low and our rainfall(4) is categorized as normal only i.e., very low is Negative Large(NL), low is Negative Small(NL) and normal is zero(ZE). So according to the above table we get the fuzzy rules are given below.

If (Temperature is verylow [2.5] and rainfall is normal [.75]) then crop growth is low

If (Temperature is low [1.5] and rainfall is normal [.75] then crop growth is low

Like this system derive the rules and after this they perform prediction using naive bayes with these fuzzy rules.

D. Dataset

A **dataset** is a collection of related sets of information that is composed of separate elements but can be manipulated as a unit by a computer. For the proposed system, they need three dataset (1) crop dataset (2) fertilizer dataset (3) soil dataset. Crop dataset contain weather parameter of each location entered by the user i.e., humidity, sunshine, rainfall, temperature. Fertilizer dataset contains the diseases and the symptoms for corresponding diseases and the cure for those diseases. In soil Dataset it contain the different components in the soil along with their value i.e., ph level, potassium, phosphorus, calcium. Table 2 given below is an example of crop dataset. For the creation of dataset, proposed system collect data from Agriculture Office Perumbavoor and from Karshika keralam.com.

Table 2
Crop Dataset

cid	name	temparature	humidity	rain_fall	sun_shine	status
1	rice	15-25	high	medium	bright	yes
2	rice	20-34	high	low	bright	yes
3	rice	15-25	low	high	warm	yes
4	Rubber	20-34	medium	medium	warm	no
5	Rubber	15-25	low	high	bright	no
6	Rubber	20-34	medium	high	warm	no
7	Rubber	20-40	low	low	bright	no
16	rice	25-30	low	medium	bright	no
17	rice	15-30	high	medium	bright	no
18	rubber	25-29	low	high	warm	no
19	mango	20-30	high	medium	bright	no
20	cocont	25-32	low	low	strong	no
21	pepper	25-30	low	medium	bright	no
22	banana	15-30	high	medium	bright	no
23	tea	30-39	low	high	warm	no

4. EXPERIMENTAL RESULT

This section gives the important screen shots from the different phase implementation of the thesis. They are: sensor application, crop prediction, fertilizer prediction, soil prediction and fuzzy prediction system.

	ANALYSIS DETAILS
	Search Place :Kottayam Temperature:31.113599999999998 Humidity:56 Pressure:1008.0 Rain fall:0 Weather condition:Partly Cloudy
	Next

Figure 7: Weather information

By using the weather information, proposed system perform crop prediction using Naïve Bayesian. Prediction screenshot is given below it contains the different crops that are suitable for the location.

	ANALYSIS DETAILS
Place	Kottayam
Sun light	warm
Crop name:Rice, result:YES Crop name:Rubber, result:NO Crop name:Coconut, result:YES Crop name:Turmeric, result:YES Crop name:Tomato, result:YES Crop name:Chilli, result:YES Crop name:Beans, result:YES Crop name:Banana, result:YES Crop name:Pineapple, result:YES Crop name:Tea, result:NO Crop name:Pepper, result:YES Crop name:Vanilla, result:YES Crop name:Cashewnut, result:NO	

Figure 8: Crop prediction

Figure 9: Fertilizer prediction result

ANALYSIS DETAILS	
Soil name is	loamy soil
Crop name is	Rice

Figure 10: Soil prediction result

	ANALYSIS DETAILS
Place	Kottayam
Another possibility of crop	cashewnut
Another possibility of crop	Cocoa
Another possibility of crop	Pappaya
Crop name:Rice, result:YES Crop name:Coconut, result:YES Crop name:Turmeric, result:YES Crop name:Tomato, result:YES Crop name:Chilli, result:YES Crop name:Beans, result:YES Crop name:Banana, result:YES Crop name:Pineapple, result:YES Crop name:Tea, result:NO Crop name:Pepper, result:YES Crop name:Vanilla, result:YES Crop name:Cashewnut, result:NO Crop name:Rubber, result:NO	

Figure 11: Fuzzy Naïve Bayesian prediction result

By clicking on the Fuzzy system gives the output of the Fuzzy Naive Bayesian prediction system, it also takes the input as the weather parameter and the gives output as crop suitable for the location. Result of the Fuzzy Naive Bayesian prediction system is given above in the Figure 11. Other than the above mentioned prediction system proposed system has one another module that is Document Annotation [10] i.e., the user can search any data through this information system related to agriculture. User may need to give only the keyword in the search box.

5. COMPARATIVE STUDY

Comparative study of the two prediction system i.e., Naive Bayesian system and Fuzzy Naive Bayesian system are given below. From the results of the proposed system, we can see Fuzzy naive bayesian has high number of outputs than simple Naive bayes. Inorder to calculate performance measure here proposed system chooses Fuzzy naive bayes prediction system and Naive Bayes by using confusion matrix method. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Let's now define the most basic terms, which are whole numbers (not rates):

- **true positives (TP):** These are cases in which predicted yes (they correctly predict the crop name)
- **true negatives (TN):** predicted no, and they incorrectly predict the crop name.

- **false positives (FP):** predicted yes, but they incorrectly predict the crop name. (Also known as a "Type I error.")
- **false negatives (FN):** predicted no, but they actually correctly predict the crop name. (Also known as a "Type II error.")

n=55	Predicted: NO	Predicted: YES	
Actual: NO	TN = 10	FP = 10	20
Actual: YES	FN = 5	TP = 30	35
	15	40	

Figure 12: Fuzzy Naïve Bayes Confusion Matrix output

n=55	Predicted: NO	Predicted: YES	
Actual: NO	TN = 10	FP = 10	20
Actual: YES	FN = 10	TP = 20	30
	20	30	

Figure 13: Naïve Bayes Confusion Matrix output

What can we learn from this matrix?

- There are two possible predicted classes: "yes" and "no". If we were predicting the crop suitable for a location, for example, "yes" would mean they correctly predict the crop name and "no" would mean they incorrectly predict the crop name.
- The classifier made a total of 55 predictions (e.g., 55 predicted value is compare with the dataset that given from the Agriculture Office Perumbavoor).
- Out of those 55 cases, the classifier predicted "yes" 40 times, and "no" 15 times.
- In reality, 35 cases are correctly predicting the crop name for each location, and 20 cases incorrectly predict the crops name.

This is a list of rates that are often computed from a confusion matrix for a binary classifier:

- **Accuracy:** Overall, how often is the classifier correct?
  - Fuzzy Naive = (TP + TN)/total = (30 + 10)/55 = 0.73
  - Simple Naive = (TP + TN)/total = (20 + 10)/55 = 0.54
- **Misclassification Rate:** Overall, how often is it wrong?
  - (FP + FN)/total = (10 + 5)/55 = 0.27
  - (FP + FN)/total = (10 + 10)/55 = 0.36

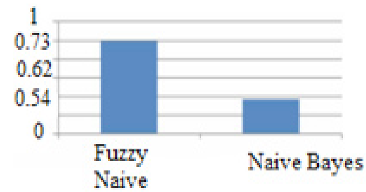

Figure 14: Performance Measure

6. CONCLUSION

The Fuzzy Bayesian model attempts to characterize crop in a more accurate and comprehensive way than simple naïve bayes. This work includes the research necessary to constructing reliable prediction through naïve Bayesian classifier. Crop growth prediction model is a very effective tool for predicting possible impacts of climatic change and yield crop growth models are useful for solving various practical problems in agriculture. This work contains about the research and the building of an effective agricultural yield forecasting system based on real-time monthly weather.

7. SCOPE FOR FURTHER WORK

In the existing implementation, the prediction is based on weather information from Yahoo API and also prediction is done by using Naïve Bayesian classifier. Another area for further work is to use any real time sensor for collecting weather data from each location and also use suitable classification algorithm. Also future work includes considering geographical area using world geographic information system for global harvest prediction

REFERENCES

[1] David B. Lobell and Marshall B. Burke(2010) On The Use Of Statistical Models To Predict Crop Yield Responses To Climate Change, Elsevier,1.

[2] Prof. M.S. Prasad Babu, N.V. Ramana Murty, S.V.N.L.Narayana (2010), "A Web Based Tomato Crop Expert Information System Based On Artificial Intelligence And Machine Learning Algorithms", International Journal Of Computer Science And InforMation Technologies, Vol. 1, (Issn: 0975-9646), Pp6-15.

[3] P.Bhargavi,M.Sc.,M.Tech and Dr.S.Jyothi, M.Sc., M.S, Ph.d.(2009), Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils, International Journal of Computer Science and Network Security, VOL.9 No.8, 117-122.

[4] Tanuja R. Patil, Shamshuddin K, Rajashekhar Patil, Sadanand P (2016) a decision support system for farmers to get high crop yield, International Conference on Computational Techniques in Information and Communication Technologies.

[5] K. Sriram and K. Suresh, Machine Learning Perspective for Predicting Agricultural Droughts Using Naive Bayes Algorithm, Special Issue on Innovations in Information, Embedded and Communication Systems): 178-184, 2016.

[6] Yoav Freund and Robert E. Schapire(2007) Artificial neural networks for rice yield prediction in mountainous regions, Journal of Agricultural Science, 145, 249–261.

[7] Zadeh, L. A., Fuzzy Sets and Information Granularity, Selected Papers by L.A. Zadeh; Pp 443-448, 1979

[8] Hasan, M.,Tsegaye,T., Shi, X., Schaefer, G. and Taylor, G.,Model for predicting rainfall by fuzzy set theory using USDA scan data, 2008

[9] Zadeh, L. A., The Concept of a Linguistic Variable and its Application to Approximate reasoning-II, Information Science 8, Pp 301-357, 1975a

[10] Eduardo J. Ruiz, Vlis Hristidis, and Panagiotis G. Ipeirotis, Facilitating Document Annotation Using Content and Querying Value, Ieee Transactions on Knowledge and Data Engineering, Vol. 26, No. 2, February 2014