ISSN:2320-0790

An International Journal of Advanced Computer Technology

Ameliorated Methodology for the Design of Sugarcane Yield Prediction Using Decision Tree

[1]Ashwinirani, [2]Dr. B. M. Vidyavathi

[1]PG Scholar, [2]Professor
Department of Computer Science & Engineering
BITM Bellary

Abstract- The productivity and quality of a Crop depends upon several parameters including but not limited to environment parameters like temperature, humidity, rainfall, wind, the quality and quantity of the pesticide, type, quantity and quality of fertilizer and so on. High productivity and yield is of utmost essentiality in a country like India where the growing population demands more grains that the field produces right now. Thus there is an urgent need to bring in more scientific studies in this direction. As amount of agricultural fields is limited and cannot be increased, it is important to get maximum productivity out of each crop. Past records of yield of a crop in different months, the known crop diseases and the weather plays a good indication as to what should be the ideal condition for high yield of a crop. However the dependencies among the parameters are so fuzzy that it is difficult to estimate the exact yield and need for exact quantity of pesticides and fertilizer to minimize the risk of crop disease, crop failure and to improve the yield. In this work a Framework is proposed that can be used to provide guidance to the farmer to detect the fault at the earlier stage to reduce losses. Decision tree, an efficient, globally competitive and vibrant classifier is used to predict sugarcane yield.

Keywords:- Pre-processing, Crop yield prediction, Data mining, classification, crop productivity, climatic factor

1. INTRODUCTION:

Crop yield prediction has been a topic of interest producers, consultants, and agricultural related organizations. The crop yield is unified bio-socio-system comprised of complex interaction among the soil, air, water, and the crops grown in it, where comprehensive model is required which are possible only through classical engineering expertise. Crop forecasting is the art of predicting crop yields (tons/ha) and production before the harvest takes place, typically a couple of months in advance. Crop forecasting is based on various kinds of data collected from different sources: meteorological data, agro-meteorological (phenology, yield), soil (water holding capacity), agricultural statistics. Based on both meteorological and the agronomic data and several indices derived which are deemed to be relevant variables in determining crop yield production, for instance crop water satisfaction, average the soil moisture, surplus, excess moisture.

Timely and accurate crop yield forecasts are essential for crop production, marketing, storage, and transportation decisions and they help managing the risk associated with these activities. Understanding the stochastic behavior of crop yield is an essential part all levels. At the country level, yield forecasting is used in the determination of national food security, import and export plans, crop insurance policy and government aid for farmers. The timely evaluation of potential yields is increasingly important because of the huge economic impact of agricultural products on world markets and the strategic planning. The work of Savin, which developed a simple method for any place of the world, combines the available local climatic and agronomic data, for producing the best possible yield forecast using multiple linear regressions. It aim sat giving support to the Food Security and Food Aid policy by improving information on the crop prospect, particularly in the regions of the world stricken by frequent food shortage. After the initial development and demonstration the developed methods and systems are now being tested on a preoperational basis.

At farm level, knowledge of the yield forecast before the harvest time gives the producers information to plan their farming activities and marketing strategies for their products. For example, predicting shortfalls of the crop yields for the coming year gives the government time to initiate appropriate policies and the farmers to make crop selection decisions.

Accurate prediction of corn yield is important for government policy making. The self-sufficiency has in general been the guiding principle for agricultural policy. Only rise in corn's importance as feed pushed the livestock sector as a major force in corn policy considerations. Thus, having an accurate prediction model generally benefits the policy making body for more robust and realistic plans.

Agriculture yield primarily depends on the weather conditions, the Water management, the soil, pests, weeding, the prop paring, the diseases and, the planning of harvest operation, and the likes. Effective management of these factors is necessary to estimate the probability of such unfavorable situation and to minimize the consequences. The accurate and reliable information about historical crop yield is thus vital for decisions relating to agricultural risk management. Historical crop yield information is also important for supply chain operation of companies engaged in industries that use agricultural produce as raw material. The livestock, the food, animal feed, the chemical, poultry, fertilizer pesticides, seed, paper and many other industries use agricultural products as intergradient in their production processes. The accurate estimate of crop size, risk that helps these companies in planning supply chain decision like production scheduling. Business such as the seed, the fertilizer, the agrochemical and agricultural machinery industries plan production and marketing activities based on crop production estimates.

Although numerous models have been developed for crop yield prediction, the majority of them only deals with climate related data. Several reasons for yield gap of crops are the following: first and foremost, the erratic, unpredictable weather conditions that affect crop growth yields. Tropical storms can easily destroy crops and adversely affect crop production. Second, Asian farmers tend to use less than the recommended amounts of fertilizers because they lack the capital to purchase inputs. Farmers also reported that yield gaps are explained by poor seed quality, low seed replacement and ineffectiveness of recommended agronomic practices. Other causes include pest incidence, poor agricultural extension services that contribute to farmers' insufficient access to improved technology and/or technical information, and poor cultural management practices.

The against this background, there is need to develop an crop prediction model framework that can (1) pre-process and fuse potential predictor raw data from multiple sources, (2) provide an accurate prediction of crop yield, and (3) Learn useful prediction policies for decision planners, particularly for Provincial Agriculturists.

After the analysis made from the existing approaches to crop yield prediction, in this paper we proposed a methodology to increase the sugar cane yield. The paper consists of the following sections, Section 2 provides review of related literature, Section 3 discusses the proposed methodology and Section 4 discusses experimental results, followed by that conclusion.

2. REVIEW OF RELATED LITERATURES

A simulation model characterizes the mathematical relationships intrinsic to the data set from previous results. This method can generate results under various conditions assuming extensive information used to develop and test model. However, in the agricultural data, information is rather sparse or hard and completely not available. Because of the limitation, the regression approach is the common approach for predicting yield across large area of India. Furthermore, the most investigated statistical crop-yield-weather models are multivariate regression models (Yu, 2011, cited by Zhu, et.al, 2012). The agro-meteorological crop yield forecasting using a multiple regression was introduced by Gommes to develop an approach used by FAO and a number of developing countries for crop forecasting that would provide a good compromise between input requirements and ease of validation. However, considering the inherent and irreparable disadvantages of the multiple regression models, such as variable interdependence, the stringent linearity and the normality assumptions, a more scientific methodology to incorporate weather data into crop yield models, is still under exploration, and remains of great importance to government, and private sector insurers, and reinsurers (Zhu, et.al, 2012)[1].

A decision tree partitions the input space of a data set into the mutually exclusive regions each of which is assigned label, a value or an action to characterize its data points (Othman, Yau, 2007) [2]. The decision tree mechanism is transparent and we can follow a tree structure easily to see how the decision is made. A decision tree is a tree structure consisting of the internal and the external nodes connected by the branches. The internal node is a decision making unit that evaluates a decision function to determine which child node to visit next. The external node, on the other hand of tree, has no child nodes and is associated with a label or value that characterizes the given data that leads to its being visited. However, many decision tree construction algorithms involve a two - step process. First, a very large decision tree is grown. Then, to reduce large size and over fitting the data, second step, the given tree is pruned. The pruned decision tree that is used for classification purposes is called the classification tree.

The work of Uno et.al. (2005)[3] used agronomic variables, that focusing on nitrogen application and the weed control strategies. The study explored the potentials of two machine learning algorithms, the Artificial Neural Network (ANN) and Decision Trees (DT), for the development of yield mapping and forecasting systems from airborne hyper spectral imagery. The experimental plots were designed to simulate various crop growth scenarios, involving combinations of three different nitrogen application rates and four different weed control strategies. The study demonstrated that the potential of machine learning algorithms for the development of in-season yield mapping and forecasting system is generally high. Particular high prediction accuracies is obtained with ANNs. However, this study focused on the performance of machine learning algorithms rather than on predictions of seasonal variations in crop yield.

The work of Veenahadri, et.al. (2011)[4] Modelled soybean productivity using Decision Tree (DT) algorithms. It was proposed to develop innovative applications of data mining techniques in predicting influence of agro-climatic factors on soybean crop production in Bhopal district. The climatic factors considered in the analysis include the rainfall, the evaporation, the maximum temperature, the maximum relative humidity, and the yield as the response attributes. The decision trees formulated were then converted to classification rules using If-Then-Else. The salient conclusions of the study were (1) the decision tree analysis indicated that the productivity of soybean crop was mostly influenced by relative humidity, temperature, and rainfall; (2) the decision trees are fast to execute and much to be desired as representations of knowledge interpretations; and (3) the rules formed from the decision tree are helpful in predicting the conditions responsible for the high or low soybean crop productivity under given climatic parameters.

In the present study an attempt has been made to study the influence of climatic parameters on soybean productivity using decision tree induction technique. The findings of Decision tree were framed into different rules for better understanding by the end users. The study findings will help the researchers, the policy makers and the farmers in forecasting the crop yield in advance for market dynamics. There are three types of decision trees based on the predicted outcomes. If the predicted outcome is the class to which the data relates it is known as a classification tree. In cases where the predicted outcome is a real number for a numeric prediction it is referred to as a regression tree. Lastly if both a class and a real number are the predicted outcomes, then the method is a Classification and Regression Tree (Carter) analysis (Witten & Frank, 2005) [5].

Decision trees have certain advantages over other data mining techniques. The decision trees are simple to understand and interpret. They are capable of handling both nominal and numeric data, and are easily explainable through Boolean logic in typical white box model fashion. They allow validation of the model through statistics and are capable of good performance on large data sets (Luger, 2005) [6].

The MODIS (Moderate Resolution Imaging Spectroradiometer) the sensor, on the board of the Orbiting platforms of the international program EOS (Earth Observing System), led by NASA (National Aeronautics and Space Administration), has been generated processed data for global studies of vegetation. MODIS data have a moderate spatial resolution of 250 m on the area that a pixel covers on surface. It has high temporal repetitiveness (two days for the satellite Terra to cover the same area of land surface), the good 12-bit radiometric quality corresponding to 4096 levels of gray to represent the image and high geometric accuracy of geographic location, with the correction for attenuation of the atmospheric effects. Further, it has a free distribution. These characteristics that provide immense potential for use in monitoring sugarcane crop, in the large areas, by classifying time series images (Xavier et al., 2006) [7].

The study [8] evaluated the feasibility to estimate the yield at municipality level in São Paulo State, Brazil, using the 10 day periods of SPOT Vegetation images and the meteorological data. Twenty municipalities, seven cropping seasons were selected between 1999 and 2006. The plant development cycle or growing was divided into four phases, according to the sugarcane physiology say, to obtaining spectral and meteorological attributes for each phase. The important attributes were selected and average yield was classified according to the decision tree. The Values obtained from the NDVI time profile from December to January next year enabled to classify yields into three classes: the below average, the average and the above average. The results were more effective for 'average' and 'above average' classes, with 86.5 and 66.7% accuracy respectively.

Monitoring sugarcane planted areas using SPOT Vegetation images allowed previous analysis and predictions on the average municipal yield trend. The SPOT Vegetation images were available from 1999, for this reason, the period analyzed was defined between August 1999 and October 2006[8].

This work [9] investigates the process of yield prediction in cotton crop production using the soft computing technique of fuzzy cognitive maps. Fuzzy cognitive map (FCM) is a fusion of fuzzy logic and cognitive map theories is used for modeling and representing experts' knowledge. The investigated methodology was evaluated for 360 cases measured during the time of six subsequent years (2001–2006) in a 5 ha experimental cotton field, in predicting yield class between two possible categories ("low" and "high"). The FCM model was constructed by experts and a description on its construction process is given at next section. The measurement data were used to be categorized by the FCM tool into two yield production categories. These data were result of six years of measurements (2001–2006) at the same cotton field in Central Greece. For comparison purposes with FCM tool, the most used machine learning techniques are applied in a large number of scientific fields are used. These algorithms include decision trees. The main advantage of this approach is its simple structure and the flexibility, representing knowledge visually and more descriptively. Hence, it might be convenient tool in predicting cotton yield and improving crop management [9].

Decision trees are commonly used in the operations research, specifically you can see in decision analysis to identify a strategy most likely to reach goal. The practice decisions have to be taken online with no recall under the incomplete knowledge; the decision tree should be paralleled by the probability model as a best choice model or online selection model algorithm. Another use of decision trees for calculating conditional probabilities [11].

3. PROPOSED METHODOLOGY

In our proposed methodology the data is collected from Agriculture University, provincial farmer, and BIDAR weather data collected from weather department of Gulbarga that data is in the form of raw so that data is preprocessed and Decision tree classifier is used foe yield prediction

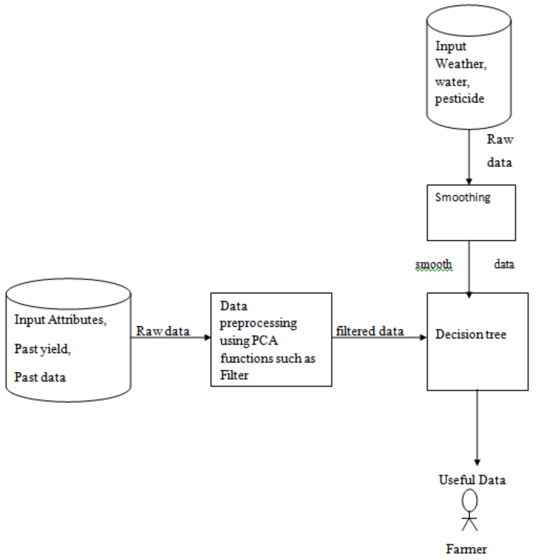


Figure 1: System Architecture

The input variables that we are going to use are Temperature, Solar radiation, Humidity, Rainfall, Soil type, Seed variety, Seed rate, Fertilizer type, Fertilizer amount, weed management, pest management, Length of the land preparation, Method used in land preparation, cropping pattern, plant spacing, plant depth, Labor utilization, Weather disturbance etc. One of the main goals of crop prediction models is to estimate agricultural production as a function of weather and soil conditions as well as crop management. Dynamic crop production model systems, as decision supporting tools, have extensively been utilized by agricultural scientists to evaluate possible agricultural consequences from inter-annual climate variability and/or climate change. Additionally, this used three sources of input variables, namely climate-related, agronomic-related, and weather disturbance as input parameters. Climate-related data and data on weather disturbance have been collected from Weather department BIDAR, while agronomic-related data were collected from municipal agriculturists.

Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. It prepares raw data for further processing. A good data preprocessing helps to create better model and will consume less time. Here PCA function filtering, smoothing is used.

Data goes through a series of steps in the pre-processing component:

Filtering: Both weather and Yield datasets are quite large and are with noise. For instance in some year there might have been slight drizzle in the month of November second week making rainfall and humidity high. However that is just a very high frequency data, meaning that such instances are rare and might repeat after several years. Such data needs to be filtered out using PCA based filtering technique such that the resulting data has no redundancy and no high frequency noise.

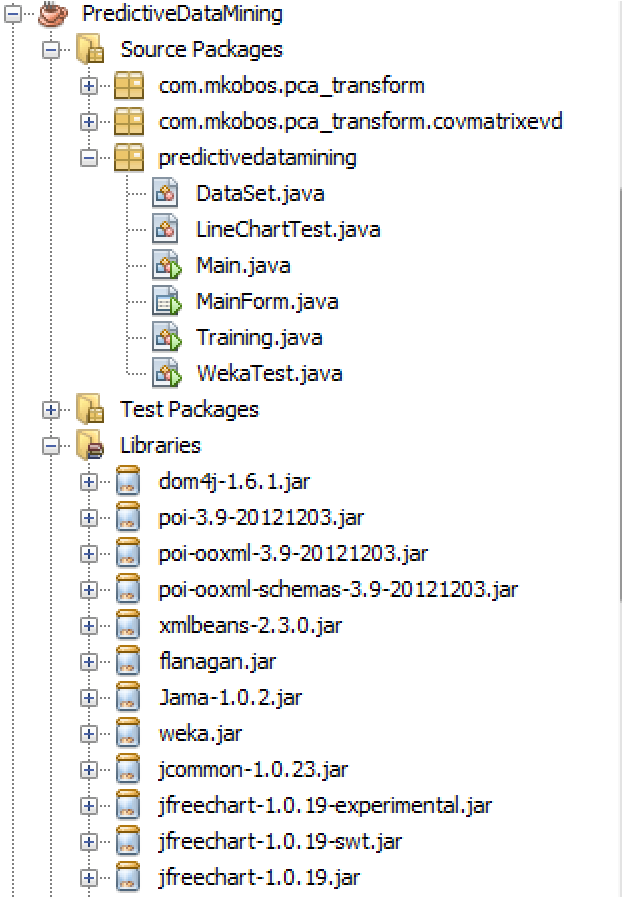
Smoothing: Smoothing deals with filtering the input observation. When yield is predicted, a week's weather, the days of the year and current level of pesticide is entered. This might suffer from abrupt variation. Hence we use median filtering to smooth the input data.

The proposed framework utilizes one of the data mining techniques called decision tree. The decision tree algorithm enables to predict the yield and enhance productivity. Decision trees are fast to execute and much to be desired as representations of knowledge interpretations and it is easy to interpret and explain. Rules formed from the decision tree are helpful in predicting the conditions responsible for the high or low sugarcane crop productivity under given climatic parameters. The main idea behind use of only decision tree based classification was bolstered many factors. Since we are working with categorical and numerical values with a various parameters associated with each sample of data. So, decision tree stands as the best candidates to classify such kind of data as decision tree themselves are nonparametric and can be generated with ease. The benefit of decision trees is that they are a non-linear method and have the ability to handle different types of data. An added benefit of classification and regression trees is their ability to handle missing data within predictor variables. Each parent node suggests a dependent variable or the cause and the child nodes become effect nodes. Thus an effect can be easily calculated by travelling in the tree through all the cause nodes.

4. EXPERIMENTAL RESULT AND DISCUSSION

4.1 Components

The project is developed in Java in NetBeans IDE. The project uses several components and external libraries. Firstly we would present the dependencies and then elaborate the coding essentials



```
PredictiveDataMining
    Source Packages
        com.mkobos.pca_transform
        com.mkobos.pca_transform.covmatrixevd
        predictivedatamining
            DataSet.java
            LineChartTest.java
            Main.java
            MainForm.java
            Training.java
            WekaTest.java
    Test Packages
    Libraries
        dom4j-1.6.1.jar
        poi-3.9-20121203.jar
        poi-ooxml-3.9-20121203.jar
        poi-ooxml-schemas-3.9-20121203.jar
        xmlbeans-2.3.0.jar
        flanagan.jar
        Jama-1.0.2.jar
        weka.jar
        jcommon-1.0.23.jar
        jfreechart-1.0.19-experimental.jar
        jfreechart-1.0.19-swt.jar
        jfreechart-1.0.19.jar
```

Figure 4.1 Shows the project architecture in Netbeans

4.2 UI Design

The GUI is built in Java swings using the swing components being offered by NetBeans

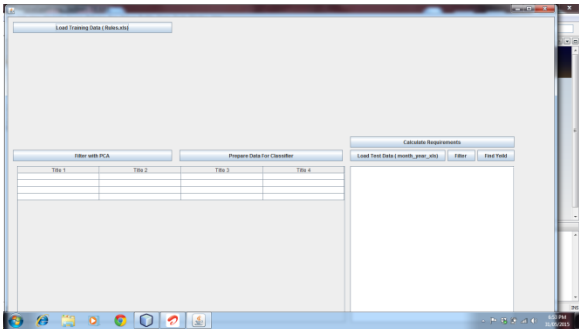
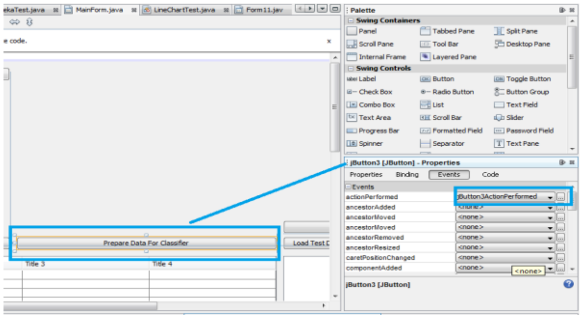

Load Training Data (Rules.xls)
Calculate Requirements
Filter with PCA
Prepare Data For Classifier
Load Test Data (month_year_xls)
Filter
Find Yeild
Title 1
Title 2
Title 3
Title 4

Figure 4.2 User Interface Design

Right side top most panel is the tool bar panel from where visual tools like EditBoxes and Buttons can be dragged and dropped on the form element. When a component is brought over the form, it's associated object is created by Netbeans. Programmer can access the properties and methods of the component using that object.

Netbeans provides easy way to add events to the components



Events:

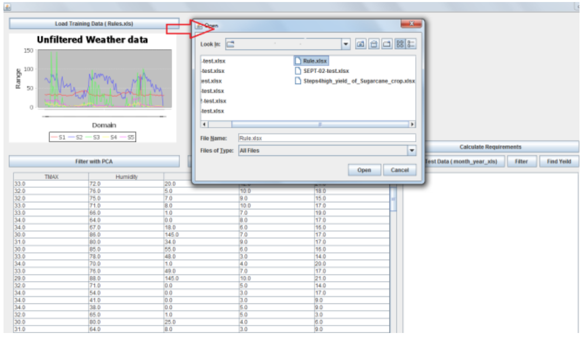
actionPerformed	jButton3ActionPerformed
ancestorAdded	<none>
ancestorMoved	<none>
ancestorMoved	<none>
ancestorRemoved	<none>
ancestorResized	<none>
caretPositionChanged	<none>
componentAdded	<none>

Figure 4.3: Adding Event Listener

4.3 shows how a component can be selected and then from the events tab in property editor and event can be added.

4.3 Load Weather Test XLS Data File

TrainingData is loaded from Load Training Data Button Click event listener. In this firstly a jFileChooser is opened to provide the user with the File open dialog to select the current weather data.



-test.xlsx
-test.xlsx
est.xlsx
-test.xlsx
-test.xlsx
-test.xlsx
Rule.xlsx
SEPT-02-test.xlsx
Steps4high_yield_of_Sugarcane_crop.xlsx

File Name: Rule.xlsx
Files of Type: All Files

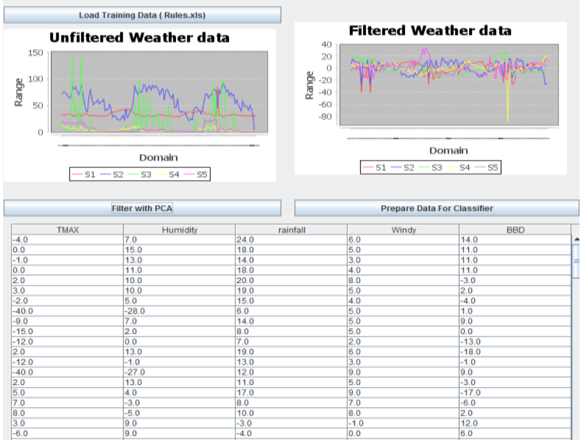
TMAX	Humidity			
33.0	72.0	20.0	[illegible]	[illegible]
32.0	76.0	5.0	10.0	18.0
32.0	75.0	7.0	9.0	15.0
33.0	71.0	8.0	10.0	17.0
33.0	66.0	1.0	7.0	19.0
34.0	64.0	0.0	8.0	17.0
34.0	67.0	18.0	6.0	16.0
30.0	86.0	145.0	7.0	17.0
31.0	80.0	34.0	9.0	17.0
30.0	85.0	55.0	6.0	16.0
33.0	78.0	48.0	3.0	14.0
34.0	70.0	1.0	4.0	20.0
33.0	76.0	49.0	7.0	17.0
29.0	88.0	145.0	10.0	21.0
32.0	71.0	0.0	5.0	14.0
34.0	54.0	0.0	3.0	17.0
34.0	41.0	0.0	3.0	9.0
34.0	38.0	0.0	5.0	9.0
32.0	65.0	1.0	5.0	3.0
30.0	80.0	25.0	4.0	6.0
31.0	64.0	8.0	3.0	9.0

Figure 4.4: Loading Training Data into project

Loaded training data is in the form of raw data that has to be preprocess using PCA.

4.4 Filtering with PCA

Input data is filtered using Filter with PCA captioned button. The event handler is as given below.



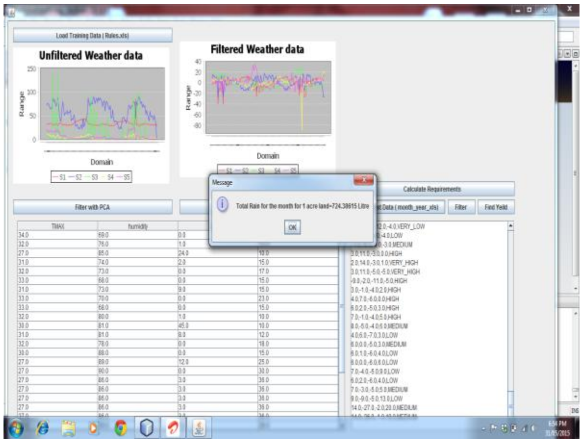
TMAX	Humidity	rainfall	Windy	BBD
-4.0	7.0	24.0	6.0	14.0
0.0	15.0	18.0	5.0	11.0
-1.0	13.0	14.0	3.0	11.0
0.0	11.0	18.0	4.0	11.0
2.0	10.0	20.0	8.0	-3.0
3.0	10.0	19.0	5.0	2.0
-2.0	5.0	15.0	4.0	-4.0
-40.0	-28.0	6.0	5.0	1.0
-9.0	7.0	14.0	5.0	9.0
-15.0	2.0	8.0	5.0	0.0
-12.0	0.0	7.0	2.0	-13.0
2.0	13.0	19.0	6.0	-18.0
-12.0	-1.0	13.0	3.0	-1.0
-40.0	-27.0	12.0	9.0	9.0
2.0	13.0	11.0	5.0	-3.0
5.0	4.0	17.0	9.0	-17.0
7.0	-3.0	8.0	7.0	-6.0
8.0	-5.0	10.0	8.0	2.0
3.0	9.0	-3.0	-1.0	12.0
-6.0	9.0	-4.0	0.0	6.0

Figure 4.5 Data Filtering with PCA

4.5 Load Test Data

Having provided with the filtered training data, the next role is to input the past year's particular month's data. The objective of this step is to build and model and find out what would be the average rainfall, temperature of the same month in this year. Based on rainfall, irrigation can be recalculated. Based on temperature and rainfall, BBD density is predicted. High density can be mapped to low requirement for pesticide and vice versa.

However one of the most important things to be noticed is that sugar production parameters are specified with respect to per acre data. So for the weather data, we calculate the total rain in terms of mm/acre.

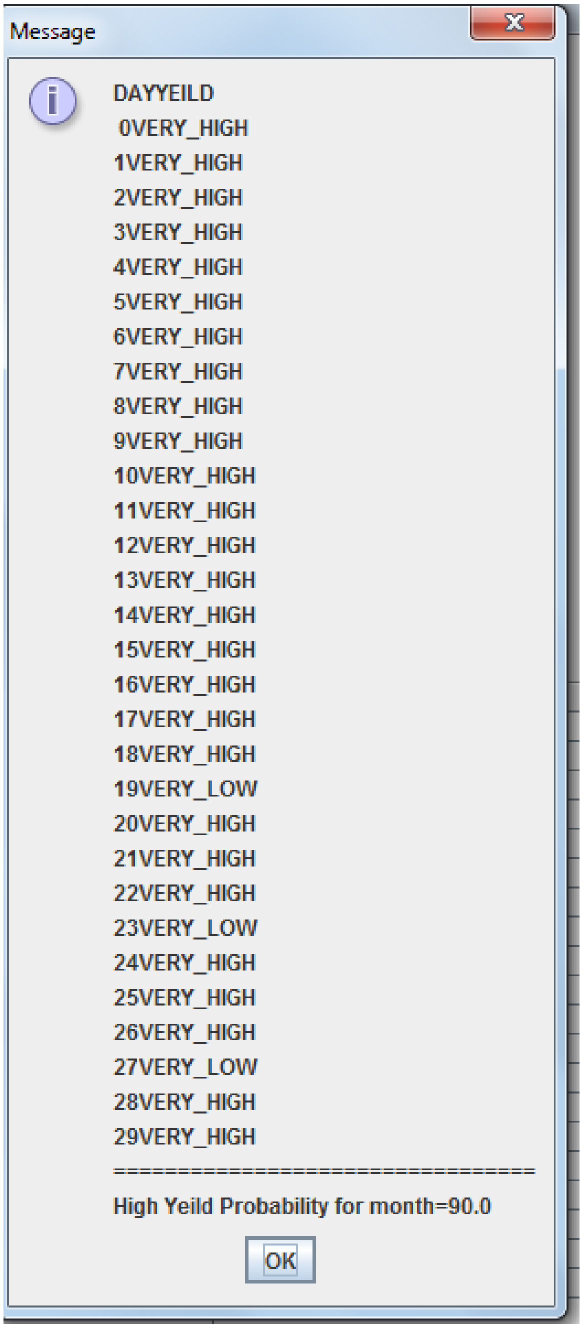


Total Rain for the month for 1 acre land=724.38615 Litre

TMAX	humidity		
34.0	69.0	0.0	[illegible]
32.0	76.0	1.0	[illegible]
27.0	85.0	24.0	10.0
31.0	74.0	2.0	15.0
32.0	73.0	0.0	17.0
33.0	68.0	0.0	15.0
31.0	73.0	9.0	15.0
33.0	70.0	0.0	23.0
33.0	68.0	0.0	15.0
32.0	80.0	1.0	10.0
30.0	81.0	45.0	10.0
31.0	81.0	8.0	12.0
32.0	78.0	0.0	18.0
30.0	88.0	0.0	15.0
27.0	89.0	12.0	25.0
27.0	90.0	0.0	30.0
27.0	86.0	3.0	36.0
27.0	86.0	3.0	36.0
27.0	86.0	3.0	38.0
27.0	86.0	3.0	36.0

```
2.0,-4.0,VERY_LOW
0,-4.0,LOW
0,-3.0,MEDIUM
3.0,11.0,-3.0,0.0,HIGH
2.0,14.0,-3.0,1.0,VERY_HIGH
3.0,11.0,-5.0,-5.0,VERY_HIGH
-9.0,-2.0,-11.0,-5.0,HIGH
3.0,-1.0,-4.0,2.0,HIGH
4.0,7.0,-6.0,0.0,HIGH
6.0,2.0,-5.0,3.0,HIGH
7.0,-1.0,-4.0,5.0,HIGH
8.0,-5.0,-4.0,6.0,MEDIUM
4.0,6.0,-7.0,3.0,LOW
6.0,0.0,-5.0,3.0,MEDIUM
6.0,1.0,-6.0,4.0,LOW
8.0,0.0,-6.0,6.0,LOW
7.0,-4.0,-5.0,9.0,LOW
6.0,2.0,-6.0,4.0,LOW
7.0,-3.0,-5.0,5.0,MEDIUM
9.0,-9.0,-5.0,13.0,LOW
14.0,-27.0,-2.0,20.0,MEDIUM
```

Figure 4.6: Screen shot of test data and filtering the data. Also average rainfall is shown as dialog box



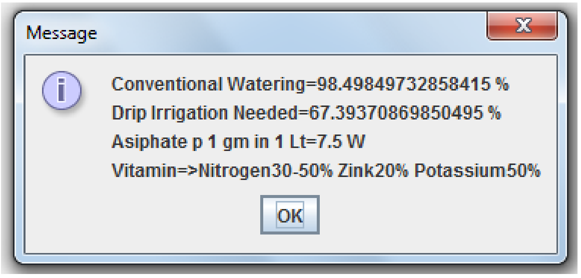
```
DAYYEILD
 0VERY_HIGH
1VERY_HIGH
2VERY_HIGH
3VERY_HIGH
4VERY_HIGH
5VERY_HIGH
6VERY_HIGH
7VERY_HIGH
8VERY_HIGH
9VERY_HIGH
10VERY_HIGH
11VERY_HIGH
12VERY_HIGH
13VERY_HIGH
14VERY_HIGH
15VERY_HIGH
16VERY_HIGH
17VERY_HIGH
18VERY_HIGH
19VERY_LOW
20VERY_HIGH
21VERY_HIGH
22VERY_HIGH
23VERY_LOW
24VERY_HIGH
25VERY_HIGH
26VERY_HIGH
27VERY_LOW
28VERY_HIGH
29VERY_HIGH
================================
High Yeild Probability for month=90.0
```

Figure 4.7: High Yeild Probability Estimated by Classifier

Based on weather of each day of the month the yield calculated as above.

4.6 Sugar Crop Yield Prediction

In this step the generalized rule for good crop prediction is simulated through the yield probability obtained through classifier. Pesticide requirement is reduced based on pest density being predicted. Irrigation requirement is adjusted based on average montly rainfall in mm/acre. Result is shown in Figure 4.8



```
Conventional Watering=98.49849732858415 %
Drip Irrigation Needed=67.39370869850495 %
Asiphate p 1 gm in 1 Lt=7.5 W
Vitamin=>Nitrogen30-50% Zink20% Potassium50%
```

Figure 4.8: Result of Yield Prediction

5. CONCLUSION

Agriculture and allied activities constitutes the single largest component of India’s gross domestic product, contributing nearly 25% of the total. Crop yield prediction has been a topic of interest for producers, consultants, and agricultural related organizations. In this work we have developed a Timely and accurate crop yield forecasts which are essential for crop production, marketing, storage, and transportation decisions as well as managing the risk associated with these activities. We have developed a framework that can (1) pre-process and fuse input raw data from multiple sources, (2) provide an accurate prediction of crop yield, (3) identify significant variables that affect crop yield. This Framework can be used to provide guidance to the farmer to detect the fault at the earlier stage to reduce losses. By using Decision tree, an efficient, globally competitive and vibrant classifier sugarcane agricultural yield can be predicted.

REFERENCES

[1] Zhu, Wenjun, et al., “Improving Crop Yields Forecasting Using Weather Data: A Comprehensive Approach Combining Principal Component Analysis and Credibility Model” 2012.

[2] Othman, MohdFauzi bin, Yau, Thomas Moh Shan. “Comparison of Different Classification Techniques Using WEKA for Breast Cancer” , 2007.

[3] Uno, Y., et.al., “Artificial Neural Networks to Predict Corn Yield from Compact Airborne Spectographic Imager Data” 2005.

[4] Veenadhari, S., et al., “Soybean Productivity Modelling Using Decision Tree Algorithms” 2011.

[5] Witten, I. H., Franke, E., & Hall, M. A. Data Mining: “Practical Machine Learning Tools and

Techniques". San Francisco: Morgan Kaufman (2005).

[6] Luger, G. F. "Artificial Intelligence (5th ed)". London: Addison Wesley (2005).

[7] Raorane, A.A., Kulkarni, R.V., "Data Mining: An Effective Tool for Yield Estimation in the Agricultural Sector" 2012.

[8] Marinkovic, et al., "Data Mining Approach for Predictive Modelling of Agriculture Yield Data" 2010.

[9] Han J and M Kamber "Data Mining Concepts and Techniques" Second Edition Elsevier Publication, 2009.

[10] Beard, D. A., Gray, D. M., & Carmody, P. Farmers' "management of seasonal variability and climate change in WA". Crop Updates 2010.

[11] Ekasingh, B. S., Ngamsomsuke, K., Letcher, R. A., & Spate, J. M. "A Data Mining approach to simulating land use decisions: Modeling farmer's crop choice from farm level data for integrated water resource management". Paper presented at the Proceedings of the International Conference on Simulation and Modeling (2005).