Cotton Crop Yield Prediction using Data Mining Technique

Amiksha Ashok Patel[1]

Research Scholar, Faculty of Computer Science and
Applications, Charotar University of Science and
Technology, Changa, Gujarat, India

Dr. Dhaval Kathiriya[2]

Director Information Technology
Anand Agricultural University
Anand, Gujarat, India

Abstract—Cotton is a very important crop, as India leads it in terms of production in the world; and also that a vast number of manpower is engaged in farming as well as post-harvest processing and management of different derivatives of it. Weather is crucial for the productivity of the crop. The challenges of climate change; availability of limited land and water for farming; lake of knowledge for good cultivation practices and judicious use of agricultural inputs with farmers are critical hindrances for improving productivity. This requires thorough research on land preparation and use, how to improve fertility of soil, good agronomic practices in lieu of variable climatic conditions, etc. All the talukas of the three districts of North Gujarat where cotton is cultivated have been selected purposively for this study. The effect of soil type, soil pH, soil organic carbon, phosphorous, potassium, precipitation and temperature were selected as independent factors. The yield of cotton crop has positive correlation with the selected parameters. The data sets were applied for analytical process to WEKA. The difference between average of predicted and actual yields of all talukas for high rainfall year 2013 was only 1.55 per cent. The difference between actual and predicted yield for the low temperature year (2015) in different talukas of all talukas was only 0.44 per cent.

Keywords—Data mining; cotton crop yield prediction; agriculture; data processing; data visualization

I. INTRODUCTION

One of the great challenges of agricultural development is to guarantee food security. Simultaneously, securing fiber requirement is also one of the necessities of human being. Cotton is an important crop for world's poor. Cotton is grown commercially in more than 80 different countries, mostly in the longitudinal band between 37°N and 32°S.

Cotton is especially adapted to semi-arid and arid environments, where it is either grown as rain-fed or with irrigation. About 53 per cent of the world's cotton growth areas benefit from full or supplementary irrigation. Cotton has certain resilience to high temperatures and drought due to its vertical tap root. The crop is, however, sensitive to water availability, particularly at the stage of flowering and boll formation. Rising temperatures favor development of the cotton plant, unless day temperatures exceed 32° C.

Climate change will affect the cotton crop in numerous ways in different areas. Temperatures are expected to increase all over India. Rainfall intensity during monsoons may become a prevalent problem. Higher temperatures in already hot areas may hinder cotton development and fruit formation. Rain-fed cotton production may suffer from higher climate variability leading to periods of drought or flooding. With respect to the production level, cotton has limited capacity to respond to heat stress, through 'compensatory growth'. Its vertical tap root also provides resilience against spells of drought, but also makes it vulnerable to water-logging.

Cotton is a natural plant fiber which grows around the seed of the cotton plant. Fibers are used in the textile industry. First, the cotton fiber is obtained from the cotton plant and then spun into yarn. Further, the cotton yarn is woven or knitted into fabric. The use of cotton has a long tradition in the clothing industry due to its desirable characteristics. The value of world cotton production in 2017-18, was around US$50 billion. Cotton is a driver of economic development and is of critical importance to the economies of developing and least developed countries. Cotton connects people to markets and provides economic opportunities on the frontiers of the world economy.

Of course, there are many factors that affect prediction of cotton. However, crop yield prediction is extremely challenging due to numerous complex factors which affect cotton crop at different growth stages (a short list is as per Annexure 1). It is very difficult to have a site specific measurement of each of these factors and to evaluate their combined effect on cotton production. As such, the current study was planned to look at the effect of the key seven parameters and their contribution to production.

II. BACKGROUND

Cotton is one of the major cash crops grown in India. The productivity of this crop can be improved dramatically if correct agro-technologies are adhered. The yield gap of research farms or potential of a variety and that of average harvest at farmers' fields is very high.

Cotton is a crucial component of the Indian economy as her textile industry is predominantly cotton based. India is one of the biggest producers and also exporter of cotton fabric. Cotton cultivation is a well-established practice in India. Gujarat, Maharashtra, Telangana, Andhra Pradesh and Karnataka are the major cotton producer states in India. Indian textile industry contributes to around 5 percent to the nation's gross domestic product (GDP), 14 percent to industrial production and 11 percent to total export earnings. After agriculture, textile industry is the second largest employer of over 510 lakh people directly and 680 lakh people indirectly.

The challenges of climate change; availability of limited land and water for farming; lake of knowledge for good cultivation practices and judicious use of agricultural inputs with farmers are critical hindrances for improving productivity. This requires thorough research on land preparation and use, how to improve fertility of soil, good agronomic practices in lieu of variable climatic conditions, etc. The analytical issues, till yesterday, were been handled by applying different statistical tools. However, using tools of data mining and other diagnostic approaches are becoming more useful in making decisions regarding production practices and also prediction of yields [4].

The comprehensive approach, which comprises of various technologies and methods, for example, statistics, Data Mining, Visual Data Mining (VDM), information handling, Data Warehousing (DW), Online Analytical Processing (OLAP) and different frameworks, was considered to be a useful approach. It can also help in better crop predictions based on historical data.

A. Role of DM in Agriculture

An accurate estimate of crop production and risk helps the country in planning supply chain decision like production scheduling. Business such as seeds, fertilizers, agrochemicals and agricultural machinery industries plan production and marketing activities based on crop production estimates [9], [14]. These are helpful for the farmers and the government in decision making namely:

1) It helps farmers in providing the crop yield record with a forecast, so as to reduce the risk of crop management.

2) It helps the government in making policies for crop insurance and supply chain operations.

In large data sets, data mining is the computational process for discovering new patterns. Data mining provides major advantage in agriculture for disease detection, problem prediction and for optimizing inputs like pesticides, irrigation, fertilizers, etc. With advancement in technological applications in agriculture, a lot of information is made available. Hence, data mining techniques in agriculture is used for pattern reorganization and crop health detection. Reliable and timely estimates of crop production are important for taking various decisions for marketing, pricing, storage, distribution and import-export. The crop yields primarily depend on diseases, pests, climatic conditions, time of harvest, etc. As such, these predictions are very useful for agriculture domains. Data mining techniques are used not only for quantifying requirements of inputs and timely executing various agricultural operations but also for pre-harvest forecasting for crop yields. Data mining is also called as knowledge discovery database (KDD).

Data mining tasks can be classified into two categories:

- Descriptive data mining.

- Predictive data mining.

Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict the direct values based on patterns determined from known results. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. As far as data mining technique is concern, in most of the cases, predictive data mining approach is used. Predictive data mining technique is used to predict future crop, weather forecasting, pesticides and fertilizers to be used, revenue to be generated and so on. [4].

III. SURVEY OF LITERATURE

Many research studies have focused on the importance of data mining to be used as a tool for analysis of big data of agriculture for meaningful conclusions [1], [2].

Ashok Kumar and N. Kannathasan dealt with various data mining procedures that can be utilized in farming. Their examination suggested that a correlation of various data mining strategies could deliver a productive calculation for soil grouping [3].

M. C. Geetha researched Data Mining Techniques in Agriculture and argued that information mining in agriculture is quite useful to forecast the productivity and production of crops. She talked about various information-digging applications for tackling distinctive farming related issues [8].

Ruß G. utilized information got from three fields of Germany. Researcher utilized regression methods on farming yield information and reasoned that help vector regression can fill in as a superior reference demonstrates for yield expectation. Likewise, the model parameters which have been built upon one informational index can be utilized for different techniques on selected agriculture data [12].

S. Veenadhari, B. Misra, and C. Singh endeavored to aggregate the examination discoveries of various scientists who took a shot at harvest profitability information. The machine learning approach of coordinating software engineering with farming will help in gauge farm yields adequately [13].

A study by Ekasingh B. S. Ngamsomsuke K. Letcher R. A. & Spate J.M. has analyzed how data mining may be applied for the purposes of crop production [7]. Majority of earlier researchers including Dunstan D. (2009) who used data mining as a supportive tool with statistical analysis, have focused on crop yield management and its' quality evaluation [6].

Raorane A.A. and Kulkarni R.V. (2012) talked about different data mining strategies, as a result of use of data mining methods; an effective production system can be derived that can take care of complex farming issues [11].

Ramesh Vamanan and K. Ramar (2011) presumed that the Data Mining method (Naïve Bayes Classifier) when connected to a farming soil profile may enhance the confirmation of legitimate soil profile, substantial examples and profile classification are contrasted with standard statistical investigation strategies [10].

IV. MATERIALS AND METHODOLOGY

A. Data Acquisition

The focal point of this research was to look at the effect of temperature, rainfall and soil parameters (namely soil type, Soil

pH, Carbon, phosphorus, and potassium) on the cotton yields for different farming locales in the study area.

For the present research, 27 talukas of the three districts of Gujarat State were taken. The soil parameter data were retrieved from the Soil Health Card data of Government of Gujarat routed through Anand Agricultural University, District Anand, Gujarat State. The production data were collected from the Department of Agriculture, Government of Gujarat, by approaching them personally and also by random access to farmers for verification of the information. The data of rainfall and temperature were collected from the Sardarkrushinagar Dantiwada Agricultural University, SKNagar, District Banaskantha, Gujarat State.

In the present study, the rainfall, temperature and five soil parameters were considered as independent factors and their effect on cotton productivity was analyzed. As a consequence, instances of cotton yield were examined against these datasets. The average rainfall, temperature and selected soil parameters of the ten years for each of the 21 talukas in the three districts were obtained from a secondary database. These data sets of ten years (from 2006 to 2015) were analyzed. The dataset was structured and combined to be managed in excel spreadsheet as talukas, average rainfall, maximum and minimum temperature, soil parameters, and cotton yield.

B. Analytical Procedure

To conduct different research experiments and calculations WEKA, Revolution R and SPSS have been used. Initially the data was collected and maintained in Microsoft Excel. Further, data transformation and other calculations were done in software like SPSS and Microsoft Excel. The total datasets was classified and kept in different folders; and further that ordered into the rainfall, temperature, and soil variation (pH, Carbon, Phosphorous, and Potassium) and yield. Different algorithms were tested in WEKA software to check and decide the most suitable among all algorithms and evaluate output with other datasets. These data were used in WEKA and R software for dept. analysis and experiments.

C. The Activity Experiments

The following examinations were done iteratively for calculating the impact of the climatic factors on cotton yield:

1) Rainfall and soil type relationship in view of the cotton crop.

2) Effect of nutritional variations of soil on cotton yield.

Further the data were restructured from the perspective of having sufficient depth and substance to be believable before initiating the processing. The lateral development happened because attributes were included from other data collection. Both the reduction and extension add up to the pre-preparing of the dataset and incorporation and delineation of outliers through Exploratory Visual Data Mining (EVDM) of the mapped information.

The compiled data were administered to techniques of combination, categorization, accumulation and statistical projection to locate elite method and related best fit design. In the auxiliary investigation through OLAP appraised for

association in the data sets. Examination of these outcomes were taken up and concluded accordingly.

At last, the two assortments of outcomes were combined and arranged diagrammatically to decide a general example that gave both a minute and apparent context of data. The datasets were then examined through a progression of tests which included cross classifications, correlation, sequencing, time series analysis and regression. The outcomes of these tests were then investigated and delineated in detail.

As a result of this study, recommendations for getting good cotton crop yields by managing the agronomical and other practices in lieu of the difficult weather and soil conditions were made for north Gujarat farming area.

D. Data Analysis

The information utilized in this investigation had diverse attributions and was comprised of five separate however related substances. The majority of the datasets were in connection with North Gujarat, India.

In the current research, different data were used with varied reasoning. As referred in Section 3.3, the different datasets comprised of temperature, soil type, rainfall and soil parameters were used. Different data sets were obtained either from secondary source of data namely Department of Agriculture, Government of Gujarat; Anand Agricultural University, Anand and Sardarkrushinagar Dantiwada Agricultural University, Sardarkrushinagar, Gujarat, India. The counter verification of these data and obtaining few other data were done by personal deliberation with scientists and farmers. All the datasets were fitted explicitly for the cotton crop production areas of the selected locale of the study.

The data extraction and pertinence involved using a number of software tools that represented an admixture of retrieving, pre-processing, scrutiny, data mining and revelation of temperature, soil type, rainfall, and soil parameters' data. The process was alienated into five fundamental stages namely data collection, pre-processing, handling, data examination and processing.

Different crop have their critical and optimum climatic requirements, for example, the increasing temperature may affect agriculture by reducing the productivity on different crops in different seasons. The analysis of climatology at the region level is most useful for the solution of practical agricultural problems. Temperature has a complex relationship to the development of plant at different growth stage.

E. Basic Model Process Flow

The entire information extraction and analysis process is demonstrated graphically in Fig. 1 below. It framed the initial segment of research in anticipation of the data collection, data preparation, data modeling and data storage in the protected investigations.
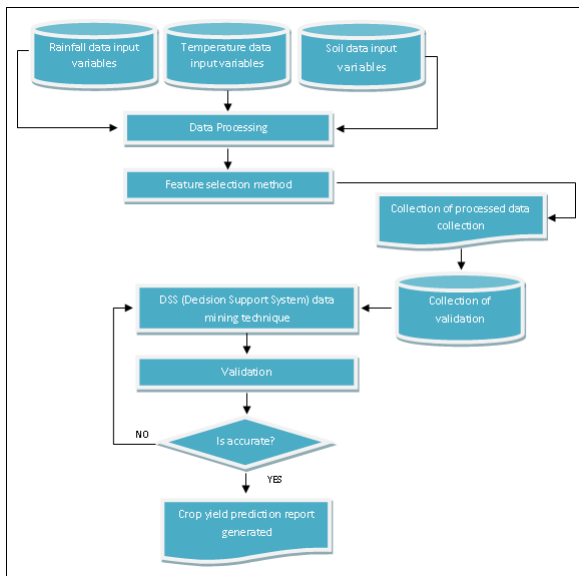
Fig. 1. Basic Model of Process Flow.

F. Analytical Procedure

To conduct different research experiments and calculations WEKA, Revolution R and SPSS have been used. Initially the data was collected and maintained in Microsoft Excel. Further, data transformation and other calculations were done in software like SPSS and Microsoft Excel. The total datasets was classified and kept in different folders; and further that ordered into the rainfall, temperature, and soil variation (pH, Carbon, Phosphorous, and Potassium) and yield. Different algorithms were tested in WEKA software to check and decide the most suitable among all algorithms and evaluate output with other datasets. These data were used in WEKA and R software for dept. analysis and experiments.

- The Activity Experiments

The following examinations were done iteratively for calculating the impact of the climatic factors on cotton yield:

1) Rainfall and soil type relationship in view of the cotton crop.

2) Impact of rainfall on cotton yield.

3) Impact of temperature on cotton yield.

4) Combined effect of rainfall and temperature on cotton yield.

5) Effect of nutritional variations of soil on cotton yield.

As a result of this study, recommendations for getting good cotton crop yields by managing the agronomical and other practices in lieu of the difficult weather and soil conditions were made for north Gujarat farming area.

V. RESULTS AND DISCUSSION

A. Taluka Wise Yield Prediction for Low and High Rainfall Years

The data sets of low and high rainfall years (2009 and 2013 respectively) were applied for analytical process to WEKA. The datasets used and calculations made in WEKA are given in Table I.

TABLE I. WEKA PREDICTION FOR YEAR 2009 AND 2013

Taluka	LR 2009		HR 2013	
	Predicted Yield	Actual Yield	Predicted Yield	Actual Yield
AMIRGHADH	649.52	488	627.46	598
BECHARAJI	481.23	584	620.87	568
BHABHAR	500.49	526	564.92	523
DANTA	401.41	456	392.28	395
DANTIWADA	440.27	556	610.46	630
DEESA	657.96	621	615.55	630
DHANERA	599.29	578	647.30	615
DIYODAR	594.51	543	645.93	599
HIMATNAGAR	386.82	599	605.44	617
IDAR	737.05	636	557.37	601
KADI	427.99	500	613.31	594
KANKARAJ	514.21	513	619.88	597
KHEDBHRAMHA	373.41	408	435.85	451
KHERALU	211.74	300	370.83	350
MAHESANA	579.97	548	645.34	602
PALANPUR	604.30	623	627.25	645
POSHINA	475.79	348	375.57	303
PRANTIJ	508.20	615	573.36	536
SATALASANA	289.91	296	369.98	366
TALOD	535.21	642	537.68	566
UNJHA	365.22	311	387.25	369
VADALI	494.03	598	633.98	565
VADGAAM	384.59	508	352.14	439
VADNAGAR	601.71	578	576.96	754
VIJAPUR	543.93	721	721.76	685
VIJAYNAGAR	448.53	568	539.17	495
VISNAGAR	597.04	654	632.21	684

The difference between actual and predicted yield for the low rainfall year (2009) in different talukas varied greatly. It is obvious that the maximum difference in predicted and actual yield was observed for Poshina taluka as 36.72 per cent, followed by Amirghadh taluka (33.10 per cent) and Unjha taluka (17.43 per cent). The average differenct between maximum value of predicted yield and actual yield was 36.72 per cent. The average difference between minimum value of predicted yield and actual yield was -35.42 per cent. However, the difference between average of predicted and actual yields of all talukas was only -5.39 per cent.

The difference between actual and predicted yield for the high rainfall year (2013) in different talukas varied greatly. It is obvious that the maximum difference in predicted and actual yield was observed for Poshina taluka as 23.95 per cent, followed by Vadali taluka (12.21 per cent) and Becharaji taluka (9.31 per cent). The average difference between maximum value of predicted yield and actual yield was 23.95 per cent. The average difference between minimum value of predicted yield and actual yield was -23.48 per cent. However, the difference between average of predicted and actual yields of all talukas was only 1.55 per cent.

B. Taluka Wise Yield Prediction for Low and High Temperature Years

The difference between actual and predicted yield for the low temperature year (2015) in different talukas varied greatly.

It is obvious that the maximum difference in predicted and actual yield was observed for Becharaji taluka as 18.03 per cent, followed by Vadnagar taluka (17.01 per cent) and Satlasana taluka (16.27 per cent). The average difference between maximum value of predicted yield and actual yield was 18.03 per cent. The average difference between minimum value of predicted yield and actual yield was -29.24 per cent. However, the difference between average of predicted and actual yields of all talukas was only 0.44 per cent.

The difference between actual and predicted yield for the high temperature year (2010) [5] in different talukas varied greatly. It is obvious that the maximum difference in predicted and actual yield was observed for Talod taluka as 34.88 per cent, followed by Prantij taluka (26.02 per cent) and Vijaynagar taluka (25.05 per cent). The average difference between maximum value of predicted yield and actual yield was 34.88 per cent. The average difference between minimum value of predicted yield and actual yield was -38.34 per cent. However, the difference between average of predicted and actual yields of all talukas was only -1.87 per cent.

The data sets of high and low temperature years (2010 and 2015 respectively) were applied for analytical process to WEKA. The datasets used and calculations made in WEKA are given in Table II.

TABLE II.   WEKA PREDICTION FOR YEAR 2010 AND 2015

Taluka	HT 2010		LT 2015	
	Predicted Yield	Actual Yield	Predicted Yield	Actual Yield
AMIRGHADH	502.43	564	629.09	752
BECHARAJI	558.29	528	634.99	538
BHABHAR	490.08	547	544.97	518
DANTA	388.33	383	394.86	558
DANTIWADA	578.35	485	665.49	621
DEESA	631.26	606	613.14	589
DHANERA	555.71	560	640.64	604
DIYODAR	534.43	630	595.15	592
HIMATNAGAR	659.87	575	638.67	623
IDAR	404.08	513	608.32	627
KADI	498.56	537	560.68	540
KANKARAJ	394.64	640	590.18	640
KHEDBHRAMHA	464.23	451	440.14	434
KHERALU	278.17	320	361.47	330
MAHESANA	539.36	574	616.16	603
PALANPUR	520.94	630	640.54	733
POSHINA	345.32	388	378.34	368
PRANTIJ	626.34	497	594.32	652
SATALASANA	303.52	315	377.89	325
TALOD	698.69	518	587.47	620
UNJHA	300.64	334	378.01	350
VADALI	594.92	526	618.91	629
VADGAAM	278.92	364	486.74	540
VADNAGAR	583.01	670	772.29	660
VIJAPUR	753.58	694	700.61	666
VIJAYNAGAR	573.98	459	542.80	589
VISNAGAR	640.13	674	702.45	673

VI. DISCUSSION

Prediction of yield of a crop is very difficult; as it is the sum of complex interrelationship of many factors. The water affects a lot on cotton crop production. Though the farmers have no control on precipitation; although if having facilities of irrigation; timely irrigation, method of irrigation, quantity of irrigation and other irrigation management issues plays very important role in cotton crop production. Even delaying irrigation by one or two days in a peak season alters the effect of insect and pest infestation on the crop. In the current research, the data mining classification function of Gaussian Processes showed strong positive correlation between the average annual rainfall and cotton crop yield for the selected 27 talukas.

If real time outputs of this information are communicated to farmers for improving their crop management, it can really contribute to sustainable as well as improved production.

VII. CONCLUSION

Data mining is the process of finding the useful outcomes from the large data sets. During this work, the time series forecasting package have been used for regression approach of Gaussian Processes for yield prediction. The Gaussian Processes algorithm using with parameters like year, crop yield, temperature, rainfall and the five soil parameters are considered within the model development.

Another important factor for cotton crop production is temperature. Temperature in air and soil; difference in day and night temperature; sudden changes in temperature; etc affects the productivity. The hot winds, at a particular growth period are harmful for and at a different growth period is useful for production. When temperatures become too hot, fertilization may be compromised, leading to fewer seeds produced per boll, smaller boll masses, and ultimately, lint yield reductions [15].Temperature affects cotton crop in a complex way on the production of the crop. However, the results of this study shows very strong and positive correlation between air temperature (minimum and maximum) on cotton crop yield.

The soil parameters, variety, farmers' management abilities and their socio-economic capabilities, etc all independently as well as their interactive effects decides the production of cotton. The yield was directly associated with improved soil water relations resulting from the cropping and tillage treatments. Application of varying levels of fertilizer in combination with bio-fertilizer positively influences to cotton yield as they improve the availability of NPK to the crop. The soil parameters including soil pH, SOC, P & K have positive correlation with cotton crop production. As the soil type in the entire operational area was identical (sandy loam type – Goradu type), its' correlation with cotton crop yields could not be assessed.

If real time outputs of this information are communicated to farmers for improving their crop management, it can really contribute to sustainable as well as improved production.

VIII. WAY FORWARD

Forecasts of production include imbedded assumptions about farmers' reactions to changes in output prices, input

prices, weather forecasts, labor availability, input availability, storage capacity, marketing opportunities, food security, and changes in technology, government policies and other innumerable factors. Forecasts of consumption are really forecasts of textile mill managers' choices in response to the welter of price information, resource constraints and government policies they face, etc.

Nevertheless, even though human behavior is highly variable and unanticipated policy shocks are common, with great advances in technology improvement in forecasts has not been. Perhaps this is because the information we are getting faster is actually degrading in quality. In other words, the statistics on which forecasts are based are becoming less accurate, thus undermining the value of getting those statistics more easily.

REFERENCES

[1] Abdullah, A., and Ansari, I. A. (2005), Discovery of cropping regions due to Global Climatic Changes using Data Mining, Paper presented at the 3rd International Symposium on Intelligent Information Technology in Agriculture, Beijing.

[2] Chien, and Chena (2008), Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry Expert Systems with Applications, 34(1, January 2008), 280-290.

[3] D, Ashok Kumar and Kannathasan, N (2011) A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining, International Journal of Computer Science Issues, Vol. (8).

[4] Developing innovative applications in agriculture using data mining-Sally Jo Cunningham and Geoffrey Holmes.

[5] Drew, J. (2010), Operating In a Change Environment – NEAR/Drought Reform.

[6] Dunstan, D. (2009), Hierarchies of sustainability in a catchment, Paper presented at the 4th International Conference on Sustainable Development and Planning, Cyprus.

[7] Ekasingh, B. S. et al (2005), A Data Mining approach to simulating land use decisions: Modelling farmer's crop choice from farm level data for integrated water resource management, Paper presented at the Proceedings of the 2005 International Conference on Simulation and Modelling.

[8] Geetha, M.C. (2018), A Survey and Analysis on Regression Data Mining Techniques in Agriculture, International Journal of Pure and Applied Mathematics, Vol 118 No. 8, 341-347.

[9] Gleaso CP. Large area yield estimation/forecasting using plant process models. Paper presentation at the winter meeting American society of agricultural engineers palmer house, Chicago, Illinois. 1982; 14–17.

[10] Ramesh Vamanan, K. Ramar (2011), Classification of agricultural land soils a data mining approach, International Journal on Computer Science and Engineering, Vol. 3(1).

[11] Raorane A.A., Kulkarni R.V. (2012), Data Mining: An effective tool for yield estimation in the agricultural sector, International Journal of Emerging Trends and Technology in Computer Science (IJETTCS), Vol (1-2).

[12] Ruß, G. et al (2009), Visualization of agriculture data using self-organizing maps, Paper presented at the Proceedings of AI-2008.

[13] S. Veenadhari, B. Misra and C. Singh (2014), "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5, doi: 10.1109/ICCCI.2014.6921718.

[14] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).

[15] William T. Pettigrew (2008), Potassium influences on yield and quality production for maize, wheat, soybean and cotton, Physiologia Plantarum, Volume 133, Issue 4.

ANNEXURE I

The following are some of the more frequent types of information that can be derived from the basic data:

A. *Air temperature*

1) Temperature probabilities;

2) Chilling hours;

3) Degree-days;

4) Hours or days above or below selected temperatures;

5) Interdiurnal variability;

6) Maximum and minimum temperature statistics;

7) Growing season statistics, that is, dates when threshold temperature values for the growth of various kinds of crops begin and end.

B. *Precipitation*

1) Probability of a specified amount during a period;

2) Number of days with specified amounts of precipitation;

3) Probabilities of thundershowers;

4) Duration and amount of snow cover;

5) Dates on which snow cover begins and ends;

6) Probability of extreme precipitation amounts.

C. *Wind*

1) Windrose;

2) Maximum wind, average wind speed;

3) Diurnal variation;

4) Hours of wind less than selected speed.

D. *Sky cover, sunshine, radiation*

1) Per cent possible sunshine;

2) Number of clear, partly cloudy, cloudy days;

3) Amounts of global and net radiation.

E. *Humidity*

1) Probability of a specified relative humidity;

2) Duration of a specified threshold of humidity.

F. *Free water evaporation*

1) Total amount;

2) Diurnal variation of evaporation;

3) Relative dryness of air;

4) Evapotranspiration.

G. *Dew*

1) Duration and amount of dew;

2) Diurnal variation of dew;

3) Association of dew with vegetative wetting;

4) Probability of dew formation based on the season.

H. *Soil temperature*

1) Mean and standard deviation at standard depth;

2) Depth of frost penetration;

3) Probability of occurrence of specified temperatures at standard depths;

4) Dates when threshold values of temperature (germination, vegetation) are reached.

I. *Weather hazards or extreme events*

1) Frost;
2) Cold wave;
3) Hail;
4) Heatwave;
5) Drought;
6) Cyclones;
7) Flood;
8) Rare sunshine;
9) Waterlogging.

J. *Agrometeorological observations*

1) Soil moisture at regular depths;
2) Plant growth observations;
3) Plant population;
4) Phenological events;
5) Leaf area index;
6) Above-ground biomass;
7) Crop canopy temperature;
8) Leaf temperature;
9) Crop root length.