

# A Survey on Crop Yield Prediction based on Agricultural Data

Dhivya B H<sup>1</sup>, Manjula R<sup>2</sup>, Siva Bharathi S<sup>3</sup>, Madhumathi R<sup>4</sup>

U.G. Student, Department of Computer Science and Engineering, Sri Ramakrishna Engineering College,  
Coimbatore, India<sup>1</sup>

U.G. Student Department of Computer Science and Engineering, Sri Ramakrishna Engineering College,  
Coimbatore, India<sup>2</sup>

U.G. Student Department of Computer Science and Engineering, Sri Ramakrishna Engineering College,  
Coimbatore, India<sup>3</sup>

Asst. Professor Department of Computer Science and Engineering, Sri Ramakrishna Engineering College,  
Coimbatore, India<sup>4</sup>

**ABSTRACT:** Agriculture is one of the major revenue producing sectors of India and a source of survival. Various seasonal, economic and biological factors influence the crop production but unpredictable changes in these factors lead to a great loss to farmers. These risks can be quantified when appropriate mathematical or statistical methodologies are applied on data related to soil, weather and past yield. With the advent of data mining, crop yield can be predicted by deriving useful insights from these agricultural data that aids farmers to decide on the crop they would like to plant for the forthcoming year leading to maximum profit. This paper presents a survey on the various algorithms used for crop yield prediction.

**KEYWORDS:** Crop yield prediction, Agriculture, Forecasting, Data Mining

## I. INTRODUCTION

Agriculture is important to human beings because it forms the basis for food security. It helps human beings grow the most ideal food crops and raise the right animals with accordance to environmental factors. India's agriculture is composed of many crops, with the foremost food staples being rice and wheat. Indian farmers also grow pulses, potatoes, sugarcane, oilseeds, and such non-food items as cotton, tea, coffee, rubber, and jute. Over 70 per cent of the rural households depend on agriculture. Agriculture contributes about 17% to the total GDP and provides employment to over 60% of the population.

Data mining is defined as a process of identifying previously unknown inferences from the huge volume of available data. It finds application in market analysis, production control, fraud detection, customer retention, E commerce etc. Data mining software analyses relationships and patterns in stored transaction data based on open ended user queries. On the basis of the nature of data being mined there are two categories of functions involved in data mining namely, Descriptive function that deals with general properties of data and Prediction function that identifies the trends based on available data. As far as agriculture is concerned predictive types that include classification, association, clustering and regression are used

Knowledge Discovery Process comprises of the following steps they are Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Presentation. Data Cleaning is used to remove the noise and inconsistent data. Multiple data sources are combined using Data Integration. In Data

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

Selection, the data relevant to analysis are retrieved from the database. In Transformation, data is transformed or consolidated into forms for the mining process by performing summary or Aggregation process. In Data mining, the data patterns are extracted by applying intelligent methods. In Pattern Evaluation patterns are evaluated. Knowledge is represented in Knowledge presentation. The following diagram gives the basic steps that are involved in knowledge discovery in data mining

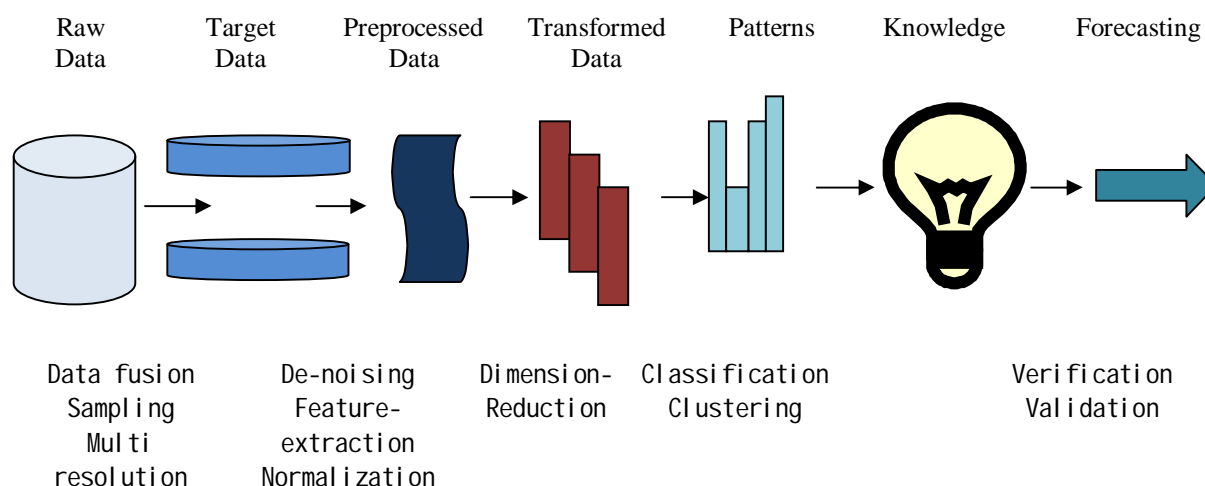


Fig 1: Steps in Knowledge discovery

## II. LITERATURE SURVEY

Raval et al discuss about the Knowledge Discovery Process and the basics of various Data Mining Techniques such as Association rules, Classification, Clustering, Prediction and Sequential Patterns [1].

Agrawal et al discuss about various Data Mining tools such as Dashboards, Text-Mining tools. They provide an overview about these tools and the various scenarios in which they can be deployed [2].

Grajales et al have proposed a web application that utilizes open dataset like historical production, land cover, local climate conditions and integrates them to provide easy access to the farmers. The proposed architecture mainly focuses on open source tools for the development of the application. The user can select location from map for which the details are available at one click [3].

Bendre et al collects data from GIS (Global Information System), GPS (Global Positioning System), VRT (Variable Rate Fertilizer) and RS (Remote sensing) and are manipulated using Map Reduce algorithm and linear regression algorithm to forecast the weather data that can be used in precision agriculture. The purpose of this study was to investigate the effective model to improve the accuracy of rainfall forecasting [4].

Hemageetha mainly focuses on using the soil parameters like pH, Nitrogen, moisture etc for crop yield prediction. Naive Bayes algorithm is used to classify the soil and a 77% accuracy is achieved. Apriori algorithm is used to associate the soil with the crops that could provide maximum yield in them. A comparison of accuracy achieved during classification using Naïve Bayes, J48 and JRIP is also presented [6].

Rub et al presents a comparative study on the regression models that could be used for predicting yield. The algorithms discussed are Multilayer perception Model (MLP), Regtree (Regression tree), RBF (Radial Basis Function Network and SVM (Support Vector Machine). They have concluded that SVM serves as a better model as far as yield prediction is concerned [7].

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

Sujatha et al describes the purpose of various classification techniques that could be used for crop yield prediction. A few of the data mining methods, such as the Naïve Bayes, J48, random forests, support vector machines, artificial neural networks were presented. A system using climate data and crop parameters used to predict crop growth has been proposed [8].

Kushwaha et al predicts the suitability of a crop for a particular climatic condition and the possibilities of improving the crops quality by using weather and disease related data sets. They have proposed an analysis, classification and prediction algorithm that helps in building a decision support system for precision farming. It is based on the Hadoop file system [9].

Fathima et al uses data mining techniques on real time data that help in knowledge discovery. They use k means clustering algorithm to cluster the farmers based on the crop type and irrigation parameters. Apriori algorithm is used to determine which two crops are selected as a frequent item set. They generally focus on the policies that government could frame by the cropping practices of farmers [10].

Veenadhari et al described the purpose of data mining methods in the area of agriculture. A few of the data mining methods, such as the k-means, ID3 algorithms, the k nearest neighbor, support vector machines, artificial neural networks were presented [11].

Sellam et al explain the various environmental parameters like the Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI) that influences the yield of crop and the relationship among these parameters is established. Using Regression Analysis (RA), Linear Regression (LR) the various environmental factors and their infliction on crop yield is analyzed [12].

Raorane et al discusses the various data mining techniques for improving the crop production in agriculture. A few of Data mining methods, such as Artificial Neural Network(ANN), Decision Tree algorithm, Regression Tree, Bayesian network, Support Vector Machine(SVM), k means are used for classification [13].

Kaur et al analyze the different Data Mining techniques to find suitable data model that helps in achieving high accuracy for price prediction. Coimbatore market price of tomato data are collected and price is predicted using BP neural network and the result is simulated using MATLAB [14].

Ankalaki et al presents a comparative study on DBSCAN and AGNES algorithm for clustering. Crop yield is forecasted using MLR (Multiple Linear Regression) and a formula is derived for each crops [15].

Gayatri et al use IOT and web services to handle large amount of data. Sensors are used to collect the data and pass the data to data centre. Agriculture field images are captured and GPS is used to accurately feed the data into repositories along with their location. Far and near nodes are communicated through cloud [16].

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

Author and publication	Techniques used	Parameters achieved	Limitations
Grajales , 2015	A web application is developed using Open source tools.	The details of the selected location from map are available at one look to the user.	Nil
Bendre , 2015	Map Reduce and Linear Regression algorithm are used for weather forecasting.	The effective model to improve the accuracy of rainfall forecasting is investigated.	The forecasting is done based on only a weather data.
Hemageethaa , 2016	Naïve Bayes, Appriori algorithm are used for yield prediction.	Focuses mainly on various soil parameters like pH, Nitrogen, moisture etc and comparison accuracy is also presented.	Only 77% of accuracy is achieved.
Rub , 2009	MLP, Regression tree, RBF, SVM are used.	Comparative study of various algorithms is presented.	Nil
Sujatha , 2016	Naïve Bayes, J48, random forests, support vector machines, artificial neural networks are implemented.	Climate data and Crop parameters are used for crop yield is predicted.	Other parameters like soil are not considered.
Kushwala , 2015	Hadoop Distributed File System (HDFS) is used.	The proposed prediction algorithm helps in building a decision support system for precision farming.	It only predicts the suitability of crop for the given soil parameters and not the yield.
Fathima , 2014	k means and Appriori algorithm are used.	Crop type and Irrigation parameters are considered.	Focus on the policies that government could frame by the cropping practices of farmers
Veenadhari , 2014	k-means, ID3 algorithms, the k nearest neighbor, support vector machines, artificial neural networks are discussed.	The purpose of Data Mining techniques in the field of agriculture is presented.	Nil
Sellam , 2016	Regression Analysis (RA), Linear Regression (LR) are cited.	Describes about various environmental factors that influence the crop yield and the relationship among these parameters is also established.	Nil
Raorane , 2012	Artificial Neural Network (ANN), Decision Tree	The techniques used for crop production is	Nil

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

	algorithm, Regression Tree, Bayesian network, Support Vector Machine (SVM) and K means are discussed.	discussed.	
Kaur , 2014	They use BP neural network and simulate the result using MATLAB.	The suitable data model for achieving high accuracy for price prediction is found.	The prediction is mainly based on only price.
Ankalaki , 2016	DBSCAN, AGNES and MLR are used.	The comparative study between DBSCAN and AGNES is presented.	The formula is derived for each crop separately.
Gayatri , 2015	IOT and GPS Image capturing are used.	Images are captured. Far and near nodes are communicated through cloud.	Focuses mainly on image processing techniques.

### III. CONCLUSION AND FUTURE WORK

There are various systems that use diverse data mining technologies to manipulate data to derive insights and help in decision making for farmers. But the major problem is that they focus either on one crop and predict or forecast any one parameter like either yield or price. This project is used to forecast the yield and price of major crops of Tamil Nadu based on historical data.

The data and predicted output are accessible for the farmers through a web application. This aids farmer to decide on the crop they would like to plant for the forthcoming year. In addition the web application also provides a forum for the farmers to merchandise the products without middlemen which help them to obtain maximum price for their products. The discussion forum enables users to clarify doubts regarding the usage of the web application.

### ACKNOWLEDMENT

The authors express their sincere gratitude to the Principal, Director Academics and Head of the Department of Computer Science and Engineering of Sri Ramakrishna Engineering College for giving constant encouragement and support to complete the work.

### REFERENCES

- [1] Kalyani, M.R., "Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, Issue 10, pp.439-442, 2012.
- [2] Agrawal, H., Agrawal, P., "Review on Data Mining Tools", International Journal of Innovative Science, Engineering & Technology, Vol. 1, Issue 2, pp.52-56, 2014.
- [3] Grajales, D.F.P., Mosquera, G.J.A, Mejia, F., Piedrahita, L.C., Basurto, C., "Crop-Planning, Making Smarter Agriculture With Climate Data", Fourth International Conference on Agro-GeoInformatics, pp.240-244, 2015.
- [4] Bendre, M. R., Thool, R.C., Thool, V. R., "Big Data in Precision Agriculture : Weather Forecasting for Future Farming", 1<sup>st</sup> International Conference on Next Generation Computing Technologies, pp.744-750, 2015.
- [5] Tan, L., "Cloud-based Decision Support and Automation for Precision Agriculture in Orchards", ScienceDirect, pp. 330-335, IFAC-Papers OnLine 49-16, pp.330-335, 2016.
- [6] Hemageetha, N., "A survey on application of data mining techniques to analyze the soil for agricultural purpose", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.3112-3117, 2016.
- [7] Rub, G., "Data Mining of Agricultural Yield Data:A Comparison of Regression Models", 9<sup>th</sup> Industrial Confrence, Vol.5633, pp.24-37, 2009.
- [8] Sujatha, R., Isakki, P., "A study on crop yield forecasting using classification techniques", International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE), pp.1-4, 2016.
- [9] Kushwaha, A.K., SwetaBhattachrya, "Crop yield prediction using AgroAlgorithm in Hadoop", International Journal of Computer Science and Information Technology & Security (IJSITS), Vol. 5- No2, pp.271-274, 2015.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 3, March 2017

- [10] Fathima, G.N., Geetha, R., "Agriculture Crop Pattern Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Engineering, Vol. 4, Issue 5, pp.781-786, 2014.
- [11] Veenadhari, S., Misra, B., Singh, C.D., "Machine learning approach for forecasting crop yield based on climatic parameters", International Conference on Computer Communication and Informatics, pp.1-5, 2014.
- [12] Sellam,V., Poovammal, E., "Prediction of Crop Yield using Regression Analysis", Indian Journal of Science and Technology, Vol. 9(38), pp.1-5, 2016.
- [13] Raorane, A.A., Kulkarni R.V., "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends & Technology in Computer Science(IJETTCS), Vol. 1, Issue 2, pp.75-79, 2012.
- [14] Kaur, M., Gulati, H., Kundra, H., "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Vol. 99– No.12, pp.1-3, 2014.
- [15] Ankalaki, S., Chandra, N., Majumdar, J., "Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Special Issue 2, pp.25-29, 2016.
- [16] Gayatri, M.K., Jayasakthi, J., Anandha Mala, G.S., "Providing Smart Agricultural Solutions to Farmers for better yielding using IoT", IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), pp.40-43, 2015.
- [17] [www.tnau.ac.in](http://www.tnau.ac.in)
- [18] [http://www.tutorialspoint.in/data\\_mining/](http://www.tutorialspoint.in/data_mining/)

