



## Rice Crop Yield Prediction using Data Mining Techniques: An Overview

**Dakshayini Patil**

PG Scholar, Dept. of CSE

B.L.D.E.A'S Dr.P.G.Halakatti College of Engineering  
and Technology, Vijayapur, Karnataka, India**Dr. M .S Shirdhonkar**

Professor, Dept. of CSE

B.L.D.E.A'S Dr.P.G.Halakatti College of Engineering  
and Technology, Vijayapur, Karnataka, IndiaDOI: [10.23956/ijarcsse/SV7I5/0135](https://doi.org/10.23956/ijarcsse/SV7I5/0135)

**Abstract**— This paper shows the overview of rice crop yield prediction. Examines Different data mining techniques utilized for foreseeing rice crop yield. Rice crop creation assumes an imperative part in sustenance security of India, contributing over 40% to general yield generation. High harvest generation is reliant on appropriate climatic conditions. Inconvenient regular atmosphere conditions, for example, low precipitation or temperature extremes can drastically diminish edit yield. Growing better strategies to foresee edit efficiency in various climatic conditions can help rancher and different partners in vital basic leadership as far as agronomy and harvest decision. This paper reports utilization of various information mining methods will anticipate rice trim yield for Maharashtra state, India. To this review, 27 regions of Maharashtra were picked on the establishment of accessible information from openly available Indian Administration records with different atmosphere and harvest parameters. Precipitation, least temperature, normal temperature, most extreme temperature, reference trim evapotranspiration, range, generation and yield for the Kharif season (June to November) were the parameters chosen for the study for the years 1998 to 2002. WEKA tool was used for dataset processing

**Keywords**— WEKA tool; Classifiers; Crop analysis; Data mining; Yield prediction

### I. INTRODUCTION

Farming structures the establishment of Indian economy. Vast territories of rural land are not accomplishing satisfactory yield generation because of both climatic and monetary difficulties. The harvest creation on a very basic level depends upon climatic conditions, which impacts on yield improvement. Water lack, nonappearance of high yielding assortments and poor innovation exchange of best agronomic practices are thought to be the vital components for low harvest yields in India. On the off chance that harvest generation is to increment both yield creation and the administration of trimming region must move forward. The utilization of innovation and different software engineering procedures to anticipate crop productivity in various climatic conditions can help farmers different partners in critical basic leadership as far as agronomy and yield decision. Varieties in atmosphere, soil, water, bother attack and different components add to the intricacy of the issue

The software engineering system of Bayesian Networks (BN) has found a huge relevance in conditions that require indicative thinking for contingent conditions. These strategies can be utilized to evaluate show structures and the instability of parameters. Furthermore, frameworks with expanding complexities, additionally observe this method to be very reasonable. To characterize the associations between the factors, BN utilizes probabilistic parts of a system rather than deterministic examinations. It is a planned non-cyclic chart. In a system, it contains a probabilistic portrayal of connections among factors. Fig 1 shows the study area and Fig 2 shows climatic parameters considered for the work.

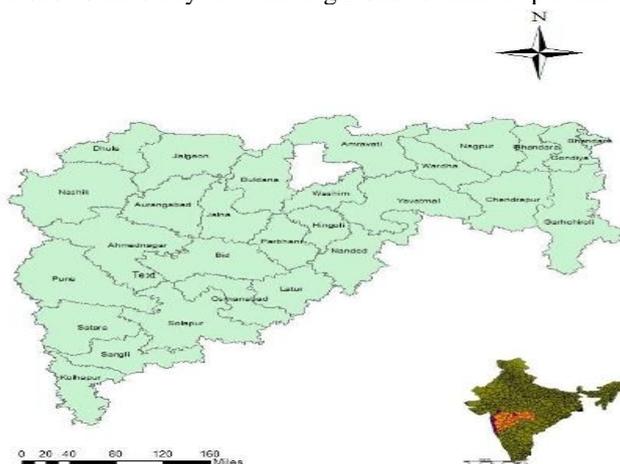


Fig. 1 Districts of Maharashtra State, India

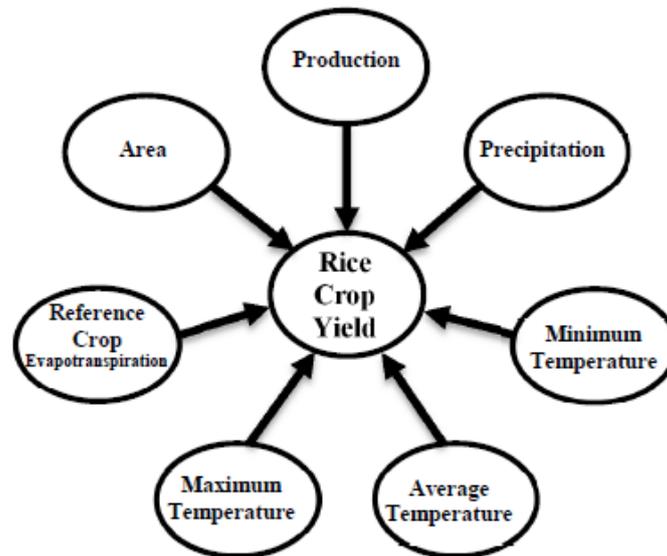


Fig. 2 climatic parameters

## II. METHODOLOGY USED

The phases in this research includes: (1) pre-processing of dataset (2) Building the prediction model utilizing WEKA and (3) Analysing the outcomes

### A. Dataset pre-processing

Broad pre-processing was done to examine all the datasets in Microsoft Office Excel. This was done to deal with the missing qualities and other information irregularities. It comprised of taking after segments sr. no, name of the locale, precipitation, least temperature, normal temperature, greatest temperature, reference edit evapotranspiration, zone, generation and yield. The document was then sorted on the premise of territory. Records with territory under 100 hectares were evacuated as they were not considered for the present review.

### B. Building the prediction model using WEKA

WEKA utilizes the Java programming dialect for less demanding cooperation with information documents and conveying visual outcomes it contains a GUI. The GUI additionally makes it less demanding to utilize the complete gathering of information pre-processing and demonstrating systems. It has a general programming interface and the WEKA libraries. The standard information mining errands like information pre-handling, grouping, arrangement, relapse, representation, and highlight choice are likewise bolstered by WEKA. Every one of WEKA's methods are predicated utilizing information is made in a solitary level record or connection, where every information point is portrayed by a settled number of traits. Access to SQL databases for preparing of the outcomes was using Java Database network. The informational collection arranged for the examination was then viewed in WEKA and spare in .arff organize for building the forecast demonstrate.

### C. Analysing the results

Cross validation is an approach to examine how an anticipated information mining model will perform on an obscure dataset, i.e how well the model sums up. For the current dataset, 10-fold cross validation technique was utilized to subset the information for preparing and testing. The information were haphazardly partitioned into 10 sections which one section was utilized for testing while the rest of the parts were utilized as the preparation information. The procedures were rehashed 10 times and the outcomes were normal.

## III. APPLICATIONS

- Predicts crop yield production of rice for Maharashtra state, India.
- Predicting productivity of crop in various climatic conditions can help farmer and other partners in essential basic leadership as far as agronomy and product decision.
- It also helps to increase the yield of crop

## IV. RELATED WORK

In [2] Vikas Chawla et al (2016), propose a data-driven approach that is 'gray box' i.e. that seamlessly utilizes expert knowledge in constructing a statistical network model for corn yield forecasting. Multivariate gray box model is developed on Bayesian network analysis to build a Directed Acyclic Graph (DAG) between predictors and yield. Starting from a complete graph connecting various carefully chosen variables and yield, expert knowledge is used to prune or strengthen edges connecting variables. Subsequently the structure (connectivity and edge weights) of the DAG that maximizes the likelihood of observing the training data is identified via optimization. The focus of this work is to

construct a corn yield predictor at the county scale. Corn yield (forecasting) depends on a complex, interconnected set of variables that include economic, agricultural, management and meteorological factors. Conventional forecasting is either knowledge-based computer programs (that simulate plant-weather-soil-management interactions) coupled with targeted surveys or statistical model based. The former is limited by the need for painstaking calibration, while the latter is limited to univariate analysis or similar simplifying assumptions that fail to capture the complex interdependencies affecting yield.

In [3] N. Gandhi et al (2016), presented the overview on utilization of machine learning system for Indian rice editing ranges. Machine learning systems can be used to enhance forecast of harvest yield under various climatic situations. This paper examines at the exploratory outcomes acquired by applying SMO classifier utilizing the WEKA apparatus on the dataset of 27 areas of Maharashtra state, India. Those dataset acknowledged to the rice trim yield forecast might have been sourced from openly available Indian organization records. The parameters recognized for those review were precipitation, base temperature, Normal temperature, most extraordinary temperature Furthermore reference trim evapotranspiration, territory, arrangement Furthermore yield for those Kharif season (June to November) for the years 1998 to 2002.

In [4] M.C.S.Geetha (2015), examined about the piece of information mining in context of cultivating field and furthermore presents around a few information mining systems and their related work by a few creators in setting to horticulture area. It additionally examines on various information mining applications in tackling the distinctive horticultural issues. Horticulture is the most noteworthy application region especially in the creating nations like India. Utilization of data innovation organization over agribusiness might change those circumstance from claiming choice making Furthermore farmers could yield clinched alongside An exceptional approach. Information mining assumes an significant part to choice making looking into a few issues identified with farming field. This paper gives a overview of different information mining systems utilized within agribusiness. Research paper expects in discovering suitability information models that attain An secondary precision What's more a secondary sweeping statement for admiration to four parameters in particular rainfall, year, handling also region from claiming sowing. To this purpose, distinctive sorts from claiming information mining systems were assessed around separate information sets.

In [5] D. Ramesh and B. Vardhan (2015), exhibited An short Investigation about crop yield prediction utilizing Density based clustering technique and Multiple Linear Regression (MLR) for the selected region. A recent development in Information Technology for horticulture field has turned into an intriguing exploration region to anticipate the harvest yield. The issue from claiming yield forecast may be a significant issue that remains will be tackled In view of accessible information. Diverse information mining strategies are utilized Also assessed in agribusiness assessing what's to come year's harvest creation. At first the factual model Multiple Linear Regression strategy is connected on existing information. The impacts so gotten were checked Also investigated utilizing the information mining framework to be particular Density-based grouping strategy. In this system the aftereffects of two strategies were looked at as expressed by particular locale.

In [6] R. Medar and V. Rajpurohit (2014), presents the various crop yield prediction methods using data mining techniques. Agricultural system is very complex since it deals with large data situation which comes from a number of factors. Crop yield prediction need been a subject sentence about premium to producers, consultants, and agricultural related associations. In this paper our focus is on requisitions about information mining systems in agricultural field. Distinctive Data Mining strategies, extremely late utilizations of information mining procedures in farming field. Information mining innovation has gotten an extraordinary advance with the quick improvement of software engineering, computerized reasoning. Information Mining is a developing examination field in agribusiness edit yield investigation. Information Mining is the way toward recognizing the concealed examples from extensive measure of information. The issue of yield forecast can be fathomed by utilizing information mining methods. With the change of information mining advancements, particularly those with no premises or people subjective, information mining can be connected in numerous territories. In this paper a few information mining strategies were embraced keeping in mind the end goal to gauge trim yield examination with existing information and their utilization in information mining.

In [7] R. Kalpana et al (2014), presents a overview on information mining techniques in agribusiness. Data processing in Agriculture is associate rising analysis field. Application of information mining techniques plays an important role in agricultural and environmental connected areas. Survey created on agriculture mistreatment data processing techniques reveals the importance to extract helpful data from dataset. Recently data processing techniques in agriculture required evaluating, storage, monitored and retrieval the resources used. Completely different data processing techniques measure terribly recent applications in agriculture. In agriculture, some issues like yield estimation and crop productivity remains to be solved supported obtainable knowledge. This survey aims to search out appropriate data processing models to realize high accuracy and prediction capabilities. It's the opinion that additional techniques and algorithms to be studied connected agricultural issues can provide smart lead to agricultural growth. Finally, mistreatment data processing techniques in agriculture could be a up to date technique to search out the answer over the normal and traditional technique.

In [8] G.N.Fatima and R.Geeta (2014), explores the utilization of different information digging strategies for learning disclosure in horticulture division. Learning disclosure in budgetary association have been worked to assess their operation and fundamentally to bolster basic leadership utilizing information as key component. Existing programming are wasteful in demonstrating such information attributes. Creators present distinctive shows for finding learning as affiliation principles, bunching, arrangement and relationship appropriate for information attributes. Proposed information mining systems, the leader can characterize the extension of agribusiness exercises to enable the diverse strengths in existing farming division. In the agribusiness division, information mining can help government to expand

yield advantage principally to bolster basic leadership, solid and auspicious data on harvest region, edit creation and land utilize is of incredible significance to organizers and arrangement producers for effective agrarian improvement and for taking choices on obtainment, stockpiling, open dissemination, fare, import and numerous other related issues to contend in the distribute of product example.

In [9] Mr. Abhishek B. Mankar and Mr. Mayur S. Burange (2014), concentrate on the uses of Data mining systems in the field of agribusiness. It is exceptionally hard to obtain the data what truly need with the aggregation of vast amount of information. Information mining innovation need been gained an incredible advancement for those fast improvement of computer science, artificial intelligence. Yield prediction will be a critical agricultural issue to rancher that remains will be comprehended in view of that past accessible information. The yield forecast issue can be illuminated by actualizing data mining procedures. Various information mining strategies were assessed looking into distinctive information sets. In this paper a portion information mining strategies were embraced in place with evaluate crop yield Investigation for existing information and their utilization Previously, information mining. Some information mining system need not yet connected to farming problem, Concerning illustration a sample GPS strategies might make utilized for finding significant majority of the data starting with agricultural-related such as soil identification.

In [10] S. Dahikar and S. Rode (2014), proposed simulated neural system approach for horticultural product yield expectation. By considering different circumstances of climatologically marvels influencing neighborhood climate conditions in different parts of the world. These climate conditions directly affect edit yield. Different examines have been done investigating the associations between extensive scale climatologically marvels and product yield. shown to be intense devices for displaying and expectation, to expand their adequacy. Edit forecast procedure is utilized to foresee the appropriate harvest by detecting different parameter of soil and furthermore parameter identified with climate parameters. For that reason we are utilized Artificial Neural Network (ANN). Authors inferred that ANN is valuable apparatus for product expectation. In this paper incorporates the parameter of their provincial soil parameter. At that point it is examine by utilizing encourage forward back engendering ANN. Break down in tangle lab ANN way to deal with make it more effective.

In [11] A.A.raorane and R.V.kulkarni (2013), proposed part of information mining in farming. Dependable and convenient gauges of harvest generation are required for different approach choices identifying with storage, distribution, valuing, promoting, import-send out, and so forth. The official forecasts significant grain and business products are issued by Directorate of Economics and Statistics. However, these advance estimates are just estimates and not the goal estimates. In working out these evaluations there is bunches of subjective appraisal in light of a few subjective components. There is along these lines a need to grow measurably solid target forecast of yield grounds and creation. The advances in processing and data stockpiling hove gave immense an a large portion of information. The test has been to concentrate information from this crude information; this has prompt new strategies and systems. In perspective of this, there is a requirement for a target strategy for pre-reap trim estimating. This includes developing reasonable estimate model(s) which has certain benefits over the conventional forecasting strategy.

In [12] Yethiraj N G (2012), made an endeavor to survey the exploration on utilization of information mining systems in the field of farming. Information mining in application in agribusiness is a moderately new way for anticipating and foreseeing of rural product and creature administration. This article investigates the utilizations of information mining procedures in the field of horticulture and partnered sciences. There is a developing number of utilizations of information mining procedures in horticulture and a developing measure of information that are as of now accessible from numerous assets. This is generally a novel research field and it is required to develop later on. There is a considerable measure of work to be done on this developing and fascinating exploration field. The multidisciplinary approach of incorporating software engineering with agribusiness will help in estimating/overseeing agrarian harvests successfully.

In [13] A.A.raorane and R.V.kulkarni (2012), intended to evaluate new information mining methods and apply them to the different factors comprising in the database to build up if important connections can be found. Farming is a business with hazard. Crop production relies on upon climatic, geological, organic, political and financial elements. On account of these elements there are a few dangers, which can be evaluated when connected suitable scientific or factual strategies. Really exact data about the way of verifiable yield of harvest is critical demonstrating input, which are useful to agriculturists and Government association for basic leadership prepare in building up legitimate strategies. It is watched that productive method can be created and investigated utilizing the fitting information, the information which is gathered from Kolhapur region to take care of complex horticultural issues utilizing Data mining systems.

In [15] N.K. Newlands and L. Townley-Smith (2010), proposed Bayesian system approach for foreseeing vitality edit yield. Common asset issues regularly should be demonstrated utilizing information that is frequently fragmented at various spatial and fleeting scales with various levels of instability. Fluctuation because of atmosphere, soil, nuisances and administration choices add to facilitate auxiliary and useful many-sided quality of oversaw environments. Bayesian systems are perfect for such circumstances by empowering symptomatic thinking on contingent conditions to survey display auxiliary and also parameter vulnerability. They apply Bayesian systems to the issue of providing territorial with an ideal, hearty supply of biomass from vitality crops. Crops have diverse ideal atmosphere, water and supplement necessities, and affectability to outrageous climate, intrusive vermin and different effects. They test an improved model form in southern Manitoba, Western Canada. We analyze affectability of ideal respect planting/reap timing under verifiable climate, water, supplements and extraordinary occasion/bug misfortune inconstancy. We look at changed classifiers in getting a system arrangement and examine future work to apply the model at higher spatial and transient determination.

## V. CONCLUSION

This paper surveys the technical achievements in the field of Rice crop yield prediction. Discusses methodology, comprehensive survey of various proposed methods to predict rice crop yield and applications. It also discusses various data mining techniques used for prediction of crop yield for rice. Growing better strategies to foresee crop productivity in various climatic conditions can help farmer and different partners in essential basic leadership as far as agronomy and product decision

## REFERENCES

- [1] Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, “ *Predicting Rice Crop Yield Using Bayesian Networks*”, Conference on Advances in Computing, Communications and Informatics, 2016
- [2] Vikas Chawla, Hsiang Sing Naik, Adedotun Akintayo, “*A Bayesian Network approach to County-Level Corn Yield Prediction using historical data and expert knowledge*”, Data Science for Food, Energy and Water, 2016
- [3] N.Gandhi, L.J. Armstrong, O. Petkar and A. Tripathy, “*Rice Crop Yield Prediction in India using Support Vector Machines*”, IEEE The 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Thailand, 2016
- [4] M.C.S.Geeta, “*A survey on data mining techniques in Agriculture*”, International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, No. 2, pp. 887-892, 2015
- [5] D. Ramesh and B. Vardhan, “*Analysis of crop yield prediction using data mining techniques*”, International Journal of Research in Engineering and Technology, vol. 4, no.1, pp. 47-473, 2015
- [6] R. Medar, V. Rajpurohit, “*A survey on data mining techniques for crop yield prediction*”, International Journal of Advance Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59-64, 2014.
- [7] R.Kalpna, N.Shanti and S.Arumugam, “*A survey on data mining techniques in Agriculture*”, International Journal of advances in Computer Science and Technology, vol. 3, No. 8, 426- 431, 2014
- [8] G.N.Fatima, R.Geeta, “*Agriculture crop pattern using data mining techniques*”, International Journal of Advanced Research in in Computer Science and Software Engineering, vol. 4, No. 5, 781-786, 2014
- [9] Mr. Abhishek B. Mankar, Mr. Mayur S. Burange, “*Data Mining - An Evolutionary View of Agriculture*”, International Journal of Application or Innovation in Engineering & Management (IJAIEM), vol. 3, No. 3, pp. 102-105, 2014
- [10] S. Dahikar and S. Rode, “*Agricultural crop yield prediction using artificial neural network approach*”, International Journal of Innovative Research in Electrical, Electronic Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683-686, 2014
- [11] A.A.Raorane and R.V.Kulkarni, “*Review-Role of data mining in Agriculture*”, International Journal of Computer Science and Information Technology, vol. 4, No. 2, pp. 270-272, 2013
- [12] Yetheraj .N.J, “*Applying data mining techniques in the field of agriculture and allied sciences*”, International Journal of Business Intelligent, vol. 1, No. 2, pp. 72-76, 2012
- [13] A.A.Raorane and R.V.Kulkarni, “*Data Mining-An effective tool for yield estimation in the agricultural sector*”, International Journal of Engineering Trends and Technology, vol. 1, No.2, pp. 75-79, 2012
- [14] S.Veenadhari, B.Mishra and C.D.Singh, “*data mining techniques for predicting crop productivity-A review article*”, International Journal of Computer Science and Technology, vol. 2, No. 1, pp. 98-100, 2011
- [15] N.K.Newlands, L. Townley-Smith, “*Predicting Energy Crop Yield Using Bayesian Networks*”, In Proceedings of the Fifth IASTED International Conference, Vol. 711, pp. 014-106, 2010