ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **10** (2021), no.3, 1425–1430
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.10.3.32

A DATA MINING APPROACH FOR CROP YIELD PREDICTION IN AGRICULTURE SECTOR

Puninder Kaur[1], Jasmeen Kaur Chahal, and Taruna Sharma

ABSTRACT. The economy of a country depends on the agriculture. The GDP growth of country is also depends on agriculture sector. It provides the employment of the nation's workforce. The development of crop depends on various factors such as climate condition, fertilization, water supply, land management. These factors are not easy to identify so the data mining technique are applied for yield prediction. Data mining techniques uses various algorithms like C4.5, CART algorithms, Naïve Bayes algorithm, EM algorithm and SVM etc. The data mining technique take row data, and then apply various data mining algorithm, and produce appropriate output. In this survey paper various algorithms are discussed and accurate results are produced that helps the formers for cultivation.

1. INTRODUCTION

Cultivating is the vital segment that follows up on the financial arrangement of India. The growth of the farmers is totally dependents upon the yield. So every farmer is very much excited to know about the status of yield. The growth of the yield depends upon following parameters like climate condition, amount

[1]*corresponding author*

2020 *Mathematics Subject Classification.* 94A16.

Key words and phrases. Data mining, Crop Yield Predictions, Prediction Models, Classification, Clustering.

Submitted: 22.02.2021; *Accepted:* 13.03.2021; *Published:* 22.03.2021.

of water supplied to field, fertilizers apply, land management, rain and important concept is property of [1]. The soil is currently abused in a wrong manner and the over water system and over utilization of manures and pesticides are corrupting it truly. So, we need to aware about the latest technology, parameters depends for yield of the crop and amount of fertilizers apply for production of crop[2]. Individuals of India are working on cultivating for a huge number of years however results are failing to satisfy because of different parameters that influence the product yield. Because of different components like soil compose, precipitation, quality of the seed, lack of technical knowledge of about crop yield are still now unaware. So, the methods like artificial neural network and data mining are developed for yield prediction in advance and classification of particular crop [3].

This research work mainly focuses on implementing on various crop production by utilizing data mining techniques by performing searching on various cultivation data properties of soil because our crops production totally depends upon soil property like appropriate amount of nitrogen, potassium, calcium and magnesium etc should be available in our soil. These analysis we achieve two goals. First this system will visualize the influence of crop movement on fertility of class of soil and second is its effect on total yield of production. So, for improvement of soil is essential and can be possible with the help of rotation on land [4]. Paddy, wheat and other grain upland movement is the very much significant farming system. Nonetheless, the current state of affairs is not very optimistic, because the cropped fields of rice and wheat have gone down in the last decade.. Data mining provides various algorithms and models that are helpful in agriculture sectors [5][6].

1.1. **Data Mining.** Data Mining means mining of data. This method is investigating, extracting and identifying the meaningful data from large amount of data and extract it in specific manner or in particular pattern. It includes the selection of data followed by preprocessing and modifies data as per requirement. Pre processing includes detecting missing and duplicates values [7][8]. The mining techniques take historical data about production of yield according to the climate change. Several predictions of data can be applied on the basis of knowledge gathered which can be turn with the help in increasing crop productivity. Decision Support System (DSS) is also implemented for the farmers and

prevent the overheads of decisions about the soil and crop to be cultivated. This type of framework can help farmers to take critical decisions. DSS is a software system that helps the analysts to predict or identify useful information from a raw data, documents or business models for analyzing problems and solve it by taking decisions".

Data mining technique use multiple linear regressions are also along with Support Vector Machine and modified non-linear regression. The prediction of yield depends on climate condition and prediction of weather can be possible data mining techniques and machine learning techniques are associate with each other. This type of framework helps to take critical decisions which were prior taken by using utilizing various methods. This framework can be applied by data mining techniques. The applications of data mining techniques in agricultural are examined by researches [9].

Table 1: Various Data Mining classes [9][10]

Data mining Classes	Data mining methods	Data Mining Algorithms	Application Area
CLASSIFICATIONS	NEURAL NETWORK(NN)	CART Alorithm	Network Management Identifying and control network traffic
	Decision Tree	Support Vector Machine	Face Detection and Classification of images
	K-Nearest Neighbor	KNN Algorithm	Data compression
	Naïve Bayes Classification	Naïve Bayes Algorithm	Real Time prediction
CLUSTERING	K-Means Clustering Technique	Expectation-Maximization (EM)	Sales
	Hierarchical clustering	Hierarchical clustering Algorithms	CRM, Marketing
	DBSCAN	DENCLUE Algorithms	Identify Outlier
REGRESSION	Linear regression	CART Algorithm	Number prediction(like height of children)
ASSOCIATION RULE	Association Rule Mining	Apriori Algorithm, FP Tree, Fuzzy FP Tree	Market Basket Analysis, protein Sequencing
OUTER	Anomaly Detection model	Outlier Detection Algorithm	Fraud Detection

2. Review of Literature

N. Gandhi et al. [1] researched and investigated different datasets for making assumption of different rice production in different conditions of zones of humid, equatorial wet and dry environment portions in India. In this research paper various classification algorithms are executed and weka as data mining

tool used for running all classification algorithms. It proposes different models for future guidance. Chlingaryan et al. [2] is emphasis on to predicting yield using combination of both machine learning and remote sensing device. The machine learning technique takes large number of input and check the status of nitrogen in the soil. This hybrid technique is cost effective and gives appropriate and fast result. Tamil et. al [3] in this research paper the author emphasis on to predict yield prediction using new algorithm is termed as Lemuria. In this paper both k-means clustering technique and naïve bayes technique with PSO are used for training and testing respectively. After this with the help of DSS the actual result is used that is used for predicting crop.

Table 2: Summary of the literature Review

Reference	Methodology	Summary	Advantages	Issues
Alberto Gonzalez-Sanchez et al. [4]	WEKA	Research on making forecast capacity of machine learning methods this paper also compare the machine learning and linear regression techniques.	ANN, regression trees, k-nearest neighbor regression and support vector for better yield	Data set can be increased.
S. Mishra et. al [5]	WEKA	This paper focus for predicting crop production model using data mining methods.	Various classification methods are used for predicting cost and fault of a model and IBK gives highest result.	Calculated Parameters are not fixed so the result varies and it has more complexity.
Pantazi et al [6]	CP-ANN, XY-Fusion, MATLAB.	The researcher focuses on the soil property for predicting yield prediction. The main target based on wheat.	The merits of this research are to use of the wireless sensors network along with Artificial neural network and it gives better results.	Models are unable to give continuous output. High resolution of data is required.
Veenadhari et al. [7]	C 4.5, Crop advisor	Modern web applications are used for making prediction the effect of various climatic objective measures on the crop production.	The user friendly software tool crop advisor is developed and it gives 75 percent result for all types of crops.	More advance features can be added.
H. Patel et al [8]	WEKA Tool	Various classification models such as J48, Naïve Bayes and smart cart are compared.	J48 gives the best result and fast result.	The size of dataset can be increased.
Dey et al. [9]	Multiple linear regression	In this paper multiple linear regression Adaboost, SVM and MNR equation is used for accuracy.	In this paper non-linear multi Regression proves better and it also gives information for whether prediction.	This model depends on amount of crop so, extension of this model is under progress.

3. Conclusion

In conclusion, the agriculture is the major components for developing a country. The utilization of proper tools and techniques use the farmer for decision making. So, in this paper various algorithms are discussed for prediction of yield. Data mining deals with big and complex data. Data mining can take large datasets It extracts the useful information from large dataset and after pre-processing, it discards the useless information. These data mining provides various tools and algorithms such as SVM, J48, C4.5, LADTree, DENCLE algorithms and other classification and clustering algorithms. These algorithms solve the complex problems of agriculture and provide the accurate results that help the formers for agriculture growth. In future more advance tools and technologies can be developed for predicting crop and more parameters can be added and tested.

References

[1] N. Gandhi, L. Armstrong: *Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India*, In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE. (2016), 1901-1906.

[2] A. Chlingaryan, S. Sukkarieh, B. Whelan: *Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review*, Computers and electronics in agriculture, **151** (2018), 61-69.

[3] M. Tamil Selvi, B. Jaison: *Lemuria: A Novel Future Crop Prediction Algorithm Using Data Mining*, The Computer Journal, (2020).

[4] A. González Sánchez, J. Frausto Solís, W. Ojeda Bustamante: *Predictive ability of machine learning methods for massive crop yield prediction*, (2014).

[5] S. Mishra, P. Paygude, S. Chaudhary, S. Idate: *Use of data mining in crop yield prediction*, In 2018 2nd International Conference on Inventive Systems and Control (ICISC), IEEE, (2018), 796-802.

[6] X.E. Pantazi, D. Moshou, T. Alexandridis, R.L. Whetton, A.M. Mouazen: *Wheat yield prediction using machine learning and advanced sensing techniques*, Computers and Electronics in Agriculture, **121** (2016), 57-65.

[7] S. Veenadhari, B. Misra, C.D. Singh: *Machine learning approach for forecasting crop yield based on climatic parameters*, In 2014 International Conference on Computer Communication and Informatics, IEEE, (2014), 1-5.

[8] U. Ayub, S.A. Moqurrab: *Predicting crop diseases using data mining approaches: classification*, In 2018 1st International Conference On Power, Energy And Smart Grid (Icpesg), IEEE, (2018), 1-6.

[9] U.K. DEY, A.H. MASUD, M.N. UDDIN: *Rice yield prediction model using data mining*, In 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, (2017), 321-326.

[10] H. PATEL, D. PATEL: *A comparative study on various data mining algorithms with special reference to crop yield prediction*, Indian Journal of Science and Technology, **9**(22) (2016), 1-8.

CHITKARA UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY
CHITKARA UNIVERSITY, PUNJAB, INDIA.
Email address: puninder.k@chitkara.edu.in

CHITKARA UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY
CHITKARA UNIVERSITY, PUNJAB, INDIA.
Email address: jasmeen.chahal@chitkara.edu.in

CHITKARA UNIVERSITY INSTITUTE OF ENGINEERING AND TECHNOLOGY
CHITKARA UNIVERSITY, PUNJAB, INDIA.
Email address: taruna.sharma@chitkara.edu.in