



CROP YEILD PREDICTION USING ML

¹M Koteswararao, ²Sirigiri Naga Harsha Vardhani, ³Desu L S R Badhrinadh, ⁴Noorbasha Imamvali,
⁵Kesari Gowtham Reddy

^{1,2,3,4,5} Assistant Professor, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

Agriculture is the most important sector that influences the economy of India. It contributes to 18% of India's Gross Domestic Product (GDP) and gives employment to 50% of the population of India. People of India are practicing Agriculture for years but the results are never satisfying due to various factors that affect the crop yield. To fulfill the needs of around 1.2 billion people, it is very important to have a good yield of crops. Due to factors like soil type, precipitation, seed quality, lack of technical facilities etc the crop yield is directly influenced. Hence, new technologies are necessary for satisfying the growing need and farmers must work smartly by opting new technologies rather than going for trivial methods. This paper focuses on implementing crop yield prediction system by using Data Mining techniques The analysis on agriculture dataset .Differnt classifiers are used namely J48, LWL, LAD Tree and IBK for prediction and then the performance of each is compared using WEKA tooling enhancement of python with machine learning For evaluating performance Accuracy in linear regressing and Random forest and KNN classifier used as one of the accurse factors. The classifiers are further compared with the values of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Relative Absolute Error (RAE). Lesser the value of error, more accurate the algorithm will work. The result is based on comparison among the classifiers.

1. INTRODUCTION

Data Mining is the process of analyzing, extracting and predicting the meaningful information from huge data to extract some pattern. This process is used by companies to turn the raw data of their customer to useful information. The process of Data Mining includes first selection of data followed by preprocessing of data and then transforming the data to get patterns which can then be used to predict useful insights. Preprocessing includes finding outliers and detecting missing values whereas transformation finds the correlation between objects. Applying the data mining techniques on historical climate and crop production data several predictions. Can be made on the basis of knowledge

gathered which in turn can help in increasing crop productivity. Decision Support System (DSS) has to be implemented for the farmers to prevent the overheads of decisions about the soil and crop to be cultivated. DSS is a software system that helps the analysts to predict or identify useful information from a raw dataset, documents or business models to analyze a problem and solve it by making decisions". This system would help farmers to make important decisions which were earlier taken by using inefficient trivial methods or by guessing. The prediction system will be implemented by using data mining



techniques. Previous researches depict the application of data mining techniques in the agricultural sector. The dataset is used which is described. Several algorithms are used for analysis namely Classification algorithms namely – J48 and LAD Tree and Lazy Learner algorithms namely – IBK and LWL which are discussed in WEKA tool is used for analysis. The experimentation steps in WEKA and also depicts the confusion 2 matrix for each of the classifier. Performance measurements and their general definitions are elaborated in e RMSE, MAE, RAE, Sensitivity, Specificity and accuracy is defined. The performance of each classifier is evaluated through factors namely Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Relative Absolute Error (RAE). Accuracy is also compared which is depicted in in the end the Conclusion along with the future work is discussed was to performance by using machine learning with different types of algorithmic crop yield prediction.

2. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, the next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system from huge

data to extract some pattern. This process is used by companies to turn the raw data of their customer to useful information. The process of Data Mining includes first selection of data followed by preprocessing of data and then transforming the data to get patterns which can then be used to predict useful insights. Preprocessing includes finding outliers and detecting missing values whereas transformation finds the correlation between objects.

Applying the data mining techniques on historical climate and crop production data several predictions can be made on the basis of knowledge gathered which in turn can help in increasing crop productivity. Decision Support System (DSS) has to be implemented for the farmers to prevent the overheads of decisions about the soil and crop to be cultivated. DSS is a software system that helps the analysts to predict or identify useful information from a raw dataset, documents or business models to analyze a problem and solve it by making decisions". This system would help farmers to make important decisions which were earlier taken by using inefficient trivial methods or by guessing. The prediction system will be implemented by using data mining techniques. This work includes several sections as follows the previous work which were accomplished by several researchers. The motivation behind this paper is discussed in For experimentation the dataset is used which is described in machine learning . Several algorithms are used for analysis namely Classification algorithms namely – J48 and LAD Tree and Lazy Learner algorithms namely – IBK and LWL which are discussed in Section V.



WEKA tool is used for analysis. Section VI presents the experimentation steps in WEKA and also depicts the confusion matrix for each of the classifier. Performance measurements and their general definitions are elaborated. Terms like RMSE, MAE, RAE, Sensitivity, Specificity and accuracy are defined. The performance of each classifier is evaluated through factors namely Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Relative Absolute Error (RAE). Accuracy is also the Conclusion along with the future work is discussed.

3. SYSTEM ANALYSIS

3.1 Existing system:

Existing in all recommended food (high stability), otherwise these recommended food have no reason to be recommended. Different research in past explains that data mining techniques can help in building a system that could effectively Solve complex agriculture problems without less human intervention analyzed on agriculture data by using different techniques and later compared the performances here used ensemble model for precision agriculture proposed the use of data mining techniques for implementing Crop Selection Method (CSM). which tells about the sequence of crops to be planted surveyed about different data mining techniques and how they can be useful in agriculture sector, analyzed agriculture dataset for prediction of rice yield in humid subtropical climate zone and tropical wet and dry climate zone in India, here worked on prediction of crop yield using Multiple Linear Regression (MLR) and Density based clustering techniques. To study analyzed crop yield prediction by using Support Vector Machine (SVM), Multiple

Linear Regression (MLR), Adobos and Modified Non Linear Regressions related to the data mining techniques which are applied on agriculture dataset for analysis of different algorithms about analysis of different algorithms and at the end comparing them through different factors such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Accuracy or factors like Average Rain in Area, Average Sowing in Area etc. In [19] the authors have discussed about a type of agriculture called Precision Agriculture. In [20], authors worked on gathering remote sensing data and then further using it to build indices for analyzing crop productivity. The comparison of classification methods like KNN, Bayesian Network, and Decision Tree

Disadvantages:

1. Performance is less.
2. Less accuracy score.

3.2 Proposed system:

Proposed that we can recommend food according to the body's Creatinine values. However, the above studies are basically carried out through long-term clinical trials, which just recommend food for certain specific diseases and they seldom study the relationship between nutritional ingredients and diseases by data mining techniques.

Advantages:

1. Performance is improved.
2. Less cost effective.

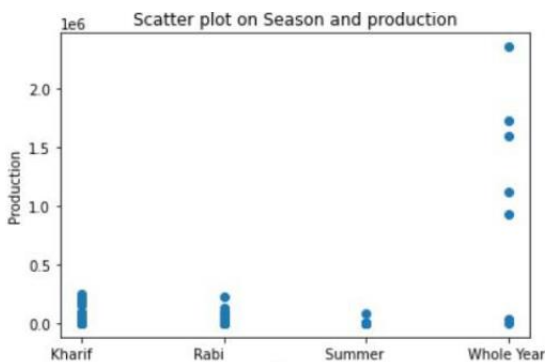
Results:

District_Name	Crop_Year	Season	Crop	Area	Production	
0	AURANGABAD	2014	Kharif	Arhar/Tur	32900	8700.0
1	AURANGABAD	2014	Kharif	Bajra	45300	24200.0
2	AURANGABAD	2014	Kharif	Cotton(lint)	422600	244300.0
3	AURANGABAD	2014	Kharif	Groundnut	4100	900.0
4	AURANGABAD	2014	Kharif	Jowar	3700	1700.0

Data main to identify dataset

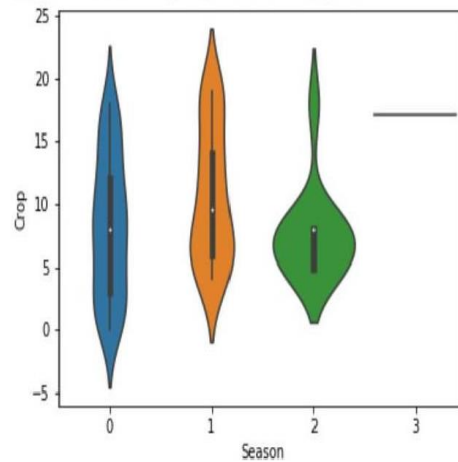
	Crop_Year	Area	Production
count	151.0	151.000000	1.350000e+02
mean	2014.0	38315.238411	7.848001e+04
std	0.0	74309.750744	3.074922e+05
min	2014.0	100.000000	1.000000e+02
25%	2014.0	1300.000000	6.000000e+02
50%	2014.0	6200.000000	4.200000e+03
75%	2014.0	41450.000000	2.495000e+04
max	2014.0	422600.000000	2.354600e+06

Data Description of dataset



Scatter Plot graph

<matplotlib.axes._subplots.AxesSubplot at 0x28c6754c1d0>



Violin plot graph

3. CONCLUSION:

The errors for different classifiers can be minimized if the dataset is pruned further by decreasing the confidence factor. Lesser the errors more accurate will be the analysis. IBK achieves highest accuracy whereas; LAD Tree has the lowest accuracy. The information that to acquired after analysis of machine learning algorithm in development can be combined in a form that is useful to the farmers for early prediction and decision making process. With the help of this information the percentage of loses and unsatisfactory yield will decrease as the management of the whole process can be done with the help of real statistics. In future we can use real time weather and soil datasets which will be gathered personally by equipments or the datasets can be acquired from trusted websites like ndiaagrstat.com or india.gov.in/data-portal-India. To further modify the model we can combine different classifiers to build one single model which is called Ensemble. By doing this we can



achieve a level of performance which could not be achieved by single algorithm. Also, the nature of Dataset affects the analysis therefore, more cleaned and pre processed can be used for better results. To better performance of the executable on data set to predict the results and made of inbuilt with library and merge the weka library in python with machine learning based on to development with classification algorithms like random forest, linear Regression, svm and KNN algorithms based on results to analysis with evaluate the best accuracy score and predict the target values.

REFERENCES

1. Mucherino A, Papajorgji P, Pardalos PM: A survey of data mining techniques applied to agriculture. *Oper Res.* 2009, 9 (2): 121-140.
2. Cunningham, S. J., and Holmes, G. (1999). Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference, 1999.
3. N.Gandhi and L.J. Armstrong, "Applying data mining techniques to predict yield of rice in Humid Subtropical Climatic Zone of India", Proceedings of the 10th INDIACom-2016, 3rd 2016 IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, 16th to 18th March 2016.
4. N. Gandhi and L. Armstrong, "Rice Crop Yield forecasting of Tropical Wet and Dry climatic zone of India using data mining techniques", IEEE International Conference on Advances in Computer Applications (ICACA), pp. 357-363, 2016.
5. N.Gandhi and L.J. Armstrong, "A review of the application of data mining techniques for decision making in agriculture", 2016 2nd International Conference on Contemporary Computing and Informatics (ic3i).
6. S.Pudumalar, E. Ramanujam, R. Harine Rajashreeñ, C. Kavyań, T. Kiruthikań, J. Nishań, 'Crop Recommendation System for Precision Agriculture', 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).
7. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and 60 anagement for Computing, Communication, Controls, Energy and Materials (ICSTM).
8. WEKA 3: Data Mining Software in Java, Machine Learning Group at the University of Waikato, Official Website: <http://www.cs.waikato.ac.nz/ml/weka/index.html>, accessed on 12nd October 2017.
9. 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.
10. Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice yield prediction model using data mining", International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.
11. A. Ahamed, N. Mahmood, N. Hossain, M. Kabir, K. Das, F. Rahman, R. Rahman, "Applying data mining techniques to predict annual yield of major crops and recommend



planting different crops in different districts in Bangladesh”, 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 1-6, 2015.

12. A.B. Mankar and M.S. Burange, “Data Mining-an evolutionary view of agriculture”, International Journal of Application or Innovation in Engineering and Management, Vol 3, No 2 pp .102-105, March 2014

13. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, “Analysis of agriculture data using data mining techniques: plication of big data”, Jouranl of Big Day, Springer Open.