Estimation of Major Agricultural Crop with Effective Yield Prediction using Data Mining

Rajesh Kumar Maurya, Sanjay Kumar Yadav, Tarun Kumar Sharma

Abstract: ***Agriculture is an area where the at most uncertainty exist. Crop production is totally depended on number of factors like geography, weather, biological, political, and economical. Indian financial system is mainly driven by farming and provides service opportunity in farming sector of Indian financial system is higher than world's average (6.4%). Agriculture (17.32%), Services (53.66%) and Industries (29.02%) are mainly contributing in the GDP of the nation.. Apart from this, a massive amount of raw agricultural data is present, but analysis these facts are very complicated for yield estimation of crop. So the most difficult task is to bring meaningful information and knowledge out of the raw agricultural data. Data mining can tailor data knowledge to estimate the crop yield. The aim of this research paper is to estimate crop yield by implementing data mining techniques.***

Keywords- Statistical Approaches, DM Technique, production estimation, tracking patterns, Classification, Cluster based Analysis, Linear regression Analysis (LRA), Multiple Regression Analysis (MRA) using SPSS.

I. INTRODUCTION

Indian agriculture recognized for different types of crop production, mainly due to change in resources. India is a country of the rural economy Crop yield projections and estimation are critical areas of research used to make sure food safety around the world and the Indian financial system. [1]. India is a global agricultural force. Each year we are getting a high production of yield from preceding year. This more production in successive years totally depends on impact of various crop factors, efficiency of estimation and forecasting technique. Crop forecasting and estimation are very important for developing new government policies that actually give impetus to crop production. The country is cultivate about 63% is treated with rain, while 37% is irrigated [www.worldbank.org].

II. LITERATURE REVIEW

Extraction from data is the process of obtain utilitarian facts from big amount of data. Aim of any extraction process is to drive knowledge from the existing dataset. The existing data set and transform it into a unique human-understood form for some pre-use.

Revised Manuscript Received on May 05, 2019.

Rajesh Kumar Maurya, Department of Computer Science and Information Technology Sam Higginbottom University of Agriculture, Technology and Sciences Allahabad, India,

Sanjay Kumar Yadav, Department of Computer Science and Information Technology Sam Higginbottom University of Agriculture, Technology and Sciences Allahabad, India,

Tarun Kumar Sharma, Assistant Professor, ABES Engineering College, Ghaziabad (Uttar Pradesh)

There is a need of effective tools and techniques that can be developed and tailored to solve complex agricultural problems by the use of data extraction [2].

For the forecasting and analysis of agriculture crop production different kind of data exploration techniques used. Non-use technologies also addressed the researchers for the issue of bi-clustering. [3].Different statistical techniques quantify the major factor that effects the agriculture production. These statistical techniques are very much useful to convert the important information received from the farmer and convert it into the knowledge which is required by the government to form the better policy form that increases the agriculture production[4][5]. This paper provides information and effective tools about data mining. Data extraction methods are ANN and decision tree whereas most efficient forecasting analysis tools used in agriculture data are KNN, K-Means, and SVM.ANN and SVM are the most widely used methods to explore the useful data extensively in the area of agriculture [6] [7].

III. CROP PRODUCTION AND FORECASTING USING DM TECHNIQUE

It is easy to set up relationship among the tedious process by using data mining techniques. DMT simplify the rule to classify and analyze the different agricultural data set to forecast the agriculture crop production. In last few years several DM techniques applied for analyzing and estimating the crop production. Nature of soil, water level and fertility gradient are the various agriculture factor influencing DM techniques results [8] [9] [10].

IV. DM AND STATISTICAL APPROACHES FOR PRODUCTION ESTIMATION

a) Tracking patterns and Classification of Agricultural data
b) Cluster Based Analysis(CBA)
c) Association Rules for yield Estimation(ARYE)
d) ANN for yield Estimation of Agricultural Crop.
e) Support Vector Machine(SVM)

a) Tracking patterns and Classification of Agricultural data

Data extraction is learning to recognize patterns in data set. This is usually to identify some of the intervals in your data that occur at regular intervals, or over time there is a flow and flow of a particular variable.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Tracking and Classification methods are capable of handling a massive volume of data in data extraction. Classification is the method of data extraction to predict the category of agricultural data. [11][12].

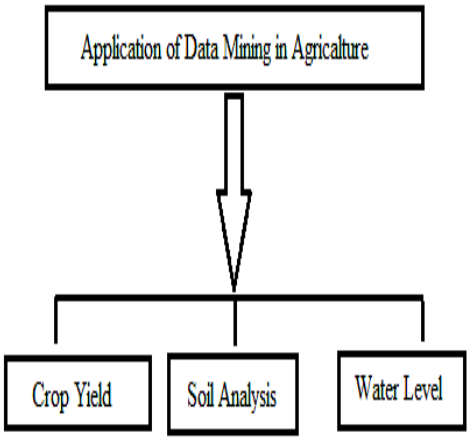

Fig. 1 Data Mining Application

The aim of tracking and classification techniques is to rigorously estimate the target class for each case in the data set of agriculture crop. Classification model could be used to find agricultural crop yield prediction as low, medium, or high risk factors. Classification constructs a form used to predict agricultural crop segment labels to differentiate objects from different categories of anonymous objects [13][14].

b) Cluster based Analysis(CBA)

CBA is most useful analysis to evolve the clusters of similar kind of crop with various type of parameter. Blocks are basically sub categories of agricultural data. Users understand the basic structure of data set. They are used as a standalone tool to gain insight into the distribution of data processing algorithms. Clustering is a group that consist same type objects. Simplification is achieved by representing the data in fewer clusters that require fine details [15] [16].

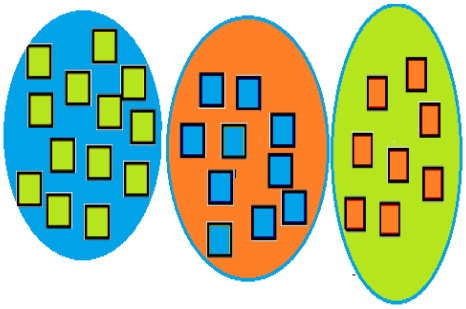

Fig. 2 Cluster Analysis of Agricultural Crop

c) Association Rules for yield Estimation (ARYE)

ARYE has a large number of applications and it is widely used to help find outcrop yield estimation and correlations in agricultural datasets [17].An association rule mining describe how frequently events have occurred together. With the use of Mining Association rule we can identify some interesting interconnection between different varieties in a large amount of agricultural crop production data [18].

d) Artificial Neural Network (ANN)

ANN is very useful algorithm in forecasting agricultural crop yield using different crop performance factors Biological nerve organization is the basis of ANN. Agriculture, farming method is very multifarious since it deals with the large data situation which comes from large number of factors. It consists of set of interconnected neurons. Connection between neurons is links and having weight associated with it [19].

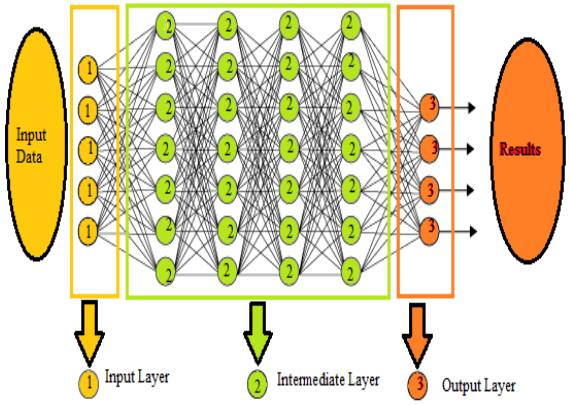

Fig. 3 ANN with 5 Input and 4 –Output Results

(e) Support Vector Machine

SVM algorithm is help to forecast the category of crop based on number of factor of crop and soil of crop yield estimation

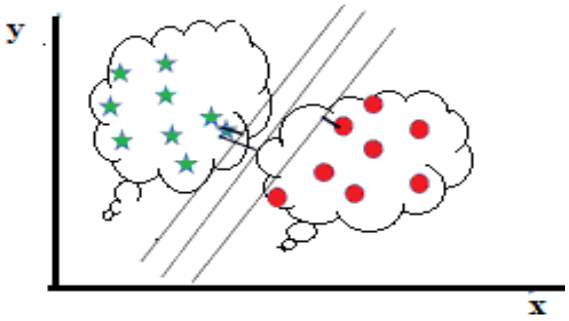

Fig. 4 Support Vector Machine

SVM is the specific discriminatory selection of various factor of crop by the excellent interval. SVM is an automated learning algorithm under supervision that can also be used for classification or regression challenges [20].

V. MULTIPLE LINEAR REGRESSIONS (MLR)

LRA expression is derived from the statistical regression model based on which the agricultural crop yield is estimated. Basically MLR is used to find linear relationship between a dependent variable and one or more independent variables. For formulation consider a polynomial of the nth degree [21][22].

$$y = a_1 + a_2x^1 + a_3x^2 + a_4x^3 + ---+a_nx^n$$

VI. EXPERIMENTAL DATA

Data refers to land use statistics, which is an important element for planning and developing policy formulation in agriculture. The Agricultural data set used for this kind of study is freely available at the website of directorate of economics & statistics of India [www.data.gov.in].For this study, we have selected Crop (Arhar, Barley, Maize, Potato).We have collected data from year 1997 to year 2013 that cover the different parameter of agricultural crop production that are Year, Season, Land area, production, yield, area under irrigation.

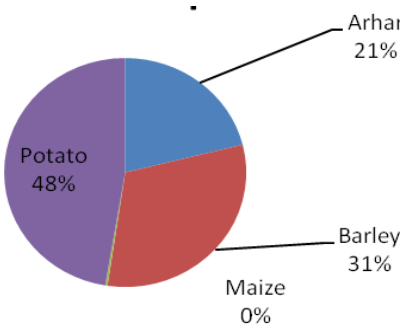

Fig. 5 Production of four crops tons

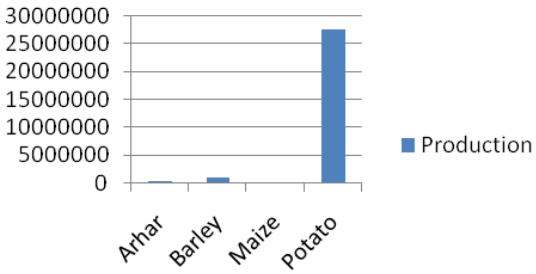

Fig. 6 Production of four crops in tons

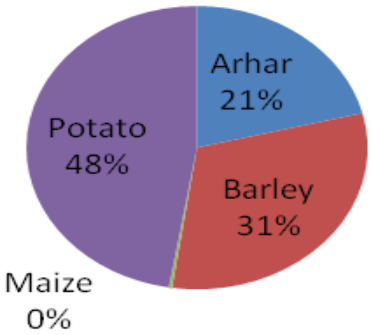

Fig. 7 Area of four crops

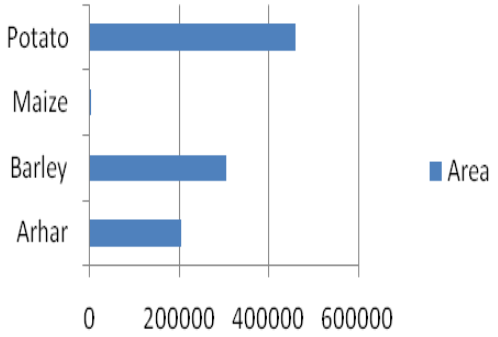

Fig. 8 Area of four crops in acre

a) Development Environment for Analysis

All agricultural experimental data set were analyzed using SSPS library. IBM SPSS offers variety of data, text analysis in research environment that has the good collection of statistical algorithms for Classification, processing and association rules.

Table. 1 ANOVA -Single Factor Table

ANOVA-SINGLE FACTOR				
OUTPUT SUMMARY				
Groups	*Count*	*Sum*	*Average*	*Variance*
Area	381	2907502	7631.238845	376915548.7
Production	379	66825459.1	176320.4726	5.34887E+11

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table. 2 ANOVA –2 factors with replication

ANOVA 2-Factor with Replication						
SUMMARY	Crop	Area	Production	Total		
Arhar						
Count	31	31	31	93		
Sum	61942	205138	482283	749363		
Average	1998.129032	6617.354839	15557.51613	8057.666667		
Variance	1.049462366	369150067.6	3071471089	1153966109		
Barley						
Count	31	31	31	93		
Sum	61963	304072	1017662	1383697		
Average	1998.806452	9808.774194	32827.80645	14878.46237		
Variance	1.827956989	570056143.7	6943005762	2623028373		
Maize						
Count	31	31	31	93		
Sum	62097	3011	6130	71238		
Average	2003.129032	97.12903226	197.7419355	766		
Variance	18.44946237	6787.916129	203586.7978	843873.9565		
Potato						
Count	31	31	31	93		
Sum	62048	460515	27712238	28234801		
Average	2001.548387	14855.32258	893943.1613	303600.0108		
Variance	8.189247312	634608818.6	2.64037E+12	1.03737E+12		
Total						
Count	124	124	124			
Sum	248050	972736	29218313			
Average	2000.403226	7844.645161	235631.5565			
Variance	11.34828219	412726192.2	7.92201E+11			
ANOVA						
Source of Variation	**SS**	**Df**	**MS**	**F**	**P-value**	**F crit**
Sample	6.10807E+12	3	2.03602E+12	9.212930026	6.91752E-06	2.62970693
Columns	4.40219E+12	2	2.20109E+12	9.959868133	6.16476E-05	3.02080007
Interaction	1.18248E+13	6	1.9708E+12	8.91779748	4.36918E-09	2.123781055
Within	7.95587E+13	360	2.20996E+11			
Total	1.01894E+14	371				

Year wise data of four major crops of Ghaziabad district UP, has been taken fromthe website of directorate of economics & statistics of India (www.data.gov.in).

Arhar(21%)

Barley(31%)

Malze(0%)

Potato(31%

Table. 3 Regression Statistics of 123 Observation

Regression Statistics	
Multiple R	0.679370042
R Square	0.461543653
Adjusted R Square	0.457093601
Standard Error	658308.848
Observation	123
Intercept	Coefficients
2071	2638.126391
	29.76553853

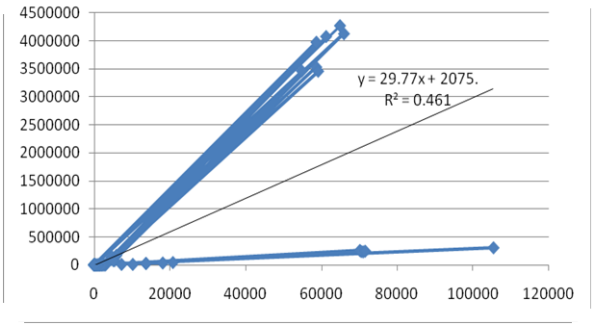

Fig. 8 Regression Line (Areavs Production)

Table. 4Regression Output(ANOVA)

ANOVA			
	Regression	**Residual**	**Total**
df	**1**	**121**	**122**
SS	**4.49477E+13**	**5.24378E+13**	**9.73855E+13**
MS	**4.49477E+13**	**4.33371E+11**	
F	**103.7164525**		
Significance F	**5.72956E-18**		

VII. CONCLUSION

To analyze and forecast the crop production, statistical model technique MLR applied on the data set having 382 records of Ghaziabad district Uttar Pradesh with different parameter such years, season, crops, area and production. In this study, four major crops production of different year and different season (rabi, kharif and summer) were compared and analyzed. For analyzing the data ANOVA Single-Factor and Two- Factor with replication based study has been conducted. Similar process can adopted for other crops to authenticate the validity of yield prediction. It has been observed that of more effective techniques that can be developed to find the solution of complex agricultural issue using DM techniques. The obtained results were verified and analyzed through statistical software IBM SPSS package.

REFERENCES

1. Fathima, G. Nasrin, and R. Geetha. "Agriculture crop pattern using data mining techniques." International Journal of Advanced Research in Computer Science and Software Engineering 4.5(2014): 781-6.ASA
2. Veenadhari, S., Dr Bharat Misra, and Dr CD Singh. "Data mining Techniques for Predicting Crop Productivity–A review article." International Journal of Computer Science and Technology IJCST 2.1(2011).
3. Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.
4. Djodiltachoumy, S. "Analysis of Data Mining Techniques for Agriculture Data." International Communications 4.2 (2016): 1311-1313.WE
5. Mucherino, Antonio, PetraqPapajorgji, and Panos M. Pardalos. "A survey of data mining techniques applied to agriculture." Operational Research 9.2 (2009): 121-140.
6. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ...& Zhou, Z. H. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37.
7. Hand, D. J. (2006). Data Mining. Encyclopedia of Environmetrics, 2.
8. Ramesh, D., &Vardhan, B. V. (2015). Analysis of crop yield prediction using data mining techniques. International Journal of Research in Engineering and Technology, 4(1), 47-473.
9. Ahamed, AT M. Shakil, et al. "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh." Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACIS International Conference on. IEEE, 2015.
10. Sangeetha, A., and M. Ravichandran. "A Survey on Data Mining Approaches to Handle Agricultural Data." Data Mining and Knowledge Engineering 8.9 (2016): 286-290.
11. López, C. J. Á., Valiño, J. A. R., & Pérez, M. M. (2008). Typology, classification and characterization of farms for agricultural production planning. Spanish Journal of Agricultural Research, (1), 125-136.
12. Sastri, A. S. R. A. S., Rao, B. R., Krishna, Y. R., & Rao, G. G. S. N. (1982). Agricultural droughts and crop production in the Indian arid zone. Archives for meteorology, geophysics, and bioclimatology, Series B, 31(4), 405-411.
13. Allen, R. C. (1999). Tracking the agricultural revolution in England. Economic history review, 209-235.
14. Wackernagel, M., Schulz, N. B., Deumling, D., Linares, A. C., Jenkins, M., Kapos, V., ...& Randers, J. (2002). Tracking the ecological overshoot of the human economy. Proceedings of the national Academy of Sciences, 99(14), 9266-9271.
15. Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research, 38(suppl_2), W64-W70.
16. Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. Information sciences, 222, 175-184.
17. Gandhi, N., & Armstrong, L. J. (2016, December). A review of the application of data mining techniques for decision making in agriculture. In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1-6). IEEE.
18. Sujatha, R., &Isakki, P. (2016, January). A study on crop yield forecasting using classification techniques. In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-4). IEEE.
19. Berson, A., Smith, S., &Thearling, K. (2000). Building data mining applications for CRM (pp. 4-14). New York: McGraw-Hill.
20. Mucherino, A., Papajorgji, P., &Pardalos, P. M. (2009). A survey of data mining techniques applied to agriculture. Operational Research, 9(2), 121-140.
21. Berson, A., Smith, S., &Thearling, K. (2000). Building data mining applications for CRM (pp. 4-14). New York: McGraw-Hill.
22. Van den Putte, A., Govers, G., Diels, J., Gillijns, K., &Demuzere, M. (2010). Assessing the effect of soil tillage on crop growth: a meta-regression analysis on European crop yields under conservation agriculture. European journal of agronomy, 33(3), 231-241.