2021 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2021

Effective use of Big Data in Precision Agriculture

Sharayu Ashishkumar Lokhande
Department of Computer Engineering
Army Institute of Technology
sharayulokhande@gmail.com

2021 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-7281-8519-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ESCI50559.2021.9396813

***Abstract* –Precision Agriculture is the key terminology in agriculture Engineering. Precision agriculture can make the use of legacy data of agriculture to make the farming better in terms of quantity and quality. To enhance the production of the agriculture, technologies such as big data analytics along with data mining tool can use the legacy agricultural data to make the future prediction. This prediction can help to enhance the Agro-Economy.**

Index Terms – Big Data Analytics, data analysis, Seasons, Agriculture crops data, Linear Regression, Multiple Linear Regression.

I. Introduction

Agriculture plays a pre-eminent part in Indian economy. Over 58% of the rural economy depend on agriculture as their principal means of income. Agriculture, along with fisheries and forestry, is one of the largest suppliers to the Gross Domestic Product. The stake of agriculture in the GDP in 2012 was 18% and in employment 50% [1]. Indian farmers are not getting estimated yield of crops due to various factors like changes in temperature, rainfall, soil condition, pest and diseases etc. There is strong need in India to predict the crop production against the ecological and soli factors. In recent years several efforts had made to developed simulation or predictive models using the data for atmospheric parameters and crop performance.

A. Agriculture in India

Agriculture is the primary source of Indian economy. It servers the 75% employment to the population such as food and clothe. It also generates the employment to all those who supplies the necessities for the farming, it includes employments, equipments, pesticides, transportation. It helps for economic growth. In the early periods of green revolution helps to increase in crop yield production. But due to the lack of technical knowledge it can't be increased beyond certain level [2].

The increasing use of technology in the domain of agriculture helps to expand the production, quality of the crops.

Indian economy mainly based on agriculture, it offers direct employment to 2/3rd of our population and it is a provider of food, clothing and other basic provisions of life for the entire population. The agriculture sector is important for food security, employment generation and economic growth.

There is a constant degradation in agricultural growth, which is the worried factor. There is enhancement growth in food grain production in the early years of green revolution. But due to the lack of knowledge of science and digital information, the growth of food grain production cannot be increased beyond certain level [2]. Technology is growing faster and faster day by day. It has been penetrated vastly from urban to rural region. Government and farmer go hand in hand for the betterment of the crop production and its supporting small scale businesses based on the farming. Government has taken many initiatives by providing the agriculture apps to help the farmer. Mobile messaging services helps the farmer to get the information about their queries related to agricultural. It will provide the agro-vendor's information to farmers. It also provides the weather information, soil contents information, market near to them to supply the crop items [20].

B. Issues in Agriculture

Eventually, there is ecological impact on the production practices used by farmers. Some of the negative environmental impact includes: [3]

Weather: Uncertainty and non-uniformity in Indian weather affects badly on crop yield production. Prediction on weather certainly helps farmer to save their crops in terms of money and efforts. Weather prediction can be done on historical data. Crop selection based on the weather forecasting information [14].

Soil: Degradation of soil parameters due to excess use of pesticides like urea, endosulfan etc. If the soil parameters are known prior of sowing, this will help the farmer to take the decision about the crop selection [15].

Crop cutting: With the help of image sensing data, farmers can take the decision when the crop cutting needs to be done. Prior information helps to saves the cost of machinery or labor required for crop cutting [15].

Pests Management: Prior weather information and soil analysis data helps the farmer to use fewer amounts of pesticides. Lesser the amount of pesticide leads to more organic crop. It gives good profits to farmer [16].

Intercropping: Experts can suggest the farmer to use intercrops based on long term and short term crop production. It gives profit to farmers. This can be done by analyzing the pattern of crops based on legacy data for varying the crop [16].

C. Problem statement

Early research has done on few crops ,few regions ,few season , some of the parameters such as temperature, soil contents, weather conditions ,raining situation etc. This may help to the respective regions. If the study has been enhance further for many regions, maximum seasons, crops which may affect directly to the farmers economy, number of crops, more parameter, this in term benefits more number of people. This may help to reduce food grain problem. [19]. If the span of the

data is increased more for the experimentation then the predicted values can be more accurate. Recent data will be used along with the big data analytics tools and methodology. It helps to improve classification and prediction of data.

D. Proposed solution

The proposed system can be categorized into two phases construction phase and operational phase. In the construction phase the preprocessing of the data has been done. In this phase the data can be categorize into training data and test data. In the operational phase decision tree generator done on the training data.

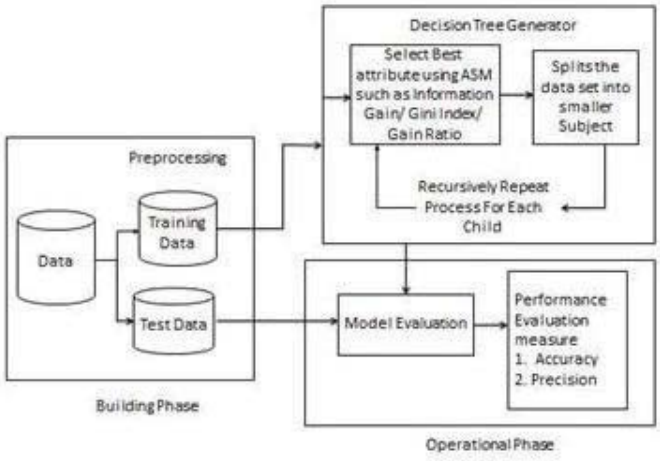


Fig. 1. Proposed System

Fig. 1 illustrates the process. The data has transfer from construction phase to operational phase to make the prediction. Here the research objective is focuses on operational phase which contains data mining algorithms. Proposed system is categorized into two phases: Construction phase and Operational phase. Preprocessing on data has done in construction phase. In the operational phase decision tree generator algorithm has been applied. This algorithm works on attribute selection measures (ASM) such as information gain, Gini index and Gini ratio. It splits the data into smaller subjects. Recursively the process has repeats till the each child.

Big data frame work allows collecting the huge volume of varieties of data. It processed the huge data required to the domain application [12]. In the application the required data such as weather data, soil analysis data, and production of the different crops per tons in different seasons has been considered. The framework processed varieties of data coming from different sources. The variety of data may contain structure, unstructured, documents, column data and images [13]. Various decisions can be taken in advance by processing Tera and Peta bytes of data which helps farmer in saving efforts, costs and increases the yield productivity.

II. LITERATURE REVIEW

A. Big data

Big data is not the buzzword in today's circumstances. Big data helps to support massive data, varieties of data different types of data such as structured, unstructured, document based, column basted data etc. [4] .

Volume: Real data is very densed. The size of the real data has been increasing exponentially up to Exabyte. Every day daily logs has been created and stored. It can be double by every year [5].

Velocity: The speed by which the data is generated is tremendous. The rapidly generated data need to store. As per the demand it has to be processed to reach to the conclusion [6].

Variety: Different variety of data generated rapidly. The variety of data contains text data, structured, unstructured, audio, video etc. Big data supports the variety of data coming from different sources like daily log, online data, and web data [7].

III. PRECISION AGRICULTURE AND BIG DATA IN INDIA

India is an agriculture centric country. Indian economy is purely based on agriculture. Rapidly increasing population and decreasing texture of the land become serious issue in the coming years. To satisfy the requirement agricultural production need to be increased in the proportion of the population. In other words, we will have to learn to produce more with less. Agriculture has already seen two major revolutions. The first, at the time of the industrial revolution when production was mechanized and the second was the green revolution when pesticides and other agrochemicals were widely adopted. Now there is the third revolution based on the Big Data [18].

The use of Big Data in agriculture thus has the potential to produce more with less. The solution is precision agriculture. It is nothing but the combination of big data and data mining technology applied on the agriculture data. When the technology used along with the agriculture data, it in turn helps to increase the crop yield production despite of the factors affecting on the agriculture such as weather, soil content, pests, etc. Actual practices need to be done on small area to the large area. The results may help the farmers to adopt the new technologies. Based on the prior knowledge the field has been divided into sub units depending on the soil texture, contour of the land and water level [8]. Based on the variability of the land sub division different seed to be sowed at different sub division .This will helps to improve the crop yield production.

IV. MODULE DESCRIPTION

Following are the modules can be used in research process:

A. Dataset collection

Analysis has done on the following dataset:

Crops versus Seasonal Production Crops versus Temperature

In proposed work, agriculture data has been collected from following sources:

Dataset in agricultural sector [https://data.gov.in/]

B. *Data preprocessing*

There are different varieties of data sources available. These may contain noise, redundancy, and consistency, etc. Unnecessarily it may increase the storage for useless data [17]. There may be the requirement of high data quality. For effective data analysis, preprocessing on data is required at most. Big data analytics tools such as Weka [12] or R [10] can be used to perform above operations on data.

In the research WEKA tool has been used for preprocessing of data. Preprocessing of data can be performed on the following way, firstly it split the data into train and test data, and secondly it trains the classifier. Big data analytics tools such as WEKA can be used to perform above operations on data.

C. *Decision tree generation*

This is the next module of the research. In this module Attribute Selection Measure (ASM) is used. It is a heuristic for selecting the splitting criterion that partition data into the best possible manner. It uses selection measures such as Information Gain, Gain Ratio, and Gini Index. This process recursively repeats for each child [11].

D. *Data classification and clustering*

Data classification is the process (or) procedure of sorting and categorizing data into various types, forms which helps to analyze the data. Data classification helps in the prediction of the data. Decision tree classification is a supervised learning algorithm. It can solve both regression and classification problems. This algorithm works on attribute selection measures (ASM) such as information gain, Gini index and Gini ratio. It splits the data into smaller subjects. Recursively the process has repeats till the each child.

E. *Methodology*

In this paper the Linear Regression and Multiple Linear Regression technique of predictive analysis has been used for crop yield prediction.

1) *Linear Regression*

TABLE I. INPUT FOR LINEAR REGRESSION

Input is taken as follows

Model Used	Linear Regression
Instances	102
Attribute	13
Classifier mode	Linear Regression with full training set
Test mode	10-fold-cross-validation
Dataset used	Precipitation monthly year wise
Sample Duration	Year 1901-2002

TABLE II. CLASSIFIER MODEL FOR FOLD -10

Classifier Model for fold -10

Instance	Actual data	Predicted data	Error
1	0.016	8.329	8.313
2	24.682	17.008	-7.674
3	0.824	7.327	6.503
4	0.016	8.237	8.221
5	49.391	8.092	-41.229
6	0.441	7.327	6.886
7	39.305	7.443	-31.862
8	0.125	8.513	8.388
9	19.298	8.548	-10.75
10	0.016	7.327	7.311

TABLE III. CROSS VALIDATION SUMMARY

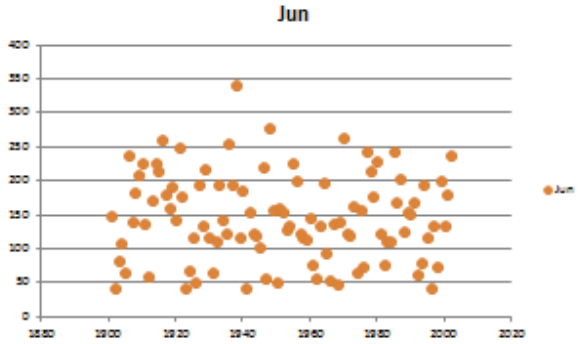
Cross Validation Summary

Correlation coefficient	-0.0982
Mean absolute error	12.3804
Root means squared error	17.9879
Relative absolute Error	101.9493 %
Root relative squared error	102.8813 %
Total number of instances	102

Correlation coefficient =-0.09

It is Significant because it lies in between +0.8 to -0.8

Visual Representation of : Precipitation Data
Model : Linear Regression
Month of Precipitation : June



Precipitation for the month of June is ranges from 42 to 340%
Which is a good condition to crop sowing.

Fig. 2. Visual Representation of Linear Regression

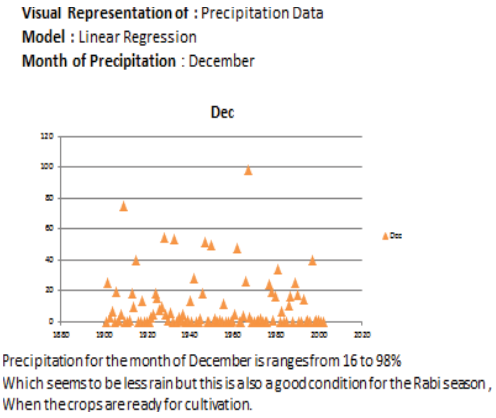


Fig. 3. Model of Linear Regression for Precipitation Data

2) Multiple Linear Regressions

Multiple linear regression (MLR) [9] is a statistical technique used to predict the outcome of a dependent variable from more than one independent variable. Multiple linear regression (MLR) basically used to find the relation between the independent and dependent variables.

The model for MLR, given n observations, is:

$yi = B0 + B1xi1 + B2xi2 + ... + Bpxip + E$ where $i = 1,2, ..., n$

In the crop yield prediction model, the use of Multiple Linear Regression technique is used to predict the production which is dependent variable using more than one independent variables namely Year, Area, Temperature, soil content, Yield (ph, Nitrogen, Phosphorus, potassium, Water)

3) Result Analysis

In this paper Multiple Linear Regression technique has been used to predict the year wise production of crop Jowar of kharif season. Table 1 displayed the sample data for year from 1997 to 2016 for the crop Jowar and its respective estimated value for the respective year.

TABLE IV. ESTIMATED YEAR WISE JOWAR PRODUCTION USING MLR

Year	**Production** (Tonnes)	***Predicted Production*** *(Tonnes)*
1997	290100	226927
1998	244000	215943
1999	133500	204959
2000	142500	193975
2001	121100	182992
2002	162500	172008
2003	161300	161024
2004	108300	150040
2005	184700	139056
2006	205800	128072
2007	210400	117088
2008	102800	106104
2009	85200	95120
2010	39400	84136
2011	124300	73152
2012	71800	62168
2013	21800	51185
2014	15100	40201
2015	10300	29217
2016	16700	18233

TABLE V. INPUT FOR MLR

Model Used	Multiple Linear Regression
Instances	20
Attribute	2
Classifier mode	Multiple Linear Regression
Dataset used	year wise production of crop Jowar of kharif season
Sample Duration	Year 1997-2016
Data taken for	district of Akola in Maharashtra
Estimated value Ranges	-2 % to +2%.

The comparison of real value of Jowar for Kharif season from the year 1997 to 2016 with the estimated value using Multiple Linear Regression (MLR) technique has been shown in figure 4. Regression statics has been shows in the below figure 5. Here R square values come as 0.65 and standard error as 48133.34. After the prediction, the production has been increased.

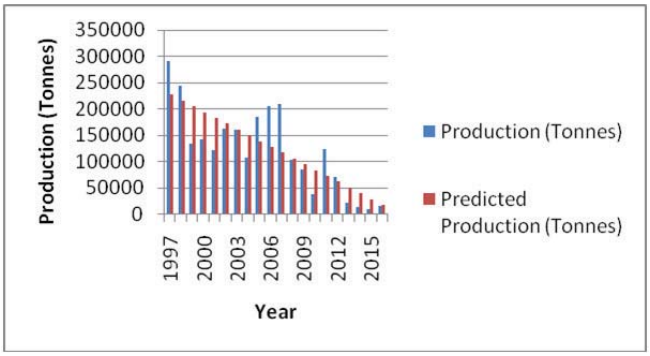


Fig. 4. Estimated values of MLR

Regression Statistics	
Multiple R	0.811163
R Square	0.657985
Adjusted R Square	0.638984
Standard Error	48133.34
Observations	20

Fig. 5. Regression Statics

V. CONCLUSION

Big Data is changing the scope of the Indian agriculture from traditional to digital. Big data definitely helps to solve the food security issues which the world is going to face in the future. Big data along with the data mining algorithms may be the solution provider for the global problem in the near future. Analytics is key success factor to create value out of these data. This paper briefly describes difference between the real values of Jowar with the estimated value using Multiple Linear

regression technique. In future it can be done for more crops using data mining technique like decision tree classifier with its algorithm Id3, C4.5 and CART.

REFERENCES

[1] Guruswamy, Mohan. “Food for Thought: India Will Have 1.7 Billion People by 2050. How Will It Feed Them All?” Scroll.in, Scroll.in, 29 May 2017, scroll.in/article/838970/food-for-thought-india-will-have-1-7-billion- people-by-2050-how-will-it-feed-them-all.

[2] CHAPTER - I 1.1. INTRODUCTION - Shodhganga. (n.d.). Retrieved from http://shodhganga.inflibnet.ac.in/bitstream/10603/15715/7/07_chapter 1.pdf

[3] “Environmental Effects of Agriculture.” Wikipedia, Wikimedia Foundation,16Mar.2011,en.wikipedia.org/wiki/Environmental_effects_o f_agr iculture..

[4] Mayer-Sch¨onberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Eamon Dolan/Houghton Mifflin Harcourt

[5] Tekiner, F.; Keane, J.A., "Big Data Framework," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on , vol., no., pp.1494, 1499, 13-16 Oct. 2013.

[6] Thomas Davenport, 2013. Text Book. Big Data at Work: Dispelling the Myths, Uncovering the Opportunities.

[7] Min Chen, Shiwen Mao, Yunhao Liu, “Big Data: A Survey”, Springer Science Business Media New York 2014.

[8] Data Collection and Analysis Methods for Data from Field ... (n.d.). Retrieved from http://www.eolss.net/sample-chapters/c10/E5-17-05-02.pdf.

[9] Proceedings of the 2nd International Conference on Computer Science and Application Engineering, ACM, dl.acm.org/citation.cfm?id=3207677. [12] https://www.cs.waikato.ac.nz/ml/weka/ (Access date 12 December 2017).

[10] https://www.r-project.org/ (Access Date 5 Nov 2017).

[11] "Incremental Induction of Decision Trees", by Paul E. Utgoff, Kluwer Academic Publishers, 1989

[12] Tekiner, F.; Keane, J.A., "Big Data Framework," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on , vol., no., pp.1494, 1499, 13-16 Oct. 2013.

[13] Chan, J. O. (2013). An Architecture for Big Data Analytics. Communications of the IIMA, Volume 13 | Issue 2, USA.

[14] Klassen, Jerry. “Weather Effects Could Have Large Impact on U.S. Crop Levels and Carryout.” Farmtario, 5 July 2019, farmtario.com/markets-business/markets/weather-effects-could-have-large-impact-on-u-s-crop-levels- and-carryout/.

[15] (PDF) Cereal Crop Production | T. N. Bhusal - Academia.edu. www.academia.edu/19211413/Cereal_crop_production.

[16] Gummadapu, Damodar Rao. “Leveraging Big Data in Agriculture.” Iamwire, 16 May 2017, www.iamwire.com/ 2017/05 /leveraging-big-data-in-agriculture/ 152495.

[17] S.Shibusawa and C.Haché. ―Data Collection and Analysis methods for Data from Field Experiments‖. Encyclopedia of Life Support Systems. ISBN : 978-1-84826-583-7

[18] Tag Archives: Green Revolution. (n.d.). Retrieved from https://actnaturallyblog.wordpress.com/tag/green-revolution/

[19] Soil Data Analysis Using Classification Techniques and ... (n.d.). Retrieved from https://www. researchgate.net/ publication/ 225279640_ Soil_Data_Analysis_Using_Classification_Techniques_and_Soil_AttributePre diction.

[20] ..., I. I. (n.d.). Study of Improvement in Efficiency and Effectiveness of Performing Farming using Internet of Things. Retrieved from https://www.academia.edu/ 35576110/ Study_of_Improvement_ in_Efficiency_and_ Effectiveness_ of_ Performing_ Farming_using_ Internet_of_ Things.