Literature Survey

Estimate the Crop Yield Using Data Analytics

ABSTRACT :

Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics. Efforts are going on to understand how big data analytics can agriculture productivity. The present study gives insights on various data analytics methods applied to crop yield prediction and also signifies the important lacunae points in the proposed area of research.

INTRODUCTION :

Agriculture forms the basis for food security and hence it is important. In India, majority of the population i.e., above 55% is dependent on agriculture as per the recent information. Agriculture is the field that enables the farmers to grow ideal crops in accordance with the environmental balance. In India, wheat and rice are the major grown crops along with sugarcane, potatoes, oil seeds etc. Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture. This domain provides employment to more than 60% of the total population and has a contribution to GDP also (about 17%). In the farm output, India ranks second considering the worldwide scenario. This is the widest economic sector and has an important role regarding the framework of socio-economic fabric of India. Farming depends on various factors like climate and economic factors like temperature, irrigation, cultivation, soil, rain fall, pesticide and fertilizers. Historical information regarding crop yield provides major input for companies engaged in this domain. These companies make use of agriculture products as raw materials, animal feed, paper production and so on. The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield. Farmer experiences was the only way for prediction of crop yield in the past days. Technology penetration into agriculture field has led to automation of the activities like yield estimation, crop health monitoring etc. Crop yield prediction has generated a lot interest in the research community and also for agriculture related organizations. Crop yield prediction helps the farmers in various ways by providing the record of previous crop yield. This is helpful to government in framing policies related to crops such as crop insurance policies, supply chain operation policies. Knowing what crops has been grown, and how much area of it had been shown historically, combined with the prices at which it could have been sold at the nearest market-place provides the income-growth profile of the farmer.

Agriculture sector is struggling to increase the productivity of crop in India. Monsoon rainfall is the main source of water for more than 60 percent of the crops. Smart agriculture driven by Information Technology is the emerging trend in the research in this area in recent days. One of the areas being explored is the problem of yield prediction which is a major concern. Data mining techniques are being widely used as a part of solution for crop yield prediction. Various data mining techniques are under evaluation for estimation of crop production of the future years. Data mining is the process in which the hidden patterns are discovered using analysis of large data sets. The data mining and data analytics techniques use artificial intelligence, statistics, machine learning and database system. In data mining, unsupervised and supervised methods are being used. In unsupervised learning, clusters are formed using large data sets and in supervised learning classification are done based on the data sets. In clustering technique, ‘data points’ are examined to group them into ‘clusters’ according to specific parameter. The data points in same cluster have less distance compared to data points of different clusters. The analysis of

the cluster divides data into well organized groups. The natural structure of the data is captured by these well-formed groups [3, 4].

REVIEW OF LITERATURE METHODS OF CROP YIELD PREDICTION :

At present we are at the immense need of another green revolution to supply the food demand of growing population. With the decrease of available cultivable land globally and the decreased cultivable water resources, it is almost impossible to report higher crop yield. Agriculture based big data analytics is one approach, believed to have a significant role and positive impact on the increase of crop yield by providing the optimum condition for the plant growth and decreasing the yield gaps and the crop damage and wastage. With this aim the present paper reviews about the various advances, design models, software tools and algorithms applied in the prediction assessment and estimation of the crop yield. India is basically agriculture based country and approximately 70% our country economics is directly or indirectly related to the agricultural crops. The principle of crop which occupies the highest (60-70%) percentage of cultivable land in the Indian soil is the paddy culture and it is the major crop especially in central and south parts of the India. Rice crop cultivation plays an imperative part in sustenance security of India, contributing over 40% to general yield generation. The enhanced yield of the rice crop depends largely on the water availability and climatic conditions. For example, low precipitation or temperature extremes can drastically diminish rice yield. Growing better strategies to foresee yield efficiency in a mixture of climatic conditions can help to understand the role of different principles factors that influence the rice crop yield. Big data analytic methods related to the rice crop yield prediction and estimation will certainly support the farmers to understand the optimum condition of the significant factors for the rice crop yield, hence can achieve higher crop yield.

CROP YIELD PREDICTION USING DATA ANALYTICS :

In India crop yield is season dependent and majorly influenced by the biological and economic causes of an individual crop. Reporting of progressive agricultural yield in all the seasons is an ample task and an advantageous task for every nation with respect to assesses the overall crop yield prediction and estimation. At present a common issue worldwide is, farmers are stressed in producing higher crop yield due to the influence of unpredictable climatic changes and significant reduction of water resource worldwide. A study was carried out to collect the data on world climatic changes and the available water resources which can be used to encourage advanced and novel approaches such as big data analytics to retrieve the information of the previous results to the crop yield prediction and estimation. Study imported that the selection and usage of the most desirable crop according to the existing conditions, support to achieve the higher and enhanced crop yield .

The accurate prediction of crop yield certainly benefits the farmers in choosing the right method to reduce the crop damage and gets best prices for their crops. A research group conducted a work with an objective of accurate prediction of crop yield through big data analytics to assess various crop yield influencing factors such as Area under Cultivation (AUC) interims of hectors, Annual Rainfall (AR) rates and Food Price Index (FPI) and to develop relationship among these parameters. Regression Analysis (RA) methodology was applied to examine the selected factors and their impact on crop prediction and final yield. RA methodology is a multivariable investigation practice which can categorize the factors in to groups such as explanatory and response variables and helps to assess their interaction to obtain a resolution. All the selected factors of the present study design known as AR, AUC and FPI were measured for a period of 10 years between the years 1990-2000. A novel method called Linear Regression (LR) is applied to analyse the relationship between explanatory variables (AR, AUC, FPI) and the crop yield considered as response variable. Study reported that the R_2 value for the studied factors clearly indicate that crop yield is principally depends on AR. Study also reported that the other two factors (AUC and FPI) screened were also found to have significant impact after the AR. Study shall be continued to analyse the impact of for other substantial factors like Minimum Support Price (MSP), Cost Price Index (CPI), Wholesale Price Index (WPI) etc. and their relationship on the yields of different crops.

Crop yield gaps, measured as difference between expected yields based on the potency and actual farm yield received. In order to achieve the higher crop yield, farmers must need to tackle the influencing factors such as influence of change in climate conditions on the prospects of crop yields, and change in the usage of agricultural land to assess and ultimately reduce the crop yield gaps. Several researchers reported the applications of bio simulation models to estimate the crop yield gaps in the last decade. The impact of the crop yield gaps assessment studies conducted through bio simulation base methodologies were negatively influenced by quality and resolution of climate and soil data, as well as unscientifically expectations about crop yield prediction systems

and crop yield assessment modelling designs calibration method. An explicit rationale model which can effectively applied at various levels of the availability of quality information for identifying data sources to analyse crop yield and measuring yield gaps at definite geographical locations and works based on the rise in titer approach. The model is highly helpful in retrieving the useful data from the available, poor quality, less rigorous data sources or if the data is not available. A case study was discussed on the application of selected model design to quantify the yield gaps of maize crop in the state of Nebraska (USA), and also at the different geographical locations representing the nations Argentina and Kenya at national scale level. Different geographical locations such as Nebraska (USA), Argentina and Kenya were identified to symbolize the distinct scenarios of Agri based data availability and the quality for the selected variables assessed to predict and estimate the crop yield gaps. The definitive aspiration of the planned method is to afford transparent, easily accessible, reproducible and technically sound and strong guidelines for predicting the yield gaps. The proposed guidelines were also relevant for understanding and to simulate the influence of change in climate conditions and usage of cultivable land changes from national to global scales. As indicated, the better understanding of data importance and usefulness for analysing crop yield and estimating yield gaps as illustrated can help in identifying the data gaps in the crop yield and allow focusing on the various efforts taken at the global level to address the most critical issue.

Analysing the yields of crop is necessary to update the policies to ensure food security. A research group conducted a study with the aim in suggesting a novel data mining method to predict the yields of crop depends on agricultural big data analytics methodologies, which were progressively contrast with conventional data mining methodologies in the process of handling data and modeling designs. Study suggested that the method employed should be user friendly, work based on progressive big-data responsive processing structure, supposed to utilize the existing agricultural based significant datasets and would still be used with the larger volumes of data growing at enormous rates. Nearest neighbours modeling is one such novel data mining technique which works on the results collected based on data processing structures form the farmers and suggest a well unbiased result on the base of accuracy and prediction time in advance. Study further discussed a case study on the assessment of actual crop dataset (numerical examples on) in China from 1995-2014. Study reported that the novel model employed has publicized an improved performance and was found to be progressive in reporting prediction accuracy percentage of the compared methodologies with conventional designs.

Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical regions. Satellite retrieved data have frequently been revealed to present data sets that, by itself or in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands. The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction. A review was conducted to discuss the applications of remote sensing technology to determine the impact and causes of yield gaps. Even though the example discussed by the research group demonstrates the usefulness of remote sensing in the prediction of yield gaps, but also many areas of possible application with respect to the crop yield assessment, prediction and improvement remain unexplored. Study proposed two less complicated, easily assessable methods to determine and quantify the yield gaps between various agricultural fields. First method works closely with the constructive maps representing the average crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation.

In coming decades, two most significant and important factors found to influence crop yield is, increase in the global population and economy, which greatly demands the higher and sustainable agricultural based crop yields. The capacities of food production at global level is going to be very limited due to the less availability of cultivable land, water resources, difficulties in maintaining the sustainable crop production levels, effects of changes in the global climatic conditions and also by various biophysical parameters which influence the crop yield. The farmers need to be educated on the application of scientifically proven methods to quantify the crop yield capacities and same need to be informed to higher authorities to maintain transparency in sharing the actual information, intern helps in making the policy based, research oriented, development and investment related decisions that aim to influence future crop yield. Crop production abilities and yield gaps can be assessed and measured by comparing the possible yields at normal conditions with respect to the crop production under, respectively, irrigated and rain fed conditions by keeping the crop yield levels limited by the less availability of the water as benchmarks. Yield gaps can be defined as the difference between the expected crop yields with respect to the actual crop yield and accurate, spatially unambiguous awareness and information about the yield gaps is necessary to achieve sustainable amplification of agricultural yields. Keeping an aim of discussing the impact of the various methods practiced in measuring the yield gaps with a spotlight on the local-to-global importance of outcomes, a research group carried out a survey on the various methods applied to estimate yield gaps. Study reported few standard operation methods, employed in quantifying the crop yield potential on the data collected from the farmers of

western Kenya, Nebraska (USA) and Victoria (Australia). Study recommended for the use of accurate and recent yield data assessed through calibrated crop model designs and further up scaling validated methods in the prediction of crop yield gaps the bottom-up application of this global protocol allows verification of estimated yield gaps with on-farm data and experiments.

DIFFERENT DATA ANALYTICS TECHNIQUES APPLIED IN CROP YIELD PREDICTIONS

SL	Title of the paper	Authors / Year of publication	Highlights
1.	A Survey on Crop Yield Prediction based on Agricultural Data	Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R/ 2017	Presented a survey on the different algorithms applied in the assessment and prediction of crop yield Discussed about the mechanism of knowledge the discovery in Agricultural data mining
2.	Big Data for weed control and crop protection	F K Van Evert, S Fountas, D Jakovetic, V Crnojevic, I Travlos & C Kempenaar/ 2017	Critically discussed about the challenges faced and the profound opportunities lies in the Big Data analytics in agriculture: Outlined Big Data analytics models with numerical algorithms applied Represent the importance of reforming the mined data in the form of understandable information to the farmers. Discussed about various advances, tools and algorithms applied in transforming the data in to easily understandable information to the framers and thrown a light on success story of Netherlands in achieving the maximum crop yield and their smart forming practices. Also discussed about the control of invasive, parasitic and herbicide-resistant weeds to improve the overall crop yield applying Big Data analytics.
3.	The Impact of Data Analytics in Crop Management based on Weather Conditions	Swarupa Rani A/ 2017	Discussed the application of mathematical model like fuzzy logic designs in optimization of the crop yield, artificial neural networks in validation studies, genetic algorithms designs in accessing the fitness of the model applied, decision trees, and support vector machines to study soil, climate conditions and water regimes related to crop growth and pest management in agriculture.
4.	A Study on Crop Yield Forecasting Using Classification Techniques	R. Sujatha, Dr.P.Isakki Devi/ 2016	Discuss the importance of comparing previous agricultural data with present to identify optimum condition favor enhanced crop yield. Envisaged the importance of best crop selection depending on the season and the climatic factors which supports enhanced crop yield.

5.	Prediction of Crop Yield using Regression Analysis	V. Sellamand E. Poovammal/ 2016	Regression analysis was carried out to find the relationship among the parameters i.e Area under Cultivation (AUC), Annual Rainfall (AR) and Food Price Index (FPI) which influences the final crop yield and reported that the crop yield principally depends on the Annual Rainfall (AR).
6.	How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis	Patricio Grassinia, Lenny G.J. van Bussel, Justin Van Warta, Joost Wolf,Lieven Claessens, d, Haishun Yanga, Hendrik Boogaarde, Hugo de Groote,Martin K. van Ittersumb, Kenneth G. Cassman/ 2015	Presented a case study (Nebraska - USA and at a national scale for Argentina and Kenya) on the application of an explicit rationale design approach in identifying the data sources which simulates Crop (maize) yield and also helps in quantifying the maize yield gaps. Suggested the robust guidelines for analyzing the crop yield gaps, accessing the climate and land use changes at global level to address the issues of crop yield.
7.	Prediction of crop yield using big data	Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua Wang Juyun/ 2015	Developed a novel model i.e Nearest neighbours modeling to calculate and predict the yield of crop depends on the available Big data sets.
8.	The use of satellite data for crop yield gap analysis	David B. Lobell/ 2013	Discussed the use of remote sensing technology to identify and measure the causes of yield gaps and the assess the impact on the overall crop yield. Reported very simple methodologies to measure the yield difference with respect to season, environment and the land use.
9.	Yield gap analysis with local to global relevance-A review	Martin K. van Ittersuma, Kenneth G. Cassmanb, Patricio Grassinib, Joost Wolfa, Pablo Tittonell, Zvi Hochmand	Discussed about the various methods used on quantifying the yield gaps at local-to-global ratio. Reported few standard operation methods, employed in quantify the crop yield potential on the data collected from the farmers of western Kenya, Nebraska (USA) and Victoria (Australia). Study recommended for the use of accurate and current yield data, with calibrated and validated crop models and up scaling methods in the prediction of crop yield.

CONCLUSION :

As a result of penetration of technology into agriculture field, there is a marginal improvement in the productivity. The innovations have led to new concepts like digital agriculture, smart farming, precision agriculture etc. In the literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to climatic conditions and crop yields data. The activities of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction etc. In this survey, the specific activity, crop yield prediction has been surveyed and the major trends have been identified.

It can be concluded that the research in the field of agriculture with reference to using IT trends like data analytics is in its infancy. As the food is the basic need of humans, the requirement of getting the maximum yields using

optimal resource will become the necessity in near future as a result of growing population. The survey outcomes indicate the need for improved techniques in crop yield analytics. There exists a lot of research scope in this research area.

REFERENCES :

[1] D Ramesh, B Vishnu Vardhan. Data Mining Techniques and Applications to Agricultural Yield Data. *International Journal of Advanced Research in Computer and Communication Engineering.* 2013; 2(9).

[2] Swarupa Rani. The Impact of Data Analytics in Crop Management based on Weather Conditions. International Journal of Engineering Technology Science and Research. 2017; 4(5):299-308.

[3] F K Van Evert, S Fountas, D Jakovetic, V Crnojevic, I Travlos, C Kempenaar. Big Data for weed control and crop protection. John Wiley & Sons Ltd on behalf of European Weed Research Society, 2017: 218–233.

[4] Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua. Prediction of crop yield using Big Data. 8th International Symposium on Computational Intelligence and Design.2015.

[5] Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. A Survey on Crop Yield Prediction based on Agricultural Data, International Journal of Innovative Research in Science, Engineering and Technology. 2017; 6(3):4177-4182.

[6] R. Sujatha, P.Isakki Devi. A Study on Crop Yield Forecasting Using Classification Techniques, IEEE, 2016.

[7] V. Sellam and E. Poovammal. Prediction of Crop Yield using Regression Analysis, Indian Journal of Science and Technology, 2016; 9(38).

[8] Patricio Grassinia, Lenny G.J. van Bussel, Justin Van Warta, Joost Wolf, Lieven Claessens, d, Haishun Yanga, Hendrik Boogaarde, Hugo de Groote, Martin K. van Ittersumb, Kenneth G. Cassman. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research. 2015; 49–63.

[9] David B. Lobell, The use of satellite data for crop yield gap analysis, Field Crops Research-143, 2013; 56–64.

[10] Martin K. van Ittersuma, Kenneth G. Cassmanb, Patricio Grassinib, Joost Wolfa, Pablo Tittonell, Zvi Hochmand. Yield gap analysis with local to global relevance-A review. Field Crops Research – 143, 2013; 4–17