LITERATURE SURVEY

(1)- In **“Year book of crop yields - Estimation of crop yields”** by **V.G.Panse** says statistics of crop yields in most countries are based on periodic reports from crop reporters. Essentially such reports represent the reporters quantitative judgment of what the yield is, based on their personal impression of the crop enquiry among farmers. In mote advanced countries such as the U.S.A. and several European countries, crop reporter's are farmers and other private individuals resident in rural areas and connected with farming, who voluntarily supply to the government agency concerned, the information called for in a mailed questionnaire. In underdeveloped countries, crop reporters are generally government officials or agents who submit reports relating to crop yields for the area under their administrative charge. These official reporters are naturally much less numerous than voluntary crop reporters in advanced countries, and consequently, the unit or area for which a report is made is usually large, such as an administrative sub-division of a district and possibly a village, but hardly ever a cultivator’s holding or a field.

Estimates of yield, derived from such reports are likely to be much more inaccurate than those for which a very large number of reporters supply data on their respective areas. Unfortunately, the most serious defect in yield estimates based on crop reporters' data, whether the reporters are official or voluntary, is that they are subjected to large and indeterminate biases. Investigations made in India have shown that the official reporters have a marked tendency to lean towards the normal, with the result that yield is underestimated in favourable seasons and overestimated in poor seasons. Voluntary crop reporter's, on other hand, might have incentives for a systematic underestimation of yield.

In the U.S.A., the regression method is used as a regular routine for eliminating 3 bias in the crop reporters estimates, the regression of these estimates on "actual" yields for past years being used for this purpose. The actual yields are provided by the revised estimates of yield late in the season, on the basis of check data from various official as well as private sources, such as marketing's, shipments, the amount of the crop processed or handled in factories, and other reasonably. complete utilization information. Such
data are available for a few crops like cotton, tobacco, sugar beet; but even here there can be no guarantee that the information is really complete. For the majority of crops, including food crops which are partly consumed locally or are stored by the farmer for his own use and for feeding his livestock, such check data are not available, and indications of yield from other sources, such as an annual farm census in a few states or the quinquennial census in all states, are used as a basis for revising crop reporters estimates of yield,
It is possible, however, that the data from these other sources also suffer from biases as the crop reporters estimates do. Under the conditions prevailing in most countries it is not possible to obtain worthwhile information on the total production of the majority of crops such as would permit the systematic application of the regression technique in order to improve the quality of the crop reporters' estimates of yield. Further, where such regression is feasible, it can remove only the bias for which a trend has been established.

Thus, experience under a wide range of agricultural and economic conditions demonstrates the inability of subjective or personal methods do estimating yield to provide reliable results and points to the necessity of replacing such methods by those involving;

(i) the selection of a representative sample of the crop for observation and
(ii) using for observation the method of physical measurement of yield at harvest.

(2)- In **"Crop yield forecasting using data mining"** by **Pallavi Kamath et al**, India is a heavily reliant on agriculture. Organic, economic, and seasonal factors all influence agricultural yield. Estimating agricultural production is a difficult task for our country, particularly given the current population situation. Crop production assumptions made far in advance can help farmers make the necessary planning for things like storing and marketing.

Crop production prediction involves a huge amount of data, making it a perfect candidate for data mining methods.

Data mining is method of accumulating previously unseen anticipated information from vast database. Data mining assists in the analysis of future patterns and character, enabling companies to make informed decisions. For a specific region, this research provides a fast inspection of agricultural yield forecast using the Random Forest approach. The process of analysing, cleaning, and modelling data to generate useful knowledge and conclusions is known as data analysis.

Methods are used to convert the customer's raw data into valuable information. This research can be extended to agriculture as well. Most farmers relied on their long-term field experience with specific crops to forecast a greater yield in the coming season. Nonetheless, they do not receive a fair price for their crops. It typically occurs because of insufficient irrigation or poor crop selection, but it may also occur when crop yields are lower than expected. Due to a variety of factors, the farmers who make up the majority do not achieve the predicted Crop yield. That data set of crop yield which consists of many components. By studying the soil and atmosphere for the specific area, by which increase crop production, optimal crop can be estimated.

Advantage of this research mainly is Farmers will benefit from this forecast. To determine which crops are best for their farm based on soil type, ph., and fertilizer. In this paper effort is made in order to know the region-specific crop yield analysis and it is processed by implementing by random forest algorithm. In this project have chosen dataset which in .csv format. For the training purpose 80% of data is used and remaining 20% of data is used for testing.

After the successful training and testing next step is finding the accuracy of the model. We have achieved a good accuracy which means this model is good for predicting yield. We have designed the Website which consists of Four Functional Modules as shown in the
1) Crop Module: This module will provide the list of available crops. On selection of each one of it will give the detailed description of the crop.
2) Soil Module: This module will provide the list of available soils. On selection of each one of it will give the detailed description of the soil.
3) Weather Module: In this module by entering the city name the user can get the live weather forecast. Open weather app is free open source weather data. By using weather API key can fetch the current or historical weather data.

(3)- Indonesian Journal of Electrical Engineering and Computer Science[Vol .12 ,No.3,December 2018 ,pp. 1087~ 1093ISSN: 2502-4752, DOI : 10.11591/ijeecs.v12.i3.pp1087-1093]

BM Sagar, Cauvery N K says the result of penetration of technology into the field of agriculture. In this literature, it has been observed that analysis has been done on agriculture soils, hidden patterns discovery using data set related to

climate conditions and crop yields data. The activity of agriculture field are numerous like weather forecasting, soil quality assessment, seeds selection, crop yield prediction has been surveyed and the major trends have been identified using ::ata Analytics”

Agriculture forms the basis for food security and it is important. In India above55% people depend on agriculture as per the recent information. In India ,Wheat and rice are the major grown crops along with sugarcane, oil seeds etc.. The non food items like rubber, cotton, jute etc…More than 70% of rural area depend on agriculture. In the farm output ,India ranks second considering the world wide scenario .This is the widest economic sector and has an important role regarding the frame work of socio-economic fabric of India .Historical information regarding crop yield provides major input for companies in planning supply chain decision like production scheduling .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 insight on various data analytics method applied to crop yield production and also signifies the important lacunae points in the proposed area of research ..

Crop yield production using Big 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 session is an ample task and an advantageous task for every nation with the respect to assess the over all crop yield prediction. The accurate prediction of crop yield certainly benefits the framers in choosing the right method to reduce the crop damage and gets best price for their crops. A research group conducted a work with an objective of accurate

prediction of crop yield through big data analytics to assure various crop yield influencing factors such as Area Under Cultivation (AUC) in terms of hector , Annual Rainfall (AR)rates and Food Price Index (FPI) and to develop relationship among these parameters. All the selected factors of the present study design known as AR, AUC and FPI were measured for a periods of 10 year between the years of 1990-2000. A novel method called Linear Regression (LR) applied to analyse the relationship between explanatory

variables (AR, AUC, FPI) and the crop yield considered.

(4)-Keerthy t et al Some of the increasing role of Remote Sensing in Crop Area Estimation. Remote Sensing may be defined as the collection of information about an object or area without being in physical contact with the object or area. Aircraft and satellites are the most common vehicles from which remote sensing observations are made. Aerial photography is the original and most familiar form of remote sensing and is widely used for topographic mapping, engineering and environmental studies, agricultural estimation, crop disease information, military observations, and exploration for oil and minerals.

The method USDA's Natural Resource Conservation Service (NRCS) has an extremely large point sample and survey known as the National Resources Inventory (NRI). The NRI collects data in all counties and parishes of the 50 states and Puerto Rico, the Virgin Islands, the District of Columbia and areas of the Pacific Basin. The NRI was conducted at five year intervals beginning in 1977 and has been conducted annually since 2000 [Boryan, 2012]. The objective of the NRI is to monitor "status, conditions and trends in soil, water and other natural resources data on non-Federal lands in the United States" [Breidt, F. and W. Fuller,1999]. Data collected include broad land use and land

cover categories such as irrigated and not irrigated cropland. These categories are then evaluated in terms of land capability class and subclass such as prime farm land; erosion and potential erodibility. Annual data collection is conducted primarily using photo interpretation and by ancillary data sources in house [Nusser and Goebel, 1997].

It was also expected to reduce the variance by approximately a factor of three and avoid the problems of incomplete PSUs. It was also considered to be more logical from the implementation perspective, and the new design has served to reduce survey costs. The base sample for LUCAS corresponds to an initial 1 Km grid of about 4,000,000 points for the entire area of the EU. The LUCAS master sample is a subset of the base sample from a 2 Km grid created by using all the even points of the base sample, consisting of around 1,000,000 points. Points located on small islands are excluded from the sample. The master sample is then stratified by land cover classes and sub-sampled for ground data collection.

Some of the issues faced by them at the time. First, the area planted for harvest of a given crop may change throughout the growing season. Such issues as use for purposes other than grain, abandonment, extreme weather damage, or unusual economic conditions may cause this change. It is usually necessary to make estimates several times throughout the crop season even for a given crop. Certainly the most significant challenge in estimating crop production is in doing so early in the season. Area estimation can create problems especially in countries prone to drought or flooding problems. Landscape factors such as elevation and/or soil type also create problems; for example, terraced crops versus those planted on a steep incline.

SAR was deemed to be necessary for this application data because of its robustness to cloudiness, and the fact that there is considerable rain and cloudiness in the paddy rice growing area during the growing season. Apparently, the flooding of the target crop during the growing season, with water signatures mixed with the green plants, made SAR imagery more useful in this effort than it might be for a non-flooded crop. Overall, the results of this crop area estimation effort seemed to be very good for the 2005 crop season.

The methodology has been expanded and implemented on an operational level in the four major summer grain producing provinces of Northwest, Mpumalanga, Free State, and Gauteng, and it is reportedly providing reliable crop area estimates.

Advantage of USDA They are an excellent and timely source for crop area estimation. However, they also serve a very important statistical goal of measuring what is not available from a list of farmers (known as NOL for ‘not on list’ for hundreds of variables. It is the combination of area and listbased information through a multiple frame sampling approach that leads to complete and cost efficient results for agricultural statistics of all sorts.

Advantages: It can be used to statistically combine results, where comparable to sample surveys, or more simply as a quality control for both the administrative and sample survey data. This is a way for saving money, reducing response burden producing figures for very detailed domains and allowing estimation of transition over time.

The test results for the six Southwest Rostov districts indicated that all methods worked reasonably well for large fields, and there was no significant difference between the methods.

The test Some of the methods used to solve. Statistical methods have made clear that the enumeration of small samples can greatly reduce the cost of the collection of agricultural statistics while increasing their accuracy. A well designed sample, for which the data are carefully collected, can provide much cheaper statistics than a census and provide more timely information on current conditions. Combining area and list frames, known as multiple frame sampling, has some very good qualities and solves some of the problems with each of them individually.

Advantage of the VITO NN tended to perform best, and, in general, better results were obtained when using VGT, rather than MODIS data. The authors concluded that further investigation should be made to clarify this latter result, and to verify the performance of NN in conditions where less extensive ground training data are available.

Shailesh Shetty S *et al.* This project supports farmers in evaluating which crop to grow in a specific area at a specific time and predicting whether it will be profitable or not. It gives the specifics by specifying whether the crop is profitable. As a result, this device aids farmers in their decision-making process, allowing them to save time.

Suvidha Jambekar *et al.* Regression analysis is applied as a predictive modelling tool to predict crop production for crop production. The regression algorithms applied were, Multivariate Adaptive Regression Splines, and then Multiple Linear Regression, Random Forest Regression. According to the results, Random Forest Regression may be used to accurately estimate wheat, and rice, and maize production.

B. Devika, B. Ananthi *et al*. Agriculture expands yield production to meet demand to limit overlapping, and the government encourages it for crop yield forecast on Tamil Nadu dataset imports. The regression method is put to the test of yield prediction capabilities in this study.

R. Vidhya *et al*. They observed accuracy rate improves when a dataset with more features is used. As opposed to other approaches, such as Decision trees, linear regression, random forest algorithm is shown to be superior to other prediction algorithms. The included dataset incorporates a lot more variables resulting in more precise prediction.

Hetal Patel, Dharmendra Patel *et al.* They measured performance of the classification algorithms Naive Bayes, J48, and Simple Cart. This crop prediction comparative analysis employs a large dataset.

At last, all the papers show the pros and cons of the estimation of crop yields. We can analyse and give the solution for the problem using **"Data analytics".**