LITREATURE REVIEW

ESTIMATION OF CROP YIELD USING DATA ANALYTICS

ABSTRACT:

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 world wide 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. Farmers experience 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.

This survey focuses on various methods being used for crop yield prediction. The methods being used are Density based clustering techniques, Multiple Linear regression, Clustering large applications (CLARA), Petitioning around Medoids (PAM) and density based clustering algorithm called DBSCAN.

CROP YIELD PREDICTION USING MACHINE LEARNING:

Agriculture, since its invention and inception, be the prime and pre-eminent activity of every culture and civilization throughout the history of mankind. It is not only an enormous aspect of the growing economy, but its essential for us to survive. Its also a crucial sector for Indian economy and also human future. It also contributes an outsized portion of employment. Because the

time passes the requirement for production has been increased exponentially. So as to produce in mass quantity people are using technology in an exceedingly wrong way. New sorts of hybrid varieties are produced day by day. However, these varieties dont provide the essential contents as naturally produced crop. These unnatural techniques spoil the soil. It all ends up in further environmental harm. Most of these unnatural techniques are wont to avoid losses.

But when the producers of the crops know the accurate information on the crop yield it minimizes the loss. Machine learning, a fast-growing approach thats spreading out and helping every sector in making viable decisions to create the foremost of its applications. Most devices nowadays are facilitated by models being analyzed before deployment. The main concept is to increase the throughput of the agriculture sector with the Machine Learning models. Another factor that also affects the prediction is the amount of knowledge thats being given within the training period, as the number of parameters was higher comparatively. The core emphasis would be on precision agriculture, where quality is ensured over undesirable environmental factors. So as to perform accurate prediction and stand on the inconsistent trends in temperature and rainfall various machine learning classifiers like Logistic Regression, NaÃ¯ve Bayes, Random Forest etc. are applied to urge a pattern. By applying the above machine learning classifiers, we came into a conclusion that Random Forest algorithm provides the foremost accurate value. System predicts crop prediction from the gathering of past data. Using past information on weather, temperature and a number of other factors the information is given. The Application which we developed, runs the algorithm and shows the list of crops suitable for entered data with predicted yield value.

CROP YIELD PREDICTION BASED ON AGRICULTURAL DATA:

Agriculture is one of the major revenue producing sectors of India and a source of survival. Various seasonal, economic and biological factors influence the crop production but unpredictable changes in these factors lead to a great loss to farmers. These risks can be quantified when appropriate mathematical

or statistical methodologies are applied on data related to soil, weather and past yield. With the advent of data mining, crop yield can be predicted by deriving useful insights from these agricultural data that aids farmers to decide on the crop they would like to plant for the forthcoming year leading to maximum profit. This paper presents a survey on the various algorithms used for crop yield prediction.

Data mining is defined as a process of identifying previously unknown inferences from the huge volume of available data. It finds application in market analysis, production control, fraud detection, customer retention, E commerce etc. Data mining software analyses relationships and patterns in stored transaction data based on open ended user queries. On the basis of the nature of data being mined there are two categories of functions involved in data mining namely, Descriptive function that deals with general properties of data and Prediction function that identifies the trends based on available data.

CROP YIELD PREDICTION USING BIG DATA ANALYTICS:

Agriculture is the backbone of Indian economy. Big data are emerging precised and viable analytical tool in agricultural research field. A modern trend in the Agriculture domain has made people realize the importance of big data. Big data is a term associated with the new dataset model and the technologies needed to alter in line with the new data source, which refers to big data analytics.

From Big data analysis farmers can identify the changes in rainfall, soil quality types of soil and forming times. So they are ready to face and get good results in harvesting. Machine learning technique increases the quality of the big data analysis. Now the crop yield predictions are automatically better even though the big data analysis is collected by machine learning techniques there may be chances for unstructured. Big data Analysis with Machine Learning Approaches by Hadoop technique is applied to analyze the large volume of data. Clustering and K- Means Algorithm is used to identify the group of data and collaborate in one cluster. Big data is used where the crops are dependent on the weather conditions by applying a different algorithm and tools. By using machine learning approaches under the

platform Hadoop we can manage the environmental conditions, such like, soil types, soil testing, crop pattern, weather, crop rotation by proper parameter.

REFRENCES:

Agrawal et al discuss about various Data Mining tools such as Dashboards, Text-Mining tools. They provide an overview about these tools and the various scenarios in which they can be deployed [2].

The proposed architecture mainly focuses on open source tools for the development of the application. The user can select location from map for which the details are available at one click [3].

K. Sabarina and N. Priya, [2] has presented an efficient strategy for crops Big data for the benefit of precision agricultural lowering data dimensionality. Predictive analytics can be used to make the smartest decision in farming by collection real-time data analysis with streaming data.

S. Rajeswari, K. Suthendran, and K. Rajakumar, [3] have presented a conceptual architecture of the big data open platform used for supporting many sectors like agriculture. Big data, mining Techniques cloud-based big data analytics and IOT Technology performs an important role in the feasibility study of smart agriculture.

Aruvansh Nigam, Saksham Garg, Archit Agrawal[1] conducted experiments on Indian government dataset and its been established that Random Forest machine learning algorithm gives the best yield prediction accuracy. Sequential model thats Simple Recurrent Neural Network performs better on rainfall prediction while LSTM is good for temperature prediction. The paper puts factors like rainfall, temperature, season, area etc. together for yield prediction. Results reveals that Random Forest is the best classier when all parameters are combined.

Leo Brieman [2] , is specializing in the accuracy and strength & correlation of random forest algorithm. Random forest algorithm creates decision trees on different data samples and then predict the data from each subset and then by voting gives better the answer for the system. Random Forest used the bagging method to trained the data. To boost the accuracy, the randomness injected has to minimize the correlation while maintaining strength.

TEAM MEMBERS:

SHRINITHI

SRIROOPA

BOOMIKA

VINISHA

DEPARTMENT OF COMPUTER SCIENCE

ST. JOSEPH'S COLLEGE OF

ENGINEERING