Crop Yield Prediction Based on Indian Agriculture Using Machine Learning

Mrs.R.Kiruthika[1], Miss.M.Sowndharya[2]
[1]Assistant Professor, Dept of Computer Science and Engineering
[2]Dept of Computer Science and Engineering
[1, 2] Mahendra Engineering College, Namakkal, India-637503

Abstract- *Crop yield prediction is extremely challenging due to its dependence on multiple factors such as crop genotype, environmental factors, management practices, and their interactions. Have to collect and employ machine learning algorithms on the data about temperature, rainfall, soil, seed, crop production, humidity, and wind speed data, which will help the farmers improve the productivity of their crops. Firstly, pre-process the data with a data engineering perspective and then apply machine learning algorithms which further analyze and processes the large volume of data. Secondly, employ clustering algorithms which are employed on results gained and provide a better result on the data in terms of accuracy. After that, use bar graphs and scatter plots to study the relationship between the crops, rainfall, temperature, and soil seed type in selected regions. Further, a recommender system based on this has been used to predict the crops and display them for use by the farmers. The system design is scalable and can be used to find the recommended crops and can be employed in a larger geographical region.*

I. INTRODUCTION

With the sudden changes in weather conditions, farmers and agriculture throughout the country suffer as they fail to produce enough crops. This leads them to take serious steps as they are unable to provide for their family and make ends meet. This also leads to a scarcity of availability of food resources in the country. The conditions of farmers in our country need to be changed. India's economy is greatly influenced by agriculture as it serves as the backbone of the country. More than 50% of the country is dependent directly or indirectly on the agriculture sector and it is responsible for the employment of the major labor force of the country, which accounts for over 40%. Agriculture produces big volumes of data every year, and hence there is a need to get rid of the obsolete traditional predicting methods by charts and use the availability of the big data collected to create a more prioritized and accurate predicting system.

Big data will help confront the challenges and enhance the understanding of the whole sector. Big data analytics is the process of examining large data sets containing a variety of data types. The influence of weather can be deemed a major priority in predicting crop yield. A lot of research work has been conducted in identifying how weather as a factor affects agriculture, but most of these studies require large complex information which is not directly available. This leads to the collection of data by estimation which can have either a negative or a positive effect. Hence improvement is needed in the methodology to compensate for the availability of data. Focuses on crop prediction using agricultural and clustering algorithm along with a recommendation function is carried out in the hope to propose crops to sow and elucidate big data applications in agricultural production.

1.1 WEATHER PREDICTION

Climate and weather influence cropping area, intensity, and yield in different ways. Plant growth is dependent on precipitation and temperature. If the precipitation level is too high or too low or if the temperature is too high or too low, plants may not grow well. Weather plays an important role in agricultural production. It has a profound influence on the growth, development, and yields of a crop, incidence of pests and diseases, water needs, and fertilizer requirements. Despite careful agronomic planning on a micro-scale to suit local climate crops experience various types of weather changes on a year-to-year basis.

Deviations from normal weather occur with higher frequencies in almost all years, areas, and seasons. The most common one is a delay in the start of the crop season due to a delay in monsoon onset in rain-fed regions of India and temperature changes or high-temperature differences or early cessation of monsoon. The effects of weather changes from a normal pattern on crops build up slowly but are often wide enough to destabilize the national agricultural production scenario.

1.2 GEOGRAPHICAL FACTORS

Soils are the uppermost part of the earth's crust, formed mainly by the weathering of rocks, formation of

humus, and material transfer. They vary in terms of origin, appearance, characteristics, and production capacity. Soil fertility is the ability of a soil to deliver nutrients needed for the optimum growth of a specified crop. Soil fertility is one of the most important factors in crop production.

1.3 DATA MINING

Machine learning is an important decision support tool for crop yield prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Several machine learning algorithms have been applied to support crop yield prediction research.

1.4MACHINE LEARNING

Machine learning (ML) plays a significant role as it has a decision support tool for Crop Yield Prediction (CYP) including supporting decisions on what crops to grow and what to do during the growing season of the crops. The present research deals with a systematic review that extracts and synthesizes the features used for CYP furthermore, there are a variety of methods that were developed to analyze crop yield prediction using artificial intelligence techniques.

1.5 RECOMMENDED SYSTEM

A large amount of weather and climate information is presently available for farmers. A portion of the information is operational or already under development, and in particular, forecasting through climatic data and formation may not be suitable for farmers when it comes to the decision-making process.

The best way to gain an advantage from natural factors is to consider them during decision-making and understand them in the best way possible. Meteorological information on agriculture, and climatic data, in particular, is an important aspect of planning in the context of agricultural production. Therefore, climatic conditions must be an integral part of the decision-making process.

These factors can be determined by recording hourly, daily, and weekly temperature data, rainfall, solar radiation, wind speed, relative humidity, and evapotranspiration. Hence improvement is needed in the methodology to compensate for the availability of data. 4 Mainly focuses on crop prediction using agricultural and clustering algorithm along with a recommendation function is carried out in the hope to propose crops to sow and elucidate big data applications in agricultural production.

II. LITERATURE REVIEW

Previous methods have always focused on analyzing past data and making an informed guess at how the crops may be chosen but fail to propose an idea to optimally find out the intricate details and take into account the ever-changing geographical and climatic changes. Hence we proposed Machine Learning.

2.1 WB-CPI: WEATHER-BASED CROP PREDICTION IN INDIA USING BIG DATA ANALYTICS [RISHI GUPTA; AKHILESH 2.1 WB-CPI: WEATHER-BASED CROP PREDICTION IN INDIA USING BIG DATA ANALYTICS [RISHI GUPTA; AKHILESH

Collecting and analyzing temperature, rainfall, soil, seed, crop production, humidity, and wind speed data (in a few regions), will help the farmers improve the productivity of their crops. Firstly, pre-process the data in a Python environment and then apply the Map-Reduce framework, which further analyses and processes the large volume of data. Secondly, k-means clustering is employed on results gained from Map Reduce and provides a mean result on the data in terms of accuracy.

After that, use bar graphs and scatter plots to study the relationship between the crop, rainfall, temperature, soil, and seed type of two regions (Ahmednagar, Maharashtra, and, Andaman and Nicobar Islands). Further, a self-designed recommender system has been used to predict the crops and display them on a Graphic User Interface designed in a Flask environment. After studying the previous work done, the main aim would be to process the data using Map Reduce and frame a recommender algorithm in Python to extract output according to the seasonal conditions and region followed by executing k-means clustering and finding the mean produce per area a group of crops will give in a particular region.

2.2 RICE CROP YIELD PREDICTION IN INDIA USING SUPPORT VECTOR MACHINES [Owaiz Petkar; Niketa Gandhi; Leisa J. Armstrong; Amiya Kumar Tripathy] 21 November 2016

The focus is on optimizing Machine learning techniques that can be used to improve the prediction of crop yield under different climatic scenarios. Presents the review on the use of such machine learning techniques for Indian rice cropping areas. Discusses the experimental results obtained by applying an SMO classifier using the WEKA tool on the dataset of 27 districts of Maharashtra state, India.

The dataset considered for the rice crop yield prediction was sourced from publicly available Indian Government records. The parameters considered for the study were precipitation, minimum temperature, average temperature, maximum temperature, and reference crop evapotranspiration, area, production, and yield for the Kharif season (June to November) for the years 1998 to 2002. For the present study the mean absolute error, root mean squared error, relative absolute error, and root relative squared error (RRSP) were calculated.

2.3 A NOVEL APPROACH FOR EFFICIENT CROP YIELD PREDICTION [P.S.Maya Gopal; R.Bhargavi] -October 2019

Extensive research in an agricultural domain has been carried out to predict better crop yield using the machine learning algorithm Artificial Neural Network (ANN) and the statistical model Multiple Linear Regression (MLR). This article examines the intrinsic relationship between MLR and ANN. A hybrid MLR-ANN model has been proposed in this research work for efficient crop yield prediction.

The proposed hybrid model is modeled to analyze the prediction accuracy when MLR intercept and coefficients were applied to initialize the ANN's input layer weights and bias. Feed Forward Artificial Neural Network with Back Propagation training algorithm was used for predicting accurate paddy crop yield. In the conventional ANN model, the weights and biases of input and hidden layers are initialized randomly.

III. EXISTING SYSTEM

In existing systems, all the ideas have mainly focused to have made use of Big Data Analytics techniques to make an automatic prediction system. First, they collected data from various sources like social media, sensor data, weather forecasts, etc., and loaded the pre-processed data into HDFS. HDFS stores datasets and provides backup features. Focused on three factors while collecting data: precipitation, temperature, and cloud cover for the state.

Implemented a prediction function for establishing forecast data through the K-means cluster algorithm. Used Apache Mahout to implement a logistic regression algorithm to predict the future based on the past data. For this, testing and training of data are performed. Then they evaluated the accuracy of the predicted result and represented the output using visualizations making use of the Flotend tool.

Made various plots showing yearly and monthly average temperature for a particular region, maximum and minimum temperature, precipitation, etc. Predict the crop yield and suggest crops based on it which would, in turn, increase the farmers and overall, the entire agriculture sector. Used a new algorithm to predict the crop yield and suggest crops based on the crop yield and considering the soil type. Used weather datasets containing information about temperature, rainfall in mm, wind speed, evaporation, humidity, etc. The weather datasets to determine the type of soil. Datasets for crop diseases to determine the ideal weather conditions which would be suitable for a particular crop to grow.

IV. PROPOSED SYSTEM

Agriculture is one of the main sectors of social concern since it provides a significant amount of food. At present, numerous nations are still hungry due to the shortage or lack of food with a rising population. The consolidated impacts of a rising population, natural weather variability, soil loss, and climate-changing demand techniques ensure crop growth and production in a timely and reliable manner. It also requires contributing to expanding agricultural food production sustainability.

These requirements indicate that land assessment, the protection of crops, and crop yield prediction are of greater importance to worldwide food production.Thus, an accurate crop yield prediction is mandatory to rely on by the nation's policymaker to obtain convenient export and import evaluations for enhancing national food security.

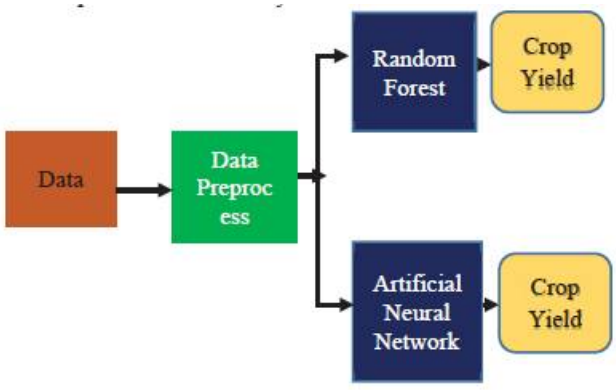

Fig 1: **ANN diagram for crop yield prediction**

- Providing the current status of crop production. Describing the fundamental aspects of the crop yield prediction process.
- An extensive critical review of the machine learning-based crop yield prediction algorithms; critical evaluation of utilized feature sets; comparative analysis of the related study.

- Detailed investigation of benefits and challenges associated with features and machine learning algorithms in the prediction of crop yield.

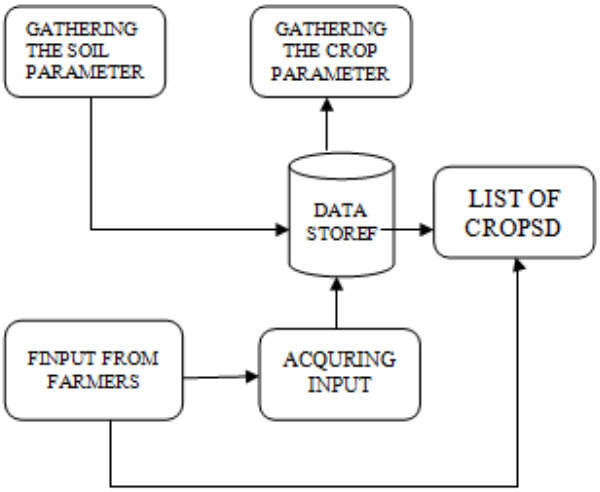

Fig 2: Crop yield prediction using machine learning

V. METHODOLOGY

Comprehensively gathering a wide range of machine learning-based crop yield prediction algorithms and combining them with other experimental articles. The machine learning-based crop yield prediction method consists of some phases, namely data collection, data preprocessing, data partition, and data analysis. A wide range of regression and classification-based prediction algorithms have been utilized to forecast crop yield.

5.1 DATA ANALYSIS

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.The ability to analyze data is critical to gaining value from the information are collecting. Learn the tools available for establishing analytics and make sure they support the results that want to achieve. The process of cleaning, changing, and processing raw data, and extracting actionable, relevant information that helps to predict the crops.

5.2 ARTIFICIAL NEURAL NETWORK

Artificial neural networks have been demonstrated to be powerful tools for modeling and prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameter of soil and also parameter related to atmosphere.An artificial neural network is an attempt to simulate the network of neurons that make up a human brain so that the computer will be able to learn things and make decisions in a human-like manner. They have been demonstrated to be powerful tools for modeling and prediction, to increase their effectiveness.

The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations.

VI. CONCLUSION

Thus, a crop recommendation system can better the results by using just Map Reduce and K-means clustering, which gives efficient results in computations. The model focuses on a wide range of crops and their products per area along with the soil type and seed types depending on the varieties used in a particular region. From the visualization graphs of K-Means clustering, we can and the mean produces for a group of crops. The algorithms that have been used for the recommender function and K-Means Clustering can be accessed. Also, the relation between parameters (like optimal temperature, seasonal rainfall, wind speed, humidity, soil availability, required seed types), crop, and the region has been studied and displayed using 2D and 3D graphs.

The system is scalable and it can be used to and the recommended crops of other states in a similar manner as described in the methodology. It can be further improved to eliminate the problem of disproportion in the production and requirement ratio if an aspect of humidity and wind speed can be added for all the regions and will give a more accurate recommendation. Factors like soil moisture, irrigation, cloud cover, etc. may be included in the system to retune its output. Also, the recommender can be modified to warn about the diseases that can occur in a crop in a particular season and suggest the types of fertilizers or nutrients needed in the soil for the crop to grow and give its best yield.

REFERENCES

[1] Mamunur Rashid; Bifta Sama Bari, Yusri Yusup; Mohamad Anuar Kamaruddin; and Nuzhat Khan; “A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction”- April 22, 2021.

[2] Gang Pan; Feng-min Li; Guo-jun Sun; "Digital camera-based measurement of crop cover for wheat yield

prediction" 2007 IEEE International Geoscience and Remote Sensing Symposium -07 January 2008.

[3] Niketa Gandhi; Leisa J. Armstrong; Owaiz Petkar; Amiya Kumar Tripathy; “Rice Crop Yield Prediction in India using Support Vector Machines” 21st Nov 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).

[4] Phusanisa Charoen-Ung; Pradit Mittrapiyanuruk; “Sugarcane Yield Grade Prediction using Random Forest and Gradient Boosting Tree Techniques” 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE).

[5] Mamunur Rashid; Bifta Sama Bari; Yusri Yusup; Mohamad Anuar Kamaruddin; Nuzhat Khan; “A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction” IEEE Access (Volume: 9) 22 April 2021.

[6] Camgirl Suresh; N.V.K.Ramesh; Syed Inthiyaz; P. Poorna Priya; Kurra Nagasowmika,1Kota.V.N.Harish Kumar; Mashkoor Shaik and 2B. N. K. Reddy; “Crop Yield Prediction Using Random Forest Algorithm” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

[7] D.Jayanarayana Reddy; M. Rudra Kumar “Crop Yield Prediction using Machine Learning Algorithm”; 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS).

[8] Rishi Gupta; Akhilesh Kumar Sharma; Oorja Garg; Krishna Modi; Shahreen Kasim; Zirawani Baharum; Hairulnizam Mahdi; "WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics" published 04 October 2021 in IEEE Access**.**