Research on Crop Yield Prediction Based on Indian Agriculture Using Machine Learning

Nitesh patel[1], Prof. Sheela singh[2]
[1, 2]Student Department of Computer Science
[1, 2]Saraswati Higher Education and Technical College of Engineering Babatpur Varanasi Uttar Pradesh 221106.

Abstract- *This paper has been prepared as an effort to reassess the research studies on the relevance of machine learning techniques in the domain of agricultural crop production. This method is a new approach for production of agricultural crop management. Accurate and timely forecasts of crop production are necessary for important policy decisions like import-export, pricing marketing distribution etc. which are issued by the directorate of economics and statistics. However, one has understood that these prior estimates are not the objective estimates as these estimate requires lots of descriptive assessment based on many different qualitative factors. Hence there is a requirement to develop statistically sound objective prediction of crop production. That development in computing and information storage has provided large amount of data. In this paper it proposes a crop selection method to maximize crop yield based on weather and soil parameters. It also suggests the proper sowing time for suitable crops using seasonal weather forecasting. Machine learning algorithms such as Recurrent neural network is used for weather prediction, and Random forest classification algorithm is used to select suitable crops. The result of proposed weather forecasting technique is compared with conventional Artificial neural network, which shows better performance results for each selected weather parameters. Time series analysis, Markov chain model, k-means clustering, k nearest neighbor, and support vector machine are applied in the domain of agriculture were presented.*

Keywords- neural network, Machine Learning, Time series analysis, Crop yield prediction

I. INTRODUCTION

In our analysis, that we tend to have discovered within side the previous studies papers are that everybody makes use of environmental condition factors like rain, daylight and agricultural factors like soil sort, nutrient possessed via the soil (Nitrogen, Potassium, etc. As Agriculture is the backbone of Indian economy.

The yield obtained primarily depends on weather conditions as rainfall patterns largely influence cultivation methodologies. With this context, farmers and agriculturalists require spontaneous advice proposition in predicting future reaping instances to maximize crop yield.

Due to insufficient involvement of technology, the throughput of agriculture is yet to reach its full glory. Every farmer is interested in knowing the yield he/she could expect at the harvest period and hence, yield prediction is an important aspect for them. On the other hand, crop yield forecasting is exceedingly challenging because of various complex aspects. Crop yield mainly depends upon climatic conditions, soil quality, landscapes, pest infestations, water quality and availability, genotype, planning of harvest activity and so on.

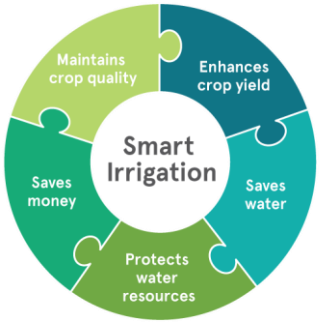

Fig.1 : Smart Irrigation

Over the years, farmers have an idea about the pattern in yield as per innate human intuition. However, rainfall as a major driver for crop raising can extensively rattle intuitive yield prediction by controlling some of the soil and environmental parameters related to the crop growth. Moreover, the right kind of soil to be employed for a crop is only known to the farmer only by on-paper advice and makes it. vital motivation behind making a managed market is to put off the undesirable exchange work out, to diminish the charges inside the commercial Centre and to offer reasonable expenses to the Farmers. A few activities have been taken to advance rural showcasing a decent method to cultivate and keep up the place of country monetary improvement. To advantage the cultivating from the new worldwide market, get admission to potential outcomes, the inward rural promoting device inside the United States of America more over wishes to be joined and strengthened.

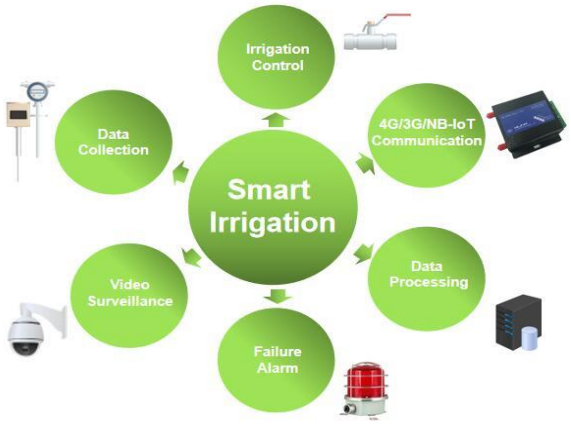

Fig.2 : Smart Irrigation
Source: W link-tech

The goal of this paper is to investigate the effect of coupling process-based modeling with machine learning algorithms towards improved crop yield prediction. The specific research objectives include:

Explore whether a hybrid approach (simulation crop modeling + ML) would result in better corn yield predictions in three major US Corn Belt states (Illinois, Indiana, and Iowa).Investigate which combinations of hybrid models (various ML x crop model) provide the most accurate predictions.
3. Determine the features from the crop modeling that are most relevant for use by ML for corn yield prediction.

Figure 1depicts the conceptual framework of this paper.

AIM & OBJECTIVE

Aim

The proposed system takes into consideration the data related to soil, weather and past year production Aim and suggests which are the best profitable crops which can be cultivated in the a propose environmental condition. As the system lists out all possible crops, it helps the farmer in deciding of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.

Objective

To increase the accuracy of crop yield prediction as well as provide an easy to use User Interface.

To analyse different climatic parameters (cloud cover, rainfall, temperature).

To derive simple decision models to demonstrate the usefulness of the stochastic yield projections in meeting specified agricultural policy goals.

To obtain estimates of aggregate physical production functions for the yields of various crops in specified states, considering various technological factors and a newly developed meteorological weather index as inputs.

Scope of study:

The scope of the project is to determine the crop yield of an area by considering dataset with some features which are important or related to crop production such as temperature, moisture, rainfall, and production of the crop in previous years. To predict a continuous value, regression models are used. It is a supervised technique. It focuses on evolution of a prediction model which may be used to predict crop yield production.

Need of The Study

When it comes to crop yield, most of us are concerned with what it will look like at harvest time. Crop yield prediction is an important agricultural problem. Each and Every farmer is always tries to know, how much yield will get from his expectation. In the past, yield prediction was calculated by analyzing farmer's previous experience on a particular crop. The Agricultural yield is primarily depending on weather conditions, pests and planning of harvest operation. Accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management.

II. LITERATURE REVIEW

1 Sonal Jain,(2020) conducted study on Machine Learning convergence for weather based crop selection in that classification and prediction techniques such as decision tree, naive Bayes, random forest are used for selection of suitable crops based upon weather and soil parameters. An ensemble technique using na¨ıve Bayes, CHAID, random forest and K-nearest neighbor is presented in which uses classification parameters such as root depth, texture, soil color, drainage etc. However, this method is able to select only one crop while more than one crop can be suitable for specified land and weather. A na¨ıve Bayes classification technique is presented in to give suitability of crops such as chili, rice, maize and cotton using weather parameters like rainfall, soil moisture,

temperature and atmospheric pressure. This method also recommends the suitable time for harvesting and sowing of a particular crop.

2 Shruti Kulkarni (2018) Predictive Analysis to Improve Crop Yield using a Neural Network Model. in this project a data-driven model that learns by historic soil as well as rainfall data to analyse and predict crop yield over seasons in several districts, has been developed. For this a particular crop, Rice is considered. The designed hybrid neural network model identifies optimal combinations of soil parameters and blends it with the rainfall pattern in a selected region to evolve the expectable crop yield. The backbone for the predictive analysis model with respect to the rainfall is based on the Time-Series approach in Supervised Learning. The technology used for the final prediction of the crop yield is again a branch of Machine Learning, known as Recurrent Neural Networks.

3 Mayank Champaneri,e.al(2020) conducted study on crop yield prediction using machine learning. In this project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Implementation of such a system with an easy-to-use web based graphic user interface and the machine learning algorithm will be carried out. The results of the prediction will be made available to the farmer. Thus, for such kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield.

4 Potnuru Sai Nishan e.al(2020) conducted study on Crop Yield Prediction based on Indian Agriculture using Machine Learning. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.

5 Ankit Singh Chauhan e.al (2019) Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. According to study Agriculture plays a significant role in the economy of India. This makes crop yield prediction an important task to help boost India's growth. Crops are sensitive to various weather phenomena such as temperature and rainfall. Therefore, it becomes crucial to include these features when predicting the yield of a crop. Weather forecasting is a complicated process. In this work, three methods are used to forecast- ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average) and ARMAX (ARMA with exogenous variables). The performance of the three is compared and the best model is used to predict rainfall and temperature which are in turn used to predict the crop yield based on a fuzzy logic model.

6. Cagatay Catal, e.al, (2020) conducted study on Crop yield prediction using machine learning: A systematic literature review, in this study, we performed a Systematic Literature Review (SLR) to extract and synthesize the algorithms and features that have been used in crop yield prediction studies. Based on our search criteria, we retrieved 567 relevant studies from six electronic databases, of which we have selected 50 studies for further analysis using inclusion and exclusion criteria. We investigated these selected studies carefully, analyzed the methods and features used, and provided suggestions for further research. According to our analysis, the most used features are temperature, rainfall, and soil type, and the most applied algorithm is Artificial Neural Networks in these models.

7. Kodimalar Palanivel (2019) conducted study on an approach for prediction of crop yield using machine learning and big data techniques. Current challenges of water shortages, uncontrolled cost due to demand-supply, and weather uncertainty necessitate farmers to be equipped with smart farming. In particular, low yield of crops due to uncertain climatic changes, poor irrigation facilities, reduction in soil fertility and traditional farming techniques need to be addressed. In this paper, an investigation has been performed on how various machine learning algorithms are useful in prediction of crop yield. An approach has been proposed for prediction of crop yield using machine learning techniques in big data computing paradigm.

8. Shivani S. Kale (2019) conducted study on A Machine Learning Approach to Predict Crop Yield and Success Rate. This research describes the development of a different crop yield prediction model with ANN, with 3 Layer Neural Network. The ANN model develops a formula to ascertain the relationship using a large number of input and output examples, to establish model for yield predictions an Activation function: Rectified Linear activation unit (Relu) is used. The backward and forward propagation techniques are used.

III. METHODOLOGY

Research methodology is the specific procedures or techniques used to identify, select, process, and analyze information about a topic. In a research paper, the methodology section allows the reader to critically evaluate a study's overall validity and reliability.

Data Pre-Processing

Data Preprocessing is a method that is used to convert the raw data into a clean data set. The data are gathered from different sources; it is collected in raw format which is not feasible for the analysis. By applying different techniques like replacing missing values and null values, we can transform data into an understandable format. The final step on data preprocessing is the splitting of training and testing data.

Fig.3 Data Proccessing

Factors affecting Crop Yield and Production. There are a lot of factors that affects the yield of any crop and its production. These are basically the features that help in predicting the production of any crop over the year. In this paper we include factors like Temperature, Rainfall, Area, Humidity and wind speed.

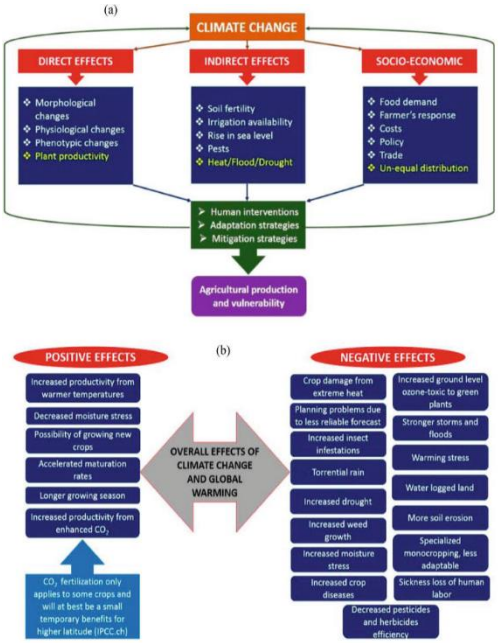

Fig.4 : Factors affecting Crop Yield and Production

Comparison and Selection of Machine Learning Algorithm

Before selecting on an algorithm to utilize, we must first analyse and compare them, and then choose the best one that matches this unique dataset. Machine Learning is the greatest technology for providing a more realistic answer to the agricultural production issue. There are several machine learning algorithms that are used to forecast agricultural production. Some of the following machine learning methods for selection and accuracy comparison:

- Logistic Regression: Logistic regression is a supervised learning classification technique used to estimate the likelihood of a target variable. The nature of the target or dependent variable is dichotomous, which implies there are only two potential classes.
- Naive Bayes: The Naive Bayes classifier implies that the existence of a specific feature in a class is independent to the presence of any other feature. The Naive Bayes model is simple to construct and is especially good for huge data sets. Along with its simplicity, Naive Bayes is renowned to outperform even the most advanced classification systems.
- Random Forest: Random Forest has the capacity to examine crop development in relation to present climatic circumstances and biophysical change. The random forest method generates decision trees based on distinct data samples, predicts the data from each subset, and then provides a better answer for the system via voting. Random Forest trains the data using the bagging approach, which boosts the accuracy of the outcome.

IV. PROBLEM STATEMENT

Crop yield prediction is critical for planning and making policy choices. Many nations still rely on ground-based inspections and reports for crop monitoring and production forecasts. These procedures are arbitrary, expensive, and time-consuming. The following are some of the most prevalent issues with present crop production forecast methodologies.

A. Seasonal weather prediction

The dataset used for weather prediction is collected from NRSA Hyderabad station, which consists of five years' data of weather parameters such as temperature, humidity, sun hours, wind speed, wind direction, etc.

The data are collected from January 2014 to September 2019 at onehour resolution. For training the prediction model, initially data pre-processing is carried out which comprises filling missing values, data transformation and data normalization. The missing values of data are filled using linear interpolation method. Following this, hourly data is converted into daily average data consist of minimum temperature, maximum temperature calculated from temperature data. Subsequently, min-max normalization is performed on dataset to convert it into a common scale. For weather prediction, recurrent neural network (RNN) is used which is a variant of neural network (NN) intended to handle sequence dependence.

B. Identification of suitable crop

It uses recurrent neural network (RNN) for seasonal weather forecasting. A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. Recurrent neural networks are well-designed machine learning techniques that stand out for their quality in different fields of activity, such as signal processing, natural language processing and speech recognition. Contrary to convolutional neural networks, RNNs clearly manage the temporal data dependencies, since the output of the neuron in time t- 1 is used with the next input, to feed the neuron itself at time t. A diagram of a typical neural RNN is detailed in Figure

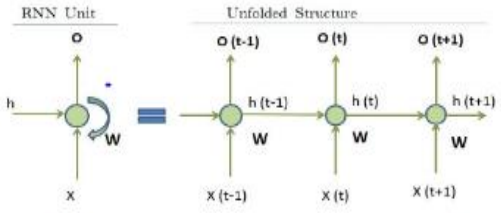

Figure 4. RNN Unit (on the left) and unfolded structure (on the right).

Fig.5: RRN Unit and unfolded structure

The advantage of RNN over conventional NN is, the output from previous step is forwarded to current step, hence the hidden state in RNN remembers the sequence while in conventional neural network all the values are treated independent of each other. The basic architecture of RNN is depicted in Fig. 2 consisting of an input layer, two hidden layers and an output layer.

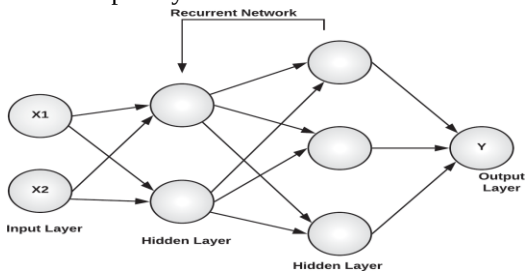

Fig.6: Hidden Layer

Result

In this research paper that are the many of outputs of shows in below:

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, Rice to nan
Data columns (total 8 columns):
2004-05    12 non-null float64
2005-06    12 non-null float64
2006-07    12 non-null float64
2007-08    12 non-null float64
2008-09    12 non-null float64
2009-10    12 non-null float64
2010-11    12 non-null float64
2011-12    12 non-null float64
dtypes: float64(8)
memory usage: 936.0+ bytes
None
```

Fig.7: output 1

For implementation of crop yield prediction using Machine learning python language is used. In this we have taken the data from year 2004 to 2012 and the data is stored into float64 data type.

```
print(crops_prod_data)
```

Crop	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	\
Rice	100.0	101.0	99.0	105.0	112.0	121.0	
Wheat	100.0	101.0	112.0	115.0	117.0	127.0	
Coarse Cereals	100.0	107.0	110.0	115.0	113.0	123.0	
Pulses	100.0	108.0	134.0	124.0	124.0	146.0	
Vegetables	100.0	109.0	103.0	118.0	113.0	124.0	
Fruits	100.0	99.0	99.0	98.0	102.0	104.0	
Milk	100.0	97.0	98.0	98.0	98.0	112.0	
Eggs, Fish and Meat	100.0	102.0	101.0	100.0	99.0	116.0	
Oilseeds	100.0	86.0	85.0	97.0	104.0	103.0	
Sugarcane	100.0	96.0	91.0	87.0	80.0	81.0	
Fibers	100.0	92.0	91.0	96.0	109.0	107.0	
All Agriculture	100.0	99.0	101.0	104.0	106.0	115.0	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Fig.8 : Output 2

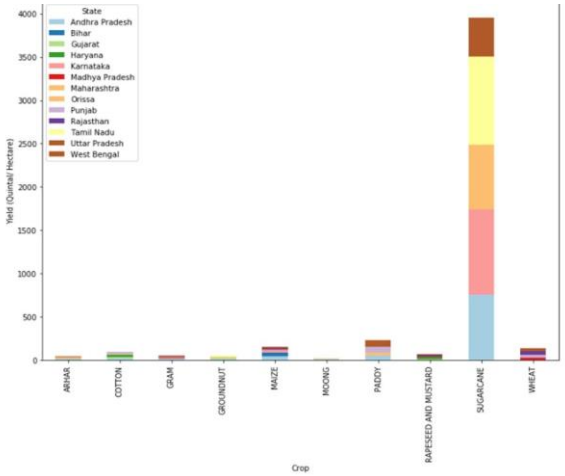

Fig.9 : Crop

```
Data columns (total 6 columns):
Crop                                 49 non-null object
State                                49 non-null object
Cost of Cultivation (`/Hectare) A2+FL 49 non-null float64
Cost of Cultivation (`/Hectare) C2    49 non-null float64
Cost of Production (`/Quintal) C2     49 non-null float64
Yield (Quintal/ Hectare)              49 non-null float64
dtypes: float64(4), object(2)
memory usage: 2.4+ KB
None
```

Fig.10: output 3

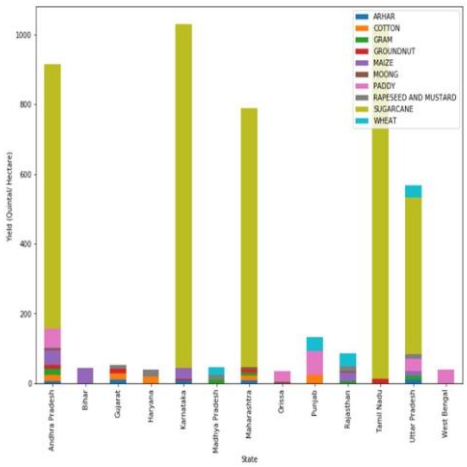

Fig.11: Static

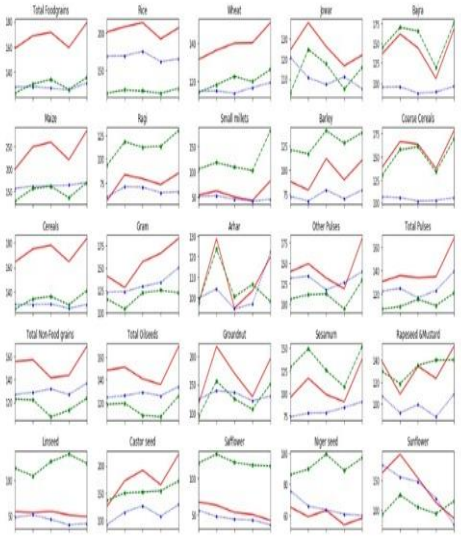

Fig.12: output 4

	0	1	2	\
Crop	ARHAR	ARHAR	ARHAR	
State	Uttar Pradesh	Karnataka	Gujarat	
Cost of Cultivation (`/Hectare) A2+FL	9794.05	10593.1	13468.8	
Cost of Cultivation (`/Hectare) C2	23076.7	16528.7	19551.9	
Cost of Production (`/Quintal) C2	1941.55	2172.46	1898.3	

	3	4	\
Crop	ARHAR	ARHAR	
State	Andhra Pradesh	Maharashtra	
Cost of Cultivation (`/Hectare) A2+FL	17051.7	17130.5	
Cost of Cultivation (`/Hectare) C2	24171.7	25270.3	
Cost of Production (`/Quintal) C2	3670.54	2775.8	

	5	6	7	\
Crop	COTTON	COTTON	COTTON	
State	Maharashtra	Punjab	Andhra Pradesh	
Cost of Cultivation (`/Hectare) A2+FL	23711.4	29047.1	29140.8	
Cost of Cultivation (`/Hectare) C2	33116.8	50828.8	44756.7	
Cost of Production (`/Quintal) C2	2539.47	2003.76	2509.99	

	8	9	...	\
Crop	COTTON	COTTON	...	
State	Gujarat	Haryana	...	
Cost of Cultivation (`/Hectare) A2+FL	29616.1	29919	...	
Cost of Cultivation (`/Hectare) C2	42070.4	44018.2	...	
Cost of Production (`/Quintal) C2	2179.26	2127.35	...	

Fig.13 : output 5

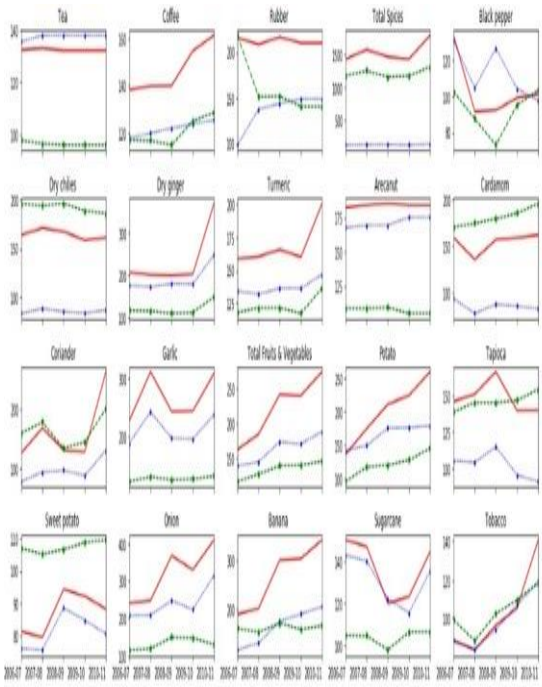

Fig.14: Output 6

V. CONCLUSION

In this review, the focus has been on science and technology, but a broad range of options including social and economic factors such as technology extension and access to technologies by farmers also needs to be pursued. Thus the system will help reduce the difficulties faced by the farmers and stop them from attempting suicides. It will act as a medium to provide the farmers efficient information required to get high yield and thus maximize profits which in turn will reduce the suicide rates and lessen his difficulties. By analyzing the issues that affect crop yield is that climatic factors such as average temperature, average humidity, average rainfall, and route map for a selected crop for particular features with date specification. This methodology to design and development of crop prediction and crop yield prediction using different machine learning techniques are used such as K-Near Neighbor, Decision Tree, Random Forest Classifier helps to increase yield and subsequent profit of agricultural production.

REFERENCES

[1] van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature 1review. Computers and Electronics in Ag.riculture, 177(January), 105709. https://doi.org/10.1016/j.compag.2020.105709.

[2] Jain, S., & Ramesh, D. (2020). Machine Learning convergence for weather based crop selection. 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2020, (February 2020). https://doi.org/10.1109/SCEECS48394.2020.75

[3] Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: A review paper. Indian Journal of Science and Technology, 9(38). https://doi.org/10.17485/ijst/2016/v9i38/95032

[4] Bouighoulouden, A. (2020). Crop Yield Prediction Using K-Means Clustering. School of Science and Engineering – Al Akhawayn University, (April), 1–35.

[5] Nishant, P. S., Sai Venkat, P., Avinash, B. L., & Jabber, B. (2020). Crop yield prediction based on indian agriculture using machine learning. 2020 International Conference for Emerging Technology, INCET 2020, (June). https://doi.org/10.1109/INCET49848.2020.9154036

[6] Sangeeta, & Shruthi, G. (2020). Design and implementation of crop yield prediction model in agriculture. International Journal of Scientific and Technology Research, 9(1), 544–549.

[7] Kulkarni, S., Mandal, S. N., Sharma, G. S.Mundada, M. R., & Meeradevi. (2018). Predictive Analysis to Improve Crop Yield using a Neural Network Model. 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018, 74–79. https://doi.org/10.1109/ICACCI.2018.8554851

[8] PALANIVEL, K., & SURIANARAYANAN, C. (2019). an Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques. International Journal of Computer Engineering and Technology, 10(3). https://doi.org/10.34218/ijcet.10.3.2019.013

[9] Journal, I. (2018). Prediction of Crop Yield using Machine Learning. International Research Journal of Engineering and Technology (IRJET) .

[10] Kale, S. S., & Patil, P. S. (2019). A Machine Learning Approach to Predict Crop Yield and Success Rate. 2019 IEEE Pune Section International Conference, PuneCon 2019, 1–5. https://doi.org/10.1109/PuneCon46936.2019.9105741

[11] Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. 2019 12th International Conference on Contemporary Computing, IC3 2019, 1–6. https://doi.org/10.1109/IC3.2019.8844901

[12] Malik, P., Sengupta, S., & Jadon, J. S. (2021). Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms. Proceedings of the Confluence 2021: 11th

International Conference on Cloud Computing, Data Science and Engineering, 1004–1007. https://doi.org/10.1109/Confluence51648.2021.9377147

[13] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697349

[14] Nisha, K. G., & Sreekumar, K. (2017). A review and analysis of machine learning and statistical approaches for prediction. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, Icicct, 135–139. https://doi.org/10.1109/ICICCT.2017.7975174

[15] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). (Paper 7 Tafur). May, 138–145.

[16] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151(November 2017), 61–69. https://doi.org/10.1016/j.compag.2018.05.012

[17] Pant, J., Pant, R. P., Kumar Singh, M., Pratap Singh, D., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46(xxxx), 10922–10926. https://doi.org/10.1016/j.matpr.2021.01.948

[18] Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. Machines, 6(3). https://doi.org/10.3390/machines6030038.

[19] Chaudhary, K. (2020). Prediction of Crop Yield Using Machine. 4(9), 153–156. www.irjet.net

[20] Ghosh, S., & Koley, S. (2014). Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks. International Journal on Recent and Innovation Trends in Computing and Communication, 2(2), 292.

[21] Gupta, J., Chauhan, A., Agarwal, A., Ranghuvanshi, A. S., & Saxena, R. (2018). Machine Learning Approach for Crop Yield Prediction and Crop Variety Recommendation in Android Application. 7(1).

[22] Bali, N., & Singla, A. (2022). Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey. Archives of Computational Methods in Engineering, 29(1), 95–112. https://doi.org/10.1007/s11831-021-09569-8