A Study on Crop Yield Prediction using Machine Learning Techniques

Kalyani M.
UG Scholar
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

Rachana R.S.
UG Scholar
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

Bhargavi K.
Assistant Professor,
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

Malashree R.
UG Scholar
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

Sindhu S.
UG Scholar
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

Praveena K.S.
Assistant Professor
Vidyavardhaka College of Engineering
Mysore, Karnataka, India

ABSTRACT
Machine learning is a powerful technique for identifying crop yields, also detecting which seeds to sow in the planting time. By studying the agricultural region, the system needs to address agricultural challenges by analyzing the agricultural region based on soil qualities and guiding farmer just on best crop to plant, allowing them to boost productivity and reduce losses. Many techniques were used to classify and predict the best suited crop for the soil. This survey provides the details on the study of different crop yield prediction techniques based on different parameters such as weather data in the past, soil parameters, and agriculture yield of previous year.

General Terms
Machine Learning, Crop Yield Prediction

Keywords
Crop Yield, Machine learning, LSTM, KNN, SVM, Random Forest

1. INTRODUCTION
Agriculture is the practice of cultivating soil, growing crops, and managing livestock. It comprises the production and sale of human-grade plant and animal products. Agriculture produces the great majority of food and textiles in the entire world. Farming is critical to the economics and progress of the country, as well as the human future. Work is, first and foremost, necessary for survival. It also provides a significant number of jobs. The demand for production has been expanding steadily over time. People are abusing technology to mass-produce massive amounts of goods. New hybrid variations are created every day. However, these variants lack crucial nutrients found in normally grown crops. These anomalies the majority of these to avoid losses, unnatural methods are employed. The loss, on the other hand, is decreased when agricultural farmers have exact information on crop output. This project will be completed. The information is provided using historical weather, temperature, and other factors. Data mining is a machine learning method for exploring and analyzing data in a variety of ways. After the data has been analyzed, it's being used to make predictions for the future. It has numerous applications. These patterns provide crop information. The project's purpose is to help producers boost their output and profit the suggested approach focuses on crop kind, yield, and weather predictions

Agriculture in India: Agriculture has been practiced in India since the Indus Valley civilization. India is the second-larger agricultural producer in entire world. In a year 2018, Farming was more than half of India's total work and given 17–18% in the entire country's GDP. In terms of net cropped area, India tops the world, followed by the United States., and India's GDP is rapidly dropping as the country's vast China. Agriculture's contribution to economic growth. Livestock management, in the other side, in our country most popular economic growth field that plays a critical part in our country social and economic fabric.

2. LITERATURE REVIEW
The prediction method shown in figure 1 is segregated into 2 phases: training phase and testing phase. All data will be collected and preprocessed throughout the training phase. The K-Means technique will be used to cluster the preprocessed data. When the number of created rules reaches a certain threshold, the training process ends. The yield value is projected during the testing phase using the generated rules. Preprocessing is the first step in the process. The data was preprocessed in this step. Some data was removed from the dataset during preparation. Some of the land will be eliminated since it is unsuitable for crop production. Basically, the four steps are used in the prediction method are Data collection, Data preprocessing, Data visualization and Exploratory data analysis (EDA). Data collection is the procedure for collecting and examining data. from various sources. It should be collected and that should be in a form that makes sense for the business problem at hand to use it to generate realistic machine learning solution. Though many methods were used for the crop yield prediction, a brief discussion of some of the most used methods is given here.

2.1 Support Vector Machine
The Support Vector Machine (SVM) is a supervised machine learning model for binary classification techniques. The purpose of this method is to generate an N-dimensional hyper plane, where N is the quantity of features in a dataset that will be used to categorize data points. a support vector machine technique, will be used to forecast agricultural production.

Support vector regression will be used to generate non-linear functions utilizing kernel functions. The radial basis function and the polynomial function are widely used parameters. Support vector regression has the advantage of avoiding the challenges of using linear functions in big datasets. input sample spaces and simplifying the optimization of difficult problems based on information from the soil, water, temperature, and rainfall, SVM is used to identify whether rice can be grown there, and a web application utilizing HTML, CSS, and JavaScript is built. We may use the web application to interact with the Machine Learning model and obtain the prediction result by providing inputs. [8]

Regardless of the number of hyper planes created, the fundamental goal of any method is to discover the plane with the largest Margin, or the greatest distance between data points of the attributes being plotted. The more distance between two points, the more precise the classification. The script also makes use of Yahoo's weather API to get the current temperature at the location, which will be one of the variables.

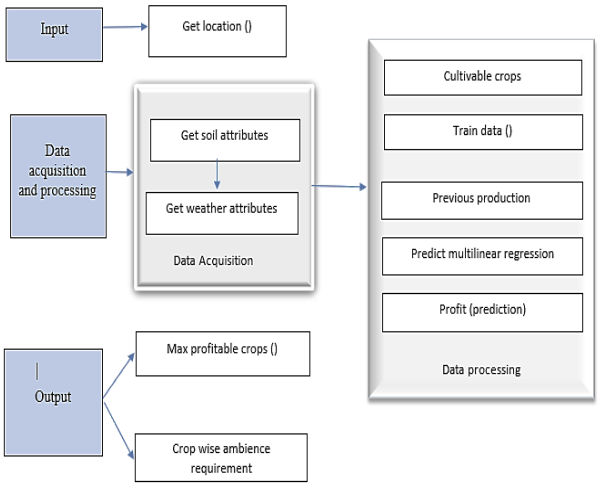

Fig 1. General Flow diagram of crop yield prediction

2.2 KNN (K- Nearest Neighbor)

KNN can be used to tackle both regression and classification problems. When KNN is used, the K-most similar group's average or median examples is used to predict regression issues. The highest frequency class from the K-most similar samples may be computed as the outcome of a KNN classification. In a nutshell, each instance votes for their class, class with the highest votes being selected as the prediction. Class probabilities may be calculated using the normalized frequency of samples belonging to each class in a collection of K most similar instances for a new data instance. For instance, consider a binary classification task (class is 0 or 1).

KNN is another name for the sample-based learning approach. To make predictions, this directly leverages the practice data. Forecasts are created for each new data point by searching the whole Summarizing the output variable across the K instances in the training set that are the K most similar examples (neighbors). This could be the mean output variable in a regression or the modal (or most prevalent) class value in a classification. A distance metric is used to determine which of the K instances in the training dataset is most similar to a new input. Euclidean distance is the most often used distance metric for real-valued input variables. Euclidean distance is calculated by taking the square root of the total of the squared differences between two points a and b.

Euclidean distance = $\sqrt{}$ [(x2 – x1)2 + (y2 – y1)2].

Various other commonly used length measurements

•Hamming distance is used to calculate distance between two binary vectors.

•Using the sum of their absolute differences, calculate the distance between actual vectors (Manhattan Distance). Another term for it is City Block Distance.

•The Makowski Distance is a combination of the Euclidean and Manhattan distances.

•KNN can be made stochastic for very large training sets by choosing a sample from the dataset and looking for the K-most comparable occurrences within it.

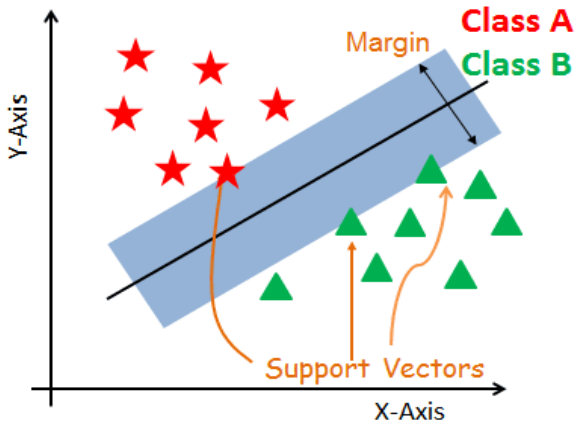

Fig 2. Classification using SVM

2.3 Random Forest Classifier

To improve the model's performance, the Random Forest technique employs numerous decision tree classifiers. The notion of ensemble learning is employed to handle complex problems in this case. It's a learning algorithm that's supervised.

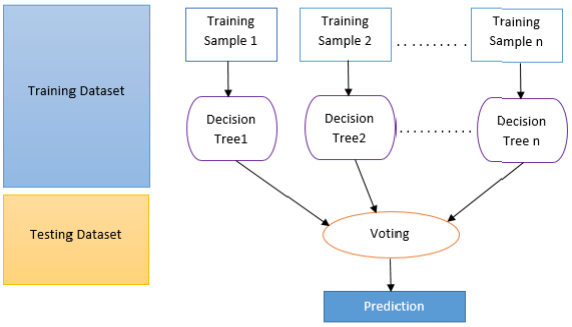

Fig 3. Random Forest algorithm [7]

Randomly generated decision trees are generated using samples from the training set. As a result, each decision tree generates a set of predictions as shown in figure 3. A majority vote determines the model's final prediction. Its popularity as a machine learning technique stems from its ability to deal with overfitting and enhance accuracy by using more trees.

Step 1: From the specified training dataset, K instances are picked at random.

Step 2: For each of the occurrences, decision trees are generated.

Step 3: The number of estimators to be developed is set to N.

Step 4: repeat step1 and step2.

Step 5: Each estimator's forecast is calculated for the new instance, and the category with the most votes is allocated.[7]

2.4 long short-term memory (LSTM)

The artificial recurrent neural network-based long short-term memory (LSTM) architecture is a deep learning architecture (RNNs). Standard feed forward neural networks, as opposed to LSTM do not contain a feedback link. LSTM is hence suitable for use in a "general-purpose computer." The crop production dataset is loaded into classification and regression algorithms to anticipate the crop's name and yield. Some of the ensemble learning approaches used are Random Forest Classifier and XG Boost, Artificial Neural Networks, Logistic Regression, Linear Regression, and KNN Classifier. The temporal features of the growth season are encoded using the LSTM model. Three LSTM layers make up the model. Each LSTM layer has 512 nodes with the activation function Leaky Elu [6].

The table 1 shows the comprehensive examination of various data sources, data sets selected, algorithms used for the prediction and the observation made on the results.

Table 1. Review of methods used in crop yield prediction

Reference	Dataset	Data source	Algorithm used	Observation
Machine learning techniques for crop yield prediction [1]	Temperature, rainfall, Production, season.	Official website of Indian government	random forest regressor algorithm Boost classifier KNN classifier algorithm Logistic regression	Experiments were conducted using a dataset given by the Indian government, and the random forest regressor was shown to have the best prediction accuracy. of 67 percent, which is higher than the Boost classifier's 63.63 percent, the KNN classifier's 43.25 percent, and the Logistic regression's 25.8%.
Machine learning techniques for crop yield prediction [2]	Crop and climate of a particular region of Maharashtra	www.data.gov.in and www.imd.gov.in are two websites that you may visit.	Random forest	Created website where the user enters information such as the district, crop, season, and hectare size, and the expected outcome is displayed. by using random forest which gives accuracy of 75% and above.
Machine learning techniques for crop yield prediction [3]	Area, production, crop name, rain fall	Data.gov.in Indian waterportal.org Power.larc.nasa.in	Logistic regression, random forest, and nave bayes	Random Forest had the best accuracy of 92.87 percent, followed by nave bayes at 91.49 percent and logistic regression at 87.81 percent among the three techniques used.
crop yields in Indian agriculture using machine learning. [4]	State, district, crop, season, region, and productivity are all factors to consider.	Indian government repository	Lasso, ENet and Kernel Ridge	Here they have applied regression stacked When those models were used singly, the outcome was improved.
Crop variety selection using Machine learning algorithms.[5]	Micro nutrients, fertilizer requirements, vulnerabilities to diseases	Online resources	Artificial neural network	choosing a crop, taking the market price into account, and choosing a crop variety are the three aspects of this algorithm.
Crop Yield Prediction and Fertilizer Efficiency [16]	State, district, crop, and season in the area	There is no specific source.	Random forest and back propagation	Analysis of crop productivity and fertilizer dose requirements
Argo-Genius: Machine Learning for Crop Prediction [26]	Crop Extent and Production, Recommended Crop, Duration of the Crop, Combination of the Crops, Current Cultivation Extent, Price	Sri Lanka's Central Agricultural Department and the Bank of Sri Lanka	GastnerNewmann Cartogram method, Auto Regressive Integrating Moving Average, Long Short-Term Memory, Linear Programming, and Auto Regressive Integrating Moving Average Model	The user enters information such as the district, crop, season, and area in hectares on a website, and the outcomes are predicted using a random forest algorithm.

	Reports, and Weather Information			
Crop Yield Prediction and Fertilizer Efficiency [16]	State, district, crop, and season in the area	There is no specific source.	Random forest and back propagation	Analysis of crop productivity and fertilizer dose requirements
Argo-Genius: Machine Learning for Crop Prediction [26]	Crop Extent and Production, Recommended Crop, Duration of the Crop, Combination of the Crops, Current Cultivation Extent, Price Reports, and Weather Information	Sri Lanka's Central Agricultural Department and the Bank of Sri Lanka	GastnerNewmann Cartogram method, Auto Regressive Integrating Moving Average, Long Short-Term Memory, Linear Programming, and Auto Regressive Integrating Moving Average Model	The user enters information such as the district, crop, season, and area in hectares on a website, and the outcomes are predicted using a random forest algorithm.
Crop Identification [27]	Color of the soil, PH, Rainfall, and Temperature Cotton, jowar, bajra, and cashew are some of the crops grown. Wheat, chickpeas, and coffee	Online Resources	Linear Support vector machine Algorithm	By providing the soil parameter, the system suggests the proper crops, probable pests that might harm the crop, and commonly used pesticides to battle the pests with an accuracy of 89.66 percent. It's conceivable to consider adding new soil attributes and expanding the data set.
Crop Yield Prediction [28]	average rainfall and production	online	The letter K stands for algorithm. The Bayes Algorithm (sometimes known as the Nave Bayes Algorithm) is a kind The Apriori algorithm is based on the linear SVM method.	Crop suggestions based on geography, rainfall, and farmer's land, as well as Yield Prediction in kilograms per acre.
Nave Bayes Map for Crop Prediction Reduce Precision on India's Region Belts [29]	Rainfall, soil air temperature, soil temperature, atmospheric pressure, moisture, and relative humidity are all factors that influence soil temperature.	sensor reports, irrigation reports	Theorem of Nave Bayes	Provide crop guidance by indicating the ideal time to harvest plants. It is also possible to have both the worst and best environmental conditions. Find out how much each crop will yield, as well as pesticide recommendations, fertilizer requirements, and irrigation requirements.
Farmers' Rice Produce Prediction Model [30]	Harvest date, Rice Seed, Scatter Date, Water Resource, and Rice Produce Quantity	Extension Service of the Department of Agriculture (DOAE)	Artificial Neural Network and Decision Tree J48, NB Tree, Random Tree, and Multilayer Perceptron	Farmers can use ANN to plan their financial strategies for their rice farms, since it has an average correctness of 81.6 percent and predicts rice production in kilograms as low, moderate, or high. Agriculturist will be

				able to calculate investment and profit while accounting for additional factors such as area size, worker cost, and material cost.
Model for Crop Prediction Using Optimization Techniques [31]	Humidity, total rainfall, temperature, and output yield are all factors to consider.	a cultural food organization	Grey Wolf Optimization and SVM	In terms of extracting feature vectors with the least amount of error, converge, and categorize as good/bad yield, SVM-GWO delivers a greater classification accuracy.
Predictions for Automated Farming. [32]	Potatoes, rice, jute, and wheat, to name a few. Average rainfall, average temperature, and crop output rate	Bangladesh Bureau of Department of Agricultural Extension, Statistics, and Agricultural Information Service and Yearly Books of Agricultural statistics,	Multiple linear regression method, K closest neighbors' algorithm	The system will determine which of the six crops is appropriate given the location of the land and the date on which the crop will be planted by determining the region, yield per unit area for that specific year, and region. In addition, depending on the option chosen, sufficient resources for the entire agricultural process, as well as prescribed timeframes, are necessary. In terms of accuracy, MLR surpassed KNN, hence it was used in the Android app.
Predictive Analytic Techniques for Yield Prediction [33]	Production in tonnage, crop, and rainfall are all factors to consider.	There is no specific source.	Linear regression, logistic regression, and ridge regression are three types of regression.	Production in tones and area in hectares have a 0.3305138 connection., indicating that as area is maximized, Crop yield or production in agriculture rises.

Table 2. Comparison Of Various algorithms

Reference	Methods	Accuracy
[1]	`SVM	78%
	RNN-LSTM	70%
[2]	Random forest	67.80%
	XG Boost	63.63%
	KNN	43.25%
	Logistic Regression	25.81%
[3]	Random forest	75%
[4]	KNN	85%
	Decision tree	80%
	Naïve bayes	82%
	SVM	78
[11]	SVM	92.6%
	Decision tree	99.8%
	KNN	99.7%
	Random Forest	81.7%
[12]	SVM	86.9%
	Random Forest	92%
[13]	Linear Regression (Rice)	90%
	Neural Networks	95%
[14]	SVM	60%
	Random Forest	93%
	KNN	86%
	Decision Tree	93.3%

3. OBSERVWTION AND COMMENTS

- Models with more and fewer features should be compared to see which one performs the best. Several algorithms have been used in different studies.
- While no conclusive conclusion can be drawn regarding which model is the best, the findings do demonstrate that some machine learning algorithms are used more frequently than others.
- Most of the studies employed a variety of machine learning models to discover which model had the best forecast.
- The most utilized models are Linear Regression, Gradient Boosting Tree, Random Forest, and Neural Networks.
- Under deep learning techniques, long short-term memory (LSTM) architecture is found to be effective in predicting the better crop for the land.

4. CONCLUSION

In this paper, a review is performed on different machine learning algorithms used for crop yield prediction. Over thirty-three articles were chosen for analysing the performance of different algorithms implemented, based on different parameters such as weather data in the past, soil parameters, and agriculture yield of previous year. Even though many algorithms were employed to foresee the crop that will suit the land the best, there is still a scope for

improvement with respect to intercrop prediction, maximum yield analysis and estimated budget for the crop growth.

5. REFERENCES

[1] Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.

[2] Mayank Champaneri, Chaitanya Chandvidkar, Darpan Chachpara, Mansing Rathod. "CROP YIELD PREDICTION USING MACHINE LEARNING". *International Journal of Engineering Research & Technology (IJERT).*

[3] Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams "Crop Yield Prediction using Machine Learning" Algorithms. *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181*

[4] Potnuru Sai Nishant ,Pinapa Sai Venkat , Bollu Lakshmi Avinash , B. Jabber "Crop Yield Prediction based on Indian Agriculture using Machine Learning". *2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020*

[5] G. Vishwa1, J. Venkatesh , Dr. C. Geetha "Crop Variety Selection Method using Machine Learning". International Journal of Innovations in Engineering and Technology (IJIET) http://dx.doi.org/10.21172/ijiet.124.05

[6] Sharma, S., Rai, S., & Krishnan, N.C. (2020). Wheat Crop Yield Prediction Using Deep LSTM Model. ArXiv, abs/2011.01498.

[7] "Crop prediction using machine learning" Madhuri Shripathi Rao et al 2022 *J. Phys.: Conf. Ser. 2161 012033. doi:10.1088/1742-6596/2161/1/012033*

[8] "Crop Yield Prediction Using Linear Support Vector Machine" N.Manjunathan1 , P.Rajesh2 , E. Thangadurai3 , A. Suresh4. *European journal of molecular and clinical medicine, ISSN 2515-8260 volume7, Issue 06 2020.*

[9] M Suganya, Dayana R, Revathi R."CROP YIELD PREDICTION USING SUPERVISED LEARNING TECHNIQUES". *International Journal of Computer Engineering & Technology (IJCET).*

[10] P Priya, U Muthaiah and M Balamurugan. "Predicting Yield of The Crop Using Machine Learning Algorithm", *International Journal of Engineering Sciences and Research Technology (IJESRT) ISSN: 2277-9655, April 2018.*

[11] Ekaanshkhosla, Ramesh Dharavath, Rashmi Priya. "Crop yield prediction using aggregated rainfall based modular artificial neural networks and support vector regression". *Published online 21aug 2019 springer nature B.V 2019.*

[12] Published online 21aug 2019 springer nature B.V 2019. "A data mining approach to crop yield prediction using machine learning". *PalArch's journal of archaeology of Egypt/Egyptology 17(12). ISSN 1567 – 214x.*

[13] Dr V Latha Jothi, Neelambi gai A, Nithish Sabari S, Santhosh K. "Crop yield prediction using KNN model". *International journal of engineering research and technology (IJERT) ISSN: 2278-0181 RTICCT-2020.*

[14] Ashwin Kowshik, Kishore Gowda HK, Rithik Somesh BR, Yashas S, Dr. Ramesh B, Nithyashree R. "Crop yield prediction based on Indian agriculture using machine learning". *International journal of multidisciplinary research and growth evaluation, volume 2, issue 4, July - Aug 2021 (230-233).*

[15] S Bharath, Yeshwanth S, Yashas B L, Vidyaranya R Javalagi, 2020, Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCETESFT – 2020 (Volume 8 – Issue 14)*

[16] S. Bhanumathi M Vineeth N Rohith, crop yield prediction and efficient use of fertilizers, *IEEE(2019).*

[17] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber' Crop Yield Prediction based on Indian Agriculture using Machine Learning.

[18] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique."*PalArch's journal of archaeology of Egypt/Egyptology 17(12). ISSN 1567 – 214x*

[19] Andrew Crane "Machine learning methods for crop yield prediction and climatechange impact assessment in agriculture" *International journal of engineering research and technology (IJERT) ISSN: 2278-0181 RTICCT*

[20] Ayush Shah, Akash Dubey, VisheshHemnani, Divye Gala and D. R. Kalbandeet.Smart Farming System: Crop Yield Prediction Using Regression Techniques.

[21] Bharath S, Yeshwanth ,Yashas B L and Vidyaranya R Javalagi 2020 Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction *International Journal of Engineering Research & Technology (IJERT) NCETESFT – 2020 vol 8 Issue 14.*

[22] Gulati P and Jha S K 2020 Efficient crop yield prediction in India using machine learning techniques *International Journal of Engineering Research & Technology (IJERT) ENCADEMS – 2020 vol 8 Issue 10.*

[23] Gupta A, Nagda D, Nikhare P, Sandbhor A, 2021, Smart crop prediction using IoT and machine learning International *Journal of Engineering Research & Technology (IJERT) NTASU – 2020 vol 9 Issue 3.*

[24] Suresh A, Ganesh P and Ramalatha M 2018 Prediction of major crop yields of Tamilnadu using K-means and Modified KNN *2018 3rd International Conference on Communication and Electronics Systems (ICCES) pp 88-93 doi: 10.1109/CESYS.2018.8723956.*

[25] Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning *2020 International Conference for Emerging Technology (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.*

[26] Anupama, C. G., & Lakshmi, C. (2021). A comprehensive review on the crop prediction algorithms. *Materials Today: Proceedings*.

[27] Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network approach *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering vol 2 Issue 1 pp 683*-6.

[28] Gamage, A., &Kasthurirathna, D. (2019). Agro-Genius: Crop Prediction Using Machine Learning. *International Journal of Innovative Science and Research Technology, 4(10).*

[29] Kumar, A., Sarkar, S., & Pradhan, C. (2019, April). Recommendation system for crop identification and pest control technique in agriculture. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0185-0189). IEEE.

[30] Bhosale, S. V., Thombare, R. A., Dhemey, P. G., & Chaudhari, A. N. (2018, August). Crop yield prediction using data analytics and hybrid approach. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1-5). IEEE..

[31] Priya, R., Ramesh, D., & Khosla, E. (2018, September). Crop prediction on the region belts of India: a Naïve Bayes MapReduce precision agricultural model. In *2018 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 99-104). IEEE.

[32] Sharma, S., Rathee, G., & Saini, H. (2018, December). Big data analytics for crop prediction mode using optimization technique. In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 760-764). IEEE.

[33] Siddique, T., Barua, D., Ferdous, Z., & Chakrabarty, A. (2017, September). Automated farming prediction. In *2017 Intelligent systems conference (IntelliSys)* (pp. 757-763). IEEE.

[34] Nagini, S., Kanth, T. R., &Kiranmayee, B. V. (2016, December). Agriculture yield prediction using predictive analytic techniques. In *2016 2nd International Conference on (IC3I)* (pp. 783-788). IEEE.