A Crop Recommendation System to Improve Crop Productivity using Ensemble Technique

Shikha Ujjainia, Pratima Gautam, S. Veenadhari

Abstract*: An integration of technology with crop yielding prediction methodology brought a major transformation in the production level globally. Machine learning concept has boosted that technology in such a manner that has further optimized the situation of farmer and agricultural industry. The combination of different types of algorithm enhances the competency of technological device to a level where the prediction becomes very effective and least deviation can be expected from the agricultural industry in the production level. The research of machine learning states about the integration of three types of models which is usually followed separately in programming the device. The study has proved the intervention of Information technology in the agricultural industry via different functions. An effective prediction by using the ensemble algorithm makes the agricultural industry competent enough to maintain the expected amount of production of crop.

***Keywords* : Machine learning, Ensemble algorithm, Integration, Agricultural Industry, Crop Prediction.**

I. INTRODUCTION

Technology has become a key element for industries to operate the business function in an effective manner. Likewise, the agricultural industry is not apart from the effect of technology and emphasized various agricultural organizations and farmers to use smart and advanced technology. Consequently, the production levels of crops are growing day by day with the increasing population of the world. However, the industry is constantly facing the issue of predicting the bio system that helps in the crop production and machine learning might resolve the issue. The purpose of this research paper is to provide a resolution in terms of technology via ESM algorithm in order to achieve the right way of predicting the crop yielding in a global as well as industrial context. The objective of this research paper would be to propose the algorithm that gives an overview of better ways of machine learning.

II. LITERATURE REVIEW

A. Requirement of integrating the ensemble algorithm in agriculture

According to [1], the population is growing at a very high rate which increases the probability of shortage of food items in the whole world because of huge consumption and less production of crops. Hence, it becomes essential for the agricultural industry to implement a technology that gives a boost to the production of crops along with quality of the food items. On the other hand, an optimum environment of unified bio-system is required for crop yielding prediction so that proper approach can be followed while producing the crops. From the bio-system of crops, it can be said that the interactions between, soil moisture, air, temperature due to sunlight and crop is highly complex. Further, the complexity enhances the likelihood of false prediction of parameters for the crops especially when no technology is used. The attributes of Ensemble algorithm technology covers up those loopholes of predicting the effect on the basis of existing parameters of the bio-system surrounding a crop. The most important feature of ensemble algorithms is the collection of various models used in machine learning. Consequently, the efficiency of artificial intelligence of the detector is very high to analyze and collect the accurate data. The industrialization is another factor that is creating problems in agriculture that consecutively affect the crop prediction as well as production negatively [2]. Due to expansion of industries in the global context, the population of the workforce involved in agriculture started degrading. As a result, the work on crop production declined in the last consecutive years. The workforce involved currently in the crop production is not able to predict the crop yielding if the artificial intelligent detector is not used in the process.

B. Importance of Ensemble algorithm in crop yielding

In order to yield the crop in a rapid process, it becomes very essential to maintain the quality of the crop. On the contrary, the limited workforce of the agricultural industry would not be able to manage quantity and quality altogether in the global context. The optimization of crop yielding prediction methodology makes it an important task for farmers and related industry for the better decision making process. Hence, the optimization process can be further improved by incorporating the technology in the crop prediction methodology. The importance of ensemble algorithms is observed in the case of technology where it covers up a major part of decision making for the crop production [3]. An integration of ensemble algorithms is actually making the device intelligent enough to record the data and analyze it in a more efficient manner. Consequently, the crop that will be yielded by the farmer would not get harmed by the biosystem.

As mentioned by [4], the collection and analysis of data crop and biosystem comprises categorization which can be segmented on the basis of three types of models.

Manuscript received on January 30, 2021.
Revised Manuscript received on February 03, 2021.
Manuscript published on February 28, 2021.

* Correspondence Author

Shikha Ujjainia*, Computer Science and Engineering, Rabindranath Tagore University, Bhopal, India. Email: shikhaujjainia90@gmail.com

Pratima Gautam, Computer Science and Information Technology, Rabindranath Tagore University, Bhopal, India. Email: pratima_shkl@gmail.com

S. Veenadhari, Computer Science and Engineering, Rabindranath Tagore University, Bhopal, India. Email: veenadhari1@gmail.com

Every model analyses the data in different manner, for example, some algorithm says about the right or wrong data, some algorithm simplifies the information by considering it's hidden structure. On the other hand, Ensemble algorithm is a combination of all the models and therefore, it contains the attribute to analyze the data on all types of aforementioned structure. As opined by [5], the agricultural industry has not put extra effort on making the electronic detector that detects the various parameters for the production of crops. Consequently, it saves a lot of the cost of the agricultural industry and farmers by preventing them from more investment on the crop yielding prediction technique if ensemble algorithm is incorporated. Hence, the ensemble algorithm is very important from the perspective of saving the cost, data quality and effort of the farmer to produce crops.

III. METHODOLOGY

A. Random Forest Algorithm

As opined by [6], Random Forest algorithm is implemented to generalize the unseen data by using the training data set and functions. Hence, a random forest algorithm follows the supervised model that is based on the hidden data and training of function by the data set. Although, the algorithm follows the combinations of models and the concept is known as bagging method which is implemented to increase the result. An ensemble or combinations of models refers to the decision tree through which the decision making process can be classified into multiple times. When a range of data is collected by estimating the unseen data then it becomes difficult for analysts to find the correct one. However, a random forest algorithm overcomes this regression problem for collecting the output variable. Likewise, the attributes of multiple models consisting of random forest manages the classification problem.

B. Gradient Boosting Algorithm

The term boosting in the Gradient Boosting algorithm indicates that transformation of machine learning from weak to strong. The concept of Gradient Boosting Algorithms (GBA) is very much similar to AdaBoost Model because the GBA is by nature gradual and sequential to find the error in the classification just like AdaBoost model [7]. The differences between GBA and AdaBoost models can be observed in the process of identifying the weak decision tree of an algorithm. GBA uses its gradient rather than weight to recognize the weak learner that is the decision tree of the loss function. Crop parameters are optimized using the loss function of the gradient boosting algorithm.

C. Gaussian Process Algorithm

The Gaussian process algorithm is the most effective method to estimate the data over the functions through which an estimation for accurate data is done by the analyst [8]. Likewise, the Bayer's rule is used also by the analyst with the help of training data to find the correct value of a biosystem's parameter. As the concept Gaussian process emphasizes the probability distribution over the function, it consequently resolves the regression problem and classification problem of the decision making trees.

D. Proposed Workflow of the Model

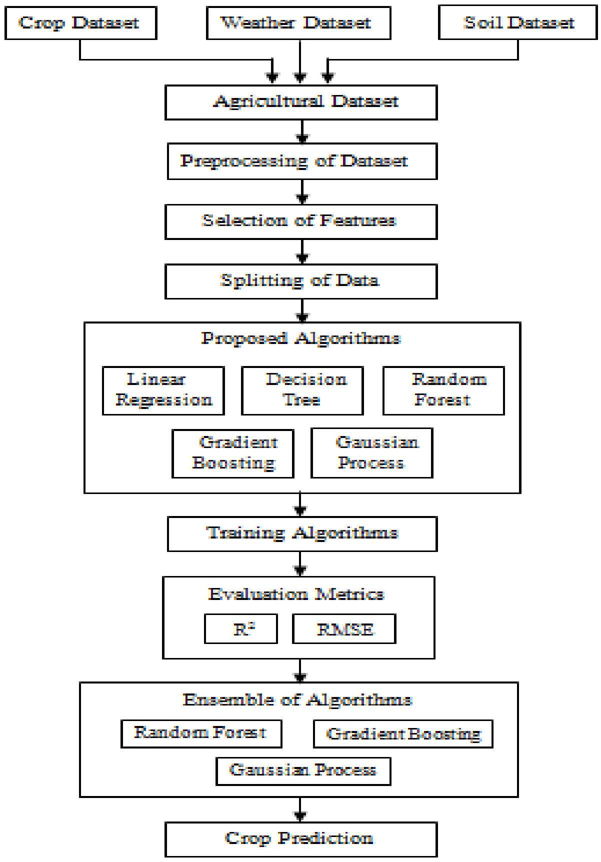

Fig. 1.Working flow of the module

From the above flowchart, it can be seen that various datasets (crop dataset, weather dataset, soil) are combined together to form an agricultural dataset. In the next level, the preprocessing of the dataset has been done to compensate for the missing data in the range of the dataset. After preprocessing, the input has to be given to the respective modules by selecting the parameters (features) in the algorithm. In addition, the data available in the dataset is divided into training and testing datasets. In the next step, different types of algorithms are proposed and machines are to be trained through models. The performance of each proposed model is evaluated by R-squared and RMSE matrices. Therefore it was found that the random forest, gradient boosting, and gaussian process algorithms are performing well and giving high accuracy. Therefore, in the next step, we ensemble these three algorithms to estimate the crop yield.

E. Ensemble Framework

The ensemble model used in this study incorporates decisions from multiple models to improve overall performance. In this study, three diverse algorithms such as random forest, gradient boosting, and Gaussian process were combined to improve model performance. Many predictions are made for each data point on average. In this approach, we take an average from all the models and use it to make the final predictions as shown in Figure 2. The averaging can be used to make predictions for regression or when calculating the probability of classification problems.

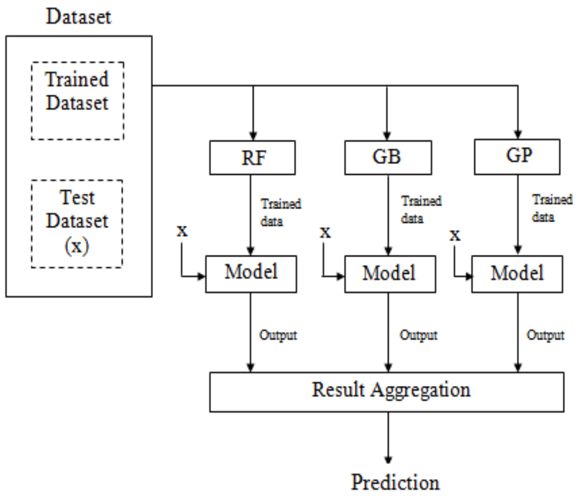

Fig. 2.Working process of the ensemble algorithms

From Figure 2 above, it is clear that the dataset is divided into two parts. Training datasets are used to train three diverse models while test data is used to examine model performance to check if it is predicting the correct outcome on a given training dataset. The outputs of all models are collected and averaged to produce a final prediction.

IV. RESULTS AND VIZUALIZATION

The performance of the model is evaluated by the coefficient of determination (R2), and root mean squared error (RMSE). R2 is a method used for the statistical measure which showing the variance's proportion for the dependent variable. Furthermore, the variance proportion explained by the variables or the independent variable in the regression model [9]. RMSE is the root mean square error used for the error prediction. The values given in Table 1 represent the R2 and RMSE values of different crops used for the ensemble model.

Table-I: Evaluation of accuracy and error of different crops

Crop Name	R^2	RMSE
Ground Nut	0.89	0.047
Maize	0.86	0.066
Moong	0.84	0.074
Onion	0.83	0.070
Rice	0.85	0.073
Sesamum	0.82	0.045
Sugarcane	0.96	0.030
Tur	0.80	0.067
Urad	0.84	0.062
Wheat	0.90	0.081

Once the model is developed, we can see our results with the help of any visualization technique such as graphs, charts etc. Figure 3 show the webapp used for crop prediction analysis where different parameter of bio system as an input and an output will be generated according to the real scenario. The input contains area (in hectare), temperature (in Celsius), rain (in mm), groundwater (in meter), Soil ph, potassium, magnesium and sodium (in kg/hectare).

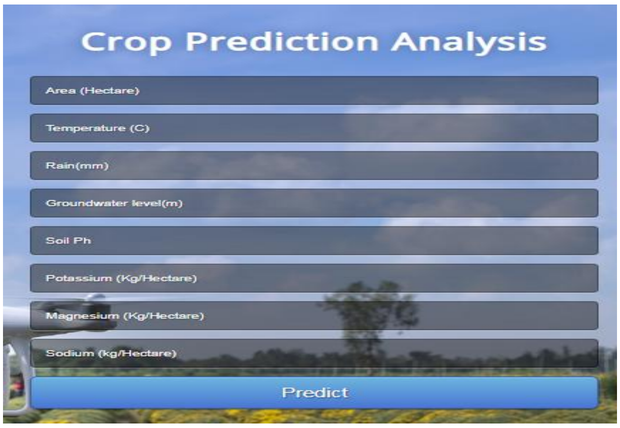

Fig. 3.**Crop Prediction interface with different parameters.**

Figure 4 show that the optimum values of parameters of bio system to produce the onion at the maximum level. It has been predicted that 96408.38 amount of onion will be produced if harvesting is done in the current scenario.

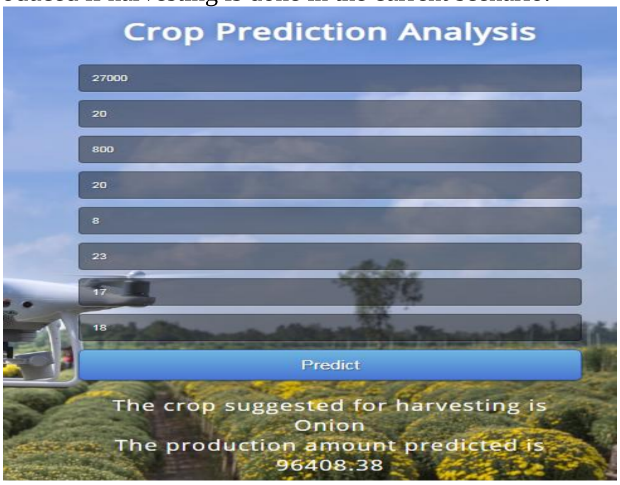

Fig. 4.**Prediction of onion crop and its production with different input parameters.**

Figure 5 shows the optimum values for producing the Maize and 313652.4 amount of Maize will be produced if harvesting is done under the provided circumstances of bio system.

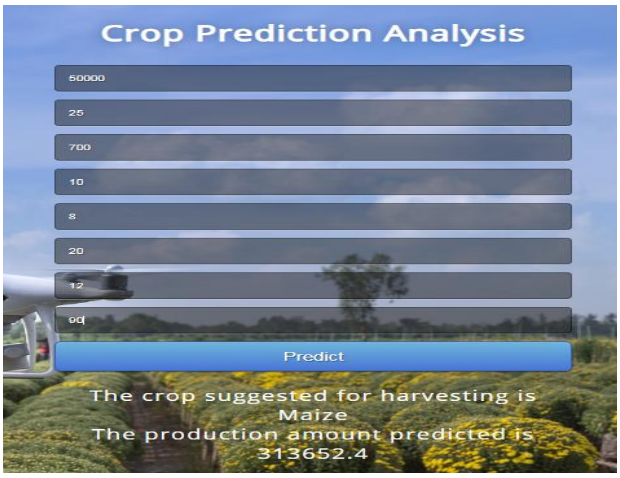

Fig. 5. Prediction of maize crop and its production with different input parameters.

V. CONCLUSION

From the above analysis, it has been concluded that technology has achieved that level of competency by which agriculture can easily predict the crop yielding production. Machine learning technology proposed the model integrated form of concept called as ensemble method that eliminated the loopholes existing in the prediction process for crop. By evaluating the different parameters of biosystem it has been understood that the technology used for making the crop yielding prediction device is very much diversified. The parameters of biosystem vary with respect to changing location and a single concept of algorithm is not sufficient to fulfill the requirement of crop prediction. Therefore, ensemble algorithm found to be the best technique for crop prediction if consider the device to be used in the global context.

REFERENCES

1. A. Nigam, S. Garg, A. Agrawal, & P. Agrawal, “Crop Yield Prediction Using Machine Learning Algorithms”, In 2019 Fifth International Conference on Image Information Processing (ICIIP), IEEE , pp. 125-130, November 2019.
2. T. van Klompenburg, A. Kassahun, & C. Catal, “Crop yield prediction using machine learning: A systematic literature review”, Computers and Electronics in Agriculture, vol. 177, p. 105709, October 2020.
3. R. A. Medar, & V. S. Rajpurohit, “A survey on data mining techniques for crop yield prediction”, International Journal of Advance Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59-64, September 2014.
4. A. J. Challinor, J. M. Slingo, T. R. Wheeler, & F. J. Doblas-Reyes, “Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles”, Tellus A: Dynamic Meteorology and Oceanography, vol. 57, no. 3, pp. 498-512, January 2005.
5. A. Ceglar, and L. Kajfež-Bogataj, "Simulation of maize yield in current and changed climatic conditions: addressing modelling uncertainties and the importance of bias correction in climate model simulations", European Journal of agronomy, vol. 37, no. 1, pp. 83-95, February 2012.
6. (Open Source Article) “A complete guide to the random forest algorithm”, Built In, Available: https://builtin.com/data-science/random-forest-algorithm
7. (Open Source Article) “Understanding Gradient Boosting Machines”, Available: https://towardsdatasecience.com/understanding-gradient-boosting-machines-9be756fe76ab
8. (Open Source Article) “An intuitive guide to Gaussian processes”, Medium, Available: https://towardsdatascience.com/an-intuitive-guide-to-gaussian-processes-ec2f0b45c71d
9. P. Vidyullatha, & D. R. Rao, “Machine learning techniques on multidimensional curve fitting data based on R-square and chi-square methods”, International Journal of Electrical and Computer Engineering, vol. 6, vo. 3, pp. 974-979, June 2016.

AUTHORS PROFILE

Shikha Ujjainia received B.Sc. degree in Computer Science from Career College, Bhopal, in 2011 and Post Graduation Degree in Master of Computer Application from Samrat Ashok Technological Institute (SATI), Vidisha, in 2014. She is currently undergoing Ph.D. programme with Rabindranath Tagore University, Bhopal, Madhya Pradesh. Her research work & interest includes machine learning, data analytics and data mining. As a research scholar, two research papers has been published in reputed journal. She is also actively involved towards online workshops, seminars, and conferences based on IoT, Machine Learning, and Big Data. She has also actively took part in similar programs organized by the university.

Dr. Pratima Gautam has PhD in Computer Application from Department of Computer Application, Maulana Azad National Institute & Technology, She is currently working as Professor in Department of Computer Science & Information Technology, Rabindranath Tagore University, Bhopal, India. With more than sixteen years of teaching experience. She has authored several research articles in International journals of repute. Besides having presented papers in several international/ national conferences. She has been invited as an expert to various national conferences as paper reviewer/ program technical committee member. She has delivered lectures in various institutions and has also participated in various training programs and attended several workshop. Her research interests include Data Mining, Machine Learning, Soft Computing and image processing. She is also affiliated with international societies like IEEE, , ACM digital library etc She has associated herself in guiding several under graduate and post graduate students in their projects and is currently providing Ph.D. supervision to six research scholars. She is also actively involved in Institutional activities like organizing Conferences/ Workshops etc.

Dr. S. Veenadhari completed her Doctoral programme from Mahatma Gandhi Chitrakoot Gramodaya Vishwavidyalaya in 2015, Master of Computer Applications from Nagarjuna University, Andhra Pradesh in 1998 and Master of Technology in Computer Science and Engineering from Makhan Lal Chaturvedi University, Bhopal in 2007. Over 15 years of academic and research experience with 45 research papers published in International and National reputed journals. Six students are pursuing their Doctoral programme and two students completed their doctoral degree under her supervision. Her research interests includes Machine leaning, Big data analytics and Cloud computing. Her expertise in agricultural data analytics through machine learning. She has published several book chapters, technical bulletins, project reports, working papers , training and teaching reference manuals. She has delivered number invited talks in many national platforms of repute. She has worked as Expert member in various committees in different institutions and member of different professional societies like IE, CSI, ISTE.