Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)
IEEE Xplore Part Number: CFP21K25-ART

Crop Recommender System Using Machine Learning Approach

SHILPA MANGESH PANDE[1], DR. PREM KUMAR RAMESH[2], ANMOL[3], B.R AISHWARYA[4], KARUNA ROHILLA[5], KUMAR SHAURYA[6]

[1]Associate Professor, Department of Information Science and Engineering

[1]Research Scholar VTU-CMRIT-CSE Research Centre

[2]Professor, Department of Computer Science and Engineering

[3,4,5,6]Student, Department of Computer Science and Engineering

[1,2,3,4,5,6]CMR Institute of Technology, Bengaluru, India and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

EMAIL: shilpa.p@cmrit.ac.in[1], premkumar.r@cmrit.ac.in[2], anmolmehta57@gmail.com[3], braishwarya878@gmail.com[4], m123kr@gmail.com[5], kshaurya8@gmail.com[6]

2021 5th International Conference on Computing Methodologies and Communication (ICCMC) | 978-1-6654-0360-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCMC51019.2021.9418351

***Abstract* — Agriculture and its allied sectors are undoubtedly the largest providers of livelihoods in rural India. The agriculture sector is also a significant contributor factor to the country's Gross Domestic Product (GDP). Blessing to the country is the overwhelming size of the agricultural sector. However, regrettable is the yield per hectare of crops in comparison to international standards. This is one of the possible causes for a higher suicide rate among marginal farmers in India. This paper proposes a viable and user-friendly yield prediction system for the farmers. The proposed system provides connectivity to farmers via a mobile application. GPS helps to identify the user location. The user provides the area & soil type as input. Machine learning algorithms allow choosing the most profitable crop list or predicting the crop yield for a user-selected crop. To predict the crop yield, selected Machine Learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbour (KNN) are used. Among them, the Random Forest showed the best results with 95% accuracy. Additionally, the system also suggests the best time to use the fertilizers to boost up the yield.**

Keywords— Crop Yield Prediction, Machine Learning, Random Forest, Crop Recommender System, Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Multivariate Linear Regression (MLR), Fertilizer

I. INTRODUCTION

Agriculture has an extensive history in India. Recently, India is ranked second in the farm output worldwide [15]. Agriculture-related industries such as forestry and fisheries contributed for 16.6% of 2009 GDP and around 50% of the total workforce. Agriculture's monetary contribution to India's GDP is decreasing [1]. The crop yield is the significant factor contributing in agricultural monetary. The crop yield depends on multiple factors such as climatic, geographic, organic, and financial elements [6]. It is difficult for farmers to decide when and which crops to plant because of fluctuating market prices [7]. Citing to Wikipedia figures India's suicide rate ranges from 1.4-1.8% per 100,000 populations, over the last 10 years [15]. Farmers are unaware of which crop to grow, and what is the right time and place to start due to uncertainty in climatic conditions. The usage of various fertilizers is also uncertain due to changes in seasonal climatic conditions and basic assets such as soil, water, and air. In this scenario, the crop yield rate is steadily declining [2]. The solution to the problem is to provide a smart user-friendly recommender system to the farmers.

The crop yield prediction is a significant problem in the agriculture sector [3]. Every farmer tries to know crop yield and whether it meets their expectations [4], thereby evaluating the previous experience of the farmer on the specific crop predict the yield [3]. Agriculture yields rely primarily on weather conditions, pests, and preparation of harvesting operations. Accurate information on crop history is critical for making decisions on agriculture risk management [5].

In this paper, we have proposed a model that addresses these issues. The novelty of the proposed system is to guide the farmers to maximize the crop yield as well as suggest the most profitable crop for the specific region. The proposed model provides crop selection based on economic and environmental conditions, and benefit to maximize the crop yield that will subsequently help to meet the increasing demand for the country's food supplies [8]. The proposed model predicts the crop yield by studying factors such as rainfall, temperature, area, season, soil type etc. The system also helps to determine the best time to use fertilizers. The existing system which recommends crop yield is either hardware-based being costly to maintain, or not easily accessible. The proposed system suggests a mobile-based application that precisely predicts the most profitable crop by predicting the crop yield. The use of GPS helps to identify the user location. The user provides an area under cultivation and soil type as inputs. According to the requirement, the model

predicts the crop yield for a specific crop. The model also recommends the most profitable crop and suggests the right time to use the fertilizers.

The major contributions of the paper are enlisted below,

1. Prediction of the crop yield for specific regions by executing various Machine Learning algorithms, with a comparison of error rate and accuracy.
2. A user-friendly mobile application to recommend the most profitable crop.
3. A GPS based location identifier to retrieve the rainfall estimation at the given area.
4. A recommender system to suggest the right time for using fertilizers.

The organization of the rest of the paper is as follows. Section II discusses the background work of researchers in the field of agriculture and yield prediction. Section III presents the proposed model for yield prediction and recommends which crop for cultivation. The model also suggests the best suitable time for the use of fertilizers. Section IV discusses the results and Section V concludes the paper.

II. Related Work

The steps taken to boost agriculture primarily involves ingraining technological expertise and inventions to make the agriculture sector more proficient and simplified for farmers by predicting the correct crops using all ML approaches. The paper discusses various algorithms such as ANN, Fuzzy Network, and various data mining techniques with their advantages. Further challenge is to have all these incorporated real-time datasets [9].

One of the early works developed a dedicated website to assess the impact of weather parameters on crop production in the identified districts of Madhya Pradesh [10]. The districts were selected on the basis of the region covered by the crop. Based on these criteria, the first five top districts with a maximum crop area were chosen. The basis of the crops selected for the study was on prevailing crops in the selected districts. The crops picked included maize, soybean, wheat and paddy, for which the yield for a continuous period of 20 years of knowledge, were tabulated. The accuracy of the established model ranged from 76% to 90% for the chosen crops with an average accuracy of 82%.

Another important work checks the soil quality and predicts the crop yield along with a suitable recommendation of fertilizers [11]. The Ph value and the location from the user were inputs used in this model. An API was used to predict the weather, temperature for the current place. The system used both supervised as well as unsupervised ML algorithms and compares the results of the two.

A classifier that uses a greedy strategy to predict the crop yield was proposed in [12]. A decision tree classifier that uses an attribute has been shown to yield better results. An ensemble model proposed suggests integrating the effects of different models, which has been shown to be typically better than the individual models. Random forests ensemble classification uses multiple decision tree models to predict the crop yield. The data are split up into two sets, such as training data and test data, with a ratio of 67% and 33%, with which the mean and standard deviation are calculated. This work also incorporates the clustering of similar crops to get the most accurate results.

Extensive work has been done, and many ML algorithms have been applied in the agriculture sector. The biggest challenge in agriculture is to increase farm production and offer it to the end-user with the best possible price and quality. It is also observed that at least 50% of the farm produce gets wasted, and it never reaches the end-user. The proposed model suggests the methods for minimizing farm produce wastage. One of the recent works presents a model where the crop yield is predicted using KNN algorithms by making the clusters. It has been shown that KNN clustering proved much better than SVM or regression [13].

In [17] predicts the crop yield for the specific year with the help of advanced regression techniques like Enet, Lasso and Kernel Ridge algorithms. The Stacking regression helped to enhance the accuracy of the algorithms.

The historical datasets are filtered to retrieve the datasets for Maharashtra state using Pandas profiling tool. The crop yield prediction model is designed using multilayer perceptron neural network and enhanced the accuracy by adjusting bias, weight and Adam optimizer. The proposed model uses ANN with three-layer neural network to predict the crop yield [18].

Supervised learning approach is used to implement crop yield prediction system. Established the correlation between multiple attributes selected from the historical which helps the system to increase the crop yield [19]. Rainfall and temperature are two factors which influence the crop yield. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) algorithms applied on these time series data to enhance the accuracy [20]. ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average) and ARMAX (ARMA with exogenous variables) methods are used to predict the temperature and rainfall using historical data. The best model among them is used in the crop yield prediction system implemented with fuzzy logic. Cloud cover and evapotranspiration are exogenous variables used in the proposed system [21].

III. Models and Methodology

Despite many solutions that have been recently proposed, there are still open challenges in creating a user-friendly application with respect to crop recommendation. The solution proposed here aims to solve these limitations, by developing a

user-friendly application that considers the parameters like rainfall, temperature, soil type etc. that directly affect cultivation. The main objective is to obtain a better variety of crops that can be grown over the season. The proposed system would help to minimize the difficulties faced by farmers in choosing a crop and maximize the yield in effect to reduce the suicide rates [16].

The proposed model predicts the crop yield for the data sets of the given region. Integrating agriculture and ML will contribute to more enhancements in the agriculture sector by increasing the yields and optimizing the resources involved. The data from previous years are the key elements in forecasting current performance. Historical data is collected from various reliable sources like data.gov.in, kaggle.com, and indianwaterportal.com. The data sets are collected for Maharashtra and Karnataka regions. The data has various attributes like state, district, year, season, type of crop, an area under cultivation, production, etc. The soil type is an attribute in other datasets with state and districts specification. This soil type column is extracted and merged into the main data set. Similarly, temperature and average rainfall are taken from a separate dataset and added to the main data sets for the specific region. The data sets are cleaned and pre-processed. The null values are replaced with mean values. The categorical attributes are converted into labels before processing the algorithms. The one hot encoding method is used to deal with categirial values in the data sets.

Figure 1 is the system architecture of the proposed model. It's a mobile app that has two modules – the prediction module and the fertilizer module. Mobile Application offers multiple services. The farmer needs to register with the app through the registration process.Once the registration is complete, the farmer can use the mobile application services. The prediction module predicts the crop yield using the selected attributes from the data sets for the specific crop. The predict module also suggests the farmer with the highest yield crops. The fertilizer module guides the farmer for the right time to use the fertilizer.

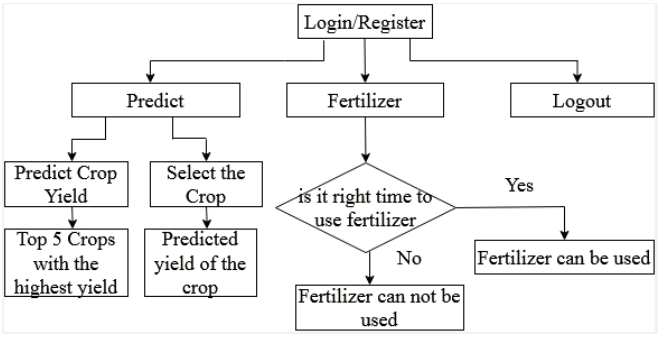

Fig. 1 System Architecture

Figure 2 illustrates the flow chart of the proposed system. It describes the whole process starting with the registration and various services provided by the mobile application.

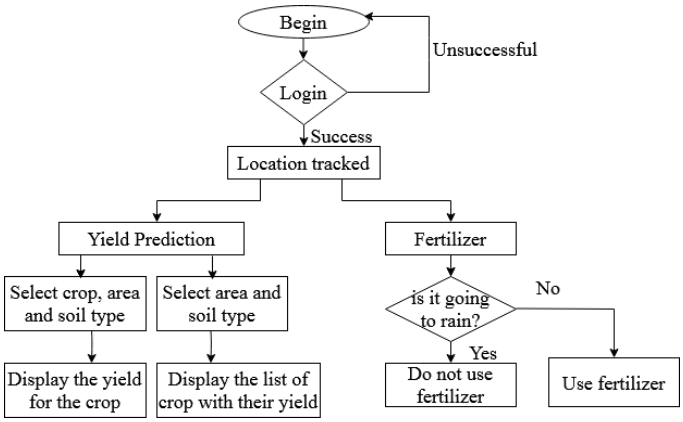

Fig. 2 Flow Chart

The very first step to use the services of the app is to register. During registration, the app locates the geographical location and identifies the region of the farmer using GPS. On successful login, the user can avail of two services. The first service is the yield prediction either for the selected crop or using a crop recommender system. The second service is the identification of the correct time to use the fertilizer. In the prediction service, the user needs to input the planned crop, soil type, and area under cultivation. The system predicts the yield for the specific crop selected. Figure 3 demonstrates the registration process to avail of the services of the app.

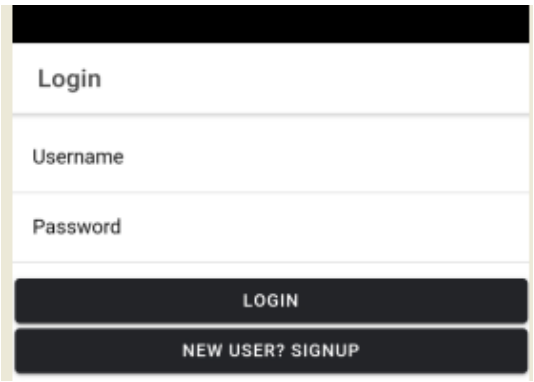

Fig. 3 Registration Process

If the farmer is not sure about the crop to be planned this year, he can use the crop recommender system. In the crop recommender system, the farmer must provide only soil type and area. The system lists the crops with their predicted yield. This makes farmers easy to decide on a crop to be planted.

The timing of applying the fertilizer is very crucial. The farmer's effort and money will get wasted if the rain comes down too early. The proposed fertilizer usage service will guide the farmer on when to use the fertilizer. The model predicts the rain for the specific location for the next 14 days with Open Weather API. If the rainfall is more than 1.25 mm then it recommends as 'not safe' to use the fertilizers.

Figure 4 demonstrates the Block diagram of Experimental Implementation. The Graphical User Interface for the proposed model is developed with the Ionic Framework with JavaScript, Angular JS, and ReactJS. The system is built and deployed across multiple platforms such as iOS, Android, desktop, and the web as a Progressive Web Apps-all with one code base [14]. The datasets and resources required for the system are hosted on firebase.

The machine learning approach is used for crop yield prediction. The patterns and correlations are discovered using ML approach. The model is trained using historical data sets where the past experience is used to represent the outcome.

Various standard machine learning algorithms are used to predict yield. Among the selected algorithms, the Random Forest regression provided the best accuracy. Random Forest builds many decision trees and then blends them together to make the most accurate and stable predictions.

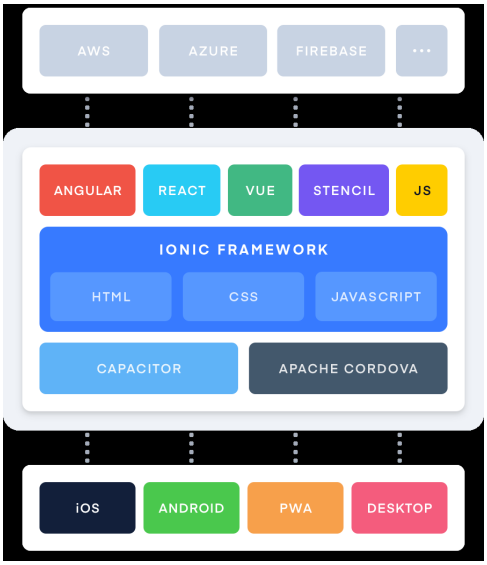

Fig. 4 Block diagram of Experimental Implementation

IV. RESULTS AND DISCUSSIONS

This section discusses the results deduced from selected algorithms for Maharashtra and Karnataka regions. The parameters used for algorithms are crop type, year, season, soil type, area, and region. For all the selected algorithms, the accuracy of the crop yield prediction is compared. Random Forest algorithm proved to be the best for the given data set with an accuracy of 95%. To predict the crop yield, selected ML algorithms such as ANN, SVM, Multivariate Linear Regression, Random Forest, and KNN are used. Table1 shows the tabulated results of the accuracy comparison of various ML algorithms. Figure 5 shows the graphical representation of the results.

TABLE I : Accuracy vs Algorithm

Algorithm	Accuracy (%)
Artificial Neural Network (ANN)	86
Support Vector Machine (SVM)	75
Multivariate Linear Regression (MLR)	60
Random Forest (RF)	95
K Nearest Neighbor (KNN)	90

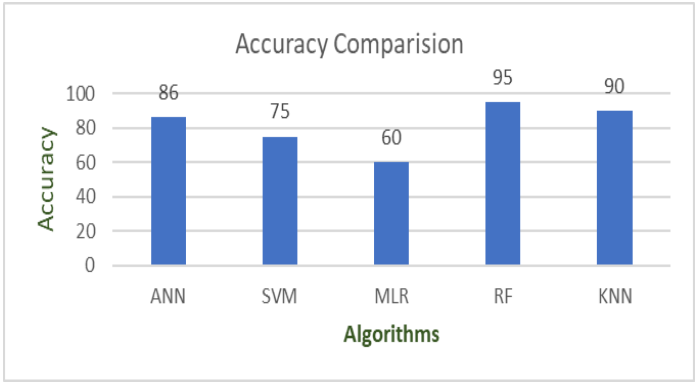

Fig. 5: Accuracy vs Algorithm

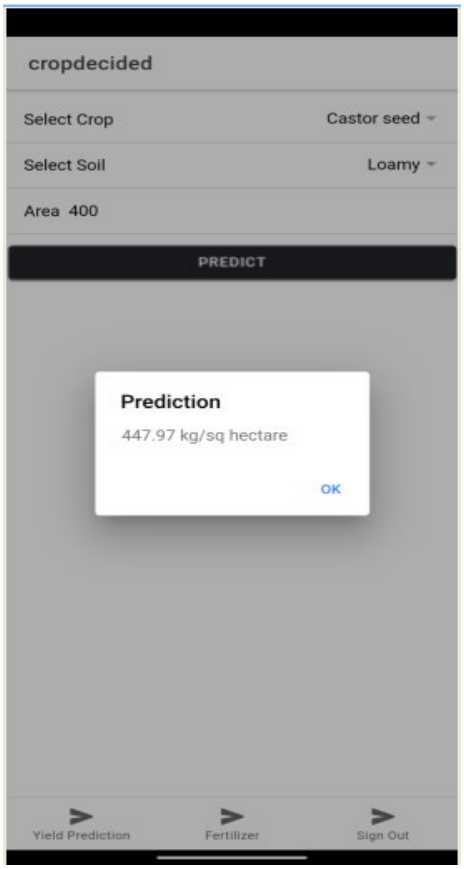

Fig. 6: Crop Yield Prediction for a specific crop

Option1: The user upfront knows the crop to be planned for this season and interested to understand the possible yield. A user will select a crop along with associated parameters such as soil type and area. The predictor block internally uses the Random Forest Algorithm to predict the crop yield for a user decided crop. Figure 6 above is a snapshot of a result predicted

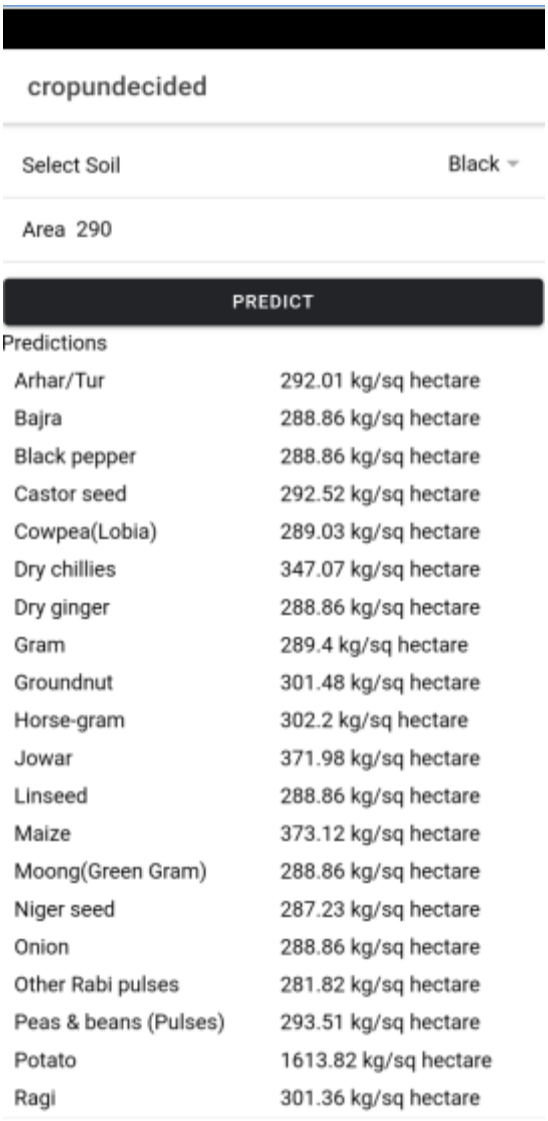

Fig. 7: The Crop Recommender system

Option2: The farmer chooses the recommender system when the user is not sure which crop to plan this year. Figure 7 shows the recommendation for the various crops based on soil type and area. Users can select from the predicted recommended list. Another feature is to get the right time for a farmer to apply the fertilizers. The system checks the weather for the next 14 days and suggests the right time to use the fertilizers, as demonstrated in Figure 8.

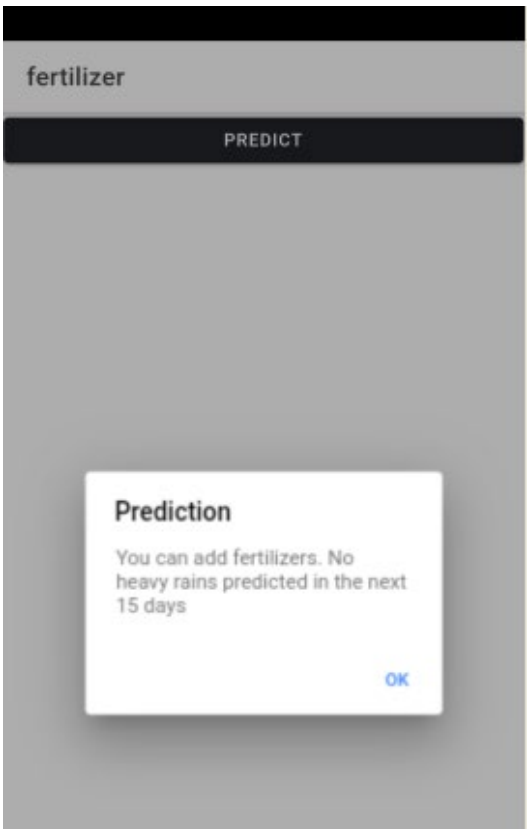

Fig. 8: Fertilizer Timing

V. CONCLUSION

This paper highlighted the limitations of current systems and their practical usage on yield prediction. Then walks through a viable yield prediction system to the farmers, a proposed system provides connectivity to farmers via a mobile application. The mobile application includes multiple features that users can leverage for the selection of a crop. The inbuilt predictor system helps the farmers to predict the yield of a given crop. The inbuilt recommender system allows a user exploration of the possible crops and their yield to take more educated decisions. For yield to accuracy, various machine learning algorithms such as Random Forest, ANN, SVM, MLR, and KNN were implemented and tested on the given datasets from the Maharashtra and Karnataka states. The various algorithms are compared with their accuracy. The results obtained indicate that Random Forest Regression is the best among the set of standard algorithms used on the given datasets with an accuracy of 95%. The proposed model also explored the timing of applying fertilizers and recommends appropriate duration.

The future work will be focused on updating the datasets from time to time to produce accurate predictions, and the processes can be automated. Another functionality to be implemented is to provide the correct type of fertilizer for the

given crop and location. To implement this thorough study of available fertilizers and their relationship with soil and climate needs to be done. An analysis of available statistical data needs to be done.

VI. ACKNOWLEDGEMENT

The author is thankful to all authors who have participated to conduct the experiment and research in the proposed research work. We are thankful to the CMR Institute of Technology for providing all resources and computing facilities to conduct this research.

REFERENCES

[1] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ, "BIoT: Blockchain-based IoT for Agriculture", 11th International Conference on Advanced Computing (ICoAC), 2019 Dec 18 (pp. 324-327). IEEE.

[2] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development, 2018;2:46-66

[3] Manjula E, Djodiltachoumy S, "A model for prediction of crop yield" International Journal of Computational Intelligence and Informatics, 2017 Mar;6(4):2349-6363.

[4] Sagar BM, Cauvery NK., "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science, 2018 Dec;12(3):1087-93.

[5] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, "Big data in smart farming–a review. Agricultural Systems", 2017 May 1;153:69-80.

[6] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Godfray HC, Herrero M, Howitt RE, Janssen S, Keating BA, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural systems", 2017 Jul 1;155:269-88.

[7] Johnson LK, Bloom JD, Dunning RD, Gunter CC, Boyette MD, Creamer NG, "Farmer harvest decisions and vegetable loss in primary production. Agricultural Systems", 2019 Nov 1;176:102672.

[8] Kumar R, Singh MP, Kumar P, Singh JP, "Crop Selection Method to maximize crop yield rate using a machine learning technique", International conference on smart technologies and management for computing, communication, controls, energy, and materials (ICSTM), 2015 May 6 (pp. 138-145). IEEE.

[9] Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudharshan K S, Vasanth S, Harish Kumar N, "A Survey on Crop Prediction using Machine Learning Approach", In International Journal for Research in Applied Science & Engineering Technology (IJRASET**),** April 2019, pp(3231-3234)

[10] Veenadhari S, Misra B, Singh CD, "Machine learning approach for forecasting crop yield based on climatic parameters", In 2014 International Conference on Computer Communication and Informatics, 2014 Jan 3 (pp. 1-5). IEEE.

[11] Ghadge R, Kulkarni J, More P, Nene S, Priya RL, "Prediction of crop yield using machine learning", Int. Res. J. Eng. Technol. (IRJET), 2018 Feb;5.

[12] Priya P, Muthaiah U, Balamurugan M, "Predicting yield of the crop using machine learning algorithm", International Journal of Engineering Sciences & Research Technology, 2018 Apr;7(1):1-7.

[13] S. Pavani, Augusta Sophy Beulet P., "Heuristic Prediction of Crop Yield Using Machine Learning Technique", International Journal of Engineering and Advanced Technology (IJEAT), December 2019, pp (135-138)

[14] https://web.dev/progressive-web-apps/

[15] https://www.wikipedia.org/

[16] Plewis I, "Analyzing Indian farmer suicide rates", Proceedings of the National Academy of Sciences, 2018 Jan 9;115(2): E117.

[17] Nishant, Potnuru Sai, Pinapa Sai Venkat, Bollu Lakshmi Avinash, and B. Jabber. "Crop Yield Prediction based on Indian Agriculture using Machine Learning." In *2020 International Conference for Emerging Technology (INCET)*, pp. 1-4. IEEE, 2020.

[18] Kale, Shivani S., and Preeti S. Patil. "A Machine Learning Approach to Predict Crop Yield and Success Rate." In 2019 IEEE Pune Section International Conference (PuneCon), pp. 1-5. IEEE, 2019.

[19] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741. IEEE, 2020.

[20] Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019.

[21] Bang, Shivam, Rajat Bishnoi, Ankit Singh Chauhan, Akshay Kumar Dixit, and Indu Chawla. "Fuzzy logic based crop yield prediction using temperature and rainfall parameters predicted through ARMA, SARIMA, and ARMAX models." In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pp. 1-6. IEEE, 2019.