**How to Cite**:
Pachade, R. S., & Sharma, A. (2022). Machine learning for weather-specific crop recommendation. *International Journal of Health Sciences*, *6*(S8), 4527–4537. https://doi.org/10.53730/ijhs.v6nS8.13222

# Machine learning for weather-specific crop recommendation

**Rahul S. Pachade**
Research Scholar, Madhyanchal Professional University, Ratibad, Bhopal. 462044, MP, India
Corresponding author email: hellorahulsp@gmail.com.

**Dr. Avinash Sharma**
Professor Dept Computer Science and Engineering, Madhyanchal Professional University, Ratibad, Bhopal -.462044, MP, India
Email: avinashvtp@gmail.com

***Abstract*---**Agriculture and its related sectors are unquestionably the most important sources of income in rural India. Additionally, having a big impact on the nation's GDP is the agriculture sector. The sector of agriculture is so large, which is great for the nation. The crop production per hectare, however, falls short of international standards. This is one of the most likely causes of the greater rate of suicide among marginal farmers in India. This research proposed best recommendation system. The proposed system recommends crops for farmers to grow based on input from the farmer's field, such as the temperature, soil, moisture, and nutrient like NPK, pH, and rainfall. Machine learning algorithms allow for optimal crop selection to be made in light of all relevant parameters. three popular machine learning algorithms were tested in this study which includes the Decision Tree, the Random Forest (RF), and the Logistic Regression. The Random Forest among them demonstrated the highest outcomes with 99.32% accuracy.

***Keywords*---**decision tree, random forest, crop yield, soil parameters, crop recommendation.

## Introduction

In the past, farmers in a certain area used their knowledge and experience to predict and forecast crops. They choose a crop that has been grown in the past, in a nearby area, or that is popular in the area, but they only grow it where they live because they don't know how to manage soil nutrients. Because of changes in technology, there are now many ways to deal with agricultural inputs. In the field


*Manuscript submitted: 9 May 2022, Manuscript revised: 18 July 2022, Accepted for publication: 27 August 2022*

of agriculture, a significant amount of work has been put in, and numerous machine learning (ML) methods have been implemented. When it comes to agriculture, one of the most challenging problems to tackle is how to boost agricultural productivity while keeping both pricing and quality at the maximum level that is practicable for the end customer. Agriculture's biggest difficulty is a lack of awareness regarding climate change. Every crop has its own ideal climate. This requires precision farming practises. Precision farming sustains crop productivity and boosts yield.

The cost of onions in our nation fluctuated last year. Many farmers in our nation decided to plant onions because of the rising price. While there was a plentiful supply of onions in some regions of Maharashtra, crop production was failed in others. This resulted from unfavourable growing circumstances. The lives of the common people were negatively impacted by a subsequent shortage in the months that followed because middle-class people couldn't buy onions, an essential food.

The crop recommendation model lets farmers choose crops based on soil and environmental conditions. It also helps farmers get the most out of their crops, which will help meet the growing demand for food in the country. The appropriate recommendation model figures out how much a crop will produce by looking at things like soil type, soil moisture, rainfall, soil pH, temperature, region, season, etc. It also helps figure out when the best time is to use fertilisers. The current crop production prediction method is either hardware-based, expensive to maintain, or difficult to use. Despite numerous new solutions being put out, there are still unresolved issues with regards to developing a user-friendly application for crop recommendation. Integrating agriculture and machine learning will lead to further advancements in the agriculture industry by raising yields and optimising resources. The information from prior years is essential for anticipating present performance. The methods utilised by the various system include increased crop yield, real-time crop analysis, parameter selection, wiser decision making, and improved crop yield.

The farmer's family's financial stability could be put in jeopardy if a decision is made without sufficient consideration. It's possible that this is one of the factors contributing to the high number of farmer suicides that have been reported in the media. Because agriculture contributes 18.8% to India's GDP, a poor choice would be detrimental to the country's economy as well as the nation's farmers. We have uncovered a dilemma for farmers regarding the best crop to cultivate. Use of IoT and machine learning technologies, which, when properly deployed, have the ability to change the game in a variety of different industries. The results of this study will show how these technologies can be utilised successfully to provide farmers with as much assistance as possible on crop ideas.

**Related work**

Machine learning algorithms might quickly and readily tackle big non-linear problems if they were given datasets from a variety of different sources. Machine learning provides better decision making in real-world circumstances with minimal need for human interaction. It offers a robust structure for decision-making that is driven by data from the past. A comprehensive knowledge of crops, soil, and the surrounding environment is necessary for the development of a crop recommendation system. [1] [9]. Nutrients, pH, and soil colour are all characteristics of a given soil [7] [3] [14] [8] [6]. The most common crop in the location should be selected as the sample crop [7] [3] [14]. Implementing precision agriculture (PA) can increase farm productivity, which might have a significant positive influence on farmers' livelihoods, the agricultural industry, and the economy as a whole. Through the use of PA, researchers have worked to develop strategies that can help farmers choose crops that will maximise their yield [18]. The proposed study method utilises chemical processes, sensors, and fertilisers to restore phosphorous, potassium, and nitrogen levels in the soil, hence promoting plant development and maximizing yield.

An extensive range of PA system has been developed by a number of authors as part of an effort to assist farmers in selecting the crop which is most suited to their soil in accordance with the characteristics of their soil [3, 7], [8]. The popular of PA systems that have been created over the course of time can be categorised as crop recommendation (CR) systems due to the fact that they recommend certain crops for individual fields depending on the characteristics of those fields [9].

Recommendation systems. are computer programmes that may determine which products will be most beneficial to individual consumers by analysing the user's past actions, product data, and interactions with the products [9]. In order to anticipate a user's interests, these systems employ a wide range of data processing techniques to examine the user's past activity and the actions of users who are similar to the user. There have been a number of attempts made by different authors to find a solution to this issue so that the results can be more accurate [10] [11] [12] [13]. They do this by utilising opinion mining techniques in conjunction with neural networks in order to get around the cold-start problem that is inherent with content-based recommendation systems.

Some crop recommendation systems are based on machine learning models [3] [15] [16] [7], while others are based on probabilistic models [2], artificial intelligence models [8], and so on. The machine learning-based models are by far the most common of these types of models. These models make use of machine learning methods in order to tackle the classification problem at hand, which is the classification of soil. This work [19] proposes a wireless sensor network for multi-parameter precision agriculture monitoring. Proposed IoT infrastructure leverages low-power Intel Galileo Gen-2 (IoT). Sensor nodes acquire field data. Each node transmits data wirelessly to the base station. The PC's screen displays temperature, humidity, sprinkler flow, and soil moisture. IoT collects data for control and decision-making.

Machine learning allows computers to learn without specific programming requirements [17]. Machine learning-based soil classification methods are preferred for CR systems [14]. In order to forecast crop yield with the available data, the authors of the study [16] use nave Bayes (NB)(accuracy98%) and k-nearest neighbours (KNN) to identify the soil. They also imply that more effective models can be created with the use of classification methods like support vector machines. This is supported by [15], who build their CR system with the help of kernel-based support vector machines (SVM), and claim an accuracy of 94.95%. When comparing SVM with RF-based models for soil classification, however, [3] discovered that the RF-based models performed better. They found that SVM had an accuracy of 75.73 percent while RF had an accuracy of 86.35 percent.

In this study [22], researchers compared the accuracy of three different algorithms—Weighted K-NN, Support Vector Machines, and Bagged Tree—for classifying soil, and found accuracy of SVM better. The authors of this paper [21] compare the accuracy of three different methods for making crop recommendations: Decision Tree (90.20%), Random Forest (90.43%), and K- NN (90.78%). K-nearest neighbour found better among the decision tree and RF.

In this work [20], a model is put out for predicting the type of soil and recommending a crop that can be grown there. The model has been put to the test using different machine learning methods, including kNN, SVM, and logistic regression, and it has acquired the highest level of accuracy. SVM's classification accuracy reached its highest level of accuracy, or 96.00%. The KNN, which stands for the K-Nearest Neighbor algorithm, is the basis for the crop prediction system that is suggested in this work [23]. As input into the model are things like the amount of nitrogen, phosphorus, and potassium in the soil, as well as the pH value. Cross-Validated kNN accuracy is 88%, kNN accuracy is 85%, Decision Tree accuracy is 81%, Naive Bayes accuracy is 82%, and SVM accuracy is 78%.

**Methodology**

The categorization of soil, which forms the basis of our crop recommendation system, is the focus of this research, in which various machine learning algorithms, including decision trees, random forests, and logistic regression, are put through their paces. According to the results obtained, the RF machine learning method in particular appears to provide the best fit. The construction of our CR system then makes use of the RF method in conjunction with a number of other machine learning strategies, such as SVM. Soil data, which includes details about nutrients like NPK and pH, is used to train these algorithms. The development of a crop recommendation system also takes into account meteorological factors including temperature, humidity, and rainfall. Figure 1 presents the flow of the procedures that were followed while carrying out the task in order to get the desired results.

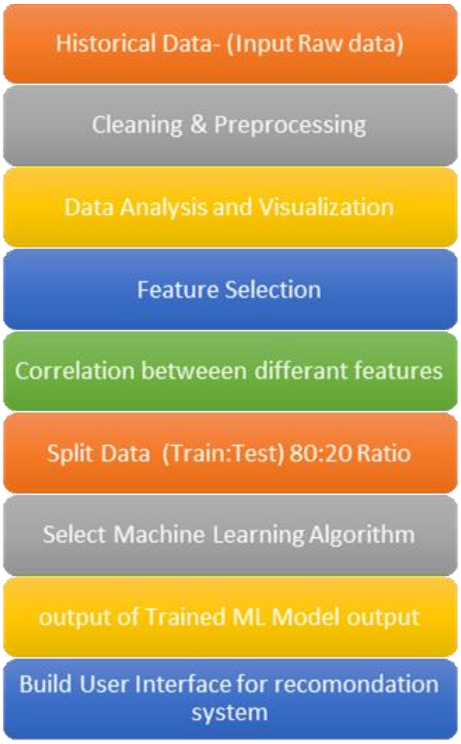


Figure 1 flow of Crop Recommendation Implementation

The first step is to verify the integrity of the dataset we will be working with. All of the gaps in the dataset must be filled with acceptable replacements. In addition, the data should be examined to verify if its characteristics follow a normal distribution. It's important to filter out the extreme cases. In addition, the data should be examined to verify if its characteristics follow a normal distribution. It's important to filter out the extreme cases. Features with a high skew value should be transformed to normalise them. Because of skewness in the features used in the dataset, we were unable to use it. In order to standardise our dataset's features, we applied a quantile transformation. In order to better comprehend our dataset, we have developed a number of visual representations. figure 2 represents crop summery in bar plot and figure 3 shows the NPK requirement comparative graph for 22 crops. To better understand the information at hand, we've made a number of charts (bar graphs, scatter plots, box plots, etc.) to see if we can spot any patterns or trends that will prove useful during implementation.

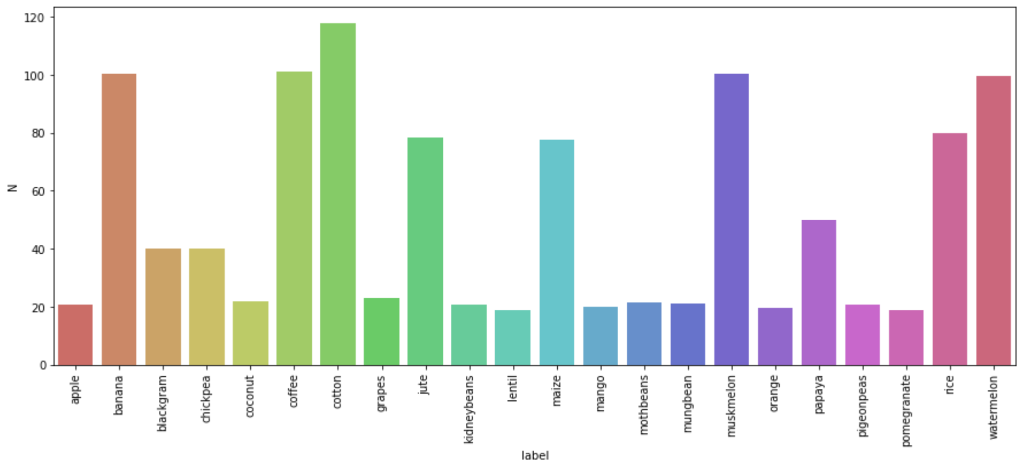


Figure 2 Crop summery

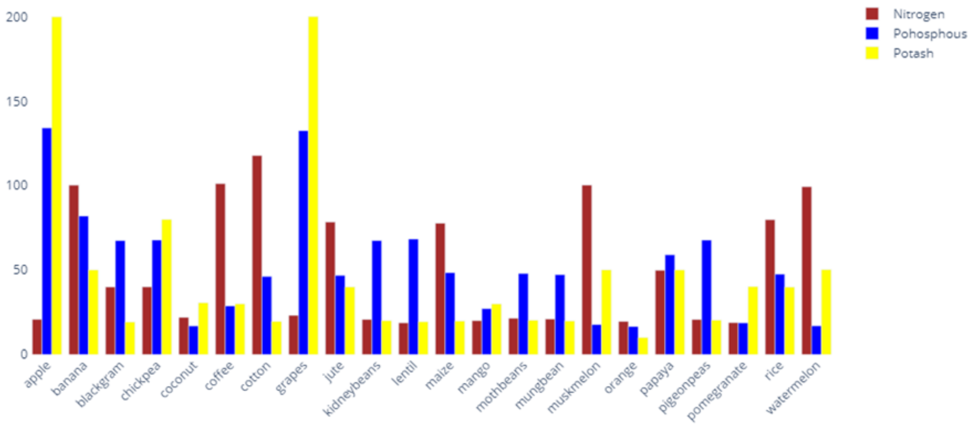


Figure 3 Comparison of NPK requirement for crops

It's critical that we limit our choices to those that will actually assist in deciding the type of crop to grow. We created a correlation matrix that graphically depicts the linear association between any two features in order to achieve this. Since the features in figure 4 are correlated with one another, using all of them to determine the type of crop to grow makes sense rather than omitting any of them.

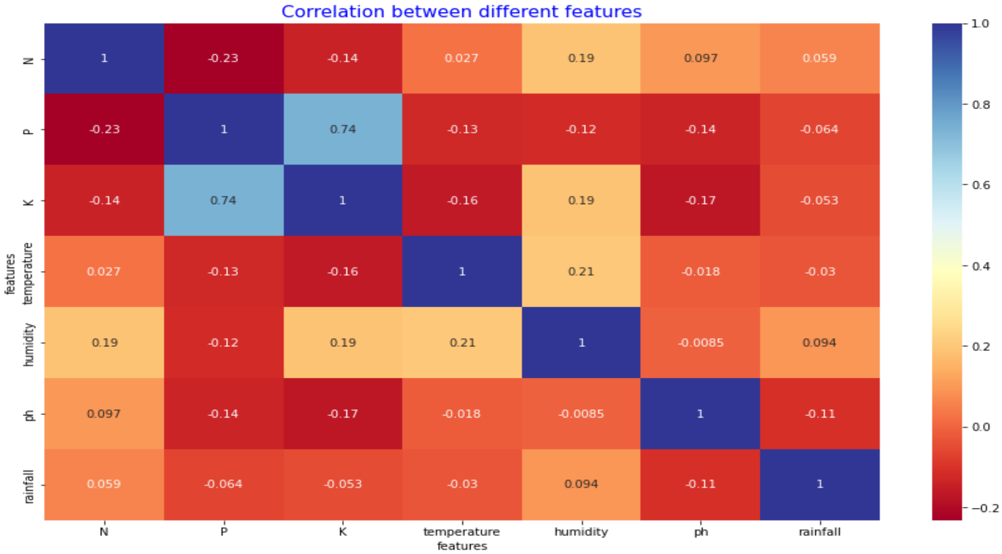


Figure 4 Correlation matrix between different features

**Results and Discussions**

We used machine learning algorithms following data cleansing and visualisation. Logistic Regression, Random Forest(RF), and Decision Tree are the three methods that were used. The features of the dataset are depicted in Figure 4.

Table 1 Summery of Classification report

| | Decision Tree | | | Random Forest | | | Logistic Regression | | |
|---|---|---|---|---|---|---|---|---|---|
| Label | Precision | Recall | f1-score | Precision | Recall | f1-score | Precision | Recall | f1-score |
| apple | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| banana | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| blackgram | 1 | 0.95 | 0.98 | 1 | 1 | 1 | 0.86 | 0.82 | 0.84 |
| chickpea | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| coconut | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| coffee | 1 | 0.88 | 0.94 | 1 | 1 | 1 | 1 | 1 | 1 |
| cotton | 1 | 1 | 1 | 1 | 1 | 1 | 0.89 | 1 | 0.94 |
| grapes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| jute | 1 | 1 | 1 | 0.95 | 1 | 0.98 | 0.84 | 1 | 0.91 |
| kidneybeans | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| lentil | 1 | 1 | 1 | 0.94 | 1 | 0.97 | 0.94 | 0.94 | 0.94 |
| maize | 0.86 | 1 | 0.92 | 1 | 1 | 1 | 0.94 | 0.89 | 0.91 |
| mango | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| mothbeans | 0.96 | 0.96 | 0.96 | 1 | 0.96 | 0.98 | 0.88 | 0.92 | 0.9 |
| mungbean | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| muskmelon | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| orange | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| papaya | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.95 | 0.98 |
| pigeonpeas | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| pomegranate | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| rice | 1 | 1 | 1 | 1 | 0.96 | 0.98 | 1 | 0.84 | 0.91 |
| watermelon | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Following the application of the three algorithms to the dataset, we are able to observe that Random Forest provides the highest accuracy out of the three algorithms. The classification report presented in table 1 summarising precision, recall, and f1 score for each algorithm used. RF achieved highest accuracy of 99.32%, followed by Decision Tree which provided an accuracy of 99.09%, and Logistic Regression which provided the lowest accuracy score: 96.82%. Python is utilised in the development of a user interface for the purpose of inputting crop recommendation parameters. Figures 5 and 6 display the outcomes of the crop recommendation system depending on the input parameter respectively. In the process of developing the Graphical user Interface, the normal range of soil nutrients is taken into consideration. In further iterations of the system, both the type of soil and the soil nutrients (NPK) will be taken into consideration in order to construct a more precise recommendation.

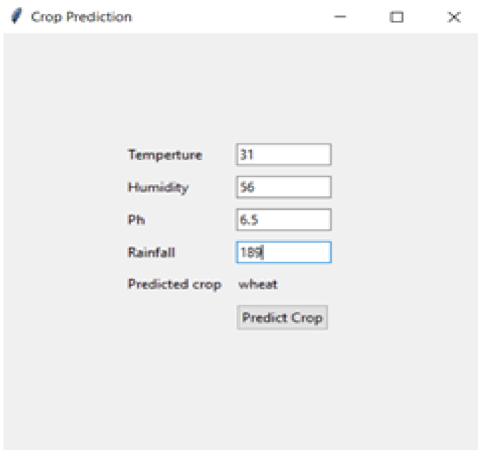


Figure 5 Prediction result for wheat

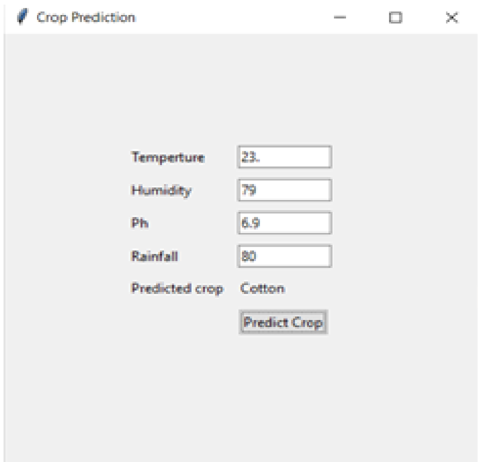


Figure 6 Prediction result for Cotton

## Conclusion

Adopting cutting-edge technologies would simplify the farmer's life. Because cultivating a crop requires in-depth knowledge and comprehension of a wide variety of factors, such as soil, temperature, and pH, amongst others, it makes perfect sense for the farmer to take advantage of new technologies. Through the use of machine learning algorithms, random forest algorithm able to determine with high precision which crop would yield the best results. Since the system only displays results for 22 different crops, in the near future, all of the data that can be collected from crops will be taken into consideration for the purpose of making the system even better and assisting farmers in making conclusions about which crops to cultivate.

## References

1. A. Kumar, S. Sarkar, and C. Pradhan, “Recommendation system for crop identification and pest control technique in agriculture,” in 2019 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2019, pp. 0185–0189.
2. A. L. Samuel, “Some studies in machine learning using the game of checkers,” IBM Journal of research and development, vol. 3, no. 3, pp. 210–229, 1959.
3. A. M. Patokar and V. V. Gohokar, "Automatic Investigation of Micronutrients and fertilizer dispense System using Microcontroller," *2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)*, 2018, pp. 1737-1739, doi: 10.1109/ICRIEECE44171.2018.9008500.
4. A. Sang and S. K. Vishwakarma, “A ranking based recommender system for cold start & data sparsity problem,” in 2017 Tenth International Conference on Contemporary Computing (IC3). IEEE, 2017, pp. 1–3.
5. D. A. Bondre and S. Mahagaonkar, “Prediction of crop yield and fertilizer recommendation using machine learning algorithms,” International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371–376, 2019.
6. G. Suresh, A. S. Kumar, S. Lekashri, and R. Manikandan, “Efficient crop yield recommendation system using machine learning for digital farming,” International Journal of Modern Agriculture, vol. 10, no. 1, pp. 906–914, 2021.
7. J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, “Collaborative filtering and deep learning based recommendation system for cold start items,” Expert Systems with Applications, vol. 69, pp. 29–39, 2017.
8. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” Sensors, vol. 18, no. 8, p. 2674, 2018.
9. K. K. Fletcher, “A method for dealing with data sparsity and cold-start limitations in service recommendation using personalized preferences,” in 2017 IEEE international conference on cognitive computing (ICCC). IEEE, 2017, pp. 72–79.
10. K. Patel and H. B. Patel, “A state-of-the-art survey on recommendation system and prospective extensions,” Computers and Electronics in Agriculture, vol. 178, p. 105779, 2020.

11. M. Paul, S. K. Vishwakarma, and A. Verma, “Analysis of soil behaviour and prediction of crop yield using data mining approach,” in 2015 International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2015, pp. 766–771.
12. M. Shinde, K. Ekbote, S. Ghorpade, S. Pawar, and S. Mone, “Crop recommendation and fertilizer purchase system,” International Journal of Computer Science and Information Technologies, vol. 7, no. 2, pp.665–667, 2016.
13. N. Kumbhar and K. Belerao, “Microblogging reviews based crosslingual sentimental classification for cold-start product recommendation,” in 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). IEEE, 2017, pp. 1–4.
14. N. Saranya , A. Mythili, 2020, Classification of Soil and Crop Suggestion using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 02 (February 2020),
15. P. Bandara, T. Weerasooriya, T. Ruchirawya, W. Nanayakkara, M. Dimantha, and M. Pabasara, “Crop recommendation system,” International Journal of Computer Applications, vol. 975, p. 8887.
16. Patokar, A.M., Gohokar, V.V. (2018). Precision Agriculture System Design Using Wireless Sensor Network. “Information and Communication Technology. Advances in Intelligent Systems and Computing”, vol 625. Springer, Singapore. https://doi.org/10.1007/978-981-10-5508-9_16
17. R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, “Crop recommendation system to maximize crop yield using machine learning technique,” International Research Journal of Engineering and Technology, vol. 4, no. 12, pp. 950–953, 2017.
18. S. A. Z. Rahman, K. C. Mitra, and S. M. Islam, “Soil classification using machine learning methods and crop suggestion based on soil series,” in 2018 21st International Conference of Computer and Information Technology (ICCIT). IEEE, 2018, pp. 1–4.
19. S. Babu, “A software model for precision agriculture for small and marginal farmers,” in 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS). IEEE, 2013, pp. 352–355.
20. S. N. Seo and R. Mendelsohn, “An analysis of crop choice: Adapting to climate change in south american farms,” Ecological economics, vol. 67, no. 1, pp. 109–116, 2008.
21. Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra” Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series” 2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018
22. Thomas Lisha Varghese Merin Mary Saji Varsha S Kevin Tom Thomas, Er. Jinu. “Crop Prediction Using Machine Learning | International Journal of Future Generation Communication and Networking.” *Crop Prediction Using Machine Learning | International Journal of Future Generation Communication and Networking*, 1 July 2020, sersc.org/journals/index.php/IJFGCN/article/view/28189.
23. Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah “AgroConsultant: Intelligent Crop Recommendation System Using Machine

Learning Algorithms ” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)