International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Intelligent Crop Selection System using Ensemble Learning with Random Forest Approach for Sustainable Farming

B Jeyashankari[1], M Dheetchanya[2], M Shree Soundarya[3], M S Bennet Praba[4]

[1,2,3,4]Department Of CSE, SRMIST, Ramapuram, Chennai
Email: [1]bz5729@srmsit.edu.in, [2]dm5170@srmist.edu.in, [3]ss8345@srmist.edu.in, [4]bennetms09@gmail.com,

Abstract

Crop analysis and prediction technology continue to expand rapidly because it serves as a critical tool for agricultural operation enhancement. In agricultural practices farmers depend on crop recommendations since these assists them in determining which crops their area and climate support most productively. Specialist knowledge operated as the main requirement in this process which demanded significant time consumption in the past. A sustainable way of living becomes necessary for continued existence. Agricultural experts believe machine learning automations should serve as a foundation for crop suggestion automation and pest detection to help farmers maximize their farming output with nutritious soil maintenance [1]. This paper develops a precise machine learning model which functions effectively for crop recommendation. A number of features within the proposed system utilize climatic data together with soil composition to make precise crop recommendations for specific areas. Agricultural crop recommendations would experience a transformative change due to this technology which leads to higher yields and sustainable farming alongside improved profitability for farmers at different scales. Through evaluation of many machine learning algorithms, we successfully measured perfect accuracy by running detailed tests over a large historical dataset. Our highest achieved accuracy level amounted to 99.5% accuracy.

Keywords: Agriculture, crop, food, environmental factors, machine learning, prediction, data analysis, recommendation, big data, and agricultural productivity.

1. Introduction

India absolutely depends on agriculture for its economic stability and human continuity continues through its agricultural sector. The occupation stands as a main necessity that sustains human life. Through farming we perform many daily activities as well as sustain human existence. When farmers lose their ability to sustain themselves the inability to repay agricultural bank loans turns into a suicide condition [2]. Evidence shows the climate is transforming in this modern age which creates continuous crop destruction leading to farmer debt and suicides. A range of mathematical or statistical analysis methods helps lower the risks in data prediction. The agricultural sector in India witnessed meaningful progress since the implementation of precision farming technology [3]. Area-specific agriculture stands as an essential element of this because it works toward farm technique optimization per geographical region. Other than its progress precision farming still faces several obstacles. Precision agriculture has made crop recommendations its most notable achievement because these recommendations depend on many different factors. The strategic suggestions derive from examining soil health together with climate conditions with water resources to determine suitable crop selections for each territorial area.

The main goal of precision agriculture includes identifying problems through location-based parameter evaluation. When implemented precision farming fails to generate precise results in every case. Precise trustworthy recommendations form a critical foundation for agriculture because faulty decisions would lead to significant financial material losses. Various researchers work to develop an improved method for crop forecasting accuracy.

The figure 1 illustrates three main algorithms of supervised and unsupervised and reinforcement learning which machine learning adopts to perform tasks depending on their respective strengths and weaknesses. Supervised learning algorithms create mathematical models through a processing method that requires both training input and predicted results. The process of building models through unsupervised learning starts with data containing only inputs since there are no labelled outputs available. The model gets more advanced functionality by using partial training data which contains specific input examples without labels through semi-supervised learning.

This research paper aims to establish which crop suits input conditions composed of soil pH together with humidity and temperature levels and rainfall intensity alongside N, P, and K fertilizer content. The research applies multiple supervised machine learning models throughout India to predict the yield accuracy across 22 different agricultural commodities. All data variables encompass different parameters. RF algorithm builds the model using features that undergo LE transformation encode the t var within this system. The important purpose of the study helps farmers select correct crops by evaluating rainfall conditions and weather patterns and soil nutritional elements. The study has ML methods to understand data through RF algo for suitable suggestion in different geographical locations. The main aims focus on enhancing crop production while decreasing financial risks along with providing better farming choices to farmers.

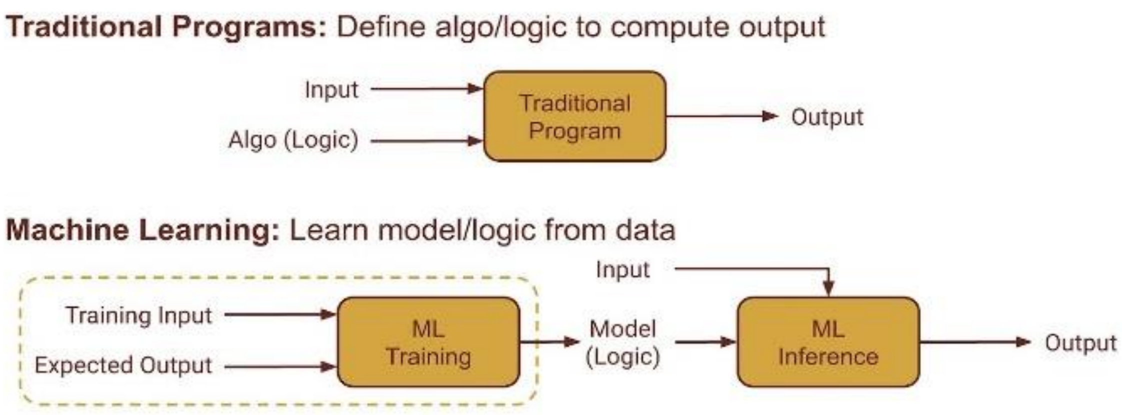

Fig. 1. Traditional Programming vs. Machine Learning

2. Related work

[4] *Aditya More et al.,* explained Neural Networks and SVM and Decision Trees that are employed to anticipate farm outputs in this study. Paper talks about outlines both the pros and cons of every method that is studied.

[5] *Ayalew Kassahun et al.,* presents a summary of crop yield prediction methods which include Neural Networks together with SVM and Decision Trees. The research outlines both the advantages and disadvantages of every method that is studied.

[6] *G. N. Srinivasan et al.,* described method that enables precise choice of suitable crop seeds according to soil characteristics alongside rainfall patterns and surface temperature properties. The proposed system performed tests using ml algorithms.

[7] *Aarti Watekar et al.,* examined key elements for crop selection including soil classification and examination together with crop yield measurement. The introduced system utilized multiple ml algorithms. Working model of this study will enable us to forecast crop types based on district together with state along with meteorological information. We have designed our proposed work to support farmers through helping them select appropriate seeds depending on soil conditions for increased national agricultural output.

[8] *Suresh et al.,* identified crop recognition through available data sources. The research obtained better accuracy together with efficiency through implementation of Support Vector Machines (SVM). This research concentrated on evaluating two data types consisting of agricultural data and geographical data. The proposed system identifies crop nutritional values for certain crops then supplies recommendations for suitable fertilizers.

[9] *Suresh Rathod et al.,* aims to recommend specific crops through analysis of soil database information. This system incorporates evaluations of factors like stratification, grain structure, porosity, water escape etc., and other characteristics when serving certain crops. Database-driven system delivers precise and fast crop decisions through various ML models.

[10] *Rashi Agrawal et al.,* described two components which include a farming suitability assessment and precipitation estimator. This system examines major agricultural crops combined with minor crops. It analyzes variables too. Network.

[11] *Daneshwari Modi et al.,* proposed smart crop advisory system using SVM algorithm to help out cultivators. Main goal of analysis is to determine crop profitability potentials because this reduces both losses and boosts production. Various soil factors require classification and the best crop prediction through an implementation of the SVM algorithm. Soil factor analysis with crop suggestions happens through the implementation of Anaconda Navigator.

[12] *Sunil Ghane et al.,* develops a crop selection system with Random Forest Models combined with Convolutional Neural Networks (CNN) to assess multiple factors such as area and soil composition and market price for determining the most suitable crop.

[13] *K. S. Shreedhara* utilizes an ensemble model combining various methods with efficient learning techniques which applies majority voting for delivering specific crop recommendations with site-related factors.

[14] *Abhishek Ray et al.,* created quick assistance systems by integrating crop data with GPS location data and machine learning programs that produce precise and efficient farming suggestions for agricultural enthusiasts. Farmers can determine suitable crops for their geographic area by leveraging individual recommendations generated through this system because it takes into account climate patterns along with soil characteristics and water resources and local environmental conditions.

3. Proposed system

As seen in Fig. 2, our framework includes a suggested process that is divided into multiple steps. The following are the various stages:

1. Data Acquisition
2. Data Cleaning and Transformation
3. Feature Selection and Extraction
4. Integration of Predictive Models
5. Development of Intelligent Crop Recommender
6. Output: Best suited Crop choice

1. Dataset Acquisition

Data includes various elements that play a major role in the farm yield assessment. The data collection includes twenty-two specific cultivated products. Available data exists as separate collections where 20% of instances serve as Test data while the rest remain in the Train segment. The sample size calculation (n) relies upon Eq 1 to determine its value.

– (1)

Here,

- Z, the z-score
- P is the proportion estimation,
- E, the margin error.

2. Data Cleaning and Transformation

Structuring of data for machine learning models operates under the name of data preparation. Model building requires this initial phase as the backbone to efficient train machine learning models. Developing machine learning projects usually does not result in data that is both clean and prepared. Anticipate data cleaning to occur before performing operations on it. To achieve this objective, we execute data preprocessing tasks. Power BI enables us to remove garbage numbers and eliminate outliers and eliminate both local min-max values and peak/downfall occurrences. in-max scaling served as the normalization method which follows Equation 2:

– (2)

Here,

- P’, the value normalized,
- P, the original value,
- Pmin and Pmax are the min and max values

3. Feature Selection and Extraction

Establishing numerical data from unstructured agricultural data demands feature extraction as a necessary initial process. The project involved implementing Label Encoding to transform class labels into integer values for encoding the categorical goal crop features. The system applies this method to satisfy Random Forest's requirement for numeric input data. The data transformation procedure preserved all original information and maintained the quantitative values of temperature, humidity, rainfall, P, K, N and pH measurement data. The conversion led to better learning capabilities because the model now identified key patterns as well as correct crop classifications based on environmental inputs.

4. Integration of Predictive Models

RF Algorthm is used for implementation for this work. Machine learning utilizes RF model as a highly optimal algorithm. An extensive dataset contained various agronomic characteristics. Various agronomic characteristics were analyzed as inputs. The data collection process included value handling and data normalization to ensure high quality and consistent results prior to training.

The Random Forest classifier implements bootstrap aggregation to generate a group of decision trees [15]. Random sampling with replacement enables the production of a subset from original datasets for each decision tree ft(x). The choice of optimal splitting criterion depends on randomly selected features during each node split evaluation. Degenerate correlations between trees are minimized by imbuing both sampling mechanisms and feature selection methods with random aspects which enhances generalization.

In classification scenarios the ensemble utilizes a voting criterion based on majority decision to produce the final prediction P^. The expression which explains this representation appears in Equation 3.

P^=m{f1(y),f2(y),…,fT(y)} – (3)

Here,

- T – total no of tress
- fT(y) – prediction

In Eq 4, the prediction uses output averaging from all trees when performing regression tasks which serves as a potential point for future development.

Y^=1T∑t=1Tft(x) - (4)

An 80:20 split method divided the data for training purposes from assessment purposes. Accuracy was used as the key criterion to assess the model alongside other important factors for measuring its effectiveness. The research adjusted model

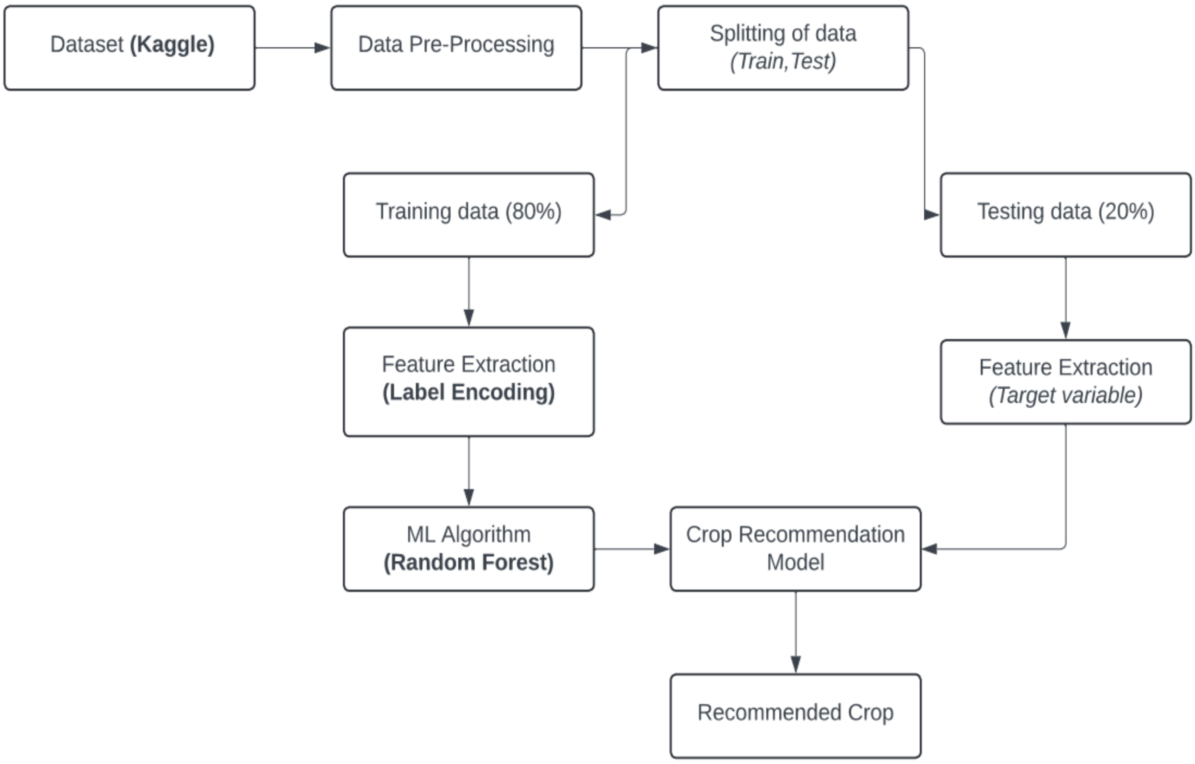

Fig.2. Proposed system Flow

criteria consisting of estimator count together with tree depth and minimum leaf sample value.

5. Development of Intelligent Crop Recommender

After the model completion of training and testing the machine learning system functions as a recommendation platform. The system utilizes trained Random Forest model by sending input information about soil nutrients and climate characteristics for identifying suitable crop selection. The model operates dependably due to ensemble decision-making which requires multiple decision trees to agree on any advice. Low latency along with scalability are core system design elements that make it possible for the platform to function as the computational core for web-based and mobile-based farmer decision-support systems. Real-time forecasting enables farmers to obtain recommendations following their instant input delivery.

6. Output: Best - suited Crop Choice

Under current conditions the recommendation system selects one crop label as the optimal cultivation choice. The recommendation depends on 22 specific crops which were selected due to their agronomic compatibility alongside historical crop choices. The most suitable crop selection with the best economic value has to be rooted in specific values. Forecast enables sustainable agriculture through better land management which leads to increased production benefits. As an operational system the tool provides straightforward scientific data which extends practical advice to farmers as well as extension specialists and agricultural planners.

4. Model accuracy and output interpretation

Evaluation was based upon Precision, Accuracy, F1 – score Recal. Our system utilized an 80:20 data separation for training and testing purposes respectively. Random Forest model achieved almost perfect accuracy at 99% when tested on the given data set which highlights strong prediction reliability. The model demonstrated excellent precision in addition to recall level above 0.98 for most crop classes thus demonstrating robust capabilities for reducing false pos and false neg results.

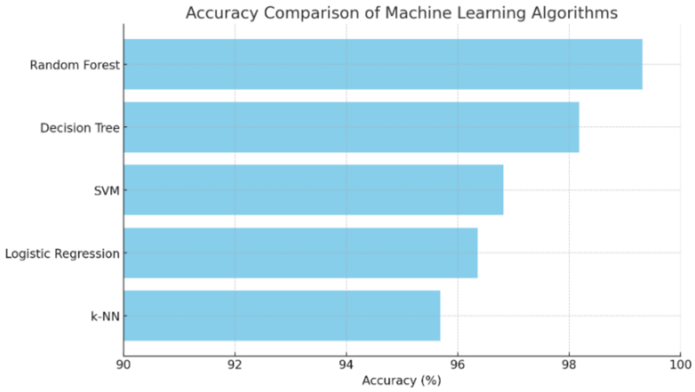

Fig.3. Accuracy Comparison

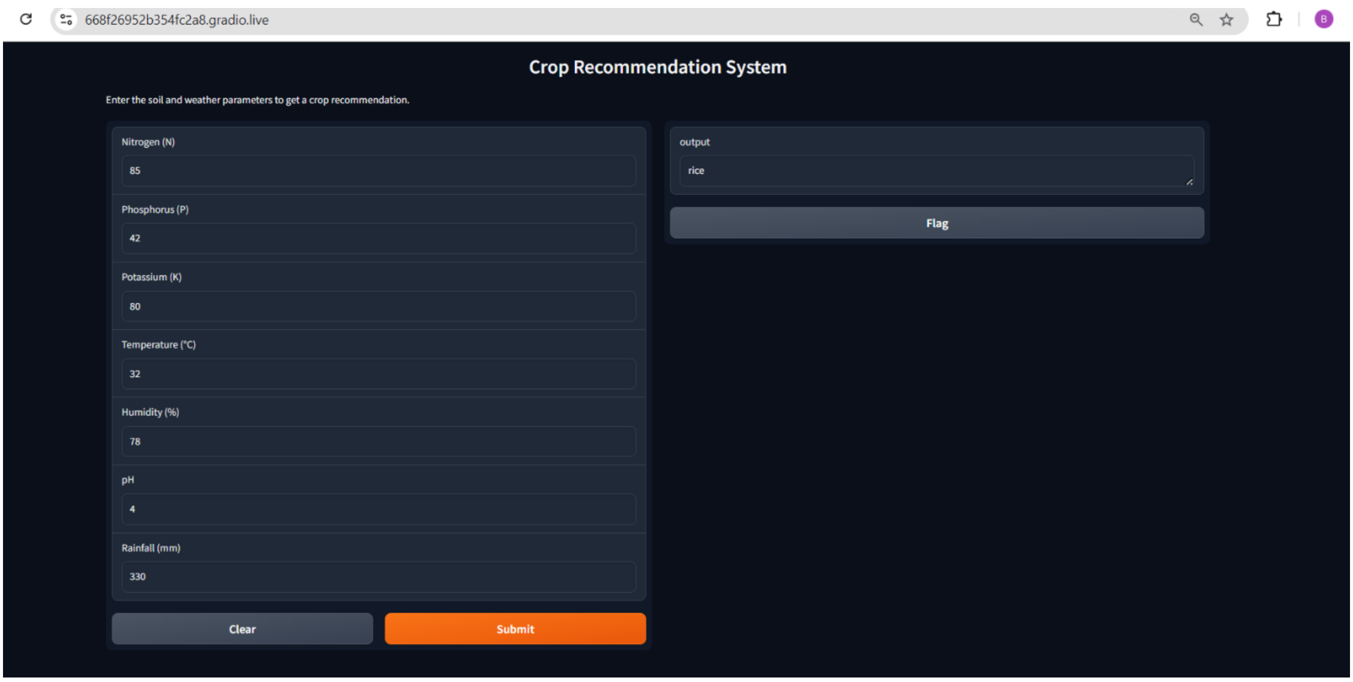

Fig.4. Output

A contingency table analysis showed model accurately identified most of the crops contained within the dataset. The classification of similar crops revealed some minor errors but these are common challenges in actual implementations of such systems. The minimization of overfitting together with better general model applicability were achieved through this implementation. The accuracy performance of Random Forest surpasses that of both Support Vector Machine and Decision Tree models as depicted in Fig 3. Random Forest demonstrated better accuracy and stability than all other classifiers based on the analysis results.

The acquired high-performance evaluation results verify that this suggested model produces dependable crop recommendations across different environmental conditions and soil types. Fig 4 displays the user interface which enables farmers to easily provide environmental parameters before obtaining real-time crop suggestions through the proposed system.

5. Future work

- **Incorporation of Advanced Machine Learning Models:**

  The performance of data predictions may benefit from implementing neural networks along with LSTM instead of traditional algorithms. The evaluation of different ensemble methods would enable optimization of model performance outcome.

- **Integration of Real-Time Data:**

  Future research should integrate real-time data about soil conditions as well as market price information because this would enable dynamically adapting recommendations. An improved system monitoring capability results when this system connects to weather forecasting APIs and real-time soil monitoring IoT sensors as presented in [16].

- **Geospatial Analysis with GIS Data:**

  The implementation of Geographic Information System (GIS) data about regional soil properties and topography and micro-climates would enhance crop recommendation precision according to [17] for various terrain conditions or regional locations.

- **Farm-Specific Customization:**

  Future developments need to concentrate on optimizing suggestions that specifically fit individual farms.

  The evaluation system takes into account particulars about farm size together with yield reports from previous seasons plus conditions of irrigation access.

  A system built to accumulate farm data chronologically will enhance decision-making capabilities for particular farming situations.

- **Market Price Prediction Integration:**

  The system should expand its crop recommendation system to consider both environmental compatibilities together with up-to-date market price data and demand Patterns. Users would obtain essential information to select profitable crops thanks to the integration.

- **Mobile App Development:**

  The arrangement of user-friendly mobile app technology makes it possible for rural farmers to access recommendations quickly. Through this application users could receive customized notifications together with specific farming information. Temperature updates together with pest warnings along with real-time pricing notifications enable customers to obtain specific relevant data.

- **Inclusion of Sustainability Metrics:**

  The future work should develop crop recommendation models according to sustainable parameters including water consumption and environmental impact measurements and carbon emissions. Such an integration would allow farmers to develop financial results while upholding their duty towards environmental responsibility.

- **Data Expansion and Enrichment:**

  The model performance can improve through an ongoing process of data enhancement which includes adding diverse superior data elements like climate projections and soil properties and pest observations. Better advice will emerge through expanded and enriched agricultural research institution collaboration on the database.

- **Incorporation of Crop Rotation and Intercropping Systems:**

  Future improvements to the model should analyse multiple-season advice through recommendations of crop cycles that combine different plants in specific planting arrangements. The strategy would protect soil fertility and develop biodiversity which supports sustainable farming activities.

- **Collaboration with Government and Policy Makers:**

  The solution would reach broader audiences when government bodies or NGOs collaborate on its expansion at scale. The implementation of agricultural subsidies linked with insurance schemes utilizing crop recommendations would deliver supplemental benefits to farmers.

6. Conclusion

The research develops an environmental and soil-oriented machine learning tool to enhance agricultural decision-making via suitable suggestions. RF learning was used to construct model which produces reliable predictions in recommending suitable crops for different conditions.

The system uses organized observation methods to monitor values thus allowing farmers to achieve maximum yields while promoting sustainable agricultural practices. Scientific analysis of the developed prognostic model demonstrated a 99% accuracy rate thus proving its capability to generate exceptional financial and agricultural productivity improvements through exact recommendations.

Subsequent developments to this project should involve the addition of real-time weather information systems together with geographic information systems (GIS), market price analytics and complex ensemble learning methods to enhance economic performance and adaptability of recommendations. The proposed system provides a sound foundation that will guide the development of upcoming smart agricultural technology which aims to supply farmers with data-based decision support.

References

1. Kamilaris, A., Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. doi:10.1016/j.compag.2018.02.016.
2. Mishra, S. (2006). Suicides of farmers in Maharashtra: An empirical analysis. Economic and Political Weekly, 41(16), 1538–1545.
3. Jat, R.A., Saharawat, Y.S., & Gupta, R. (2011). Conservation agriculture in India: Problems, prospects and policy issues. International Journal of Agricultural Sustainability, 9(1), 93–101.
4. Patil, P. *et al.* (2023) 'Crop selection and yield prediction using machine learning approach', *Current Agriculture Research Journal*, 11(3), pp. 968–980. doi:10.12944/carj.11.3.26.
5. Van Klompenburg, T., Kassahun, A. and Catal, C. (2020) 'Crop yield prediction using Machine Learning: A Systematic Literature Review', *Computers and Electronics in Agriculture*, 177, p. 105709. doi:10.1016/j.compag.2020.105709.
6. Kulkarni, N.H. *et al.* (2018) 'Improving crop productivity through a crop recommendation system using ensembling technique', *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp. 114–119. doi:10.1109/csitss.2018.8768790.
7. Reddy, D.A., Dadore, B. and Watekar, A. (2019) 'Crop recommendation system to maximize crop yield in Ramtek region using machine learning', *International Journal of Scientific Research in Science and Technology*, pp. 485–489. doi:10.32628/ijsrst196172.
8. Suresh, G., Kumar, A.S., Lekashri, S., Manikandan, R. and Head, C.O., 2021. Efficient crop yield recommendation system using machine learning for digital farming. *International Journal of Modern Agriculture*, *10*(1), pp.906-914.
9. Rajak, R.K., Pawar, A., Pendke, M., Shinde, P., Rathod, S. and Devare, A., 2017. Crop recommendation system to maximize crop yield using machine learning technique. *International Research Journal of Engineering and Technology*, *4*(12), pp.950-953.
10. Doshi, Z., Nadkarni, S., Agrawal, R. and Shah, N., 2018, August. AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.
11. Modi D., Sutagundar, A.V., Yalavigi, V. and Aravatagimath, A., 2021, October. Crop recommendation using machine learning algorithm. In *2021 5th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-5). IEEE.
12. Motwani, A., Patil, P., Nagaria, V., Verma, S. and Ghane, S., 2022, January. Soil analysis recommendation using machine learning. In *2022 International Conference for Advancement in T (ICONAT)* (pp. 1-7). IEEE.
13. Akshatha, K.R. and Shreedhara, K.S., 2018. Implementation of machine learning algorithms for crop recommendation using precision agriculture. *International Journal of Research in Engineering, Science and Management (IJRESM)*, *1*(6), pp.58-60.

14. Senapaty, M.K., Ray, A. and Padhy, N., 2024. A decision support system for crop recommendation using machine learning classification algorithms. Agriculture, 14(8), p.1256.
15. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324.
16. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in Smart Farming – A review. Agricultural Systems, 153, 69–80.
17. Ben-Dor, E., Irons, J.R., & Epema, G.F. (2018). Hyperspectral Remote Sensing for Precision Agriculture: State of the Art, Challenges, and Future Trends. Remote Sensing of Environment, 113(1), S197–S209.