SN Computer Science (2023) 4:516
https://doi.org/10.1007/s42979-023-01995-8

REVIEW ARTICLE

Voting Classifier-Based Crop Recommendation

Raswitha Bandi[1] · M. Sai Surya Likhit[1] · S. Rajavardhan Reddy[1] · Sathwik Raj Bodla[1] · Vempati Sai Venkat[1]

Received: 7 September 2022 / Accepted: 10 January 2023

Abstract

The three most important necessities for human life are food, shelter, and clothing. Young people who are technologically savvy have witnessed a significant scientific increase in the latter two areas. Despite this, farming is still regarded as a labor-intensive endeavour. Most farmers are uneducated and lack a scientific understanding of farming practices. Crop cultivation anywhere in the world is dependent on the climate, also known as seasons, and soil properties; however, increasing crop production is dependent on a variety of factors, most notably temperature. This work proposes a crop recommendation system to address the issue of increasing crop production. A vision of the perfect harvest before planting would be extremely beneficial to farmers and other stakeholders in making appropriate decisions about improving yields for local use, and it may inspire increased capacity and a wider range of product options for businesses. Precision agriculture is a modern farming strategy that advises farmers on the sorts of crops they should plant based on data collected through studies on soil features, soil types, and crop yields. This style of agriculture is also known as "high-intensity agriculture". Our system employed Machine Learning procedures to recommend the appropriate crops. This system then reduces the financial losses experienced by farmers because of establishing the ominous harvests. This problem is addressed in this paper by proposing a recommendation system using an ensemble model with majority voting and an accuracy of 99.4 percent.

Keywords Crop recommendation · Machine learning · Ensemble model · Voting classifier · Recommendation system

Introduction

Agriculture is one of the oldest and earliest industries, and it suffers from a wide gamut variety of problems, including climate change, unpredictable rainfall, pollution, personnel shortages, and the list goes on [1]. The burgeoning population growth in the current world creates a great demand for agricultural products which ultimately creates an opportunity for farmers to grow their businesses. To feed the world population, it is a requisite to increase agriculture production, to boost agriculture production, researchers are utilizing various technologies such as sensor networks, image processing, remote sensing, machine learning, and other related technologies [2]. However, Machine learning is constantly changing day by day and due to this it is a rapidly evolving approach for making machines intelligent and also trains the machines to be able to work without any instructions. Machine learning techniques are used in a diversity of fields including smart cities, health care, and automobiles. Various machine learning algorithms and techniques are now being applied in agriculture to handle a variety of convoluted problems [3]. A farmer's choice of crop is substantially and significantly influenced by his intuition as well as some other unimportant considerations such as generating quick money, being unobservant of market demand, overestimating a soil's capacity to maintain a specific crop, and so on. If the farmer makes a bad judgment, his financial situation may suffer significantly, resulting in challenges that are tough to overcome. Perhaps this is one of the many and main reasons leading to the many farmer suicide instances reported in the news daily. In countries like India which are overly populated and diversified in many sectors, where agriculture and agri-allied sectors account for the most percentage of the rural income and economy, Such an incorrect decision would reflect on the livelihoods of the farmers and their families [4]. Coming

This article is part of the topical collection "Enabling Innovative Computational Intelligence Technologies for IOT" guest edited by Omer Rana, Rajiv Misra, Alexander Pfeiffer, Luigi Troiano and Nishtha Kesswani.

Raswitha Bandi
raswitha_b@vnrvjiet.in

[1] Department of Information Technology, VNR VJIET, Bachupally, Telangana, India

Published online: 06 July 2023

to the inference a farmer's choice of crop to plant during a given growing season is taken very seriously by us. At this time, the most important goal is to create a system that can provide Indian farmers with foresight with which they can take better decisions about which crops to cultivate. With this in mind, we propose an intelligent system, which would consider all the climatic characteristics, such as temperature, rainfall, and geographical position in terms of state. In addition to this, before offering the ideal crop to the customer, we consider the soil's properties, such as the pH value, the kind of soil, and the concentration of nutrients. Not all agriculture systems provide accurate results [5]. Many research works are being carried out, but most of them do not give accurate results. To achieve an accurate and efficient crop forecast model. CNN is one of the methodologies involved in many research studies. This paper provides a system that employs the voting classifier approach to construct an efficient and accurate model that yields the best results possible out of all the machine learning techniques currently employed in this sector [6].

Literature Review

Crop Recommendation System Based on Machine Learning [7]: In rural India, agriculture and related industries provide the majority of employment and income. Agriculture contributes significantly to the country's GDP (gross domestic product). The massive agriculture industry is a boon to the country. However, the agricultural yield per hectare is unacceptable in comparison to international standards. This may contribute to a higher suicide rate among India's small-scale farmers. This study aims to provide farmers with a practical and simple yield forecasting method. Under the proposed system, farmers can be connected via a smartphone application. GPS can be used to determine the user's location. The user specifies the size and type of soil. Machine learning algorithms can be used to identify the most profitable crop list or to forecast the yield of a user-selected crop. It is possible to estimate agricultural yields using Machine Learning techniques such as Random Forest (RF), Multivariate Linear Regression (MLR), or K-Nearest Neighbour. Random Forest had the highest accuracy rate in this group, at 95%. To boost output even more, the system recommends the best time to apply fertiliser.

Agricultural Crop Recommendation System Using Various Machine Learning Regression Methods [8]: Machine learning (ML) is a technique for discovering new models from massive datasets. To achieve this goal, multiple regressive approaches will be used. The random forest method, as well as linear and polynomial regressive methods, is examples of these. Among the meteorological information produced by necessary data are area and production. This study calculates crop yield suggestions based on an exact comparison of several machine learning regressions, with an overall improvement over various current techniques of 3.6%.

Crop and yield prediction in agriculture using machine learning techniques [9]: Farmers used to forecast crop yields based on their own experience. Crop yields are at a crossroads as population growth outpaces supply. Machine Learning (ML) approaches are now being used in a variety of industries to achieve practical and profitable results. In machine learning, classification, clustering, and neural networks can be used to estimate agricultural yields. In this study, the K-Nearest Neighbors (KNN) algorithm is used to evaluate soil quality and forecast the best crop to cultivate. As inputs, our system considers temperature and soil quality. We also recommend fertiliser based on crop yield projections. According to the test results, our technique successfully forecasts crop selection and production, which greatly benefits farmers.

Machine Learning-Based Intelligent Crop Recommendation System [10]: Machine learning techniques such as Decision Tree, KNN, and Linear Regression were used to develop a recommendation system that was 81%, 88%, and 91% accurate, respectively. This system was suggested in this paper. To assist growers in making informed crop choices and avoiding failure, we developed a recommendation system based on a variety of soil, rainfall, and surface temperature characteristics. This paper proposed a recommendation system based on Random Forests and SVMs.

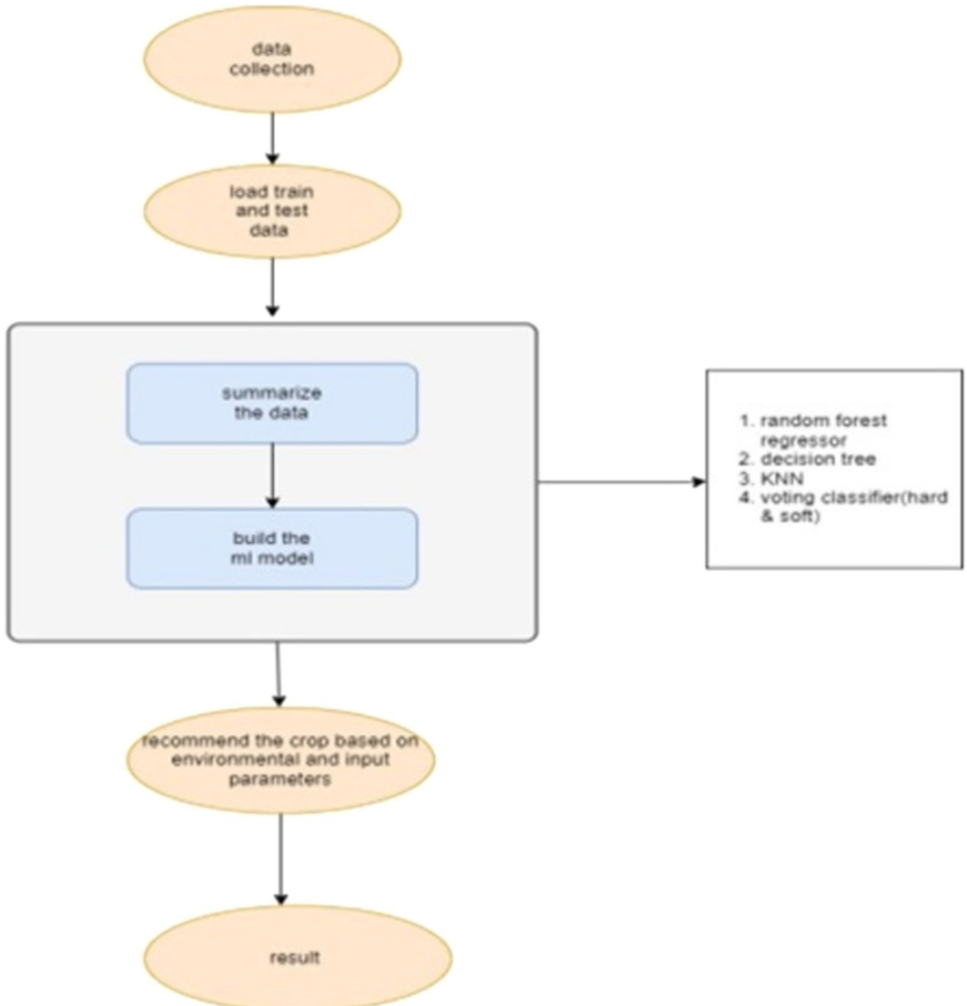
Improving Crop Productivity Using an Ensembling Technique for Crop Recommendation [11]: To assist growers in making informed crop choices and avoiding failure, we developed a recommendation system based on a variety of soil, rainfall, and surface temperature characteristics. This paper proposed a recommendation system based on Random Forests and SVMs.

Precision Agriculture Crop Recommendation System [2]: An ensemble model with majority voting is proposed in this study that uses Random Tree, K-Nearest Neighbor, and Naive Bayes as learners to recommend a crop effectively and precisely.

Methodology

A user-friendly application for crop recommendation is still in the works, despite the various solutions lately provided [12]. Here, we tackle these problems by developing an easy-to-use user application that incorporates characteristics such as rainfall, temperature, soil type, and other factors that have an impact on crop cultivation. The main objective is to determine a crop that can give farmers a better yield over the season [13]. The system that has been proposed would aid farmers in reducing the challenges they encounter in

Fig. 1 Flowchart of the system



selecting a crop and aiding farmers in the process. The suggested model forecasts the crop for the given region's data sets. Integrating agriculture with machine learning will lead to further improvements in the agriculture industry by maximizing the resources required. The data from prior years is crucial in anticipating present performance [14]. Kaggle.com is used to collect historical data. The data includes various attributes such as N, P, K, temperature, humidity, ph, rainfall, and crop. The data sets are pre-processed, cleaned. In addition, the mean values are used to replace null values. Before running the algorithms, the category characteristics are translated into labels to deal with categorical variables in data sets [15]. Utilizing the crop recommender method is an option for the farmer to consider taking if he is unsure which crop he should sow in the upcoming growing season. For the crop suggestion system to function properly the farmer is required to provide information regarding the local climate and soil and in return the system provides the farmers with an indication of what to harvest. This makes it simple for farmers to determine which crop to sow. Using ML various patterns and relations are discovered. Using historical data sets, the model is trained to predict future outcomes based on prior experience [16]. Various standard and general ML algorithms are used to develop the crop recommender system as shown in Fig. 1. Among the selected algorithms, the Voting Classifier of type hard provided the best accuracy. The algorithms we considered here are Decision Tree, Random Forest, KNN, Voting Classifier (hard), and Voting Classifier (soft). For Classification and prediction Decision tree is one of the most powerful tools, KNN stands for “K-Nearest Neighbor” [17]. It is one of the supervised machine learning algorithms. Which can be used for both the classification and regression? However, According to our literature review, many individuals have failed to find a solution using various ensemble methodologies. Our technique for making recommendations for crops is to look for a single model that can reliably forecast the desired result.

Ensemble techniques analyse a large number of models and average them to achieve a single final model, rather than building a single model and hoping it is the best/most accurate prediction possible [18]. As a result, the Voting

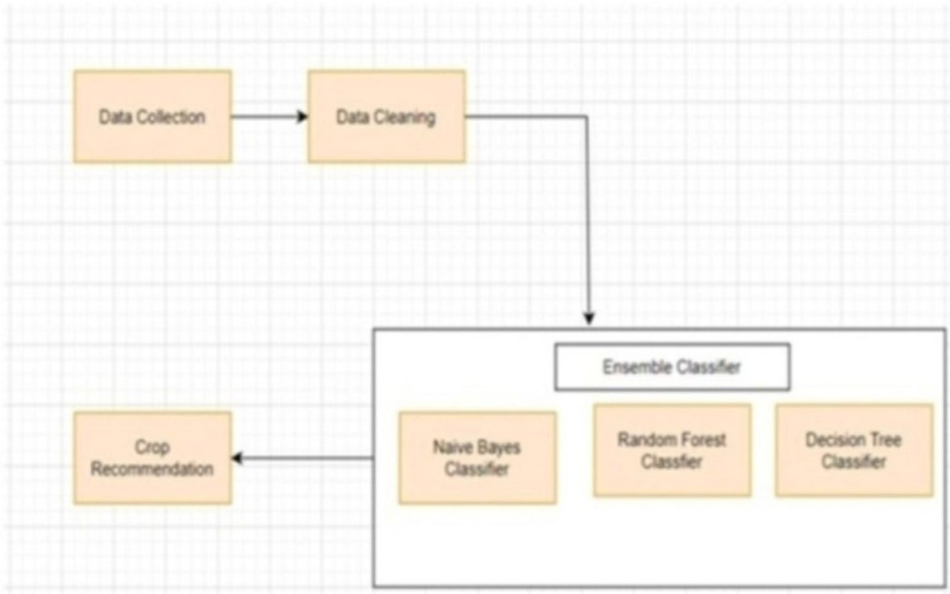


Fig. 2 Estimators taken in our model

Classifier Ensemble Technique was created as shown in Fig. 2 [19].

The Voting Classifier is a learning estimator that trains a number of different base models or estimators and then forecasts based on their output.

If desired, the aggregating criteria for each estimator output can be merged with voting decisions. There are two sorts of voting criteria:

- Hard Voting: Voting is based on the projected output class in hard voting [20].
- Soft Voting: Voting is based on the output class's expected probability.

The voting classifier compiles the predicted class or projected probability on the basis of either hard voting or soft voting. Therefore, if we feed the voting classifier a variety of different base models, it guarantees that the mistake will be fixed by one of those models [21]. To carry out our work plan in a logical manner, we begin by developing a recommendation model that can use the crop dataset as input [22]. The dataset is subsequently exposed to the following steps as part of the recommendation model's validation and evaluation:

- The crop dataset is uploaded to the model which contains the features and the target recommended crop.
- The data in the dataset is transformed into data frames followed by training and testing of the model.
- Once model training is completed, the generated model can now be used for the prediction of new incoming data values as shown in Fig. 3.

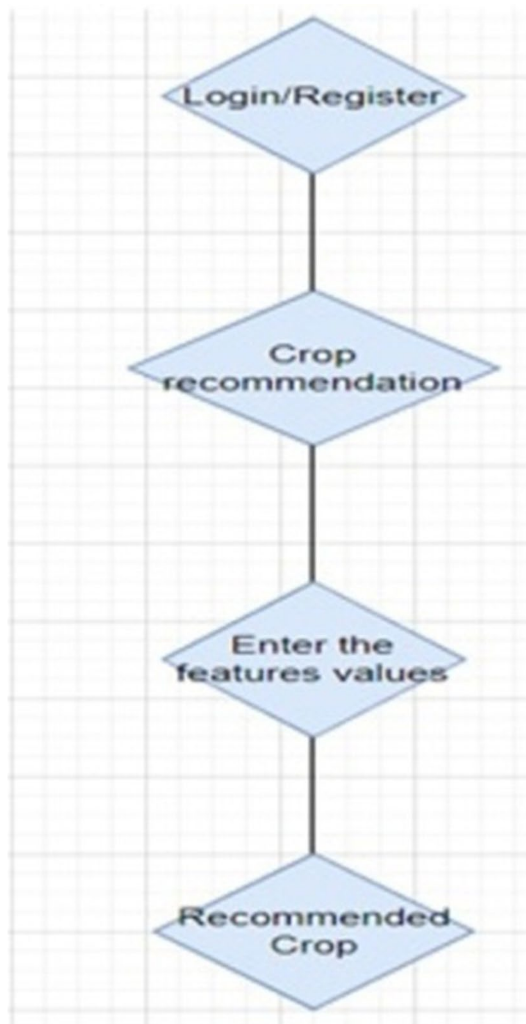


Fig. 3 Flowchart of the website

- The result generated casually indicates the crop to be cultivated depending upon the highest accuracies of each individual ml algorithm.

We shall now look into the hardware and software requirements of our proposed method.

Hardware Requirements

- Minimum 4 Gigabyte (GB) RAM (used for processing).
- 60 MB Memory space (approximate value).
- System type—64-bit Windows 10 operating system, × 64-based processor.

Software Requirements

As a consequence of its extensive support for various Learning tasks through leveraging preexisting packages, Python is a language of choice for Machine Learning and Deep Learning. We used the following packages in our codebase [23]:

- NumPy: NumPy has a broad variety of mathematical functions, such as random number generators, linear algebra routines, and Fourier transformations, in addition to other useful tools. NumPy is a library for the Python programming language that provides additional support for enormous, multidimensional arrays and matrices, in addition to a large number of high-level mathematical functions that may be used to work on these arrays [24].
- Matplotlib: Matplotlib is a library for the Python programming language that enables the creation of visualizations that can be either static, animated, or interactive. Matplotlib is a graphing tool for the Python programming language that works in conjunction with NumPy, which is an extension of numerical mathematics in Python [25].
- Pandas: Pandas is a collection of applications for manipulating and analyzing data that was developed specifically for the Python programming language. It has procedures and data structures that allow for the manipulation of numerical tables and time series.
- Seaborn: Seaborn is a Python module for creating statistical visuals [22].

Algorithms

- Decision Tree Classifier Decision trees are used to classify both regression and classification problem statements. They are tree-like structures created in a top-down fashion where each path from the root node to the leaf node represents a sequence of rules. The decision-making process starts from the root node and continues until leaf nodes have appeared [23].
- KNN Classifier, The KNN model checks for the similarity between new data points and the existing data points, and assigns the former a category that is most related to the existing categories. The model initially prepares all the known data and stores the data. Once the model encounters a new data point, it classifies the data point based on similarity [24].
- Random Forest Classifier Random Forest is one of the types of ensembling technique that makes use of a number of decision tree models to build a final model. It aggregates these models where each model is built using subsets of a single dataset so that the anomalies are reduced. This improves managing the outliers and reduces deviations caused by them.
- Every classifier votes for a class in hard voting (majority voting), and the class with the most votes wins. The mode of the distribution of individually predicted labels is the ensemble's expected target label. The mode of predictions from multiple classifiers is used to categorize input data. When the weights of the multiple classifiers are equal or not, majority voting is processed differently.
- To determine if a data point belongs to a certain target class, each classifier in soft voting gives a probability value to each data point. Following weighting by the classifier's relevance, the predictions are added together. The target label with the greatest weighted probability total will thereafter get the most votes [25].

Results and Discussion

Algorithms that yielded these findings are discussed in this section. Nitrogen (N), Phosphorous (P), Pottasium (K), humidity, temperature, rainfall and ph are the variables employed. The accuracy of the crop suggestion is compared with all the algorithms which have been specified. With a 99.3 percent accuracy rate, the Hard Voting Classifier was found to be the most effective for the supplied dataset. A variety of machine-learning algorithms are employed to make crop recommendations [26]. These include KNN, Decision Tree, Random Forest, and the Voting Classifier. Comparison of the accuracy of several machine learning methods is shown in the Table below. There is a visual depiction of the data in the below Figs. 4 and 5.

Metric-scores:

Fig. 4 Accuracies for various algorithms

	Accuracy	Precision	Recall	f1-score
KNN	97.1	97	97	97
Decision Tree	97.3	98	97	97
Random Forest	98.8	99	99	99
Voting Classifier(Hard)	99.0	99	99	99
Voting Classifier(Soft)	98.3	98	98	98

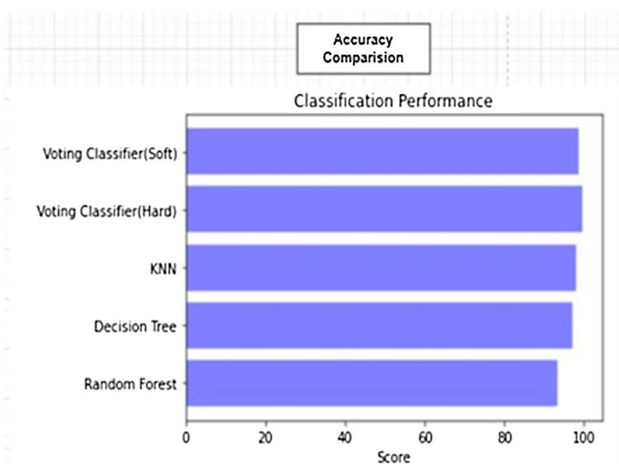


Fig. 5 Accuracy comparison bar plot

Conclusion

When compared to other machine learning models, the majority voting algorithm has the highest accuracy of 99.0%. This technology aids the farmer in making the appropriate crop choice by educating them about the information and other factors that the farmers do not keep track of, hence reducing crop failure and boosting output. This keeps them from racking up debts as well. Millions of farmers around the country will be able to use the system if it is made available online. Integration of the crop recommendation system with another subsystem is a planned next step in development. Once this has been accomplished, the farmer will be able to provide information regarding the weather and size of the field, and we will be able to estimate the crop that will be produced from that area of land.

Data availability Not applicable.

Declarations

Conflict of interest Raswitha Bandi declares that he/she has no conflict of interest. M Sai Surya Likhit declares that he/she has no conflict of interest. Rajavardhan Reddy S declares that he/she has no conflict of interest. Sathwik Raj Bodla declares that he/she has no conflict of interest. Vempati Sai Venkat declares that he/she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: systematic literature review. Comput Electron Agric. 2020;177: 105709.
2. Pudumalar S, Elangovan R, Rajashree R, Kavya C, Kiruthika T, Nisha J. Crop recommendation system for precision agriculture.

Int J Res Appl Sci Eng Technol. 2017. https://doi.org/10.1109/ICoAC.2017.7951740.
3. Kumar R, Singh MP, Kumar P, Singh JP. Crop selection method to maximize crop yield rate using machine learning technique. In: International conference on smart technologies and management for computing, communication, controls, energy and materials. 2015.
4. Lee H, Moon A. Development of yield prediction system based on real-time agricultural meteorological information. In: 16th international conference on advanced communication technology. 2014.
5. Lekhaa TR. Efficient crop yield and pesticide prediction for improving agricultural economy using data mining techniques. Int J Mod Trends Eng Sci IJMTES. 2016;03(10):11–28.
6. Gholap J, Ingole A, Gohil J, Gargade S, Attar V. Soil data analysis using classification techniques and soil algorithms for formulating yield prediction accuracy in precision agriculture', innovations in information, embedded and communication systems (ICIIECS). Attribute prediction. Int J Comput Sci Issues. 9(3).
7. Shariff S, Shwetha RB, Ramya OG, Pushpa H, Pooja KR. Crop recommendation using machine learning techniques. Int J Eng Res Technol IJERT ICEI 2022;10(11).
8. Garanayak M, et al. Agricultural recommendation system for crops using different machine learning regression methods. IJAEIS. 2021;12(1):1–20.
9. Nigam A et al. Crop yield prediction using machine learning algorithms. In: 2019 fifth international conference on image information processing (ICIIP), pp. 125–130. 2019.
10. Priyadharshini A et al. Intelligent crop recommendation system using machine learning. In: 2021 5th international conference on computing methodologies and communication (ICCMC), pp. 843–848; 2021.
11. Kulkarni NH et al. Improving crop productivity through a crop recommendation system using ensembling technique. In: 2018 3rd international conference on computational systems and information technology for sustainable solutions (CSITSS), pp. 114–119. 2018.
12. Savla A, Dhawan P, Bhadada H, Israni N, Mandholia A, Bhardwaj S. Survey of clas- sification algorithms for formulating yield prediction accuracy in precision agriculture. In: Innovations in information, embedded and communication systems (ICIIECS). 2015.
13. Parikh DP, Jain J, Gupta T, Dabhade RH. Machine learning based crop recommendation system. 2021;6(1).
14. https://www.researchgate.net/publication/325107675_Machine_Learning_in_Soil_Classification_and_Crop_Detection.
15. Formulating yield prediction accuracy in precision agriculture. In: Innovations in information, embedded and communication systems (ICIIECS). http://www.ijirae.com/volumes/Vol8/iss-03/03.MRAE10082.pdf.
16. Portugal I, Alencar P, Cowan D. The use of machine learning algorithms in recommender systems: a systematic review. Expert Syst Appl. 2018;97:205–27. https://doi.org/10.1016/j.eswa.2017.12.020.
17. Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review. Comput Electron Agric. 2020;177(January):105709. https://doi.org/10.1016/j.compag.2020.105709.
18. Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput Electron Agric. 2018;151:61–9. https://doi.org/10.1016/j.compag.2018.05.012.
19. Burke R. Hybrid recommender systems: survey and experiments. User Model User-Adap Interact. 2002;12:331–70. https://doi.org/10.1023/A:1021240730564.
20. Poriya A, Bhagat T, Patel N, Sharma R. Non-personalized recommender systems and user-based collaborative recommender systems. Int J Appl Inf Syst. 2014;6(9):22–7.
21. Yang L. Classifiers selection for ensemble learning based on accuracy and diversity. Selection and/or peer-review under responsibility of [CEIS]. Elsevier Ltd.; 2011.
22. Baitharua TR, Panib SK. Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. In: International conference on computational modeling and security (CMS). 2016.
23. Khedr AE, Kadry M, Walid G. Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector applied case on Food Security Information Center Ministry of Agriculture, Egypt. In: International conference on communications, management, and information technology (ICCMIT'). 2015.
24. Paul M, Vishwakarma SK, Verma A. Analysis of soil behaviour and prediction of crop yield using data mining approach. In: International conference on computational intelligence and communication networks. 2015.
25. Bandi R, Tejaswini K. Machine and deep learning techniques for internet of things based cloud systems. Data Sci Data Anal Oppor Chall. 2021;331.
26. Savla A, Dhawan P, Bhadada H, Israni N, Mandholia A, Bhardwaj S. Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture. In: Innovations in information, embedded and communication systems (ICIIECS). 2015.