A Crop Recommendation System Based on Nutrients and Environmental Factors using Machine Learning Models and IoT

Anishka Chauhan
ACT Lab, Department of CSE,
SRM University-AP, India
anishka_anil@srmap.edu.in

Anuraag Tsunduru
ACT Lab, Department of CSE,
SRM University-AP, India
anuraagt_sunduru@srmap.edu.in

Kishwar Parveen
ACT Lab, Department of CSE,
SRM University-AP, India
parveen_shaik@srmap.edu.in

Srilatha Tokala
ACT Lab, Department of CSE,
SRM University-AP, India
srilatha_tokala@srmap.edu.in

Koduru Hajarathaiah
ACT Lab, Department of CSE,
SRM University-AP, India
hazarathaiah_koduru@srmap.edu.in

Murali Krishna Enduri
ACT Lab, Department of CSE,
SRM University-AP, India
muralikrishna.e@srmap.edu.in

2023 International Conference on Information Technology (ICIT) | 979-8-3503-2006-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIT58056.2023.10226131

***Abstract*—With the ever-increasing population of the world, enough crop production is the biggest concern for the human race. This issue is more pressing than ever as the world population has surpassed the 8 billion mark. Smart farming has become a popular option as it solves the problem by suggesting ways to increase the quality and quantity of crop yield. It is a term associated with the practice of automating farm-related activities. This paper proposes a crop recommendation system based on machine learning algorithms for agricultural fields in India. A sensor system is also prepared to collect first-hand data from fields. These IoT sensors are then used to record levels of soil moisture content, Temperature, and the three most important macro-nutrients required for soil growth: Nitrogen (N), Phosphorus (P), and Potassium (K), from different fields. Additionally, other variables such as rainfall, sowing season, and pH value of soil are also considered to build the proposed crop recommendation system that recommends the best-yielding crop based on the other environmental factors. Multiple machine learning algorithms including Artificial Neural Networks (ANN), Random Forest, Logistic Regression, and K-Nearest Neighbor (KNN) are used and compared to identify the most efficient algorithm for the crop recommendation system. The proposed system aims to develop a model that can help farmers increase their crop yield and quality by providing personalized recommendations based on environmental variables.**

***Index Terms*—Machine learning algorithms; Recommendation system; Sensors; Soil nutrient based.**

I. INTRODUCTION

Agriculture plays a vital role in the global economy, providing food, fiber, and raw materials for various industries. With the ever-growing population and the need for sustainable agricultural practices, there is a pressing need to optimize crop production and resource utilization. One way to achieve this is through the use of advanced technologies and intelligent systems that can assist farmers in making informed decisions.

Agricultural decision-making is a complex task that involves considering various factors such as soil quality, nutrient levels, climate conditions, and crop characteristics. Traditionally, farmers rely on their experience and local knowledge to determine which crops to cultivate in a particular area. However, these conventional methods often lack precision and fail to take advantage of the vast amount of data available today.

India, the homeland of more than 140 crore people is now the most populated country in the world. Along with providing education, housing, etc to this huge population, the biggest and foremost challenge is to feed them all. With an area of 394.6 million acres, India has the second-largest cultivable land in the world. India is a developing nation with three-quarters of its population employed as agricultural laborers [1]. Agriculture is currently experiencing one of the least returns because of an array of concerns, including the current population, pollution, environmental factors, in addition to economic factors. Even though the Indian plains are incredibly fertile, only a few states produce more grain than others, hence productivity varies between states. It all comes down to the quality of the soil. Understanding the process of what to cultivate in which type of land and how to maintain the cultivating land may prevent damage to soil quality. Ineffective crop irrigation systems and unpredictable weather patterns are some variables that can damage the soil and reduce the efficiency of farming [2].

To address these challenges, the integration of Machine Learning models and the Internet of Things (IoT) has emerged as a promising approach to developing intelligent agricultural systems. Machine Learning algorithms can analyze large data sets and identify patterns, enabling accurate predictions and recommendations. On the other hand, IoT devices can collect real-time data on environmental factors like temperature, humidity, and soil moisture, providing valuable insights for decision-making.

Smart Farming is a new idea in the field of agriculture. It is associated with the amalgamation of technology and agriculture to increase the quality and production of crops in a country. With our country facing the problem of the increasing population to be fed from almost the same area of cultivable land available as we had 5-10 years ago, it is our utmost priority to come up with methods and solutions that can increase our crop yield. By analyzing a crop's ecological characteristics, machine learning algorithms and Internet of Things (IoT)-based sensors can help farmers by recommending the best crop to cultivate on their fields [3]. We can understand soil health and nutrition by using IoT sensors. This data can be extracted using these sensors and then can be analyzed using various machine learning models to come up with a solution that can tell us which crop can produce the best yield when grown on specific soil with specific environmental conditions. Through this paper, we are making a crop recommendation system that will suggest which crop to be grown for the best yield when given inputs about specific environmental variables. We are implementing 9 different ML algorithms for the same dataset and then comparing them all based on their prediction accuracy. The algorithm giving the best accuracy will be used to recommend the crops. The data that we have used for this paper is collected in different ways. We have collected the data ourselves using IoT Sensors. The rest of the data is taken from the internet from Kaggle.com. The data consists of Indian crop Information. By increasing the crop yield we can also move towards self-sufficiency and not have to depend on other countries for feeding our people. This in turn will make us economically well off as a country and we can further use this surplus money to develop the farming sector of our country.

II. Related works

Many scientists and academics are working to apply machine learning techniques in the sector of agriculture [4]. Machine learning is bringing about a huge transformation in the tech industry. Machine learning algorithms are decision tree [5], logistic regression [6], random forest [7], K-nearest neighbour. A decision tree is an algorithm commonly used for classification and regression tasks in machine learning. It utilizes a tree-like structure to make predictions or decisions based on input data. Logistic regression, despite its name, is a classification algorithm within the realm of machine learning. It is a statistical model that predicts the probability of a binary outcome by utilizing input features. Random Forest is an ensemble learning algorithm that enhances prediction accuracy and robustness by combining the outputs of multiple decision trees. During the training phase, the algorithm constructs a "forest" of decision trees, where each tree is constructed using a random subset of the training data and a random subset of input features. This randomization injects diversity among the trees, reducing the risk of overfitting the training data. K-nearest neighbors (KNN) is a machine learning algorithm that predicts by measuring the similarity between an input data point and its k-closest neighbors in the training dataset. The value of k, determined by the user, specifies the number of neighbors taken into account for prediction. For classification, KNN assigns the input data to point the majority class label among its k nearest neighbors, employing a voting mechanism. In regression, the algorithm predicts the average or weighted average of the target values from the k nearest neighbors.

The application of machine learning to predict agricultural yields based on soil data is discussed below. The yield of mustard crops has been predicted using ML algorithms based on soil analysis, according to a method put out by Vishal Pandith [8]. For this process, multiple machine learning algorithms [9], [10] like Multinomial Logistic Regression, Artificial Neural Network (ANN), and Random Forest were employed and they checked using the parameters accuracy, recall, precision, specificity, and f-score. In comparison to Naive Bayes' [11], lowest accuracy prediction of 72.33%, KNN [12] projected a maximum accuracy of 88.7% and a random forestlands prediction of 94.13%. While checking precision the lowest value was 24.17% which was predicted by Logistic Regression [13] and the maximum value of 99.94% was predicted by ANN. All the classifiers under consideration predicted recall values of over 90% for Naive Bayes. By then, we can see that higher specificities and f-scores were obtained by ANN and KNN. This shows that Naive Bayes had the highest false negative rate, whereas Logistic regression had the highest false positive rate and the lowest true negative rate. This approach leads to the conclusion that the most accurate approaches for estimating mustard crop production are KNN and ANN.

To assist farmers in selecting the best crop to grow, Murali Krishna Senapati suggested an Internet of Things (IoT)-enabled soil nutrient classification and crop recommendation model [14]. This data includes GPS position, moisture, temperature, water level, and soil nutrient information. Using this, a farmer can learn more about the soil and receive updates on its condition. The data on the soil also provides recommendations for the crops that should be planted. Farmers can obtain and examine pre-processed soil data using the MSVM-DGA-FFO method. An Android application was created to access and assess the crop, and it also aids in determining the best crop to plant. The MSVM model is tuned using the FFO approach by choosing the appropriate kernel functions. Four distinct crops' real-time data were used for the experiments, and SVM, SVM Kernel, decision trees, and MSVM-DAG-FFO were used. It enables the maintenance of periodic soil health information in a cheap cloud, aiding farmers in crop selection and enabling them to provide pertinent feedback on the use of minerals.

The utilization of multinomial logistic regression [15] in diverse agricultural contexts offers a statistical framework for examining and forecasting categorical outcomes. This technique empowers farmers, researchers, and policymakers to acquire valuable insights about crop production, disease management, pest control, and other crucial aspects of agriculture. Consequently, it facilitates enhanced decision-making and promotes sustainable agricultural practices. Artificial Neural Networks (ANNs) [16] present a robust solution for data-driven decision making in agriculture. Through their capacity to learn and generalize patterns from intricate datasets, ANNs play a pivotal role in enhancing crop management, disease control, resource utilization, and overall operational efficiency within the agricultural sector. The valuable role of Random Forest in agriculture [17] stems from its capacity to handle extensive datasets, capture intricate relationships, and deliver reliable predictions. By harnessing these capabilities, farmers and researchers can leverage data- driven insights to optimize resource allocation, enhance disease control measures, and bolster overall agricultural productivity and sustainability.

A similar smartphone app was introduced by Shilpa Mangesh Pande in 2021 to assist farmers by forecasting agricultural yields [18]. By using the user's location and soil type as input, the GPS in this program helps the user. The most lucrative crop that can be grown there is identified after processing the data using machine learning (ML) algorithms. They can also predict crop yields for crops that the user has chosen. The Support Vector Machine (SVM) [19], [20], Artificial Neural Network (ANN) [21], [22], and Random Forest (RF) [23] are employed in this to assess crop productivity. With a 95% accuracy rate, this device produces the best results and may also suggest when to apply fertilizers to boost yield.

Some of the limitations that can be observed from the related work are the usage of IoT for gathering data on soil nutrients and environmental factors is briefly mentioned in the research review. The intricacies of IoT integration, such as the kinds of IoT devices used, data transmission protocols, data processing methods, and scalability issues, are however rarely discussed. These elements are crucial for comprehending the viability and usefulness of putting into place an IoT-enabled crop recommendation system. does not specifically address the method used to gather the data, its quality, or any potential biases or restrictions related to the data used for training and evaluation.

III. Methodology

In this paper, environmental factors that lead to crop growth and impact its products are evaluated. The following are the attributes that are considered to build a prediction model, proposed in this paper: Nitrogen quantity in soil, Phosphorus quantity in soil, Pottasium quantity in soil, Soil Temperature, Humidity, PH value of soil, Rainfall, Label/crop name, and Sowing Season. To build this model several datasets from the internet are merged. 25% of the data that is used to build the model is first-hand collected from various fields in Vijayawada, Andhra Pradesh, and fields around. The rest 75% of the data is gathered from the internet and is specific to the fields in India.

To collect this first-hand data, IoT sensors are used. Data is collected from these sensors by making a sensor circuit system. The sensors used in making this sensor circuit are: 1. NPK sensor: It is used for detecting the content of nitrogen, phosphorus, and potassium in the soil. 2. Soil Moisture sensor: It is used to calculate moisture value in soil and the sensor. 3. Soil temperature sensor: It is used to monitor the temperature of soil.

The device that we made to procure data from fields consists of the NPK sensor, soil moisture sensor, and soil temperature sensor as shown in Fig. 1. All three sensors are connected to the breadboard using wires and resistors. The same line is connected to the Arduino Uno. These sensor connections are made to an Arduino Uno board connected to the laptop.

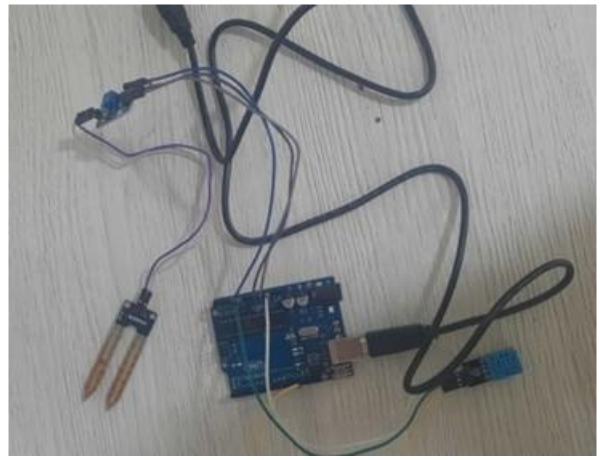

Fig. 1. Sensor circuit with soil moisture sensor and soil temperature sensor

These sensors are placed in the ground. The data is collected via the sensors for a fixed set of time. After collecting the data, it is merged with the data collected from the web. Since we have not collected the pH level of data we collected using IoT sensors, those null values are covered by replacing the null values with the mean of the pH level of that particular crop. For example, for data that we collected for rice from rice fields, the value for these tuples of pH attribute is taken as the mean of pH of the rest of the rice tuples. We have divided the entire dataset into training and testing datasets. Here, 30% of the dataset is used as a testing database and 70% as a training dataset. This dataset is then analyzed using multiple Machine learning algorithms. We have used the following algorithms to make a recommendation system from this dataset, in this paper: Decision Trees [24] provide a versatile and interpretable foundation for recommendation systems, allowing personalized recommendations, dealing with complex reasoning, giving users transparency, Naive Bayes can produce acceptable results in recommendation tasks despite its simplicity and some assumptions (such as feature independence), particularly in circumstances where interpretability, scalability, and efficiency are crucial factors, Logical Regression delivers insights into feature relevance, manages sparse data, and solves the cold start issue. It may be a useful and scalable technique for tailored suggestions across a range of topics, Random Forest (RF) can make precise and trustworthy suggestions across a range of disciplines by utilising the advantages of ensemble learning, K Nearest Neighbour (KNN) can manage sparse data, solve the cold start issue, and give people recommendations that are easy to understand, Catboost [25] is a potent method for recommendation systems due to its specific handling of categorical variables, tolerance to noisy data, prevention of overfitting, and scalability, LinearSVC [26] can nevertheless offer reliable advice in situations where a clear distinction between positive and negative examples is necessary, LightGBM [27] is a potent method for recommendation systems due to its effective handling of categorical variables, quick training and scoring, scalability, and capability to handle sparse data, and Artificial Neural Networks (ANN) are useful tools for recommendation systems because they are flexible, can capture complicated associations, and can handle a variety of input types. The process of the proposed model is shown in Fig. 2.

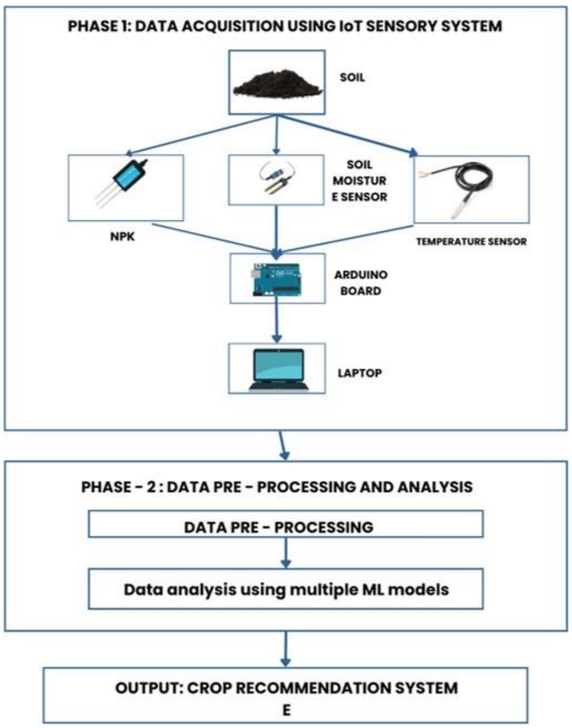

Fig. 2. Flow chart for proposed model

Our research uses IoT sensors to collect data, takes into account a wider range of variables, employs a particular set of machine learning algorithms, and offers insights into the experimental procedure. contains a range of methods designed to handle various data types, including categorical variables, sparse data, and complex associations.

IV. Dataset Statistics

The dataset that is used to build this model consists of 16 crops namely, Rice, Maize, Chickpea, Kidney beans, Black gram, Lentil, Banana, Mango, Grapes, Watermelon, Muskmelon, Apple, Orange, Coconut, Cotton, and Jute. and 9 features as shown in Fig. 3.

We have collected the data using sensors from fields around Vijayawada, Andhra Pradesh, India. But due to transportation and sowing season constraints, the data that we collected was not enough to make an effective crop recommendation system and so we have added data to our dataset by extracting from the internet. The representation of the heat map for the dataset is shown in Fig. 4.

	N	P	K	temperature	humidity	ph	rainfall	label	sowing season
0	90	42	43	20.880	82.003	6.503	202.936	rice	kharif
1	85	58	41	21.770	80.320	7.038	226.656	rice	kharif
2	60	55	44	23.004	82.321	7.840	263.964	rice	kharif
3	74	35	40	26.491	80.158	6.980	242.864	rice	kharif
4	78	42	42	20.130	81.605	7.628	262.717	rice	kharif
5	69	37	42	23.058	83.370	7.073	251.055	rice	kharif
6	69	55	38	22.709	82.639	5.701	271.325	rice	kharif
7	94	53	40	20.278	82.894	5.719	241.974	rice	kharif
8	89	54	38	24.516	83.535	6.685	230.446	rice	kharif
9	68	58	38	23.224	83.033	6.336	221.209	rice	kharif

Fig. 3. Crops Dataset.

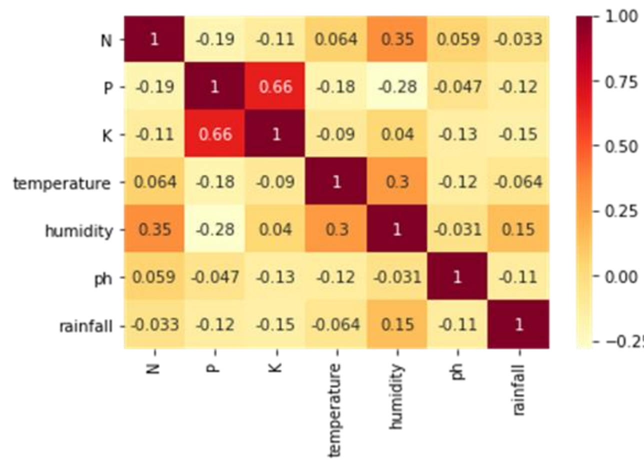

Fig. 4. Heat Map

The entire dataset is cleaned and checked for any noise. During the cleaning process, we rounded up the values of columns ‘Temperature’, ‘Humidity’, and ‘pH’ to 3 decimal places as all the values were heterogeneous, rounded up to different decimal places. So to make the data uniform we have done this step. The dataset used for making the model does not contain any null values. The dimensions of the dataset are 1440*9. Post-checking for any noise we then added another attribute to the dataset extracted from the Internet for better evaluation. We recognized which crop is majorly grown in which sowing season and then gave all the crop values from Kharif, Rabi, and Zaid under the column name 'Sowing Season'. The data that we have collected through sensors do not have pH values as we did not collect the pH value of crops while collecting data for NPK, temperature, and soil moisture. Due to this, we have filled the Null values present for the dataset's pH attribute with the mean pH value of that particular crop. For example, all the null values present in the pH attribute for the label 'Rice' was replaced with the mean value of the rest of the data with the label Rice i.e. 6.425.

This clean and processed data is then divided into training and testing data with 70% and 30% of data division to each, respectively. The training data is fed to multiple ML models for learning purposes. Post this target values are predicted for the testing data by these algorithms. These predictions are checked with the actual values of the target or label or crop name to calculate the accuracy of predictions made by the algorithm. We initiated an empty list before analyzing the data through different ML models and then appended the accuracy value of all the algorithms to that list. This list is then used to plot a bar graph and compare which algorithm provides the best accuracy rate.

V. Results

The following evaluation measurements can be applied in the context of a crop recommendation system based on nutrients and environmental parameters using machine learning models and IoT.

Accuracy: Accuracy evaluates the system's crop suggestion's overall correctness. It determines the percentage of recommended crops that were appropriately chosen. The accuracy would be 80%, for instance, if the system suggested 100 crops and 80 of them were appropriate.

Precision: Precision calculates the percentage of accurately advised crops out of all advised crops. It shows how accurate the system is at advising crops.

Recall: Recall also known as sensitivity, calculates the percentage of actual crops that were correctly advised out of all the actual crops that were eligible for recommendation. It shows how successfully the system can recognize the pertinent crops.

F1-Score: The harmonic mean of recall and precision is known as the F1-score. By considering both precision and recall, it offers a fair assessment of the system's performance.

In this research work, our focus was on developing a Crop Recommendation System that leverages machine learning algorithms to predict the most suitable crop based on specific parameters. To achieve this, we implemented nine different algorithms, each trained and tested on the same dataset. To evaluate the performance of these algorithms, we considered several metrics, including F1 score, recall, precision, and accuracy. These metrics provide insights into the algorithms' ability to correctly classify and predict the appropriate crop based on the given parameters.

For each algorithm, we stored the accuracy, precision, recall, and f1-scores in a list and plotted them against the algorithm names, as shown in Fig. 5. The graph provides a visual representation of the comparison across all nine algorithms. From the accuracy comparison, it is evident that the LightGBM algorithm outperformed the other algorithms, exhibiting the highest accuracy among the nine. In terms of precision, recall, and F1-scores, it is observed that the CatBoost algorithm performs with the highest score. In our recommendation, we have taken an algorithm based on the accuracy score, and the LightGBM algorithm is used for our crop recommendation system.

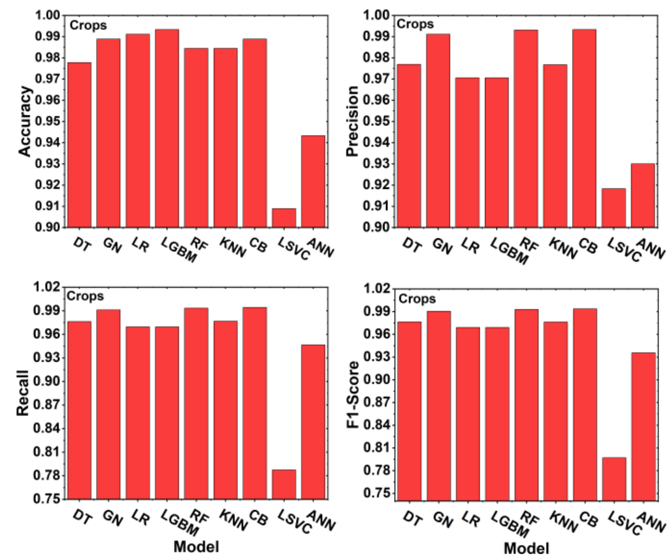

Fig. 5. Accuracy, Precision, Recall, and F1-score for different ML algorithms.

The implemented crop recommendation system will employ the LightGBM algorithm to make crop suggestions when given inputs such as N, P, and K nutrient levels, humidity, temperature, and rainfall. By leveraging the capabilities of LightGBM, the recommendation system can provide accurate and reliable crop recommendations, assisting farmers in optimizing their crop selection process. By using LightGBM as the underlying algorithm, the Crop Recommendation System benefits from its ability to handle large-scale datasets efficiently and effectively. LightGBM's optimization techniques and gradient boosting approach enable it to make accurate predictions while maintaining computational efficiency. The evaluation of nine different algorithms for our Crop Recommendation System has led us to select LightGBM as the primary algorithm due to its superior accuracy. This algorithm will serve as the foundation for our recommendation system, providing farmers with precise and reliable crop suggestions based on the given parameters.

VI. Conclusion and Future Work

In conclusion, the development of a Crop Recommendation System based on machine learning models and IoT integration has significant implications for the farming sector as a whole. By leveraging real-time data analysis and forecasting technology, this system offers valuable insights and benefits not only to crop producers but to the entire farming community. Through the utilization of this system, we can provide the computer with relevant data, enabling it to make accurate predictions about a crop's potential production. By considering factors such as NPK levels, soil moisture, temperature, and other external elements relevant to farming, we have presented an optimum strategy for predicting crop output productivity. The implementation of this technology holds immense potential in reducing resource waste in the farming sector. By accurately predicting crop yields and productivity, farmers can make informed decisions about resource allocation, thereby optimizing resource usage and minimizing waste. This, in turn, contributes to sustainable farming practices and helps overcome the challenges posed by scarce fertile land. Furthermore, the adoption of this Crop Recommendation System has broader implications for the agricultural industry and the nation's economy. By growing suitable crops in the right environment, farmers can maximize their income and improve their livelihoods. This, in turn, has a positive impact on the overall agricultural sector and contributes to the nation's GDP growth.

REFERENCES

[1] J.-J. Dethier and A. Effenberger, "Agriculture and development: A brief review of the literature," *World Bank policy research working paper*, no. 5553, 2011.

[2] N. G. Rezk, E. E.-D. Hemdan, A.-F. Attia, A. El-Sayed, and M. A. El-Rashidy, "An efficient iot based smart farming system using machine learning algorithms," *Multimedia Tools and Applications*, vol. 80, pp. 773–797, 2021.

[3] K. S. P. Reddy, Y. M. Roopa, K. R. LN, and N. S. Nandan, "Iot based smart agriculture using machine learning," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 130–134.

[4] 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.

[5] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in *2011 IEEE control and system graduate research colloquium*. IEEE, 2011, pp. 37–42.

[6] E. Bisong and E. Bisong, "Logistic regression," *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*, pp. 243–250, 2019.

[7] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *The Stata Journal*, vol. 20, no. 1, pp. 3–29, 2020.

[8] V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis," *Journal of scientific research*, vol. 64, no. 2, pp. 394–398, 2020.

[9] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[10] S. Ray, "A quick review of machine learning algorithms," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE, 2019, pp. 35–39.

[11] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 403, p. 412, 2018.

[12] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine*, vol. 4, no. 11, 2016.

[13] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of biomedical informatics*, vol. 35, no. 5-6, pp. 352–359, 2002.

[14] M. K. Senapaty, A. Ray, and N. Padhy, "Iot-enabled soil nutrient analysis and crop recommendation model for precision agriculture," *Computers*, vol. 12, no. 3, p. 61, 2023.

[15] C. Kwak and A. Clayton-Matthews, "Multinomial logistic regression," *Nursing research*, vol. 51, no. 6, pp. 404–410, 2002.

[16] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1–12, 2019.

[17] A. O. Ok, O. Akar, and O. Gungor, "Evaluation of random forest method for agricultural crop classification," *European Journal of Remote Sensing*, vol. 45, no. 1, pp. 421–432, 2012.

[18] S. M. Pande, P. K. Ramesh, A. ANMOL, B. Aishwarya, K. ROHILLA, and K. SHAURYA, "Crop recommender system using machine learning approach," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021, pp. 1066–1071.

[19] Z. H. Kok, A. R. M. Shariff, M. S. M. Alfatni, and S. Khairunniza-Bejo, "Support vector machine in precision agriculture: a review," *Computers and Electronics in Agriculture*, vol. 191, p. 106546, 2021.

[20] S. Suthaharan and S. Suthaharan, "Support vector machine," *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, pp. 207–235, 2016.

[21] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th international conference on ICT and knowledge engineering (ICT&KE)*. IEEE, 2017, pp. 1–6.

[22] X. Dai, Z. Huo, and H. Wang, "Simulation for response of crop yield to soil moisture and salinity with artificial neural network," *Field Crops Research*, vol. 121, no. 3, pp. 441–449, 2011.

[23] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[24] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285, 2004.

[25] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.

[26] S. Z. Mishu and S. Rafiuddin, "Performance analysis of supervised machine learning algorithms for text classification," in *2016 19th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2016, pp. 409–413.

[27] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using gbdt, xgboost, and lightgbm algorithms," *Mathematics*, vol. 8, no. 5, p. 765, 2020.