International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

Crop Recommendation System Using Improved Apriori Algorithm

Satyanarayana Nimmala[1], K Venkatesh Sharma[2], Ravikumar Ch[3]
[1]Associate Professor, Department of Artificial Intelligence & Data Science
Chaitanya Bharathi Institute of Technology, Hyderabad, India-500075
satyauce234@gmail.com
ORCID: https://orcid.org/ 0000-0002-9433-1489
[2]Professor, Department of Computer Science and Engineering
CVR College of Engineering, Hyderabad, India- 501510
venkateshsharma.cse@gmail.com
ORCID: https://orcid.org/ 0000- 0003- 4333- 0491
[3]Assistant Professor, Department of Artificial Intelligence & Data Science
Chaitanya Bharathi Institute of Technology, Hyderabad, India-500075
chrk5814@gmail.com
ORCID: https://orcid.org/0000-0003-0809-5545

Abstract— A crop suggestion system is a technologically advanced tool that helps farmers choose which crops to plant in a certain location or with precise environmental circumstances. These systems use a variety of data sources and analytical methods to give farmers customized crop recommendations. This work presents the Crop Recommendation System Using an Improved Apriori Algorithm, which is an Apriori-based crop recommendation system. The goal of the system is to assist farmers in making well-informed choices about which crops to grow and what fertilizers to use depending on the properties of the soil and environment. With consideration for crop variety, climate, and soil nutrient content, the suggested method is an enhanced version of the Apriori algorithm. Tests of the updated algorithm on a dataset of soil samples from different parts of India revealed that it could correctly suggest the optimal crop. The model's output, association rules, is a suggestion system that farmers can use to boost crop productivity while lowering input costs. The method suggested operates in three phases: In the first stage, preprocessing the data is carried out to gather the input parameters that are crucial for determining the recommendation system. In Stage 2, the recommendation system's association rules are extracted by using an iterative approach to determine the threshold support count and confidence. Stage 3: The recommendation system's knowledge base is formed by pruning the top 8 apriori rules depending on priority. From the experiments, it is evident that the improved apriori algorithm-based extracted recommendation system is an interesting development in precision agriculture that could raise farming practices' sustainability and efficiency.

Keywords- Rice, Maize, Mango, Crop Recommendation, Apriori Algorithm.

I. INTRODUCTION

India has a diversified agricultural economy, with many different crops growing in different parts of the nation due to differing agro-climatic conditions. India's farmers select their crops taking into account the country's climate, soil composition, consumer demand, and personal tastes [1]. The following are some of the several crops that Indian farmers love to grow: **Rice:** Grown extensively throughout India, especially in places like West Bengal, Punjab, and Uttar Pradesh, rice is a staple food. It is usually grown in areas with heavy rainfall or access to irrigation since it needs a lot of water. **Wheat:** Grown in Punjab, Haryana, and Uttar Pradesh, among other northern Indian states, wheat is another staple crop. It is sown in the Rabi season and is a winter crop. **Corn, or maize:** Cultivated across India, but especially in Andhra Pradesh, Karnataka, and Madhya Pradesh. It is utilized as cattle feed as well as for human consumption. **Millets** are drought-resistant crops that are cultivated in arid and semi-arid regions of India, such as Rajasthan, Maharashtra, and portions of Karnataka and Andhra Pradesh [1]. Millets include finger millet (ragi), pearl millet (bajra), and sorghum (jowar). **Pulses:** Several states, including Madhya Pradesh, Uttar Pradesh, and Maharashtra, grow pulses like chickpeas (gram), lentils (masoor), and pigeon peas (tur). In Indian cuisine, pulses are a vital source of protein. **Cotton:** Grown in areas like Gujarat, Maharashtra, and Andhra Pradesh, cotton is a significant cash crop. India ranks among the world's top producers of cotton [2]. **Sugarcane:** Grown in states like Uttar Pradesh, Maharashtra, and Karnataka, sugarcane is widely farmed for its sugary yield. **Oilseeds:** Edible oil production is the purpose of cultivating oilseeds such as sunflower, groundnuts, and soybeans. Many states, notably Gujarat, Maharashtra, and Madhya Pradesh, grow these crops. **Spices:** Known as the "Land of Spices," farmers in states like Kerala, Karnataka, and Andhra Pradesh grow a vast array of spices, including black pepper, cardamom, turmeric, and chili peppers.

India is one of the world's leading producers of fruits and vegetables, with areas like Punjab and Haryana focusing on the production of citrus and apple-based foods [2]. **Coffee and tea:** Karnataka, Kerala, and Tamil Nadu are the main growing regions for coffee, while Assam, West Bengal, and Kerala are known for their tea plantations. **Cash Crops:** Various states cultivate cash crops for commercial use, such as sugarcane, rubber, and tobacco. analyses large amounts of data and generates recommendations using data analytics and machine learning algorithms [3]. fertilizers to use on their crops. To provide personalized recommendations to farmers, the system considers various The system can be linked to sensors and other data sources to provide real-time advice based on current weather and soil conditions. In recent days' modern agriculture is using sensors to know crop and soil details.

A crop recommendation system is a computer-based tool that assists farmers in making informed decisions about factors such as soil type, weather conditions, crop type, and nutrient requirements.

Increased crop yields, lower input costs, improved soil health, and reduced environmental impact are all advantages of using a crop and fertilizer recommendation system. Farmers can optimize their fertilizer use and reduce waste and pollution by receiving personalized recommendations. AgroCares, CropIn, and the International Plant Nutrition Institute are among the companies and organizations developing and implementing crop and fertilizer recommendation systems. These systems are being used to help farmers increase productivity and sustainability in various parts of the world, including the United States, Europe, and Asia. Overall, a crop recommendation system is a promising tool for farmers looking to optimize their use of fertilizers and improve their crop yields while reducing their environmental impact [4]. As technology advances, we can expect more sophisticated and accurate recommendation systems to be developed to meet the needs of farmers worldwide.

Several kinds of sensors that are intended to track and gather data about various facets of farming and the environment can detect agricultural data. Modern precision agriculture relies heavily on these sensors to assist farmers in making data-driven decisions that maximize crop yields, minimize resource use, and boost overall productivity.

Fig. 1. Crop and Fertilizer recommendation based on soil and climate changes

II. MOTIVATION

Many studies have shown that crop and fertilizer recommendation systems can significantly increase crop yields while also lowering farmers' input costs. A study conducted in India, for example, discovered that using a recommendation system increased yields by 14% while decreasing fertilizer use by 22%. The quality and quantity of data used, the algorithms and models used to generate recommendations, and the user interface and ease of use for farmers all have an impact on the accuracy and effectiveness of crop and fertilizer recommendation systems [5].

While crop and fertilizer recommendation systems have many advantages, they also have some drawbacks and limitations. These systems, for example, can be costly to implement and maintain, and they may necessitate significant technical expertise to function properly. Overall, the literature suggests that crop and fertilizer recommendation systems can be a valuable tool for farmers looking to optimize fertilizer use and improve crop yields. More research is needed, however, to fully understand the factors that influence the accuracy and effectiveness of these systems, as well as to develop more sophisticated and user-friendly recommendation systems that farmers can easily adopt [6]. Crop classification and yield prediction are common applications for decision trees. These models are based on a hierarchical structure of decision nodes that divide data into smaller subsets based on specific criteria, resulting in a prediction. Neural networks are a type of machine learning model inspired by the human brain's structure. These models can predict crop yield, analyze nutrients, and categorize crops. Support vector machines

(SVMs) are a type of machine learning algorithm that can be used to predict crop yield and classification. SVMs work by determining the best decision boundary between different classes of data and mapping the data into a higher-dimensional space using a kernel function [7].

Random forest is an ensemble learning technique for improving prediction accuracy by combining multiple decision trees. These models are frequently used to forecast crop yields and assess nutrient levels. Deep learning is a subfield of machine learning that uses multiple-layer neural networks to learn complex data representations. Deep learning models have been used to predict crop yield, detect disease, and characterize plant phenotypes [8]. Bayesian networks are probabilistic graphical models that represent the relationships and conditional dependencies between variables. These models have been used to predict crop yields and manage nutrients. Overall, machine learning techniques have yielded promising results for crop and fertilizer recommendation, with many studies demonstrating increased accuracy and efficiency over traditional methods. However, the technique used is determined by the specific problem and data available, and more research is required to determine the most effective approaches for various applications [9]. Though the decision trees are good enough in crop recommendation the outcome of these models are complex to decode and need some prerequisite knowledge to understand. The proposed work in the paper conveys the outcome of apriori association rules in the form of classification rules, like if –then –else rules which are easily understood by the farmers. This set of rules recommends that farmers select the best crop based on soil and weather features [10].

III. LITERATURE REVIEW

Crop recommendation systems are based on the collection of data from various sources, such as soil samples, meteorological conditions, past crop yields, and local agricultural practices. Sensors, satellite imagery, and other methods are utilized to facilitate data collection. (Savary et al., 2012; Zhang et al., 2017) [10] [11]. Then, using sophisticated algorithms and machine learning models, data analysis is conducted while accounting for variables including crop-specific requirements, temperature, precipitation, pH levels, and type of soil (Anaya-Romero et al., 2017; Alippi et al., 2019) [12][13. Then, using the data that has been analyzed, these systems determine which crops are suitable for which farms or places. They take into account factors such as pest resistance, nutrient requirements, and crop growth cycles (Braun et al., 2010; Tekinerdogan et al., 2019) [14]. Personalized crop recommendations are produced after the study, and farmers are frequently provided with a list of appropriate crops that are graded according to their potential yield and profitability in the specified location (Das et al., 2017; Pan et al., 2019) [15]. According to Liu et al. (2019) and Nalwanga et al. (2020), these guidelines provide a basis for well-informed decision-making concerning crop selection, planting schedules, and agronomic practices, ultimately assisting in resource optimization and risk reduction. Aside from giving real-time guidance on pest management, fertilization, and irrigation based on continuously updated data and meteorological information, some systems also provide continuous monitoring and feedback throughout the growing season (Wu et al., 2018; Yang et al., 2019) [16]. To improve accessibility, these systems can be accessed via many platforms such as web-based interfaces, Smartphone applications, or SMS services. This guarantees that farmers with different levels of technological literacy can still use the systems (Savary et al., 2012; Zhang et al., 2017) [17].

Crop recommendation systems encounter several obstacles and constraints. First, as their efficacy is largely dependent on the quantity and quality of data, they are prone to errors when there is a lack of data or when the data is not representative of the real world [18]. Second, some systems suggest intricate machine learning models or algorithms, which could be difficult for farmers or organizations with little technical know-how or funding to install and maintain [19]. Furthermore, the high resource needs of some technologies, like as satellite images or sensor networks may prevent small-scale or resource-constrained farmers from adopting them. Furthermore, when studies fail to sufficiently explore the applicability of their methods in broader agricultural settings, scalability difficulties may surface. Moreover, the reliance on internet connectivity for data updates and recommendations may result in the inefficiency of these systems in locations with sporadic or restricted internet access [20]. A critical component, model correctness, depends on the caliber and volume of training data, and some studies might not fully validate their models. Precision agriculture raises several ethical and environmental issues that should be carefully considered, such as the overuse of chemical inputs and the replacement of traditional farming methods [20]. Farmers may also accept technology differently than other users due to cultural norms, lack of faith in technology, and availability of support and training. The dynamic and unpredictable character of agriculture, which is impacted by pests, market fluctuations, and climate change, needs appropriate system adaptation, which some publications may not sufficiently address.

IV. METHODOLOGY

In this paper, we have proposed an advanced Apriori algorithm that extracts the classification rules to recommend the best crop to farmers based on soil and properties. An important factor in identifying significant links between items in transactional databases is the Apriori algorithm, which is the foundation of association rule mining. Known for its use in

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

market basket analysis in particular, this algorithm is excellent at detecting products that are commonly bought together, providing insightful information about consumer behavior.

Fundamentally, Apriori makes use of the "apriori property," which states that all of an item set's subsets must likewise be frequent if an item is frequent. This characteristic facilitates the rapid creation of candidate item sets while eliminating those that don't meet the minimal support requirement. Using pseudo code, this procedure is concisely described as follows: candidate and frequent item sets are carefully created, counted, pruned, and aggregated until no new frequent item sets appear. An essential tool for identifying patterns and correlations in a variety of fields, such as market analysis and recommendation systems, the algorithm's power comes from its capacity to sort through enormous amounts of transactional data. It does this by using the apriori characteristic to expedite the search process.

The data set (crop recommendation data set) was obtained from Kaggle which includes seven input parameters and one class label attribute. The experimental design is based on 2200 records from 22 different crops. Table 2 lists the input parameters considered for the experiments. Table 1 tabulates the various crops and the number of records with a specific crop's class label. Using a heuristic approach, all numerical values of input parameters are converted into categorical values during the data preprocessing step. A range of soil and environmental factors are presented in the table, each having continuous and categorical values. These variables include temperature, humidity, pH value (pH), rainfall, nitrogen content (N), phosphorous content (P), and potassium content (K). While the category values divide these parameters into "Low," "Medium," and "High" categories based on predetermined criteria, the continuous values show the actual measurements or ratios of these parameters. These classifications make it easier to understand and use these vital measures in practice when evaluating soil quality and environmental conditions for things like land management, environmental monitoring, and agriculture.

The data set is divided into 22 subsets. Each subset contains 100 records, one for each crop chosen for the experiment. A heuristic approach is a problem-solving or decision-making strategy that uses practical rules of thumb or shortcuts to find solutions or make decisions more quickly and efficiently, often in situations where a perfect or exhaustive solution is not feasible or would require too much time and resources. Heuristics are mental strategies or cognitive tools that help individuals simplify complex problems and navigate through uncertainty. They are particularly useful in situations where people need to make quick decisions or when faced with incomplete information.

TABLE 1. DATA SET DETAILS

S. No.	Class Label Name	Number of Records
1	Rice	100
2	Apple	100
3	Banana	100
4	Chickpea	100
5	Coconut	100
6	Coffee	100
7	Cotton	100
8	grapes	100
9	jute	100
10	Kidney beans	100
11	Lentil	100
12	Maize	100
13	Mango	100
14	Moth beans	100
15	Moonbeam	100
16	Muskmelon	100
17	Orange	100
18	Papaya	100
19	Pigeon peas	100
20	Pomegranate	100
21	Watermelon	100

TABLE 2. INPUT PARAMETERS

S. No.	**Input Parameter Name**	**Details**	**Continuous values (Categorical value)**		
1	N	The ratio of Nitrogen content in the soil	<=50(Low)	>50 &&<=100 (Medium)	>100(High)
2	P	The ratio of Phosphorous content in the soil	<=50(Low)	>50 &&<=100 (Medium)	>100 (High)
3	K	The ratio of Potassium content in the soil	<=70(Low)	>71&&<=140 (Medium)	>140(High)
4	Temperature	Temperature in degrees Celsius	<=20(Low)	>21&&<=32(Medium)	>32 (High)
5	Humidity	Relative humidity in %	<=35(Low)	>35&&<=70 (Medium)	>70 (High)
6	Ph	ph value of the soil	<=5(Low)	>5&&<=7 (Medium)	>7 (High)
7	Rain Fall	Rainfall in mm	<=90(Low)	>90&&<=180 (Medium)	>180 (High)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

Pseudo code:

The Apriori algorithm is based on the "apriori property," which states that if an item is frequent (i.e., it occurs with a minimum support threshold), all of its subsets must also be frequent. This property is leveraged to efficiently generate candidate item sets and prune those that do not meet the minimum support threshold.

Initialization:

Initialize a list of candidate itemsets, C, with frequent items of size 1.

Initialize a list of frequent itemsets, L, with frequent items of size 1.

Generate Frequent Itemsets:

while L is not empty:

// Generate candidate itemsets of size k+1 from frequent itemsets of size k.

Ck+1 = Apriori-Gen(L)

// Count the support of each candidate itemset by scanning the transaction database.

for each transaction in the database:

for each candidate item in Ck+1:

if the candidate item is a subset of the transaction:

Increment the support count of the candidate itemset.

// Prune candidate itemsets that do not meet the minimum support threshold.

Lk+1 = {candidate in Ck+1 | support(candidate) >= min_support}

// Add frequent itemsets of size k+1 to the list of frequent itemsets.

L = L ∪ Lk+1

Repeat the process until no new frequent itemsets can be generated.

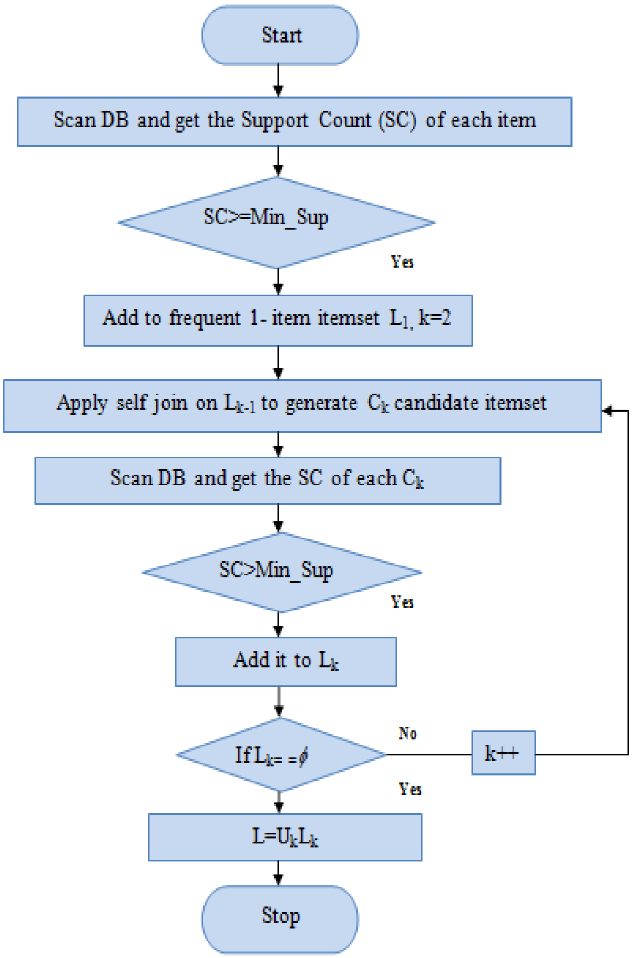

Figure 2: Flow diagram of Apriori Algorithm

V. IMPLEMENTATION & RESULTS

The Apriori algorithm is a well-known association rule mining technique for identifying frequent itemsets in a transactional database. It seeks out interesting relationships between different items in a dataset, such as products that are frequently purchased together in a store or items that are frequently viewed together on an online retail website. The Algorithm is driven by two metrics: support and confidence. Support quantifies an itemset's frequency and is defined as the proportion of transactions that contain the itemsets. Confidence is defined as the proportion of transactions containing both items over the number of transactions containing the first item and measures the likelihood of one item being purchased given the purchase of another item. The Experimental setup is done initially at support 0.6 and confidence 0.9, and an iterative heuristic approach is used to extract the association rules. The best top 8 rules are extracted at support count 0.4 and confidence 0.9. They are considered as best rules as their consequent part of the rule covers almost all input parameters considered for the experiments and also covers most of the input records. The rules recommending the best crop for the farmers are given below.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8
***Article Received**: 20 May 2023 **Revised**: 28 July 2023 **Accepted**: 16 August 2023*

a) R1:N=Medium&&K=Low &&temperature=Medium &&humidity=High&&ph=Medium &&rainfall= High ➔rice

b) R2: N=Low && P=High &&ph=Medium && rainfall=Medium ➔ apple

c) R3:P=Medium && K=Low && temperature=Medium && humidity=High &&ph=Medium && rainfall=Medium ➔banana

d) R4: N=Medium && P=Low && K=Low && temperature=Medium &&ph=Medium ➔maize

e) R5: N=Low && P=Low && K=Low && temperature=High&& humidity=Medium && rainfall=Medium ➔mango

f) R6: K=Low && humidity=High &&ph=Medium && rainfall=Medium && P=Low ➔orange

g) R7: P=Medium && temperature=High && humidity=High &&ph=Medium ➔papay

h) R8:temperature=Medium&&humidity=High &&ph=Medium && rainfall=Low && K=Low ➔watermelon

VI. CONCLUSION

Based on the Apriori algorithm's experimental setup and association rule mining, the algorithm has generated eight recommended rules for suggesting the best crop for farmers based on certain input parameters. Each rule consists of a set of conditions (antecedent) that indicate the input parameter values and a consequent part that suggests the best crop to grow. The rules are ranked based on their support and confidence values. Example Rule 1 (R1): If Nitrogen (N) is Medium, Potassium (K) is Low, Temperature is Medium, Humidity is High, pH is Medium, and Rainfall is High, then the recommended crop is Rice. Rule 2 (R2): If Nitrogen (N) is Low, Phosphorus (P) is High, pH is Medium, and Rainfall is Medium, then the recommended crop is Apple.

These rules are extracted from the transactional database and have been chosen as the best rules based on their ability to cover a significant portion of the input parameters considered in the experiments and cover a substantial number of input records. Farmers can use these rules as guidelines to make informed decisions about the most suitable crops to grow based on specific environmental conditions and soil nutrient levels.

REFERENCES

[1] Birthal, P. S., et al. "Indian Agriculture: Status, Importance, and Role in Indian Economy." Agricultural Economics Research Review, Vol. 27, 2014.

[2] Ramesh Chand. "Agriculture in India: Issues for Policy Analysis." Indian Journal of Agricultural Economics, Vol. 58, No. 3, 2003.

[3] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," IEEE Xplore, Aug. 01, 2018.

[4] Nimmala, S., Vikranth, B., Muqthadar Ali, S., Usha Rani, R., & Rambabu, B. (2022). An intelli MVC: An intelligent majority voting and heuristic-based hybrid classifier to predict high blood pressure. Journal of Intelligent & Fuzzy Systems, 43(3), 3003–3020.

[5] G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," IEEE Xplore, Dec. 01, 2021.

[6] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," IEEE Xplore, Apr. 01, 2019.

[7] K. Bakthavatchalam et al., "IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms," Technologies, vol. 10, no. 1, p. 13, Jan. 2022.

[8] P. Akulwar, "A Recommended System for Crop Disease Detection and Yield Prediction Using Machine Learning Approach," Recommender System with Machine Learning and Artificial Intelligence, pp. 141–163

[9] Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 151, pp. 61–69, Aug. 2018.

[10] Savary, S., Ficke, A., Aubertot, J. N., &Hollier, C. (2012). Crop losses due to diseases and their implications for global food production losses and food security. Food Security, 4(4), 519-537.

[11] Zhang, X., Li, W., & Zhang, M. (2017). Crop yield prediction under soil constraints. Computers and Electronics in Agriculture, 135, 25-33.

[12] Anaya-Romero, M., Fajardo, J. O., Sanchez-Ortiz, A., Fuentes-Angel, J., & Vargas-Vazquez, D. (2017). Crop yield prediction based on machine learning. Procedia Computer Science, 120, 161-168.

[13] Alippi, C., Boracchi, G., Roveri, M., &Benassi, G. (2019). Robust online cropping season classification of time series of multi-spectral images. Computers and Electronics in Agriculture, 157, 322-330.

[14] Braun, M., Fritz, C., & Schulz, H. (2010). A model-based system for agricultural decision support. Environmental Modelling & Software, 25(10), 1225-1232.

[15] Tekinerdogan, B., Tosun, A., &Kuscu, I. (2019). An ontology-based expert system for precision agriculture. Computers and Electronics in Agriculture, 156, 304-315.

[16] Das, A., Ghosh, S., & Ghosh, A. (2017). A novel fuzzy logic and neural network-integrated framework for crop yield prediction. Computers and Electronics in Agriculture, 136, 44-53.

[17] Pan, L., Wang, W., Chen, J., Shi, W., & Wei, L. (2019). Deep crop classification using 3D convolutional neural networks. Computers and Electronics in Agriculture, 162, 1069-1076.
[18] Liu, B., Sun, F., & Ma, W. (2019). A hybrid model for maize yield prediction using machine learning methods. Computers and Electronics in Agriculture, 157, 282-292.
[19] Nalwanga, R., Walker, S., & May, K. (2020). Fertilizer recommendation for maize using machine learning. Computers and Electronics in Agriculture, 177, 105712.
[20] Satyanarayana, N., Ramadevi, Y., Manasa, K., & Rambabu, B. (2020). An intelli AFM: An Intelligent Association-based fuzzy rule miner to predict high blood pressure using bio-psychological factors. Intelligent Decision Technologies, 14(2), 227–237.