IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Intelligent Machine Learning-Based Suggestion For Crop Systems

Gokul S
Post Graduate
Department of Computer Applications
Hindustan Institute of Technology and Science
Chennai, India

Ms. Kalpana K
Assistant Professor
Department of Computer Applications
Hindustan Institute of Technology and Science
Chennai, India

Abstract* - India is predominantly an agricultural country, with agriculture having a major impact on the nation's economy and people's quality of life. The majority of Indian farmers share the common problem of choosing crops that are not best suited to their land's natural conditions. They will thus witness a significant drop in their overall production. In order to address this issue, crops are advised depending on soil, weather, humidity, rainfall, and other geographical conditions, all of which have a significant impact on crop quantity and quality. Therefore, the goal of our research is to analyze all of these variables, match them to the specifications needed for each crop, and provide farmers the best alternative available. As a result, crop selection mistakes are reduced and production is raised. In order to accomplish our goal, we will employ an ensemble model with a majority voting technique that employs a Decision Tree, Logistic Regression, Naive Bayes, CNN, Random Forest (RF), and Extreme Gradient (XG) Boosts learners to highly accurately and efficiently recommend a crop for the site-specific parameters.

Keywords: Recommendation System, XGBoost, Random forest, CNN, Decision Tree, Fuzzy Algorithm

I. INTRODUCTION

India is among the oldest nations in the world to continue farming. Globalization has, however, resulted in a significant shift in agricultural techniques in recent years. Agriculture is one of the main areas of attention. Because they produce so much food, they make a major contribution to society. This remains true in a lot of nations. There is a food scarcity that many people are facing because of population growth. Modern agriculture usually makes use of advanced technologies. Using this advanced agricultural technology may help farmers become more profitable, efficient, safe, and environmentally friendly. The IDE for image processing is called MATLAB. The Hybrid Neural Network approach is used to train the model and increase the program's accuracy in determining the PH and nutrient levels of the soil. The area of artificial intelligence that doesn't involve explicit programming is machine learning.

Our model's output or outcome will be predicted using input data that may be retrieved from previous years. Here, the project's goal is to create a machine learning model that can be used to gather data on the soil, weather, and other geographical elements that are relevant to the growth of a certain crop. The intended model will be built using machine learning methods, specifically Random Forest, which uses decision trees, together with XGBoost and Linear Regression. The model will be able to produce extremely precise forecasts because to these advanced algorithms, which carefully examine the input data. The Python Django web framework will be used to create a user-friendly web interface that presents the predicted results to users. The combination of web development and machine learning guarantees a stable and intuitive platform that people can easily use and understand to make well-informed decisions in agriculture.

II. RELATED WORKS

Deep learning is used in agricultural yield prediction research, which faces issues such as feature quality reliance and indirect data-to-yield mapping. Deep reinforcement learning is integrated in an innovative way. By maintaining data distribution, the suggested Deep Recurrent Q-Net, which combines RNN with Q-Learning, achieves exceptional accuracy (93.7%).

In order to increase farmers' crop yields, this research focuses on gathering and analyzing agricultural and environmental data. The procedure entails preprocessing the data in Python, performing k-means clustering for accuracy, and using the Map Reduce framework for large-scale analysis. Using bar graphs and scatter plots, the study investigates relationships between crops, rainfall, temperature, soil, and seed types in two regions: Ahmednagar, Maharashtra, and the Andaman and Nicobar Islands. Crop predictions are shown using a self-designed recommender system on a Flask-based graphical user interface. In the future, the scalable system can propose crops to additional states.

This is addressed by precision agriculture, which increases total production by taking into account crop yield statistics, soil types, and features to recommend appropriate crops. Improved farming decision-making and efficient resource usage are two advantages of this approach. A recommendation system using random tree, CHAID, K-Nearest Neighbor, and Naïve Bayes as learners is presented to implement precision agriculture. Based on soil factors, this approach guarantees great accuracy and efficiency in crop recommendations. The resultant categorized picture allows for accurate projections of crop yields under certain weather circumstances since it incorporates important variables such as weather, crop yield, and state and district-specific crops.

We are all aware that agriculture is India's main economic sector. The yield of practically every type of crop cultivated in India is forecast in this study. Simple factors like State, district, season, and area are used in this script to create novelty, and the user may anticipate the crop's production in whatever year they want.

III. PROPOSED SYSTEM

The central objective of this initiative is to offer substantial support to end-users, particularly farmers, by presenting innovative ideas and addressing a critical issue in agriculture – profit and cost loss. The project's focal point involves the development of an advanced decision-making XG Boost model. By leveraging insights from two research papers documented in IEEE, the project aims to encapsulate cutting-edge strategies and methodologies. Ultimately, the go a list empower farmers with a reliable and efficient tool that enhances decision-making processes, mitigates profit and cost losses, and contributes to overall agricultural sustainability. This project incorporates Weather API technology to get input of weather parameters from 'WB- CPI: Weather Based Crop Prediction in India Using Big Data Analytics' paper and soil input parameters gathering using the reference from 'Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications' paper. Combining both these ideologies helped us to develop a system that produces result with more accuracy and efficiency. The output of the project plays the major role in crop selection for the field under study. This plays a major role in increasing the crop yield thereby increasing the profit.

XGBoost Model Training and Hyper parameter Tuning

i. Split the data set into testing sets and training sets.

ii. Choose XGBoost algorithm for its advantages in handling complex relationships.

iii. Fine-tune hyper parameters (learning rate, tree depth, regularization) through cross-validation.

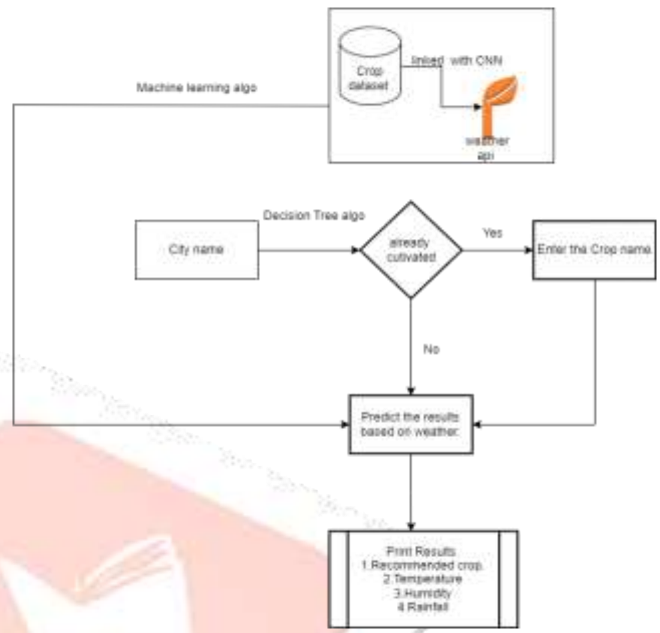
$$f(x) \approx y(t-1) \qquad (1)$$

This represents the prediction from the previous iteration(t-1), denoted as y(t-1).

$$Z(q) \approx \Sigma\chi(x)^2 * (y(t-1) - f(x))^2 + \lambda * T * sum(\sum \chi^2(x))^2 \qquad (2)$$

This formula represents the objective function that XGboost aims to minimize during training. Using(**1**)and (**2**)

Figure: 1 Architecture Diagram



$$g(x) \approx y(t-1) + T * sum(\Sigma\, \chi(x)^2 * (y(t-1) - f(x)))^2)$$

This formula represents the update to the prediction for the current iteration.

T is a hype parameter that controls the step size of the update.

These formulas describe the prediction, regularization, and update steps in the XG Boost algorithm. The algorithm seeks to find the optimal combination of weak learners by minimizing the objective function, which includes terms for prediction accuracy and regularization.

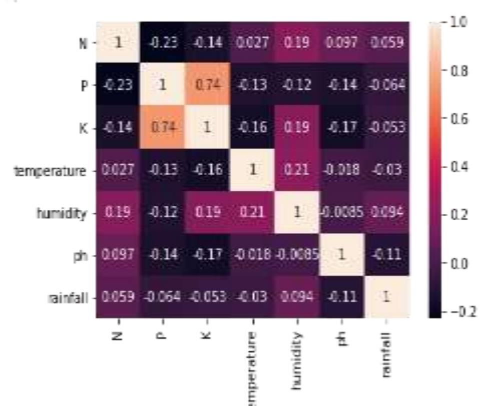


Figure:2 Shows XGBoost model is trained with the collected data that is Temperature, Rainfall, Humidity and land details.

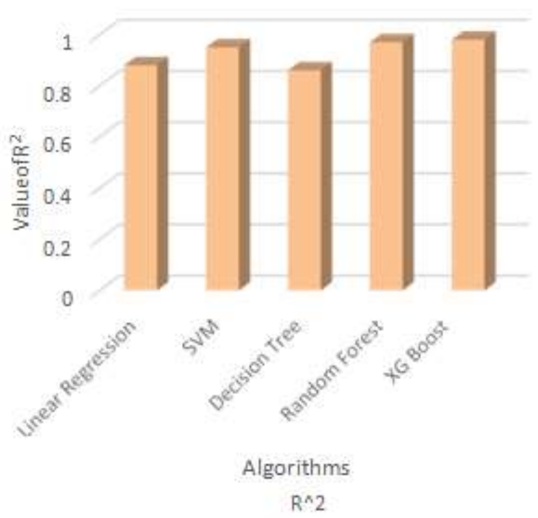


Figure:3 Shows the comparison of R2 for XG Boost algorithm with other algorithms.

Weather Integration

Real-time Weather Updates: Integrate APIs or services to

Provide real-time weather updates for accurate decision-making.

Figure:4 Weather Integration

IV. IMPLEMENTATION AND RESULTS

Five distinct areas make up the process: data intake, preprocessing, data segmentation, classification, and output.

With the primary goal of improving the system's accuracy in supplying the very positive anticipated values or levels of the nutrients and the pH of the soil, the Hybrid Neural Network is included as the training model.

i. **Data Collection**

Nutrients in the soil The dataset needed for this project was gathered from the Department of Agriculture in a few Tamil Nadu districts, including Ariyalur, Salem, and Trichy. The gathered dataset includes details such as characteristics and soil sample values that match the district in which the dirt was extracted. The dataset consists of about 12 characteristics, and the suggested system uses a total of 1676 soil sample instances. The table named Table 1 below shows the attribute description of the dataset that was gathered.

ii. **Data Pre-Processing**

This phase explains how to eliminate unnecessary data from the dataset so that the necessary result data may be extracted.

iii. **Data Conversion**

This phase explains how to eliminate unnecessary data from the dataset so that the necessary result data may be extracted..

iv. **Classification**

This data mining method is based on machine learning with algorithmic techniques. XG Boost and Random Forest algorithms are used in this work to classify soil nutrient datasets.

v. **Prediction**

The classification system is renowned for its accuracy and performance analysis, and it offers farmers recommendations for the best crop to plant depending on the soil.

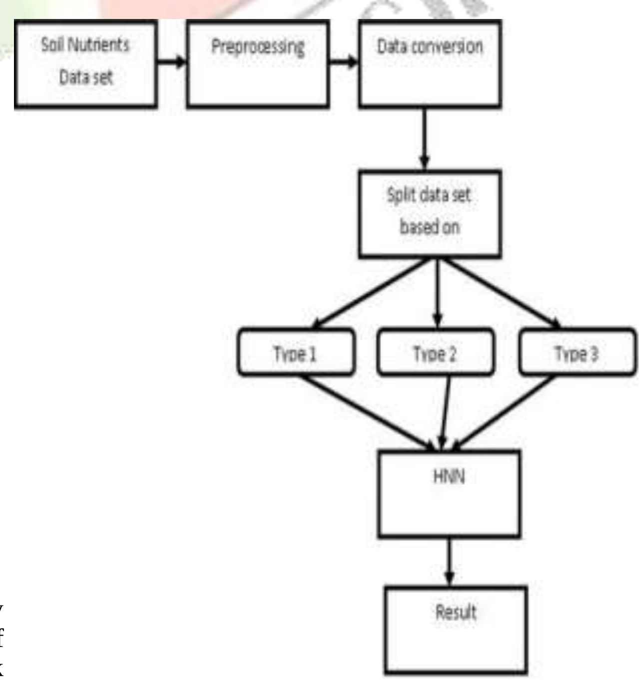


Fig 5. Proposed Workflow

The gathered datasets were pre-processed, and any unnecessary information was eliminated. Following the pre-processing stage, the entire data set was classified as LOW, MEDIUM, or HIGH based on the amount of nutrients found.

Next come algorithms such as Hybrid Classification, RNN, and CNN. The nutrients in the soil are then categorized into the following levels: Very High, High, Medium, Low, and Very Low based on these algorithms.

The comparative examination of classifiers is shown in Table:2. The following graphs demonstrate the accuracy of the rate of those algorithms using a set of three distinct nutrients, as well as the time it takes to execute the various classification algorithms by grouping the three types of nutrients.

Fig 6. Accuracy for execution time of classifiers

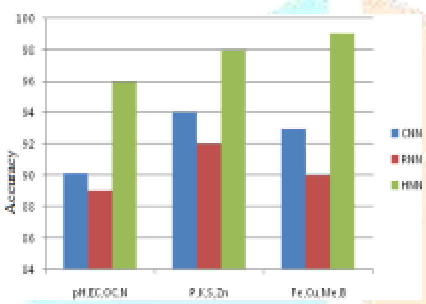


Table1: SAMPLE NUTRIENT DATA-SET

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	8.30	0.26	0.54	43.2	17.6	789	10.5	2.45	2.45	3.56	3.67	3.4
Ariyalur	5.10	0.17	0.40	60.60	16.25	60	17.50	0.53	5.23	0.17	3.38	1.0
Coimbatore	6.50	0.38	0.18	231.0	18.0	400	8.10	0.80	5.10	2.20	2.60	0.30
Coimbatore	7.80	0.18	0.18	173.0	13.6	284	17.24	1.0	5.30	1.40	2.60	0.41
Karur	7.90	0.6	0.35	80.00	2.70	189	8.10	0.84	1.40	0.02	1.08	1.10
Karur	8.10	1.50	0.15	45.00	2.70	233	0.11	0.10	2.60	0.03	1.70	1.10
Salem	7.90	0.10	0.28	162	7.50	315	29.2	0.23	10.5	0.72	7.23	2.30
Salem	7.80	0.10	0.11	189	10	378	33.20	0.32	13.5	0.74	0.32	1.0
Thanjavur	6.90	0.15	0.75	205	75	250	25	1.30	6.80	1.80	2.0	1.0
Thanjavur	7.70	0.17	0.69	189	75	300	20.60	1.60	7.40	0.90	2.0	1.0
Trichy	7.0	1.10	0.25	179	53.0	195.0	12.20	1.06	5.72	0.96	3.48	0.10
Trichy	6.80	0.40	0.19	147	25.0	185.0	12.60	0.95	6.27	0.91	5.76	0.10

District	pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B
Ariyalur	H	L	M	L	M	H	M	H	L	H	M	H
Ariyalur	L	L	L	L	M	L	H	L	M	L	M	M
Coimbatore	L	L	L	L	M	H	L	L	M	H	M	L
Coimbatore	H	L	L	L	M	H	H	L	M	M	M	L
Karur	H	L	L	L	L	M	L	L	L	L	L	H
Karur	H	M	L	L	L	M	L	L	L	L	L	H
Salem	H	L	L	L	L	H	H	L	H	L	H	H
Salem	H	L	L	L	L	H	H	L	H	L	L	M
Thanjavur	M	L	H	L	H	M	H	L	M	H	M	M
Thanjavur	H	L	M	L	H	H	H	M	M	L	M	M
Trichy	M	M	L	L	H	M	M	L	M	L	M	L
Trichy	M	L	L	L	H	M	M	L	M	L	H	L

Table2:COMPARATIVE ANALYSIS OF CLASSIFIERS BASED ON NUTRIENTS

V.CONCLUSION

Maintaining the soil resource base's health is essential for long-term growth because it Farmers will have the ability to The generation of yield is mostly the result of the nutrients in the soil. Determining the kinds of crops that may be planted in a given soil type to provide a satisfactory yield depends on an analysis of the nutrients present in the soil. The healthiest and most abundant food comes from the best soils. Based on the literature, we draw the conclusion that the fields of data mining and remote sensing offer a wide range of algorithms and prognostication techniques for the categorization of soil attribute data. Thus, utilizing algorithms like K-Nearest Neighbor, the temperature and humidity data acquired aided in the process of climate prediction and categorization. It facilitates the discovery of hidden patterns. Several different satellites are utilized to derive water characteristics from satellite pictures. Some of the results are discussed. As a result, a proposal was made to make an attempt by taking into account both the current problems and the potential for extraction from water bodies in the future.

References

[1] Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEEaccess,8, 86886-86901.

[2] Bhuyar, V. (2014). Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District. Int. J. Emerg. Trends Technol. Comput. Sci, 3(2), 200-203.

[3] Baskar, S. S., Arockiam, L., & Charles, S. (2013). Applying data mining techniques on soil fertility prediction. International Journal of Computer ApplicationsTecnologyandResearch,2(6),660-662.

[4] Gholap, J. (2012). Performance tuning of J48 Algorithm for prediction of soil fertility. arXiv preprint arXiv:1208.394*3.*

[5] Gholap,J.,Ingole, A.,Gohil,J., Gargade,S.,& Attar, V. (2012). Soil data analysis using classification techniques and soil attribute prediction. ArXiv preprint arXiv:1206.1557.

[6] Suman,B.B.N.,&Naib,B.(2013).Soil classification and fertilizer recommendation using WEKA. IJCSMS International Journal of Computer Science & Management Studies, 13(05).

[7] Hemageetha, N., &Nasira, G. M. (2016). Analysis of soil condition based on pHvalue using classification techniques .IOSRJ Compute Eng (IOSR-JCE),18(6), 50-54.

[8] Bhargavi, P., &Jyothi, S. (2011). Soil classification usingdataminingtechniques:acomparative study. *International Journal of Engineering Trends andTechnology*,*2*(1),55-59.

[9] Rajeswari, V., &Arunesh, K. (2016). Analysing

soil data using data mining classification techniques. Indian journal of science and Technology, 9(19), 1-4.

[10] Noble,I., Bolin,B., Ravindranath,N., Verardo, D., &Dokken, D. (2000). Land use, land use change, and forestry. Environmental Conservation, 28(3),284-293.

[11] Henderson,F.M.(1995).Environmental factors and the detection of open surface water areas with X-band radar imagery. International Journal of Remote Sensing, 16(13),2423-2437.(13),pp.2423–2437,September1995.

[12] Nath,R.K.,& Deb,S.K.(2010).Water-body area extraction from high resolution satellite images- an introduction,review, and comparison. International Journal of Image Processing (IJIP), 3(6), 265-384.

[13] Prakash, A., & Gupta, R. P. (1998). Land-use mapping and change detection in a coal mining area- a case study in the Jharia coal field, India. International journal of remote sensing, 19(3), 391-410.

[14] Zhang, Q., Wang, C., & Shinohara, F. (2007, November). Automatic extraction of water body based on EOS/MODIS remotely sensed imagery. In MIPPR 2007: Automatic Target Recognition and Image Analysis; and Multispectral Image Acquisition (Vol. 6786, pp. 1068-1074). SPIE..

[15] Nath, R. K., & Deb, S. K. (2010). Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. International Journal of Image Processing (IJIP), 3(6), 265-384.

[16] Li,S.,Peng,S.,Chen,W.,&Lu,X.(2013).INCOME: Practical land monitoring in precision agriculture with sensor networks. Computer Communications, 36(4), 459-467.

[17] Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What are we assessing when we me asure food security? A compendium nd review of current metrics. Advances in nutrition, 4(5), 481-505.

[18] Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., &Hutjes, R. W. (2018). Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. Agricultural and forest meteorology, 250, 243-261.

[19] Holzman, M. E., Carmona, F., Rivas, R., &Niclòs, R. (2018). Early assessment of crop yield from remotely sensed water stress and solar radiation data. ISPRS journal of photogrammetric and remote sensing, 145, 297-308.

[20] Singh, A., Ganapathy subramanian, B., Singh, A. K., & Sarkar,S.(2016).Machine learning for high-throughput stress phenol typing in plants. Trends in plant science, 21(2), 110-124.