International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*

Local Industrialization Based Lucrative Farming Using Machine Learning Technique

Sakshi A. Patil[1], Dr. Mrunal S.Bewoor[2], Mrs. Sheetal S. Patil[3], Dr. Rohini B. Jadhav[4], Dr. Avinash M. Pawar[5], Mrs. Sonali D. Mali[6], Dr. Amol K. Kadam[7]

[1]Department of Computer Engineering,
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
sapatilpg21-comp@bvucoep.edu.in

[2]Department of Computer Engineering,
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
msbewoor@bvucoep.edu.in

[3]Department of Computer Engineering,
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
sspatil@bvucoep.edu.in

[4] Department of Information Technology
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
rbjadhav@bvucoep.edu.in

[5]Bharati Vidyapeeth's College of Engineering for Women, Pune, India
avinash.m.pawar@bharatividyapeeth.edu

[6]Department of Information Technology
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
sdmali@bvucoep.edu.in

[7]Department of Computer Engineering,
Bharti Vidyapeeth Deemed to be university College of Engineering Pune, India
akkadam@bvucoep.edu.in

Abstract—In recent times, agriculture have gained lot of attention of researchers. More precisely, crop prediction is trending topic for research as it leads agri-business to success or failure. Crop prediction totally rest on climatic and chemical changes. In the past which crop to promote was elected by rancher. All the decisions related to its cultivation, fertilizing, harvesting and farm maintenance was taken by rancher himself with his experience. But as we can see because of constant fluctuations in atmospheric conditions coming to any conclusion have become very tough. Picking correct crop to grow at right times under right circumstances can help rancher to make more business. To achieve what we cannot do manually we have started building machine learning models for it nowadays. To predict the crop deciding which parameters to consider and whose impact will be more on final decision is also equally important. For this we use feature selection models. This will alter the underdone data into more precise one. Though there have been various techniques to resolve this problem better performance is still desirable. In this research we have provided more precise & optimum solution for crop prediction keeping Satara, Sangli, Kolhapur region of Maharashtra. Along with crop & composts to increase harvest we are offering industrialization around so rancher can trade the yield & earn more profit. The proposed solution is using machine learning algorithms like KNN, Random Forest, Naïve Bayes where Random Forest outperforms others so we are using it to build our final framework to predict crop.

Keywords-prediction; machine learning models; cultivation; feature selection models; industrialization.

I. INTRODUCTION

The use of machine learning technologies in different fields for the growth of industry & country is rapidly increasing nowadays. Growth of country also depends on food industry because with increasing population and urbanization available agricultural land is limited & will remain same. Nowadays farmers financial situation is also highly unstable because not taking right crop at right time and right place. Because of this they have to face financial loss very often. So, the aim is to get more and more yield from that available land with maximum profit. For this we need to make changes in farming methodology which we have been using traditionally. Problem with this is we don't exactly know about many factors which affect the growth of crop.

In recent years, researchers have been working on multiple technologies which can help us to understand and control those factors. One of the technologies is machine learning with which we can analyze previously available data and can predict for future. Here factors affecting crop growth like soil health, pH, humidity, soil composition, temperature, rainfall are highly area

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*

prone. They change from place to place and to predict crop based on these requires collection of huge datasets which is difficult task. We can tell from this that crop prediction is difficult task & requires many steps to be followed. Recently there are many new crop prediction models are being introduced but better performance is still desirable. So, after getting motivated from machine learning models like feature selection and classification we have decided to help farmers by suggesting which crop to grow based on analyzing data of their region like soil fertility, composition, humidity, rainfall, temperature, with available industries nearby for maximization of profit. We will also suggest the fertilizers to be used for better yield of crop. For easy use of proposed system, we are creating web application using Django.

II. RELATED WORK

This section gives overall analysis of diverse available procedures used by scholars for crop recommendations.

In this paper Sahu S., et al. [1], proposed a model which predicts suitable crop after considering various parameters of soil & atmosphere. Model is divided in 4 stages where data is collected & classified using Random Forest algorithm on which Hadoop (2.6) framework is applied to handle the huge data using map-reduce which will finally predict the crop. The model gave accuracy of 0.9143.

Model by Paulo Vitor Duarte de Souza, et al. [2], proposes models of multilayer perceptron's to predict corn productivity by considering climatic data, soil features, soil-water balance.

Model by S. M. Usha, et al. [3] presents the efficient watering and monitoring method to increase crop yield using IOT, neural networks, and image processing.

In this paper Ngozi Clara Eli-Chukwu [4], provides review various of AI technologies which have been used in field of agriculture for managing several sectors like disease, weed, soil, crop by analyzing their strengths & limitations.

Model projected by Monika G., et al. [5], focuses on suggesting crop to be taken in field based on soil contents and climatic parameters using models like SVM, Naïve Bayes (NB), Decision Trees & Logistic Regression. Where NB shows best accuracy.

In this paper Rakesh K. R., et al. [6] have proposed three-tiered architecture. First module is data pre-processing & feature extraction where they have selected features like nitrogen, phosphorus, potassium, temperature, rainfall, humidity, pH level. Second module is analysis & classification of data using Naïve Bayes, random forest, decision tree, SVM. Third module is model testing & validation using visualization of confusion matrix.

In the experiment proposed by G. Mariammal, et al. [7] they have evaluated the efficiency of MRFE technique using KNN, NB, DT, SVM, RF and bagging classification techniques to predict suitable crop for farming. This work analyses performance of different classifier with MRFE technique to achieve better result.

The model proposed by Nischitha K, et al. [8] predicts rainfall and crop to cultivate using AI. They have used SVM and Decision tree algorithms for it. They also have prepared a user interface for easy usability of technology for user with some explanation of predicted crop to make it understandable.

III. PROPOSED FRAMEWORK

- Dataset:
  The system works on crop prediction based on soil properties and other factors like area and industrial belt where farmers can sell the crop faster to increase the profit margin. So, first step for us was to get an appropriate dataset related to our work. We obtained the dataset from Kaggle.com [9], [10] and altered it with data available on government websites for Industrialization information. Here we have considered data related to soil's chemical composition such as pH, Nitrogen, Phosphorus, Potassium & atmospheric features such as Humidity, Rainfall, Temperature. Provided input data will be in CSV format for further procedure.

- Pre-processing:
  The dataset obtained from the online websites for our work was in a raw CSV (comma-separated values) format, which was containing lot of missing values and unrelated data. Which might become false indicator for end result. Hence in this module, we need to work on the gathered dataset. To align it in proper format we need to do its pre-processing using various python libraries such as pandas, NumPy, etc. It will to eliminate the noise, irregularities, and unrelated information. This phase is crucial in guaranteeing that the data is precise and reliable.

- Data Analysis:
  In this module, once the dataset is cleaned and ready for further actions, but before that we need to analyze the dataset to decide on which features are affecting the result, which features are more crucial in crop prediction procedure & which features are not affecting our end goal. Using graphical analysis with the help of Seaborn and matplotlib libraries of python, we did all the necessary analysis of the data for further framework implementation.

- Feature Selection:
  Feature engineering is a machine learning technique that influences data to generate new variables that aren't in the training set. It can produce new features for both supervised

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*

and unsupervised learning, with the aim of making things easier and fastmoving data alterations while also improving prototype accuracy. Feature Selection is needed when working with machine learning methods. Irrespective of the data or architecture, a dreadful feature will have a direct influence on final model. This step involves identifying the most momentous features that have the maximum impact on growth of the crop. These factors affect the yield and profit at last.

- Model Development:
  In this stage, we applied various machine learning algorithms on our modified dataset which is processed to check their individual performances. We have implemented Naïve Bayes, K-Nearest Neighbor & Random Forest to check the performance of algorithm so that we can select the best method for our final model deployment. We have compared result of every algorithm by generating confusion matrix based on precision, recall & F-measure values.
- Model Deployment:
  In this final module, our machine learning model is ready, so that we can feed new data to our model for prediction of crop based on the input dataset. Here we have implemented a website using Django to make the crop prediction procedure user friendly so that it will attract farmers.

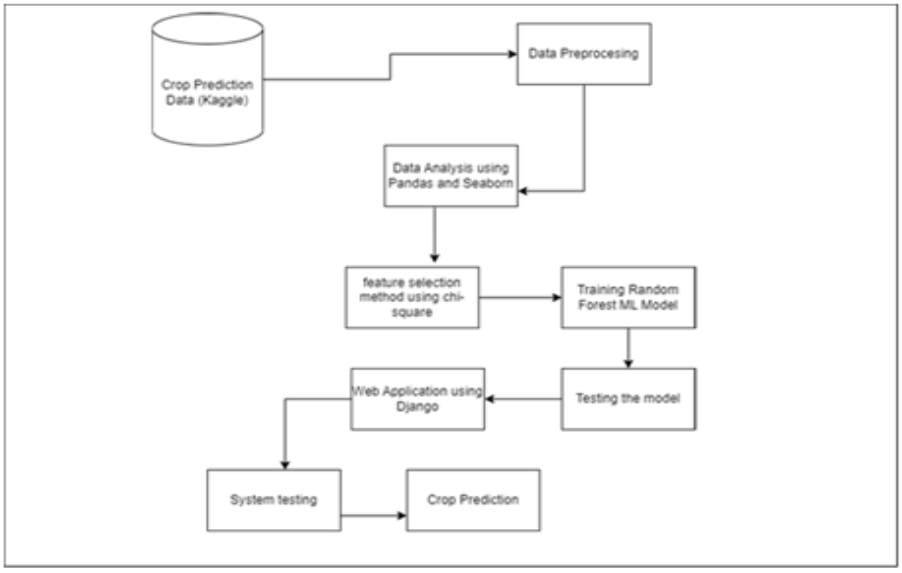


Fig. 1. Workflow of proposed framework

IV. TECHNIQUE USED

Algorithm 1: KNN (K-Nearest Neighbor)

Step 1. After fixing dataset and dividing it in training (80%) and testing (20%) choose the value of X which is nearest data point. X can be any integer.

Step 2. For each point in test data do:

2.1- Find detachment between test data & every row of training data with any of the following method:

Euclidean Distance –

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad \textit{(Equation 1)}$$

Manhattan Distance –

$$d(p, q) = \sum_{i=1}^{n} |p_i - q_i| \qquad \textit{(Equation 2)}$$

Minkowski Distance –

$$d(p, q) = \left(\sum_{i=1}^{n}(p_i - q_i)^r\right)^{\frac{1}{r}} \qquad \textit{(Equation 3)}$$

Most commonly used on is Euclidean method.

2.2- Now based on value of detachment sort them in non-decreasing order

2.3- Now it will choose top x rows from sorted array.

2.4- It will assign class to the test point based on most frequent class of these rows.

Step 3. Once the KNN algorithm has been trained on the labeled data, it can be used to predict the labels of new data points

Algorithm 2: Naïve Bayes

Bayes Theorem is stated as following equation:

$$P(P|Q) = \frac{P(Q|P)P(P)}{P(Q)} \qquad \textit{(Equation 4)}$$

Here P & Q are events & P(Q) is not equal to Zero.

P(P|Q): Probability of occurrence of event P when event Q is true.

P(P) & P(Q): Probabilities of occurrence of event P & Q respectively.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*

P(Q|P): Probability of occurrence of event Q when event P is true.

P is class variable called as prior probability and Q is dependent feature vector called evidence.

P(P) is prior probability of proposition.

P(Q) is prior probability of evidence.

P(P|Q) is posterior

P(Q|P) is the likelihood.

$$Posterior = \frac{(Likelihood).(Proposition\ prior\ probability)}{Evidence\ prior\ probability} \quad \textit{(Equation 5)}$$

Algorithm 3: RF (Random Forest)
Step 1. Casually select "r" features from all the "x" features. Only condition is that r<<x.

Step 2. Out of the "r" features find node "n" using finest split point.

Step 3. Again, split the node into daughter nodes using best split.

Step 4. Repeat above steps till "l" number of nodes been reached.

Step 5. Build forest by repeating steps 1 to 4 for "m" number of trees.

To predict using trained random forest algorithm:
Step 1. Take test features and use rules of each randomly created DT to predict outcome & stores predicted outcome.

Step 2. Calculate votes for each predicted target.

Step3. Consider high voted predicted target as final prediction from random forest algorithm.

Each random forest will predict different targets for same test feature. Then after considering each predicted target votes will be calculated.

Here if 1000 random DTs are forecasting 3 unique targets x, y, z and x is getting high votes say 600 trees are forecasting target x. then final random forest returns x as projected target. This is called majority voting.

V. TECHNIQUE EVALUATION PARAMETERS

The summary is evaluated with co-selection methods such as Precision, Recall and F-measure. Precision and recall are quality measure which considers all the retrieved documents. In case of performance evaluations of the summarizers precision can be used to estimate to what degree the result obtained satisfy the user requirement. The summary generated through this technique considers user issued query. These can be signified precisely as revealed in the following equations.

$$Precision = \frac{\text{No .Of term} \in \text{Summary}}{\text{No.of different terms} \in \text{the Query}} \quad \textit{(Equation 6)}$$

$$Recall = \frac{\text{No.Of matching sentences} \in \text{Summary}}{\text{No.of relevant sentences} \in \text{the whole data}} \quad \textit{(Equation 7)}$$

F-measure is choral mean of Precision and Recall factors; both have equal importance.

$$F\text{–}measure = \frac{2*(\text{Precision}*\text{Recall})}{\text{Precision}+\text{Recall}} \quad \textit{(Equation 8)}$$

F-measure varies between 0 and 1. With 0 as worse and 1 for better having larger value of F-measure nearer to 1.

$$Compression\ Ratio = \frac{\text{No,of terms} \in \text{T}}{\text{No.of total terms} \in \text{data}} \quad \textit{(Equation 9)}$$

$$Retention\ Ratio = \frac{\text{Information} \in \text{summary}}{\text{Information} \in \text{data}} \quad \textit{(Equation 10)}$$

Here 'T' is the summarized text. The summary can be considered as good if the compression ratio CR is low and the retention ratio (RR) is high. This compression ratio and retention ratio are the quantity measuring metrics to evaluate the methods. Along with all these measures the algorithms are evaluated for the CPU time required to complete the task.

VI. RESULT & DISCUSSION

After careful implementation of algorithms on derived & processed dataset we have come up with Fig. 2. which graphically shows accuracy comparison means performance difference between different algorithms. Here on X-axis, we are taking accuracy on the scale of zero to one while on Y-axis we are comparing different AI algorithms.

Similarly, as revealed in Fig. 3. We are comparing this accuracy numerically considering different parameters like precision, recall, f1-score & support factors. Naïve Bayes & KNN both gives 97.85% accuracy where Random Forest gives 100%.

Thus, from analysis we can see that Random Forest outperforms other two. This is why we have implemented the website by applying Random Forest algorithm to show final & more precise result for crop prediction.

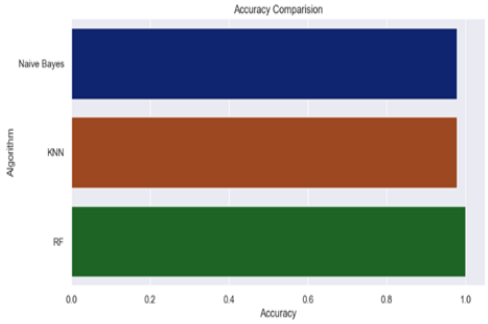
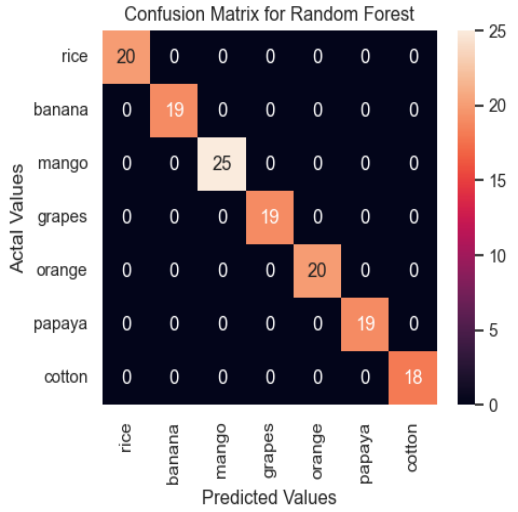
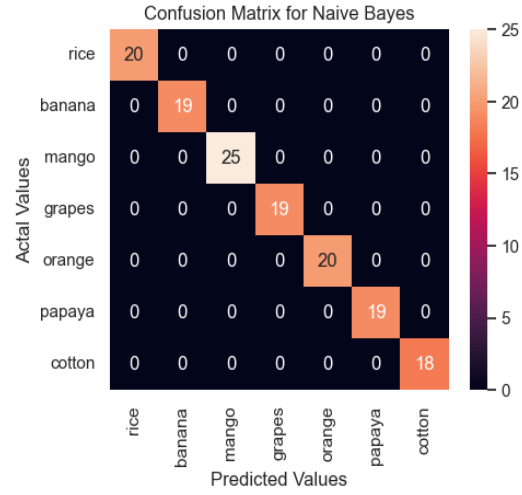


Fig. 2. Performance Analysis of different algorithms

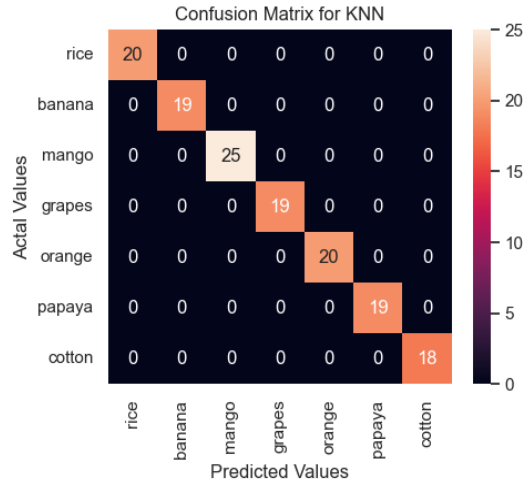
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*




```
RF's accuracy is:  100.0 %
              precision    recall  f1-score   support

      banana       1.00      1.00      1.00        20
      cotton       1.00      1.00      1.00        19
      grapes       1.00      1.00      1.00        25
       mango       1.00      1.00      1.00        19
      orange       1.00      1.00      1.00        20
      papaya       1.00      1.00      1.00        19
        rice       1.00      1.00      1.00        18

    accuracy                           1.00       140
   macro avg       1.00      1.00      1.00       140
weighted avg       1.00      1.00      1.00       140
```




```
Naive Bayes's Accuracy is:  97.85714285714285 %
              precision    recall  f1-score   support

      banana       0.91      1.00      0.95        20
      cotton       0.95      1.00      0.97        19
      grapes       1.00      1.00      1.00        25
       mango       1.00      1.00      1.00        19
      orange       1.00      1.00      1.00        20
      papaya       1.00      0.89      0.94        19
        rice       1.00      0.94      0.97        18

    accuracy                           0.98       140
   macro avg       0.98      0.98      0.98       140
weighted avg       0.98      0.98      0.98       140
```




```
KNN's Accuracy is:  97.85714285714285 %
              precision    recall  f1-score   support

      banana       0.91      1.00      0.95        20
      cotton       0.95      1.00      0.97        19
      grapes       1.00      1.00      1.00        25
       mango       1.00      1.00      1.00        19
      orange       1.00      1.00      1.00        20
      papaya       1.00      0.95      0.97        19
        rice       1.00      0.89      0.94        18

    accuracy                           0.98       140
   macro avg       0.98      0.98      0.98       140
weighted avg       0.98      0.98      0.98       140
```

Fig. 3. Accuracy comparison of algorithms

Also, Fig. 4. shows web application created using Django to make the system attractive for end user and helps to make system popular.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
Article Received: 01 June 2023 Revised: 26 July 2023 Accepted: 10 August 2023

Fig. 4. Web-Application UI

VII. CONCLUSION AND FUTURE WORK

This study recommends a system to predict crop using machine learning. This system uses an approach of classification type of machine learning. First the input is given in the csv format to the system and output is given by type of crop prediction which will help the farmer to select right type of crop for production purpose based on various soil properties and industrial area nearby. It studies the various soil properties and not only that it studies various nearby industries in that area which will help the farmer to sell the goods produced easily, increasing their profit margin.

The developed system can be expanded in any desired fields to increase profit margin of the farmer by helping them to select

right kind of crop to produce based on various soil properties. Moreover, the system can be further expanded for all types of crops and all types of area across nation. Currently we have only taken 3 districts Sangli, Satara, Kolhapur from Maharashtra state for study purpose.

ACKNOWLEDGMENT

An effective crop prediction model has been prepared by miss Sakshi Patil and Professor Mrs. Mrunal Bewoor. Author wants to express gratitude to all the staff members as well as entire section, family, and friends for their backing and assurance for sharing lot of knowledge during the groundwork of this article.

REFERENCES

[1] S. Sahu, M. Chawla and N. Khare, "An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, pp. 53-57, 2017, https://doi.org/10.1109/CCAA.2017.8229770.

[2] P. V. Duarte de Souza, L. Pereira de Rezende, A. Pereira Duarte, and G. V. Miranda, "Maize Yield Prediction using Artificial Neural Networks based on a Trial Network Dataset," Engineering, Technology & Applied Science Research, vol. 13, no. 2, pp. 10338–10346, Apr. 2023, https://doi.org/10.48084/etasr.5664.

[3] Usha, S. & Mahesh, H., "Monitoring and Analysis of Agricultural Field Parameters in Order to Increase Crop Yield through a Colored Object Tracking Robot, Image Processing, and IOT," Engineering, Technology & Applied Science Research, vol. 12, no. 4, pp. 8791-8795, Aug. 2022, https://doi.org/10.48084/etasr.5028.

[4] Eli-Chukwu, Ngozi, "Applications of Artificial Intelligence in Agriculture: A Review," Engineering, Technology and Applied Science Research, vol. 9, no. 4, pp. 4377-4383, Aug. 2019, https://doi.org/10.48084/etasr.2756.

[5] M. Gupta, S. K. B. V, K. B, H. R. Narapureddy, N. Surapaneni and K. Varma, "Various Crop Yield Prediction Techniques Using

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10s
DOI: https://doi.org/10.17762/ijritcc.v11i10s.7626
***Article Received:** 01 June 2023 **Revised:** 26 July 2023 **Accepted:** 10 August 2023*

Machine Learning Algorithms," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, pp. 273-279, Feb. 2022, https://doi.org/10.1109/ICAIS53314.2022.9742903.

[6] Muhammad Yusuf R. Siahaan, Rakhmad Arief Siregar, Faisal Amri Tanjung. (2023). Optimized Flexural Strength of Aluminium Honeycomb Sandwiches Using Fuzzy Logic Method for Load Bearing Application. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 466–472. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2704

[7] R. K. Ray, S. K. Das and S. Chakravarty, "Smart Crop Recommender System-A Machine Learning Approach," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022, pp. 494-499, https://doi.org/10.1109/Confluence52989.2022.9734173.

[8] G. Mariammal, A. Suruliandi, S. P. Raja and E. Poongothai, "Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique with Various Classifiers," IEEE Transactions on Computational Social Systems, vol. 8, no. 5, pp. 1132-1142, Oct. 2021, https://doi.org/10.1109/TCSS.2021.3074534.

[9] Mahendra N, Dhanush Vishwakarma, Nischitha K, Ashwini, Manjuraju M. R, "Crop Prediction using Machine Learning Approaches," International Journal of Engineering Reasearch & TECHNOLOGY (IJERT), vol. 9, no. 8, pp. 23-26, Aug. 2020, https://doi.org/10.17577/IJERTV9IS080029.

[10] https://www.kaggle.com/datasets/gdabhishek/fertilizer-prediction

[11] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset