International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 11, November 2025)

Machine Learning for Smart Harvest Agri Powered System for Ecological Farming

Dr. S. Brindha[1], Dr. V. R. Elangovan[2], Dr. D. Helen[3], Sajith Balaji Sarvesh[4]

[1,2,3]Assistant Professor, Department of Computer Applications, SRMIST Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM University, KTR, Tamilnadu, India.

[4]Student, Department of Computer Science and Engineering, KarpagaVinayaga College of Engineering and Technology, Chengalpattu, Tamilnadu, Chennai, India

***Abstract*-- Agriculture stands as the lifeblood of humanity, feeding nations, driving economies, and fostering livelihoods worldwide. Amidst escalating challenges, this innovative Agri-system focuses on optimizing crop management while ensuring environmental sustainability. In order to transform farming practices in line with Sustainable Development Goal 2: Zero Hunger, Smart Harvest, an inventive Agri-system, combines machine learning (ML), deep learning, and data-driven intelligence. Acknowledging the critical significance of agriculture, it concentrates on four fundamental foundations. First of all, it provides farmers with insights for resource allocation and strategic planning by properly forecasting crops using historical data, weather patterns, and soil analyses. Second, it uses machine learning algorithms to create crop fertilizers that help farmers choose environmentally friendly crops. Thirdly, it detects and forecasts possible plant illnesses using image recognition and machine learning algorithms, allowing for preemptive interventions to maintain crop productivity. Fourth, by encouraging cooperation among farmers via an easy-to-use interface, it places a high priority on community involvement. Smart Harvest highlights the vital role that agriculture plays in creating a successful and nourished world by redefining farming methods, empowering farmers, ensuring global food security, and promoting sustainable agricultural practices. The purpose of this study is to present a technological app for smart harvesting that can determine the appropriate crops, soil level, disease prediction, and harvesting technique. The method of text identification and image recognition is also used when we utilize this smart harvesting program, and it is incredibly user-friendly.**

***Keywords*-- Machine Learning, Web Crawling, Web Usage Mining, BIRCH, Advanced Level KNN.**

I. INTRODUCTION

The agriculture sector plays a pivotal role in sustaining global food security and economic stability. With advancements in technology, the agricultural landscape is witnessing a transformative shift towards smart and data-driven practices.

The Smart Harvest App emerges as a beacon of innovation, empowering farmers and agriculture enthusiasts with cutting-edge tools and insights to optimize productivity and mitigate risks. This comprehensive platform integrates various features and functionalities to cater to the diverse needs of modern agriculture.

1. Crop Prediction

One of the core features of the Smart Harvest App is its ability to predict the most suitable crops based on environmental parameters. Leveraging a sophisticated Random Forest model, the app analyzes factors such as nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall. By providing personalized crop recommendations, the app enables farmers to make informed decisions about crop selection, leading to improved yields and resource utilization.

2. Plant Disease Detection

Early detection and management of plant diseases are critical for preserving crop health and optimizing yields. The Smart Harvest App integrates a pre-trained deep learning model capable of accurately identifying and classifying plant diseases. By uploading images of diseased plant leaves, users can receive prompt diagnoses and recommended treatment strategies, thereby minimizing crop losses and promoting sustainable farming practices.

3. Fertilizer Prediction

System Optimal nutrient management is essential for ensuring healthy plant growth and maximizing yields. The app's Fertilizer Prediction System utilizes machine learning algorithms to recommend suitable fertilizers based on soil type, crop selection, and nutrient levels.

4. Agri News

Updates Staying informed about the latest agricultural trends, policies, and market developments is crucial for decision-making.

The Smart Harvest App features an integrated news feed that delivers realtime updates on agriculture-related news, innovations, and best practices. Users can access relevant articles, reports, and insights to stay abreast of industry trends and make strategic decisions.

5. Interactive Agri

Chatbot The app's interactive chatbot serves as a virtual agricultural assistant, offering personalized advice, recommendations, and solutions to users' queries. Whether it's optimizing crop management practices, accessing government schemes, or seeking expert guidance, the chatbot provides valuable support and fosters knowledge-sharing within the agricultural community. The Smart Harvest App represents a paradigm shift in agricultural technology, empowering farmers with data-driven solutions, predictive analytics, and real-time information.

II. RELATED WORKS

Numerous clustering algorithms have been created for various uses. Partitioning clustering, hierarchical clustering, density-based methods, grid-based methods, and model-based clustering methods are the several types of clustering techniques. K-means, K-medoids, PAM, CLARA, and CLARANS are examples of partitioning clustering algorithms that divide items into K clusters and then iteratively reallocate those objects to enhance the quality of the clustering results.

In order to protect the data points in the legislature of low level clusters, the hierarchical clustering algorithms distribute the content among three pre-prepared clusters. The concept behind density-based clustering techniques is that there are at least a minimal number of clusters in the neighborhood of a given unit of distance for every cluster point [7]. Density-based clustering is a concept used in machine learning techniques to identify each cluster and determine which neighborhood of a given unit distance has the fewest possible points. Various forecasting techniques have been created and assessed by researchers worldwide in the agricultural field.

Agriculture data for 2015–2023 is analyzed in a number of papers by Ramesh and Vishnu Vardhan. The K means clustering method is used to group rainfall data into four clusters. The technique for modeling the linear relationship between a dependent variable and one or more independent variables is called multiple linear regressions, or MLR.

Rainfall is a necessary variable, and the year, the sowing area, and the production are self-governing factors. Finding appropriate data models that accomplish both high accuracy and high simplification in terms of capitulate prediction skills is the rationale behind this effort [9].

III. EXISTING SYSTEM

Traditional manual farming methods are the mainstay of the current agricultural system, and farmers base their decisions on observational data, their own experience, and guidance from regional agricultural services. The majority of data gathering techniques are done by hand, with very little technology used for pest identification, crop monitoring, or soil sample. While disease and pest control tactics frequently rely on visual examination and conventional remedies, crop prediction and management are dependent on historical trends rather than sophisticated predictive algorithms.

Rather than using a precise study of soil nutrients, fertilizer application is guided by general guidelines. There are few digital venues for real-time information access, and information sharing takes place through traditional knowledge networks and local extension organizations. Limited data-driven insights, wasteful resource consumption, and susceptibility to market and climate volatility are among the difficulties.

IV. PROPOSED WORK

A paradigm change in agriculture is represented by the planned Smart Harvest system, which uses cutting-edge technologies to transform farming methods. The system's primary function is to collect real-time data on crop status, soil health, and environmental conditions by integrating IoT (Internet of Things) devices including weather stations, drones, and soil sensors. To give farmers useful insights and suggestions, this data is further processed using sophisticated machine learning algorithms. Accurate predictions of crop types and growth patterns are made possible by the system's crop prediction model, which has been trained on historical data and environmental factors. A deep learning algorithm that examines plant photos to detect and categorize diseases early on improves disease diagnosis and enables focused interventions. To maximize nutrient use efficiency and crop yields, the system also provides customized fertilizer recommendations based on soil nutrient analysis, crop requirements, and environmental conditions.

Farmers can easily access information, receive warnings, and engage with the Smart Harvest system thanks to its user-friendly interface, which was created utilizing web technologies and the Stream light framework. Scalability, data storage, and accessibility from any location are guaranteed via cloud integration, while system performance and dependability are ensured by routine upgrades and maintenance. All things considered, the suggested system ushers in a new era of smart agriculture by providing farmers with data-driven decision-making tools, precision farming methods, and sustainable agricultural practices.

4.1 Random Forest Classifier Used for Crop and Fertilizer Prediction

A potent machine learning approach for classification and regression applications is the Random Forest Classifier. The Random Forest Classifier is used in your Smart Harvest project to predict crops and propose fertilizer. This is a thorough description of the Random Forest Classifier's operation and function in your project:

1. Random Forest Classifier Overview: Several decision trees are combined in Random Forest, an ensemble learning technique, to generate predictions. The reason it's termed "random" is that it uses a random subset of training dataset attributes and samples to generate each tree in the forest. This unpredictability enhances generalization and lessens overfitting.

2. Training Process: Training Procedure: The Random Forest Classifier creates a forest of decision trees during the training stage. A random feature pick and a portion of the training data are used to build each decision tree. This unpredictability guarantees that every tree is distinct and identifies distinctive patterns in the data.

3. Decision Making: Based on the input features, each Random Forest tree makes an individual prediction about the target variable (crop kind or fertilizer suggestion). The majority voting process is employed for classification tasks such as crop prediction, where the most common prediction among all trees determines the final prediction. The average forecast from all trees is taken into account in regression tasks such as fertilizer suggestion.

4. Significance of Features: The capacity of Random Forest to evaluate the significance of features in the prediction process is one of its main advantages. Based on how much each feature helps reduce impurity (such as Gini impurity for classification) across all decision trees, the algorithm determines the relevance of each feature. Knowing which characteristics have the biggest effects on crop prediction and fertilizer advice is made easier with the help of this information.

5. Managing Outliers and Missing Values: Random Forest can withstand outliers and missing values in the data. Missing values can be handled by imputing them using the dataset's other features. Additionally, compared to single decision tree models, Random Forest's usage of several trees and feature subsets reduces the impact of outliers on the overall model.

6. Scalability and Performance: Random Forest can effectively manage big datasets with many of features and is very scalable. Random Forest is appropriate for both small-scale and huge data applications due to the parallel nature of decision tree training. It is also a popular choice for a variety of machine learning tasks because it typically performs well without requiring a lot of hyperparameter modification.

Agricultural data, including soil characteristics, weather patterns, nutrient levels, and past yield data, are used to train the Random Forest Classifier. Based on input environmental data, it forecasts the most likely crop type. It makes recommendations for appropriate fertilizers according to crop type, soil type, and nutrient needs. Random Forest's ensemble nature and randomness help produce reliable and accurate predictions, which improves the Smart Harvest App's usability and functionality for farmers.

4.2 Objectives of Smart Harvest App

The Smart Harvest App's main goal is to transform the agriculture industry by utilizing cutting-edge technologies and data-driven solutions to solve major issues that farmers confront and develop sustainable farming methods. The following goals guided the design of the app:

1. Optimizing Crop Management: By offering precise crop type forecasts based on environmental variables including soil composition, temperature, humidity, and rainfall, the app seeks to optimize crop management procedures. This assists farmers in choosing crops, scheduling plantings, and allocating resources in an informed manner.

2. Improving Disease Detection: Improving plant disease detection is one of the main goals of the Smart Harvest App. The software can detect plant diseases early on by using deep learning algorithms and picture recognition technology, which enables farmers to reduce crop losses and take prompt preventive action.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 11, November 2025)

3. Improving Fertilizer Recommendations: By examining crop types, environmental factors, and soil nutrient levels, the app's Fertilizer Prediction System aims to enhance fertilizer recommendations. This minimizes nutrient waste, guarantees the best possible use of fertilizers, and encourages sustainable farming methods. 4. Encouraging Sustainable Farming: Maintain

4. Promoting Sustainable Agriculture: The Smart Harvest App places a strong emphasis on sustainable agriculture. It strives to encourage techniques that conserve natural resources, limit environmental damage, and boost long-term agricultural productivity. The app helps the agriculture industry become more resilient overall by promoting sustainable farming practices.

5. Empowering Farmers with Technology: Giving farmers access to technology-driven tools and insights is one of the app's other goals. The software helps farmers make better decisions, be more efficient, and earn more money by giving them access to real-time data, predictive analytics, and practical suggestions.

6. System Requirements

OS	**Windows 11**
Processor	**AMD Ryzen 5 5500U with Radeon Grap**
RAM	**8.00GB**
Hard Disk	**500GB**
IDE	**Anaconda Navigator and Google Colab**
Coding Language	**Python**

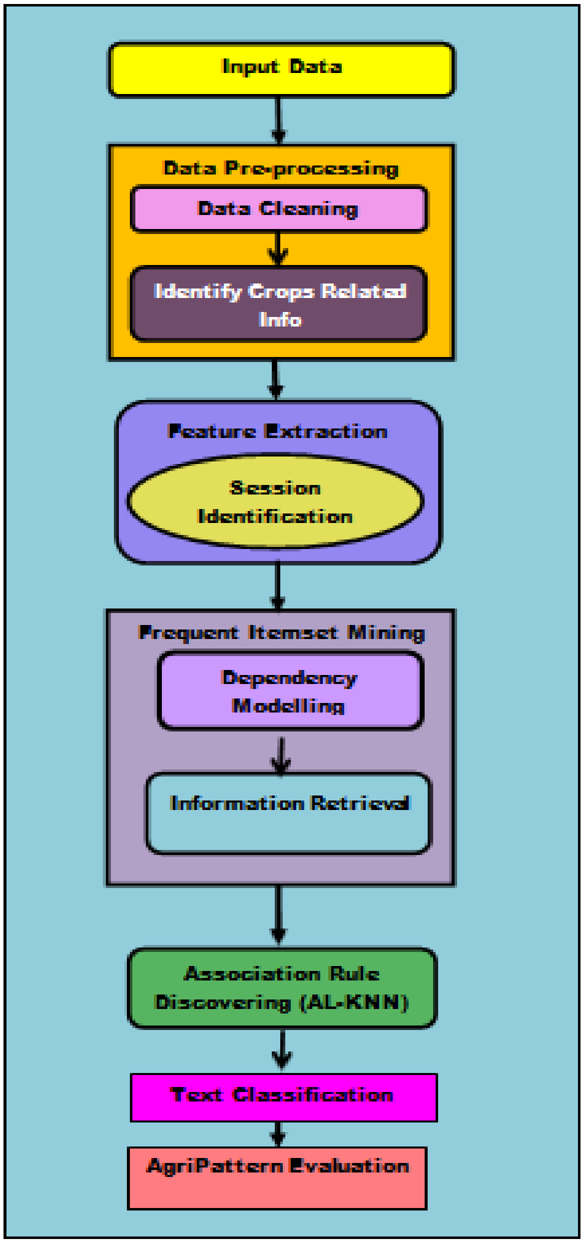

Fig. 3.1Advanced Level K-Nearest Neighbors (AL-KNN) Method usingPattern Discovery Method for smart Harvesting

4.3 Data Cleaning

The most crucial component of superior information management is data quality. Data quality management issues can occur anywhere in information systems. Businesses have been confused about whether they should move forward with data analytics to find out if their data is dirty or whether they should clean their data first. Data cleansing fixes these problems. It is the process of identifying inaccurate, lacking, or inappropriate data and then fixing the errors and omissions found to improve the quality of the data. In general, data cleaning improves data quality and reduces errors. Although it cannot be prevented, fixing inaccurate data and eliminating incorrect entries can be a tedious and time-consuming process.

4.4 Feature Extraction

By automatically evaluating, the bi-level feature extraction-based text mining for fault diagnosis solves the above described problems. Essentially, we take a syntactic and semantic method to extract fault features, which are then combined to achieve the desired results. The proposed feature fusion of two levels may increase the accuracy of fault identification for all fault classes, especially minority ones, since the extracted features at each level have constraints and focus specifically on a particular component of feature spaces. Farmers can easily access information, receive warnings, and engage with the Smart Harvest system thanks to its user-friendly interface, which was created utilizing web technologies and the Stream light framework. Scalability, data storage, and accessibility from any location are guaranteed via cloud integration, while system performance and dependability are ensured by routine upgrades and maintenance. All things considered, the suggested system ushers in a new era of smart agriculture by providing farmers with data-driven decision-making tools, precision farming methods, and sustainable agricultural practices.

4.5 Feasibility Study

As the name suggests, a feasibility study is used to assess an idea's viability, such as making sure this project is technically and legally possible and economically justified.This shows if the project is worth the money. In certain situations, a project could not be feasible.Many factors can contribute to this, such as the need for excessive resources, which not only diverts those resources from other duties but also may be more expensive than an organization would recoup by taking on a project that isn't viable.

In addition to finding a solution, the feasibility study aims to gain an understanding of the problem's extent. The problem definition is solidified during the study, and the elements of the problem that should be incorporated into the system are identified. Benefits are therefore calculated more precisely at this point. The main things to think about are:

- Economic feasibility
- Technical feasibility
- Operational feasibility.

4.5.1 Economic Feasibility

Studies of economic viability Here, the advantages in the form of lower expenses are taken into account in addition to the price of the gear and software. This project will undoubtedly be advantageous since the suggestion system will boost productivity and decrease manual labor. It will save a great deal of time and doesn't require any extra hardware resources.

4.5.2 Technical Feasibility

Technical viability assesses the software, hardware, and manpower needs, among other things. Technically, this project is definitely doable. The equipment specification is a major focus of this project, which will effectively meet nearly all user needs. Although the technical requirements for this system may differ greatly. The ability to generate results immediately upon entering input in the tab. b. When a patient has diabetes, heart disease, or Parkinson's disease, it is used to anticipate the disease. c. The capacity to handle data with ease. As a result, every data's fundamental input and output are identified. Therefore, the project will be technically feasible and easy to build up.

4.5.3 Operational Feasibility

Only if the suggested system can be transformed into information systems that satisfy organizational needs will it be advantageous. This approach helps with project decision-making, minimizes manual labor, and promotes the production of quality outcomes.

4.5.4 Frequent Item Set

The three categories of statistical techniques are univariate, multivariate, and hybrid. Feature filtering techniques, also referred to as univariate procedures, take features into account separately. Information gain (IG), chi-square, occurrence frequency, log probability, and minimum frequency threshold are a few examples of this category.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 14, Issue 11, November 2025)

Despite their computational efficiency, univariate techniques ignore attribute relationships. Multivariate methods consider a set of attributes and select attributes using a wrapper model. Examples of these methods include decision tree models, recursive feature removal, and evolutionary algorithms. Compared to univariate methods, multivariate methods require more computing power because they look at attribute interactions. To increase precision and effectiveness, hybrid approaches integrate univariate, multivariate, and other techniques.

4.5.5 Advanced Level K- Nearest Neighbour

A popular supervised learning technique for developing different machine learning models that aim to predict class labels in a linked dataset is classification. The machine learning literature has presented a number of classification methods for a variety of uses. Neural networks, support vector machines, decision trees, random forests, and naive Bayes are the most popular classification techniques. Moreover, a strong classification method that may also be applied to regression issues is the Advanced Level K-Nearest Neighbors (AL-KNN) approach. It shows that until a certain test question is encountered for prediction using this method, the training dataset will not be completely processed. In fact, a comprehensive search in large dimensions may be carried out by Advanced Level K-Nearest Neighbors (AL-KNN). To find the K-closest tuples to a test query, the Advanced Level K-Nearest Neighbors (AL-KNN) classifier envisions the training set as an m-dimensional space of patterns.

V. Results And Discussions

The ability of an IR system to return pertinent documents, as well as the accuracy and precision of these retrieved documents, is commonly measured.

$$Precision \frac{|relevant\ \ documents\ \ \cap retrieved\ \ documents\ |}{|retrieved\ \ documents\ |} \quad \text{---- (1)}$$

The second measure is Recall. It is the proportion of documents that are related to the query and have been found.

$$Recall = \frac{|relevant\ \ documents\ \ \cap retrieved\ \ documents\ |}{|relevant\ \ document\ |} \quad \text{------(2)}$$

These binary measures benefit to computeadditional information retrieval metrics which is F-measure

$$F - measure = \frac{2 * precision\ * recall}{Precision\ + recall} \quad \text{-------(3)}$$

Accuracy is used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition is

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{-----}(4)$$

The accuracy of system is calculated utilizing the cross validation in this method and calculate the values utilizing given formula. The results obtained by the system in first five experiments. Locating the relevant files is without problems utilising of keyword looking.

Table. 5.1
Advanced Level K-NN Classification Method for smart harvesting

Methods	Precision (%)	Recall (%)	F-Measure (%)	Accuracy (%)
SVM	**89.24**	**87.45**	**88.23**	**89.54**
PDM	**90.12**	**90.34**	**87.59**	**91.67**
Advanced Level K-NN	**90.23**	**90.16**	**92.67**	**95.12**

In table 5.1 describes the existing methods as Support vector machine and pattern discovery method and Advanced Level K-NN is proposed method is better accuracy rate. While comparing the above methods the proposed method gives high accuracy and its F-Measure value also increased.The text recognition and image recognition processes are also applied while we are utilizing this smart harvesting application due to the fact that this app is completely user-friendly. By providing a comprehensive solution to address major issues encountered by farmers and encouraging sustainable farming methods, the Smart Harvest.

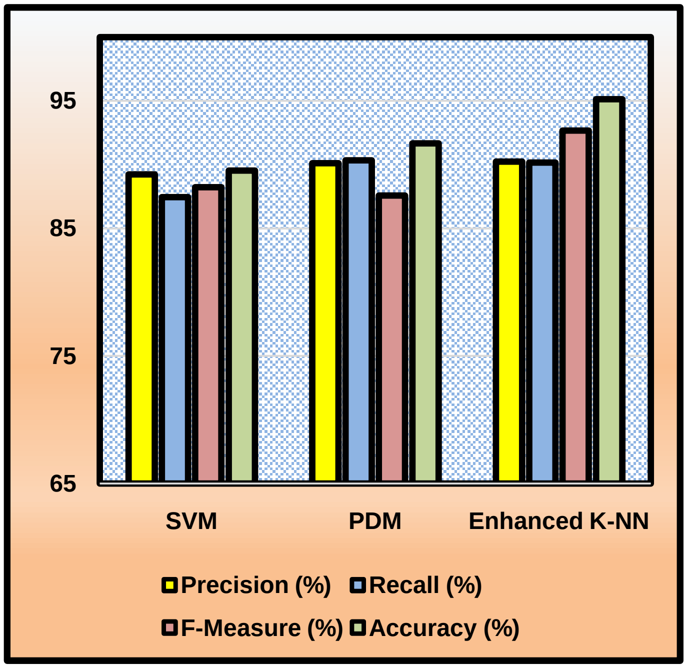

Fig.5.2Advanced Level K-NN Classification Method for smart Harvesting

In Fig.5.2 explains the comparative chart on K-NN classification method and existing methods. The Advanced Level K-NN gives 95.12 % of accuracy and 92.67 value in F-measure. While comparing the existing methods the proposed method produce the high accuracy results.

VI. Conclusion

The purpose of this study is to create a smart harvesting system that utilizes a technology application to determine the soil level, crops that are appropriate for harvesting, and to forecast the disease and harvesting operation. App marks a significant advancement in the agriculture industry. The app accomplishes its goals of improving crop management, boosting disease detection, enhancing fertilizer recommendations, encouraging sustainability, and providing farmers with technology-driven tools by utilizing cutting-edge technologies like machine learning, deep learning, and data analytics. By offering precise crop type predictions based on environmental variables, early plant disease diagnosis, and optimal fertilizer recommendations, the app greatly boosts agricultural productivity, ensures food security, and preserves natural resources.

Additionally, the app is essential in providing farmers with the information, perspectives, and tools they need to make wise decisions, increase productivity, and improve their standard of living. As a whole, the Smart Harvest App supports food security, sustainable agriculture, and the welfare of farming communities, all of which are in line with SDG 2's "Zero Hunger" objective. It establishes the groundwork for a more robust, effective, and sustainable agricultural ecosystem and is evidence of the ability of technology to bring about positive change in the agriculture sector.

REFERENCES

[1] Abdelgwad MM, Soliman THA, Taloba AI, Farghaly MF (2021) Arabic aspect based sentiment analysis using bidirectional GRU based models. J King Saud Univ–computInf Sci. https://doi.org/10.1016/j.jksuci.2021.08.030

[2] Blackley SV, Huynh J, Wang L, Korach Z, Zhou L (2019) Speech recognition for clinical documentation from 1990 to 2018: a systematic review. J Am Med Inform Assoc 26(4):324–338. https://doi.org/10.1093/jamia/ocy179

[3] Dr.S.Brindha, Dr.S.Sukumarn, Relevance Pattern Discovery for Text Classification Using Taxonomy Methods" in International Journal for Science and Advance Research in Technology (IJSART)" Volume 4 Issue 11 –November 2018 ISSN [online]:2395-1052.

[4] Dr.S.Brindha, Dr.S.Sukumaran, An Analysis on Big Data Interrogation Explore Problems and Tools, InternationalJournal of Novel Research and Development (IJNRD), ISSN:2456-4184, Volume 5, Issue 6,June 2020.

[5] Cha, G.-W.; Choi, S.-H.; Hong, W.-H.; Park, C.-W. Developing a Prediction Model of Demolition-Waste Generation-Rate via Principal Component Analysis. Int. J. Environ. Res. Public Health 2023, 20, 3159. [CrossRef]

[6] Corso, M.P.; Perez, F.L.; Stefenon, S.F.; Yow, K.-C.; GarcíaOvejero, R.; Leithardt, V.R.Q. Classification of Contaminated Insulators Using k-Nearest Neighbors Based on Computer Vision. Computers 2021, 10, 112. [CrossRef]

[7] Gnanasanjevi G, Rajagopalan S, Arutkumaran M, Information Use Patterns of Farmers in Rice-based Cropping Systems in Puducherry, Volume 18, Issue 4, Page 60-66, 2025; Article no.ARJA.128835ISSN: 2456-561X.

[8] Fan, G.-F.; Guo, Y.-H.; Zheng, J.-M.; Hong, W.-C. Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting. Energies 2019, 12, 916.

[9] Ismail A, Abdlerazek S, El-Henawy IM (2020) Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping. Sustain (switz). https://doi.org/10.3390/su12062403

[10] Lee, C.-Y.; Huang, K.-Y.; Shen, Y.-X.; Lee, Y.-C. Improved Weighted k-Nearest Neighbor Based on PSO for Wind Power System State Recognition. Energies 2020, 13, 5520.

[11] Lu, J.; Qian, W.; Li, S.; Cui, R. Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery. Appl. Sci. 2021, 11, 919. [CrossRef]

[12] SakshiVasantGondane, S. N. Ingle, P, D. Peshattiwar, Jitendra B. Durge, "Comparative Technical Efficiency and Constraints in Seed and Grain Production of Chickpea Variety JAKI-9218 in Amravati Districtof Maharashtra, India"Volume 18, Issue 4, Page 54-59, 2025; Article no.ARJA.146388ISSN: 2456-561X, Asian Research Journal of Agriculture (ISSN: 2456-561X)

[13] Nguyen, L.V.; Vo, Q.-T.; Nguyen, T.-H. Adaptive KNN-Based Extended Collaborative Filtering Recommendation Services. Big Data Cogn. Comput. 2023, 7, 106. [CrossRef]

[14] Wang, J.; Zhou, Z.; Li, Z.; Du, S. A Novel Fault Detection Scheme Based on Mutual k-Nearest Neighbor Method: Application on the Industrial Processes with Outliers. Processes 2022, 10, 497. [CrossRef]

[15] Zhang, L.; Zhu, Y.; Su, J.; Lu, W.; Li, J.; Yao, Y. A Hybrid Prediction Model Based on KNN-LSTM for Vessel Trajectory. Mathematics 2022, 10, 4493. [CrossRef]