Grenze International Journal of Engineering and Technology, June Issue

Smart Agriculture - Crop Suggestion System

Divya H N[1], Ravikumar S[2], Rajasab B Maidur[3], Mohammed Shakeeb[4] and Sumanth R[5]
[1]Assistant Professor, Department of Computer Science, JSS Science and Technology University, Mysore, India
[2-5]Department of Computer Science, JSS Science and Technology University, Mysore, India
Emails: divyahn@jssstuniv.in, rr4104209@gmail.com, rajasabmaidur@gmail.com,
mohammedshakeeb966@gmail.com, sumanthr121@gmail.com

***Abstract*— Agriculture is the foundation of India's economy, and it plays a vital role in the growth of the Indian economy field. The technologies are emerging as improved, but in the agricultural field, there are still issues which are not been solved yet. In agriculture, crop selection plays a vital role in the yield, Sometimes the crop selection follows and mainly depends on a farmer's experience, commonly failing to consider critical soil factors that determine the crop's suitability, based on its functions or the soil properties. This project was an innovative agriculture project. Under this, the first module was a brilliant suggestion system based on soil and weather conditions.**

This project demonstrates a crop suggestion system which is based on the soil conditions, such as pH, temperature, Nitrogen(N), Phosphorus(P), Potassium(K) values and weather conditions such as temperature and rainfall. Our project helps the farmer's to select the most suitable crop scientifically. For our project, we consider parameters such as N, P,K, pH, temperature, and rainfall to recommend the most accurate crop. For this project, we have collected the data from the Organic Farming Research Centre, Nagenahalli, Karnataka, combined with published agriculture data sets to ensure reason-specific accuracy Again, with this data, we have performed the supervised learning models, such as random forest classifier,decision tree, support vector system and gradient boosting were tested and Comparing the accuracy, F1-score, confusion Matrix, and evaluation metrics , we come up with the most accurate model, a random forest classifier, which achieved the highest precision and stability in recommending suitable crops for the land.

After model selection we deploy the model through a flash based interface, where the user can give input to the system thereby get output that is the recommended crop will be displayed with the explanation and the PDF option. It also enabled to generate print of that crop for explanation we used graphical visualisations, feature distributions, correlation, heat maps, and SHapley Additive exPlanations (SHAP) based explainability plots to enhance the interrupability. This model demonstrated high reliability across different soil conditions and environmental factors, and it was tested against ambiguous cases, which are near decision boundaries. Future enhancement includes manual reading or automatic reading of data on sensor data integration devices or mobile application development. We have a project on smart agriculture. In this model, we have a crop suggestion system, and another model that is a prediction for future enhancement."Agriculture is not just about crops; it is about the roots of a nation."

***Index Terms*— Crop recommendation, Machine learning, Random Forest, Soil and weather analysis, SHapley Additive exPlanations explainability.**

Grenze ID: 01.GIJET.12.2.156_21

I. INTRODUCTION

Agriculture is the foundation of India's economic growth, playing an indispensable role in creating new jobs and employment, ensuring food security, and contributing to national development. "The future of agriculture lies in the intelligent use of data". In a large portion of India's agriculture, various advanced technologies across multiple sectors still depend on conventional crop selection practices based on farmers' experience and past yields of the crop. These typical methods often overlook fundamental aspects, such as soil condition, particularly soil fertility, nutrient composition, and dynamic weather conditions, resulting in inefficient crop yields and ineffective resource utilisation [3][4]. To overcome these limitations, there is a growing need for intelligent, data-driven, analytics-based tools capable of inspecting soil and environmental attributes to support informed decision-making data driven in agriculture. This work proposes a machine learning-based crop suggestion system, which falls under the category of smart agriculture. This system recommends suitable crops by assessing key agricultural parameters, including nitrogen, phosphorus, potassium, pH, temperature, and rainfall. Machine learning models are capable of identifying hidden sequences or trends in large and complex data set structures that traditional approaches often fail to recognise by using algorithms such as random forest, decision tree, KNN, support vector machine, and gradient boosting. Search data driven insights help improve the crop selection and prediction promote optimised resource utilisation as well as management there by reducing the wastage of water and fertilizers while supporting eco-friendly agricultural practices[8][9] .The system is designed with the simple web based interface to ensure the availability for formal with minimal technical expertise and experience, ultimately aiming and objective of the project is to enhance productivity and fortify the technological framework of India's agriculture field.

II. RELATED WORK

Machine learning has significantly advanced agricultural decision-making by enabling optimised product crop prediction systems that assist farmers in selecting crops using evidence-based insights. Several studies have explored and demonstrated the role of soil nutrients, temperature and rainfall in crop Appropriateness modelling, demonstrating that anticipatory analytics can improve the yield of outputs. Although many existing systems depend on fixed data sets that do not adapt well to evolving climate conditions.[3] [4] [8]. Multiple research works have evaluated machine learning algorithms, such as decision trees, K-nearest neighbours (KNN), and support vector machines (SVM), for automating crop selection. These models have shown acceptable baseline prediction accuracy; however, their limited capacity to handle fluctuating and irregular environmental variations reduces their impact in real-world agricultural scenarios. Assemble learning approaches, including random forest and gradient boosting, have gained recognition due to their Robustness,for higher forecasting performance and the ability to generalise over diverse agricultural data sets. The features mentioned above are considered prominently to analyse this project and achieve optimised outcomes.

Some recent development systems incorporate explainable AI methods to make model decisions more understandable for end users. Several attempts have integrated soil nutrient analysis with rainfall forecasting. However, Challenges remain particularly the lack of area-focused data sets, inadequate real-time sensor integration, and minimal focus on intuitive interfaces,which are crucial for practical adoption by farmers. [3] [4] [9]. Despite technological advancements, many existing models remain restricted to research prototypes without development and deployment at scale. The integration of real time environmental data streams, Targeted soil information, Explainable outputs can significantly enhance the usability and credibility of these systems this presents challenges employing random forest based predictions incorporating local agricultural knowledge , analysed data using several machine learning methods and providing web based interface which is easy to use by the farmers which does not require more expertise and also enables immediate and explainable crop recommendations for improved smart decision-making results in more suitable crop was suggested by the software.

III. SYSTEM DESIGN

System design is the outline of a project that defines how the entire system will work including physical components, software programs ,data transfer, structure modules and communication between the components. In this proposed system, it is developed as low overhead web based smart decision support Advisory Framework that recommends most ideal crop types based on instantaneous agricultural parameters. It requires general purpose computing environment equipped with the Parallel processing CPU, at least 4GB RAM and a stable network connection to handle instantaneous inference efficiently. The software stack employs Python 3.8+, Flask as backend minimal framework for routing and model execution and a modern web browser for the user interface. A random forest classifier is utilized Due to its Reliability ,Consistency and Superior performance in modelling irregular agricultural data sets when compared to traditional single classifiers and individed Learning algorithms.[1][6][7]

The trained machine learning model, Along with preprocessing encoders and SHAP based interpretability components is serialized into pickle files to enable fast and reusable inference without reinitializing the full pipeline. These pickle files helps reduce the execution of the algorithms every time whenever the user is input to the web page. This significantly reduces computational system burden and improves the latency during deployment. The workflow begins when user input agricultural parameters through the websites interface. These inputs transmitted to the back end using RESTfull POST request for further processing. The system cleans, scales and arranges the input vector before performing prediction process using the trained random forest model, followed by interpretability model explanation using SHAP, it quantify the contribution of each attributes towards the final recommendation.[2][7]

To ensure area specific applicability the generated crop prediction is cross checked with stored agricultural metadata derived from the local soil data sets and expert knowledge bases [3] [4]. The server then construct's a structured JSON response then display organized response payload which contains the predicted crop. Explainable values and quality metrics. At the front end this response is rendered with Visual output and optional Download support using JSPDF and HTML canvas. For production deployment the system is hosted using gunicorn to ensure expandability, Minimal delay and Multi user handling by integrating Machine learning interface explainable AI and web engineering principles, the system functions as a reliable and effective smart decision-support tool for data driven forming in agriculture field.

The Fig.1 explain the complete workflow of the crop suggestion system showing the interaction among the user front-end, back-end server and the machine learning model. The user provides soil-weather data on the front-end,which delivers this input to the server through a POST request. The back-end post this input to the machine learning model which executes or analyze's the data and returns forecasted and estimated crop. Then back-end generates result page and transmits output to the web interface where it is rendered and shows the result to the intended user.

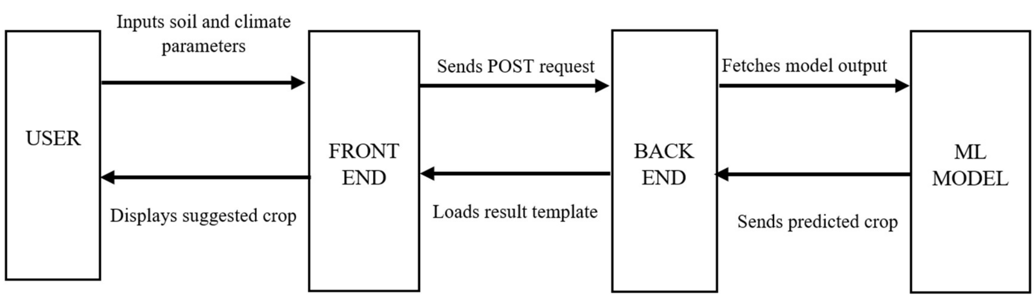

Figure 1. Dataflow and System Design of the Crop Recommendation System

IV. METHODOLOGY

A. Data Collection

Data collection serves in developing the crop suggestion system as the caliber applicability and precision of the collected data directly influence the system performance in generating reason specific prediction. The primary data-set was acquired from the organic farming research station Naganahalli,Karnataka which provided reliable field level information on soil fertility nutrient makeup and eco-friendly organic-farming practices specific to the local agriculture climate condition [3]. This data set offer value of an inside into soil nutrient dynamic and green cultivation, a brochure that reflect real world agriculture-settings.

Additionally, a secondary dataset was sourced from the publication Krushi Belegala Sudharitha Besaya Paddhatigalu by Krushi Mahithi Ghataka, Vistharana Nirdheshanalaya, Krushi Vishwavidyanilaya, Bengaluru–560 065 [4]. This reference provided extensive agronomic standards soil category and crops specific operational method and handling technique essential for the strength of the recommendation model. By merging evidence base field data with the documented agricultural-knowledge the system ensures higher accuracy flexibility and sustainability in crop prediction . Combined approach enables the model to aligned effectively with real world condition resulting in more reliable and regionally optimised crop suggestion.

B. Data Preprocessing

Data Pre-processing is a preparation of collected data which is suitable for model learning. Data pre- processing is a fundamental stage in developing the crop suggestion system, ensuring that the agriculture data set is clean, consistent and appropriately organised for machine learning analysis. The data obtained from Organic Farming Research Centre Nagenahalli and other verified agriculture publication exhibited variation that could impact model performance. To address these issues, multiple preprocessing techniques were systematically applied. missing values commonly arising from the manual data entry errors or sensor malfunctions were treated using mean estimation it is defined mathematically as:

$$X = \{ x, \quad \text{if } x \text{ is valid}$$

$$\bar{x}=\sum x/n, \quad \text{if } x \text{ is missing}\} \tag{1}$$

Where x represent the imported value and x̄ denotes the mean of available sample this approach keeps data set integrity while reducing information loss. Category variable or flashlights such as crop names were converted into encoded values using label encoding where each category is assigned to an integer integer-label. Since the numerical attributes nitrogen, phosphorus, potassium , PH, temperature and rainfall exist on different scales minmax normalisation was applied to normalise values into uniform range between 0 and 1. This helps to transfer the details from the data to the model correctly. This modification prevent attributes with larger magnitude from dominating model training improves stabilization and balances the learning process.

Finally, feature selection or variable filtering was performed to identify the most cultivated critical parameters these features directly influence soil fertility and crop yield and potential stability. Forming the optimised input vector X. Overall, the preprocessing pipeline enhanced data uniformity, also reduced duplication and optimised the data set for dependable machine learning full stop as a result, the system achieved improved reliability, predictive and accuracy along with reasons specific adaptability and redundancy formation for crop recommendation.

C. Data Splitting

Data splitting is a fundamental step in the development and evaluation of the Crop Suggestion System, ensuring that the trained machine learning model applies broadly, well to unseen data rather than memorizing existing sequences. The complete data source was partitioned and separated into two unique and non-intersecting-subsets. We split the datset into 80:20 ratio where 80% data will be used as training dataset and 20% dataset will be used as testing dataset.This split ratio is widely recognised for achieving training acccuracy and testing dependibility. The training subset enabled the model to learn the associations among soil fertility parameters and environmental factors there by estimating a predictive function. This data splitting is crucial step where different types of same datasets are used for evaluation of the model. The testing subset was then used to evaluate the model's adaptation ability, avoid overfitting and ensure uniform and accurate forecasts for unseen field new situations which broadly forecasted in model and testing the performance.

D. Model Selection

Model selection is a vital component in designing and accurate precise and optimised crop suggestion system, as it determines which algorithm best suits the attributes and organization of the collected data. The principal target is to identify a machine learning model capable of delivering high inference correctness specificity, recall and stability for multi class crop classification. To accomplish this, several supervised algorithms were evaluated using the processed dataset, each model was trained and validated under identical testing setup. Table I. shows comparison of machine learning algorithms based on evaluation metrics.Among all the models, Random Forest Classifier demonstrated higher efficiency. It's ability to manage complex patterns and resist overfitting made it the most appropriate selection. In a Fig.2 explains Psuedocode of Random Forest Classification. Thus, Random Forest was choosen as the optimal model for the deployment due to its stability, robustness, and consistent performance.

E. Model Training

Finally, Random Forest Classifier was chosen because it's a multi-model system, the combined model approach enhances both predictive accuracy and generalisation ability on unseen data. The algorithm constructs multiple decision trees and merges the predictions. Each decision tree is trained using a resampled dataset, and at every mode, a randomly selected subset of features is used in the algorithm for accurate performance. This method reduces dependency among trees and minimises overfitting. Random forest computes attribute contribution scores, showing the most dominant factors. This aligns with farming relevance seen in previous studies. Its high inference performance, alignment with multiple day types and interpretability and transparency through feature importance make it suitable for agricultural applications and function-based field work in farming.

F. Deployment

The deployment integrates the trained model with a responsive UI that enables real-time predictions. After training, the model was refined and serialised for quick model loading.It uses pickel files where model training details and result was loaded and it will be fetched by Flask-Backend. The backend function acts as a low-overhead backend where interaction between the user and the prediction model occurs, enabling minimal delay, When a user submits the form, the backend check all parameters are valid,if they are valid then it verifies inputs and computes the result.The result will be displayed on frontend screen. SHAP improves interpretability and makes model decisions more transparent, which is crucial for a decision support system. This rollout setup creates an applicable approach with an explainable output for farmers and agricultural practitioners.

```
Algorithm: RandomForest_Classification

Input: Dataset D, number of trees n

Output: Final class prediction ŷ

1. Initialize an empty list Forest

2. For i = 1 to n:

a. Create a bootstrap sample Di from dataset D

b. Train a decision tree Ti on Di

- At each split, randomly select subset of
features

c. Add Ti to Forest

3. To classify a new instance x:

a. For each tree Ti in Forest:

Predict class ci = Ti(x)

b. Compute majority vote among all ci

4. Return:

ŷ = mode{T1(x), T2(x), …, Tn(x)}
```

Figure 2. Pseudocode of Random Forest Algorithm

TABLE I. COMPARISON OF MACHINE LEARNING ALGORITHMS BASED ON EVALUATION METRICS

Model	Accuracy	Precision	Recall	F1-Score
RandomForestClassifier	0.8404	0.8340	0.8404	0.8333
GaussianNB	0.8310	0.8353	0.8310	0.8278
DecisionTreeClassifier	0.8310	0.8463	0.8310	0.8311
BaggingClassifier	0.8263	0.8381	0.8263	0.8234
GradientBoostingClassifier	0.8028	0.7958	0.8028	0.7945
ExtraTreeClassifier	0.7371	0.7641	0.7371	0.7318
Logistic Regression	0.7183	0.7666	0.7183	0.6996
SVC	0.6808	0.6864	0.6808	0.6524
KNeighborsClassifier	0.6761	0.7308	0.6761	0.6833

V. RESULTS

The test-based assessment was performed on the crosfinal.csv dataset containing soil, climatic and ecological surroundings features. Several-supervised learning algorithms were tested, and their efficiency was evaluated using Accuracy, precision, recall and F1 score, among all models, the Random Forest Classifier achieved the highest accuracy, i.e 0.8404 and F1 score 0.8333, demonstrating superior generalisation and robustness compared to other classifiers. The relative performance of all models is presented in Table I. The performance of the crop suggestion system was evaluated using sequence of typical evaluation testing samples and unclear or ambiguous test cases listed in Table II.

TABLE II. TEST CASES TABLE

N	P	K	pH	temperature	rainfall	label
95	42	48	5.8	22.5	1175	Rice
112	59	38	6.65	21.8	725	Wheat
128	65	43	6.45	24.2	690	Maize
97	53	57	6.35	27.8	1020	Cotton
238	112	122	7.35	31.5	1380	Sugarcane
42	56	38	6.82	25.7	680	Soybean

These testcases used to measure the accuracy, trustworthiness and generalization competence of the Random Forest Classifier across heterogeneous soil and climatic circumstances. In typical cases the model demonstrate High forecast performance. In test case 1, the inputs resulted in correct prediction of Rice which perfectly matched with training data as well as testing (unseen) data. It indicates the model's understanding of high rainfall and slightly acidic soil conditions favourable for paddy cultivation. In test case 2, the inputs led to an accurate wheat prediction, which indicates a crop with medium nitrogen content, neutral soil pH and a mild weather environment suitable for wheat crop growth. The model also evaluated using ambiguous or borderline test cases,where multiple crops matched the given conditions. In Ambiguous Case 1 (Cotton or Rice) the model predicted cotton representing a typical edge scenario. In Ambiguous Case 2 (Maize or Wheat), the model predicted wheat, demonstrating its ability to differentiate slight environmental variations. Even in ambiguous scenario,the predictions aligned with recognised agricultural methods demonstrate the model's capability capture detailed interactions in the data.

VI. CONCLUSION

In this study a smart crop recommendation system was realized in order to assist farmers in decision-making by taking into account local soil conditions and climatic conditions. The plantings are determined in terms of the N, P, K, pH, temperature, and rainfall patterns by some of the most prevalent parameters and recommend the crops at particular sites. The random forest classifier was obtained with the best performance among different machine learning methods. The model emerged good and workable as well, with a good performance not only on training data but also on new, unseen cases. This reproducibility is important in real-world practice as there is significant diversity in the conditions that farmers encounter. We are developing a user-friendly web application that the farmers can access easily. This interface will alert with instant crop suggestions (and provide visual explanations for the recommended crops). This transparency will work wonders for farmers' trust in what a system recommends. Our approach is one that we feel uses real field data from the Organic Farming Research Centre in Naganahalli, Karnataka and is complemented by an extensive base of agricultural research. We place our system into realistic application, so it can fit for any soils and any farming in the area, and therefore is of significant use rather than theoretical. The latter has enormous potential for growth. For instance, we could potentially introduce IoT sensors that track soil in real-time, and create a mobile app that gets this technology directly between farmers and their machines. We will have to integrate a disease prediction functionality to build the framework as a modern decision support instrument in Indian agriculture. We expect that these practical, farmer-friendly technologies will play an important role in making agricultural practices more sustainable and productive.

REFERENCES

[1] L. Breiman, "Random Forests," Machine Learning vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[2] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[3] Department of Agriculture, Government of Karnataka, "Soil Fertility and Organic Cultivation Data Records," Organic Farming Research Station, Naganahalli, Mysore, Karnataka, India, 2023.

[4] Krushi Mahithi Ghataka, "Krushi Belegala Sudharitha Besaya Paddhatigalu," Vistharana Nirdheshanalaya, Krushi Vishwavidyanilaya, Bengaluru, 2022.

[5] S. Rasool, R. Ramesh, and K. Kumar, "Data preprocessing techniques for efficient machine learning in agricultural systems," International Journal of Advanced Computer Science and Applications, vol. 13, no. 7, pp. 205–212, 2022. [Online]. Available: https://thesai.org/Publications/ViewPaper?Volume=13&Issue=7&Code=IJACSA&SerialNo=25

[6] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 4th ed. San Francisco, CA: Morgan Kaufmann, 2022.
[7] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA: O'Reilly Media, 2023. [Online]. Available: https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/
[8] P. Singh and A. Verma, "Comparative study of machine learning algorithms for crop yield prediction," Computers and Electronics in Agriculture, vol. 204, pp. 107–128, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923000099
[9] S. Rasool, R. Ramesh, and K. Kumar, "Performance analysis of ensemble models for agricultural data classification," International Journal of Advanced Computer Science and Applications, vol. 13, no. 9, pp. 189–198, 2022. [Online]. Available: https://thesai.org/Publications/ViewPaper?Volume=13&Issue=9&Code=IJACSA&SerialNo=22