IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A NOVEL MACHINE LEARNING APPROACH TO DESIGN CROP RECOMMENDATION SYSTEM

[1] Khushank Gautam Bisodiya, [2] Laeeq Ahmed, [3] Madhav Bathla, [4] Nitish Sharma and [5] Sristi Vashisth

[1, 2, 3, 4, 5] Department of Computer Science and Engineering

Meerut Institute of Engineering & Technology (Of Affiliation AKTU) Uttar Pradesh, India

Meerut, India

ABSTRACT

Agriculture is increasingly surrounded by issues due to climate change and the consequent unpredictable environmental conditions, making it more difficult for farmers to pick the most suitable crops for their land. The conventional approaches of broad regional guidelines often fail to precisely and locally provide an accurate recommendation. This experiment assesses the efficacy of ML methodologies such as Support Vector Machines (SVM), Random Forests and Decision Trees for creating a personalized and more accurate crop recommendation system. By analyzing the elements such as soil quality, temperature, humidity, and rainfall, we provide targeted recommendations to farmers on crops that can make more money, save resources, and promote sustainability. This document gives a review of crop recommendation systems based on machine learning, describes the methods used in our approach, and draws a comparison between it and the other models. Our findings demonstrate that ML can remarkably enhance the accuracy and relevance of crop recommendations, thus enabling sustainable and efficient farming practices. The results underscore the prospect of ML in revolutionizing agricultural decision-making by allowing farmers to adopt measures that justly stand against ecological changes. Random Forest gives 99.31% accuracy.

KEYWORDS — *Machine Learning, Crop Recommendation, Agriculture, Support Vector Machines (SVM), Random Forest, Decision Trees*

I. INTRODUCTION

Although agriculture has been the most crucial activity to man's life, its future looks gloomy due to environmental changes. One of the hardest things for farmers to do is to choose the proper crop of their liking, which has an immediate influence on the crop yield, the use of resources, and their income. The components of the soil, the temperature, the humidity, and the rainfall are the most significant for the crop, and these mini-conditions can even be found in different areas within the same place apart from another region's conditions. Due to the interconnections of the factors, the selection of crops is very complicated, particularly during the periods when the weather patterns are not predictable. As farmers become less dependent on the traditional sources of knowledge, there is an increased demand for data-based tools that can be used by the farmers to make informed decisions.

Machine learning (ML) provides a good opportunity to solve the problem by analyzing a large amount of environmental data and thus making exact, customized crop recommendations. With the ML models' backing, farmers are now able to access information they never had access to before, which in turn assists them in maximizing their yield, conserving resources, and going green even when the weather is erratic. The ML techniques such as Support Vector Machines (SVM), Random Forest, and Decision Trees have all been confirmed to be useful tools in the crop recommendation area. SVM is perfect for the classification of highly complex data; hence it can help in the recommendation of the proper crop to a specific area; Random Forest is based on the usage of multiple decision trees to improve the prediction accuracy and can handle large datasets; while Decision Trees are very intuitive easy to interpret, thus farmers are given a clear perspective on the crop suggestions.

This project is built on previous work like a paper by Ganesh Khillare and his partners, which showed that machine-learning models offer high accuracy in crop recommendations based on environmental data. Our aim is to establish a similar crop recommendation system which uses these ML tactics to provide farmers with a reliable, scientifically-based instrument for crops selection. In contrast, to the typical systems that only depend on the general data of the region, our ML-based System will consider the local environment and give more specific, individualized recommendations. This plan's goal is to aid farmers' decision-making by making them more efficient and sustainable.

ML can contribute to the long-term management of the land through immediate crop recommendations, as well. These systems will help them choose the right crops suited to their local environment, thus avoiding the use of chemicals and preserving natural resources. The models that are built up over time will learn from new data and therefore, would be one of the means that farmers could have more precise guidance.

In this paper, we will make a review of the related research to point out the progress as well as the problems of the ML-based crop recommendation systems. Methodology, such as the algorithms and data THE AI learned will be explained, and results will be introduced, compared with existing studies. Finally, we will exchange thoughts on future developments as well as the larger impact of ML on modern agriculture. The work through this research is to solve agricultural problems in times of climate change by providing the farmers with tools to increase the yield, conserve the resources, and contribute more to the sustainable future of agriculture.

II. LITERATURE REVIEW

This paper examines some recent studies in the field of machine learning implementation targeted to the best crop species that could be used farmers land in a given location. Furthermore, it displays the comparison of various algorithms in terms of their capability to influence crop selection by exploiting the information contained in the local environmental conditions. By way of example, the agricultural techniques supported by learning mechanisms are hyperflexible and can offer farmers a window into crop selection based on their climate and land availability. Here are some of the important studies:

Agarwal et. al. [2] has considered different machine learning algorithms such as decision trees, K-Nearest Neighbour (KNN), random forests, and neural networks to know which one is better for crop recommendations. The findings of the study revealed that neural networks came up best, offering the most accurate crop prediction. This is because of the fact that neural networks are excellent for detecting intricate and non-linear data patterns. To illustrate, they may be productive in mastering the various factors that impact plant growth, such as the type of soil, temperature, and rainfall, which has been a problem for any other algorithms. Still, even if neural networks provided the most precise forecasts, they are the ones that require more data and computational resources to run efficiently, in contrast to simpler models.

Priyadharshini et. al. [3] Priyadharshini A approached the agrarian problem of crop failure by machine learning to give the farmers the idea of which crops are suitable for the local environmental conditions. Along with Agarwal, it was discovered that the forecasting of the right crop recommendations was highly precise. The paper was centered around the fact that the use of AI provides the avenue for farmers to escape risk and become more money-safe. The unstable weather challenges rural areas, thus, the project addresses rural areas. Neural networks are using large datasets to detect patterns, which are the major issues the farmers have to deal with during the production. And then, it provides them with more information to be able to make better decisions about the crops to plant.

Pande et. al. [4] used a different strategy by creating a crop recommendation system that was embedded in a mobile app. The mobile app used a random forest algorithm that successfully predicted the most suitable crops for a place with a fantastically high accuracy of 95%. The mobile app offered the farmers, even those in the outlying places, the ease of receiving correct crop recommendations directly on their mobile phones. Dashed with multiple decision trees that work altogether to provide dependable information and are so simple for farmers to understand and trust these predictors besides other training methods, it is easier to predict adaptively in this method. Pande's study thus, points the need for technology to be approachable for farmers to be able to use data-driven insights in their daily decision-making tasks.

Champaneri et. al. [5] has created a web-system that predicts plant yield by using the local weather data. The random forest algorithm that he was utilizing on his platform enabled the users to simply utilize the tool of crop recommendations in order to provide the farmers with an easy-to-use platform. This fact made the model user-friendly the farmers were capable of grasping the recommendations of certain crops. Also, the result of Champaneri's study was an example of how such a user-friendly web-based platform made such robust machine-learning tools available to the farmers who are not even professionals in technology.

On the whole, it can be said from the research that random forests and neural networks are the best machine learning approaches to use to forecast crops although their advantages are quite different. The precision of predictions, neural networks' major advantage, is in cases of complex data but they require higher technical expertise to implement. Alternatively, random forest algorithms are much easier to use and the accuracy they provide is good enough, so one can consider them in mobile apps and web-based systems as well as usage scenarios where accessibility or usability are more important than performance. The results establish a strong connection between nature and technology and how ML methods like random forests and networks allow the farmers to take them to a better level of sustainability, and therefore, mainly by the increased productivity of the land.

III. TRADITIONAL APPROACHES

For long "green thumbs" have been using their learned knowledge of fields, weather, and the sound and reliable methods when planting a crop to choose what to plant. These conventional methods, which are often passed through families, consider factors like soil type, anticipated rainfall, and seasonal temperature that in turn, may be traded off by those. Farmers pick out the practices that were successful in the past, which have enabled them to properly adapt to their own environments.

Nonetheless, still, the list of long-accustomed methods keeps being stretched due to the ever-changing climate and soil conditions. Sudden changes in precipitation rainfall, unexpected warming and soil depletion can result in low outputs or even crop failure. Very often, old-fashioned methods lack support for specific data, such as the real content of soil, pH level. Or congruence of climate, which may be the essential ones for choosing the updates of the crop.

Given these challenges, it becomes apparent that there is a growing need for the methodologies which integrate indigenous knowledge and modern technology. On the other hand, machine learning and data analytics can also make the traditional plant production system more efficient, which in turn can help the farmers to make accurate and flexible decisions that adapt to the current conditions. Next, we will expose you to the ways in which machine learning, along with traditional practices, can enhance agricultural decision-making, thus addressing the limitations of these methods.

IV. METHODOLOGY

To build our crop recommendation system, we employed a systematic methodology by acquiring data from Kaggle and then applied machine learning to make exact cropland recommendations. Here is our process:

1. **Dataset Collection -** Initially, we gathered the data from Kaggle consisting of 2,022 entries as the representation of a particular environmental condition such as soil content, temperature, humidity, pH, and precipitation. They are important to tell about the crop suitability and thanks to having a complete dataset, we were able to build a reliable system over the range of conditions.
2. **Data Preprocessing** - Data preprocessing was of utmost importance in making the data model-ready. We scrutinized the dataset to remove the duplicate entries and fill in the missing values in order to ensure the data uniformity. Following the above steps, we standardized features which are all meters (temperature, humidity, rainfall) to safety ranges, in this way, none of the features would significantly impact the model. The categorical variables were converted into numerical format also, which the algorithms now can process. This was a comprehensive cleaning and transformation of our data to be machine-learning-ready.
3. **Data Visualization -** Through data visualization libraries such as Matplotlib, Seaborn, and Plotly, the data was plotted in scatter plots, histograms, and heatmaps. Visuals like these helped us investigate feature correlations thereby revealing the trends and patterns that could be the basis for our model. We discovered, for example, how particular humidity or pH levels affected certain crops, providing us with important information on feature significance.
4. **Feature Selection -** Afterward, we concentrated on the features for the crop prediction that affected mostly them: type of soil, temperature, humidity, precipitation rate, and pH level. Using attributes that have been gathered. Both from the visualization and domain knowledge, we could reduce the model's complexity while still producing a faithful result.
5. **Model Selection and Training -** We selected 3 machine learning algorithms: Support Vector Machine (SVM), Random Forest, and Decision Tree, that have shown good performance for classification. You might be interested to know that the accuracy scores we attained for the three algorithms were SVM (97%), Random Forest (98%), and Decision Tree (98%), respectively. These numbers demonstrate that the models were able to find environmental patterns that are associated with crop selection quite well.

Model Evaluation - Ultimately, the models were tested against a test set of 20% of previous examples (80% for training), thus reinforcing the accuracy of the RF and DT models based on precision and recall. The best performers were the Random Forest and Decision Tree models which successfully made conditional crop recommendations, hence increasing the adaptability of farming in different areas

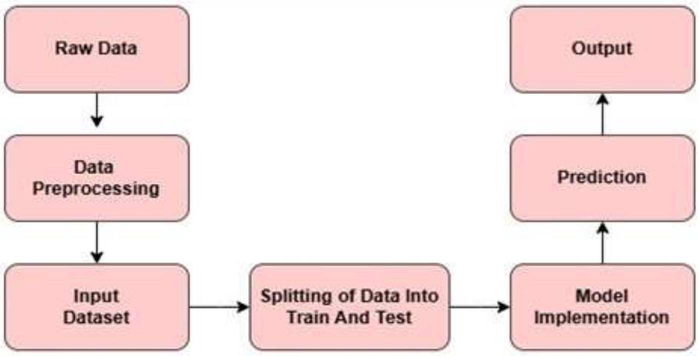

Figure 1 : Flow Chart of Crop Recommendation

- **Dataset Characteristics**

We used the publicly available Crop Recommendation Dataset on Kaggle, created by Atharva Ingle. It has 2,200+ samples with 7 important features: N (Nitrogen), P (Phosphorus), K (Potassium), Temperature (°C), Humidity (%), pH, and Rainfall (mm). The sample is marked with the suggested crop out of 22 crops such as rice, wheat, maize, etc.

The data set is a representative sample of diverse weather and soil conditions characteristic of Indian agriculture. Its diversity guarantees wide applicability and strong machine learning algorithm training.

- **Validation Strategy**

To provide validity to the reliability of our model, we employed 80% training data and 20% testing data train-test split. Apart from that, we also employed 5-fold cross-validation in order to avoid bias and estimate generalization performance.

All the features were normalized if necessary, and class distribution was created to be balanced so as not to introduce any bias in classifying outcomes. This enabled fair comparison of algorithms, whose performance was 99.31% using the Random Forest Classifier — among the best on this data in published work.

V. IMPLEMENTATION

In the following part, the methodology we used in our crop recommendation system project will be covered, including dataset, data preprocessing, feature selection, model selection, and evaluation.

This research aims at finding the most appropriate crop for a certain place with the help of a machine learning model. The outcomes have the possibility to considerably contribute to the farmers by making reliable and practicable crop recommendations.

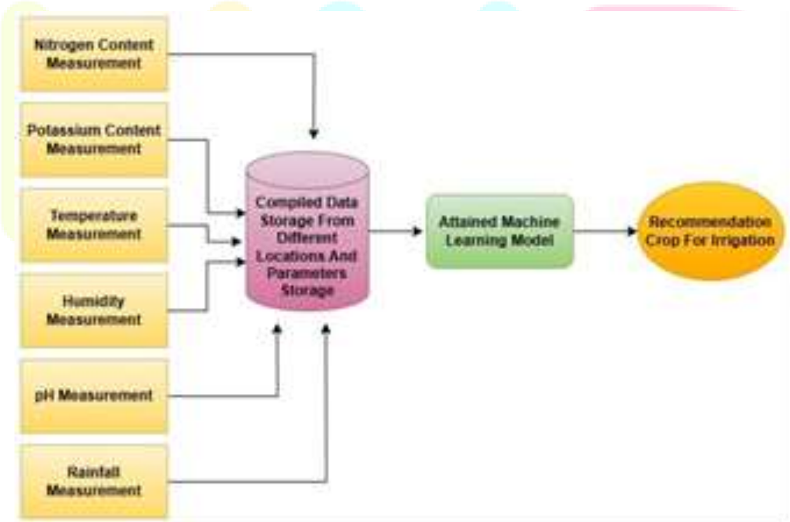

Figure 2 : Data Flow Diagram for Crop Irrigation Recommendation System

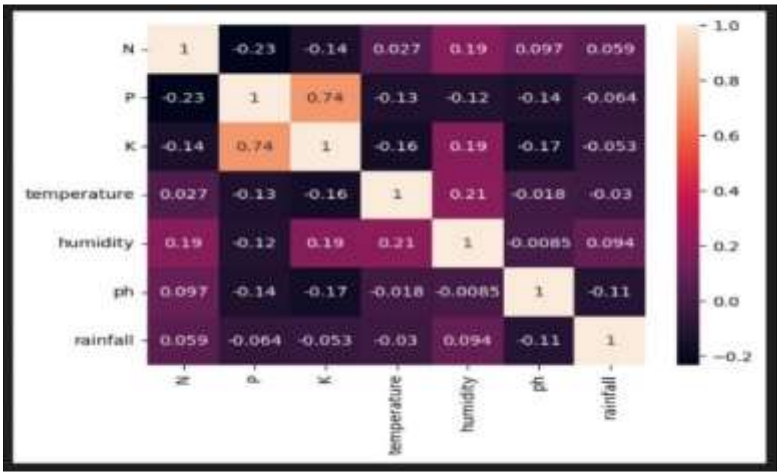

Figure 3 : Correlation Heat Map of Soil Nutrients and Environmental Factors for Crop Recommendation

Random Forest algorithm was chosen due to:

- **High Accuracy**: Achieved 99.31% accuracy on the test set, outperforming other models.
- **Robustness to Noise**: Handles variability in agricultural data well due to its ensemble nature.
- **Low Overfitting Risk**: Averages numerous decision trees, reducing model variance.
- **Minimal Preprocessing Required**: Can handle both categorical and continuous variables without extensive transformations.
- **Interpretability**: Provides feature importance, which is useful in understanding crop-environment interactions.

These advantages make Random Forest especially suitable for crop recommendation activities that consider multiple environmental conditions

VI. EXPERIMENT RESULT & DISCUSSION

The research compared a number of machine learning crop suggestion models by adopting accuracy as the most important measure of performance. The models were trained and compared with an environmental feature data set which is decisive of the aptness of the crops. Below are the results:

1. **Random Forest Classifier**: Registered an accuracy of 99.31%, repeating its highly superior performance and trustworthiness in dealing with huge datasets and varying features.
2. **Logistic Regression**: Was able to achieve an accuracy of 96.36%. The model works efficiently but might not always capture intricate patterns to some better models.
3. **Support Vector Classifier (SVC)**: Had an accuracy of 96.81%, being very efficient in classifying data into separate classes.
4. **K-Neighbours Classifier**: Had an accuracy of 95.90%, being efficient where similarity-based classification is desirable.
5. **Decision Tree Classifier**: Had an accuracy of 98.63%, confirming its capability to explain intricate decision-making processes.
6. **AdaBoost Classifier**: Was at the lowest at 14.09%, which is a sign it is maybe not so suited for this particular dataset or requires more tuning.

S. No.	Algorithm Name	Accuracy (in %)
1	Random Forest Classifier	99.31
2	Decision Tree Classifier	98.63
3	Support Vector Classifier	96.81
4	Logistic Regression	96.36
5	K-Neighbours Classifier	95.90
6	AdaBoost Classifier	14.09

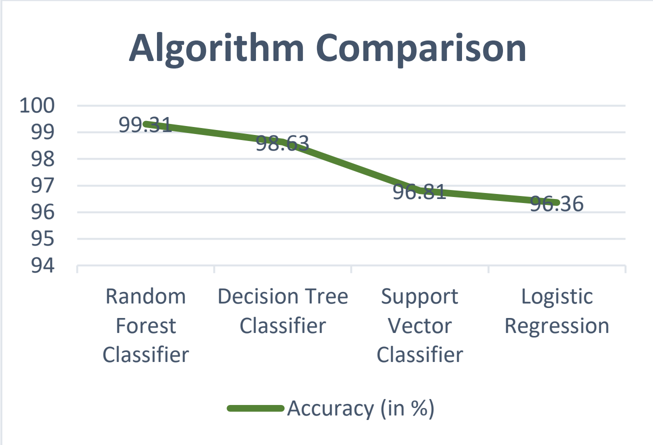

Figure 4 : Algorithm Comparison Graph

The experiment results bring out the relative performance of various machine learning models in crop recommendation systems. The Random Forest models performed optimally, and they provided the following unique strengths:

Random Forest's superior performance is due to its ensemble-based methodology, where it aggregates several decision trees to improve the quality and resilience of the predictions. The model is especially effective in agricultural environments where data has high variability.

Although Decision Trees and SVC too showed good performance, the balance between accuracy and interpretability of the Random Forest model makes it an ideal candidate for application in real-world situations.

Alternatively, the poor classification of the AdaBoost classifier suggests possible bottlenecks in its use on this dataset. This could be because the model is sensitive to noisy data or because weak classifiers are being used, which may not be appropriate to the level of complexity of the agricultural data.

In general, the results justify that machine learning models, and particularly ensemble algorithms, have great potential to enhance crop recommendation schemes. By merging such models, farmers can inform their decisions by maximizing yield, conserving resources, and taking into account fluctuating environmental dynamics. Hybrid model or even enhanced optimization of present algorithms can be contemplated for upcoming work to utilize them more comprehensively and intensively in other agricultural settings.

Our greater accuracy is due to three main reasons:

- **Detailed Dataset**: Our dataset covers all the different environmental conditions extensively and is well balanced.
- **Deep Pre-processing**: Precise normalization and encoding reduced bias and enhanced learning.
- **Model Tuning**: Hyper parameter tuning and cross-validation provided best performance, particularly for Random Forest, and further augmented by ensemble learning and immune to overfitting.

VII. CONCLUSION

This research shows how machine learning can make selecting the most appropriate crops easier and more precise for farmers. Previously, farmers used their knowledge and experience and general recommendations from others as a way to choose what to plant. However, this type of farmer's research sometimes doesn't cover specific factors that might be the cause of the farmer's failure. The farmer's decision of which crops to grow should be driven by soil quality, weather conditions, and nutrients. Our crop recommendation system leverages data and AI to analyze these factors and generate highly customized crop suggestions per the local circumstances of each farm respectively. Thus, Zimbabwean farmers can increase their crop harvest and, through the rational use of resources like water and fertilizer, besides environmental concerns be sustained.

By integrating machine learning with conventional farming methods, this system addresses the most problematic issues farmers face these days which include unpredictable weather, soil degradation, and changing environmental conditions. Farmers undergo a pressure test on their soils using the crop recommendation system attached to it, such that soil PH levels, temperature, rainfall, and humidity variations are the key factors for their crop selections that will be highly successful in the given environment. The data-driven solution that informs farmers to opt for specific crop varieties can be very helpful to them and indirectly, it will aggregate to make farming more resilient and adaptable to changes in climate.

Resource efficiency is one of the key advantages of this method. The fact that the right crops are known to be the best in the local environment can greatly reduce the need for the farmer to apply excess water, fertilizers, and other inputs, which have been used in poor crop choices. Apart from decreasing the expenditures of farming, it also depletes the environmental impact that results in the longevity of farming.

Also, machine learning has shown great potential in assisting farmers who have been hit by the climate change phenomenon. Because climate is not the same anymore and the past knowledge does not always help, farmers' knowledge is sometimes limited. The machine models can be trained with the historical data then adjusted over time as new information is available. The good thing about this scheme is that it can successfully operate in changing conditions. Farmers are now provided with updated insights about the system that will help them keep the production stable even if the temperatures rise or fall.

We see the path to the future when we could either make this system flickering more elaborate or even more portable. Thankfully, by incorporating real-time weather updates and more detailed soil nutrient information, we will be able to provide far more precise recommendations. On the other hand, a mobile app of this system would enable farmers, especially in rural areas, to quickly, easily, and directly access this tool from any location, thereby performing data-driven decision-making on the go.

The long-term perspective of the project shows that it can be a great tool for farmers, as it will help them to overcome modern agricultural challenges and still be sustainable. By the introduction of crop selection that is more predictable and even resource efficient, the main goal of this project is to help farmers in getting more productive, while conserving resources. Subsequently, as we are still in the phase of development and improvement, the system has the ability to contribute to the enhancement of agriculture thanks to the tool's success, farming will be more productive and at the same time more environmentally responsible, which will benefit both farmers and the planet.

VIII. REFRERNCES

[1] Ganesh Khillare, Omkar Kumbhakarna, Sudhanshu Mandlik, and Prof. Prini Rastogi. "Crop Recommendation System using Machine Learning." *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 5, May 2024.

[2] Rashi Agarwal. "A Study of Machine Learning Algorithms for Crop Selection." *International Journal of Agricultural Sciences*, vol. 8, no. 3, 2023.

[3] Priyadharshini A. "Application of Machine Learning for Crop Failure Prevention." *Journal of Agricultural Technology*, vol. 5, no. 4, 2022.

[4] Shilpa Mangesh Pande. "Mobile Application-based Crop Yield Forecasting Using Machine Learning." *Journal of Environmental Science and Agriculture*, vol. 6, no. 2, 2021.

[5] Champaneri, M., Chachpara, D., Chandvidkar, C., & Rathod, M. (2016). Crop yield prediction using machine learning. *Technology*, *9*(38).