Rainfall Prediction and Soil Based Crop Recommendation Using Machine Learning

[1]Mr. PATAN VAZID, [2] Mr. P.SAI SATISH, [3] Mr. P.NIKHILESWARA RAO

[4]Mrs. R. SHOBA RANI, [5]Dr.J.JAYA PRAKASH, [6]Mrs. CHINCHU NAIR

[1,2,3] Students, [4,5,6] Professor

vazidcse@gmail.com, saisatishpaluri@gmail.com,puthanikhilrao@gmail.com

Department of Computer Science and Engineering,

Dr. M.G.R. Educational And Research Institute,

Maduravoyal, Chennai-95, Tamil Nadu, India.

Abstract*- India, being an agrarian nation, relies vigorously upon farming yields and ago-modern items for its economy. High return the board is significant for ranchers who generally need to anticipate the normal yield. To accomplish this, different significant qualities, for example, the area of soil alkalinity file and soil pH esteem are examined. The percentage of vital elements such as potassium (K), phosphorus (P), and nitrogen (N) is also taken into consideration. Utilizing outsider applications, for example, climate and temperature APIs, it is feasible to assemble data about precipitation and soil organization in a particular region, as well as soil type and supplement values. By examining the qualities of this information, AI calculations are utilized to prepare the information and make a prescient model. The model intends to furnish ranchers with exact suggestions on proper manure costs in view of the environment and soil boundaries of their territory, in this way expanding ranchers' yields and pay. What's more, the framework gives extra guides, for example, gel, oil or different synthetic compounds expected for better harvest development. This coordinated methodology empowers ranchers to get exact exhortation custom fitted to their particular cultivating conditions and advances proficient and useful cultivating rehearses by incorporating rainfall prediction models, the system enhances agricultural planning, promotes resource efficiency, and supports sustainable and profitable farming practices.

Keywords: Machine Learning, Rainfall, Crop, nitrogen (N), phosphorus (P), and potassium (K), Climate, Temperature.

INTRODUCTION

In the present innovation driven world, powerful data sharing can extraordinarily enable agronomists and help them understand and work on their true capacity. Sharing important and ideal data through formal and casual channels advances receptiveness among ranchers, which decides the extension and viability of data scattering. Utilizing web innovations, for example, HTML and CSS, you can make a web application that combines crop-soil recommendations with rainfall prediction to support farmers in making informed agricultural decisions. The crop-soil recommendation system suggests suitable crops based on the selected soil type and, conversely, identifies the most appropriate soil for a chosen crop. If both crop and soil are provided, the system explains their compatibility, ensuring a better understanding of the selection process using decision tree algorithm ,K-Nearest Neighbor (KNN). The rainfall prediction module analyzes past weather data using polynomial and linear regression techniques. It considers factors such as temperature, humidity, wind speed, and previous rainfall patterns to predict whether it will rain the next day. This helps farmers plan their activities effectively and take necessary precautions. By integrating data-driven crop selection with weather forecasting, our system provides farmers with valuable insights, allowing them to optimize their farming strategies, improve productivity, and enhance resource management.

RELATED WORK

One of the literature review is one of the most crucial phases in the software development process. Prior to growing the device, it is crucial to ascertain the time component, cost savings, and commercial enterprise stability. Once these are met, the next step is to determine which operating system and language can be used to expand the device. Once they start creating a device, programmers need a lot of outside help. You can get this help from veteran programmers, books, or websites. When building the system, the previously listed concerns are taken into consideration. in order to expand the suggested device. Examining and reviewing all of the challenge improvement's needs is the core function of the assignment improvement department.

Literature evaluation is the phase of the software development process that is most important. for any task. Prior to expanding the equipment and associated layout, time considerations, resource requirements, labour, economics, and organizational electricity must be identified and examined. The next phase is to determine the operating system needed for the project, the software program specifications of the particular computer, and any software that needs to be carried on after those factors have been met and thoroughly investigated. A step similar to expanding their tools and related capabilities.

Indonesia experiences unpredictable weather patterns due to its tropical climate. In order to make decisions about future weather information, a study that forecasts rainfall is required. One of the elements influencing local weather variations is rainfall. This study focused on the climate of Semarang, which is located in the Central Java region and has topographical elements including lowlands and mountains that influence rainfall. Rainfall is influenced by temperature, humidity, wind speed, and the length of sunlight. These variables are processed using data from BMKG Semarang. Comparing the research findings with data that had already been gathered was the aim of the data analysis for this study. The study used K-Nearest Neighbor comparison and multiple linear regression with rainfall data as the dependent variable and other features as independent factors techniques [1].

Natural disasters like floods are caused by rainfall. Predicting rainfall is a big issue everywhere. A number of industries, including agriculture, manufacturing, construction, power generation, and tourism, are impacted by rainfall. These days, machine learning is highly helpful in forecasting rainfall, which is essentially necessary for today. Predicting rainfall accurately in the current situation is challenging. Rainfall prediction is an effort to forecast a region's rainfall. This work employs a range of techniques, including classifiers and regression, to forecast rainfall, including the decision tree algorithm, support vector regression (SVR), linear regression, random forest regression, and random forest classifier. In agriculture, effective rainfall has a big impact on how quickly crops grow. Machine learning will be used to determine the amount of predictable rainfall that will be used for crop productivity, water resource management, and water usage planning. Additionally, a comparison of the outcomes of various machine learning algorithms is part of the research. The suggested machine learning (ML)-based approach will assist in forecasting the annual quantity of rainfall. With an accuracy score of 89%, the random forest regression clearly performs better than other algorithms when a different approach is applied [2].

Predicting rainfall is crucial in a number of sectors like disaster relief, agriculture, and water resource management preparation. Machine learning algorithms have garnered significant attention due to their ability to analyze rainfall data from the past and produce accurate forecasts. This work uses machine learning algorithms to estimate rainfall patterns based on a number of meteorological factors, including pressure, temperature, humidity, and wind speed. A number of procedures, including feature selection, model training, evaluation, and data preprocessing, are used to identify the approach that best meets the prediction aim. By examining the complex relationships between meteorological variables and rainfall patterns using historical rainfall data, the models can be used to produce accurate and timely forecasts. The use of machine learning methods to forecast rainfall supports decision-making in a range of weather-dependent businesses by offering a data-driven method that increases forecast accuracy and reliability [3].

In China, rainfall is the main cause of landslides; hence, it is extremely important to anticipate the spatial-temporal hazards of landslides caused by rainfall. The majority of nations and areas now base their landslide hazard prediction systems solely on rainfall data, which leads to a high rate of false alarms and poor spatial precision in the hazard forecast outputs. Based on multi-modal earth observation data, this research suggests a hazard prediction model that considers landslide triggering causes, landslide predisposing settings, and the spatial regularity of previous landslides. Hong Kong's rainfall-induced natural terrain landslides' spatial-temporal hazard prediction performance has been considerably enhanced by the suggested model [4].

Real-time heavy rainfall event (HRE) prediction is a major difficulty in India, especially in areas with complicated landscape, such as Assam, where severe hydro-meteorological events are frequently connected to flash floods that have detrimental effects on the entire region. The catastrophic HREs in June 2022, which resulted in numerous fatalities, extensive damage, and economic losses totaling 200 crores, forced the evacuation of almost 4 million people. Severe flooding occurred in Assam as recently as June 2023. In order to overcome the limitations of deterministic numerical weather models in accurately forecasting these events, the study explores the inclusion of deep learning (DL) models, particularly U-Nets, using simulated daily accumulated rainfall outputs from various parameterization techniques. With a mean absolute error (MAE) of less than 12 mm, the U-Net-based model forecasted rainfall at the district scale for four days in June 2022 better than individual and ensemble model outputs. When the DL model was compared to the weather research and forecasting (WRF) model, the MAE over Assam decreased by a significant 64.78% projections [5].

The prudent distribution of water supplies, flood forecasting, and other factors may depend on accurate rainfall forecasting. The machine learning-based data self-driven model is more flexible when dealing with this type of erratic data. This research proposes VMD-GRA-Elman is a group sum prediction model based on variation modal decomposition (VMD), gray correlation analysis (GRA), and Elman neural network (Elman). Raw rainfall data was first separated using VMD, and then the split data was correlated using grey relational analysis. The data were then combined into high-frequency and low-frequency data. While combining data can save processing time, the forecast accuracy of combining high-frequency data is obviously reduced, while the prediction accuracy of combining low-frequency data is essentially unaffected [6].

This research suggests a novel RR-DP-PSO algorithm model for rainfall prediction in order to address the issue of previous algorithms' poor accuracy. To identify characteristics in rainfall data and produce accurate predictions, the algorithm combines the Ridge regression, PSO, and DP-PSO algorithms. The RR-DP-PSO method clearly has benefits in rainfall prediction, according to the experimental results on the rainfall dataset in the Baiyun area of Guangdong Province. The RR-DP-PSO algorithm shows a reduced mean square error (MSE), a greater accuracy of 86.67%, and a more accurate prediction when compared to the conventional techniques of SVM, LSTM, and ANN. This method effectively increases the precision and dependability of rainfall forecasts by combining a number of strategies. The RR-DP-PSO algorithm's successful trials have shown its prospective uses in meteorology and natural disaster forecasting, and they have also yielded fresh concepts and techniques for upcoming advancements in meteorological prediction models [7].

Accurate rainfall data forecasting can successfully stop flood-related losses. Many algorithmic models have failed to predict rainfall effectively in recent years. Based on the PLS-ARMIA algorithmic model, this is suggested. Initially, the partial least squares approach is used to process the historical non-smooth rainfall data in order to obtain smooth rainfall data. Rainfall prediction is then carried out using the model of the ARMIA algorithm. Lastly, by comparing the experiments based on the PLS-ARMIA algorithm model with the BP algorithm model, LSTM algorithm model, and ANN algorithm model, the final experimental results compare the MSE, RMSE, and MAPE of the PLS-ARMIA algorithm model with the other three models. The PLS-ARMIA algorithm model has the best MSE of 32.21, RMSE of 4.25, and MAPE of 32.15% among these four models, demonstrating that the PLS-ARMIA algorithm model can accurately predict rainfall. The study's findings can provide crucial data support and a scientific foundation for decision-making in a variety of departments, including agricultural, meteorological, and water conservation [8].

Predicting rainfall accurately is essential for agriculture, water resources, and disaster preparedness. In order to forecast rainfall in Australia, we employ two sophisticated research designs in this example the Random Forest classifier and logistic regression. Kagle, which offers a comprehensive overview of climatic changes across time, served as the source of the data set used in this study. The goal of our study is to create a reliable prediction model for the many forms of precipitation that may occur in Australian territories. While Logistic Regression gives a baseline model, Random Forest Classifier offers an ensemble-based approach that improves accuracy. To improve the performance of the model, we search for content engineering and hyper parameter tuning. The study's findings are significant for stakeholders and decision makers because precise rainfall predictions can improve risk mitigation, resource management, and sustainability planning across a range of domains. The outcome of our models is that they will help us better understand the rainfall patterns in Australia, which will aid in preparedness and decision-making [9].

Accurate rainfall forecasting in dynamic meteorological prediction is a riddle. We present a novel approach that uses the Gooseneck Barnacle Optimizer and the Support Vector Machine for Least Squares (LSSVM) to anticipate weekly average rainfall patterns in a complex and dynamic natural setting with unpredictable sky movements. The wide-ranging effects of rainfall prediction, notably on public health, demonstrate its importance. However, our preparation is limited since traditional single-model and machine-learning approaches must account for variations in rainfall patterns. We investigate the potential of this new combination using a 2014–2018 dataset from credible meteorological sources. The foundation of our approach is a new evolutionary algorithm called the Gooseneck Barnacle Optimizer (GBO) was inspired by the intricate mating behaviours of gooseneck barnacles. Wind direction, food availability, food attraction, navigational sperm throwing properties, and intertidal zone wave movement are all dynamically intertwined during mating is well captured by GBO, which leads to two essential optimization stages: exploration and exploitation [10].

EXISTING SYSTEM

Due to the system's inability to accurately forecast crop yields and soil fertility, farmers face significant obstacles. This vulnerability prompts loss of efficiency, as ranchers can't conclude which yields are appropriate for their property. Additionally, they frequently struggle to select the appropriate seeds, making their farming decisions even more difficult. Without precise gauges, ranchers face misfortunes and it becomes hard to work on their cultivating rehearses and accomplish improved results. Also, inability to appropriately quantify how much manure required frequently prompts abuse, waste, and damage to the climate. Lacking sufficient data and apparatuses, ranchers will generally utilize over the top measures of composts, which inflates costs as well as damages soil wellbeing over the long haul. Tending to these difficulties with cutting edge prescient models and information driven proposals can fundamentally work on farming efficiency and manageability.

Disadvantages:

- Fertilizer wastages.
- Expensive.
- Less yield.

REQUIREMENT ANALYSIS

Evaluation of the Rationale and Feasibility of the Proposed System

The prediction of the nutrients required to fertilize crops the main goal of this research is to use machine learning techniques. Utilizing machine learning, the suggested solution gives farmers accurate fertilizer recommendations by forecasting the nutrients that are required by various crops. It assists ranchers with decreasing wastage, set aside cash and increment yields by utilizing the perfect proportion of compost. By reducing the environmental impact of excessive fertilizer use, it supports sustainable agriculture. Modern technology made the system possible, it is simple easy to use, and it may be combined with current farm equipment, making it a practical and affordable option for farmers.

PROPOSED SYSTEM

The proposed framework utilizes AI procedures, for example, choice trees and irregular timberlands to anticipate explicit supplements expected by various kinds of yields. This approach empowers more proficient and savvy ID of agrarian requirements, furnishing ranchers with exact proposals customized to their yields. The system makes it possible to quickly analyze data about crops and soil to provide accurate recommendations for nutrient requirements, which helps farmers make better decisions. By allowing farmers to tailor fertilizer application based on the unique needs of respective crops, the system also significantly reduces waste. Farmers can now optimize their use of fertilizers, ensuring that each crop receives the appropriate nutrients, as opposed to blanket application. By reducing the environmental impact of excessive fertilizer use, this targeted approach promotes more sustainable agricultural practices in addition to cost savings and resource conservation.

Advantages:

- Easy to predict
- Good yield
- Inexpensive.
- Low wastage.

SELECTED METHODODLIGIES

Machine Learning:

The machine learning (ML) area of AI and computer technology focuses on using statistics and algorithms to mimic how AI follows human study and progressively improves its accuracy. Decision-Making Process Usually, computer learning algorithms are used to make predictions and categorize data. Your calculation evaluates the example in the records based on different information measurements that may be named. An error function is an error attribute that assesses the model estimate. Models can be used to examine the accuracy of the issue representation model through correlations. Method of improving the model the loads are adjusted to reduce the difference between the known event and the expected rendition, assuming the model best captures the reality covered in the preparation dataset. This "assessment and advancement" process is repeated in the calculation, which continuously refreshes the loads until an exactness edge is achieved. Given that deep learning and artificial intelligence are typically used in tandem, the differences between the two are quite important. Neural networks, two subcategories of artificial intelligence are machine learning and deep learning. But brain networks are a subset of artificial intelligence, and deep learning is a subset of brain organizations. Deep learning and AI differ in how each computation learns. Directed learning, sometimes known as "profound" computational learning, can use named informational indexes to shed light on its principles, but it is not actually a described informational index. From raw data, like text or photos, a deep learning method may reliably find a collection of consistent features that differentiate one kind of information from another. This removes the need for human interaction and allows for the usage of large amounts of data. In this MIT session, Lex Friedman discusses how profound learning can be viewed "at the contraption learning level" (link is external to IBM.com).

SYSTEM ARCHITECTURE

The portrayal of the general characteristics of the product is connected to the meaning of the prerequisites and the laid-out request of a serious level of the contraption. Numerous web pages and their connections are described and designed during architectural design. Key software components are defined, broken down into processing modules and conceptual records systems, and the connections that exist between them are explained. The proposed framework characterizes the accompanying modules

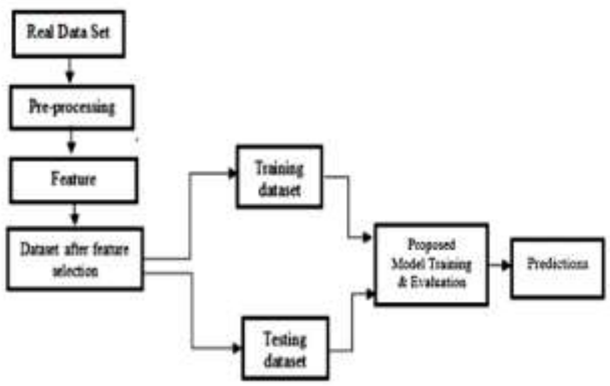


Fig 1: System Architecture

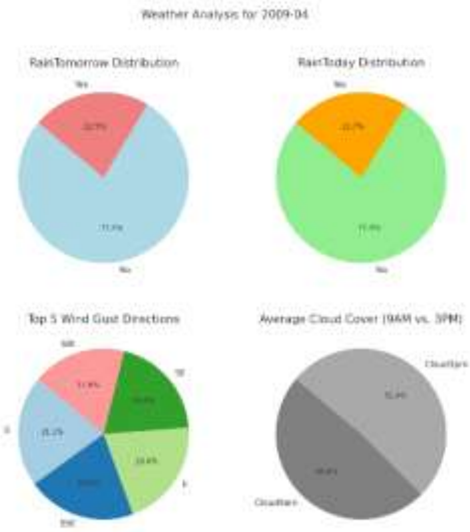
SYSTEM MODULE

- Data Collection Module.
- Data Pre-processing Module.
- Machine Learning Module.
- Recommendation Engine Module.
- User Interface Module.

Data Collection Module:

The data necessary for precise nutrient estimations is gathered by the data collection module. This incorporates soil assortment information, for example, pH levels and supplement content (nitrogen, phosphorus, potassium), climatic information like precipitation and temperature, and explicit yield subtleties. This data comes from sensors, weather APIs, and farmer collaborations to make a complete dataset that can be used for more research.

Data Pre-processing Module: The information pre-processing module cleans and readies the gathered information for examination. It handles absent or



conflicting information, standardizes the information to guarantee all values are a similar size, and sorts out them in a configuration reasonable for AI. This step guarantees that the information is precise and prepared to construct solid prescient models.

Machine Learning Module:

The AI module utilizes calculations, for example, choice trees and arbitrary backwoods to investigate pre-handled information. It trains models utilizing this information to learn examples and connections between soil properties, atmospheric conditions and yield prerequisites. When prepared, these models can precisely anticipate the particular supplements expected by various yields. This aides in giving precise manure proposals.

Recommendation Engine Module:

The proposal motor module utilizes prepared AI models to produce explicit supplement necessities for various harvests. It considers information, applies expectations and gives customized manure suggestions to ranchers. In addition, it suggests additional inputs that are necessary for crop growth to guarantee that each crop receives the appropriate nutrients in the appropriate quantities.

User Interface Module:

The UI module is the piece of the framework through which ranchers communicate. It provides farmers with a straightforward and user-friendly platform for entering soil and crop data. The point of

Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

interaction shows customized supplement and compost proposals, making it simpler for ranchers to comprehend and apply the suggestions to further develop crop yields.

WEATHER ANALYSIS

The analysis for April 2009 includes an evaluation of rainfall, wind gust directions, and cloud cover to gain insights into weather trends. The findings are visualized using pie charts representing key distributions:

Rainfall Prediction:

The "RainTomorrow" distribution shows that 22.5% of the days experienced rainfall the next day, while 77.5% remained dry.Similarly, the "RainToday" distribution indicates that 22.7% of the days had rainfall, whereas 77.3% did not.

Fig 2: Weather Analysis for April 2009

Wind Gust Direction:

The top five dominant wind gust directions include South (21.1%), South-Southeast (20.6%), East (20.6%), Southeast (19.8%), and Southwest (17.8%). This distribution suggests that winds predominantly originated from the southern and southeastern directions during this period.

Cloud Cover Analysis:

A comparison of average cloud cover between 9 AM and 3 PM reveals a relatively balanced distribution, with 48.6% cloud cover in the morning and 51.4% in the afternoon.

These insights provide valuable information for understanding weather patterns, aiding in rainfall prediction models, and optimizing agricultural decision-making.

RESULTS & DISCUSSION

We looked at statistics from the Indian state of Bihar from 1901 to 2017. The various graphs' visualizations aid in our comprehension of the data and guide our decision-making over the best course of action. It offers significant perspectives. Using precise numerical and scientific approaches, forecasting provides relevant and trustworthy information on past, present, and future activity. Predicting the numerical numbers for a given assignment involves a few steps. The first step is to identify the issue with a thorough study, and the second is to gather the necessary information to assess the issue for more estimation. Compare the estimated and real numbers after making any necessary adjustments. The rainfall is plotted by year thanks to the arrangement of the data.

SCREEN SHOTS

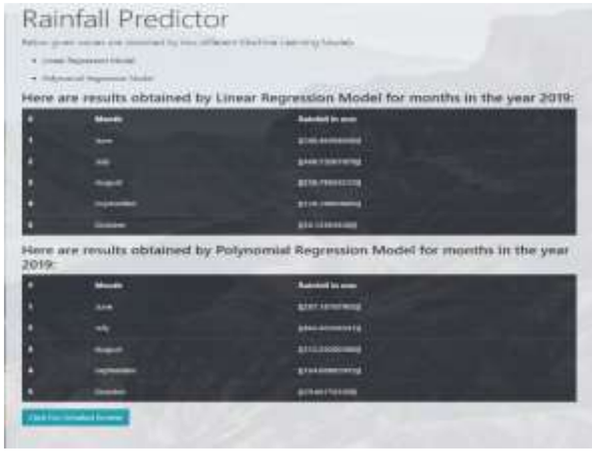


Fig 3: Image of Rainfall Predictor

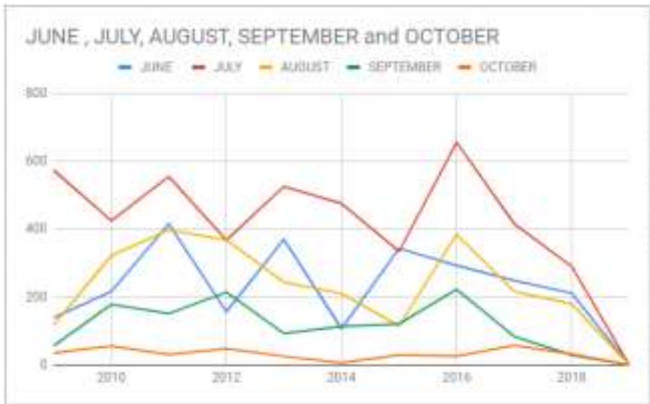


Fig 4: Image of Statistics & Models

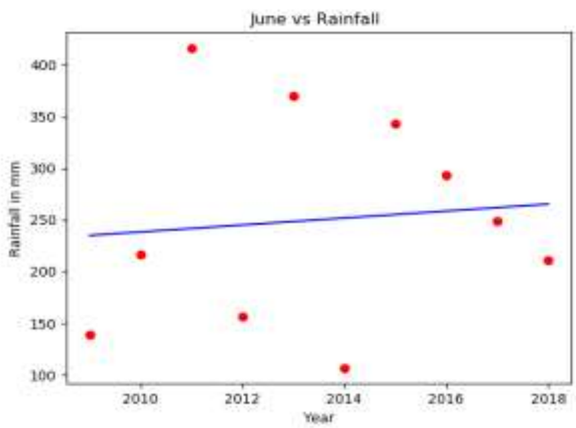


Fig 5: Image of Linear Model

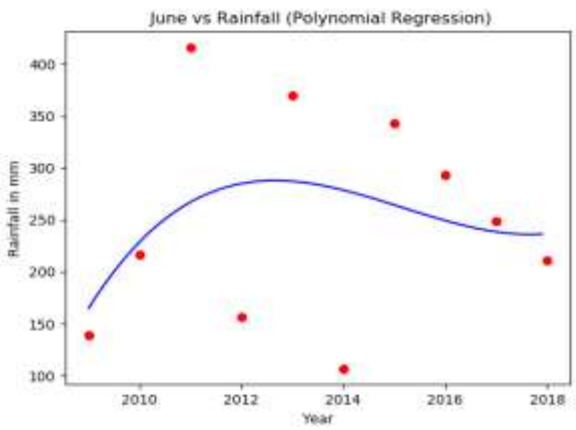


Fig 6: Image of Polynomial Model

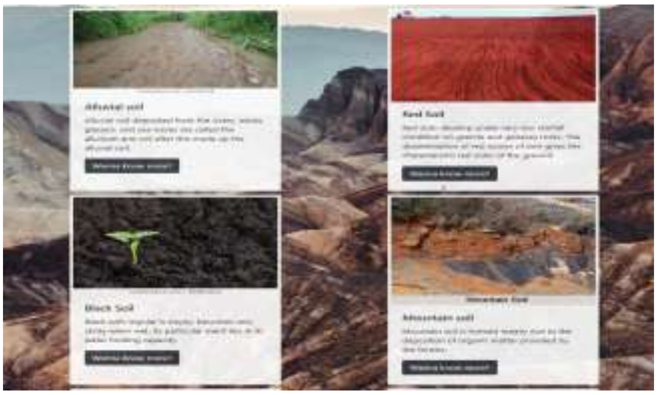


Fig 7: Image of Types of Soils

Fig 8: Image of Soil & Crop Recommendation

Fig 9: Image of Crop Recommendation for Soil

CONCLUSION

All in all, the presentation of AI methods, for example, choice trees and irregular timberlands to anticipate crop supplement necessities addresses a critical improvement in farming practices. The proposed framework gives exact and customized proposals to ranchers, bringing about additional effective and prudent cultivating. The system contributes to reducing waste and optimizing resource use by quickly and precisely determining fertilizer requirements. It builds yield and efficiency as well as advances reasonable cultivating rehearses. By and large, this innovative methodology offers a promising answer for a portion of the significant difficulties confronting present day horticulture, empowering ranchers to settle on informed choices and accomplish improved results.

FUTURE SCOPE

A machine learning-based flood prediction system has a lot to offer communities that are susceptible to floods. Machine learning algorithms can accurately predict the likelihood and intensity of a flood using past data and real-time monitoring, allowing authorities to take preventative action to minimize the effects on the neighbourhood. Machine learning algorithms, like Random Forest, can be trained on large datasets of historical flood data, meteorological data, topographical data, and other relevant parameters to accurately predict the likelihood of future floods. A machine learning-based flood prediction system can save lives, improve emergency response times, and stop property damage. Overall, considering its potential benefits, any town at risk of floods would be smart to invest in such a system.

REFERENCES

[1] M. Ahmed, M. H. Sulaiman, M. Rahman and A. J. Mohamad, "A Novel Gooseneck Barnacle Optimizer Based Machine Learning Approach for Rainfall Prediction," 2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI), Dhaka, Bangladesh, 2023, pp. 1-6, doi: 10.1109/STI59863.2023.10464846.

[2] B. Setya, R. A. Nurhidayatullah, M. B. Hewen and K. Kusrini, "Comparative Analysis Of Rainfall Value Prediction In Semarang Using Linear And K-Nearest Neighbor Algorithms," 2023 5th International Conference on Cybernetics and Intelligent System (ICORIS), Pangkalpinang, Indonesia, 2023, pp. 1-5, doi: 10.1109/ICORIS60118.2023.10352274.

[3] A. Raut, D. Theng and S. Khandelwal, "Random Forest Regressor Model for Rainfall Prediction," 2023 International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/ICCAMS60113.2023.10526085.

[4] P. S. Sai, T. M. Thiyagu and P. G. Soumya, "Rainfall Prediction using Machine Learning Algorithms," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 677-681, doi: 10.1109/ICAAIC60222.2024.10575126.

[5] Y. Chen et al., "Spatial-Temporal Hazard Prediction of Rainfall-Induced Landslides Using Multi-Modal Earth Observation Data," IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 2024, pp. 3761-3764, doi: 10.1109/IGARSS53475.2024.10640781.

[6] O. Sharma, D. Trivedi, S. Pattnaik, V. Hazra and N. B. Puhan, "Improvement in District Scale Heavy Rainfall Prediction Over Complex Terrain of North East India Using Deep Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-8, 2023, Art no. 4106408, doi: 10.1109/TGRS.2023.3322676.

[7] S. Hu and X. Wang, "Monthly Rainfall Prediction Based on VMD-GRA-Elman Model," 2023 4th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Nanjing, China, 2023, pp. 95-102, doi: 10.1109/ICBASE59196.2023.10303054.

[8] P. Deng, L. Yao, S. Huo and H. Wei, "Refined Rainfall Prediction Using the RR-DP-PSO Algorithmic Model," 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), Shanghai, China, 2024, pp. 577-582, doi: 10.1109/ICAACE61206.2024.10548598.

[9] S. Huo, T. Lu, D. Wei and P. Deng, "A Research on Rainfall Prediction Based on PLS-ARMIA Algorithm Modelase," 2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qiangdao, China, 2023, pp. 668-672, doi: 10.1109/ICFTIC59930.2023.10456096.

[10] R. Pandey, M. Upadhya and M. Singh, "Rainfall Prediction Using Logistic Regression and Random Forest Algorithm," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 663-668, doi: 10.1109/IC2PCT60090.2024.10486681.