Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Agriculture Data Analysis Using Random Forest And Linear Regression

[1]P Bhargava ,[2]P Sudheer Kumar, [3]P Sai Durga Naveen,[4]Dr. D. Usha, [5]Dr.T. Kumanan

[1,2,3] Final Year CSE, [4,5]Professor
[1,2,3,4,5]Department of Computer Science and Engineering
[1,2,3,4,5]Dr. M. G. R Educational and Research Institute, Maduravoyal, Chennai, India
Pathakamuri1111@gmail.com,usha.cse@drmgrdu.ac.in

Abstract
India is a horticultural nation and its economy is to a great extent founded on harvests and precipitation. To keep an eye on crop yield, all farmers need to know how much rain will fall. Guaging is the use of science and innovation to anticipate the condition of the air. Planning for water quality, crop productivity, and efficient use of water resources all depend on accurate rainfall measurements. Utilizing different learning strategies, the machine can foresee precipitation. AI strategies are utilized to assess precipitation. This article centers around some famous AI calculations for precipitation estimating. This article compares algorithms such as random forest, polynomial regression, and simple linear regression. From this examination, it is feasible to dissect which strategy gives the best precipitation exactness.

Keywords: Agriculture, Data Analysis, Machine Learning, Rainfall Prediction, Crop Production

I INTRODUCTION

Expecting the rush is a troublesome errand. Albeit a few components have previously been proposed, precise precipitation determining is extremely challenging. In provincial regions like India, crop achievement or disappointment at whatever year and water shortage are dependably principal. An agricultural region can suffer catastrophic consequences from even minute variations in seasonal rainfall. Preventing losses and shocks from natural disasters can be made easier with accurate rainfall forecasts. In specific circumstances, for example, floods and dry seasons, profoundly exact precipitation estimates are helpful for farming administration and fiasco anticipation. Various solutions are provided in this article. Rainfall forecasting can benefit from the use of machine learning methods.

II OBJECTIVE

It is vital and significant for all ranchers to trust that the precipitation will break down the exhibition of the gather. Prescient anticipating is the utilization of science and advancement to foresee the situation. Precise assurance of precipitation is vital for fruitful usage of water assets, expanding crop yields and early preparation of water structures.

III RELATED WORK

1. Expectation of Indian Summer Storm Precipitation (ISMR) utilizing time series information: A specialist framework in view of fluffy entropy and neurons.

The creators introduced a prescient model for Indian summer rainstorm precipitation at month to month and occasional scales. A period series dataset covering the period 1871 to 2020 was utilized for determining. The dataset is partitioned into two sections (1) 1871-1960 utilized as preparing information and (2) 1961-2020 utilized as testing information. Measurable investigation shows elements of rainstorm which can't be actually figured out by numerical and factual models. Thusly, the creators of this study recommended involving three techniques for this kind of expectation: fluffy set, entropy and fake brain organization. Utilizing these three strategies, a prescient model was created thinking about the powerful idea of ISMR. In the proposed model, despite the fact that he put it in principle, he utilized it to manage the vulnerabilities innate in certain information. The idea of entropy

estimation is adjusted in this model and used to give contribution to the entropy capability as enrollment levels. Fuzzy Information Gain is the name of this entropy function (FIG). Using ann, both rules are fixed. The fluffy worth of each state is then utilized as contribution to the ANN. The proposed model is named "Fluffy entropy and neuron-based master framework for ISMR expectation" since it is a mix of entropy and ANN. The accompanying accuracy measures were utilized to assess the proposed model exhibition: performance parameters (PP), the correlation coefficient (CC), the standard deviation (SD), and the root mean square error (RMSE). In light of the outcomes, the proposed model is viable and effective contrasted with other existing models.

2. Complete assessment of seven AI techniques for precipitation estimating utilizing meteorological subsidiaries.

The scientists analyzed the anticipated execution of the new and complex technique that drawn out Markov affix with precipitation estimating to other generally utilized learning devices: support vector relapse, hereditary programming, M5 rules, M5 trees model, spiral brain premise. Ligula organizations and K neighbors. Everyday precipitation video information were gathered from
42 urban communities on two mainlands with various environment frameworks. 20 urban areas chose from across Europe, 22 urban areas from across America. There were two reasons why two continents were chosen for data extraction: one, to carry out the experiment in a variety of climates and under a variety of weather conditions; the other, to serve as local areas because the selected cities are far apart. The ultimate objective is not to connect the experiment to a particular climate or region. Contingent upon the outcomes, combined precipitation gives improved results contrasted with anticipating utilizing day to day precipitation information. Utilizing bunched information, vector relapse, spiral premise capabilities, and hereditary programming by and large perform well, yet outspread premise capabilities beat complex Markov chain techniques. For all chose informational indexes, every technique utilized similar boundaries, so it isn't ensured that all strategies utilize the best boundaries. In the trial, the analysts noticed the connection among precision and environment estimate attributes, for example, factor precipitation designs, most extreme precipitation sums, and spatial scope of precipitation. There was likewise no massive distinction in the forecast mistake of the algorithms using of water structures.

3. A mixed statistical model of the daily precipitation decreases.

The creators proposed a crossover strategy to gauge day to day precipitation by joining two techniques:
1) arbitrary leap and 2) support vector machine. RF was picked for its vigor in characterization and was utilized to anticipate whether it would rain, while SVM was picked for its capacity to fit the information non-straightly and was utilized to anticipate how much downpour. Downscaling daily rainfall at three stations on the east coast of Peninsular Malaysia was used to evaluate the proposed model: Kemaman, Duncan, and Besut the Malaysian Irrigation and Drainage Department's collection of daily rainfall data from 1961 to 2000 is arranged chronologically. A sum of 26 environment informational indexes were gathered by the Public Place for Natural Expectation Reanalysis dataset and utilized as indicators for model downscaling. The reliability of the rainfall time series was evaluated using a variety of quality control procedures. Histograms were created for the informational indexes to mirror the inquiries, what's more, Understudy's t test was likewise used to distinguish contrasts between two gatherings of informational indexes, which were at last observed to be comparative in each of the three regions. As per the outcomes, the crossover strategy can decrease precipitation with Nash-Sutcliffe productivity in the scope of 0.90-0.93, which is higher than RF and SVM models.

IV EXISTING SYSTEM

Our farming is the strength of Indian economy. The culture of the Etesia farmer is all that matters. Crops need great soil, excrement and great air supply for good yield. Weather conditions determining is vital for the requirements of each and every rancher. Because of abrupt environment/environment changes, individuals endure financially and truly. Weather conditions guaging is quite possibly of the most troublesome issue in the cutting edge world.

The principal focal point of this paper is weather conditions estimating utilizing different AI strategies. Arrangement, grouping, independent direction, brain networks and so forth. Data connected with weather

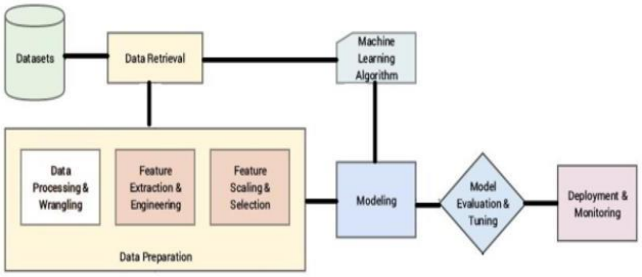
conditions is otherwise called meteorological information. In this article, the most regularly utilized meteorological boundaries are precipitation, wind speed, temperature, and cold.

V PROPOSED SYSTEM

Precipitation is significant for food creation projects, water the executive's tasks and all open air exercises projects.Delayed dry season or weighty precipitation during development and advancement phases of the harvest can prompt critical decrease in crop yield.India's economy is based on major crops and is an agricultural nation.Hence precipitation estimating turns out to be vital in rural nations like India.Precipitation determining has been one of the most difficult logical and innovative issues of the last 100 years.

VI SYSTEM ARCHITECTURE

Figure 6.1 System Architecture



DATA FLOW DIAGRAM

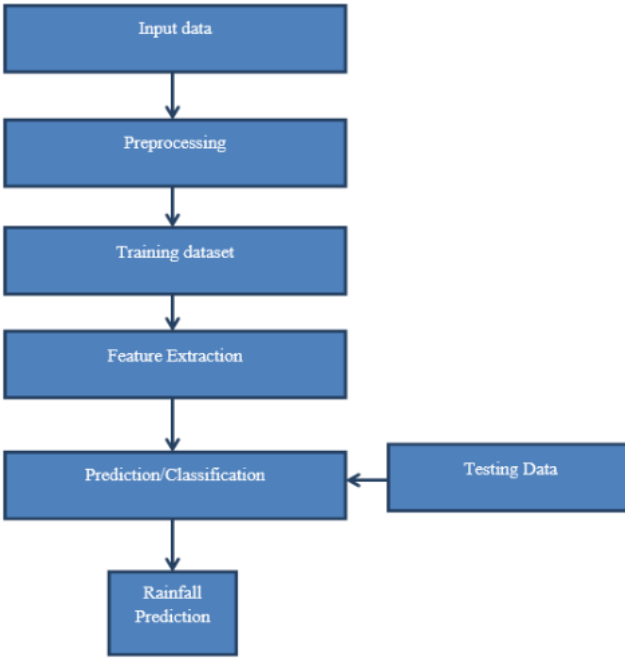


Figure 6.2 Data Flow Digram of Rain fall Prediction

VII MODULES

1. Data Collection
2. Data Cleaning
3. Data Selection
4. Data Transformation
5. Machine Learning Stage

1. Data Collection

The information utilized for this work were gathered at a weather conditions station. Case information covers the period 2012 to 2015. The accompanying techniques were utilized in this piece of the review: information cleaning, information choice, information change and AI.

2. Data Cleaning

As of now, a standard information design is fostered that permits looking for missing information, looking for copy information, and eliminating invalid information. At last, the cleaned PC information is changed into a structure reasonable for AI.

3. Data Selection

In this step, pertinent information for examination, for example, dropping tree is chosen and removed from the dataset. The climate dataset comprised of ten ascribes, two of which were utilized to anticipate what's in store. Due to the fact that cloud data does not contain a high percentage of values and has a format in which all values are identical, neither was used in the analysis.

4. Data Transformation

"It's additionally called information coordination." Here chose information is changed into structures reasonable for AI. Information documents were put away in comma-isolated design (CVS) and informational indexes were standardized to limit the impact of scaling on the information..

5. Machine Learning Stage

The hardware review stage is separated into three stages. At each stage, all strategies were utilized to investigate the meteorological information. The testing strategy utilized for this review is parted rate: preparing on a specific level of the informational index, cross-approving it, and testing on the leftover rate. After this, they found intriguing examples that address information.

VIII RESULT ANALYSIS

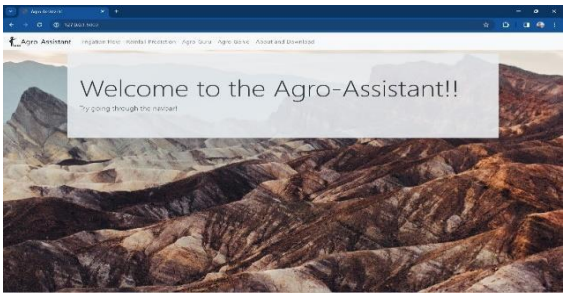


Figure 8.1 Homepage of Web Application

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy.

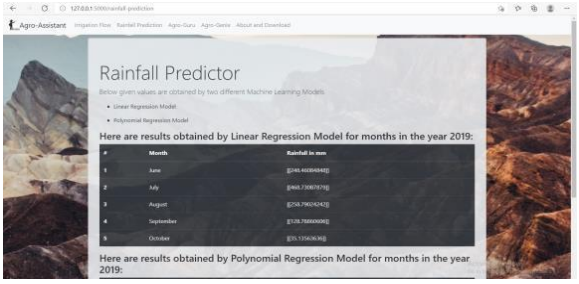


Figure 8.2 Rainfall Predictor Page

Statistics and Models

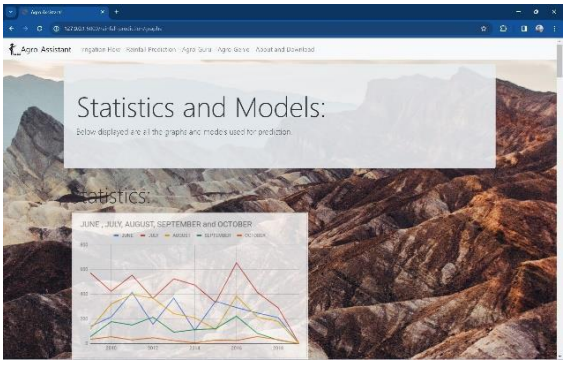


Figure 8.3 Graph of Rainfall Prediction

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

Crop Prediction

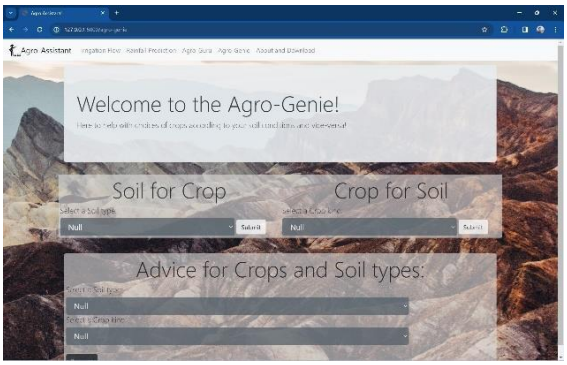


Figure 8.4 Crop and Soil Prediction Page

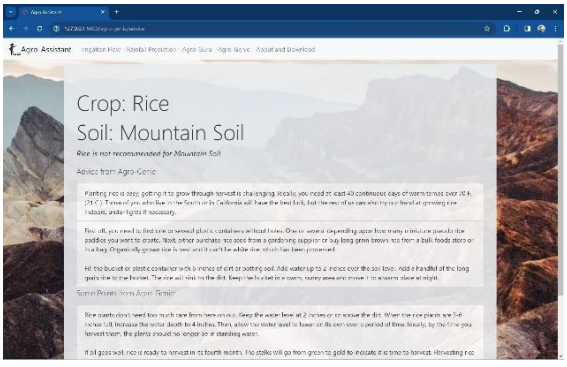


Figure 8.5 End User Output

IX CONCLUSION

Determined weather conditions is a meteorological errand that can be effectively adjusted by specialists utilizing a mathematical weather conditions estimating framework. Weather conditions determining is finished utilizing different AI strategies, particularly basic direct and polynomial order methodology. The fundamental objective is to improve characterizations and expectations for legacy; In this work, a weather conditions estimating model was planned and created. However, prior to implementing the suggested approach, it is necessary to examine the

model's other limitations in the future. In the field of rainfall forecasting, there are a number of issues that can make machine learning techniques more effective.

X FUTURE SCOPE

The framework can be improved by adding the accompanying. Image processing for crop disease detection. Their clients can transfer pictures of yield illnesses and Get proposals on pesticides. Presentation of astute water system framework Screen climate and soil conditions, oceanic plants application, and so on. To consequently change the water system plan.

REFERENCES

[1] Xiong, Lihua, and Kieran M. OConnor. "An empirical method to improve the prediction limits of the GLUE methodology in rainfallrunoff modeling." Journal of Hydrology 349.1-2 (2008): 115-124.

[2] Schmitz, G. H., and J. Cullmann. "PAI-OFF: A new proposal for online flood forecasting in flash flood prone catchments." Journal of hydrology 360.1-4 (2008):1-14.

[3] Riordan, Denis, and Bjarne K. Hansen. "A fuzzy casebased system for weather prediction." Engineering Intelligent Systems for Electrical Engineering and Communications 10.3 (2002): 139- 146.

[4] Guhathakurta, P. "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network." Current Science 90.6 (2006): 773-779.

[5] Pilgrim, D. H., T. G. Chapman, and D. G. Doran. "Problems of rainfall-runoff modelling in arid and semiarid regions." Hydrological Sciences Journal 33.4 (1988): 379-400.

[6] Lee, Sunyoung, Sungzoon Cho, and Patrick M. Wong. "Rainfall prediction using artificial neural networks." journal of geographic information and Decision Analysis 2.2 (1998): 233-242..

[7] French, Mark N., Witold F. Krajewski, and Robert R. Cuykendall. "Rainfall forecasting in space and time using a neural network." Journal of hydrology 137.1- 4 (1992): 1-31.

[8] Charaniya, Nizar Ali, and Sanjay V. Dudul. "Committee of artificial neural networks for monthly rainfall prediction using wavelet transform." Business, Engineering and Industrial Applications (ICBEIA), 2011 International Conference on. IEEE, 2011.

[9] Noone, David, and Harvey Stern. "Verification of rainfall forecasts from the Australian Bureau of Meteorology's Global Assimilation and Prognosis(GASP) system." Australian Meteorological Magazine 44.4 (1995): 275-286.

[10] Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators." Neural networks 2.5 (1989): 359-366.

[11] Haykin, Simon. Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[12] Rajeevan, M., Pulak Guhathakurta, and V. Thapliyal. "New models for long range forecasts of summer monsoon rainfall over North West and Peninsular India." Meteorology and Atmospheric Physics 73.3-4 (2000): 211-225.