IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 4, May 2023

Rainfall Forecasting and Density Estimation Using Machine Learning

Esther Gandhimathi, Ruthesh Prem Kanth CH, Boyella Sumanth, Shaik Sohail, Syed Masthan

Dhanalakshmi College of Engineering, Chennai, India

Abstract*: India is agrarian and its economy is primarily based on crop productivity and rainfall. Rainfall forecasting is very important for all farmers to analyse crop productivity. Utilising science and technology to forecast atmospheric conditions is known as precipitation forecasting. For effective use of water resources, crop productivity, and long-term planning of water resources, accurate rainfall estimation is essential. It can predict rainfall using a variety of machine learning techniques. equipment that is used to estimate rainfall. The prominent machine learning techniques for predicting rain are the main topic of this essay. In this article, several methods are examined, including the random forest algorithm, simple linear regression, and polynomial regression. This comparison allows us to determine which algorithm forecasts rainfall with more accuracy.*

Keywords: Forecasting.

I. INTRODUCTION

Forecasting precipitation is one of the most challenging tasks. Although many algorithms have already been proposed, accurate precipitation forecasting is still very difficult. The success or failure of the crop and the availability of water in any given year are always viewed with the highest concern in an agricultural nation like India. Agriculture can be severely harmed by even small variations in seasonal rainfall. Precipitation forecasting that is accurate has the ability to reduce the likelihood of natural disasters and the harm they do. Highly precise rainfall forecasting is helpful for agricultural management and disaster prevention in some situations, such as floods and droughts. In this study, many algorithms have been examined. Forecasting rainfall successfully makes use of machine learning techniques.

II. METHODOLOGY

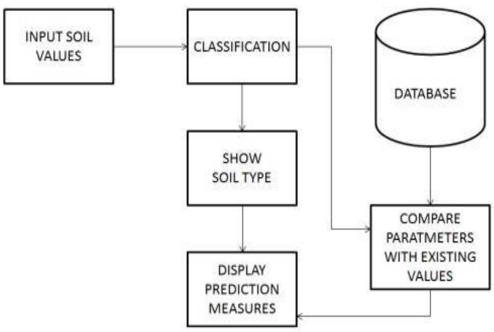

Figure 1: Block Diagram of Rainfall & Soil Analysis System.

The main purpose of this method is to classify soil according to nutrients. For this we obtained a database of soil samples. Soil is called victim classification, tree algorithm and soil type to be displayed. We measure square footage to accurately predict crops for a given soil. Also, until now, we tend to improve the soil if the farmer wants to produce a clear yield on the same soil, suggesting the nutritional requirements for the same soil. ID3 (Iterative Dichotomise) is used to generate a call tree 3) supporting categories (eg low, high, medium). Naive Bayes examines the files and trains the algorithm using this knowledge. ID3 categorises square footage into low, medium, and high categories. A "decision point" is a weighted category, and a point is its square measurement. Depending on the circumstance, a split (leaf point), a portion of the field, and a corresponding "class" child point may be used. For instance, the nitrogen value (N-value) may be less than 40% and be considered "low". The N-value of the call point's location in this instance and the category that belongs to the "low" soil pattern. An ordered set of files can be produced through entropy. As a result, the amount of information needed for categorization is inversely correlated with the rise in entropy. Entropy ought to be equal to potential, then. Before training, this is frequently accomplished by manually filtering data in response to a set of computer commands. Information, on the other hand, is corrupted or turns into random interference. During event classification, the remaining value is known as information gain (I.G.). the classification that most closely resembles the previous one (accidental destruction or degradation of an object).

Algorithm:

Step 1: Using the data set, determine the likelihood that each characteristic (an object being harmed or degraded by chance) will occur.

Step 2: Cut the collection into bits that are as little as you can (where something breaks or becomes a random mess).

Step 3: Decide which tree node that attribute is located in.

Step 4: Repeating data in pieces, residual attributes must be categorised using the ID3 algorithm.

III. LITERATURE REVIEW

A Fuzzy-Entropy-Neuro Based Expert System for Forecasting Indian Summer Monsoon Rainfall (ISMR) Using Time Series Data is found in [30].

AUTHORS : Pritpal Singh(30 July 2017)

The authors present a model to predict Indian summer rainfall on a monthly and seasonal basis. Data from 1871 to 2014 were used for estimation. The dataset used as training data is (1) 1871-1960 and (2) 1961-2014 as test data. Statistical analysis informs the dynamic behaviour of monsoon rainfall, which cannot be effectively predicted by mathematical and statistical models. Therefore, the authors of this study recommend using three methods for this type of prediction: Fuzzy Set, Entropy, and Artificial Neural Network. Using these three

methods, a prediction model was developed to deal with the dynamic nature of ISMR. The proposed model uses complex ensemble theory to resolve the uncertainty in the database. The concept of entropy computation is modified in this model to introduce the entropy function as a degree of membership. This approach modifies the idea of entropy computation by introducing the entropy function as a degree of membership. Fuzzy Information-Gain is the name of the entropy function (Figure). After that, ANN is used to resolve each fuzzy rule. Each fuzzy set's fig value serves as the input for the ANN. Due to the inclusion of a fuzzy set, entropy, and ANN, the suggested model is known as an "entropy-neural-based artificial intelligence system for ISMR prediction". Performance parameters (PP), root mean square error (RMSE), standard deviation (SD), and correlation coefficient (CC) are used to assess the performance of the suggested model. The results show that the suggested model outperforms existing models in terms of efficiency and effectiveness.

An In-Depth Analysis of Seven Machine Learning Techniques for Predicting Rainfall in Weather Derivatives

AUTHORS : Sam Cramer, Michael Kampouridis , Alex A. Freitas,

Antonis K. Alexandridis b/ 2017

With other widely used machine learning techniques, such as Support Vector Regression, Genetic Programming, M5 Rules, M5 Model trees, Radial Basis Neural Networks, and k-Nearest Neighbours, the researchers in this study compared the predictive performance of the most recent and cutting-edge method, "Markov chain extended with rainfall prediction." Datasets for daily precipitation were gathered from 42 cities on two continents with a wide range of climate characteristics. 20 cities from around Europe and 22 from across the USA were chosen. Two continents were chosen for data extraction for two reasons: first, to conduct the experiment in distinct climates with varying weather, and second, because of their geographic positions because the chosen cities were spread out far from one another. The experiment's main objective was to be impartial in terms of climatic type and geographic location. According to the findings, accumulated rainfall amounts can produce better predictions than those made using daily rainfall data. Support Vector Regression, Radial Basis Functions, and Genetic Programming all performed well when applied to the gathered data, however Radial Basis

Functions outperformed the contemporary "Markov chain" technique. Each technique used the same parameters for all of the chosen datasets, hence it was not certain that each technique employed the optimal set of parameters. A correlation between predicted accuracy and meteorological characteristics, such as the volatility of rainfall, the amount of maximum rainfall, and the interquartile range of rainfall, was discovered during the experiment. Additionally, there was no discernible difference in the algorithms' prediction error between the cities of the two continents (the USA and Europe). With the aid of accumulated rainfall amounts, a problem with the discontinuity in rainfall data was resolved.

Statistical Downscaling of Daily Rainfall Using a Hybrid Model
AUTHORS : Sahar Hadi Poura , Shamsuddin Shahida, Eun-Sung Chungb / 2016
The authors suggest a hybrid approach that combines random forest and vector machine to decrease daily precipitation. While SVM was chosen due to its suitability for non-linear data and was used to predict the amount of rain that will appear, RF was chosen due to its robustness in classification and was used to predict whether it would rain or not. At three stations on the east coast of Peninsular Malaysia, Dungun, Besut, and Kemaman, the proposed model was assessed with reduced daily rainfall. The Department of Irrigation and Drainage Malaysia provided the time series of daily rainfall from 1961 to 2000. The National Centre for Environmental Prediction provided a total of 26 climate
characteristics that were gathered and used as predictors for the recession model. To evaluate the variation in rainfall timing, a number of qualitative observations were made. Additionally, a
histogram is created for the database to show the issue. The average difference between two database segments that were discovered to be identical in three copies was calculated using the Student's t-test. The results show that, compared to RF and SVM models, the hybrid approach can minimise rainfall with a Nash-Sutcliff efficiency of 0.90 to 0.93.

IV. RESULT AND DISCUSSION

Performance testing is performed on the database. We collect rainfall data from January to December every year. This project has two main goals.
First, it recommends suitable plants for a particular soil, as well as this project suggests suitable soil for growing a particular crop. It will also advise that if the soil and crop are given, the crop can be planted on the given soil. If the plant is not suitable for the given soil, the result will indicate that the given plant is not recommended for the given soil and at the same time will suggest a suitable soil. The project has a database of different soils that are searched based on their ability to grow different crops.The second goal is to consistently forecast rainfall over a period of time. This project collects various data related to the amount of rainfall and predicts rainfall by analyzing data from previous records. This project measures rainfall in mm. The project should forecast rainfall in mm from 1949 to 2030.
Below are three steps. The first step asks you to select a soil type, the second step asks you to select a crop type, and the third step asks you to select a soil type and a crop type.

CULTIVATION FOR RED SOIL

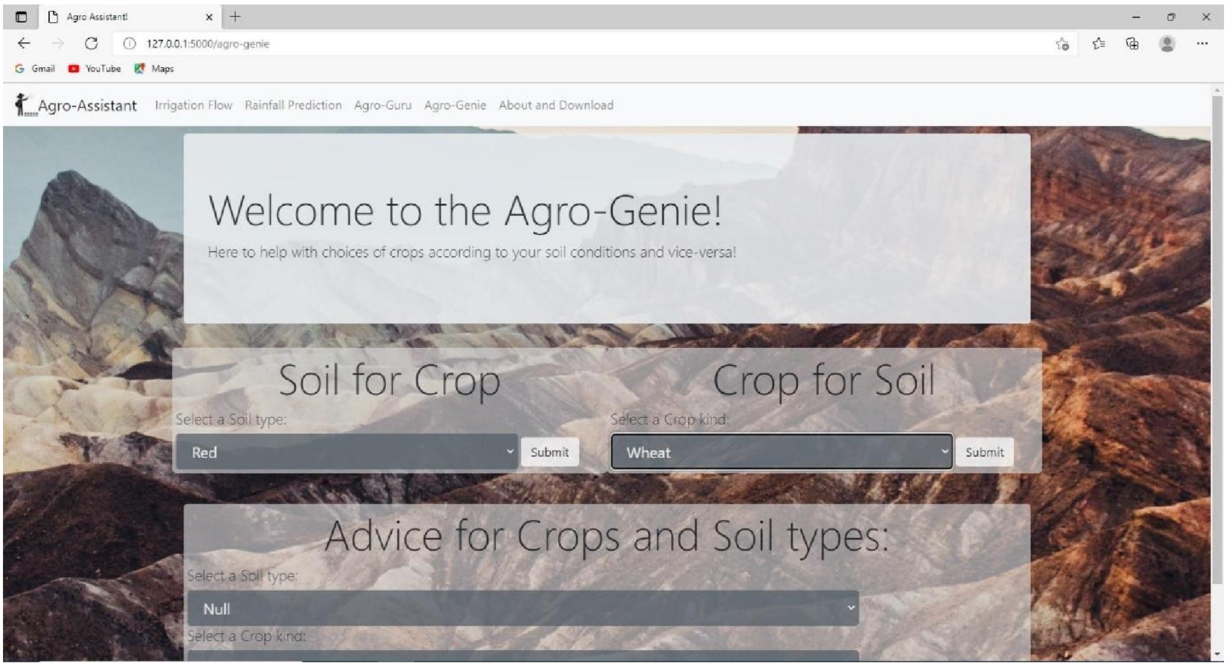

Figure 2

Figure 2 above defines the first task that asks for soil type (red) and will display the different plants you want to grow in a given soil type, as in Figure 3.

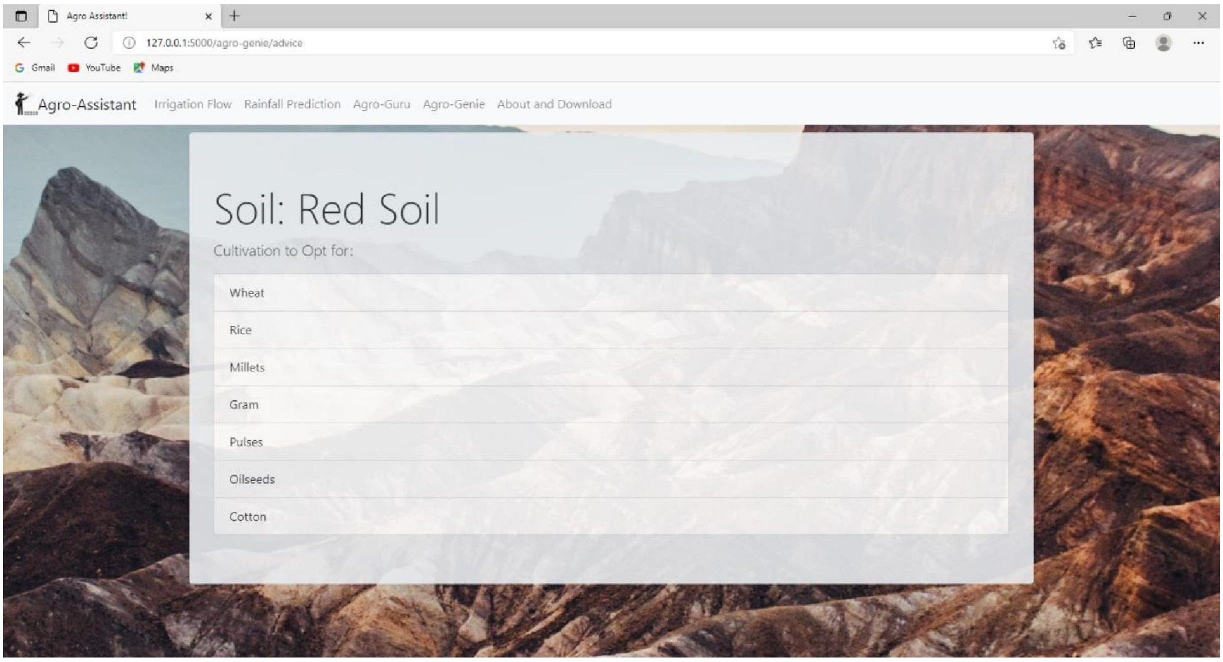

Figure 3

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 4, May 2023

SOIL FOR CROP

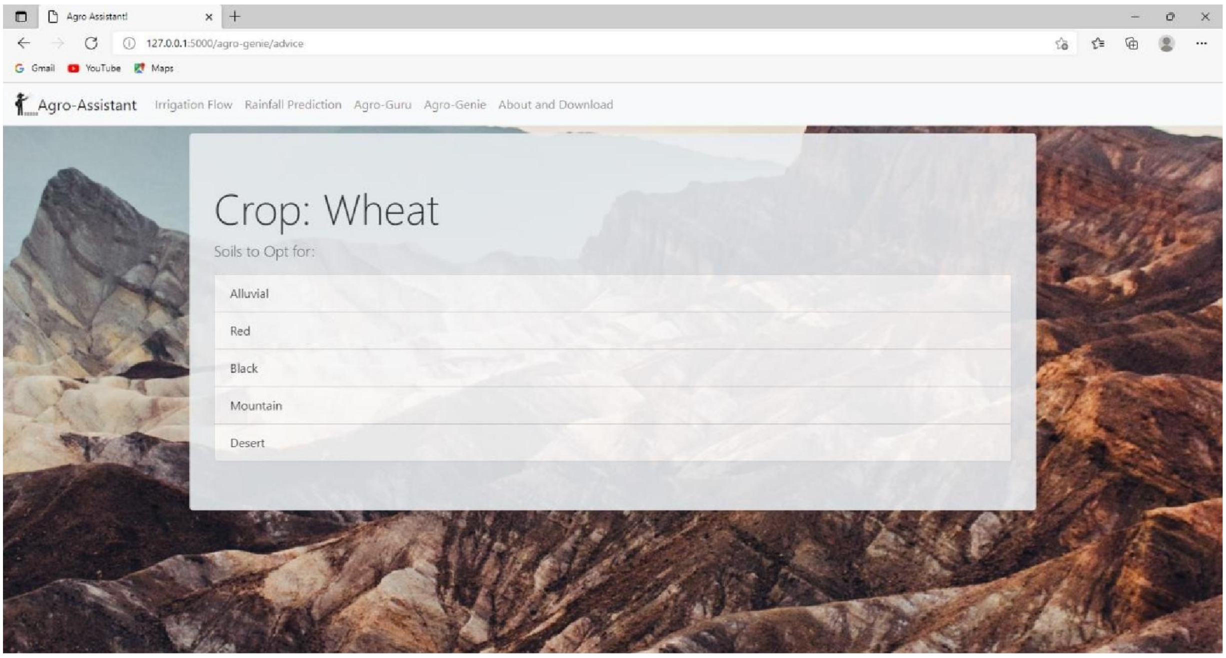

Figure 4

Figure 4 above defines a second application that asks for a type of crop (wheat) and will show different soils that will grow a certain crop.

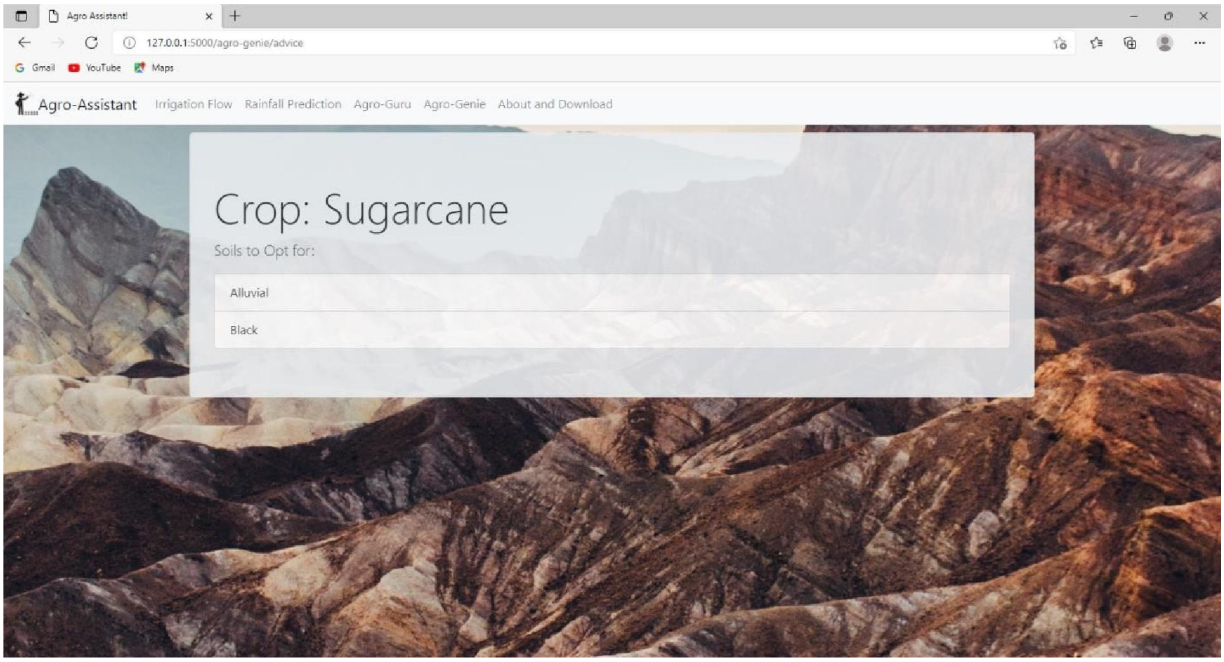

Figure 5

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

ADVICE FOR CROPS AND SOIL TYPES

Figure 8 shows the types of plants and soil. It refers to the third process that asks for the type of soil and yield, if the plant is planted accordingly in that soil, it indicates that the given plant is recommended for the given soil, if not it indicates that the given plant is not recommended for the given plant. soil and also provide the necessary information about non-recommended soil.

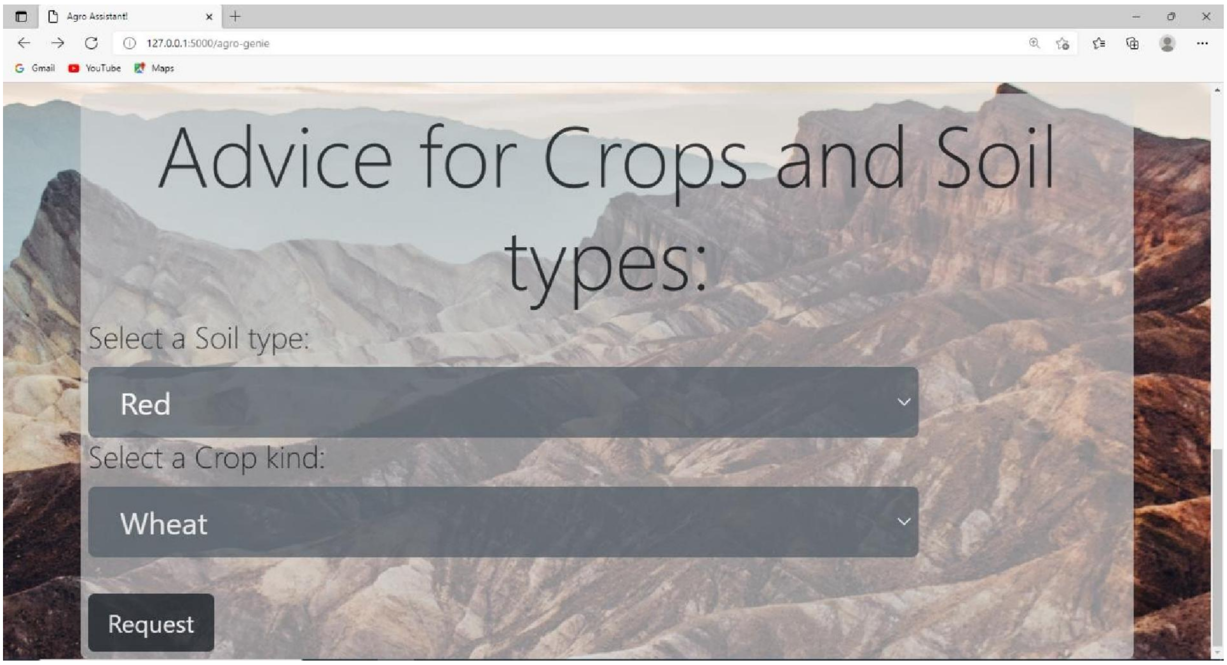

Figure 6

RECOMMENDED SOIL

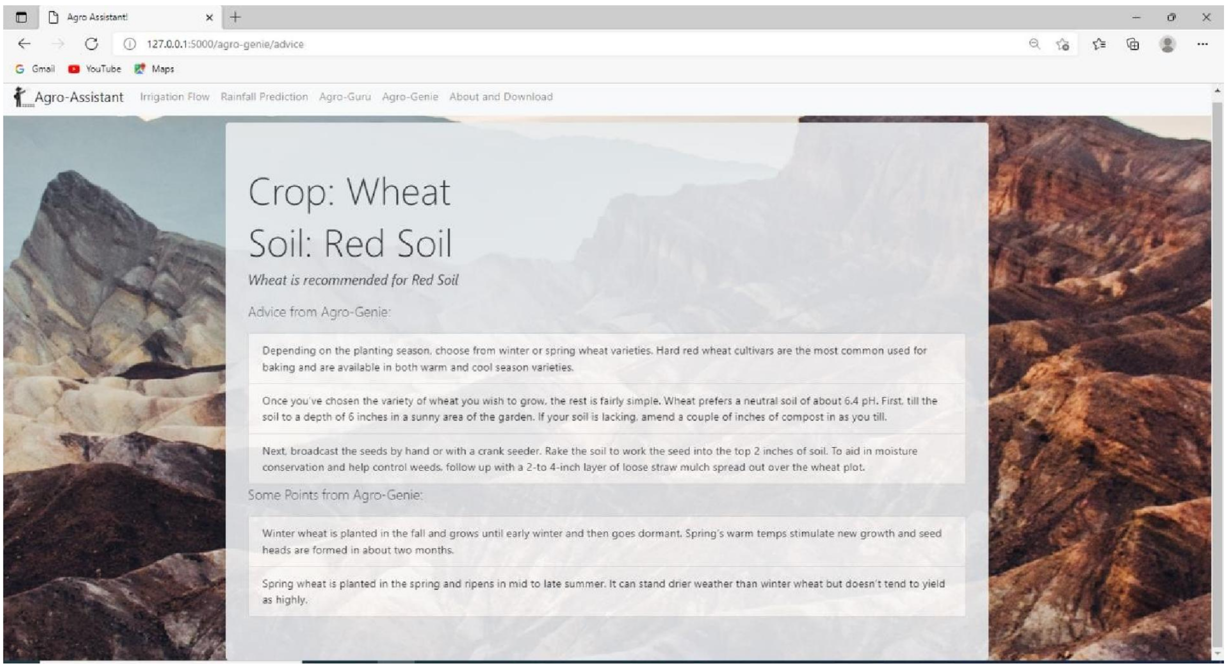

Figure 7

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

RAINFALL PREDICTION

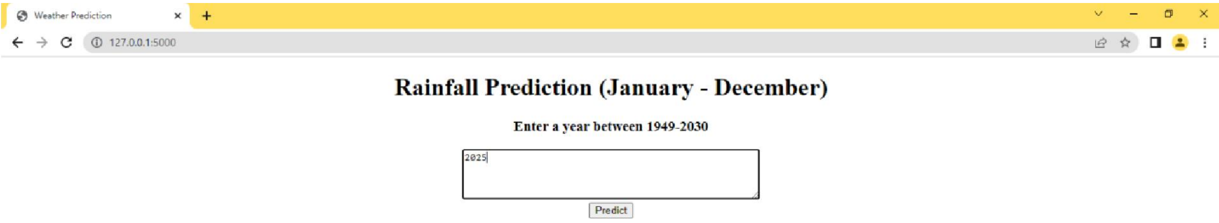

Figure 8

Figure 8 asks for valid years between 1949 and 2030. By using linear and polynomial regression algorithms and various databases collected from previous records. The output is then plotted for January to December as shown in Figure 9.

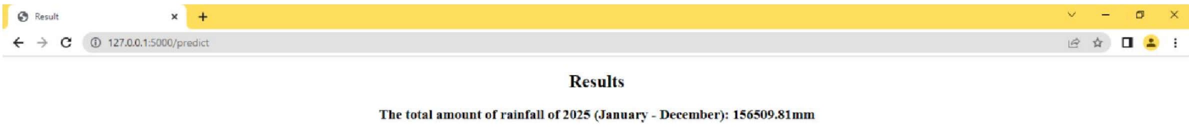

Figure 9

GRAPHICAL REPRESENTATION OF RAINFALL

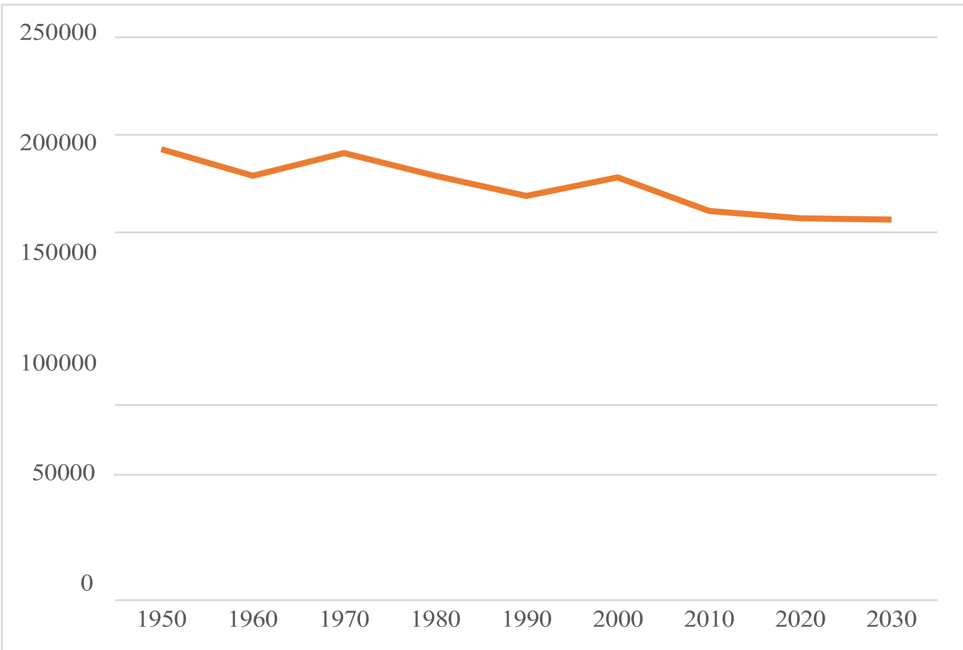

Figure 10. Rainfall graph from 1950 to 2030

Figure 10 shows a graphic representation of the 10-year variation of the above rainfall density in mm from 1950 to 2030.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

STATISTICS AND MODELS

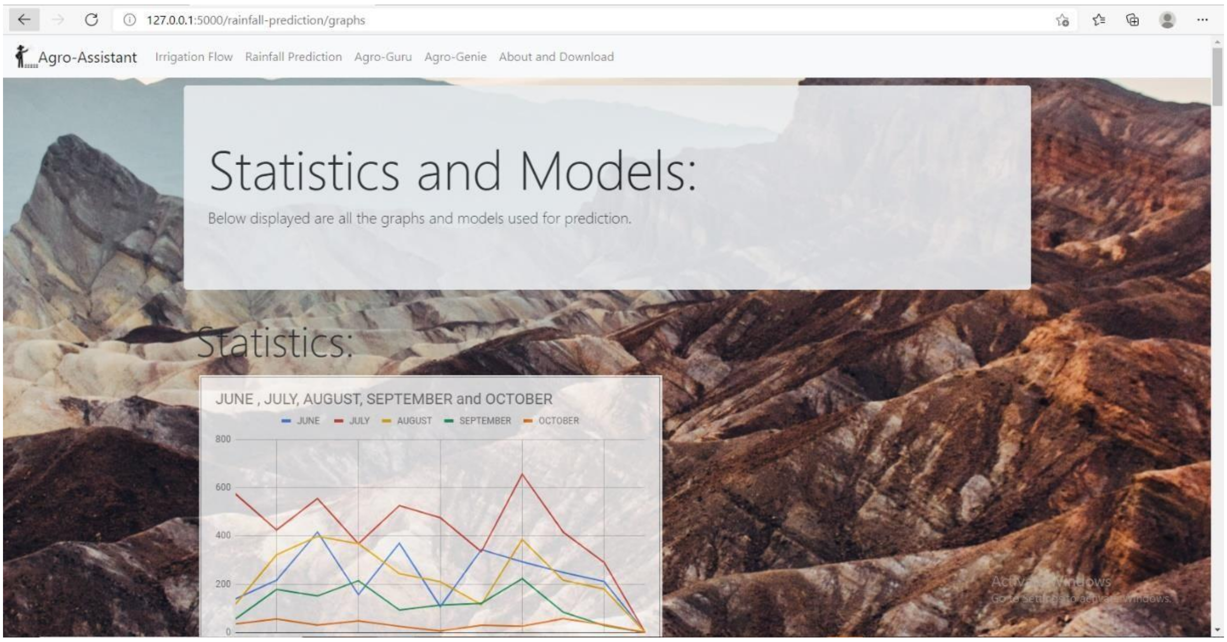

Figure 11

Fig 11 shows a graphical representation of rainfall in mm from June to October from 2010 to 2018

V. CONCLUSION

Weather forecasting is the business of meteorology, and it is easy to change the work of researchers using numerical weather forecasting techniques. The weather is mainly estimated using linear regression and simple polynomial statistical methods. The main goal for now is to improve classification and prediction performance; In this work, a climate data model was developed and developed. However, some limitations of the model are also noted, which should be reviewed before using the proposed method in the future. Also, there are some problems and challenges for a better application of Machine Learning methods in rainfall forecasting systems.

ACKNOWLEDGMENT

We extend our Sincere thanks to prof. Dr. Devasena M.E, PhD.
We also would like to thank our project guide prof. Esther Gandhimathi.

REFERENCES

[1]. Xiong, Lihua, and Kieran M. OConnor. "An empirical method to improve the prediction limits of the GLUE methodology in rainfall runoff 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.