

Maharashtra Rainfall Prediction using Machine Learning

Prof. T. P. Raju, Tejeswini Naidu, Herambh Jatadhar, Shraddha Manerao, Ashish Armarkar
Master of Computer Applications
Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, India

Abstract: Machine learning is the foundation of the attempt to predict rainfall. Here, our research will use historical data analysis to forecast rainfall. In this research, a genetic algorithm is utilized. Various researchers are doing a wide range of studies to forecast rainfall before it really occurs. The usage of genetic algorithms has been heavily utilized for a variety of research projects, and weather prediction is no different. With the use of genetic algorithms, we have created a system in this study that could anticipate rain. Genetic algorithm is used to extract features from the input data set. When compared to other approaches use of genetic algorithms seems to be more efficient for predicting the rainfall as hence it is used in the current work. The evaluation results are performed based on evaluating various parameters and the proposed model seems to provide a better efficiency when compared to the rest of the previous traditional rainfall prediction systems

Keywords: Rainfall Prediction, Genetic Algorithm, Crossover, Accuracy, Evaluation, SVM, Naïve bias, RBFN..

I. INTRODUCTION

In this project, we have put forth a genetic algorithm-based model for predicting rainfall using historical data. The correctness and effectiveness of the entire suggested model are also determined using the genetic algorithm. One of the most significant climatic changes in our country that needs to be predicted in advance is rainfall. Due to this intense rainfall, many people have died in several countries. The task of predicting rain is difficult, and the outcomes should be reliable. The use of weather variables including temperature, humidity, and pressure allows for the prediction of rainfall using a variety of hardware tools. SVM, Naïve bias, and RBFN monetary unit were classifiers employed in this research that were more appropriate than other applied mathematical and numerical techniques. For the purpose of predicting rainfall, we used a CSV data set of various national regions. carried out a comprehensive analysis to forecast the local area's rainfall.

II. LITERATURE REVIEW

For predicting rainfall in a specific geographic area, numerous studies have been put out by various researchers in their articles. The algorithm's performance is tracked using Matlab. The input dataset's features are extracted using a genetic algorithm. The incorrect area in the graph representation was then shown. Consequently, employing these dataset values. They are capable of making precise and reliable rainfall predictions.

Arima Model (Auto Regressive Integrated Moving Average)). For time series forecasting, analysis, and prediction, this model is employed. Box and Jenkins's proposal includes four ways. The ARIMA model goes through the four processes listed below. In general, sensors are used as input, such as wind sensors and light sensors. But in this case, convolution neural networks are being used to take input from historical data. The data is divided into training, testing, and calculating in this. From this NN training, the mean absolute error. Calculating the number of iterations and determining the ideal Epoch value. Using data from the Multiple Linear Regression (MLR) for Indian dataset, this research suggests a model for predicting rainfall. Multiple meteorological variables are included in the input data to help estimate rainfall more accurately. The parameters used to validate the suggested model are the Mean Square Error (MSE), accuracy, and correlation. outcomes in the suggested MLR-based technique for predicting rain precipitation. The chosen data set contains 4116 total data points. The MLR prediction's accuracy is 0.99.

When the criteria for a heavy rain advisory are anticipated to be met within the next three hours, the EWS sends out a warning signal. We developed a method of selective discretization that changes a portion of continuous input variables to nominal ones. The PCA reduced the dimensionality of input variables by developing a new 370 coordinate system that offers an insightful perspective of the data, and the selective 368 discretization approach selectively discretized input variables that have a nonlinear relationship with the extremely shortterm heavy rainfall. The application of this approach to the forecasting of the various rainfall intensity series gathered by 135 stations dispersed throughout the Tiber basin yielded some interesting findings, which were discussed in this work. SVM: There are other machine learning algorithms used for classification, but SVM is superior to most of the others since it produces results with higher accuracy. Support Vector Machine Algorithm is more accurate and provides a faster forecast than other classification methods. SVM is a simple classifier technique that is very effective and frequently used for pattern identification. Additionally, it may perform classification better than any other classifier. The primary goal of an SVM classifier in the SVM approach is to identify the function $f(x)$, which determines the decision boundary or hyper plane. Both Approaches C and D produce effective outcomes.

III. ADVANTAGES AND DISADVANTAGES

3.1 Advantages

- The forecasting of rainfall is under the supervision of modern technology.
- The ML program could result in significant time savings. And in today's world, everyone could use more energy.
- Monitoring rainfall forecast successfully with ML.
- We can program the system to irrigate a portion of efficient rainfall.

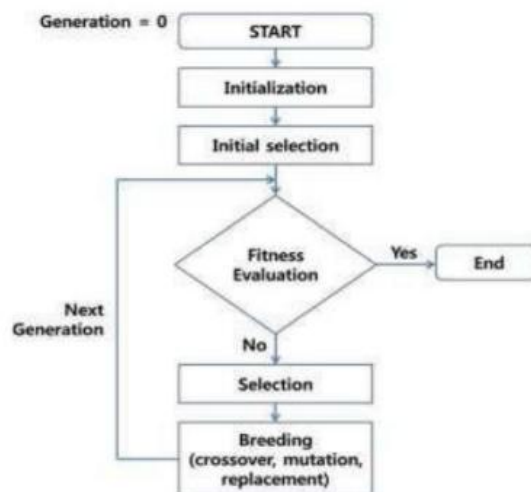
3.2 Disadvantages

- Weather is exceedingly challenging to predict accurately.
- Monitoring so many variables from so many different sources is expensive.
- The cost of the computers required to do the requisite millions of calculations is high.
- If the weather does not match the forecast, the weather forecasters are held responsible.
- The price of upkeep is expensive.

IV. METHODOLOGY

The process of machine learning involves examining the features that are extracted, then testing the URL with the classifier. We utilized the UCI Phishing Website Dataset to train the classifier.

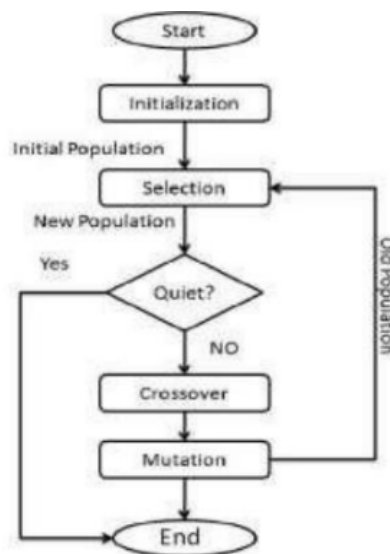
4.1 Genetic Algorithm



The evolutionary algorithm, which is heavily employed in feature selection, is shown in a block diagram. The chosen feature is combined with the feature from the input dataset. Three modules make up the model. The first is the preprocessing module, which takes the input and converts it for feature extraction. The Genetic Algorithm is the second. The provided input dataset is subjected to data processing. We have a genetic algorithm flow. The values from the rainfall dataset were predicted using this technique. for selecting the best from the region or space of choice.

4.1.2 Module 1: Genetic flow

Using the selection, crossover, and mutation operations, it will assign the fitness. The next prediction will be made if the value is true. If it is false, the three operator will once more be used to assign the fitness. Each of these three operators has a special operation. The rainfall dataset values were predicted using this technique. for selecting the best from the chosen location or region. It made it possible for our suggested system to anticipate values more quickly and provide the best performance validation.



4.1.2 Module 2: Dataset

For our investigation, we used the CSV dataset for India. Navy bias, RBFN Classifiers, and SVM. Machine learning algorithm SVM is a supervised that may be applied to classification or regression issues. It transforms your data using a method known as the kernel trick, and based on these modifications, it determines the best border between the potential outputs. Navy prejudice A similar approach is used by Naive Bayes to forecast the likelihood of various classes based on various attributes. The main use for this approach is text classification, which has issues with multiple class problems. By assessing the input's similarity to examples from the training set, an RBFN conducts classification.

An example from the training set is stored as a "prototype" in each RBFN neuron. Each neuron calculates the Euclidean distance between the fresh input and its prototype when we want to categorize it. Project goals versus project objectives genetic programming This algorithm mimics the natural selection process, in which the most physically fit individuals are chosen to reproduce in order to give rise to the next generation of humans. Classifiers- Svm In SVM, we take the output of the linear function and, if it is larger than 1, we classify it into one category, and, if it is less than 1, we classify it into a different category. We achieve this reinforcement range of values $[-1, 1]$ since the threshold values are modified to 1 and -1 in SVM. Which acts as margin Navy prejudice The classifier known as Naive Bayes makes use of the Bayes Theorem.

It forecasts the likelihood that a given record or data point belongs to a certain class for each class, for example. The most likely class is the one with the highest likelihood. Traditionally, Gaussian functions are used to implement RBFN Activation functions in RBFNs.

V. SYSTEM CONFIGURATION

Requirements for analysis

- Algorithm
- Classifiers
- Dataset

Software Requirement

- Operating System – Windows 7, Windows 8, Android
- Languages – Python
- Compiler
- Browser – Chrome, Firefox, Opera, UC-Browser, etc.
- Database Requirement
- CSV

VI. CONCLUSION

With the development of the modern period, several technologies are expanding. The field of weather prediction is rapidly expanding in the upcoming technological era. The use of genetic algorithms is essential for categorization approaches as well as for accurately forecasting the weather for a certain region or country during a given time period. In order to prevent the loss of life and property damage, we have suggested an effective approach of employing a genetic algorithm to predict rainfall in a specific area well in advance. The model suggested using a genetic algorithm pulls the features from the dataset that is provided. All of the previously available models appear to perform worse than the proposed model. Future work might involve the usage of any security protocols to protect the system as a whole.

REFERENCES

- [1]. <https://www.geeksforgeeks.org/ml-rainfall-prediction-using-linear-regression/>
- [2]. <https://www.geeksforgeeks.org/rainfall-prediction-using-machine-learning-python/>
- [3]. <https://towardsdatascience.com/predicting-rain-with-machine-learning-2acf80017c44>