IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

RAINFALL PREDICTION USING AI AND ML

Prakul Singla, Harshit Jain, Yashashvi Kumar
Department of Computer Science and Engineering
SRM Institute of Science and Technology, Modinagar, Delhi NCR

Abstract:
Agriculture is the key point for survival. For agriculture, rainfall is most important. These days rainfall prediction has become a major problem. Prediction of rainfall gives awareness to people and they can get to know in advance about rainfall to take certain precautions to protect their crop from rainfall. Machine Learning algorithms are mostly useful in predicting rainfall. Some of the major Machine Learning algorithms are ARIMA Model(Auto-Regressive Integrated Moving Average), Artificial Neural Network, Logistic Regression, Support Vector Machine and Self Organizing Map. Two commonly used models predict seasonal rainfall such as Linear and Non-Linear models. The first models are the ARIMA Model. While using Artificial Neural Network(ANN) predicting rainfall can be done using Back Propagation NN, Cascade NN or Layer Recurrent Network. Artificial NN is the same as Biological Neural Networks. We will use a kaggle dataset to train and test a Decision Tree Classifier, Logistic Regression & Random Forest Classifier model to predict whether there is going to be rainfall tomorrow or not. We will split the dataset into training and testing data and will be using the training dataset to train the model and testing dataset to observe the accuracy of the model.

Keywords: Linear Regression, Decision Tree Classifier, Random Forest Classifier.

I. INTRODUCTION

Rainfall forecasting is very important because heavy and irregular rain can cause many impacts, such as crop and farm destruction, property damage. Therefore, a better predictive model is needed for early warning, which can reduce risks to life and property, and help manage agricultural farms in a better way. This prediction mainly helps farmers, and water resources can also be used efficiently. Predicting precipitation is a difficult task and the results must be accurate. There are many hardware devices for predicting precipitation using weather conditions such as temperature, humidity, and pressure. These traditional methods cannot work effectively, so by using machine learning techniques, we are able to produce accurate results. We can do this simply by analyzing historical rainfall data and can predict rainfall for future seasons. We can apply many techniques like classification, regression based on requirements, and we can also calculate the error between reality and prediction as well as the accuracy. Different techniques produce different accuracy, so it is important to choose the right algorithm and model it as required.

In this chapter, we gave a brief introduction to machine learning, the different algorithms used, and how it can be useful in predicting precipitation using different parameters like weather, time, interval way, etc. We used different machine learning algorithms like decision tree, random forest classifier, logistic regression, etc. to train the system, then found the most accurate rainfall prediction algorithm, which was then used to forecast the rainfall. Precipitation depends on parameters like temperature, pressure, wind speed, humidity, visibility, etc. We also try to find correlations between different columns of the data set.

II. LITERATURE SURVEY

The ability to anticipate rainfall has long been a topic of study, and the techniques used to do so have developed over time. In the past, people used natural indicators, like changes in animal behavior and the appearance of clouds, to forecast rainfall. These techniques were ineffective because they mainly relied on observation and intuition. However, they were useful for survival and rudimentary agricultural planning. In the nineteenth century, scientists started to create increasingly complex techniques for predicting rainfall. In the early 20th century, technological developments helped to develop systems for rainfall prediction that were more precise. A. Kala and Dr. S. Ganesh Vaidyanathan has used the Artificial Neural Network for implementing their idea. Their procedure for rainfall prediction includes gathering the weather data, then preprocessing it, with training data and then validating it with testing data and in the end evaluating the model by comparing desired and actual output [1]. There are various algorithms used for the prediction of rainfall. Fundamentally, there are two approaches to predict Rainfall. They are Empirical and Dynamic methods. The Empirical approach is based on the Analysis of past data of weather and its relationship to different atmospheric variables. In a dynamical approach, predictions are generated based on numerical methods, like using mathematical equations. They have surveyed different algorithms that are helpful for rainfall prediction like, Linear Regression, AutoRegressive Integrated Moving Average [ARIMA]. These are some of the algorithms that are reviewed for their applicability to predicting rainfall [2]. There was 0.935 accuracy using the Artificial Neural Network. R. Kingsy Grace and B. Suganya in their paper have said that they have implemented their model using machine learning. They have compared various models such as Deep Convolutional Neural Network, ANN, Linear Regression, There was 99% accuracy of the Multiple Linear Regression [3]. The authors have selected a supervised machine learning algorithm, for rainfall prediction. The algorithm analyzes the previous data related to temperature, region, area, and year. It takes historical data as input and output as a result. The algorithm has proved to simplify and reduce manual work and also provides smooth workflow [4]. The authors have made their analysis using the correlation data and then prediction using the Linear Regression model. The model has selected parameters like precipitation, average temperature, vapor pressure [5]. They used machine learning techniques to implement this model. Their methodology includes data exploration and analysis, data preprocessing, modeling and evaluating. In the end they concluded that Australian rainfall is not certain there is not a particular relation between time and shower. Still, they are able to find certain patterns and develop high performance models [6].

III. SYSTEM ANALYSIS

1. We'll talk about the several experiments conducted in this chapter in an effort to determine the most reliable method for forecasting rainfall. We'll talk about the issue at hand as well as the solution we're developing to address it. The various machine learning algorithms, including decision trees and random forest classifiers, will be addressed. We'll examine the model to see how various factors—including temperature differences, wind speeds, humidity levels, etc.—affect rainfall.

2. EXISTING PROBLEM

Forecasting rainfall is a highly sought-after service since it is essential to many different industries and uses. Planning and management of the agricultural, transportation, energy, and water industries depend on accurate rainfall forecasts. The present issue with our project is figuring out how to compute the precision of precipitation levels based on numerous parameters.

3. PROPOSED SOLUTION

For the purpose of developing an accurate model for forecasting future rainfall, the rainfall prediction system will examine historical weather data as well as other pertinent information. To evaluate the model's accuracy, a number of methods including decision trees, random forest classifiers, and logistic regression models will be used. Future rainfall predictions will be made using the most precise algorithm.

IV. METHODOLOGIES

This chapter will cover the procedures for providing accurate input data, including how to gather information from numerous sources that offer real-time data, Explore the data set to comprehend the

parameters that make up the data set, then clear it to see how it divides and is utilized to train our prediction model.

1. DATA COLLECTION

Data is gathered from the open source Kaggle platform, which is used to obtain data for machine learning and artificial intelligence. Kaggle is also free to use. The years 2008 through 2017 are represented in this data set.

2. DATA EXPLORATION

We must have a clear understanding of the information we have gathered. We must tally the total number of days that the data is gathered. include the year and month to indicate which state the data belongs to, The amount of rainfall in millimeters for the specific state is shown in a separate column called Annual Precipitation.

3. DATA CLEANING

The data must first be filtered, so we will look for any rows that have missing values before removing them if they won't significantly affect the forecast. If not, we will try to fill the gaps with the mean or median values, whichever is less significant. The dataset's empty values will be swapped out for zeros. Then, we will remove all the columns that are not necessary.

4. MODEL TRAINING

In this case, after we train the model and validate it using the first filtered data, we will try to make the model and make predictions using a more optimized data set. To do this, we will further exclude non-affecting data, such as the name of the city or state for a specific state, since it will be repeated endlessly without adding any additional information, making it unnecessary.

5. TESTING

The models created in the previous step will be tested using historical test data to calculate precision. In our project, we will use that model whose accuracy is the highest.

V. RESULTS

1. RESULTS OBTAINED FROM LOGISTIC REGRESSION MODEL

```
[[16694   785]
 [ 2661  2445]]
              precision    recall  f1-score   support

           0       0.86      0.96      0.91     17479
           1       0.76      0.48      0.59      5106

    accuracy                           0.85     22585
   macro avg       0.81      0.72      0.75     22585
weighted avg       0.84      0.85      0.83     22585

0.8474208545494798
```

2. RESULTS OBTAINED FROM DECISION TREE CLASSIFIER MODEL

```
[[14930  2549]
 [ 2403  2703]]
              precision    recall  f1-score   support

           0       0.86      0.85      0.86     17479
           1       0.51      0.53      0.52      5106

    accuracy                           0.78     22585
   macro avg       0.69      0.69      0.69     22585
weighted avg       0.78      0.78      0.78     22585

0.780739428824441
```

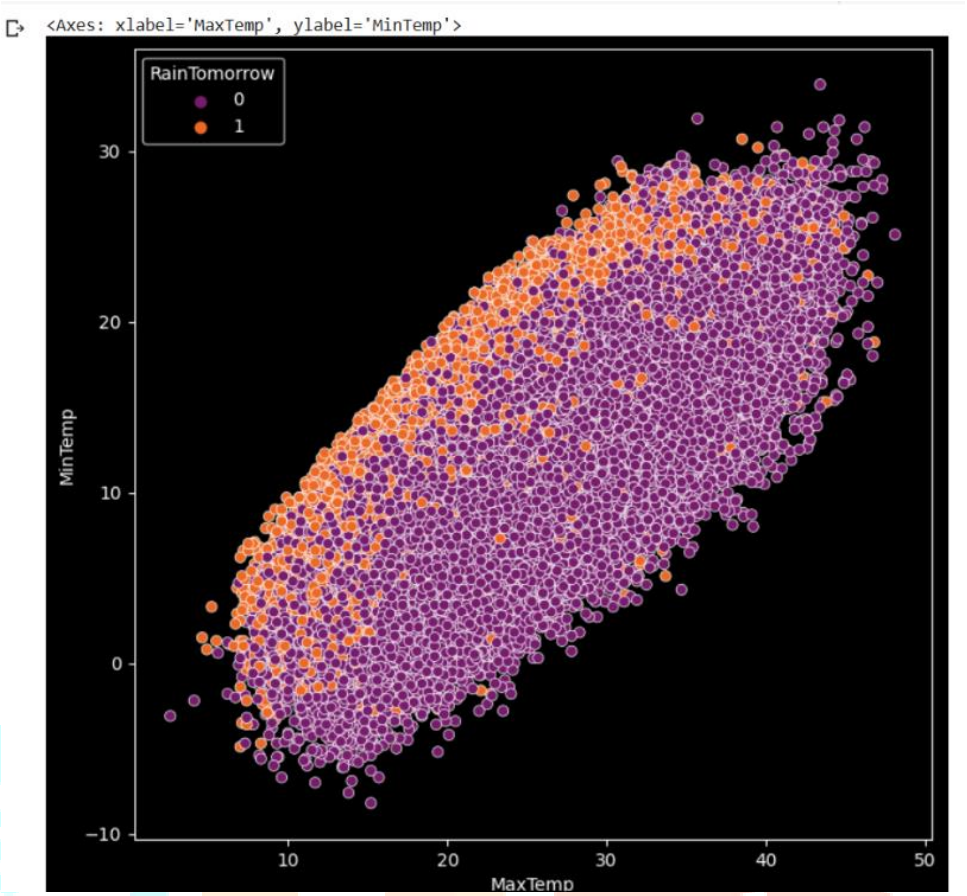
3. RESULTS OBTAINED FROM RANDOM FOREST CLASSIFIER MODEL

```
[[16666   813]
 [ 2456  2650]]
              precision    recall  f1-score   support

           0       0.87      0.95      0.91     17479
           1       0.77      0.52      0.62      5106

    accuracy                           0.86     22585
   macro avg       0.82      0.74      0.76     22585
weighted avg       0.85      0.86      0.84     22585

0.8552579145450521
```

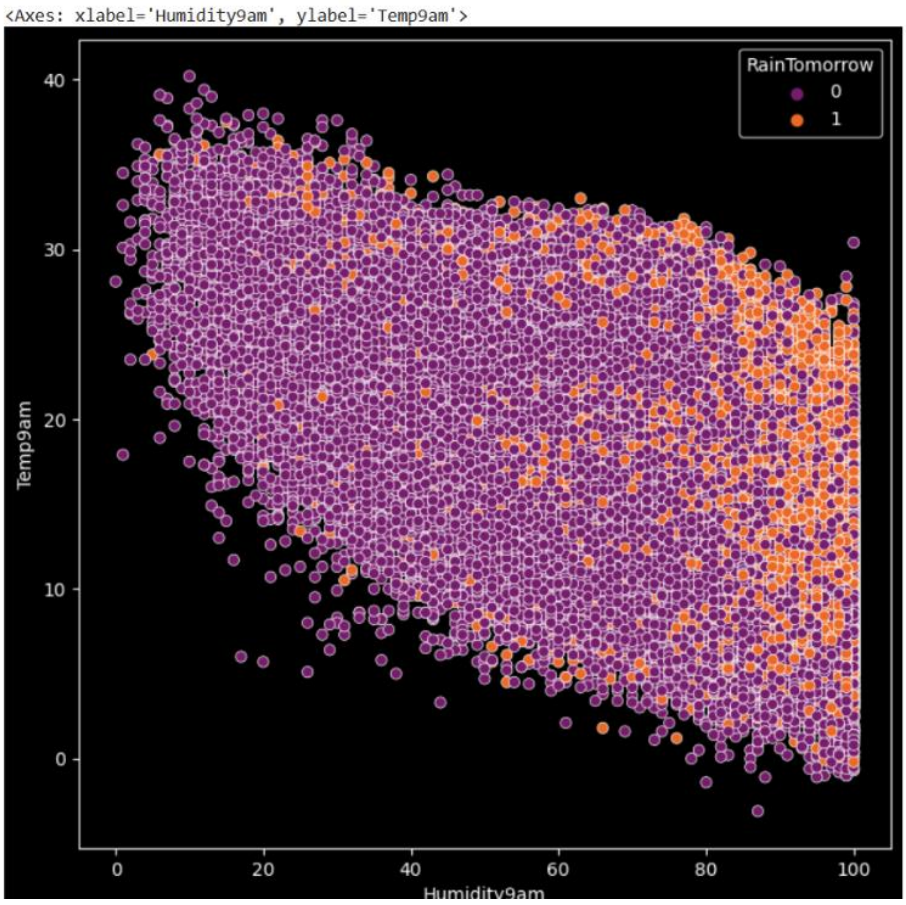
4. RESULTS OBTAINED FROM XGBOOST MODEL

```
acc 0.8572946646003985
f1              precision    recall  f1-score   support

           0       0.88      0.95      0.91     17479
           1       0.75      0.55      0.64      5106

    accuracy                           0.86     22585
   macro avg       0.81      0.75      0.77     22585
weighted avg       0.85      0.86      0.85     22585

matrix [[16542   937]
 [ 2286  2820]]
```

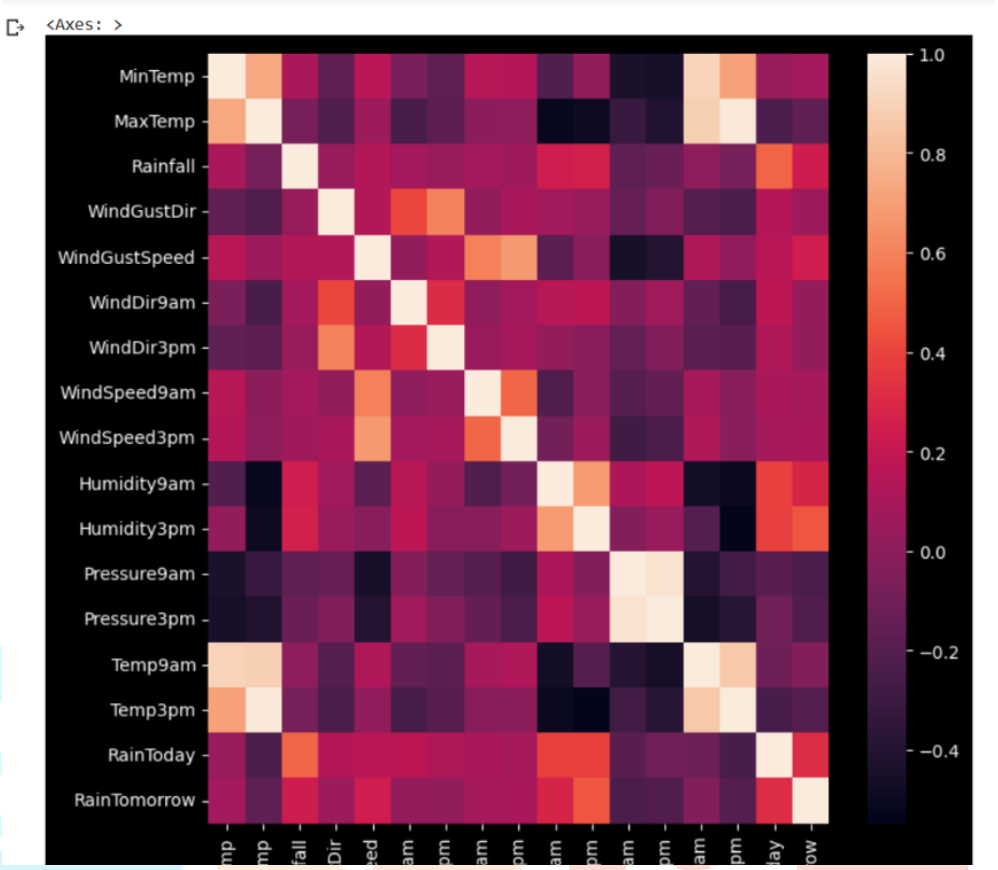
5. GRAPH FOR PRECIPITATION W.R.T. MIN & MAX TEMPERATURE



6. GRAPH FOR PRECIPITATION W.R.T. TEMPERATURE AND HUMIDITY



7. HEATMAP



VI. CONCLUSION AND FUTURE SCOPE

1. CONCLUSION

The project's goal is to choose a specific algorithm to forecast rainfall, considering the variables that influence it. It has been demonstrated that the Random Forest Classifier Model can be a flexible prediction method. In this research, we looked at different algorithms and how they responded to the target variable's varied inputs. Instead of using conventional models, machine learning may provide us with smarter models. Additionally, less computing power and manual labor are needed. Since regression is the most effective method of prediction, we have looked at the logistic regression model. We have gained knowledge of several preprocessing methods that are necessary to get the dataset ready. The dataset must be devoid of any noise, consistency issues, overfitting, or other irregularities that could compromise the accuracy of the model. Additionally, we looked into a few regression algorithms that can adapt their predictions to changes in the input data. We now have a thorough understanding of numerous algorithms thanks to the comparison study.

2. FUTURE SCOPE

The future enhancements we can apply to our project to improve its performance and accuracy will be covered in this chapter. For rainfall prediction, we have only suggested one technique, and it has both advantages and disadvantages compared to other algorithms. Even so, we were unable to predict rainfall at various locations with a 100% accuracy rate. In order to accurately predict when it will rain, a mix of algorithms must be constructed and applied in real time. Due to the dynamic nature of rainfall factors and how many distinct real-world scenarios they are affected by, a model that has superior data pretreatment and optimization techniques may yield better results. In the future, there will be additional

REFERENCES

1. Kala, A., & Vaidyanathan, S. G. (2018, July). Prediction of rainfall using artificial neural networks. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA).

2. Mr. Dhawal Hirani, Dr. Nitin Mishra, “A survey on Rainfall Prediction Techniques”, International Journal of Computer Application (IJCA)".

3. Grace, R. K., & Suganya, B. (2020, March). Machine Learning based Rainfall Prediction. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS).

4. Mrs. P. Chaya, I. Kavya, S. Likitha, M. Tanushri, C. Roshni Poovanna, "A Prognostic Rainfall Using Machine Learning Technique", International Journal For Research In Applied Science & Engineering Technology (IJRASET).
www.ijraset.co

5. Nikhil Sethi, Dr. Kanwal Garg, “Exploring Data Mining Technique for Rainfall Prediction”.

6. Oswal, N. (2021). Predicting Rainfall using Machine Learning Techniques. Institute of Electrical and Electronics Engineers (IEEE).

7. The dataset used has been taken from Kaggle Url-https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-packag

8. Geeksforgeeks - For information about models. https://www.geeksforgeeks.org/rainfall-prediction-using-machine-learning-python/