Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

20

RainFall Prediction System for Mumbai

Aditya Nikhade, Rahul Khetale

Dr. VISHWANATH KARAD

MIT-WORLD PEACE UNIVERSITY

ABSTRACT

These days, climate change is accelerating due to global warming, which has a major influence on humanity. Sea levels are rising, the atmosphere and ocean are warming, and there are more floods and droughts as a result. Uneven rainfall or precipitation is one of the main effects of it. Today, most of the important global authorities are taking into consideration the laborious problem of precipitation forecasting. One climatic factor that has an impact on the many human activities is precipitation. like manufacturing, production, and tourism in the agricultural sector. Rainfall becomes extremely problematic as a result, necessitating more accurate forecasts. Accurate rainfall forecasting is crucial for all of these reasons. There are several ways to forecast it, but the one that is chosen for the objective of this assignment is to analyze and compile rainfall data from the past 12 months, gathered over a period of 5 years. The goal is to utilize this data to forecast rainfall for the following day. To achieve this, the project aims to optimize the results by employing a random forest classifier as a machine learning model for predicting rainfall.

Keywords: Accuracy, Forecasting, Machine Learning Algorithms, Rainfall, random forest classifier

1. INTRODUCTION

Forecasting rainfall is a difficult and Complex issue that has a big impact on human society since reliable predictions can lessen the number of people and money lost due to natural disasters like floods and droughts. Although machine learning and artificial intelligence (AI) techniques have been shown to perform better than conventional statistical methods in terms of accuracy, heavy rain forecasts continue to be a challenge for Meteorological Bureaus

around the world, particularly in nations lacking in the necessary technology. As a result, this project aims to do comparative studies on machine learning methods, make precipitation prediction techniques easily accessible to non-experts, and give early precipitation forecasts and statistical data on rainfall patterns in Mumbai.

Precautionary actions that can be considerably lessen the effects of sudden and heavy rainfall. Accurate rainfall forecasting is important in many ways. Yet, precise rainfall forecasting has proven to be difficult due to climatic changes. The accuracy of rainfall forecasting could be increased by using machine learning techniques to discover hidden patterns from previous data on meteorological characteristics.

In order to provide irrefutable statistics and deliver helpful information for agriculture, health, and drinking purposes, the system analyzes rainfall data from all locations. Predictions must be accurate,

2. MATERIAL AND METHODS

Dataset Preparation:

The purpose of this inquiry is to forecast Mumbai's typical rainfall. Data from India's Ministry of Earth Sciences and Meteorology Department show that rainfall in Maharashtra increased by 85%.

Data on rainfall for the years 2016 to 2020 were taken from the Meteorological Department of India's Annual Climate Report report.

It is planned to use daily rainfall data from 2016 to 2020 to examine average rainfall patterns and analyze trend variations. To eliminate noisy records, the dataset is pre-processed.

Method:

A well-liked machine learning technique called the random forest classifier is frequently utilised in many different study fields, including bioinformatics, finance, healthcare, and social sciences. In classification problems where the objective is to predict the class labels of fresh observations based on a collection of input features, random forest classifiers are frequently utilised in academic articles.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Since they can handle highly dimensional and complicated data and are robust to noise and outliers, random forest classifiers are frequently used in research articles. Also, the random forest classifier is renowned for its capacity to offer rankings of key features, which can assist researchers in locating the dataset's most pertinent variables.

It is crucial to choose the algorithm's hyperparameters—such as the number of trees, the maximum depth of each tree, and the amount of features taken into account at each split—carefully when utilizing random forest classifiers in research. To make sure that the model generalizes well to new data, do this using cross-validation or other model selection techniques.

Overall, the random forest classifier is an effective tool for classification problems in research, and due to its adaptability and reliability, it is a preferred choice for many applications.

3. DATA PREPROCESSING:

1. Formatting: Three approaches are typically included in data pre-processing: formatting, cleaning, and sampling. The process of transforming data into a readable format is referred to as formatting. This could entail transferring data from a non-relational database to a database system or from an unique file format to a plain text file.
2. Cleansing: Cleaning involves locating and resolving any incomplete or missing data. If data instances don't have the necessary data, it could be essential to eliminate them in specific circumstances.
3. Sampling: In addition, some attributes might include confidential data that needs to be erased or anonymized. Instead of using the complete dataset for analysis, sampling entails choosing a representative selection of data. This can make it simpler to assess and test solutions and reduce the computational and memory requirements.

5 FEATURE EXTRACTION:

A method for reducing the amount of attributes in a dataset is feature extraction. Feature extraction entails changing the attributes themselves, as opposed to feature selection, which ranks existing attributes according to their predictive relevance. Usually, linear combinations of the original qualities make up the converted attributes.

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

The Classify module of Python's Natural Language Toolkit library's classifier algorithm is used to train our models after preparing the data. We gather a named dataset and divide it into training and testing sets. The performance of the models is assessed using the testing set. We used a variety of machine learning methods, including Random Forests, which are frequently employed in text classification applications, to categorise the pre-processed data.

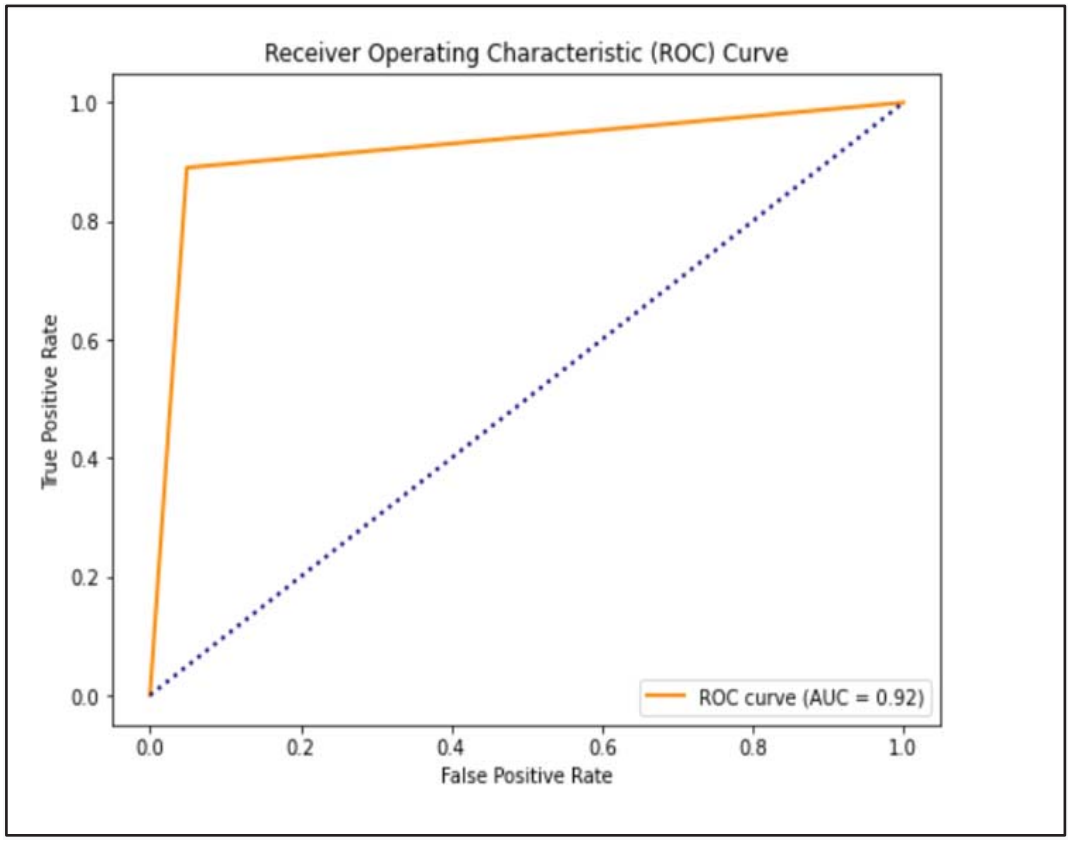
6. APPLYING ALGORITHMS:

Using training data that were divided in the previous module to train the test data, a random forest technique is employed in this module. With training data, a forecast or fraud percentage will be obtained as an output. Using this fraud, percent can develop a confusion matrix that will display the number of fraudulent and legitimate transactions. This module will employ a different algorithm in an effort to increase the accuracy of fraud detection by increasing the percentage of fraud predictions. The better fraud accuracy method will be used based on the dataset once both algorithms have been compared. We used the random forest technique and the neural networks algorithm in this project.

7.RESULTS AND DISCUSSIONS:

	temp	dew	humidity	sealevelpressure	winddir	solarradiation	windspeed	class
count	1781.000000	1781.000000	1781.000000	1781.000000	1781.000000	1781.000000	1781.000000	1781.000000
mean	28.342560	21.644975	69.565974	1008.878046	201.304267	228.169175	22.309882	0.456485
std	1.960027	4.532065	14.533809	3.706157	47.075559	57.034865	6.360669	0.498243
min	20.200000	5.000000	28.600000	994.100000	65.600000	52.800000	9.400000	0.000000
25%	27.200000	18.100000	58.000000	1006.200000	163.700000	194.600000	18.400000	0.000000
50%	28.500000	23.800000	71.600000	1009.100000	204.500000	232.300000	22.300000	0.000000
75%	29.700000	25.300000	81.200000	1011.800000	240.900000	274.400000	24.600000	1.000000
max	32.800000	27.500000	98.200000	1017.400000	316.000000	330.900000	128.100000	1.000000

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar




```
accuracy = accuracy_score(y_test, y_pred)
print(r"Accuracy:", {accuracy})
```

```
Accuracy: 0.9215686274509803
```

```
print(r'confusion matrix =',confusion_matrix(y_test, y_pred))
```

```
[[175   9]
 [ 19 154]]
```

```
print(r"ROC-AUC Score:", {roc_auc})
```

```
ROC-AUC Score: 0.9206301834631818
```

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.90	0.95	0.93	184
1	0.94	0.89	0.92	173
accuracy			0.92	357
macro avg	0.92	0.92	0.92	357
weighted avg	0.92	0.92	0.92	357

8.CONCLUSION:

Rainfall forecasting in Mumbai using the Random Forest Classifier Algorithm. The five years of historical weather data from January 1, 2016, to November 15, 2020, are used in this study to make predictions. Results of the Random Forest Classifier technique are presented in tables and graphs. In order to make accurate predictions, a classification system is utilised, where the input data is cleaned and standardised before categorization. Based on the trained data, this model predicts rainfall with a 92% accuracy rate. For the prediction, nine parameters were taken into account. Due to missing values and the absence of important climatic characteristics, some elements have not been taken into account. It is proposed that additional procedures be used for future work because carried out taking into account more variables with more precise data and climatic characteristics on various weather dates. From this point forward, accuracy is predicted using Random Forest.

REFERENCES:

1. L. Houthuys, Z. Karevan and J.A. Suykens, “Multi-view LSSVM regression for black-box temperature prediction in weather forecasting”, International Joint Conference on Neural Networks, pp. 1102-1108, May 2017, IEEE
2. Ali Haidar and Brijesh Verma. "Monthly rainfall forecasting using a one-dimensional deep convolutional neural network." IEEE Access 6, pp. 69053-69063, Nov 2018

3. S. Manandhar, Y.H Lee and S. Dev," GPS derived PWV for rainfall monitoring", IEEE International Geoscience and Remote Sensing Symposium, pp. 2170-2173, Jul 2016. IEEE.
4. S. Chatterjee, B. Datta, S. Sen, N. Dey and N.C Debnath," Rainfall prediction using hybrid neural network approach",2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing, pp. 67-72. IEEE
5. S. Dev, F.M. Savoy, Y.H. Lee, and S. Winkler," Design of lowcost, compact and weather-proof whole sky imagers for HighDynamic-Range captures", IEEE International Geoscience and Remote Sensing Symposium, pp. 5359-5362, Jul 2015. IEEE.
6. M. Fujita and T. Sato," Observed behaviours of precipitable water vapour and precipitation intensity in response to upper air profiles estimated from surface air temperature", Scientific Reports, vol. 7, no. 1, pp. 1-6, Jul 2017
7. D.R. Nayak, A. Mahapatra, and P. Mishra," A survey on rainfall prediction using artificial neural network", International Journal of Computer Applications, vol. 72, no. 16, Jan 2013.
8. S. Chatterjee, S. Ghosh, S. Dawn, S. Hore, S. and N. Dey, "Forest Type Classification: A hybrid NN-GA model-based approach", In Information systems design and intelligent applications, pp. 227- 236, Springer, New Delhi,2016.
9. A.D Dubey, "Artificial neural network models for rainfall prediction in Pondicherry", International Journal of Computer Applications, vol 120, no. 3, Jan 2015.
10. S. Manandhar, Y.H. Lee, Y.S. Meng, and J.T. Ong," A simplified model for the retrieval of precipitable water vapor from GPS signal", IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 11, pp. 6245-6253, Jul 2017.
11. S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A.S. Ashour and V.E. Balas, "Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings", Neural Computing and Applications, vol. 28, no. 8, pp. 2005-2016, Aug 2017.