Turkish Journal of Computer and Mathematics Education *Vol.12 No.6 (2021), 3467-3474*

Research Article

Efficient Rainfall Prediction and Analysis using Machine Learning Techniques

Gowtham Sethupathi.M[a], Yenugudhati Sai Ganesh [b], Mohammad Mansoor Ali[c]

[a]AssistantProfessor,Departmentof Computer Science and Engineering,SRMinstituteofscienceandtechnology,Chennai,India[b]UG student, Departmentof Computer science and Engineering, SRM institute of science and technology, Chennai, India[c]UG student, Department of Computer science and Engineering, SRM institute of science and technology, Chennai, India

Abstract: Rainfall prediction is a beneficiary one, but it is a challenging task. Machine learning techniques can use computational methods and predict rainfall by retrieving and integrating the hidden knowledge from the linear and non-linear patterns of past weather data. Various tools and methods for predicting rain are currently available, but there is still a shortage of accurate results. Existing methods are failing whenever massive datasets are used for rainfall prediction. This study provides efficient rainfall prediction methods of machine learning techniques: random forest and logistic regression methods that provide an easy and accurate prediction and determine which one is more effective in comparison. This study would assist researchers in analyzing the most recent work on rainfall prediction with an emphasis on machine learning techniques and providing a reference for possible guidance and comparisons. Anaconda framework is used, and the coding language used is Python, which is portable and dynamic. Numpy, matplotlib, seaborn, and pandas are the libraries used for the implementation.

Keywords: Rainfall prediction, Classification, Random forest, Machine learning,Logistic regression.

1. Introduction

In the current scenario, rainfall is a significant factor for most essential things happening throughout the world. The farming sector is regarded as one of the most critical factors determining the country's economy, and farming relies entirely on rainfall. This research uses machine learning techniques for rainfall prediction and conducts the comparative analysis of two machine learning techniques, respectively, depicting an efficient rainfall prediction method. Rainfall prediction facilitates water resources management, flood alerts, flight operations management, limiting transportation, construction activities, and other factors that are most important to humankind. Rainfall data for forecasting is collected using weather satellites, wired and wireless instruments, and high-speed computers are used. Rainfall prediction has been a fascinating and captivating sector since the dawn of civilization, and it remains one of the most complex and enticing domains.Scientists use various methods and techniques to predict rainfall, some of which are more precise than others. Weather forecasting gathers atmospheric conditions such as humidity, temperature, pressure, rainfall, wind direction &speed, evaporation, etc.

Presently, Rainfall prediction is the most crucial factor for most water storage schemes worldwide. The uncertainty of rainfall data is one of the most complex problems. Today, most rainfall forecasting methods are incapable of detecting hidden patterns or non-linear trends in rainfall data. This research would help discover all hidden patterns and non-linear trends, which would be necessary for predicting accurate rainfall [1]. Due to the presence of complex issues in existing methods that cannot find the hidden patterns and non-linear trends efficiently the majority of the time, the forecast predictions were incorrect, resulting in massive losses. Thus, this research aims to find a rainfall prediction system that can solve all issues, find complexity and hidden patterns present, and provide proper and reliable predictions, therefore assisting the country in developing agriculture and the economy. [2]

1.1 Machine Learning techniques used

Random forest is an ensemble learning method and supervised algorithm of machine learning technique. The ensemble learning method is a learning process in which different algorithms or the same algorithm are repeatedly combined to provide the most accurate prediction technique. It combines several similar algorithms, namely the various decision trees, to produce trees like a forest. Thus, the term "Random Forest" applies to both classification and regression functions. The basic procedures required in carrying out the random forest algorithm are first to choose K records at random from the data collection. And then create a decision tree using these K records. Decide how many trees you want to use in your algorithm and replicate the previous steps. Each forest tree estimates the division under which the new record belongs when it comes under classification. Eventually, the new form is granted to the class that receives the most votes. Since several trees and every tree are trained on a sample of data, the random forest algorithm is not biased.

Essentially, the random forest algorithm depends on the capacity of the crowd, which decreases the algorithm's overall bias. This algorithm is highly robust. And whenever new information is applied to the database, the total algorithm is unaffected since further information may impact one tree. Still, it is highly uncertain that it will affect all trees. The random forest algorithm works best when you have both numerical and categorical attributes. Random forest works efficiently for rainfall prediction in massive datasets. Logistic regression examines the relationship between dependent or independent variables to determine the probabilities of a function. It does not treat only the association between the study variables as if it were a straight line. Also, Logistic regression, on the other hand, uses the natural logarithmic equation to determine the relationship between variables and test results to determine the coefficients. Since it includes actual identifiers for training data, it is also a supervised machine learning algorithm. Logistic regression is an extension of linear regression. It can be helpful in both classification and regression functions, but it was mainly used for classification purposes. It is termed as a classification algorithm widely.The critical distinction between the two is that logistic regression is helpful whenever the dependent variables are binary function, and linear regression is practical whenever dependent variables are continuous. The advantage of logistic regression has been increased from the past decade for prediction purposes.

1.2 Libraries and Platforms used

NUMPY is a numerical Python application that provides fast mathematical calculation functions for computation. It can be used to read data in arrays and for computing procedures. PANDAS can read and write various files and directories. Furthermore, data processing of information frameworks makes the data source file highly performed and easy to use. SEABORN is based on matplotlib, which is used for data visualization in python data that provides a high-level platform for displaying appealing and detailed statistical graphics. MATPLOTLIB is a Python two-dimensional plotting library that produces high-quality reports in various hardcopy and graphical formats. Matplotlib is used in Python scripts, Jupyter notebooks, IPython shells, Web frameworks, and four graphical user interface toolkits. Matplotlib strives to make things easy, fast and also make complex things possible. Using a few code lines, you can generate graphs, histograms, bar charts, scatterplots, etc. Anaconda Navigator is a user interface software application that allows you to quickly launch applications and access conda packages, configurations, and channels without using command-line functions. It is compatible with Linux, Windows, and OS. Jupyter notebook is available in anaconda navigator. It is an open-source computational notebook that allows researchers to combine source code, computational performance, descriptive language, and multimedia tools into a single document.

2. Related Works

Rainfall is one of the most critical factors determining the life behaviour of ecosystem members, which influences the country's economic factor and citizens in the area. To avoid disasters caused by rainfall activities, an action that predicts rainfall instability behaviour in the future must be triggered. For rainfall prediction, two processes are typically used. One is to use the vast amounts of data gathered over time and analyze the data to acquire information about future rainfall. The other consists of the create equations by defining various parameters and substituting the values to produce the desired outcome.

Researchers in paper [3] proposed a novel approach for predicting rainfall. A two-step procedure was used. The biased forward classification method is used to narrow the list of features and consider the prospective segments for rain forecasting. In the training process, the dataset is first aggregated using the k-means method, and then a different Neural Network is trained for each cluster. The proposed two-step prediction approach was compared to the classifier related to multiple statistical efficiency evaluation parameters. Over the years 1989-1995, the Dumdum meteorological department collected data for experimental purposes.

One of the most recent researches includes a paper [4] in which researchers expressed rainfall expectation is concerned with the guarantee of precipitation patterns for a specific region. It is regarded as critical for the horticultural industry and other businesses. As far as everyone is concerned, this is the first attempt at applying a somewhere down in expecting monthly precipitation. Compared to Australia Public weather report and Earth-System Prediction model, the suggested method is a gauging design published by the meteorology service. The proposed method was evaluated by comparing the Australian Weather and Earth-System Prediction Model, a forecast model posted by the weather service, a prediction model published by the weather department. The mean absolute error, Pearson's correlation coefficient, root mean square error, and proposed Nash-Sutcliffe efficient coefficient were enhanced. More research revealed that months with higher yearly rainfall averages are typically outperforming with lower yearly rainfall averages in months. The efficiency is good and has a lot of potential in this sort of application.

The data mining technique assists in the discovery of secret patterns, which leads to the accurate prediction of

rainfall. This method takes into account all of the variables that influence rainfall and forecasts potential rain. To predict rainfall, a customized, optimized, and improved data mining technique is used. Several climate attributes are considered to indicate rain. Weather information such as evaporation, atmosphere, wind direction, maximum and minimum temperature, humidity, etc., are collected. It is stated that using more attributes will not guarantee that the forecast will be highly reliable. For prediction, both supervised and unsupervised methods can be considered. In paper [5], it discussed predictions in Malaysia, Columbia, Australia, and India. Previous weather data used to train the algorithm is one of the most important variables that influenced the outcome. the climate attributes are used as predictors and the area where the prediction will occur. Some of the most recent research trends in this domain are discovering correlations between weather characteristics, while others emphasize training the algorithm. Several hidden factors can influence rainfall prediction, and data mining techniques can identify all of the hidden patterns. This approach must be adapted and optimized such that all projections are more error-free. So, future research aims to improve, refine, and integrate this data mining strategy so that all of the current problems associated with discovering secret patterns are overcome. A proper association between weather variables is found.

The deep learning process is divided into three phases. The first phase is concerned with determining the correct algorithm. The algorithm is chosen based on the dataset available, such as which algorithm fits the given data information well. The second phase is concerned with locating the one that works well with the algorithm. Several Metrics products for predictive precipitation modelling, the proportion of false reports, and hazard percentages are being implemented in the final phase. The calculations in the first two phases are based on the initially expected time sequence, including negative values. The benefit of this approach is that it identifies all non-linear patterns and suitably calculates correlation coefficients. Still, since it uses actual predicted data for estimation, the calculation often takes more time and does not provide an accurate result. [6].

The linear regression method is used to discover a relationship between dependent and non-dependent variables. So this method will provide a reasonable estimate of rainfall for a given time frame. It is dealing with the analysis of data sets and their subsequent processing. The obtained data is further analyzed, and the outcome is expected. The key benefit of using this method is that it outperforms data mining methods.

In contrast to linear regression, which provides approximate value, the data mining approach provides generalized value. The only drawback of this strategy is that it does not work well with long-term forecasting. This method will be improved in the future by using multiple regressions to identify rainfall prediction [7].

The hybrid classifier is divided into three sections. The sections are as follows: simulation section, training section, and testing section. The simulation method is concerned with data analysis. The training section is concerned with determining the best algorithm to use, and the dataset is evaluated to obtain the predicted result. Following research, it was discovered that this technique produced a good outcome. This algorithm can be improved in the future by enhancing series data rainfall estimation and successfully hybridizing the support vector regression analysis [8].

Although various data mining methods are available for use, it is essential to choose a technique acceptable for the domain to which it is applied. In some cases, the regression technique is more efficient, while in others, the rule-based technique and decision tree algorithms provide reliable results at a low computational cost. Researchers analyzed various data mining techniques and compared the performance of C4.5, CART, k means clustering, ANN, and MLR techniques when used for weather prediction in [9]. They concluded that k-means and decision tree algorithms outperform other methods in weather prediction.

Model SARIMA in Sudan had been tested. Several calculations are made, and the charts are traced, and outcomes are found based on these charts. As a sequence, three types of diagrams are drawn. The figures reveal that this is not true of the stationary hypothesis. It can't be fixed, but it might be twelve seasonal. It shows a secular horizontal tendency. This model has the most significant disadvantage: it is restricted to a smaller region. Improving this strategy in the future can extend the field [10].

Bayesian network method for a monthly average rain forecast of 21 Assam stations, India. This method can be helpful for improved water supply control. For analysis, which has been taken from various sources, monthly data from 1981 to 2000 are used for all atmospheric parameters. Rainfall at a location for this model is considered a vector, and the Bayesian network displays correlations between rainfall at various stations. In this study, the researcher uses the K2 algorithm, and the conditional probability with cross-validation approximations is found. Further research will improve this model [11].

3. Proposed Methodology

The proposed system forecasts rainfall with two machine learning techniques: logistic regression and

random forest, for a more precise solution. First, the system compares the procedure and provides the best algorithm to the output. Data entry, pre-processing of data, data division, algorithm training, data set checking, comparisons between both algorithms, prediction of the most reliable algorithm, and results at the end are the steps related to the proposed scheme.

The collection of data included in this study consists of many parameters and the known class of output. The output class will be predicated on the other values and perspectives. Using the output class in the dataset compares the reliability and consistency of the Machine Learning techniques. The effects of processing are equal to the known output class, and reliability is calculated by the sum of accuracy, recall, and f test of output.

The classification method used in this research is divided into four stages: database collection, pre-processing, forecasting, and results from estimation. The input data collection for Rainfall prediction is taken from both the weather prediction website and contains several atmospheric parameters. Inadequate information would affect the accuracy of the outcome because the parameter with a missing value would be unable to participate in the prediction process. Aside from missing values, there was also uncertainty in the database, where the value remained below or exceeded thresholds. It is suggested that the values be held under such limits for good data analysis performance.

The pre-processing of the dataset is a crucial step in the classification scheme that ensures the consistency of extraction performance. A classification system is used to predict rainfall, in which the dataset was cleaned and structured before the procedure. Cleaning aims to deal with the missing attributes, and the goal of incarceration is to protect the estimations of the characteristics within particular limits. Such pre-planning tasks, including wins, are essential to a better classification operation. Prediction implementation of pre-owned AI strategies is based on accuracy, analysis, and estimate, necessary data recovery measures. The numbers are then shown in tables and maps.

The data collection is made up of data that must be cleaned before being separated into test and train data. After that, it must go through pre-processing and simulation, and it must be predicted using an algorithm in modelling before being analyzed and implemented.

As seen in Figure 1, the dataset comprises information that must be cleaned before being separated into the test and train data, pre-processing and modelling, and prediction using an algorithm in modelling before being analyzed and deployed.

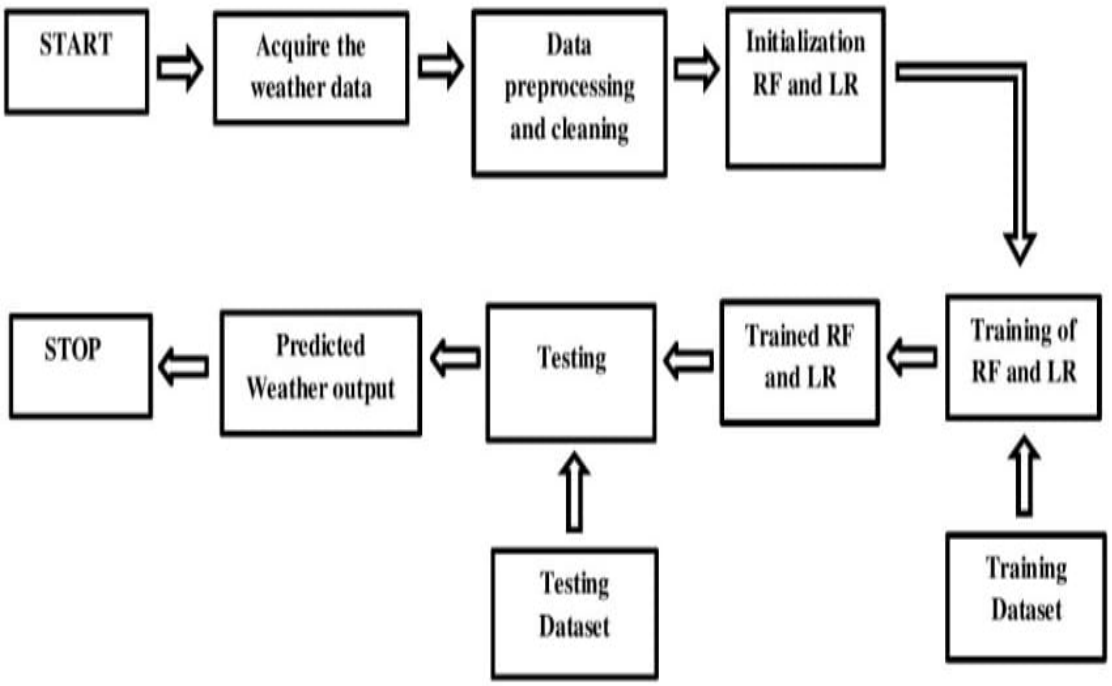
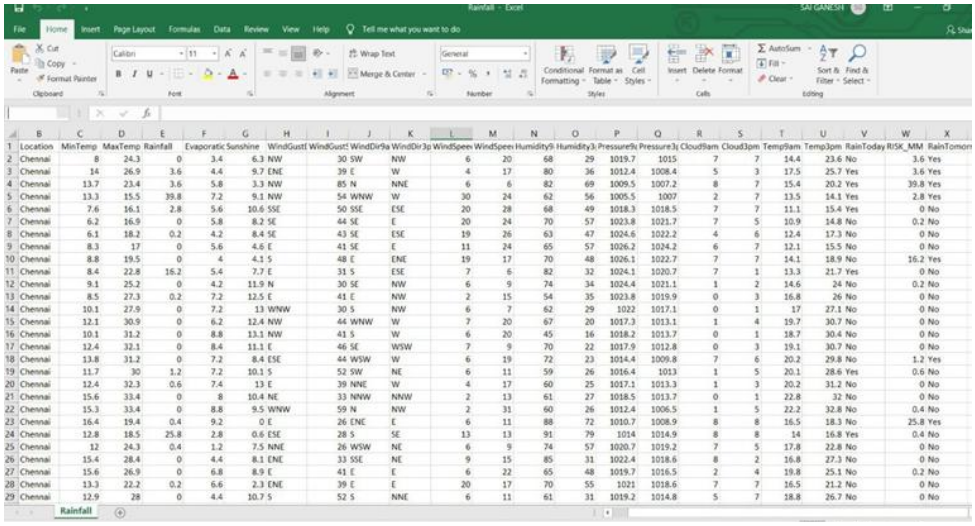


Figure 1. System Architecture

3.1 Collection of Data set



Location	MinTemp	MaxTemp	Rainfall	Evaporatic	Sunshine	WindGust[	WindGust!	WindDir9a	WindDir3p	WindSpeen	WindSpeen	Humidity9	Humidity3	Pressure9(	Pressure3(	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorr
Chennai	8	24.3	0	3.4	6.3	NW	30	SW	NW	6	20	68	29	1019.7	1015	7	7	14.4	23.6	No	3.6	Yes
Chennai	14	26.9	3.6	4.4	9.7	ENE	39	E	W	4	17	80	36	1012.4	1008.4	5	3	17.5	25.7	Yes	3.6	Yes
Chennai	13.7	23.4	3.6	5.8	3.3	NW	85	N	NNE	6	6	82	69	1009.5	1007.2	8	7	15.4	20.2	Yes	39.8	Yes
Chennai	13.3	15.5	39.8	7.2	9.1	NW	54	WNW	W	30	24	62	56	1005.5	1007	2	7	13.5	14.1	Yes	2.8	Yes
Chennai	7.6	16.1	2.8	5.6	10.6	SSE	50	SSE	ESE	20	28	68	49	1018.3	1018.5	7	7	11.1	15.4	Yes	0	No
Chennai	6.2	16.9	0	5.8	8.2	SE	44	SE	E	20	24	70	57	1023.8	1021.7	7	5	10.9	14.8	No	0.2	No
Chennai	6.1	18.2	0.2	4.2	8.4	SE	43	SE	ESE	19	26	63	47	1024.6	1022.2	4	6	12.4	17.3	No	0	No
Chennai	8.3	17	0	5.6	4.6	E	41	SE	E	11	24	65	57	1026.2	1024.2	6	7	12.1	15.5	No	0	No
Chennai	8.8	19.5	0	4	4.1	S	48	E	ENE	19	17	70	48	1026.1	1022.7	7	7	14.1	18.9	No	16.2	Yes
Chennai	8.4	22.8	16.2	5.4	7.7	E	31	S	ESE	7	6	82	32	1024.1	1020.7	7	1	13.3	21.7	Yes	0	No
Chennai	9.1	25.2	0	4.2	11.9	N	30	SE	NW	6	9	74	34	1024.4	1021.1	1	2	14.6	24	No	0.2	No
Chennai	8.5	27.3	0.2	7.2	12.5	E	41	E	NW	2	15	54	35	1023.8	1019.9	0	3	16.8	26	No	0	No
Chennai	10.1	27.9	0	7.2	13	WNW	30	S	NW	6	7	62	29	1022	1017.1	0	1	17	27.1	No	0	No
Chennai	12.1	30.9	0	6.2	12.4	NW	44	WNW	W	7	20	67	20	1017.3	1013.1	1	4	19.7	30.7	No	0	No
Chennai	10.1	31.2	0	8.8	13.1	NW	41	S	W	6	20	45	16	1018.2	1013.7	0	1	18.7	30.4	No	0	No
Chennai	12.4	32.1	0	8.4	11.1	E	46	SE	WSW	7	9	70	22	1017.9	1012.8	0	3	19.1	30.7	No	0	No
Chennai	13.8	31.2	0	7.2	8.4	ESE	44	WSW	W	6	19	72	23	1014.4	1009.8	7	6	20.2	29.8	No	1.2	Yes
Chennai	11.7	30	1.2	7.2	10.1	S	52	SW	NE	6	11	59	26	1016.4	1013	1	5	20.1	28.6	Yes	0.6	No
Chennai	12.4	32.3	0.6	7.4	13	E	39	NNE	W	4	17	60	25	1017.1	1013.3	1	3	20.2	31.2	No	0	No
Chennai	15.6	33.4	0	8	10.4	NE	33	NNW	NNW	2	13	61	27	1018.5	1013.7	0	1	22.8	32	No	0	No
Chennai	15.3	33.4	0	8.8	9.5	WNW	59	N	NW	2	31	60	26	1012.4	1006.5	1	5	22.2	32.8	No	0.4	No
Chennai	16.4	19.4	0.4	9.2	0	E	26	ENE	E	6	11	88	72	1010.7	1008.9	8	8	16.5	18.3	No	25.8	Yes
Chennai	12.8	18.5	25.8	2.8	0.6	ESE	28	S	SE	13	13	91	79	1014	1014.9	8	8	14	16.8	Yes	0.4	No
Chennai	12	24.3	0.4	1.2	7.5	NNE	26	WSW	NE	6	9	74	57	1020.7	1019.2	7	5	17.8	22.8	No	0	No
Chennai	15.4	28.4	0	4.4	8.1	ENE	33	SSE	NE	9	15	85	31	1022.4	1018.6	8	2	16.8	27.3	No	0	No
Chennai	15.6	26.9	0	6.8	8.9	E	41	E	E	6	22	65	48	1019.7	1016.5	2	4	19.8	25.1	No	0.2	No
Chennai	13.3	22.2	0.2	6.6	2.3	ENE	39	E	E	20	17	70	55	1021	1018.6	7	7	16.5	21.2	No	0	No
Chennai	12.9	28	0	4.4	10.7	S	52	S	NNE	6	11	61	31	1019.2	1014.8	5	7	18.8	26.7	No	0	No

Figure 2.Data Set CSV File

As Shown in the Fig 2 data set CSV file, The collection of data used in this system includes rainfall data from many regions within India. It includes rainfall data from 2015 to 2018. Along with that, average rainfall and rainfall between the transition of two months have been included. The dataset contains a total of 966 rows. The dataset was obtained from the Kaggle website, which is a data collection and publishing website. It includes attributes such as date, location, minimum temperature, maximum temperature, wind direction and speed, wind gust direction and speed, humidity, pressure, temperature, rain today, rain tomorrow, etc.

3.2 Pre-processing

The collection of data used in this system includes rainfall data from many regions within India. It includes rainfall data from 2015 to 2018. Along with that, average rainfall and rainfall between the transition of two months have been included. The dataset contains a total of 966 rows. The dataset was obtained from the Kaggle website, which is a data collection and publishing website. It includes attributes such as date, location, minimum temperature, maximum temperature, wind direction and speed, wind gust direction and speed, humidity, pressure, temperature, rain today, rain tomorrow, etc.

There are four different data pre-processing stages:

- Data Arrangement: The material we've chosen is unlikely to be in a format that allows you to interact with it. The details could be in a social database, but we needed it in a simple text or a restricted record configuration. So, it is formatted in a data framework or material document.
- Import dataset and libraries: The formatted data is imported into a CSV file So that jupyter notebook can read the file and continue the process. All-important libraries required like Numpy, pandas, matplotlib, seaborn are imported for reading, visualization, and manipulating the data.
- Removing null values: Sometimes, the information of data may miss. In that case, we can perform two methods in removing the null values either deleting the row which contains the null value or calculating the mean value of the particular column or a row and replaces the missing value with the mean value. Therefore, it gives better results than the previous method.
- Splitting the data: Choose the independent variables or feature columns of the database, represent them as x, and define the target or dependent variable rain tomorrow as y. The database is separated into two separate sets - train data and test data using function train_test_split(). Typically, the dataset is divided into 7:3 or 8:2 ratios. That means we can use 70-80% of the data for training the algorithm while leaving out the remaining 20-30% for test data. The splitting of data depends on the form and size of the dataset.

3.3 Applying algorithms

After splitting the database into training data and testing data, we will train the logistic regression and random

forest models. From the sklearn and sklearn.ensemble package, we will import logistic regression function and random forest classifier. The training data is a part of a database used to train the algorithms, and test data is the remaining part of the database used to validate the algorithms. By using training data, the algorithms understand the relationship between dependent and independent variables.

Now in the implementation part, First, we will construct an instance of logistic regression function as "lf" and random forest classifier as "RF." The fit() function allows us to train the model; once we have fit the training model, the algorithms use the test data to forecast outcomes according to its methods. Results of rain tomorrow prediction are generated in binary form 1 or 0 that represents yes or no options. We can create a confusion matrix table to validate the predicted outcomes with actual outcomes and find the accuracy score. The confusion matrix table is used to assess a classification model's results. The output of an algorithm can also be visualized. The basics of a confusion matrix are the total of accurate and inaccurate prediction results.

4. Results and discussion

Results are shown in the form of figures after comparing the predicted data with actual data. Accuracy scores of algorithms are calculated using efficient computation and imported into confusion matrixes for understanding purposes. By using accuracy scores, a comparison of algorithms is shown. The confusion matrix shows values along with colour representations in which more values represent darkish colours, and fewer values represent faded colours.

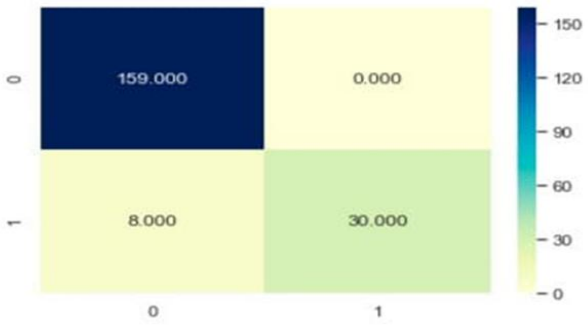


Fig 3: Confusion matrix for Logistic Function

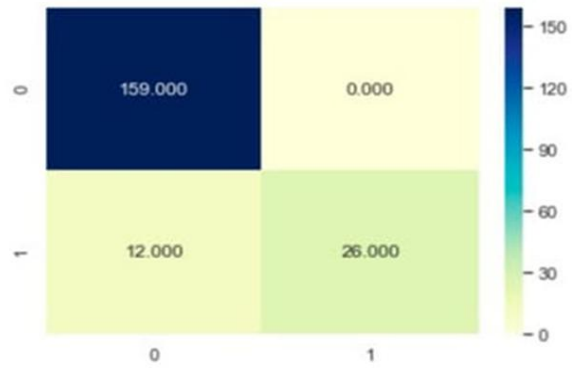


Fig 4: Confusion matrix for Random Forest

The Confusion matrix for Logistic regression function and Random forest algorithms are shown in Figures No. 3 and 4. Here, we can able to see an array object or dynamic view as the confusion matrix. Diagonal elements are correct estimates, while incorrect estimates are nondiagonal values.

As the test data contains around 20% of the total data set, 197 predictions are made. In Figure No.3, 159 and 30 are the correct predictions, and 8 are incorrect predictions. In Figure No.4, 159 and 26 are the correct predictions, and 12 are wrong predictions.

```
acc1 = accuracy_score(y_test ,RF)
acc1
```

```
0.9441624365482234
```

```
from sklearn.metrics import confusion_matrix
```

```
cn1 = confusion_matrix(y_test , mod3.predict(y_train))
```

Fig 5: Accuracy score for Random Forest

```
acc = accuracy_score(y_test , lf)
acc
```

```
0.9593908629441624
```

```
from sklearn.metrics import confusion_matrix
```

```
cn = confusion_matrix(y_test , lf)
```

Fig 6: Accuracy score for Logistic Regression

The accuracy score for rainfall prediction by Random forest and Logistic regression are given in Figures No. 5 and 6. The accuracy score percentage for the random forest is 94.4%, and logistic regression is 95.9% approximately. The accuracy score is calculated by dividing correct predictions with total predictions.

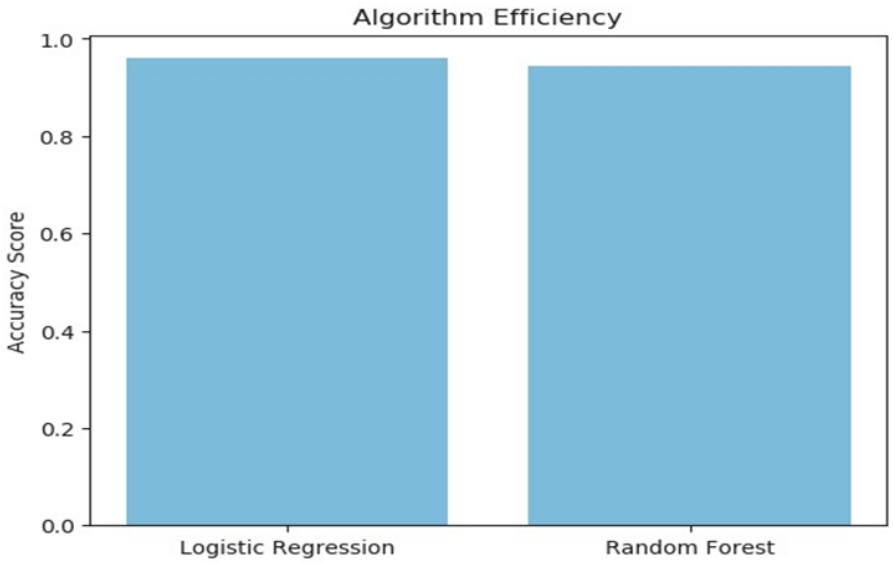


Fig 7: Comparison of algorithms

As seen in Figure No. 7, both algorithms performed well depending on their technique and evaluated with great accuracy, high speed in less time.

Results have shown that both algorithms performed well. The confusion matrix layout allows visualization of the performance of both algorithms and offers less inaccurate predictions. The accuracy score for the logistic regression algorithm is slightly more efficient than the random forest algorithm.

5. Conclusion

Rainfall is primarily responsible for the majority of India's economy. It should be considered as the main issue for the majority of us. The existing methods for rainfall forecasting fail in the most complicated situations because they cannot forecast the hidden patterns present, which are yet to be understood to perform an accurate prediction. Two approaches are being compared to obtain a precise method of predicting rainfall. The first is a random forest technique, and the second is logistic regression. Techniques for mining user data, The results of the quality analysis are shown in graphs and figures. A classification method is used for accurate prediction in which the given dataset is cleaned and normalized before the classification process begins. The results reveal that the characterization procedures used worked well for the no-downpour class but not so well for the downpour class. The reasons for the lower downpour class exactness may include missed qualities, the lack of major climatic characteristics in the dataset, and a lower overall precipitation rate in the areas. Further predictions can be carried out by evaluating additional classification methods and climate characteristics on various weather dates. Accuracy is now based on random forest, and logistic regression to predict rainfall is very efficient and provides accurate results.

References

1. MasafumiGoto, Faizah Cheros, Nuzul Azam Haron, Nur-AdibMaspo, Aizul Nahar Bin, MohdNawiHarun, and MohdNasrun,(2020)," Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review." IOP Science journal.
2. V.P Tharun, Ramya Prakash, S.Renuga Devi.(2019)"Prediction of Rainfall Using Data Mining Techniques" ieeexplore.ieee.org.
3. Ali Haidar and Brijesh Verma. (2018). "Monthly rainfall forecasting using one-dimensional deep convolutional neural network." IEEE Access 6, pp. 69053-69063.
4. S. Chatterjee, B. Datta, S. Sen, N. Dey, and N.C Debnath. (2018) "Rainfall prediction using hybrid neural network approach," 2nd International Conference on Recent Advances in Signal Processing, Telecommunications & Computing, pp. 67-72. IEEE.
5. Shabib. A Muneer. A Naureen. H, Mohamad Salman, Ifthikar.A& Jawed. Z. (2018) "Rainfall prediction using data mining techniques: A systematic literature review." International Journal of Advanced Computer Science and Applications.
6. R.Vinayakumar, P.Geetha, S.Aswin. (2018) "Deep Learning Models for the Prediction of Rainfall," IEEE.
7. MAI.Navid and MH.Niloy. (2018) "Multiple Linear Regressions for Predicting Rainfall for Bangladesh," Science Publishing Group,Volume 6, Issue 1.
8. A. Jesudoss, K. Harsha Vardhan, S. Prince Mary, K. V. Sai Sandeep, and P. Asha. (2020)"An Efficient Hybrid Machine Learning Classifier for Rainfall Prediction," IOP Science journal.
9. Divya Chauhan, Jawahar Thakur.(2014) "Data Mining Techniques for Weather Prediction: A Review," IJRITCC, ISSN: 2321-8169- 2184 – 2189.
10. Harrison. E and Tariq Mahgoub. M.(2014) "Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods," International Journal of Scientific Research in Knowledge.
11. Manish Kumar Goyal and Ashutosh Sharma. (2015)"Bayesian network model for monthly rainfall forecast," IEEE, International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).
12. A.D Dubey. (2015)"Artificial neural network models for rainfall prediction in Pondicherry," International Journal of Computer Applications, vol 120, no. 3.
13. D.R. Nayak, A. Mahapatra, and P. Mishra. (2013)" A survey on rainfall prediction using artificial neural network," International Journal of Computer Applications, vol. 72, no. 16.