

Rainfall Prediction Using Machine Learning

Jasmin Joseph

Department Of Computer Applications
Amal Jyothi College Of Engineering ,Koovapally
Kottayam,Kerala
Jasminjoseph2022@mca.ajce.in

Dr. Manju K Mathai

Asst. Professor in Computer Science
Amal Jyothi College Of Engineering ,Koovapally
Kottayam,Kerala
manjukmathai@amaljyothi.ac.in

Abstract- Rainfall prediction is the demanding and mysterious tasks which has a extraordinary impact on human society. Timely and accurate predictions can help to decrease human and financial loss. This study represents a set of experiments which involve the use of widespread machine learning techniques to build models to picture out whether it is going to rain tomorrow or not depends on weather data for that specific day . This relative study is managed using Logistic Regression Algorithm to build a predictive model to predict whether it will rain tomorrow or not. **Keywords-** machine learning, logistic regression, rainfall prediction **Analysis, report**

I. INTRODUCTION

Rainfall forecasting continues to be a serious study attracting the attention of governments, industries, risk management institutions and the scientific community. Rainfall is a weather factor that works on many human activities like agricultural production, construction, power generation, forestry and tourism, among others. In this journalism, I want to implement predictive model on Rain Dataset to predict whether it will rain tomorrow or not. The Dataset belongs about 10 years of daily climate observations of many locations. By the end of this article, we will able to fit a predictive model. To this range, rainfall prediction is vital since this variable is the one with the largest correlation with unfavorable natural events such as landslides, flooding, mass movements and avalanches. These events have affected society for years. Therefore, having an Suitable approach for rainfall prediction makes it possible to take protective and reduction measures for these natural truth.

As global warming has increases so earth's temperature also increases and due to which our local area's yearly rainfall patterns have also been forced and this destroy the population living in the region, as farmers and other people who very much depend on rainfall for all their water-based needs, so in these areas correct predictions of rainfall is of very important.

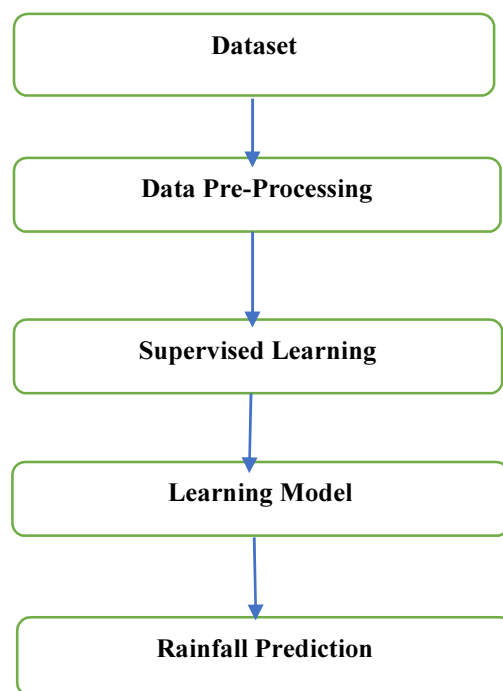
II. MACHINE LEARNING ALGORITHM

Machine learning is the method of analysing the data that automatize analytical model structure. It is associated with artificial intelligence based on the idea that system can grasp from data and recognize patterns and make decisions with very small human association. Machine learning was made from the

pattern recognition mod. Machine learning is a learning of artificial intelligence with computer science concentrate on data and algorithms to improve the human learn accuracy.

III. PROPOSED SYSTEM

One of the main advantages of this system is its capability to expand the rank of algorithm and dataset. The further the data is used for prediction will be effective, the more will be accuracy of the prediction. Same goes with the accuracy of the algorithm. Machine learning method predictions allow businesses to make more correct guesses as to the likely outputs of a question based on past data. A dataset with finite set of data cannot predict accurate output so we need to teach the model with further data in dataset. The results get from this algorithm method can be examine to create a rainfall prediction classification. The models like Logistic Regression, SVM, Random Forest will be used in rainfall prediction



The steps we need to follow are:

1. Data Collection
2. Defining Data
3. Pre-processing
4. Building model
5. Analysis
6. Results

With the algorithm we also follow these steps:

1. Import the libraries
2. Import the datasets
3. Define the data sets
4. Dataset testing
5. Algorithm execution
6. Comparison of results

Table 1: Dataset Description

S No.	Attributes
1.	Temperature
2.	Humidity
3.	cloud
4.	Rain Today
5.	wind
6.	cloud
7.	sunshine
8.	pressure

A. Data Processing

This is the crucial method in the machine learning method. The lost data and the pollution in the dataset can decrease the worth of the output that can be formed from the dataset. Data pre-processing can be carried out in order to extend the effect of the data that is retrieved after data processing. In the dataset we can execute data pre-processing by following methods.

1. **Removing the missing values-** All the arrival that have 0 want to be removed. Because 0 is not a valid entry.
2. **Data Splitting-** Once the data is cleared it needs to modify the training dataset and test models. once the split data is taken the training data set is used to train the algorithm. A training model is established on the attribute of the training is generated after the training action.

A. Machine Learning

One time data is organized we work machine learning procedure. We put in many groupings and algorithms to predict rainfall. The performance of these methods is felt to find their accuracy and meet the most features which can help us in our rainfall prediction. The following methods can be used.

1. Logistic Regression

Logistic regression is the day-to-day sort of machine learning algorithm used. It has got a high-level accuracy and is normally very final .0s and 1s are the results that are normally pointed with logistic regression. It is usually used when data needs to be grouped. Sigmoid function is used in logistic regression to predict the probability of valid and invalid class.

$$\text{Sigmoid Function } P = 1/1+e^{-(a+ bx)}$$

Here P =probability, a and b are model parameters.

2. Support Vector Machine

It is known as Supervised machine learning algorithm. It makes a hyperplane that divides different classes. This hyperplane is also used for grouping and also regression. It can point out arrival in particular classes and grouping the instances which are not supported by the data.

Algorithm:

- The hyperplane which separates the class the best is selected.
- The distance between the planes and the data called the Margin is calculated. This is used to find out the best plane.
- Increase the distance between the classes lower the chance of miss formation.
- We need to choose the class with large margin.
- Margin=Distance to negative point +Distance to positive point.

3. Random Forest

It is a machine learning algorithm used for grouping and regression data. It has got the highest mark of accuracy in compare to other models. Any dataset of any size can be simply monitor by this method. The performance of the decision tree algorithm can be broadly improved by Random Forest tree algorithm by decreasing the variance

Algorithm:

1. Select 'R' features from 'M' total features where $R \ll M$
2. Find the better node from R features
3. Split the node in to sub nodes using the best split method.
4. Repeat the above steps as late as 'l' number of nodes is extended.
5. Repeat the above steps to build forest for 'a' number of times to make 'n' number of trees.

IV.RESULTS

With our speak to we use differing grouping and models to implement our methods using the python language. By using this model, we are looking for a machine learning algorithm with the highest accuracy. By comparing working models, we can see that the Random Logistic Regression has biggest overall performance in comparison with other classifiers.

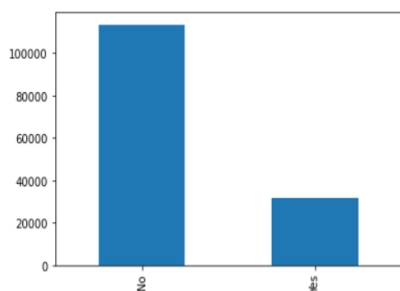


Figure: Exploratory data analysis

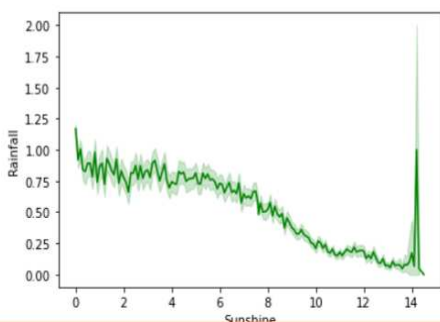


Figure 2: Sunshine vs Rainfall

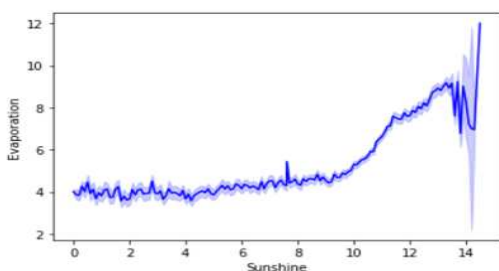


Figure 3: Sunshine vs Evaporation

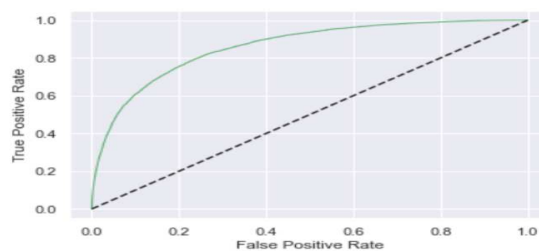


Figure 4: ROC curve for Logistic Regression Model

V.CONCLUSION

This Paper has presented a supervised rainfall learning model which used machine learning algorithms to grouping rainfall data. We used several machine learning algorithms to check the accuracy of rainfall prediction. We have compared SVM, Random Forest classifiers we can conclude that Random Forest is the Machine learning algorithm which is suitable for rainfall prediction in India.

currently machine learning used in no. industries. As the data expand the complexity of that data will increase and for that we are using machine for the better understanding of that data. In Weather predictions it's pretty helpful with good accuracy score and in rainfall also its gives pretty good predictions. In future we are planning to increase our work in Storm predictions and Crop prediction with the rainfall prediction

VI.REFERENCES

- [1] Janani, B; Sebastian, P. (2014). Analysis on the weather forecasting and techniques. *International Journal of Advanced Research in Computer Engineering & Technology*, 3(1), 59–61. <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-1-59-61.pdf>
- [2] Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I., & Nawaz, Z. (2018). Rainfall prediction using data mining techniques: A systematic literature review. *International Journal of Advanced Computer Science and Applications*, 9(5), 143–150. <https://doi.org/10.14569/IJACSA.2018.090518>
- [3] Wahyuni, E. G.Fauzan, L. M.F.Abriyani, F.Muchlis, N. F., & Ulfa, M. (2018). Rainfall prediction with backpropagation method. *Journal of Physics: Conference Series*, 983(1). <https://doi.org/10.1088/1742-6596/983/1/012059>