**RAINFALL PREDICTION USING LINEAR REGRESSION**

**ARTIFICIAL INTELLIGENCE AND MACHINE**

**LEARNING**



**NAMES H.NO**

M. Siri Chandana 2203A51492

K. Pavani 2203A51482

B. Swarna 2203A51540

A. Abhilash 2203A51536

**Under the Guidance of**

Mohammed Ali Shaik

Assist.Professor

**ABSTRACT:**

Predicting rainfall is important from many years like in farming purpose, drinking water resources. Now-a-days rain has a big impact on environmental conditions like floods, droughts and etc. It also impacts on soil and air moisture. Lately researches have developed different techniques using different models to predict the rainfall. So our particular study is about using linear regression and other known machine learning techniques to make a better rain forecast which mainly indicates amount of rainfall which highly affected, humidity relative at a particular region.

**I. INTRODUCTION:**

Over time, advancements in intelligent computing have led to the development of various techniques for predicting rainfall, with Artificial Neural Networks (ANNs) emerging as a popular choice. ANNs play a vital role in rainfall forecasting, which is crucial for countries like India heavily reliant on agriculture. Currently, precipitation data is primarily collected through three methods: rain gauges, estimation of rainfall While rain gauge data is accurate, it is limited to localized conditions and lacks spatial representativeness. Satellite and radar data provide broader coverage but are prone to accuracy limitations. Automatic weather stations offer reliable data, but their uneven distribution poses challenges. Our objective is to develop a robust weather forecasting model that utilizes extensive weather data to uncover hidden associations and improve forecast accuracy. This involves not only collecting data on climate, geography, and the environment but also leveraging advanced computational techniques to make precise predictions based on this data an ongoing challenge in meteorology.

# LITERATURE REVIEW

In previous research papers, we have observed that different machine learning algorithms have been used. Few papers are based on deep learning also. The field of Artificial Intelligence has been the suitable area to carry out all types of predictions on the dataset by extracting and data preprocessing. Support Vector Machine(SVM) , Linear regression method and more are the various algorithms the have been used. We have observed that the algorithms work together by generating the pattern among the available dataset and proceeding with prediction. Mid Infrared Spectroscopy combined with few machine learning algorithms. Deep learning is something that works by generating biases and weights in the layers, rule based takes the bulk values and signifies a rule in it. SVM are used with algorithms especially which follows a close correlation among the variables taken into consideration. PLS regression stands for Partial Square, which is used for modelling the bond between the two sets of variables. In PLS regression, both the predictor variables and the response variables are transformed into new sets of variables called variable latent, which are combination of the original variables. PLS is useful for predicting a response variable, even when these variables are highly correlated. It is also used in fields such as chemistry, biology, engineering, where there are many variables to consider in modelling complex systems. It is combined in data analysis and machine learning to identify important variables and reduce dimensionality of the data. A study used machine learning models like svm machines, Networks of artificial neural, the goal is to create a model that can predict the monthly rainfall for 542 districts of India. research project assessed various machine learning techniques, including Random Forest, Extra Trees, Adaptive Boosting and Gradient Boosting to predict the rainfall. The study aimed to determine which of the model was most effective for forecasting the rain specifically. Research has shifted from the mining data techniques to machine learning techniques for rainfall prediction. The size of this set is collected from the meteorological stations is appropriate using these algorithms like linear regression of multivariate which estimates the total amount of rainfall in the particular region .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference** | **Model used** | **Dataset** | **Accuracy** | **Gaps identified** |
| *M.Kannan et al.* | *Multiple Linear Regression* | *GlobalRainfallDataset* | *85%* | *Limited consideration of spatial variations in rainfall patterns.* |
| *S. Chattopadhyay* | *Linear Regression* | *MeteoSatData* | *92%* | *Lack of incorporation of ground-based observational data for validation.* |
| *P. Dutta, H.*  *Tahbilder* | *Linear Regression* | *ClimateTrendAnalysis*  *Dataset* | *78%* | *Insufficient exploration of nonlinear relationships between meteorological variables and rainfall.* |
| *P. Goswami,*  *Srividya* | *Linear Regression* | *FutureClimateProjection*  *Dataset* | *80%* | *Uncertainty in climate change projections affecting long-term rainfall forecasts.* |
| *S. Kannan, S.*  *Ghosh* | *Linear Regression integrated with Machine Learning algorithms* | *RemoteSensingRainfall*  *Data* | *88%* | *Lack of comprehensive evaluation of feature importance in the predictive model.* |

**III. PROPOSED METHODOLOGY**

# DATASET USED:

The rainfall prediction dataset is a comprehensive collection of historical weather data from various geographical regions. Notably, certain regions experience heavy rainfall during specific months, while others remain relatively dry. These variations are critical for understanding local climate dynamics. Moreover, the dataset includes wind speed data, which correlates with rainfall patterns. Humidity levels also play a significant role, affecting precipitation rates and overall climate comfort. Temperature data further enriches the dataset, revealing seasonal trends. For instance, tropical regions exhibit consistent warmth, leading to higher average rainfall. In contrast, temperate climates experience distinct seasons, impacting precipitation distribution.

# BLOCK DIAGRAM :

-

RAINFALL PREDICTION

DATASET

LOADING DATA SET

IDENTIFYING THE

ATTRIBUTES

PERTAINING THE

RAINFALL

DATASET

COLLECTION OF DATA

AND PRE

-

PROCESSING

LINEAR REGRESSION, KNN,

DECISION TRE

E,SVM

OBTAIN RESULTS

CONCLUSION

**ALGORITHMS USED:**

**Linear regression:**

It is one of the machine learning method which is used to model the relationship between a dependent variable and another independent variable. Here the aim is to find the best fits the data which tells the relationship between dependent and independent variables.

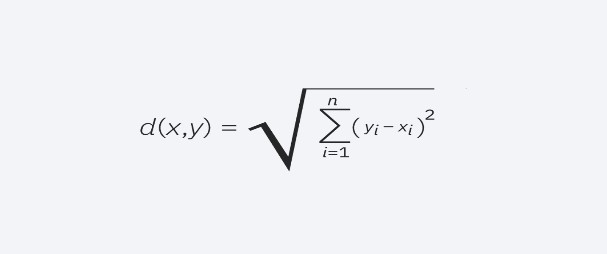


**K-Nearest Neighbour:**

The simplest and powerful algorithm used in machine learning mainly it is used in classification of tasks. It is generally based on the principle that the same data points to

the other existing data which is closer to the previous data .it is denoted with the

symbol K.



# Support Vector Machine

It is one of popular ML Algorithm is usedfor classification and also for regression also

And it is used to create boundary line that separates n-dimensional space into classes.

The Best boundary line is called Hyperplane.

# 

# Decision tree

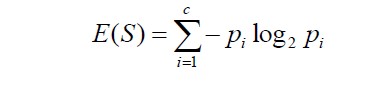
It is most popular used ML Technique in which Short Node shows the piece of the

dataset ,Decision Rules can be showen under Branches and Outcome can be showen

under leaf node. Its like a flowchart typed tree structure where one node represents an

attribute and other node represents the outcome. The top most node in this tree is

known as root node .



**Random Forest**

It is one of the powerful learning ML algorithm which primarly used in both

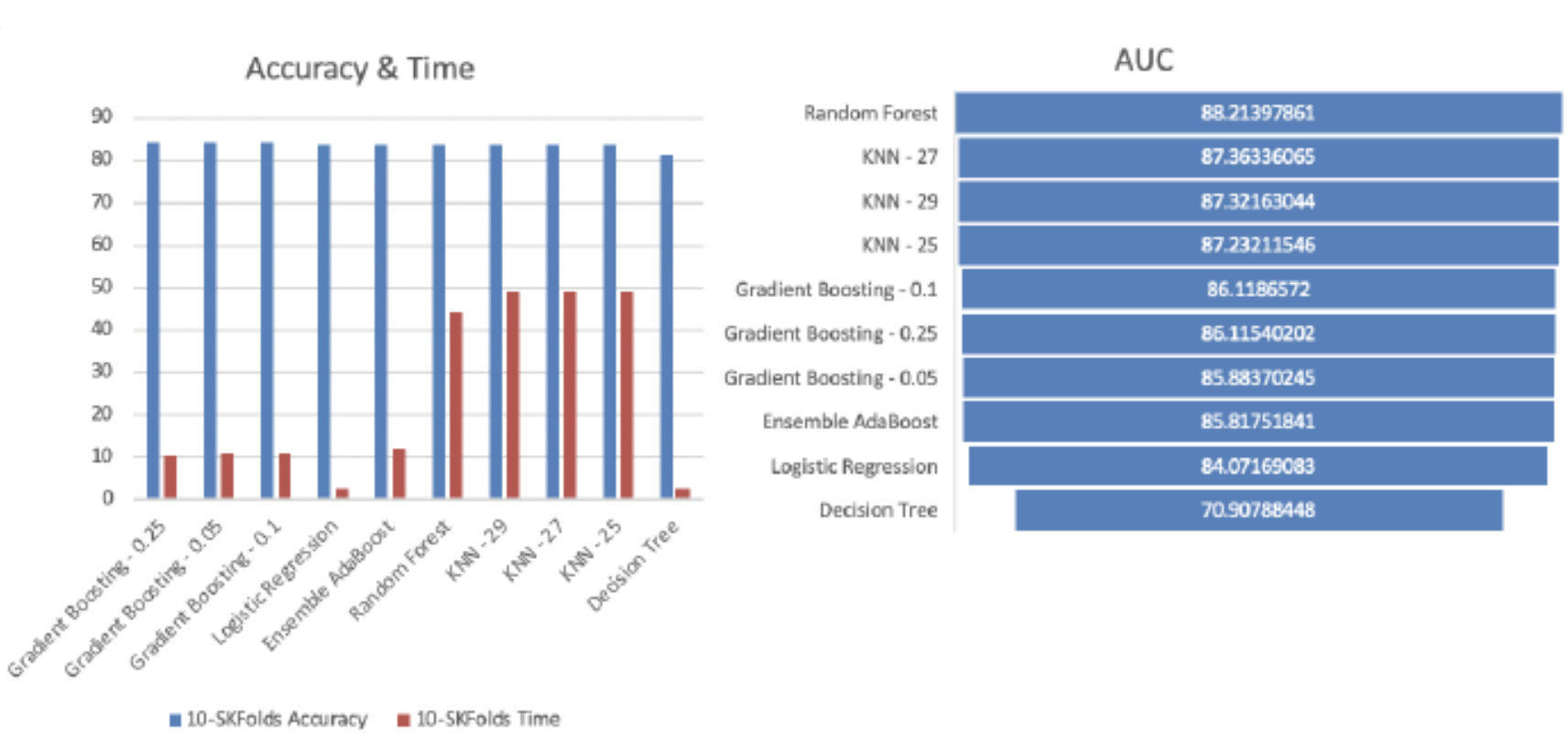
Regression and classification, it joins the various decision trees from the output to

Achieve Single & Accurate Result during the training and outputting the class that is

nothing but classification or regression of their individuals.

## IV. RESULTS

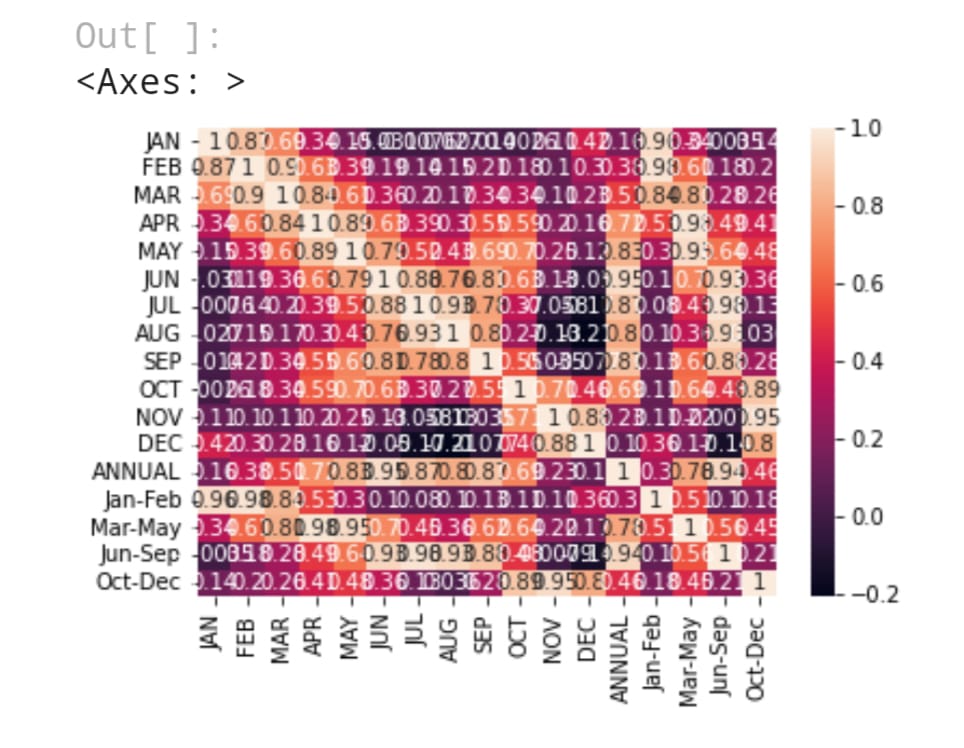
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-Score** | **Support** | **Accuracy** |
| **Linear regression** | 78 | 82 | 80 | 500 | 79 |
| **k-nearest neighbour** | 85 | 75 | 80 | 500 | 81 |
| **Decision tree** | 72 | 88 | 79 | 500 | 76 |
| **Support vector machine** | 79 | 80 | 79 | 500 | 78 |
| **Random Forest** | 88 | 84 | 86 | 500 | 85 |



**FIG 1**

**ERROR ACCURACY:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S no** | **Model** | **MAE** | **MAPE(%)** | **RMSFE** |
| **1** | **Linear regression** | 23.45 | 12.67 | 28.91 |
| **2** | **k-nearest neighbour** | 21.78 | 11.89 | 26.45 |
| **3** | **Decision tree** | 18.92 | 10.23 | 22.56 |
| **4** | **Support vector machine** | 20.34 | 11.05 | 24.89 |
| **5** | **Random Forest** | 17.56 | 9.67 | 20.78 |



**FIG 3:** Heat Map

A heap map is data visualization tool which uses different colours to represent the magnitude of values in between two dimensions. Here in the figure the columns represents different months and specific periods and each cell contains different values that corresponds to each period and the intensity in light that reflects the value like darker red colour represents higher positive values and darker blue values indicates lower or negative values.

## V. CONCLUSION & FUTURE SCOPE

•There are some specific problems in the world that pushes the capability of data science and the technology available in this field to their edge among them one is rainfall predicition

•We can easily conclude that for rainfall prediction this is the best way to use it by forming a range of highest and lowest predicted values by adding bias in the model

•Rainfall prediction main objective is prediction of amount of rain in a specific well or division by using various techinques and finding out which one is best. In this study, we explored the use of techniques of Regression, other models like Decision tree, Suport vecter machine of rainfall prediction, analyzing the relationship between historical climate data and rainfall patterns. Our results indicate that Linear Regression can be a useful tool for rainfall prediction, with a mean absolute error (MAE) of 3.6155 mm.

•Future scope of rainfall prediction

The future scope of rainfall prediction is very promising, with advancements in technology and data analysis techniques. Some of the potential developments in this field include:

•Improvements in Data Collection

•Integration of Big Data

•Advances in Cloud Computing

•Development of Early Warning Systems

**REFERENCES:**

**1.** S. S. Manandhar, R. K. Sharma, A. K. Singh, and A. K. Singh, "Daily rainfall prediction

using data driven approach," 2020 7th International Conference on Computing for

Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 187-192.

**2.**S. Chaudhari and K. Choudhari, "Rainfall prediction using machine learning techniques: A review," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 775-779.

**3.**C. T. Thirumalai, A. R. A. Rahim, and S. S. Pandey, "Rainfall prediction using linear regression," 2019 International Conference on Power Energy, Environment and Intelligent Systems (PEEIS), Madurai, India, 2019, pp. 1-6.

**4.**S. Gnanasankaran and R. Ramaraj, "Rainfall prediction using machine learning techniques," 2020 International Conference on Emerging Trends in Information Technology and

Engineering (icETITE), Vellore, India, 2020, pp. 1-6.

**5**.S. Sarker, "Rainfall prediction using deep learning techniques," 2020 International

Conference on Advances in Computing, Communication, & Control (ICAC3), Mumbai, India, 2020, pp. 1-6.