**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 Guidence of**

Mohammed Ali Shaik

Assist.Professor

**ABSTRACT :**

Rainfall prediction plays a crucial role in various domains, including agriculture, water resource management, and disaster preparedness. It is considered to be one of the important weather forecasting related research since rainfall heavily affects our nature and surroundings. Natural phenomenon such as flood, draught, weather indicators such as relative humidity, etc. are highly affected by rainfall. Leveraging machine learning techniques, particularly regression models, offers promise in improving accuracy. In recent years, researchers have explored machine learning models to enhance the accuracy of rainfall predictions. This study focuses on leveraging linear regression, along with other popular models, to improve rainfall forecasting

**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, satellite-derived rainfall data, and radar rainfall estimation. 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. Logistic Regression, Support Vector Machine, Naïve Bayes Classification, Linear regression and ridge regression etc. are the various machine learning 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. Artificial Neural Network inspired by the structure and function of the human brain. PLS regression stands for Partial Square regression, which is a statistical technique used for modelling the relationship 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 latent variables, which are linear combination of the original variables. PLS regression is useful for predicting a response variable from a large number of predictor variables, even when these variables are highly correlated. It is commonly used in fields such as chemistry, biology, and engineering, where there are many variables to consider in modelling complex systems. It is also used in data analysis and machine learning to identify important variables and reduce dimensionality of the data. A study used machine learning models like Support Vector Machines, Artificial Neural Networks, and Multiple Linear Regression to propose a rainfall prediction model that can predict the monthly rainfall for 542 districts of India ¹. Another study evaluated and compared several machine learning models such as Random Forest, Extra Trees, Adaptive Boosting, Gradient Boosting, Multilayer Perceptron, and Gaussian naïve Bayes for rainfall prediction . Research has shifted from data mining techniques to machine learning techniques for rainfall prediction. The size of the data set collected from meteorological stations is appropriate for using machine learning algorithms like multivariate linear regression to estimate the daily amount of rainfall in the 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:**

In this project Dogecoin price prediction and prediction, we use three approaches:

•Linear regression

* K-Nearest Neighbour
* Support Vector Machine
* Decision Tree
* Random Forest

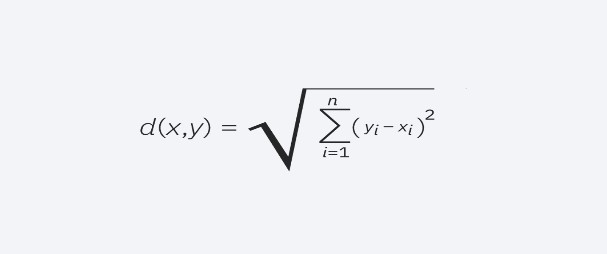
**Linear regression:**

Linear regression is a supervised machine learning method that is used by the Train Using AutoML tool and finds a linear equation that best describes the correlation of the explanatory variables with the dependent variable. This is achieved by fitting a line to the data using least squares. The line tries to minimiz the sum of the squares of the residuals. The residual is the distance between the line and the actual value of the explanatory variable. Finding the line of best fit is an iterative process.

**K-Nearest Neighbour:**

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.



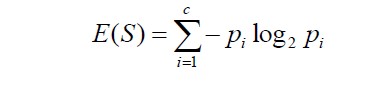
# Support Vector Machine

# 

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

# Decision tree

Decision trees are a nonparametric supervised learning method used for classification and regression. The deeper the tree, the more complex the decision rules and the fitter the model. Decision tree uses the tree representation to solve the problem. In which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. The primary challenge in the decision tree implementation is to identify the attributes. There are two popular attribute selection measures they are Entropy and Gini index.



**Random Forest**

Random Forest is an ensemble learning method that combines multiple decision trees.

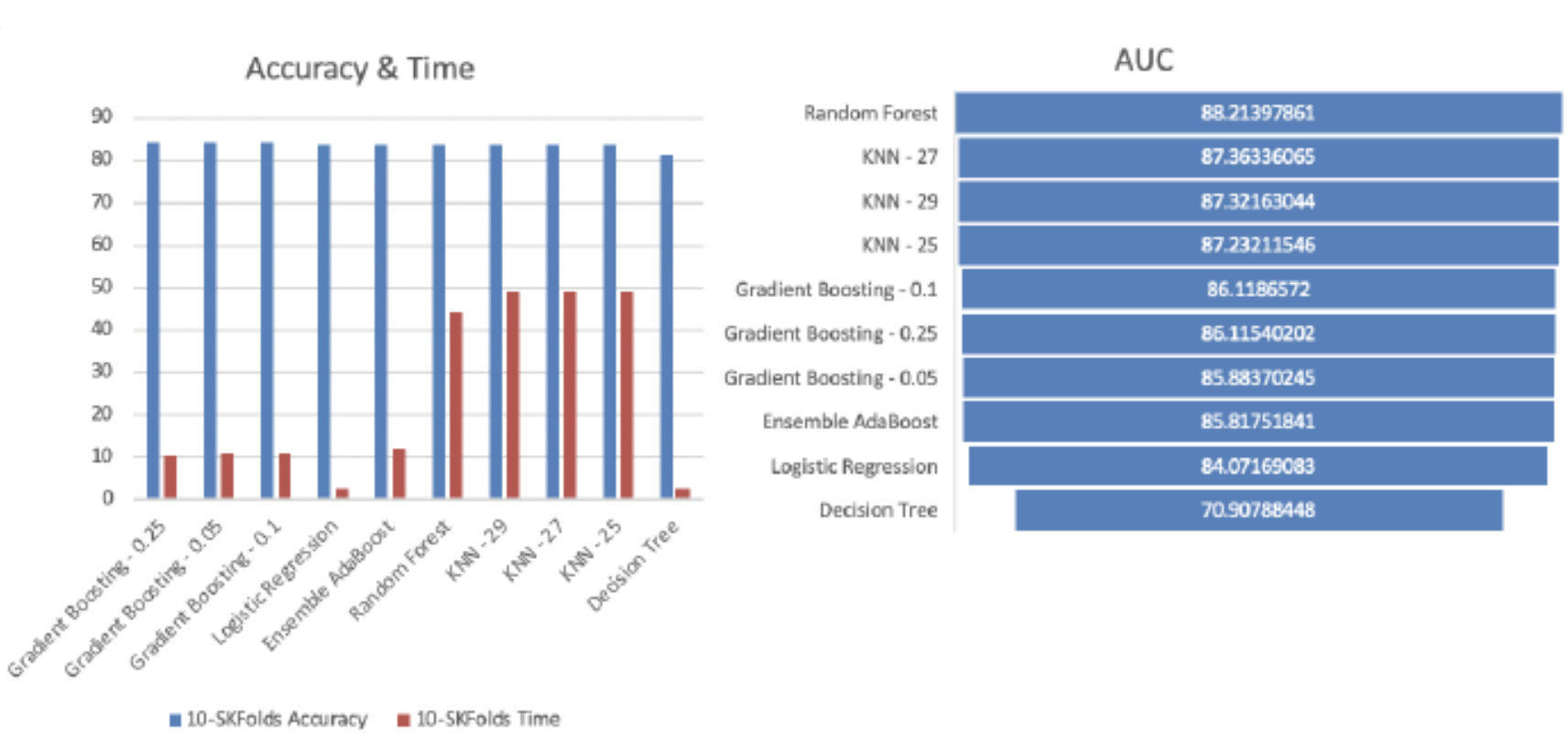
It reduces overfitting by averaging the predictions of individual trees. Each tree is trained

on a random subset of the data and features. Hyperparameters to tune include the number

of trees (estimators) and the maximum depth of each tree.

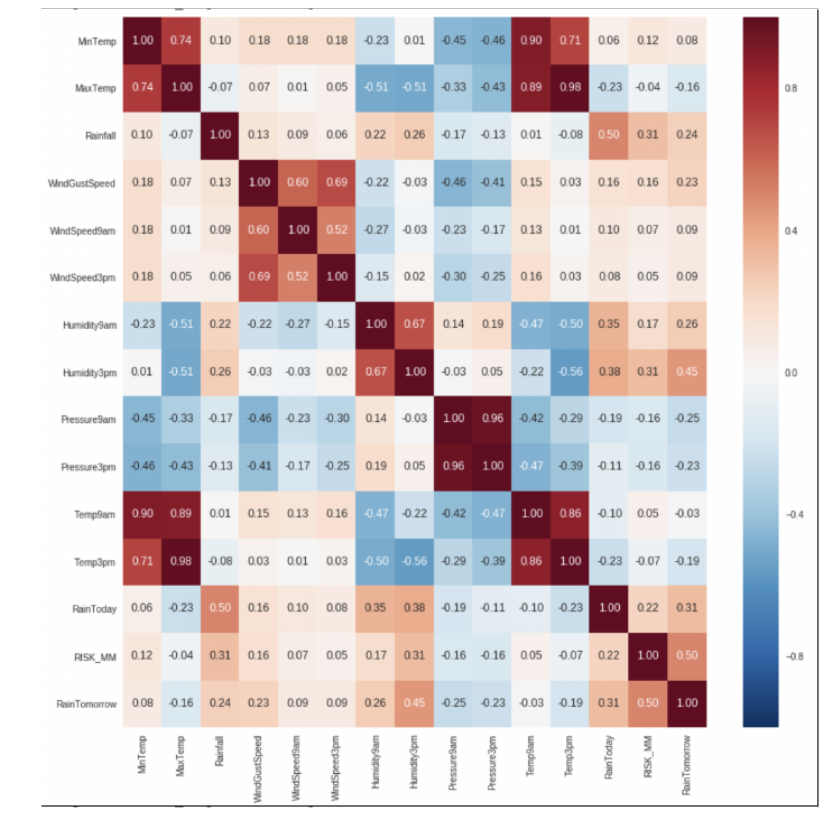
## IV. RESULTS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |



**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 |



Heat Map

## 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 Linear Regression , K nearest neighbour, Decision tree, Suport vector machine for 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

•In summary, the future of rainfall prediction looks bright, and with continued research and innovation, we can expect more accurate and reliable predictions that can help people and communities prepare for extreme weather events.

**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.**A. Garg and K. Kanchipuram, "Rainfall prediction using machine learning techniques,"

2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 780-784.

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

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