# Rainfall Prediction Using Machine Learning Techniques

**[1]Abhishek ML, [2]Harshavardhan V, [3]Sudesh Bhat, [4]Sujay J, [5]Vinay K M**
**Co-author. MAHENDRA KUMAR B,**

**[1,2,3,4,5] PG SCHOLORS, DEPT. OF MCA, DSCE**
**CA -ASST. PROF., DEPT. OF MCA, DSCE**

**Abstract –**

**This paper describes the development of a system to provide information on rainfall characteristics of rainfall and its prediction from the historical data sets, with a case study of Karnataka State, India. The analysis of rainfall year-wise is carried out. The developed system, evolved from the analysis of 50+ years' rainfall data collected from the Govt [1]. website. Prediction can be carried out by using various machine learning algorithms like linear regression, SVM, K NN method and decision tree algorithm out of which SVM is giving the highest efficiency.**

*Key Words:*Random Forest, Machine Learning, Support Vector Machine(SVM), Support Vector Regression(SVR), Decision Tree.

## I. INTRODUCTION

Rainfall is the most important climate variable that affects the crops in the tropical regions. People wish to know whether there would be rainfall in the coming months or year. To fulfil this requirement, one needs a forecast, well ahead of the commencement of the Rainy season. Such rainfall forecasts can be used by the Government in planning for the ensuing year. This work is an attempt to analyse the available data for development of a system for rainfall forecast [1]. Indian meteorological department provides forecasting data required for project. In this project we are planning to work on long term predictions of rainfall. The main motive of the project is to predict the amount of rainfall in a particular division or state well in advance. We predict the amount of rainfall using past data [2].

## II. RELATED WORKS

### Machine Learning Technique

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed [11]. Under Artificial Intelligence machine learning is the main area where a machine works like an intelligent human brain. Machine has to initially think and then learn like a brilliant man. As like a person in society learns from experiences and former data that it is exhibit to and according to that the machine takes decisions in upcoming events [1]. There are some methods of algorithms and learning in machine learning:

**1.Supervised Learning:** In this technique of machine learning few training datasets are given to the particular algorithm on the support of these datasets machine can analyse the input values and their output values.

**2. Unsupervised Learning:** In this method no output values are given to the algorithm, this method of learning is considered as unsupervised learning. We do not have any outcome value to estimate or predict.

**3. Reinforcement Learning:** To make specific decisions, the machine is trained

using this algorithm. In this method, the machine is trained continuously using trial and error methods by exposing it to the past dataset environment.

### III.    IMPLEMENTATION DATASET

The Indian Rainfall Prediction Dataset contains of 19 different attributes of 116-year data. The detailed description of the dataset is shown in Table [9]

| SLNO | ATTRIBUTES | ATTRIBUTE TYPE |
|------|------------|----------------|
| 1 | REGION | Nominal |
| 2 | YEAR | Numeric |
| 3 | JAN | Numeric |
| 4 | FEB | Numeric |
| 5 | MAR | Numeric |
| 6 | APR | Numeric |
| 7 | MAY | Numeric |
| 8 | JUN | Numeric |
| 9 | JUL | Numeric |
| 10 | AUG | Numeric |
| 11 | SEP | Numeric |
| 12 | OCT | Numeric |
| 13 | NOV | Numeric |
| 14 | DEC | Numeric |
| 15 | ANNUAL | Numeric |
| 16 | Jan-Feb | Numeric |
| 17 | Mar-May | Numeric |
| 18 | Jun-Sep | Numeric |
| 19 | Oct-Dec | Numeric |

**Table-1** Dataset Description

### DATA-PREPROCESSING

Data pre-processing is an important step of solving every machine learning problem. Datasets used with machine learning [9] may contain irrelevant, noisy data and missing parts. Thus data need to be clean by filling missing value.

### CLASSIFICATION TECHNIQUES

Here in this step we split the data into two separate tables training andtesting respectively. Training table is used for analysis and alteration purpose, whereas testing table is usedtotest the data in the final step.

### RESULTS AND EVALUATION

Monsoon prediction is clearly of great importance for India. Two types of rainfall predictions can be done; they are - Long term predictions: Predict rainfall over few weeks/months in advance. - Short termpredictions: Predict rainfall a few days in advance in specific locations [2]. We done our project using long term prediction.
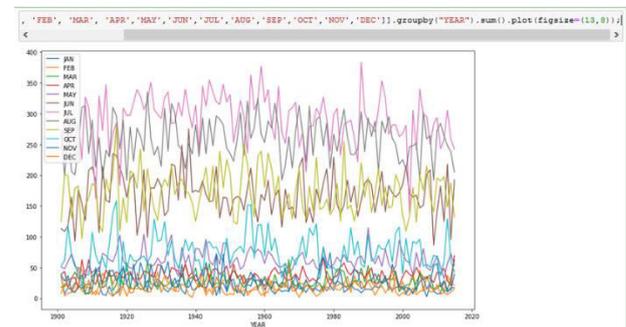
### MEAN ABSOLUTE ERROR

**mean absolute error (MAE)** is a measure of difference between two continuous variables. Assume $X$ and $Y$ are variables of paired observations that express the same phenomenon. Examples of $Y$ versus $X$ include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement [6].
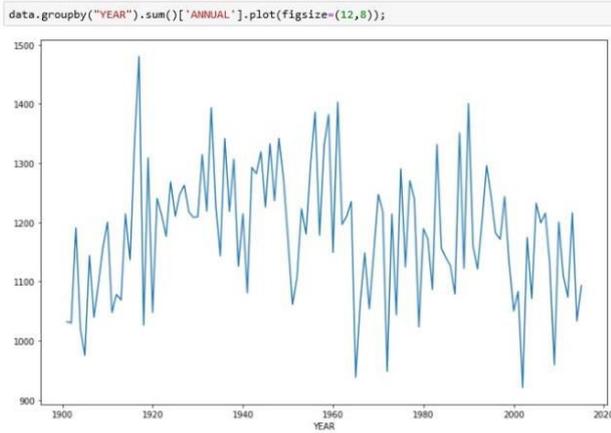
### Support Vector Regression(SVR)

The SVR uses the same principles as the SVM for classification, with only a few minor differences. First of all, becauseoutput is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities [7]. SVR allows float value in output.
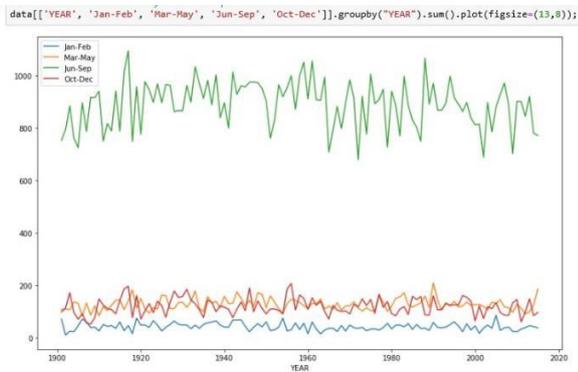
### SUPPORT



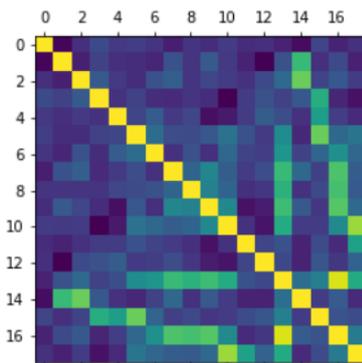-    Graph to demonstrate the appropriation of rainfall over months

```
data.groupby("YEAR").sum()['ANNUAL'].plot(figsize=(12,8));
```

- Shows distribution of rainfall over years.
- Observed high amount of rainfall in 1950s.

```
data[['YEAR', 'Jan-Feb', 'Mar-May', 'Jun-Sep', 'Oct-Dec']].groupby("YEAR").sum().plot(figsize=(13,8));
```

- Graphs clearly show that amount of rainfall in high in the months Jun, July, Aug, Sep which is monsoon season in India [2].

```
import matplotlib.pyplot as plt
plt.matshow(dd.corr())
plt.show()
```

- Correlation Matrix Plot

```
from sklearn.metrics import mean_absolute_error
from sklearn.svm import SVR
clf = SVR(gamma='auto', C=0.1, epsilon=0.2)
clf.fit(X_train, Y_train)
y_pred = clf.predict(X_test)
print (mean_absolute_error(Y_test, y_pred))
```

95.90203424666683

## IV.    PREDICTION ALGORITHM

There are various methods for prediction, which include,

- Linear Regression
- DecisionTree
- Random Forest
- Support Vector Machine(SVM)

### Linear Regression

**Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression [3].

### DecisionTree

Decision Tree Analysis is a tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks [4].

### Random Forest

Random forests, also known as random decision forests, are a popular ensemble method that can be used to build predictive

models for both classification and regression problems. Ensemble methods use multiple learning models to gain better predictive results — in the case of a random forest, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer [5].

### Support Vector Machine(SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries [12]. They were extremely popular around the time they were developed in the 1990s and continue to be the go-to method for a high-performing algorithm with little tuning[10].

### V. CONCLUSION

The prediction will help the farmers to choose whether the particular crop is suitable for specific rainfall and crop price values. This approach is to increase the net yield rate of the crop, based on rainfall.

### VI. REFERENCES

[1] www.ijritcc.org

[2] Submitted to Manipal University

[3] www.geeksforgeeks.org

[4] www.hackerearth.com

[5] Submitted to Harrisburg University of Science and Technology.
[6] en.wikipedia.org
[7] Jigar Patel, Sahil Shah, Priyank Thakkar, K Kotecha. "Predicting stock market index using fusion of machine learning techniques", Expert Systems with Applications, 2015
[8] data.sarawak.gov.my
[9] www.irjet.net
[10] www.ijert.org
[11] www.expertsystem.com
[12] www.internationaljournalssrg.org