# SYNOPSIS REPORT ON

**RAINFALL PREDICTION USING LINEAR REGRESSION**

## Introduction to project:

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way. This prediction mainly helps farmers and also water resources can be utilized efficiently. Rainfall prediction is a challenging task and the results should be accurate. There are many hardware devices for predicting rainfall by using the weather conditions like temperature, humidity, pressure. These traditional methods cannot work in an efficient way so by using machine learning techniques we can produce accurate results. We can just do it by having the historical data analysis of rainfall and can predict the rainfall for future seasons. We can apply many techniques like classification, regression according to the requirements and also we can calculate the error between the actual and prediction and also the accuracy. Different techniques produce different accuracies so it is important to choose the right algorithm and model it according to the requirements.

## Purpose(need) of System

1. **Proposed System:**

The predictive model is used to prediction of the precipitation. The first step is converting data in to the correct format to conduct experiments then make a good analysis of data and observe variation in the patterns of rainfall. We predict the rainfall by separating the dataset into training set and testing set then we apply different machine learning approaches and statistical techniques and compare and draw analysis over various approaches used. With the help of numerous approaches, we attempt to minimize the error.

**Dataset Description:** The dataset consists of the measurement of rainfall from year 1901-2015 for each state.

• Data consists of 19 attributes (individual months, annual, and combinations of 3 consecutive months) for 36 sub divisions.

• The data is available only from 1901 to 2015 for some of the subdivisions.

• The attributes are the amount of rainfall measured in mm.

As the dataset is very large, feature reduction is done so that it improves the accuracy, reduces the

computation time and also storage. Principal Component Analysis (PCA) is a technique of extracting necessary variables from a huge set of variables. It extracts low dimensional set with a motive to capture the maximum amount of information. With few variables, visualization becomes more significant.

## Project Overview. (Functionality of Models)

The machine learning algorithm called linear regression is used for predicting the rainfall using important atmospheric features by describing the relationship between atmospheric variables that affect the rainfall. The correlation study is conducted, and identified solar radiation, perceptible water vapor, and diurnal features are important variables for daily rainfall prediction using a data-driven machine learning algorithm. The future work identified by Manandhar et al. is studying the impact of using different atmospheric features using a larger data set. The researches address the relationship between independent and dependent features to identify which features impact the rainfall to rain or not to rain. The amount of daily rainfall was not found or addressed in this research, it may reduce the performance of the system. Tharun et al. performed the accuracy measure of the comparative study of statistical modeling and regression techniques (SVM, RF & DT) for rainfall prediction using environmental features. According to the result of the study, the regression techniques of rainfall prediction outperformed the statistical modeling. The experimental result showed that the RF model performed and predicted accurately than the SVM and DT. Hence, rainfall prediction is accurate, it shows high performance in machine learning models than the traditional models. This research used different machine learning techniques rather than statistical methods to predict daily rainfall amounts.

## Technology and Platform used:

**Tools/Technology:** Python

**Platform:** Jupyter

**Packages**: numpy, pandas.

**Libraries**: sklearn, matplotlib

**Modelling:** Machine learning (Linear Regression model

## Hardware & Software Requirements

2.1 At Server Side:

**Hardware Configuration**

|  |  |
| --- | --- |
| **Name** | **Details** |
| **Processor** | **Intel Pentium IV or more** |
| **RAM** | **At least 128 MB RAM is required (formally 4 GB**  **RAM)** |
| **Hard Drive** | **1 TB** |

**Software Environment**

|  |  |
| --- | --- |
| **Name** | **Details** |
| **Operating System** | **Windows 10,11** |
| **Language Used** | **Python Programming** |
| **Browser** | **Localhost (Chrome)** |
| **Tool** | **Jupyter Notebook** |

## Future Scope of Project:

Rainfall forecasting is useful in avoiding floods, which saves lives and property. It also aids in the management of water resources. Rainfall data from the previous year assists farmers in better managing their crops, resulting in increased economic growth for the country.