**TREND ANALYSIS AND PREDICTION OF RAINFALL IN TAMILNADU**

**A REPORT**

**JCS1731: PROJECT WORK PHASE - 1**

**IV YEAR / VII SEM**

**R2019**

***Submitted by***

|  |  |
| --- | --- |
| **SAFEENA THALATH** | **[130719104065]** |
| **SIVASANKARI R** | **[130719104079]** |
| **SHANMUGA PRIYA S** | **[130719104072]** |

***in partial fulfillment for the award of the degree***

***Of***

**BACHELOR OF ENGINEERING**

***in***

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**JERUSALEM COLLEGE OF ENGINEERING**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

NBA & NAAC ACCREDITED INSTITUTION

Velachery Main Road, Narayanapuram, Pallikaranai, Chennai - 600100

**DECEMBER 2022**

## BONAFIDE CERTIFICATE

Certified that this project report **“TREND ANALYSIS AND PREDICTION OF RANIFALL IN TAMILNADU”** is the bonafide work of **SAFEENA THALATH (130719104065), SIVASANKARI R (130719104079), SHANMUGA PRIYA S (130719104072)** who carried out the project work under my supervision.

|  |  |
| --- | --- |
| SIGNATURE | SIGNATURE |
| Dr.G.SASIKUMAR, M.E.,Ph.D., | Dr.T. DHANALAKSHMI, M.E.,Ph.D., |
| HEAD OF THE DEPARTMENT | Assistant Professor, |
|  |  |
| Department of ComputerScience and Engineering, | Department of ComputerScience and Engineering, |
| Jerusalem College of Engineering, Pallikaranai, Chennai – 600 100. | Jerusalem College of Engineering, Pallikaranai, Chennai – 600 100. |

## Submitted for the University examination held on \_\_\_\_\_\_\_\_\_\_\_\_\_.

**INTERNAL EXAMINER EXTERNAL EXAMINER**

# ABSTRACT

Rainfall is a key input to the hydro logical cycle, and shows temporal variability over the land surfaces. Agriculture is dependent on rainfall but occurrence of rainfall is unpredictable and erratic which produces an adverse effects on the cropping system and also causes negative effects on natural water resources . In India, rainfall pattern in last century during different seasons indicate decreasing tendency in the summer monsoon rainfall and increasing trend in the rainfall during pre-monsoon and post-monsoon months .Globally water resources of any region has become a major concern for planning and development of projects such as sustainable development of agricultural and food production projects, effective water resource management techniques and erosion and flood estimation and control methodologies etc. Rainfall forecasting has gained utmost research relevance in recent times due to its complexities and persistent applications such as flood forecasting and monitoring of pollutant concentration levels, among others. Prediction of time series data in meteorology can assist in decision-making processes carried out by organizations responsible for the prevention of disasters. Rainfall data of 30 years (1990-2019) is collected and trend analysis is performed using Mann-Kendall and Sen’s slope . Further prediction of rainfall is performed based on Recurrent Neural Network (RNN). It is essential to capture the true scenario of rainfall patterns over the years and predict for the future years to prepare in well advance and formulate effective agricultural and planning according to that.

**ACKNOWLEDGEMENT**

We extend our warmest gratitude to **Dr. M. Mala,** Chairperson, Jerusalem College of Engineering for her enduring support.

We express our sincere thanks to **Dr. Ramesh S, Ph.D.,** Principal, Jerusalem College of Engineering for his kindness, which enabled us to do this project.

We express our sincere thanks to **Dr. G. Sasikumar, Ph.D.,** Professor and Head, Department of Computer Science and Engineering, Jerusalem College of Engineering for his support throughout the project.

We would like to take this opportunity to express our sincere thanks to our supervisor, **Dr. T. Dhanalakshmi, M.E., Ph.D.,** AssistantProfessor, Department of Computer Science and Engineering, Jerusalem College of Engineering for her valuable guidance, inspirations and technical support throughout the project.

We express our gratitude to our project coordinators, **Mrs. H. Mercy, M.E.,** Associate Professor and **Mrs. S. Vinitha, M.E.,** Assistant Professor, Department of Computer Science and Engineering, Jerusalem College of Engineering for their valuable guidance and support.

We thank all the faculty and supporting staff of the Department of Computer Science and Engineering, Jerusalem College of Engineering, for their co-operation and assistance in the successful completion of the project.

**SAFEENA THALATH (130719104065)**

**SIVASANKARI R (130719104079)**

**SHANMUGA PRIYA S (130719104072)**

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE**  **NO.** |
|  | **ABSTRACT** | **iii** |
|  | **LIST OF FIGURES** | **vii** |
|  | **LIST OF TABLES** | **viii** |
| **1.** | **INTRODUCTION** | 1 |
|  | 1.1 General | 1 |
|  | 1.2 Monsoons | 2 |
|  | 1.3 Trend Analysis | 3 |
|  | 1.4 Summary | 4 |
| **2.** | **LITERATURE SURVEY** | 5 |
|  | 2.1 General | 5 |
|  | 2.2 Related Works | 5 |
|  | 2.3 Summary | 10 |
| **3.** | **SYSTEM ANALYSIS** | 11 |
|  | 3.1 General | 11 |
|  | 3.2 Existing System | 11 |
|  | 3.3 Proposed System | 12 |
|  | 3.4 Block Diagram | 13 |

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | 3.5 Software Requirements | 13 |
|  | 3.6 Hardware Requirements | 14 |
|  | 3.7 summary | 14 |
| **4.** | **SYSTEM DESIGN** | **15** |
|  | 4.1 General | 15 |
|  | 4.2 Collection of data | 15 |
|  | 4.3 Module-1 - Description | 15 |
|  | 4.4 Module-2 - Description | 19 |
|  | 4.5 UML Diagrams | 20 |
|  | 4.5 Conclusion | 23 |
| **5.** | **IMPLEMENTATION & RESULTS** | **24** |
|  | 5.1 General | 24 |
|  | 5.2 Implementation | 24 |
|  | 5.3 Summary | 29 |
| **6.** | **CONCLUSIONS AND WORK SCHEDULE FOR PHASE II** | 30 |
|  | 6.1 Summary of Phase - 1 | 30 |
|  | 6.2 Conclusion | 30 |
|  | **REFERENCES** | 31 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE TITLE** | **PAGE NO.** |
| 1 | Block Diagram | 13 |
| 2 | Mann - Kandell Workflow | 18 |
| 3 | Sen’s Slope Estimator Workflow | 20 |
| 4 | Use Case Diagram | 20 |
| 5 | Sequence Diagram | 21 |
| 6 | Class Diagram | 21 |
| 7 | Activity Diagram | 22 |
| 8 | Architecture Diagram | 23 |
| 9 | South-West | 26 |
| 10 | North-East | 27 |
| 11 | Hot weather | 27 |
| 12 | Winter | 28 |
| 13 | Whole year | 29 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **TABLE NO.** | **TABLE NAME** | **PAGE NO.** |
| 1 | Statistical Analysis of Rainfall data along with Mann-Kendall Trend and Sen's Slope | 25 |

### CHAPTER 1

### INTRODUCTION

**1.1 GENERAL**

### 

### Tamil Nadu, located in southeast peninsular India, receives the major part of its annual rainfall during the northeast monsoon season (the three month period from October to December) and southwest monsoon season (the four month period from June to September). Coastal Tamilnadu receives about 60% of its annual rainfall and interior Tamilnadu receives about 40-50% of its annual rainfall during northeast monsoon. In comparison with Indian summer monsoon, the Northeast monsoon is characterized by limited aerial extent and average lesser rainfall amount. During northeast monsoon season, Tamilnadu generally receives rainfall due to the formation of tough of low, cyclonic circulation, easterly waves, low pressure area, depression and cyclonic storm over Bay of Bengal, because, the northeast monsoon season is the major rainy season. The vicissitudes of the rainfall of Tamilnadu state has led to considerable and widespread interest among the public, farmers as well as in government circles in recent years, in view of the frequent failure of northeast monsoon rainfall over Tamilnadu.

### The variability of then tamilnadu northeast and southwest monsoon affects the agriculture food grain production, industry and generation of hydroelectric power, causing severe strain on the national economy. Large parts of the district are severely affected due to deficit monsoon rainfall. Importance of monsoon to Indian economy and as major global circulation parameter has motivated many scientists to study the variability of the tamilnadu monsoons, in the past. Therefore, a study, for the area, is essential to capture the true scenario of rainfall pattern over the years and predict for the future years. The primary source of livelihood for majority of people living in this area is purely agro-based and growing of all the major crops is mainly dependent on

### timely occurrence of rainfall. Irrigation facilities also depend, directly (canal) or indirectly (groundwater), heavily on sufficient amount of precipitation. As the region experiences, from time to time, both excessive as well as insufficient rainfall, agricultural production are often been seriously affected. A proper study on trend of rainfall is therefore needed to predict such situations in well advance and formulate effective agricultural and planning according to that.

### 1.2 MONSOONS

### 

### Monsoon is traditionally defined as a seasonal reversing wind accompanied by corresponding changes in precipitation, but is now used to describe seasonal changes in atmospheric circulation and precipitation associated with the asymmetric heating of land and sea. Usually, the term monsoon is used to refer to the rainy phase of a seasonally changing pattern, although technically there is also a dry phase

### 1.2.1 MONSOON SEASON PROCESS IN TAMILNADU

### 

### The Tamilnadu state has three distinct periods of rainfall: advance rainfall; rainfall from the tropical cyclones emerging in the Andaman Islands during the Retreat of Monsoons (October–November); and the North East monsoon during the months of October–December, with dominant northeast monsoon winds from the western disturbances emerging over the Mediterranean Sea. The dry season is from February to early June. Tamilnadu has rain during the monsoon season due to the southwest trade winds which blow towards the northern hemisphere. Tamilnadu receives rainfall in the winter season due to northeast trade winds. The normal annual rainfall of the state is about 945 mm (37.2 in) of which 48% is through the North East monsoon, and 32% through the South West monsoon. Since the state is entirely dependent on rains for recharging its water resources, monsoon failures lead to acute water scarcity and severe drought. Tamilnadu is classified into seven agro-climatic zones: north-east, north-west, west, southern, high rainfall, high altitude hilly, and Cauvery Delta (the most fertile agricultural zone).

### 1.3 TREND ANALYSIS

### 

### A trend can be defined as long-term change in the mean level. Although the subject area of climate change is vast, the changing pattern of rainfall is a topic within this field that deserves urgent and systematic attention, since it affects both the availability of freshwater and food production Hence, it is an important parameter to detect climate change and asses its impact on various sectors including agriculture, industry etc.

Prediction of rainfall remains a severe concern and has grabbed the attention of industries, governments, risk management entities, and the scientific community as well. One main complex meteorological phenomenon is Rainfall. Rainfall is a random event, and the cause of its occurrence is very complex. Regarding weather predictions, among the climatic factors, even under similar weather conditions, there is a possibility that it might rain at this moment but not at another moment. Therefore, prediction becomes crucial to gain knowledge of the atmosphere’s condition.

Predictive Analytics is an innovative analytics scheme to perform such forecasting and predict actions on future events and happenings using the previous datasets. It is a factor of climatic condition through which numerous human activities such as power generation, construction, agricultural production, tourism and forestry among others are acted. Traditional forecasting of weather conditions reminds about meteorologists making predictions in view of their experience by sitting with weather charts spread in front of them. This experience is the knowledge gathered from years of observations and from background on weather theory. To this extent, rainfall prediction is vital since this variable is the one that has maximum correlation with adverse natural events like landslides, flooding, mass movements and avalanches. These incidents have a acted society for years. This appears to be the main motivation for the evolution of numerical weather prediction and weather prediction by machine learning algorithm.

### 1.4 SUMMARY

### Tamil Nadu is a state that entirely dependent on rains for recharging its water resources thus it is important to analyse and predict rainfall patterns for better assistance in case of monsoon failures which lead to acute water scarcity and severe

### drought. This study helps us understand rainfall patterns in Tamil Nadu

### 

### CHAPTER 2

### LITERATURE SURVEY

**2.1 GENERAL**

This chapter gives the overall description of the reference papers through which the problems of existing methodology are identified. Also the methods to overcome such problems can be identified.

**2.2 RELATED WORKS**

Lot of research work has been carried out in providing service at home, some of the recent works are listed below.

**Author:** Shyam Lochan Bora, Kalyan Bhuiyan, Partha Jyoti Hazarika3, Junmi Gogoi and Kuldeep Goswami

**Title:** Analysis of rainfall trend using non-parametric methods and innovative trend analysis during 1901–2020 in seven states of North East India

### Year: 2022

**Description:**

### The paper presents the analysed variability and trends in annual as well as seasonal rainfall in the seven states of North East India for the period 1901–2020, using non-parametric tests like Mann–Kendall, trend-free pre-whitening Mann–Kendall, modified Mann–Kendall (MMK), as well as using the innovative trend analysis (ITA). The study revealed the variabilities in annual and seasonal rainfall in these seven states. In most cases, the results of all the tests were identical. However, significant differences were observed in the case of post-monsoon rainfall of Assam and Meghalaya, pre-monsoon rainfall of Arunachal Pradesh, Mizoram and Tripura, as well as in winter rainfall of Arunachal Pradesh and monsoon rainfall of Tripura.

**Author:** Lince Rachel Varghese

**Title:** A Time-series based Prediction Analysis of Rainfall Detection

### Year: 2020

**Description:**

This paper presents a time series analysis is used to extract the trends in seasonal rainfall and its effect is analysed. A time series forecasting is used to predict the future rainfall from 2020 to 2030 of all districts of Kerala. Data mining, data analysis can be used on meteorological data to find hidden patterns inside the data. Here the monthly rainfall data of 14 districts of Kerala form year 2008 to 2019 is taken into consideration. The data is classified as seasons consisting of 4 quarters. The purpose of this classification and analysis, is to find the climatic variations of districts that affect the Hevea (Rubber) cultivation. Hevea brasiliensis, the rubber tree is considered as the main source of Natural Rubber (NR) which is the most versatile raw material of nature, having multifarious uses Meteorological factors and soil fertility are the critical factors that influence the growth of Hevea.

**Author**: S. Poornima and M. Pushpalata

**Title:** Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units

### Year: 2019

**Description:**

### In this paper deep learning approach is carried out for rainfall prediction. Intensified LSTM based Recurrent Neural Network is constructed to predict rainfall. This approach is compared with other methods, namely, Holt–Winter, ARIMA, ELM, RNN and LSTM, in order to demonstrate the improvement of rainfall prediction in the proposed approach. For attaining good accuracy, it is necessary to consider previous datasets. Since the proposed Intensified LSTM network could hold large data in its memory and could avoid the vanishing gradient, it appears to show better accuracy in prediction. Although the improvement in accuracy is little in the proposed Intensified LSTM compared to existing LSTM model based RNN, the proposed prediction model preserves that accuracy for further epochs, while exhibiting reduced loss, RMSE and learning rate.

**Author**: Rawshan Ali, Alban Kuriqi, Shadan Abubaker and Ozgur Kisi

**Title:** Long-Term Trends and Seasonality Detection of the Observed Flow in Yangtze River Using Mann-Kendall and Sen’s Innovative Trend Method

### Year: 2019

**Description:**

### This study presences of possible trends in the annual, seasonal, maximum, and minimum flow of Yangtze River at Cuntan and Zhutuo stations in China for the period 1980 to 2015. The assessment was carried out using the Mann–Kendall trend test, and the innovative trend analysis, while Sen’s slope is used to estimate the magnitude of the changes. The results of the study revealed that there were increasing and decreasing trends at Cuntan and Zhutuo stations in different months. The result showed nonstationary increasing and decreasing flow trends over the region. The innovative trend analysis method has the advantage of detecting the sub-trends in the flow time series because of its ability to present the results in graphical format. The results of the study indicate that decreasing trends may create water scarcity if proper adaptation measures are not taken.

**Author**: Mislana, Haviluddin, Sigit Hardwinarto, Sumaryono, Marlon Aipassa

**Title:** Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan -Indonesia

### Year: 2015

**Description:**

This paper uses an Artificial Neural Network (ANN) with the Backpropagation Neural Network (BPNN) algorithm. In this experiment, the rainfall data were tested using two-hidden layers of BPNN architectures with three different epochs which were [2-50-10-1, epoch 500]; [2-50-20-1, with epochs 1000 and 1500]. The mean square error (MSE) is employed to measure the performance of the classification task. The experimental results showed that the architecture [2-50-20-1, epoch 1000] produced a good result with the value of MSE 0.00096341. Furthermore, BPNN algorithm has provided a good model to predict rainfall in Tenggarong, East Kalimantan - Indonesia.

**Author**: Nan Gu, Ding sheng Wan

**Title:** Trend Analysis of Extreme Rainfall Based on BP Neural Network

### Year: 2010

**Description:**

### In this paper data mining and using the basic principles of artificial neural networks to establish an average extreme rainfall prediction model which is based on the BP neural network This model only uses the extreme precipitation indexes as the factors to predict the average extreme rainfall in the coming year. The model combined with stepwise regression to select input vectors and used the Bayesian regularization method to further improve the generalization ability, thereby increasing the forecast accuracy of the trend of extreme rainfall. It is proved that the model is indeed valid and reliable by experimenting on many years of daily precipitation data of two sites in the Yangtze River

### The BP network use PX5D, PFL90, PNL90, PXCDD four extreme precipitation indexes as the input vectors and the annual average extreme rainfall as the same output vector, use 4-9-1(4 nodes in input layer,9 nodes in hidden layer,1 node in output layer) network structure and Bayesian regularized training function, the forecast errors are all below 10 mm, that is ,it can be believed that the network effectively forecast the average extreme rainfall

**Author**: Fiaz Hussain, Ghulam Nabi, Muhammad Waseem Boota

**Title:** Rainfall Trend Analysis by Using the Mann-Kendall test & Sen's Slope Estimates: A Case Study of District Chawal Rain Gauge, Barani area, Northern Punjab Province for Pakistan.

### Year: 2015

**Description:**

### This paper focuses on a key climatic variable i.e. the precipitation for analyze changing trend of rainfall of a Barani Area Chakwal District located in Northern Punjab Province Pakistan. In this area agriculture is dependent on rainfall but occurrence of rainfall is unpredictable and erratic which produces an adverse effect on the cropping system and also causes negative effects on natural water resources. Rainfall data of 37 years (1977 to 2013) on daily basis was processed to find out the monthly, seasonally and annual rainfall variability (trend and slope magnitude) by using non parametric analysis. The Mann-Kendall (MK) Test for monotonic analysis of trend together with non-parametric Sen’s Slope Estimator was used to estimate the magnitude of trend for time series data. Monthly, seasonally and annually precipitation trends has been drawn using 37 years daily data.

**Author**: Salma, S., S. Rehman, M. A. Shah

**Title:** Rainfall Trends in Different Climate Zones of Pakistan.

### Year: 2012

**Description:**

In this paper, the study was conducted across the country to assess the rainfall trend in different climate zones of Pakistan over the past three decades. For this purpose data set comprising 30 years for the period 1976 to 2005were acquired from 30 meteorological observatories from different parts of the country. The whole data was analyzed through Analysis Of Variations (ANOVA) along Dunnett T3 test. The result has shown a decreasing trend (-1.18mm/decade) all over the country, which may be attributed to the presence of drought period during 1998- 2001. Stations located in different zones of the country mainly from North, North West, West and Coastal areas respectively show overall significant decreasing trend whereas plain areas and South West of the country have been observed with no significant trend. Adverse consequences of the rainfall have already been observed in Pakistan in the form of droughts and super floods which have badly affected human settlements, water management and agriculture sectors.

**2.3 SUMMARY**

This chapter describes all the papers which contain existing system of analysis of trend in rainfall data and prediction of rainfall using RNN technologies.

### 

### CHAPTER 3

### SYSTEM ANALYSIS

**3.1 GENERAL**

The system analysis focuses on Existing System, Proposed System, Block diagram, Hardware Requirements and Software Requirements. The technologies in trend analysis is discussed existing and proposed section.

### 3.2 Existing System

### 

Several studies have been carried out in the past to analyse rain-fall trends in India using Mann–Kendall (MK) test. Although this test is being widely used for trend detection, it has some flaws regarding serial correlation in the time-series data. To overcome this problem modified Mann–Kendall (MMK) and trend free pre-whitening Mann–Kendall (TFPW–MK) tests have been introduced, which provide more reliable results than the MK test; however, they are also highly dependent on sample size and data distribution14. Therefore, innovative trend analysis (ITA) was introduced to solve the issues in the MK test15. Recently, Serenade *et al*.16 have reported that ITA is also dependent on sample size and data distribution. Despite being limited by these drawbacks, ITA has become popular due to its capability to delineate hidden trends while using other methods and thus going a step further in analysing the trends17–19. It may be noted that the use of such advanced techniques of trend detection is sporadic in NE India. Thus, it becomes significant to study the rainfall trends using non-parametric methods and the ITA method to better understand rainfall variability in the study region by comparing the results with those obtained using other techniques.

Other techniques used are temporal trend analysis which used to analyse the trend within or across user groups defined by specific time period(s) or change over time. Helpful in figuring relationships between user groups from different generations. Helpful in predicting future events based on those of the past. It has several drawbacks such as historical data may not be an accurate representation of trends and the trend may not be replicable

**3.3 Proposed System**

The proposed system uses the Mann Kendall test to analyse the trend over the region of Tamilnadu. This test has several advantages such as it does not assume the data to be distributed according to any particular rule, i.e., it does not require that the data be normally distributed. It is not affected by missing data other than the fact the number of sample points is reduced and hence might affect the statistical significance adversely. It is not effected by irregular spacing of the time points of measurement. It is not effected by the length of the time series. This analysis is then followed by Sen’s slope is used to estimate the magnitude of the changes. Further prediction of rainfall is done using neural network.

Neural networks are of two types, namely, Feed Forward Neural Networks (FFNN) and RNN. Neural networks employed in feed forward structure are not suitable for the prediction of rainfall since they do not take previous state into account. RNN on the other hand has shorter memory to hold long back status of the datasets. Hence, we go for RNN in combination with Long Short-Term Memory. The LSTM can hold a large number of previous datasets, thereby mitigating the vanishing gradient, which leads to better accuracy in prediction. The data is collected for the selected region and the trend is analysed for the region Then trend is analysed in detail by considering the various features such as temperature and humidity for the region. The analysed trend gives the reason for the increased or decreased trend of rainfall and further prediction of rainfall is done using the neural network as show in figure 3.1

### 3.4 Block Diagram

### 

FIGURE 3.1 BLOCK DIAGRAM

### 3.5 Software Requirements

### R Studio

### R Studio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. R Studio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to R Studio Server or R Studio Server Pro (Debian/Ubuntu, Red Hat/Cent OS, and SUSE Linux). R Studio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics. Pre-packaged distributions of R Studio Desktop are available for Windows, OS X, and Linux.

### Excel

### Microsoft Excel is one of the most popular applications for [data analysis.](https://www.simplilearn.com/data-analysis-methods-process-types-article) Equipped with built-in pivot tables, they are without a doubt the most sought-after analytic tool available. It is an all-in-one [data management](https://www.simplilearn.com/what-is-data-management-article) software that allows you to easily import, explore, clean, analyse, and visualize your data

### 3.6 Hardware Requirements

### Windows 11 (RAM-4GB)

#### 3.7 SUMMARY

This chapter gives an overview of the existing system and proposed system and the approaches carry out to implement the proposed model and the software and hardware requirements to implement the proposed model were briefed.

### CHAPTER 4

### SYSTEM DESIGN

### 4.1 GENERAL

System design is the most important and vital part of any project for the development purpose. It includes modules, modules description, formulae, algorithms and technologies to analyse or predict the result. In this chapter we give a detailed description of module 1 and module 2.

### 4.2Collection of data

### Rainfall data for 30 years (1990 – 2019) for the state of Tamilnadu was collected from India Meteorological Department (IMD). The data includes the seasonal rainfall in Tamilnadu. The observations are measured in milli meters (mm). For seasonal analysis, each year is divided into four climatic seasons viz. pre-monsoon (March-May), monsoon (June-September), post-monsoon (October-December) and winter (January-February).

### 4.3 Module -1 Description

**Trend Analysis**

### Statistically, trend is a significant change over time that is detectable by parametric and non-parametric procedures while trend analysis of a time series consists of the magnitude of the trend and its statistical significance. In this study statistical significance trend analysis was done by using Man- Kendall test while the magnitude of the trend was determined by the non-parametric Sen’s estimator method.

### Mann-Kendall Test

### This is a statistical method that is mostly used to check the null hypothesis of no trend versus the alternative hypothesis of the existence of a monotonic increasing or decreasing trend of hydro-climatic time series data. The non-parametric Mann-Kendall test is fit for those data series where the trend may be assumed to be monotonic (i.e. mathematically the trend consistently increasing and never decreasing or consistently decreasing and never increasing) and no seasonal or other cycle is present. MAKESENS performs two types of statistics depending upon the number of data values i.e. S – statistics is used if the number of data values are less than 10 while Z – statistics (normal approximation/distribution) for data values greater than or equal to 10.

### The statistic S is calculated as shown in equation (1)

### (1)

Where,

### and are annual values in years j and i, j > i respectively, n is the number of data points and (xj-xi ) is calculated using equation (2)

### (2)

### The mean of *S* is *E*(*S*) = 0 and the variance of *S* is given by

### Mann-Kendall variance (3)

### where *t*P is the number of ties for the *P*th value and *q* is the number of tied values. The second term in the variance formula is for tied censored data. The standardized test statistic *Z* is calculated by

### (4)

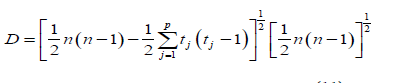
### To test for a monotonic trend at significance level, the null hypothesis of no trend is rejected if the absolute value of *Z* is greater than *Z*1–/2, which is obtained from the standard normal cumulative distribution tables21,22.

**Kendall’s Tau**

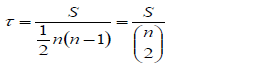
A statistic that is closely related to (ii) is Kendall’s tau (Kendall, 1975) defined by

 (5)

Where,

 (6)

When there are no ties in the data, (4) collapses to (5)

 (7)

### Due to relationship between tau and *s*in (3), the distribution of tau can be easily obtained from the distribution of *s* . It is a measure of correlation, and so measures the strength of the relationship between two variables. Kendall's tau is analogous to the correlation coefficient in regression analysis and has a range of -1 to +1.

**Mann - Kendell Workflow**

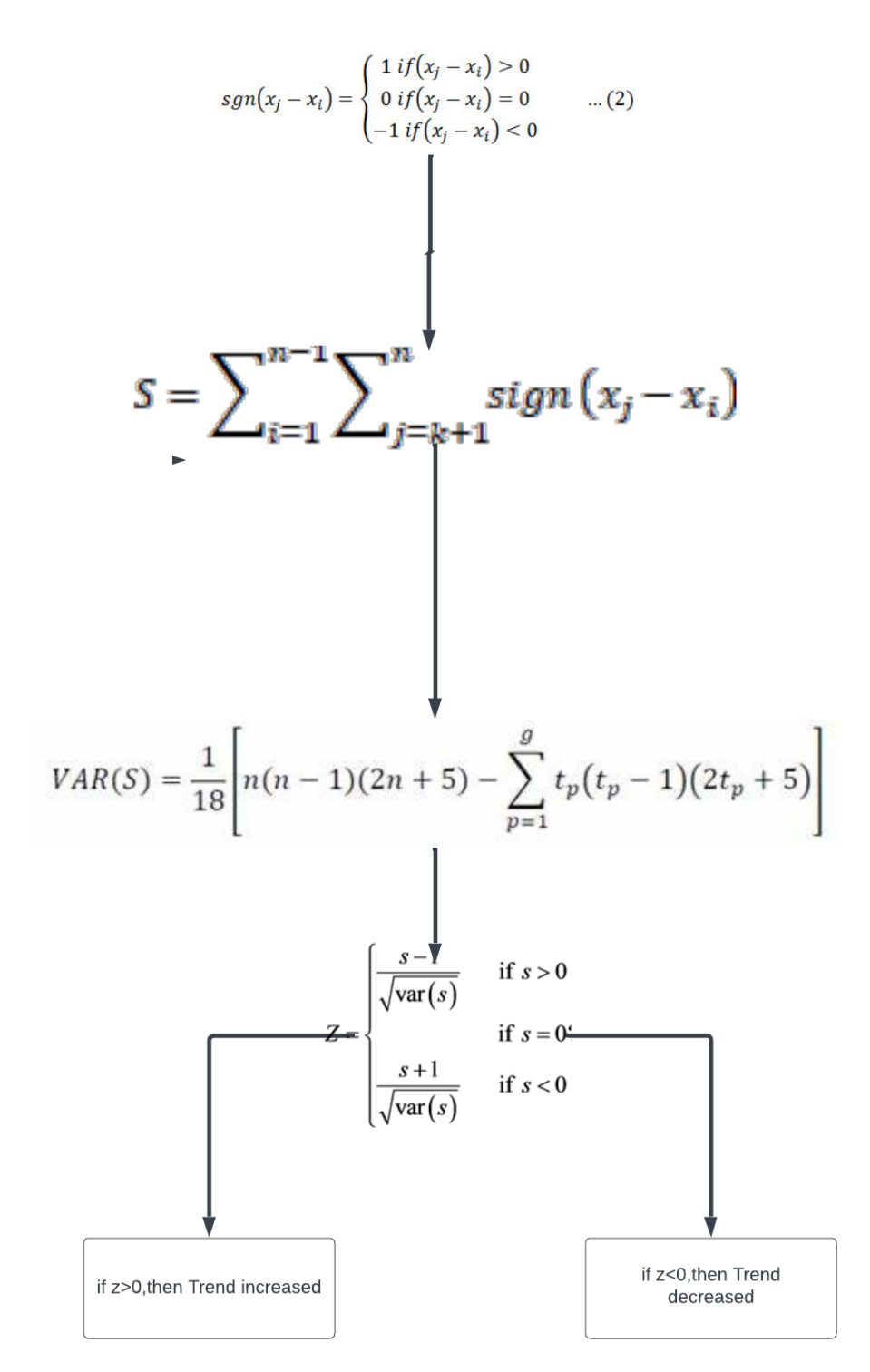


FIGURE 4.1 Mann - Kendall Workflow

### 4.4 Module 2 Description

### Sen’s Slope Estimator

### In this study, the magnitude of the trend in the time series is determined using a non-parametric method (Theil, 1950) known as Theil–Sen estimator or Sen’s slope estimator (Sen, 1968). Sen’s method assumes a linear trend in the time series and has been widely used for determining the magnitude of the trend in hydro-meteorological time series (Mondal et al., 2012; Jain et al., 2013; Kumar et al., 2014). By using this method, the true slope (change per unit time) can be estimated by using a simple non-parametric procedure. To derive an estimate of

### the slope Q in a linear model

### f (t)= Qt+ B (8)

### where is a constant, the slopes of all data pairs are calculated as

### (9)

### Where, and are data values at time j and k (j > k) separately. The median of these n values of Ti is represented by Sen’s slope of estimation (true slope) which is calculated using the equation

### (10)

Sen’s estimator is calculated using the above equation depending upon the value of n is either odd or even and then is computed using 100 (1 – α) % confidence interval using a non-parametric test depending upon normal distribution. A positive value of indicates an increasing (upward) trend while a negative value of representing a downward or decreasing trend of time series data. A positive value of indicates an upward (increasing) trend and a negative value indicates a downward (decreasing) trend in the time series.

**Sen’s Slope Estimator Workflow**

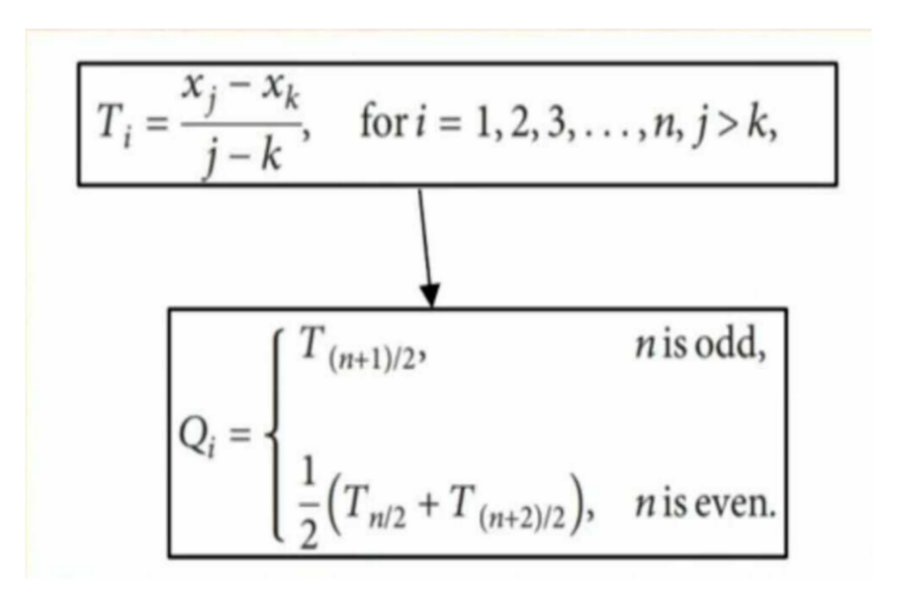


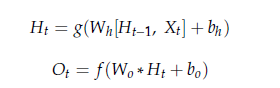
FIGURE 4.2 Sen’s Slope Estimator Workflow

**4.5 MODULE 3**

**Method For Future Forecasting**

Several probabilistic and deterministic methods like ARMA, ARIMA, SARIMA is usually employed to predict the hydrological and climatic datasets. These techniques have many drawbacks like serial correlation, non-linearity, and biases to predict the non-linear hydro-climatic data sets. Therefore, the newly developed artificial intelligence (AI) models can able to overcome this drawback. Hence, the application of AI models is now widely popular to solve environmental problems. However, in the present study, we employed the Recurrent neural network, a popular AI model, to predict and forecast the rainfall of 5 stations or districts of Tamil Nadu.

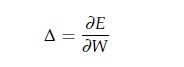
The prediction of cumulative measures from the sequences of vectors variable length including a component ‘time’ is extremely significant in machine learning. RNN has the power of learning previous-term dependencies. Previous-term dependencies are very much essential when it comes to predicting weather patterns. RNN is useful in that perspective since it is an artificial neural network utilized especially for developing prediction networks using long-term time series datasets The general RNN performs prediction at a time t as follows,

(1) (2)

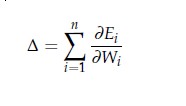
where Ht is the hidden state, Ot is the predicted output, Wh and Wo are the weights assigned in hidden layer and output layer, bh and bo are the bias for hidden and output layer, g is the activation function used in hidden layer and f is the output function for prediction. Prediction performance of RNN majorly depends on the activation function, which takes the current input and previous state as input vector and predicts output based on the hidden state result. Based on this model, much research work has been carried out with certain improvements in the model as discussed in related work section.

Elman network is a category of RNN comprising one or many hidden layers. The first layer contains the weight vectors that are attained from the input layer. Every layer receives weight from the previous layer. RNNs are designed to capture temporal contextual information along with time series data. Unlike the traditional FFNNs, RNNs possess feedback loops in their structure which are used for feeding the outcome of the preceding time steps as input to the current time step. This construction enables RNNs to develop complex temporal contextual figures along a time series dataset.

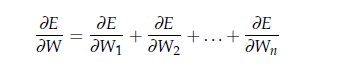
Usually, the change in weight (D) during back-propagation in a neural network is calculated based on the error (E) occurred for the weight (W) assigned in the previous state as follows,

 (3)

The Back-Propagation Through Time (BPTT) is a technique that is generally used for training the RNNs, which adjusts the weight of the network to reduce the error by back-propagating to certain count of states n constantly, and finally, the summation of error to the respective weights of the back-propagated states are calculated for the change in weight as,

(4)

The partial derivatives of the Recurrent Neural Networks for n number of states is represented as,

(5)

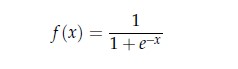
Nevertheless, it is challenging to train the traditional RNNs using BPTT because of the exploding and vanishing gradient problems. Errors from later time steps are di\_cult to propagate back to previous time steps for appropriate changes inf network parameters. To address this problem, the LSTM unit has been developed by reducing the loss at every stage significantly.

4.5.1 Functions of activation functions

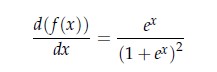
As specified in the previous section, activation functions are the major cause for the prediction and entire network performance. Hence, choosing the activation function is a major challenge for a suitable application. Assigning and adjusting the weights for the activation function has a great impact on the results of activation functions since the neurons learn the relationship between input vector and output based on the weight tuning. During the training phase, the deference in change of weights is high at the initial stage due to high learning rate, it gradually decreases and learning stops when the weight is nearest to zero, which conveys there is no change in output for the change in input. The relationship between weight update Wt. and learning rate Lt at a time t is given as

Wt. = Lt × Ew (6)

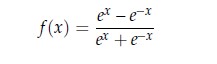
Weight tuning helps the neural networks to restyle the steepness of the curve whereas bias value applied to the network finds the best fit for the curve jointly with the data through shifting. Many activation functions have been developed in recent years, but the most commonly used activation functions are sigmoid, tanh and Relu. Every activation function has its own pros and cons, so let us discuss them in detail. Sigmoid is the function habitually used for hidden layers of Artificial Neural Networks which maps any range of input values between 0 and 1. The function squashes even a large input values to a small range due to which a substantial change in input reflects as a small change in output. Therefore, this function can be used for those applications that need to compute probability as outcome, and it is represented as below,

 (7)

During back-propagation, the gradient of the function decreases exponentially, which leads to very small derivatives and weight tuning in the initial layers of the neural network. This is defined as vanishing gradient problem, in which the neurons tend to stop learning at a certain level. The derivative of the sigmoid function is

 (8)

One more important issue in the sigmoid function is that the sigmoidal curve is not centered to the origin; instead, it is centered by 0.5. Thus computations related to transitions are difficult in the curve, and the usage of the tanh function is loftier than sigmoid. Tanh activation function also called hyperbolic tangent is more eminent than the sigmoid in several ways. It accepts a wide range of inputs such that the function ranges between 􀀀1 and +1, thereby computing the output for negative inputs also. The tanh function is defined as

(9)

For this reason, the hyperbolic tangent curve is steeper than the sigmoidal curve, and hence, derivatives are sturdy, which minimizes the vanishing gradient compared to the sigmoid. The derivative of hyperbolic tanh is

(10)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network that can learn order dependence. The output of the previous step is used as input in the current step in RNN. Hoch Reiter & Schmid Huber created the LSTM. It addressed the issue of RNN long-term dependency, in which the RNN is unable to predict words stored in long-term memory but can make more accurate predictions based on current data. RNN does not provide an efficient performance as the gap length rises. The LSTM may keep information for a long time by default. LSTM has feedback connections, unlike conventional feed-forward neural networks. It can handle not only single data points but also complete data streams.

The LSTM is made up of four neural networks and numerous memory blocks known as cells in a chain structure. A conventional LSTM unit consists of a cell, an input gate, an output gate, and a forget gate. The flow of information into and out of the cell is controlled by three gates, and the cell remembers values over arbitrary time intervals. The LSTM algorithm is well adapted to categorize, analyse, and predict time series of uncertain duration The cells store information, whereas the gates manipulate memory. There are three entrances namely Input Gate which it determines which of the input values should be used to change the memory. The sigmoid function determines whether to allow 0 or 1 values through. And the tanh function assigns weight to the data provided, determining their importance on a scale of -1 to 1. Forget Gate which It finds the details that should be removed from the block. It is decided by a sigmoid function. For each number in the cell state Ct-1, it looks at the preceding state (ht-1) and the content input (Xt) and produces a number between 0 (omit this) and 1 (keep this) and finally the Output Gate where the block’s input and memory are used to determine the output. The sigmoid function determines whether to allow 0 or 1 values through. And the tanh function determines which values are allowed to pass through 0, 1. And the tanh function assigns weight to the values provided, determining their relevance on a scale of -1 to 1 and multiplying it with the sigmoid output.

Input Gate | Long Short Term Memory (11)

Forget Gate | Long Short Term Memory(12)

Output Gate (13)



In the present study, we divided the whole rainfall datasets into training and testing datasets. Then we normalized the data and applied Bidirectional LSTM model for predicting the rainfall. In Bidirectional LSTM model each training sequence is presented forwards and backward to two independent recurrent nets, both of which are coupled to the same output layer in Bidirectional Recurrent Neural Networks (BRNN). This means that the BRNN has comprehensive, sequential knowledge about all points before and after each point in a given sequence. We applied Bidirectional LSTM model on rainfall datasets of five stations and by changing the model’s parameters like seed, momentum, learning rate until the best LSTM model achieved for prediction rainfall. In addition, we evaluated the performance of Bidirectional LSTM for predicting using Root Mean Square Error (RMSE) techniques. When we achieved the best Bidirectional LSTM model, we fixed the model parameters and applied on the rainfall data to forecast rainfall up to 2031.

**4.5 UML DIAGRAMS**

**4.5.1 USE CASE DIAGRAM**

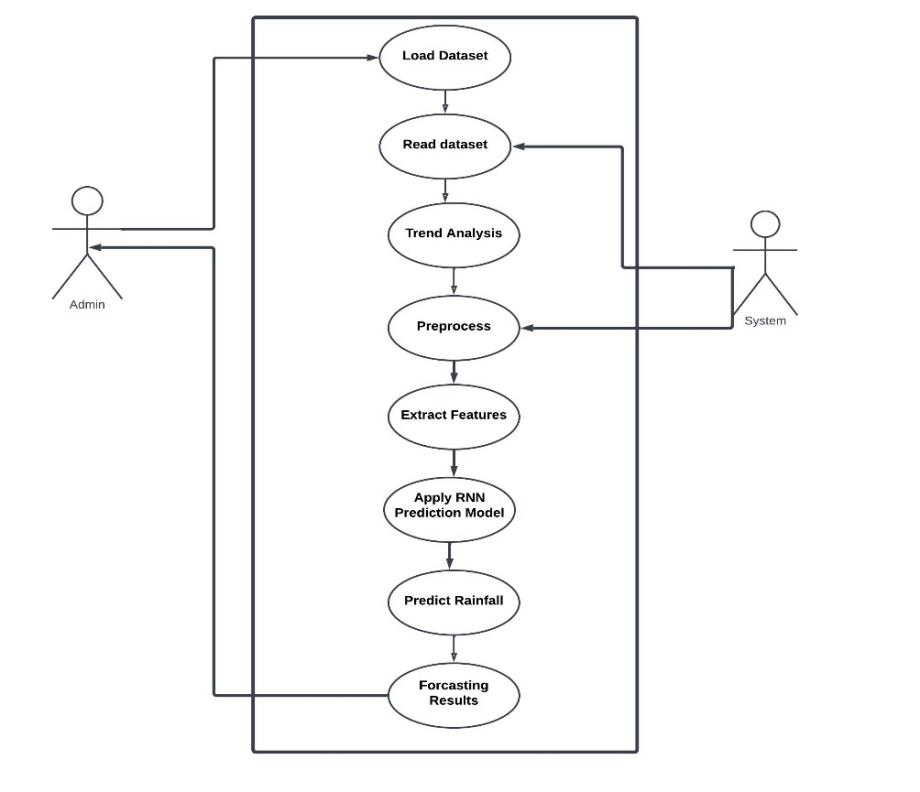
****

FIGURE 4.3 USE CASE DIAGRAM

**4.4.2 SEQUENCE DIAGRAM**

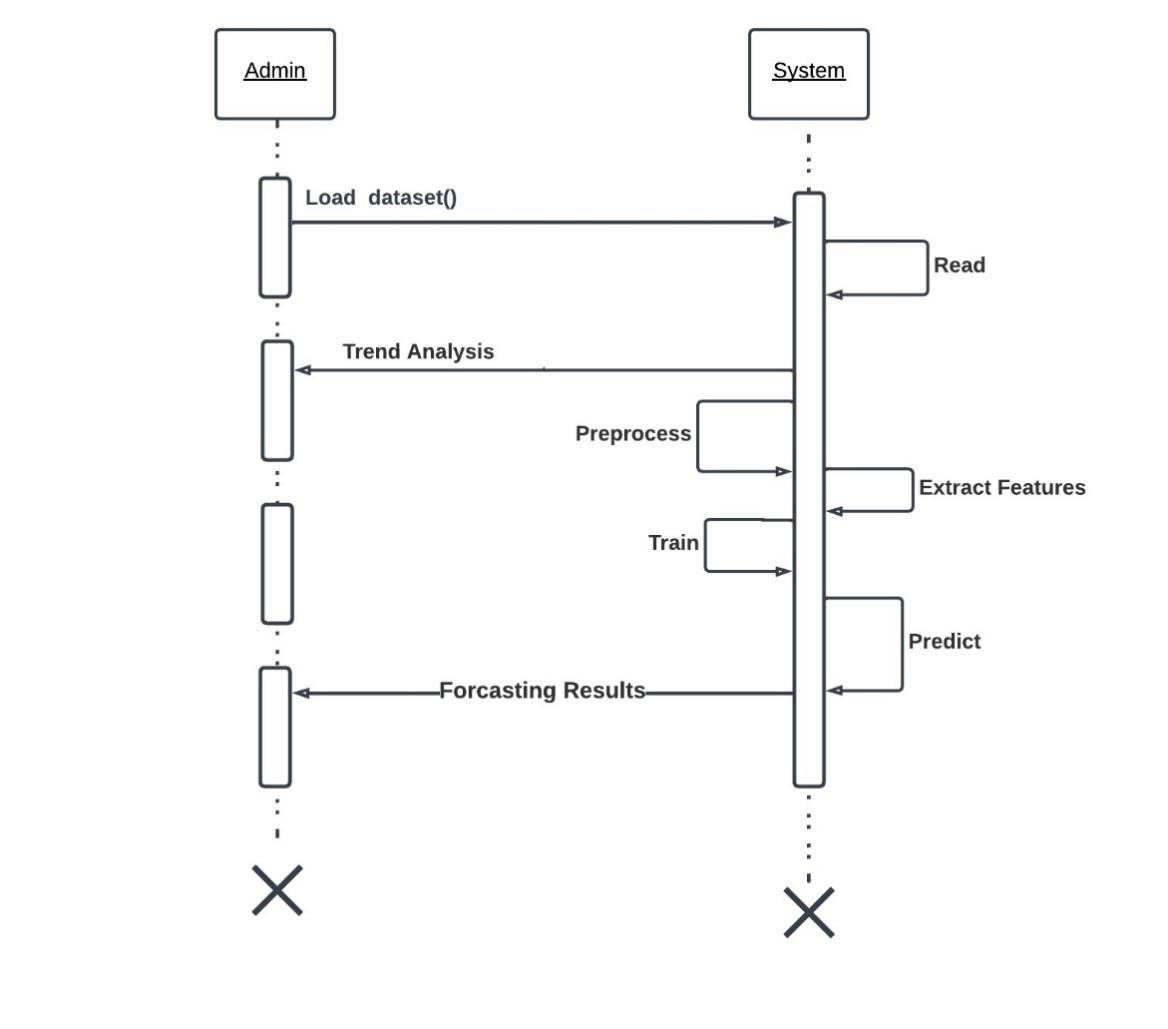
****

FIGURE 4.4 SEQUENCE DIAGRAM

**4.4.3 CLASS DIAGRAM**

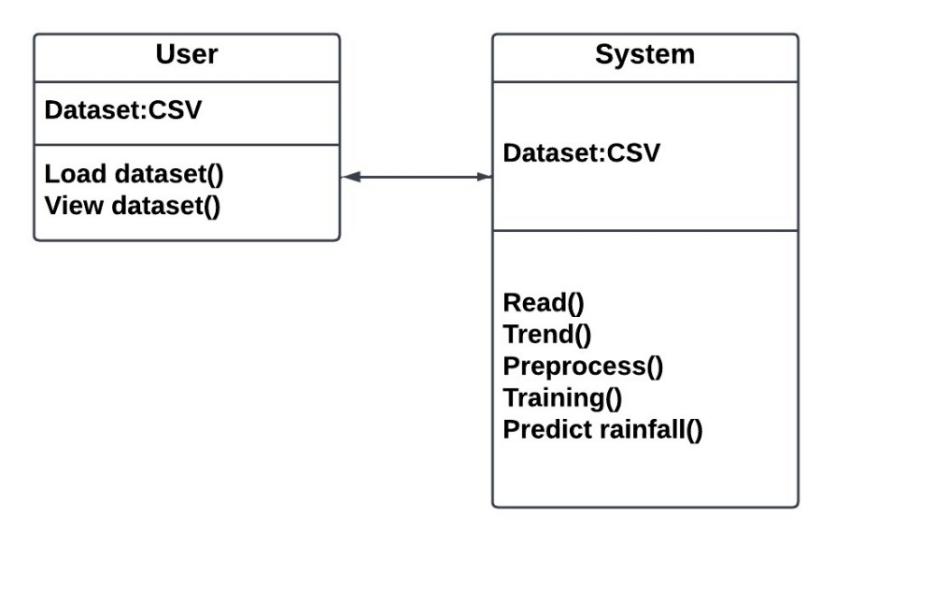
****

FIGURE 4.5 CLASS DIAGRAM

**4.4.4 ACTIVITY DIAGRAM**

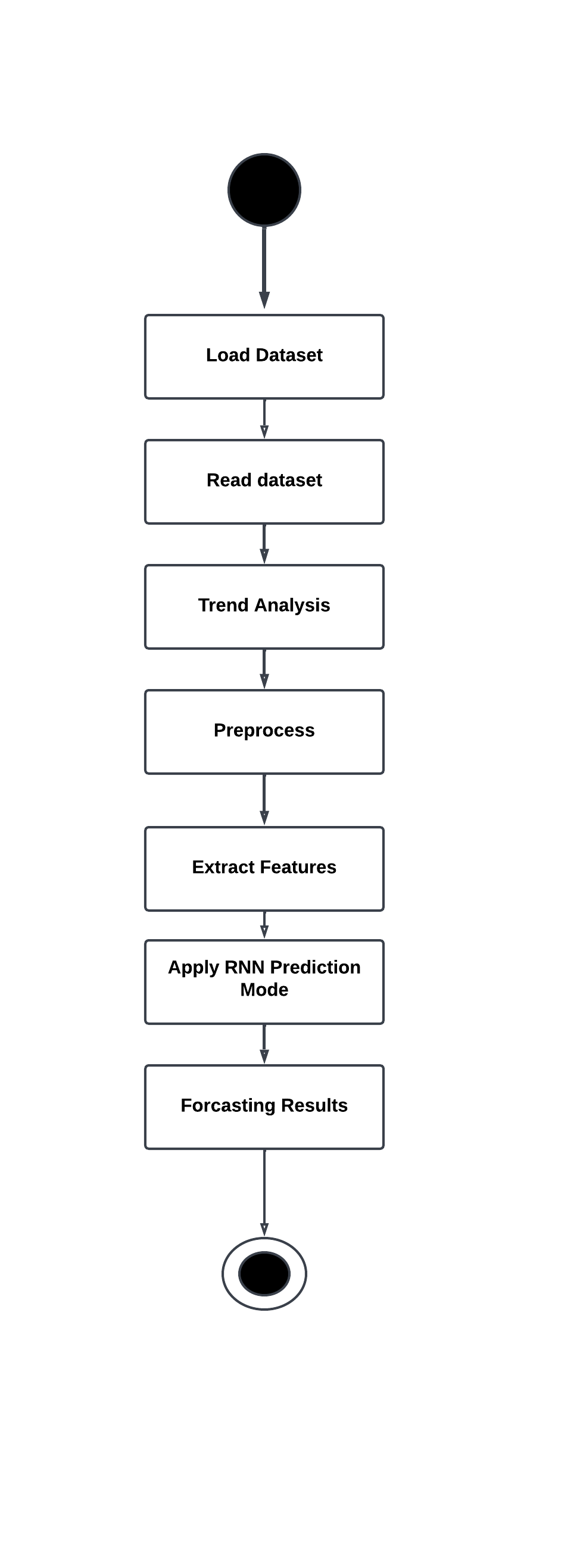
****

FIGURE 4.6 ACTIVITY DIAGRAM

**4.4.5 ARCHITECTURE DIAGRAM**

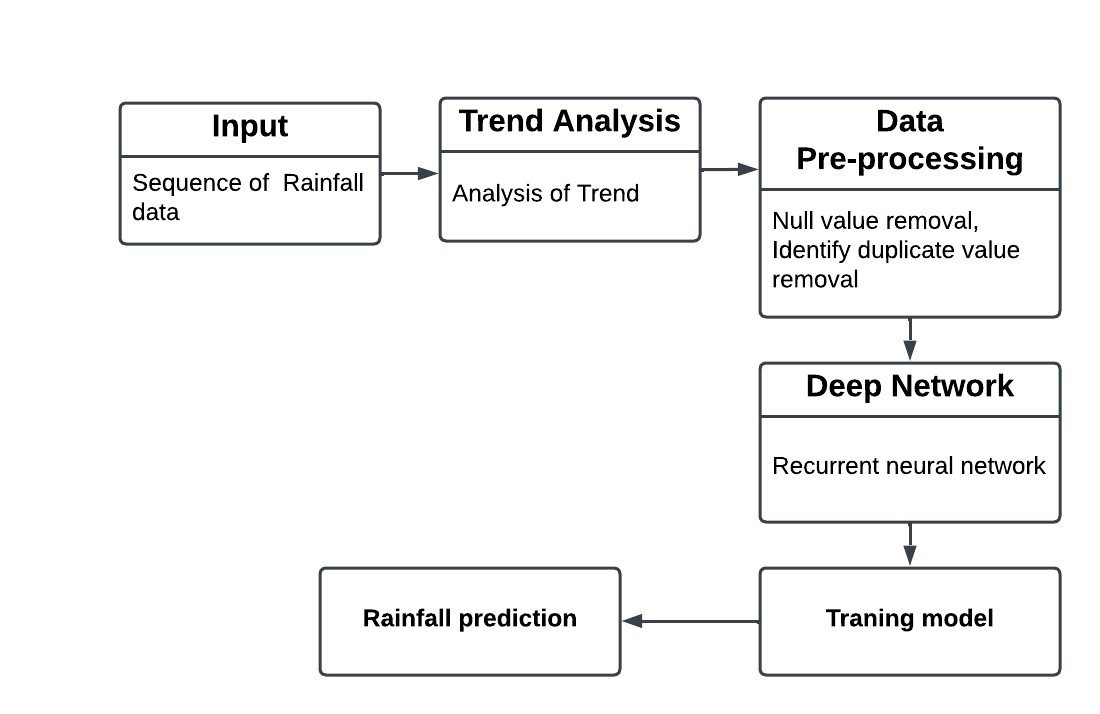


FIGURE 4.7 ARCHITECTURE DIAGRAM

**4.5 CONCLUSION**

In this chapter, we discussed about the working model of module 1 and 2. We portrayed the UML diagrams for the above modules. Also depicted a architecture of proposed work

### 

### CHAPTER 5

**IMPLEMENTATION & RESULTS**

**5.1 GENERAL**

Implementation is an important part in any analysis or prediction. In this chapter the statistical analysis od season wise rainfall data is carried out. The trend analysis of the same were discussed

### 5.2 IMPLEMENTATION

### In the present study trend analysis of rainfall of for the state of Tamil Nadu was done using Mann-Kendall to gather with Sen’s slope estimator for 30 years of time series data (1990- 2019) on seasonal and yearly basis.

The variation in rainfall data (trend) on a seasonal basis is calculated individually for each season using Mann-Kendall statistical method and the magnitude of the slope is calculated with Sen’s slope estimator as represented in graphs

Table 4.1 Statistical Analysis of Rainfall data along with Mann-Kendall Trend and Sen's Slope.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Max** | **Min** | **Mean** | **Std div** | **Z** | **Q** |
| **Actual Rainfall in South-West Monsoon (in mm)** | 454.8000 | 185.4000 | 312.0433 | 62.9602 | 0.7850 | 1.1611 |
| **Actual Rainfall in North-East Monsoon (in mm)** | 828.800 | 174.000 | 485.0526 | 144.564 | -0.3925 | -0.4583 |
| **Actual Rainfall in Winter Season(in mm)** | 119.5000 | 2.9000 | 23.3366 | 23.6713 | 0.0356 | 0.01 |
| **Actual Rainfall in Hot Weather Period (in mm)** | 283.4000 | 63.2000 | 132.2733 | 57.5880 | 1.8022 | 1.275 |
| **Actual Rainfall in Whole Year (in mm)** | 1304.1000 | 598.1000 | 952.7133 | 168.1310 | 0.1784 | 1.21 |

**South-West Monsoon**

It was analysed that there are significant changes in seasonal rainfall data some of the results showed increasing (upward) trend and some showed decreasing trends. The trend analysis for the South-West season for Tamil Nadu is showing an increasing upward trend with the value of Z statistics of 0.78501 and the magnitude of this trend from the test of the sen slope estimator resulted to be 1.1611. For this season, a relatively Kendall’s tau value (0.10344) also implies a positive correlation between rainfall precipitation and time i.e. rainfall increases considerably as the year progresses between 2000 to 2005 and continues to increase till 2019. When the normal rainfall for the southwest season for the state of Tamil Nadu has considered the analysis showed a Z statistics value of 3.3692 which also indicates an upward trend.

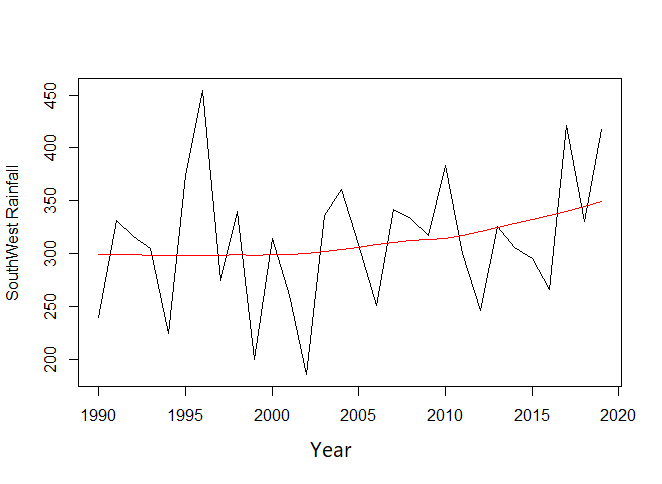


FIGURE 5.1 South-West

**North-East Monsoon**

The analysis for the northeast monsoon is performed for 30 years of rainfall for the state of Tamil Nadu. The North East is usually from October month to December month The results show from Mann Kendall test show a decreasing trend as shown in figure 5.2 in rainfall with the value of Z statistics of -0.3925 and the estimated Sen’s Slope (Q) was also calculated which resulted in a value of -0.4583. Kendall’s tau value which results in -0.05287 represents a correlation between rainfall. When compared to normal rainfall during the season which results in a Z statics value of 0.10768 and Z statistics value for the percentage of deviation for this season shows a negative value which indicates a falling trend as the year progress.

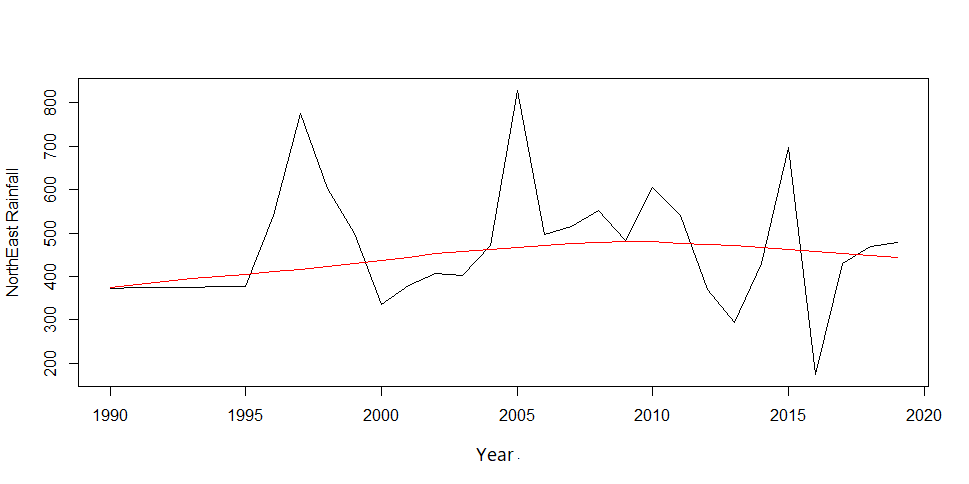


FIGURE 5.2 North-East

**Hot weather**

The Hot weather season was normally from May to June in Tamilnadu. During this season we analysed a trend (upward or downward) using data that is collected between these months. The trend is analysed by calculated Z-statistics and Q- statistics. Z value is basically from Mann Kendall and Q value reis presented as Sen's slope. Using these techniques, the Z statics showed an upward trend with positive value of 0.035693, ad variance of 3.1396 and a Q is 0.01. which shows that Z and Q have positive values which showed a significant increasing trend as shown in figure 5.3. If the values will be in means negative means to conclude that the significance-decreasing trend.

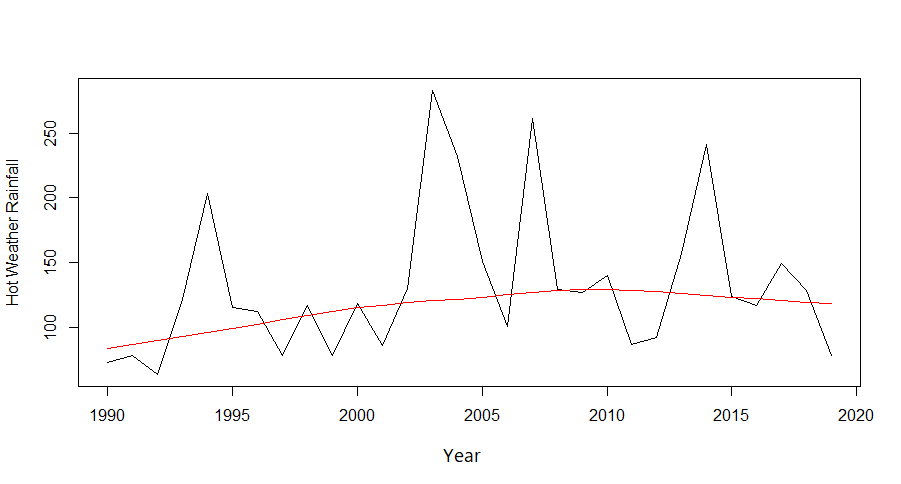


FIGURE 5.3 Hot weather

**Winter**

Winter season was normally from December to January in Tamil Nādu. In this time we analyse a trend (upward or downward) using data that is collected between these months taken for Z-statistics and Q- statistics. Z value is basically from Man Kendall and the Q value is represented as Sen's slope. Using this techniques z=0.035693, variance= 3.1396 and the Q= 0.01. which means that Z and Q will have positive values which showed a significant increasing trend as shown in figure 5.4 .If the values will be in means negative means to conclude that the a significant decreasing trend.

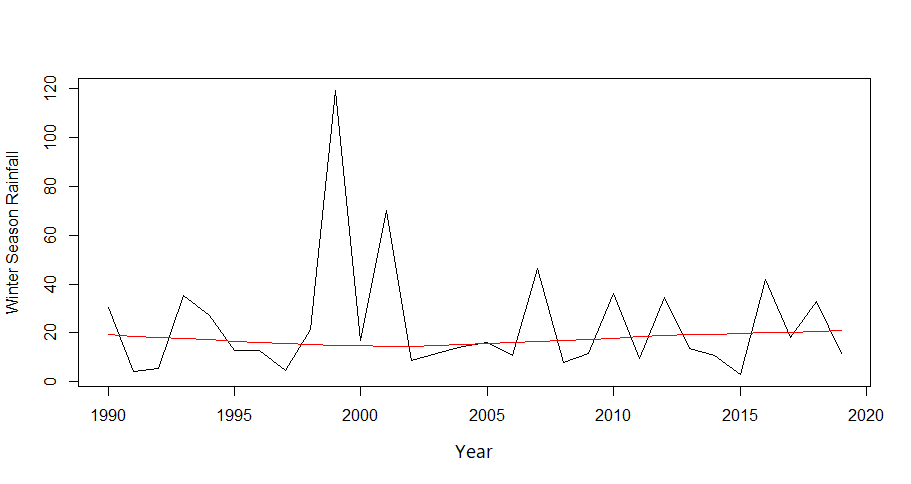


FIGURE 5.4 Winter

**Whole year**

The trend of total annual rainfall of 30 years data is represented in Figure 5.5 the Z statistics value is 0.1784 and Q value is 1.21 which indicates an upward trend. The tau value is 2.52873 which the correlation between rainfall and year. The highest rainfall was observed in 2007 of 1304.10 mm and lowest rainfall was observed in 2016 of 598 mm.

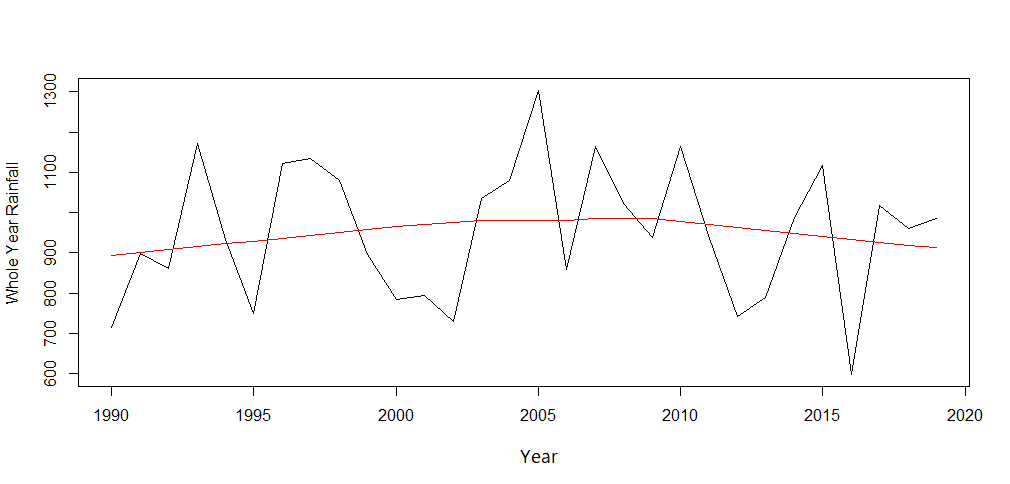


FIGURE 5.5 Whole year

**5.2. RAINFALL PREDICTION APPROACH**

**5.2.1 Data Preprocessing**

In this section, the climatic features included in the weather datasets are first described. Subsequently the feature selection process carried out as part of the pre-processing procedure to prepare the time-series data for use in the training of the rainfall forecast models is given. As with any Machine Learning approach, a processing procedure is required to prepare raw data for use in model training and testing processes. The pre-processing procedure carried out had as its objective the elimination of categorical data, the deletion or replenishment of incomplete data, and the structuring of one month of observations in one vector row.

Data is collected from NASA POWER data access viewer a data source provider. Specifically, historical data from January 1, 1981, to January 1, 2021, from the cities of Ariyalur, Coimbatore, Cuddalore, Thothukoodi and Villupuram were used. The complete dataset comprises monthly recorded weather measurements of temperature, pressure, humidity, wind speed and direction, percentage of clouds, the volume of rain, and volume of snow.

Data Preprocessing is carried out by separate weather data by city. This step aims to preserve the weather singularities of each city during the training process. The interquartile range is then used to measure or find the outliers in the data. From the data or on a box plot a fence is used to identify and categorize the type of outliers. Respective outliers were removed and data is normalized.

**5.2.2 Feature Selection Method**

Feature selection process is one of the main components of a feature engineering process. This is how a predictive model is developed by reducing the number of input variables. Feature selection techniques are employed to reduce the number of input variables by eliminating redundant or irrelevant features. It then narrows the set of features to those most relevant to the machine learning model. A feature selection objective in machine learning identifies the most helpful group of features that can be used to build useful models of the phenomena being studied

In this study, three feature selections were employed **Pearson’s correlation, forward selection and recursive feature elimination. Pearson’s Correlation** It is used to quantify linear dependence between two continuous variables, X and Y. Its value ranges from -1 to 1.Forward selection is an iterative method in which we start with having no feature in the model. In each iteration, we keep adding the feature which best improves our model till an addition of a new variable does not improve the performance of the model. **Recursive feature elimination** algorithm aims to find the best performing subset of features using greedy optimization. Each iteration creates a new model and keeps aside the best or worst performing features. The next model is constructed using the remaining features until all the features have been exhausted. The features are then ranked according to their elimination order. **It is found that forward feature selection performed the best of three and was used in prediction model.**

**5.2.3 Prediction Model**

Rainfall data of five regions starting from 1981 until 2021 is the dataset considered for the prediction process. This dataset is split into training dataset and testing dataset. Rainfall data (34 years from 1981 to 2018) is taken as the dataset for training the Bidirectional LSTM based RNN model. This trained model is then tested with the dataset of the year 2019 to 2021.

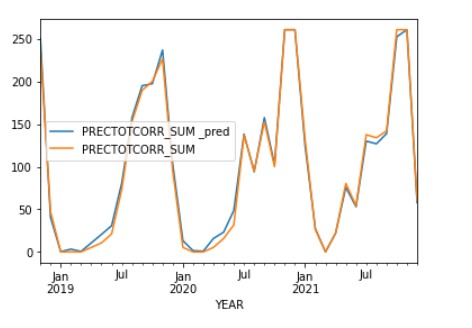
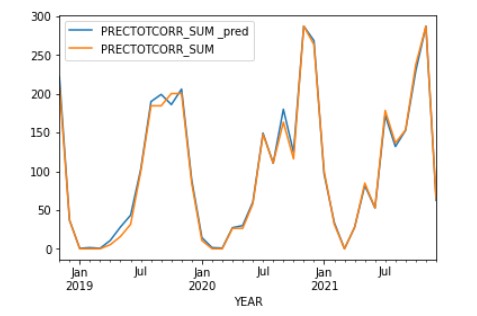
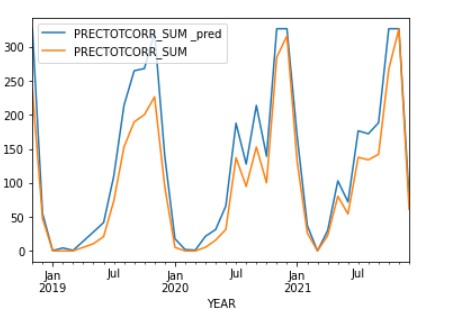
The model has the following layers: Bidirectional layer with an LSTM (Long Short-Term Memory) layer of 200 units and return sequences=True. This means that the LSTM layer returns the full sequence of hidden states instead of just the last one. The input shape is which indicates that the model expects input sequences of length window length with number of features. Dense layer with 20 units and hyperbolic tangent activation function (tanh). This layer adds non-linearity to the model. Another Bidirectional layer with an LSTM layer of 150 units. This layer doesn't have return sequences=True, which means that it only returns the last hidden state. Another Dense layer with 20 units and tanh activation function. Another Dense layer with 20 units and tanh activation function. Dropout layer with a rate of 0.25, which randomly drops out 25% of the units during training. This helps prevent overfitting. Final Dense layer with a single unit and no activation function. This layer produces the model's output, which is a single value. The model uses the Adam optimizer and mean squared error (mse) loss function, which are common choices for regression problems.

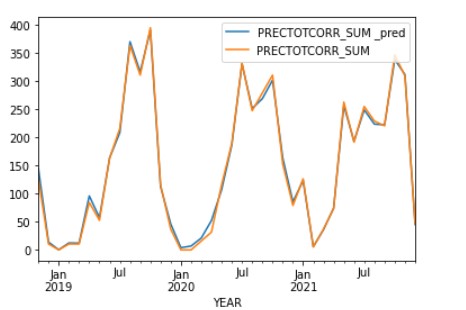
Overall, this is a fairly complex model that uses two layers of bidirectional LSTMs to capture the temporal dependencies in the input sequence, followed by several dense layers with non-linear activations to transform the LSTM outputs into a final prediction.

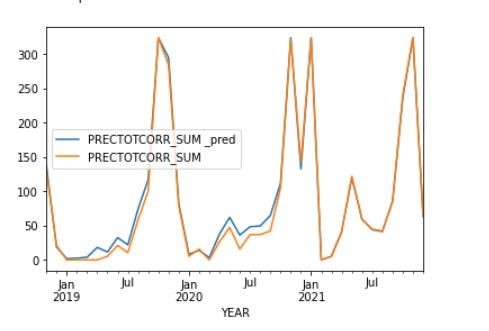
The above-described model is used and the below results were archived

**5.3 SUMMARY**

From the results of the analysis of Mann- Kendall, and Sen slope, the following points were discussed.

* The North-East season showed a decreasing trend.
* The South-West, Winter and Hot summer seasons showed an increasing trend.
* The trend analysis for the whole year also shows an increasing trend.





### CHAPTER 6

### CONCLUSIONS AND WORK SCHEDULE FOR PHASE II

### 6.1 SUMMARY FOR PHASE I

The Rainfall data collected for the years (1977 – 2019) for the state of Tamil Nadu was collected from India Meteorological Department (IMD). The data includes the seasonal rainfall in Tamil Nadu. From the obtained data Statistical Analysis of Rainfall data along with the Mann-Kendall Trend and Sen's Slope helps to understand the pattern of rainfall in the selected regions.

### 6.2 CONCLUSION

The present study concluded that the Tamil Nadu region showed substantial changes in rainfall patterns during the last century. The seasonal-wise trend of rainfall volume, in many cases, is not the same as observed for the previous region. The northeast monsoon shows a decreasing trend while other seasons show an increasing trend. The decreasing trend of North-East rainfall, can cause a drought situation and hamper agricultural activities like land preparation, particularly in non-irrigated areas.

Further trend analysis of selected features is carried out in phase two. Since the state is entirely dependent on rains for recharging its water resources, monsoon failures lead to acute water scarcity and severe drought. Based on rainfall trends, Resolutions are required to cope with the changing pattern of precipitation and the crop calendar needs to be modified if the same trend is to be stable and prolonged.

**REFERENCES**

[1] D. R., Droughts/floods in the summer monsoon season over different meteorological subdivisions of India for the period 1871–1984. *J. Climatol.*, 1987, **7**; https://doi.org/10.1002/joc.3370070106

[2] Fiaz Hussain, Ghulam Nabi, Muhammad Waseem Boota Rainfall Trend Analysis by Using the Mann-Kendall test & Sen's Slope Estimates : A Case Study of District Chawal Rain Gauge, Barani area, Northern Punjab Province for Pakistan.(2015)

[3] Lince Rachel Varghese A Time-series based Prediction Analysis of Rainfall Detection (2020)

[4] Mislan, H.; Hardwinarto, S.; Sumaryono, M.A. Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan, Indonesia. Procedia Comput. Sci. 2015, 59, 142–151.

[5] Nan Gu, Dingsheng Wan Trend Analysis of Extreme Rainfall Based on BP Neural Network (2010)

[6] Rawshan Ali, Alban Kuriqi , Shadan Abubaker and Ozgur Kisi Long-Term Trends and Seasonality Detection of the Observed Flow in Yangtze River Using Mann-Kendall and Sen’s Innovative Trend Method (2019)

[7] Salma, S., S. Rehman, M. A. Shah Rainfall Trends in Different Climate Zones of Pakistan. (2012)

[8] S. Poornima and M. Pushpalatha Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units (2019)

[9] Shyam Lochan Bora,,Kalyan Bhuyan, Partha Jyoti Hazarika3, Junmi Gogoi and Kuldeep Goswami**:** Analysis of rainfall trend using non-parametric methods and innovative trend analysis during 1901–2020 in seven states of North East India(2022)

[10] T.F.C.C., “Climate change in Pakistan: TFCC (*Task Force for Climate Change*) recommends various adaptations, mitigation measures” (2010).