# Rainfall Analysis of India

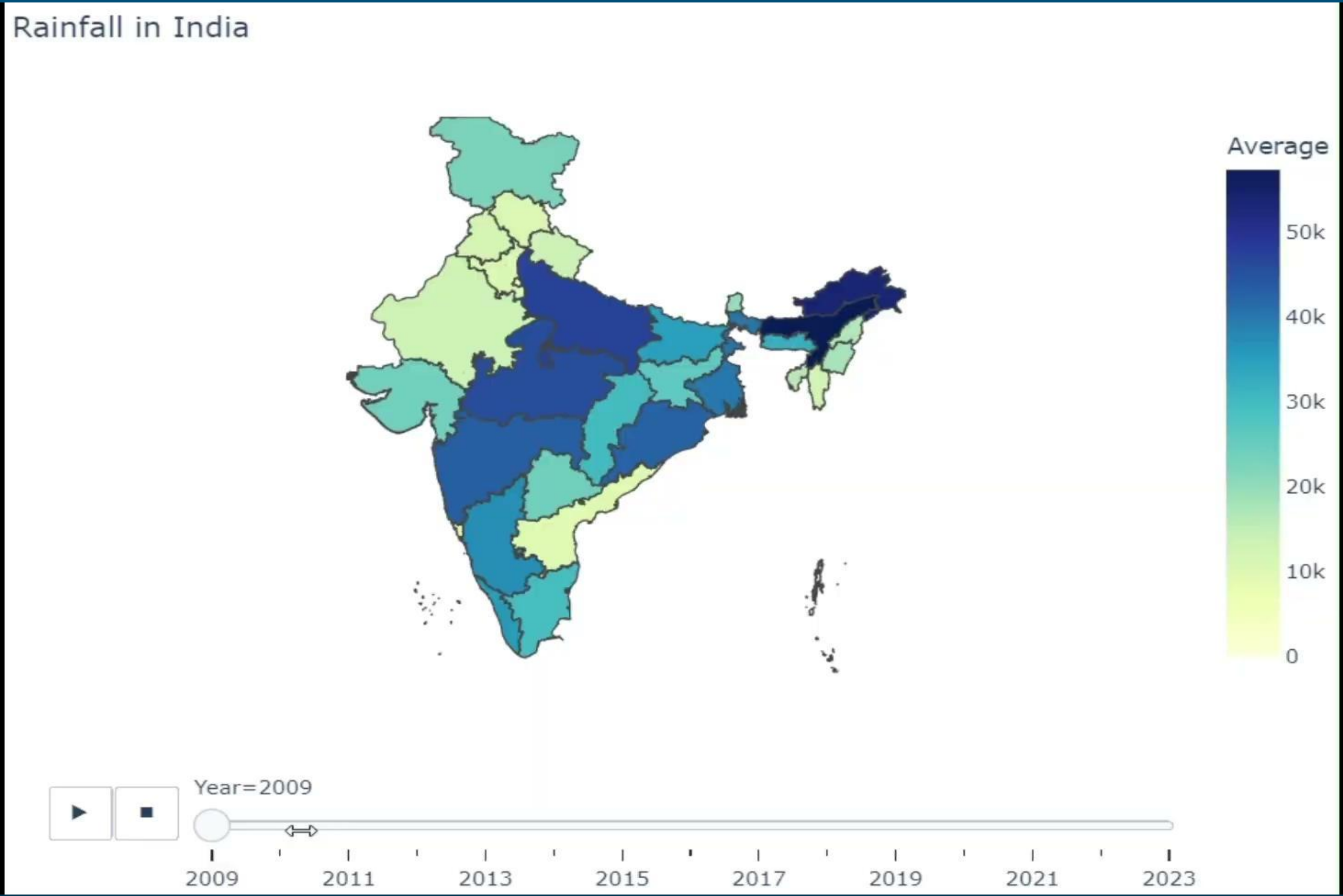
Manasa Swamireddy
(CS20B1022)

Guided by: Dr. Priodyuti Pradhan

# Introduction

- ❖ Rainfall patterns in India exhibit considerable variability across regions and time periods. India experiences a wide spectrum of rainfall regimes.
- ❖ They influence the frequency and intensity of natural hazards such as floods, droughts, and landslides.
- ❖ There is a need to analyze the rainfall patterns using historical rainfall data in order to take proactive measures to mitigate their impacts.
- ❖ The main aim is to provide valuable insights into the complex dynamics of rainfall variability in India and to understand the climate of India

# Rainfall in India over the past 15 Years



# Literature Review

## Comparison of long-term and short-term trends in India

Reference: Gangarde, A., Dauji, S., & Londhe, S. (2023). Comparison of long-term and short-term trends of annual rainfall in India: a case study. ISH Journal of Hydraulic Engineering, 29(3), 411–424. https://doi.org/10.1080/09715010.2022.2084351

- ❖ This study analyzes long-term and short-term rainfall trends across India's subdivisions.
- ❖ Utilized Mann-Kendall test, Spearman's Rank Order Correlation test for trend detection.
- ❖ These methods detected increasing trends in three subdivisions and decreasing trends in four subdivisions, but failed to detect the other 19 subdivisions.
- ❖ The average of the total rainfall of a subdivision is taken into account which results in misinterpretation, rather than analyzing each district's trend.

# Literature Review

## Study of LSTM for Rainfall Prediction in India

Reference: Zoremsanga, C., Hussain, J. (2023). A Comparative Study of Long Short-Term Memory for Rainfall Prediction in India. In: Singh, S.N., Mahanta, S., Singh, Y.J. (eds) Proceedings of the NIELIT's International Conference on Communication, Electronics and Digital Technology. NICE-DT 2023.

- ❖ The study applied four LSTM models to predict average monthly rainfall in India and compared their performances with a benchmark model found in the literature.
- ❖ Average monthly rainfall data from 1871 to 2016 was used for training and testing.
- ❖ The LSTM Model-4 achieved an RMSE of 245.30, whereas the existing benchmark model achieved an RMSE of 251.63.
- ❖ To conclude, this model is not accurate enough as the RMSE values are high.

# Literature Review

## Annual Rainfall Prediction Using Time Series Forecasting

- ❖ The study reviews univariate forecasting techniques for rainfall prediction, namely regression analysis, clustering, ARIMA, ETS and ANN.
- ❖ The results indicate that ARIMA performs well on the given data.
- ❖ However, this study considered only a few regions of India.
- ❖ When taking the large data which comprises all the districts of India, the ARIMA model fails to correctly forecast annual rainfall.

# Problem Statement

- ❖ Trend analysis commonly aggregates data at the state level, overlooking district scale variations.

- ❖ High-dimensional rainfall data from numerous districts can hinder model performance. Dimensionality reduction helps address this by transforming the dataset while preserving crucial information.

- ❖ Existing LSTM model has high value of RMSE, i.e. 245.30.

# Proposed Methodology

- ❖ Trend Detection, taking into account all the districts trends of a particular state and calculating the overall trend.

- ❖ Dimensionality reduction of the features(districts) of a state, while preserving important information using Principal Component Analysis (PCA).

- ❖ Forecasting Rainfall using Long Short-Term Memory (LSTM).

# Trend Detection

## Dataset:

❖ The dataset is extracted from the Indian Meteorological Department (Government of India) https://mausam.imd.gov.in/.

❖ It has daily rainfall values of each district of India from 2009 to 2023.

❖ As the main focus is on the trend observed by the rainfall data, the data is sliced to contain rainfall values only for their respective monsoon period.

# Trend Detection - Monsoon Periods

| Groups | States |
|---|---|
| Group - 1 (May to October) | Assam, Meghalaya, Nagaland, Sikkim, Mizoram, Tripura, Arunachal Pradesh, West Bengal, and Manipur. |
| Group 2 (June to October) | Andhra Pradesh, Bihar, Gujarat, Maharashtra, Telangana, and Uttar Pradesh |
| Group 3 (June to September) | Jammu and Kashmir, Himachal Pradesh, Haryana, Punjab, Jharkhand, Rajasthan, Uttarakhand, and Chhattisgarh |
| Group 4 (June to November) | Karnataka, Madhya Pradesh, and Odisha |
| Group 5 (June to December) | Puducherry and Tamil Nadu |
| Group 6 (June to August, October to November) | Kerala |

# Trend Detection - Architecture

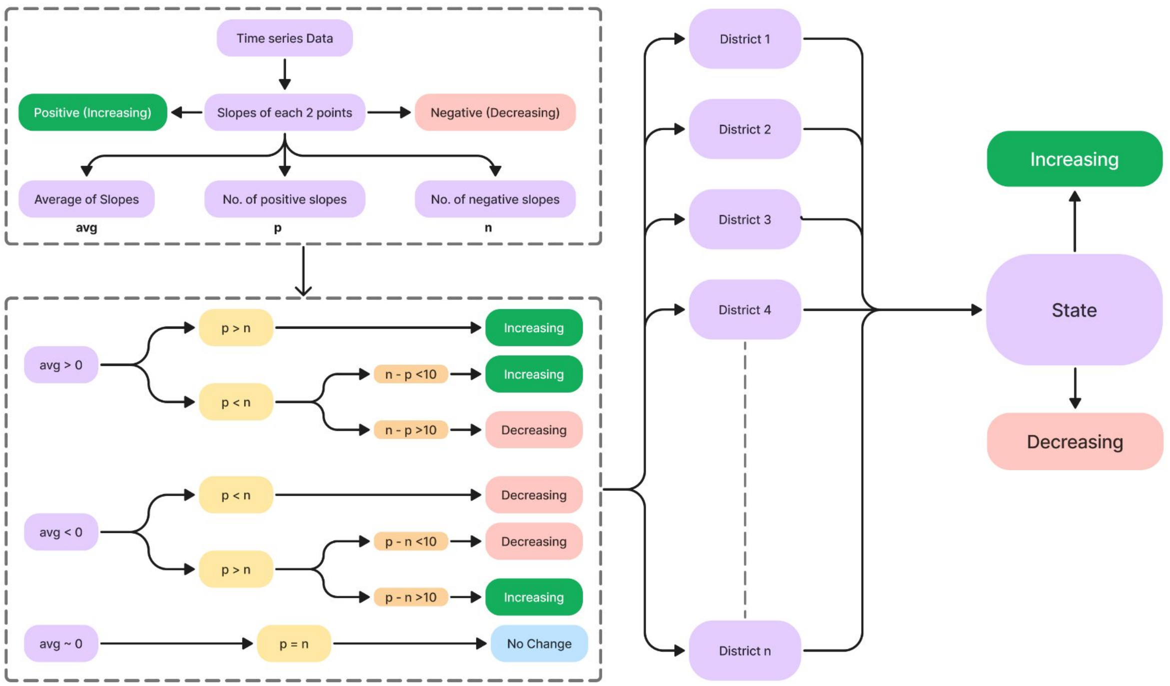



**Figure 1: Architecture of Trend Detection**

Figure 1.1 depicts the proposed Trend Detection Methodology

# Result

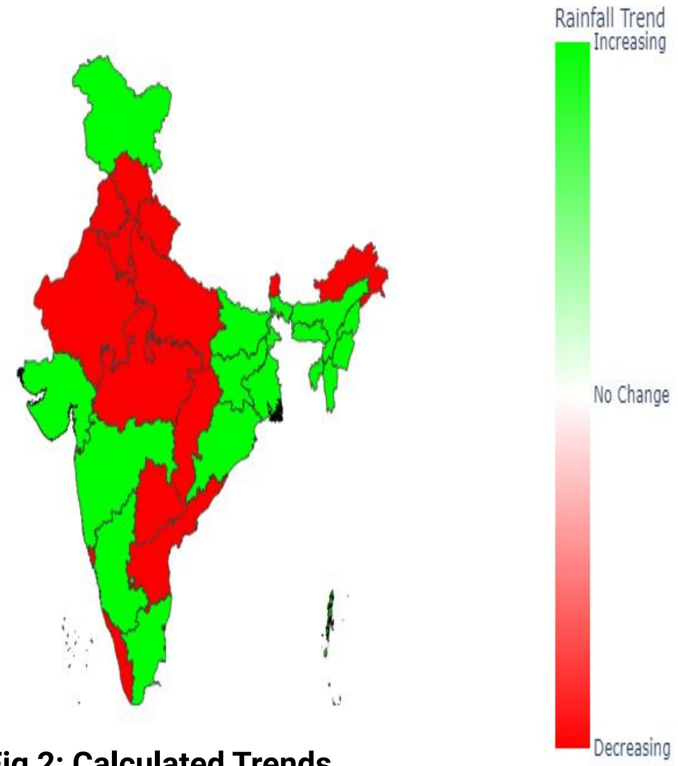



**Fig 2: Calculated Trends**

❖ From the results, it is observed that 15 states show increasing trend, and 14 states show decreasing trend and no change trend is not observed.

# Validation

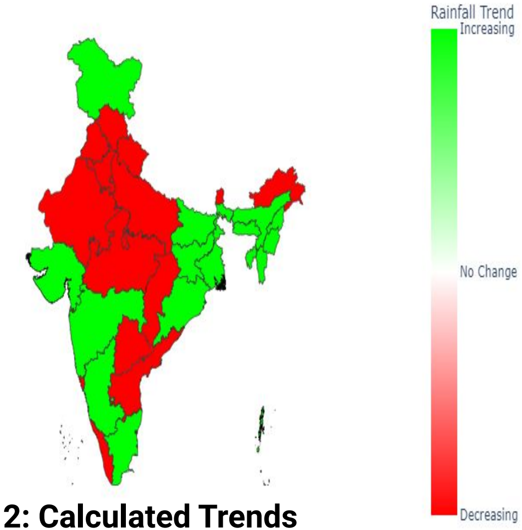



Fig 2: Calculated Trends

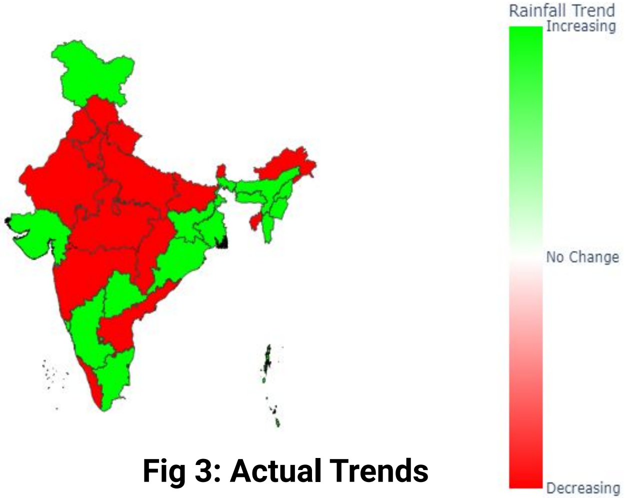



Fig 3: Actual Trends

- ❖ The actual trends are calculated using existing trend detection method Mann Kendall Test.
- ❖ Out of 29 states, 25 states show the same trend exhibited by the actual trends.

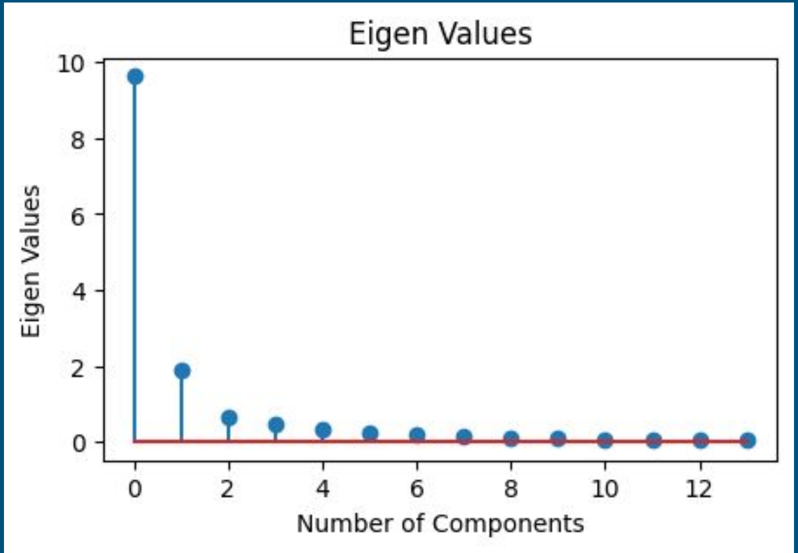
# Dimensionality Reduction using PCA

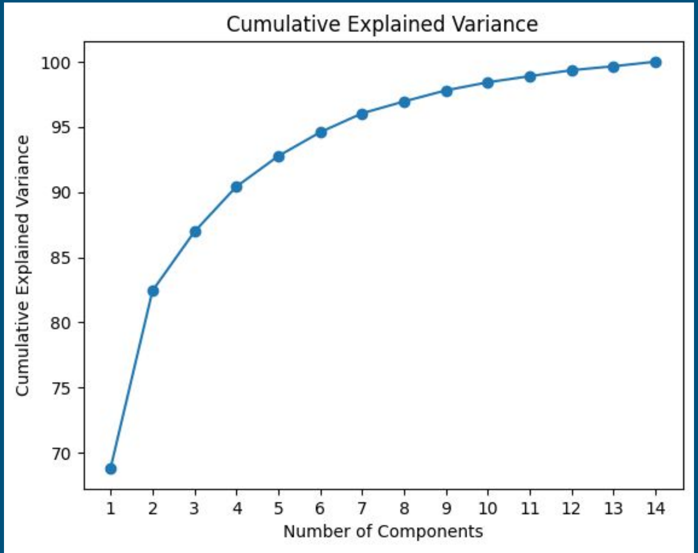
- PCA transforms high-dimensional data into a lower-dimensional space while retaining most of the important information.
- It does this by identifying the directions (principal components) in which the data varies the most.
- The principal components are calculated such that they are uncorrelated with each other.
- The first principal component captures the direction of maximum variance in the data.

# Implementation of PCA

- Data Normalization
- Covariance Matrix Computation
- Eigenvalue Decomposition
- The eigenvectors with the highest eigenvalues capture the most variance in the data and are referred to as principal components.

# Eigenvalues and Explained Variance




From the above figures, we observe that taking upto three eigenvalues, it expresses 85% variance.

# Reduced Dimensionality

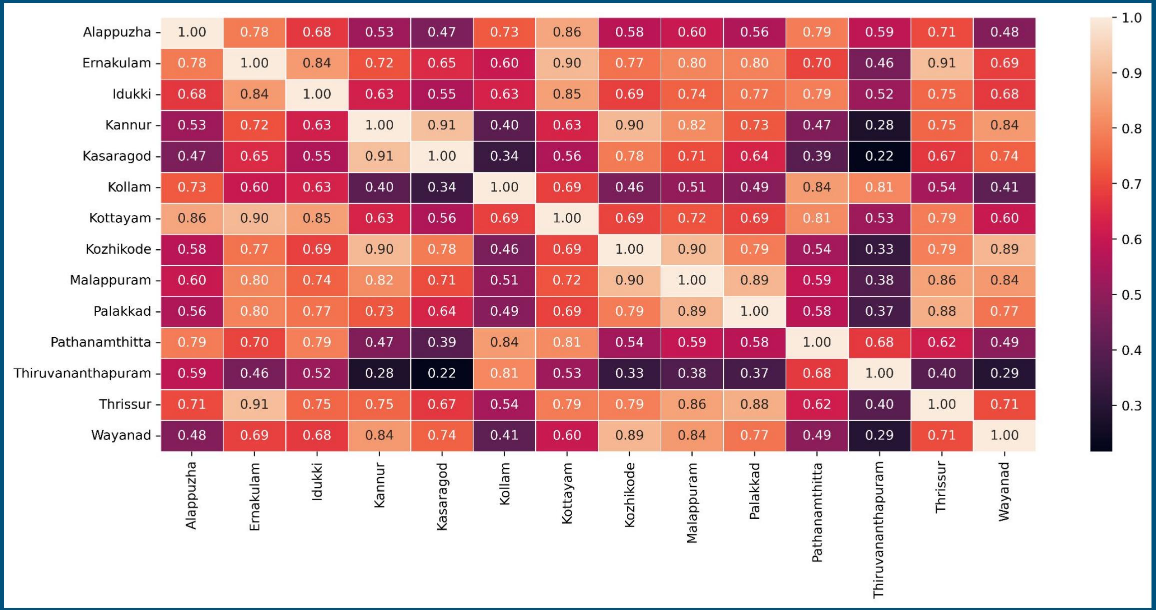



**Figure 4: Original features**

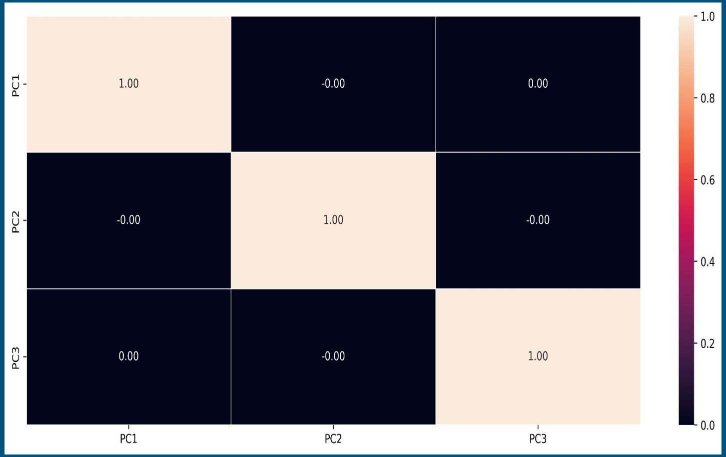
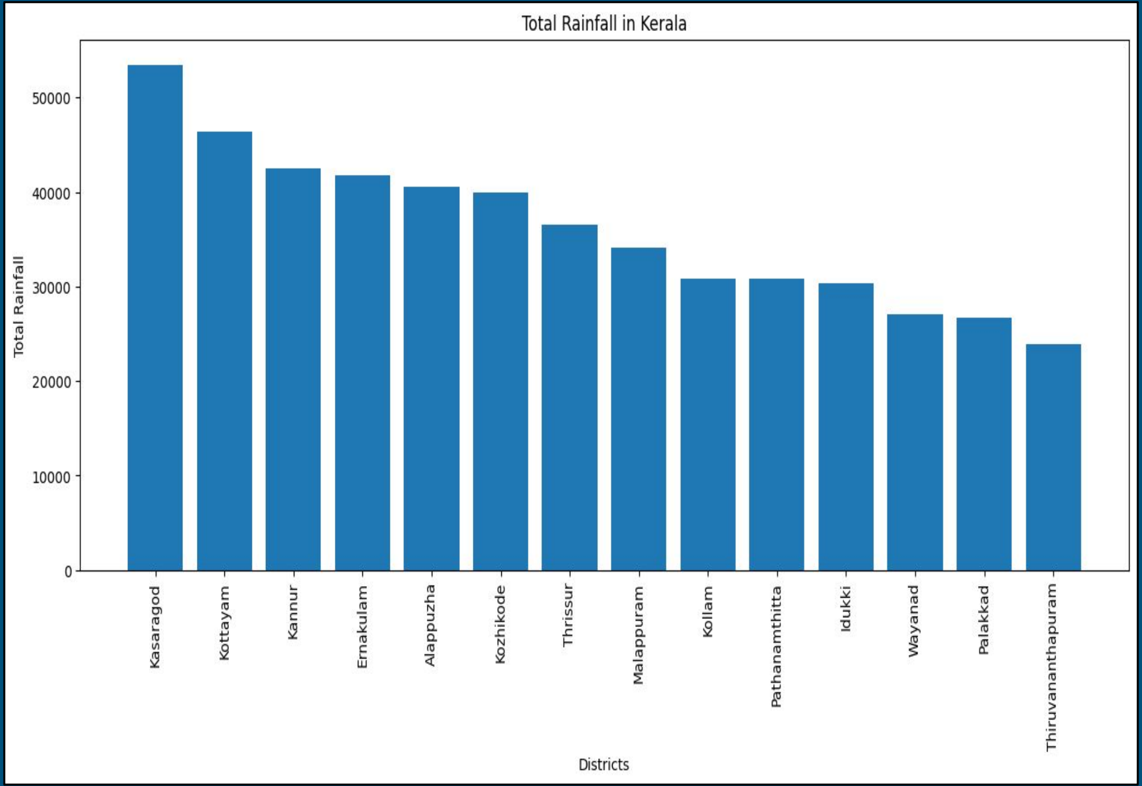
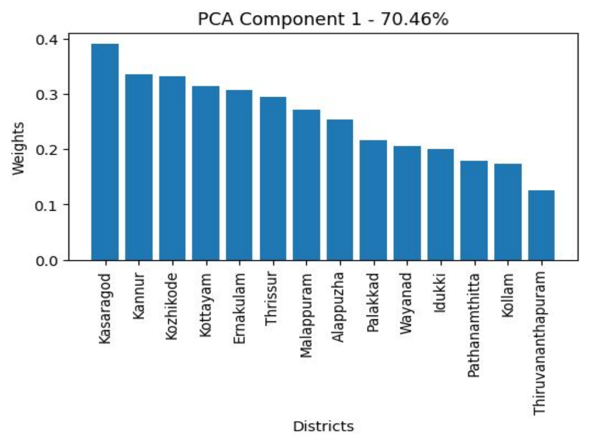


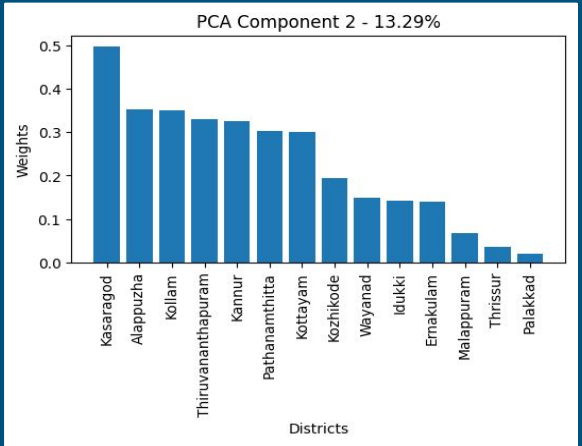

**Figure 5: Reduced features**

- ❖ As depicted in Figure 4 and 5, the dimensions are reduced and now the features are uncorrelated with each other.

# Correlation of Districts in a state
# Example 1 - Kerala



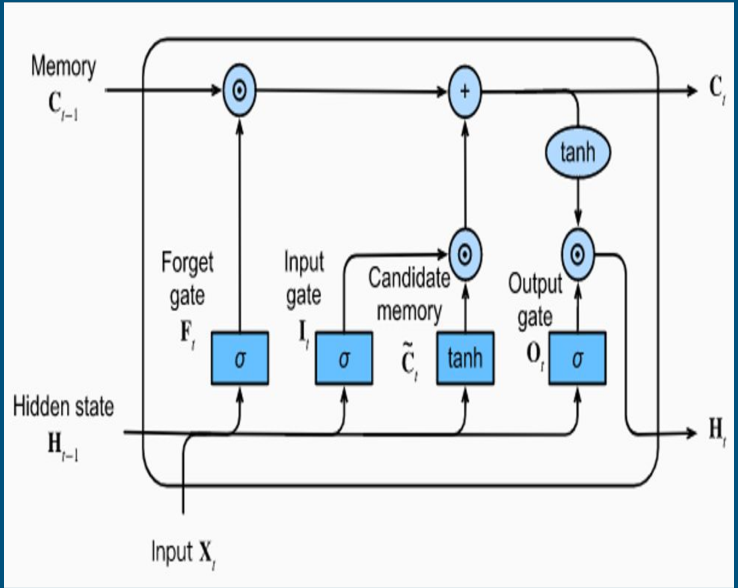



# Forecasting Rainfall using LSTM

- ❖ LSTM networks are designed to capture long-range dependencies in sequential data, making them well-suited for modeling temporal patterns in rainfall data.

- ❖ Key components include memory cells and gates (input, forget, and output), enabling the network to retain and utilize relevant information over time.



# Forecasting Rainfall using LSTM

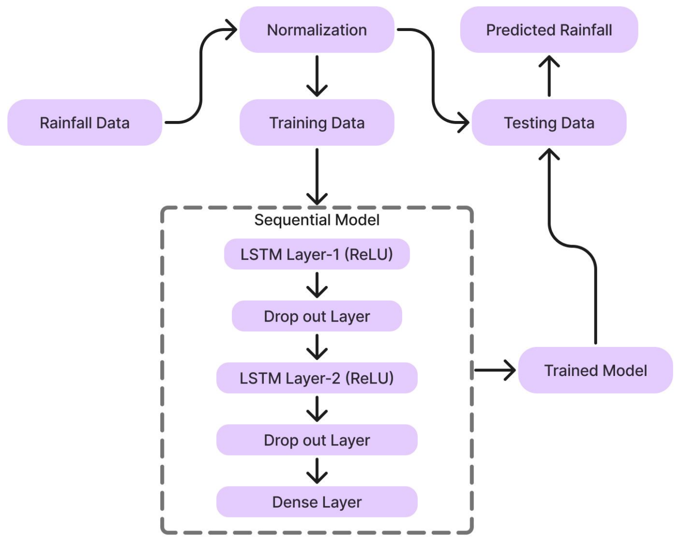



**Figure 6: Architecture of LSTM**

- ❖ The dataset has month-wise rainfall for each sub division of India from 1901 - 2015
- ❖ Figure 6 shows the architecture of the implemented LSTM model.
- ❖ Root Mean Square Error to evaluate the performance of the model

# Prediction on Training Data

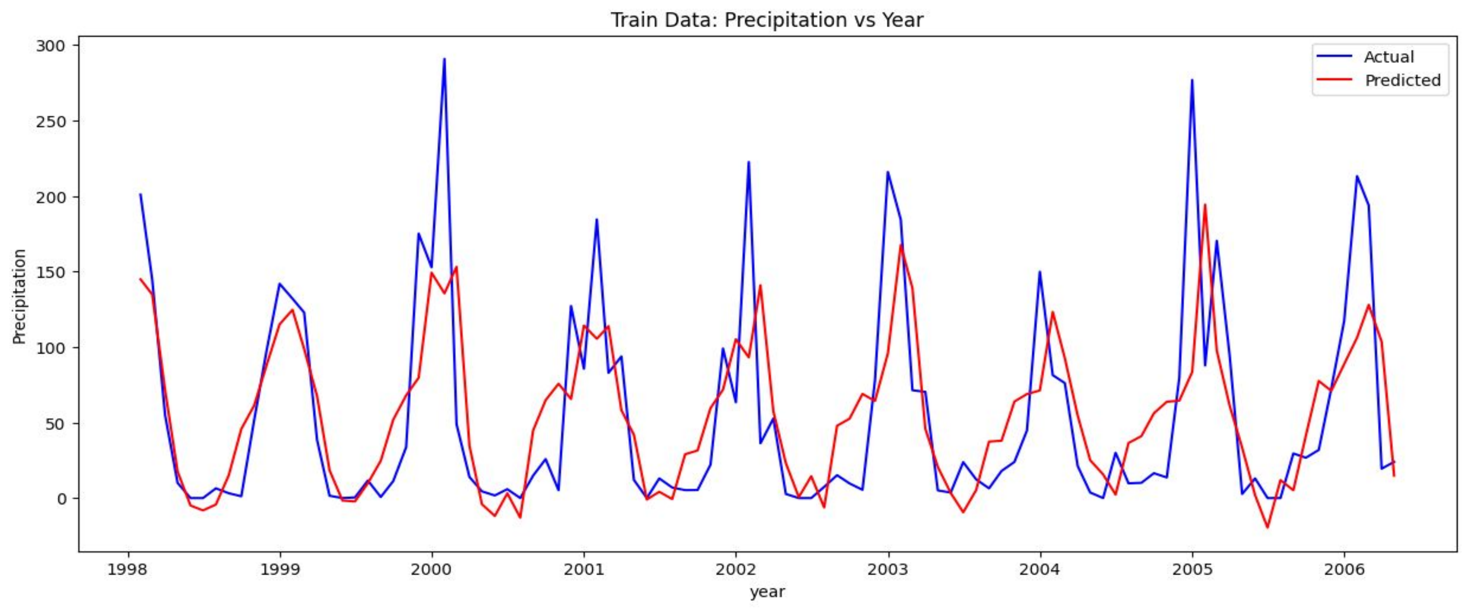



Figure 7: Predicting Rainfall using Train Data,
Train RMSE: 88.72

# Prediction on Test Data

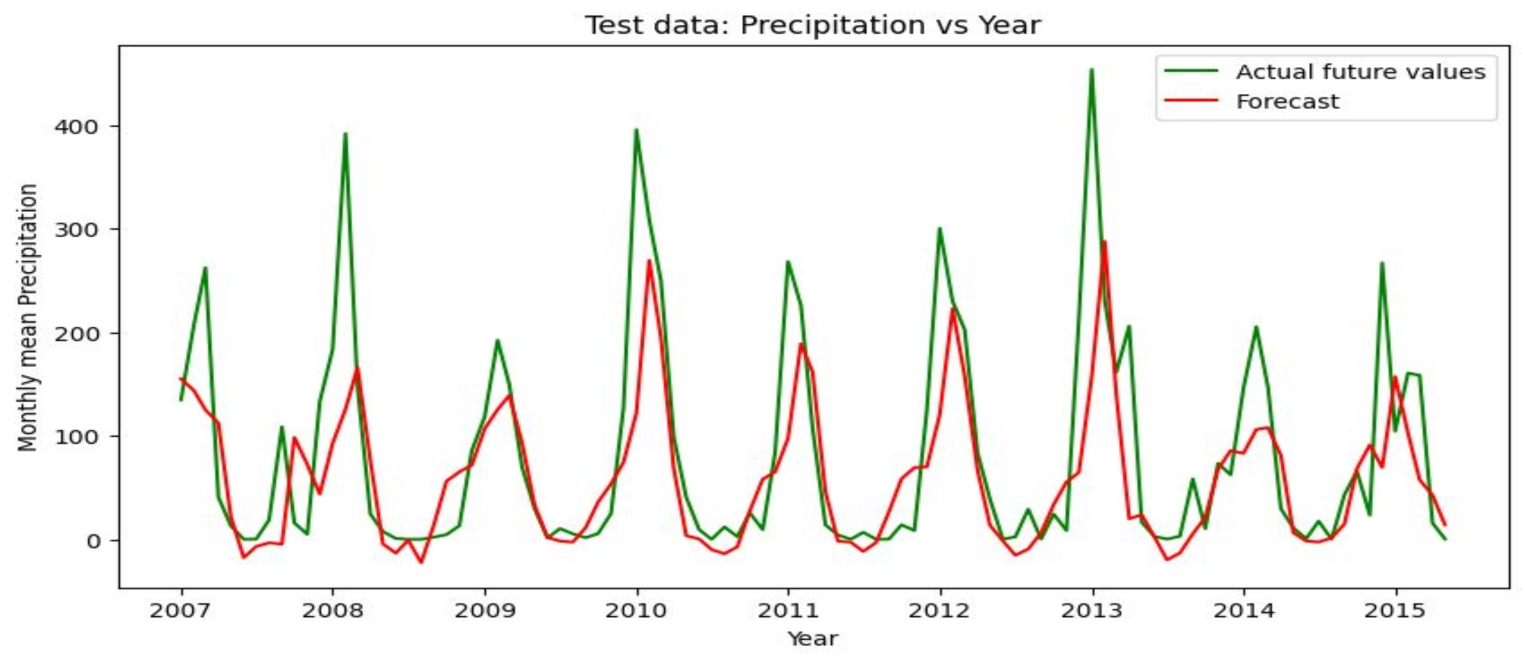



Figure 8: Predicting Rainfall using Test Data,
Test RMSE: 132.01

# Loss vs Epochs

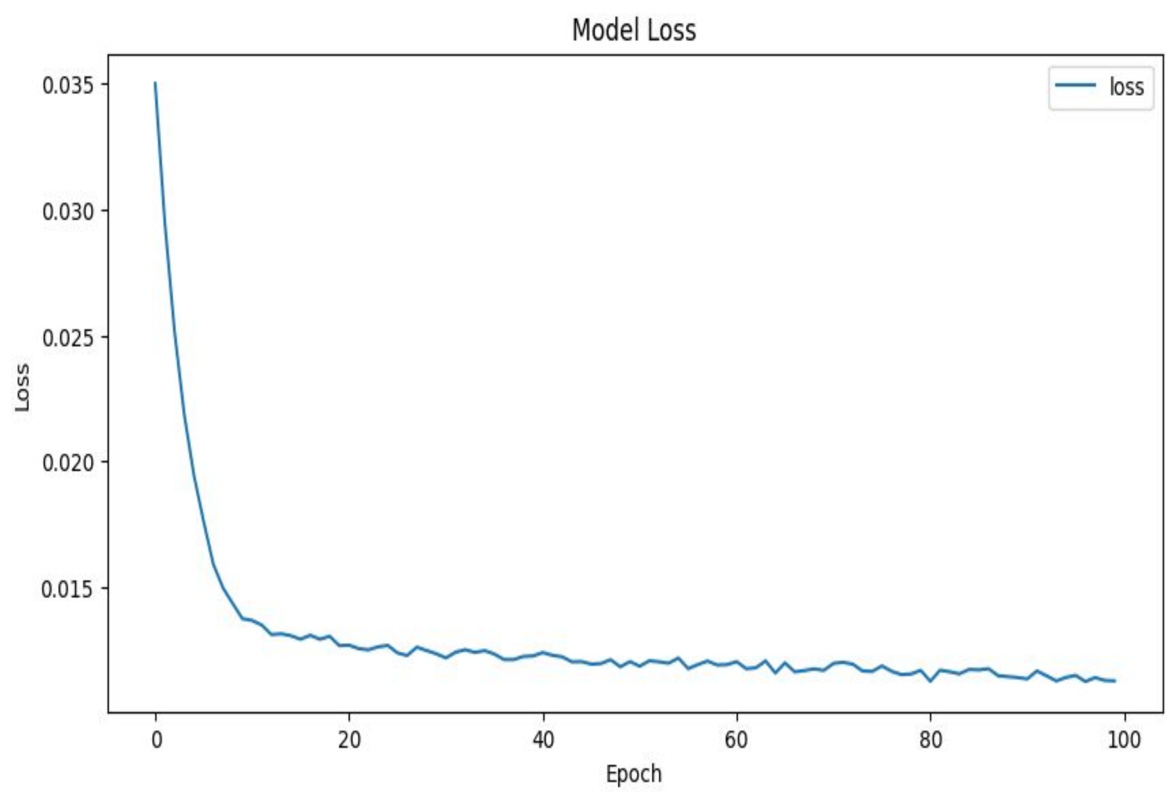
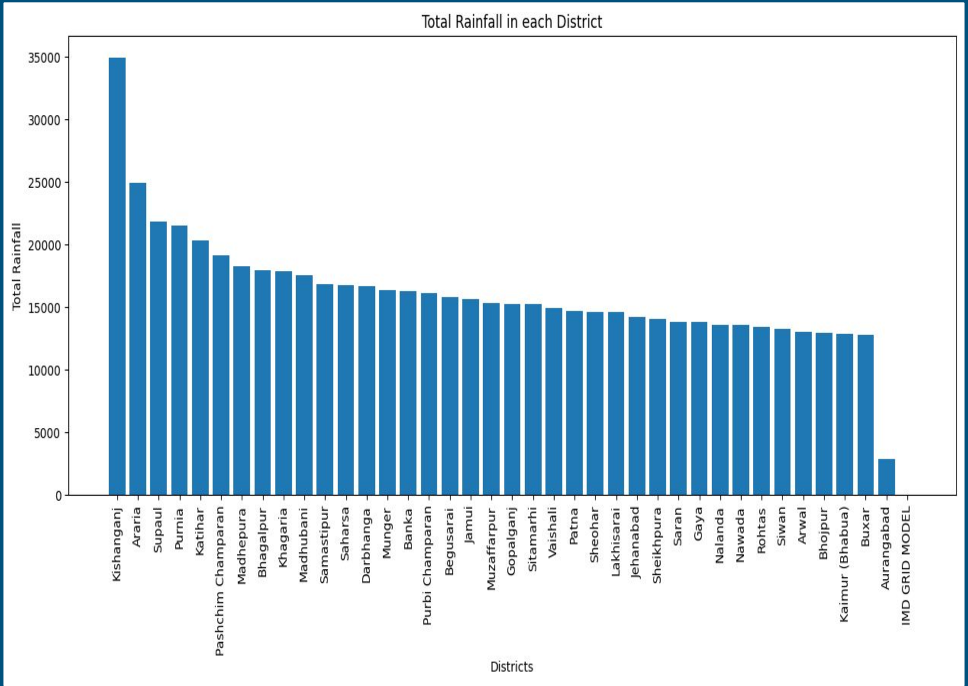
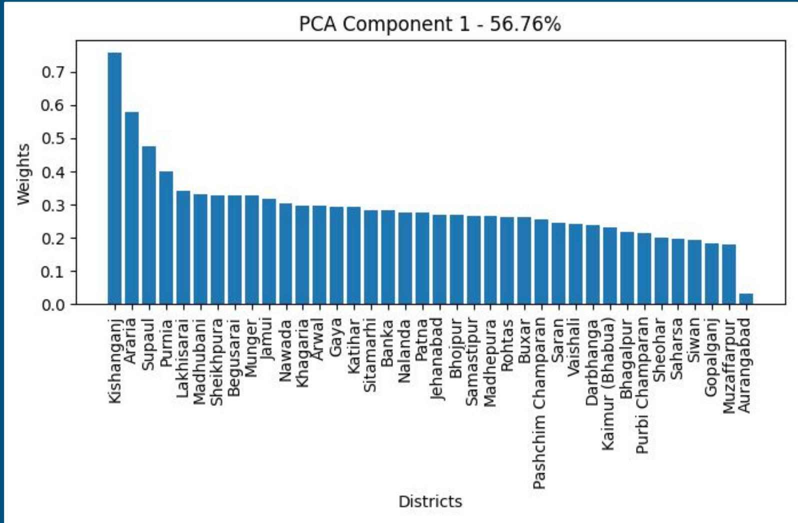



**Figure 9: Epoch vs Loss graph**

❖ Figure 9 shows the loss vs epochs graph during the training time.

# Example 2 - Bihar



# Conclusion

- ❖ The trend analysis serves as a foundational understanding of how precipitation evolves over time, crucial for climate assessment.
- ❖ The investigation into district-level features using Principal Component Analysis has highlighted pivotal districts significantly influencing rainfall patterns.
- ❖ The integration of LSTM model for rainfall forecasting has shown promising results in capturing complex temporal dependencies in precipitation data.
- ❖ Future research work includes refinement of accuracy of the forecasting models, incorporating additional environmental variables.