Data Analysis by Recurrent Neural Network, to Predict Seasonal Rainfall for Meteorological Monsoon Season and Trend of Deviation from Normal

Sumana Chatterjee[*]

Ph.D. Scholar (Computer Science), Nirwan University, Jaipur, Rajasthan, India

*Corresponding Author: sumana.spssaha.chatterjee@gmail.com

ABSTRACT

This paper is based on data analysis by python programming language on google collaborator platform, model used as recurrent neural network, to predict the seasonal rainfall total during monsoon season ,for near future years for the period June to September. Other than prediction of seasonal total rainfall for future years ,a study with data analysis in python had been done also for understanding the deviation of seasonal total for meteorological monsoon season in comparison with historical record of normal. The source of data is online data collection platform 'INDIA METEOROLOGICAL DEPARTMENT, PUNE'. The surface data for Alipore (42807) as available there, had been collected in csv file format, then uploading the csv file in google collaborator platform, executed analysis by python recurrent neural network model technique, in keras, tensorflow environment to get the predicted output for future years for the dependent variable, in this study which was 'total seasonal rainfall' within meteorological monsoon period. Trend of deviation of rainfall seasonal total for the meteorological monsoon season, in comparison with recorded normal data of the same was also studied since historical period, 1969 to present years, along with trend of deviation of the same for predicted future value for future years. One new column was created to define derived data for year wise seasonal total for meteorological monsoon season and performed the data analysis to find the result. From analysis of big data, determined predicted output for monsoon period seasonal total rainfall for near future which is the span of 2025 to 2030, with the help of recurrent neural network, with train-test split ratio 80%-20% and minimum loss.

***Keywords**: Python, AI, Google Collaborator, Train-Test, Neural Network Model, Activation Function, Optimiser.*

Introduction

The months from June to September is defined as the meteorological monsoon season for each year. From the portal, IMD, DSP, PUNE, data supply portal of India Meteorological Department, Pune, obtained two surface data file in csv format, table 2 and table 3. Then combined these two csv files suitably for the purpose of analysis with maximum number of relevant columns. The total rainfall occurred during four months June to September, is called as seasonal total rainfall for monsoon season for that year. With the help of the python code, created new column with total seasonal rainfall for monsoon, group by year column, starting from the historical year, 1969. Two downloaded csv files were used for analysis, one with daily rainfall data since historical year along with other daily observation data for weather parameters like maximum temperature, minimum temperature, average wind, rainfall, sunshine hours, weather event noted in four quarters on each day, station level pressure, mean sea level pressure, dry bulb temperature, wet bulb temperature, dew point, relative humidity, visibility, vapour pressure, wind direction, low cloud type, amount of low cloud etc. and on the other hand, another csv file was the downloaded file with monthly rainfall data. Data analysis in python had been done with both files to get output with seasonal rainfall for predicted future years. From this existing data with this created new column, predicted the seasonal rainfall for monsoon season for approximately five future years 2025

onwards to get the pattern of seasonal rainfall by recurrent neural network RNN. From the portal, collected the record of normal of rainfall data, compared this with actual occurrence of rainfall during this season under study and observed the deviation with output of value as well as graphical interpretation. Determined the deviation for predicted output for near future years also.

Literature Review

The analysis of data in this study is based on recurrent neural network, RNN. This is basically applicable for sequential data ,which performs same task for every element with sequence having memory of neural network model for the process executed previously and using back propagation techniques ,updating weights after each epoch to obtain better accuracy after each execution . Neural network model is model similar to human brain. The seasonal rainfall prediction for monsoon season with help of RNN could be done by RNN back propagation technique updating weights and biases each time to get output with maximum accuracy. The equation is with hidden layer, weights related with time t-1,t and t+1. Various papers based on rainfall prediction and other weather prediction were studied to compare idea and process of analysis of this study. Basically concept of execution of python code based on neural network on google collaborator platform was applied to get the output with prediction of amount of seasonal rainfall during monsoon in future years.
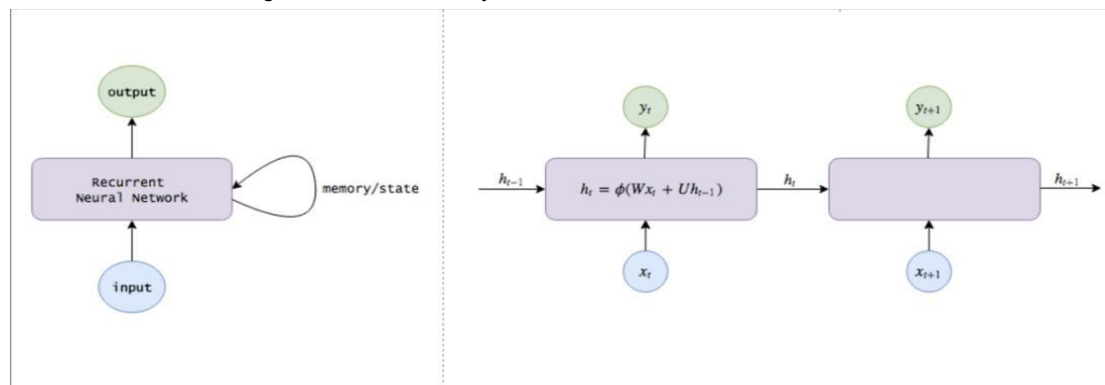

Image of RNN process flow

Research Gap

In this research study ,the data analysis had been done with two csv files ,one file consists of all types of meteorological weather parameter with daily observation and other file consists of year and month wise seasonal rainfall data for the station under study . Data analysis was done with both of these csv data files to get assurance with result of output and observed almost same result in each case with amount of seasonal rainfall total for predicted future years during monsoon season. Big data since historical years as obtained from two csv files, the file with all weather parameters since 1969 to 2023 and another file with monthly rainfall data from 1969 to 2024, both file was subjected to thorough analysis by recurrent neural network to get prediction with amount of seasonal rainfall. At the end of analysis for each case of the two csv files, obtained almost same result with amount of seasonal rainfall for few future years. The analysis was done not only with these two csv data files but analysis was also executed in five python files, some times data set was taken under consideration with the data file with all weather parameters and some times file with monthly rainfall data was taken under consideration. Chatgpt code was also subjected to analysis run on google collaborator platform, modified as required, to compare the output, where also these two csv files were taken under consideration separately. Moreover with the file with all types of weather observation data, the analysis with LSTM model was done also and there also it was predicted with the same result.

Research Questions /Hypothesis

For this research study, execution was done in total five python files, on the google collaborator platform for each. The way followed was of two types, taking consideration of two types of csv files and another type was consideration of two types of code i.e. manual python code and code provided by chatgpt and GEMINI. The study was to determine the quantity of the seasonal rainfall for the monsoon season that is the seasonal rainfall for the season with months June, July, August, September for future few years. Analysis by python neural network model could make the process easy to get the output of the same for the station under study, Alipore, station code 42807. The analysis was done with two csv files,

as mentioned, uploaded in google collaborator content folder, two separate files were executed, both with code of neural network but with two different csv files, one with daily meteorological observation data, seasonal rainfall column as derived feature in terms of daily rainfall data and for other csv file, the seasonal rainfall column as downloaded file in the csv file as reported by the department. Analysis with these two csv data files completed and almost nearly same output was obtained at the end. The another two files on google collaborator platform had been executed with the code provided by AI, CHATGPT and GEMINI to cross verify with the output using both of the csv files as mentioned. The result with output as obtained was noticed to be almost same, in the range 1300-1400 mm. Another python file with LSTM model was also executed with csv file with all the weather parameters where the output variable was considered as the difference between daily rainfall data and seasonal rainfall. Based on this extrapolation was done to determine the deviation of seasonal total rainfall for monsoon season for future few years and there also noticed the deviation by 100 mm, with increasing trend.

Methods

The whole analysis had been done with two csv files, downloaded from the 'data supply portal IMD Pune'. The analysis had been done with five python files, executed on google collaborator platform. The first two python files had been used with code of neural network as usual according as compatibility of the data files and the other code provided by CHATGPT and GEMINI, uploaded on google collaborator platform, modified as required there. The fifth python file was executed to cross verify the result, it was done by LSTM, considering csv file with daily weather observation data and the output column, considered there was difference amount from recoded data for monsoon seasonal rainfall.

- **Analysis through the python file run on google collaborator platform, csv file uploaded on content folder, the csv file with daily meteorological observational data**. In this python file as usual, at first, imported necessary python packages from python library, like numpy, panda, matplotlib, seaborn, etc for execution of python file in google collaborator. At the first step uploaded csv data file from content folder to python file. After successful uploading, checked the data columns, data type, which columns are of integer type and which are float type or object type. In the data file, year, month and date value were obtained as four different columns. So for the purpose of analysis one new column 'Date' in date format was to built taking all necessary columns related to date. Sorted the data with respect to year value of 'YEAR' column. Converted the month value in 'Date' column to proper month format. As in this research, the analysis was subjected to the data for meteorological monsoon period, that means duration for the months, June to September, so filtering of the data set was done for this span. Then arranged this filtered data group by year since historical year to present. Then performed necessary feature engineering and scaled the data. After that split the data set in 80-20 % ratio to build the model with RNN, recurrent neural network model to compile the model with 'Adam' optimiser, activation function as 'relu' and loss as 'MSE' or mean squared error. Mean squared error is simply the average of the square of difference between real value and the model predicted value. Training of the model was with epochs 100 to predict the seasonal rainfall total for monsoon season for few future years along with prediction of seasonal total for future years. Future years had been considered from 2025 to 2030. Simple recurrent neural network model performance was executed to get probable value of total seasonal rainfall for monsoon period for the years 2025 to 2030. As obtained from record of normal data for rainfall for the monsoon season, the data for recorded seasonal total for monsoon season is 1345.4 mm. Year wise deviation of actual monsoon season rainfall since historical period was determined along with visual understanding by image plot, to get idea about the deviation of the total seasonal rainfall during monsoon, from recorded normal data. The deviation with future prediction also was obtained from analysis along with visual image plot also. The model accuracy was good as obtained there and obtained as 83%.

- **Analysis through the python file run on google collaborator platform, csv file uploaded on content folder, the csv file with year and monthwise rainfall data.** This analysis was done with the csv file, obtained from the IMD Pune data supply portal. This file consists of month wise rainfall data. This data file initially as obtained was in text format. So converted this to csv delimited type and during this conversion, some 'unnamed' column was created automatically which had to be dropped out. Some rows without year value were also created during conversion. So filtered the data set with year not equal to zero. Then sorted the data set year wise. Then sum up the value of column indicating months June to September to some newly created column. Arranged this data with sum-up rainfall data with calculated value of seasonal

rainfall for monsoon, grouped by year and ultimately determined seasonal rainfall for monsoon with these months subjected to study, by the process of neural network to predict the same for future years. Obtained almost same output with amount of rainfall nearly 1400 mm for future years, as obtained from the model prediction, the accuracy rate also same as first analysis. The deviation from normal was also interpreted by the visual graph in the form of image plot. Here also same type of feature engineering was done for compliance of the python code of model performance. The model with simple recurrent neural network was performed with train-test split ration as 80-20 % ratio, activation function as relu, optimiser as adam, future years were considered as same from 2025 to 2030. Here also comparison or deviation with seasonal normal was tested along with visual plot to understand the output for historical as well as future year predicted value.

- **Analysis through the python file run on google collaborator platform, csv file uploaded on content folder, the csv file with historical weather observation data for all weather parameters, code provided by chatgpt and that code corrected as necessary for correction.** The code as obtained from chatgpt, was copied and pasted on the google collaborator platform. This time the analysis was done with the first file with the weather parameters of daily observation. After uploading this csv file, converted the date column into date time format and filtered the months for meteorological monsoon season similar as before for analysis. Grouped year wise, normalised the data and built model of neural network. Here also optimiser was used as 'adam', same activation function was used. 'relu', epoch was used as '200', loss function as 'mse'. Here the predicted value of seasonal total rainfall for monsoon for years 2025 to 2030 was around 1400 mm.

- **Analysis through the python file run on google collaborator platform, csv file uploaded on content folder, the csv file with year and monthwise rainfall data, code provided by chatgpt and that code corrected as necessary for correction.** The analysis has been done here with the second file of weather data consisting of monthly rainfall data. As described earlier this data file was obtained as text format and converted into delimited csv file for analysis compatibility. After suitable feature engineering and normalization of data, built sequential neural network model with activation function 'relu', optimiser 'adam', loss as 'mse', epochs as '200' and train-test split ratio as 80%20 %. Obtained the amount of rainfall prediction, total seasonal rainfall for monsoon season as more than 1300 mm   but less than 1400 mm.

- **Analysis through the python file run on google collaborator platform, csv file uploaded on content folder, the csv file with historical weather observation data for all weather parameters, analysis done by LSTM model.** In this case the deviation of rainfall data from the data of recorded normal was the dependent variable, subjected to analysis. The trend of deviation as obtained from this analysis, supported the result obtained from other analysis as mentioned above. In the csv file with daily weather observation data, created one new column with rainfall difference which is the difference between 'daily rainfall' and 'rainfall normal for that month'. The span of months under analysis is June to September, for the period 1969 to 2023, so the data was filtered for this period for the analysis purpose. After this with the help of neural network, considering the filtered data for this period, predicted the difference of seasonal rainfall for future years with neural network model, start year 2025 and end year 2030 with analysis frequency month wise. Thus predicted the difference as shown in the image 9.4 below. Thus, obtained yearly rainfall difference was obtained to reach the value to nearly 41 mm approximately on 2030, starting from 2025 with average rainfall deviation from normal as 11.6 mm.

Significance of the Study

The seasonal rainfall for monsoon season in Kolkata is one of the most important factor, upon which all other vital aspects such as ground water supply, agriculture, greenery, balance of temperature and humidity, ecosystem health and potential challenges depend actively. As mentioned, monsoon rainfall is the great source of water, which supports agriculture, can maintain city's greenery, can control balance of temperature and humidity after tremendous heat of summer. The hazard of water logging during excessive heavy rainfall in monsoon can be caused also damaging roads and low lying areas which require careful management and planning with precautionary measures for electrical short circuit etc. So, the amount of rainfall during monsoon season is certainly significant affecting all other vital factors which matter a lot. For this reason the prediction of total seasonal rainfall amount during monsoon is certainly have a vital role.
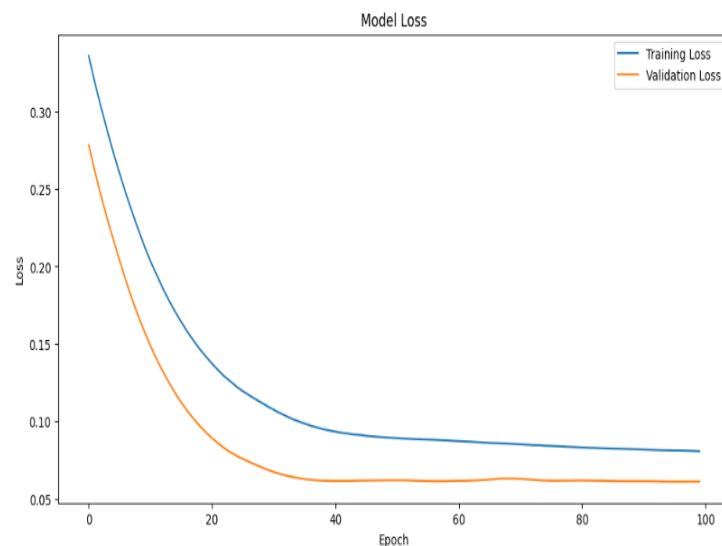
Timeline

   Time taken to collect the data from the online data supply platform IMD Pune, collection of csv data files ,time taken to process these files according as the compliance of the necessary analysis and also the time taken to execute the python programming properly to get the output with total seasonal rainfall as well as deviation with seasonal normal for monsoon months ,span during meteorological monsoon season ,determined by suitable model .Total five python files with two csv data files had to be executed ,so time taken to execute these python files in google collaborator with all necessary system - compatibility was the exact required time for this research study.
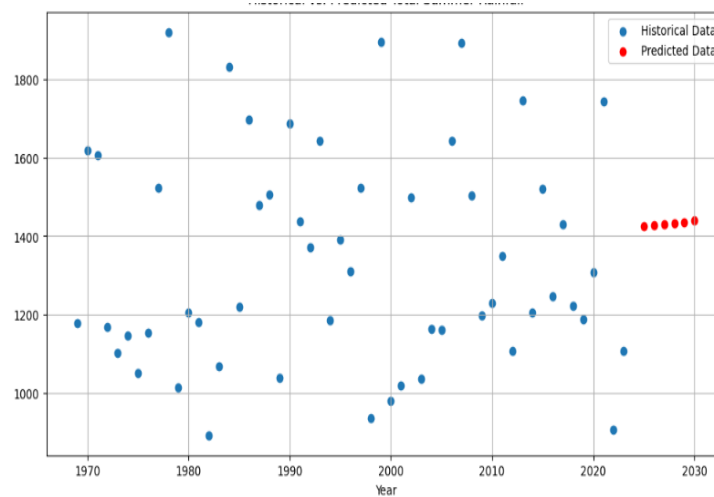
Conclusion and Future Work

   From this study it was to predict the total rainfall during monsoon season. In future it may be tried to predict, to determine the onset date of monsoon, from analysis of historical data with date of monsoon, data since historical past years along with other data as relevant with this factor such as sea surface temperature etc., data to be collected from suitable online platform.
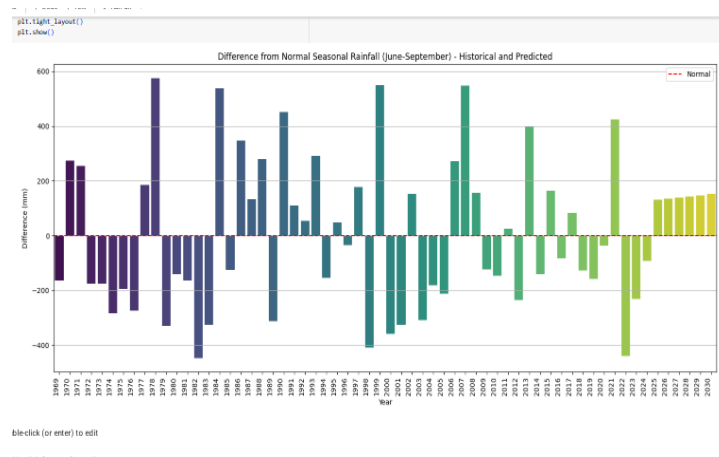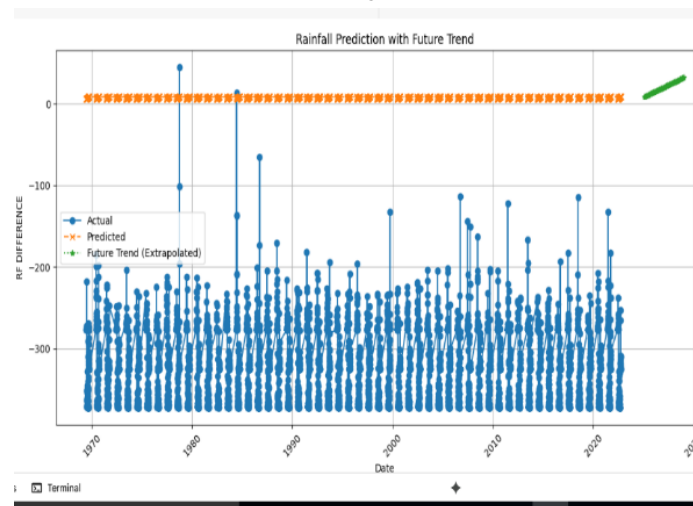
Image of Models and Few Screenshots of Plots as Obtained

The Plot of Model Accuracy/Loss

The Plot of Monsoon Season Rainfall with Prediction(mm)

Model by RNN

Model by LSTM

The plot of difference of monsoon seasonal total rainfall, difference with historical record of normal along with prediction of same for future years

References

1. Ferreira, D. D. (2021). Rainfall Prediction by Articial Neural Networks Trained using Different Climate Variables. *Research Square*, 1-17. doi:doi.org/10.21203/rs.3.rs-569733/v1

2. Kumar, A. (2022). Analysis and prediction of rainfall using machine learning techniques. *IRJET*, 1908-1915.

3. Kumar, P. (2020). Rainfall prediction using artificial neural network and sequential modelling. *IJRTI*, 1-9.

4. Mislan. (2015). Rainfall Monthly Prediction Based on Artificial Neural Network:A Case Study in Tenggarong Station, East Kalimantan -. *ScienceDirect*, 142 – 151. doi:10.1016/j.procs.2015.07.528

5. Nikalje, S. D. (2022). Rainfall Prediction and Analysis: A Review. *IJAEM*, 728-731. doi:DOI: 10.35629/5252-0403728731.

○○○