Rainfall Time Series Analysis Report

1. INTRODUCTION

This report presents an analysis and forecast of daily highest recorded rainfall using a time series model. The objective is to predict future rainfall trends for decision-making and climate study purposes.

2. BUSINESS CASE

Develop a Machine Learning (ML) model to predict the rainfall for the next one year with the help of previous data, Also make forecast about highest rainfall occurred in a day for a single month.

3. OBJECTIVES

- Prediction of rainfall accurately or next one year.
- Forecasting the rainfall for occurring highest rainfall in a single day.

4. PROJECT GOAL

1. Data understanding and preprocessing.
2. ML model to Forecast.

5. DATA OVERVIEW

- **Dataset:**
  - rainfall-monthly-highest-daily-total.csv
  - rainfall-monthly-number-of-rain-days.csv
  - rainfall-monthly-total.csv

- **Columns:**
  - Month: The date of recorded rainfall
  - Total rainfall: Highest Rainfall
  - Maximum_rainfall_in_a_day: Maximum rainfall recorded in a day (in mm)
  - Number_of_rainy_days: Total number of rainy days

- **Time Period: Historical daily rainfall data.**

6. METHODOLOGY

❖ **Data Extraction:**

- Data extraction is a fundamental process in data management, and it plays a critical role in preparing data for analysis, machine learning.

❖ **Data preprocessing:**

- Data Cleaning: Cleaning data from dataset.
- Handling Missing Values: Removal of rows or columns with excessive missing values.

❖ **Data Transformation:**

- Normalization and Scaling: Rescaling numerical data to a specific range (e.g., 0 to 1) or standardizing it to have a mean of 0 and a standard deviation of 1.

❖ **Time Series Line Plots:**

- This is the most common plot. It shows rainfall amounts over time, allowing us to see how rainfall changes from day to day, month to month, or year to year.

❖ **Models:**

1. ***ARIMA (Autoregressive Integrated Moving Average):***
   - ARIMA is a statistical model.
   - It's designed to capture linear dependencies in time series data.
   - It relies on the idea that past values of a time series can be used to predict future values.

2. ***LSTM (Long Short-Term Memory):***
   - LSTM is a type of recurrent neural network (RNN).
   - It's designed to capture complex, non-linear dependencies in sequential data.
   - It excels at handling long-term dependencies.

3. ***GRU (Gated Recurrent Unit):***
   - GRU is also a type of RNN, and a simplified version of LSTM.
   - It aims to achieve similar performance to LSTM with fewer parameters.

❖ Performance Metrics

- Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between predicted and actual values.
- Mean Squared Error (MSE): MSE calculates the average of the squared differences between predicted and actual values.

❖ Accuracy Plots

➢ ***Predicted vs. Actual Plots:***

- It plots the predicted rainfall values against the actual observed rainfall values over time.
- This is the most direct way to visualize accuracy.

➢ ***Scatter Plots (Predicted vs. Actual):***

- A scatter plot where the x-axis is the actual rainfall and the y-axis is the predicted rainfall.
- If the model is perfect, the points will fall along a straight diagonal line.

- ***Residual Plots:***
  - Residuals are the differences between the predicted and actual values.
  - These plots visualize the distribution of residuals.

- ***Error Distribution Plots (Histograms, Box Plots):***
  - These plots visualize the distribution of errors.
  - They show the frequency of different error magnitudes.

7. CHALLENGES

- Data Variability*:* Rainfall patterns can be highly variable and unpredictable.

- Model Complexity: Capturing the complex dynamics of rainfall requires sophisticated models.

8. SUMMERY

Summery of Rainfall Time Series project that depends on LSTM, GRU and ARIMA models below is the summery report of our models that gives us correct information.

Date, ARIMA Forecast, ARIMA Lower, ARIMA Upper, GRU Forecast, GRU Lower, GR U Upper, LSTM Forecast, LSTM Lower, LSTM Upper

Date	ARIMA For	ARIMA Lo	ARIMA Up	GRU Forec	GRU Lowe	GRU Uppe	LSTM Fore	LSTM Low	LSTM Upp
01-02-1982	0.387678	-0.0941	0.869461	0.378948	-1.62105	2.378948	0.364032	-1.63597	2.364032
01-03-1982	0.447397	-0.05172	0.946511	0.389453	-1.61055	2.389453	0.364929	-1.63507	2.364929
01-04-1982	0.48481	-0.02304	0.992663	0.405249	-1.59475	2.405249	0.37125	-1.62875	2.37125
01-05-1982	0.441278	-0.07626	0.958818	0.388939	-1.61106	2.388939	0.380237	-1.61976	2.380237
01-06-1982	0.467757	-0.07912	1.014636	0.387578	-1.61242	2.387578	0.383942	-1.61606	2.383942
01-07-1982	0.41842	-0.14599	0.982833	0.403353	-1.59665	2.403353	0.388844	-1.61116	2.388844
01-08-1982	0.441283	-0.15159	1.034154	0.410443	-1.58956	2.410443	0.407066	-1.59293	2.407066
01-09-1982	0.445864	-0.16342	1.055149	0.385951	-1.61405	2.385951	0.430196	-1.5698	2.430196
01-10-1982	0.453419	-0.17269	1.079525	0.362664	-1.63734	2.362665	0.430027	-1.56997	2.430027
01-11-1982	0.441238	-0.20085	1.083324	0.365367	-1.63463	2.365367	0.382654	-1.61735	2.382654
01-01-1983	0.445853	-0.21528	1.10699	0.369029	-1.63097	2.369029	0.361178	-1.63882	2.361178
01-02-1983	0.441508	-0.23613	1.119144	0.375009	-1.62499	2.375009	0.355307	-1.64469	2.355307
01-03-1983	0.445314	-0.24926	1.139889	0.380175	-1.61983	2.380175	0.348787	-1.65121	2.348787
01-04-1983	0.444559	-0.26551	1.154628	0.386779	-1.61322	2.386779	0.360851	-1.63915	2.360851
01-05-1983	0.44578	-0.27994	1.171498	0.399085	-1.60091	2.399086	0.355606	-1.64439	2.355606
01-06-1983	0.44377	-0.29708	1.184617	0.384086	-1.61591	2.384086	0.350027	-1.64997	2.350027
01-07-1983	0.44467	-0.31144	1.200775	0.380488	-1.61951	2.380488	0.347108	-1.65289	2.347108
01-09-1983	0.44421	-0.32653	1.214954	0.37717	-1.62283	2.37717	0.373694	-1.62631	2.373694
01-10-1983	0.444813	-0.3404	1.230029	0.402291	-1.59771	2.402291	0.375611	-1.62439	2.375611
01-11-1983	0.444487	-0.35482	1.243794	0.394958	-1.60504	2.394958	0.388106	-1.61189	2.388106
01-01-1984	0.444693	-0.36856	1.257951	0.39987	-1.60013	2.39987	0.388797	-1.6112	2.388797
01-04-1984	0.444411	-0.38251	1.271336	0.37132	-1.62868	2.37132	0.38912	-1.61088	2.38912
01-05-1984	0.444583	-0.39584	1.285005	0.370587	-1.62941	2.370587	0.401842	-1.59816	2.401842
01-07-1984	0.444508	-0.40915	1.298167	0.389418	-1.61058	2.389418	0.421366	-1.57863	2.421366

Summery of forecast of highest rainfall occurred in a day for a single month.

Date, Forecast, Lower Bound, Upper Bound

Date	Forecast	Lower Bound	Upper Bound
02-01-1970 00:00	0.387678446	-0.09410369	0.869460582
03-01-1970 00:00	0.447396606	-0.051718059	0.946511272
04-01-1970 00:00	0.484809693	-0.023043211	0.992662597
05-01-1970 00:00	0.441278084	-0.076262046	0.958818214
06-01-1970 00:00	0.467757198	-0.079121744	1.014636139
07-01-1970 00:00	0.418419826	-0.145993687	0.982833339
08-01-1970 00:00	0.44128307	-0.15158832	1.03415446
09-01-1970 00:00	0.445863562	-0.163422003	1.055149127
10-01-1970 00:00	0.453419151	-0.172686746	1.079525047
11-01-1970 00:00	0.441238052	-0.200847671	1.083323775
12-01-1970 00:00	0.44585311	-0.215284136	1.106990357
13-01-1970 00:00	0.441507812	-0.236128855	1.119144479
14-01-1970 00:00	0.445314003	-0.249260534	1.139888539
15-01-1970 00:00	0.444558875	-0.265510072	1.154627823
16-01-1970 00:00	0.445779656	-0.279938814	1.171498126
17-01-1970 00:00	0.443769977	-0.297077113	1.184617067
18-01-1970 00:00	0.444669549	-0.311436338	1.200775435
19-01-1970 00:00	0.44420989	-0.326533776	1.214953555
20-01-1970 00:00	0.444813003	-0.340402751	1.230028758
21-01-1970 00:00	0.444487037	-0.35481994	1.243794014
22-01-1970 00:00	0.444693164	-0.368564296	1.257950624
23-01-1970 00:00	0.444410901	-0.382513774	1.271335577
24-01-1970 00:00	0.444583256	-0.395838251	1.285004763
25-01-1970 00:00	0.444508426	-0.409149745	1.298166598
26-01-1970 00:00	0.444599651	-0.422113305	1.311312607
27-01-1970 00:00	0.444527446	-0.435030724	1.324085616
28-01-1970 00:00	0.444566018	-0.447669401	1.336801438
29-01-1970 00:00	0.444527301	-0.460198052	1.349252654
30-01-1970 00:00	0.444557964	-0.472494067	1.361609995
31-01-1970 00:00	0.444542599	-0.48466597	1.373751168

9. CONCLUSION

- The ARIMA model successfully predicts the highest rainfall for the next month.
- LSTM models are a powerful tool for working with sequential data.
- GRUs are a powerful and efficient type of RNN that excels at processing sequential data. They offer a good balance between performance and computational efficiency.
- All of three models LSTM, GRU, and ARIMA have nearly identical performance based on Mean Absolute Error and Root Mean Squared Error.
- The forecast can assist in weather preparedness, disaster management, and agricultural planning.