Meteorological Data Analysis and Prediction

A Data-Driven Assessment Using Automated Weather Stations at Aries and Machine Learning Algorithms

By

Asmita Chhabra

Under the guidance of

Mr. Samaresh Bhattacharjee

July 1, 2025

Department of Data Science & Economics,

FLAME University

Table of Contents

1. Introduction………………………………………………………….. 1

2. Project Outline………………………………………………………. 2

2.1 Data Acquisition and Cleaning ………………………………………………… 2
2.2 Database Design and Setup …………………………………………………… 2
2.3 Web Application Development………………………………………………… 2
2.4 Predictive Modeling Module…………………………………………………… 3
2.5 Reporting and Export Features………………………………………………... 3

3. Objectives ……………………………………………………… 4

4. Data Collection & Preprocessing…………………………………….. 5

4.1 Data Sources ……………………………………………………………… 5
4.2 Raw Data Characteristics……………………………………………………….. 5
4.3 Quirks and Inconsistencies …………………………………………………….. 5
4.4 Cleaning Steps ……………………………………………………………….. 6
4.5 Derived Features and Aggregations…………………………………………….... 6
4.6 Sample: Raw Data vs Cleaned Data ……………………………………………. 8

5. Database Schema and Design ……………………………………… 11

5.1 Overview 11
5.2 Table Diagram ……………………………………………………………... 11
5.3 Table Structures ……………………………………………………………. 11
5.4 Aggregation and Classification Views (Columns Names) ……………………… 12
5.5 Design Choices & Highlights …………………………………………………. 13

6. Web Application Architecture ……………………………………… 14

6.1 Technology Stack ………………………………………………………….... 14
6.2 Application Structure & Navigation ………………………………………….... 14
6.3 Main Modules/Pages 15
6.4 Backend Logic & Data Access …………………………………………….... 17

7. SQL Queries and App Functionalities ……………………………… 18

7.1 Getting Started …………………………………………………………… 18
7.2 Dashboard Overview 18
7.3 Common Dashboard Layout ………………………………………………….. 18
7.4 Detailed Tab Guide ……………………………………………………………. 19
7.5 Sidebar Features & Exports ……………………………………………….... 28
7.6 How AWS1 and AWS2 Dashboards Differ ……………………………….... 29
7.7 Notes on Machine Learning Model …………………………………………... 29

8. Machine Learning Model …………………………………………… 30

8.1 Overview …………………………………………………………………………… 30
8.2 Step-by-Step Guide ……………………………………………………………… 30
8.3 Rationale for Model Selection ………………………………………………... 35
8.4 Tips & Best Practices ………………………………………………………. 36

9. Data Analysis & Findings ……………………………………………... 37
9.1 Average Daily Temperature Trends (2018–2024) ………………………………. 37
9.2 Monthly Rainfall Trends and Monsoonal Patterns ……..…..……………………… 38
9.3 Year-on-Year Rainfall Comparison: 2021 vs 2022 ………………………………… 39
9.4 Extreme Rainfall Events (2018–2024) ……………………………………………... 40
9.5 Temperature Extremes and Shifting Baselines (2017–2025) ……………………. 41
9.6 Daily Temperature Comparison: 2019 vs 2023 …………………………………….. 43
9.7 Monthly Rainfall Anomalies and Distribution: 2019 vs 2021 …………………… 44
9.8 Machine Learning Model Evaluation for Temperature Prediction ………………. 46
9.9 Synthesized Key Insights and Implications ……………………………………… 48

10. Challenges & Learnings ……………………………………………….. 49

11. Conclusion 50

12. Future Scope …………………………………………………… 51

13. Appendix ……………………………………………………… 52
13.1 AWS Data Field Descriptions ……………………………………………….. 52
13.2 Sample Data Records ……………………………………………………… 53
13.3 Abbreviations and Acronyms ……………………………………………… 53
13.4 Additional Resources ……………………………………………………. 54

1. Introduction

Understanding the evolving dynamics of local climate is crucial for effective environmental management, disaster preparedness, and sustainable development. Nainital, a prominent hill station in the Indian Himalayas, is characterized by its unique weather patterns, which play a significant role in shaping its ecology, water resources, and livelihoods. Recent years have witnessed concerns about shifting rainfall patterns, temperature extremes, and the potential impacts of climate variability and change.

This report presents a comprehensive analysis of weather data collected from two automated weather stations (AWS1 located at the top of Manora Peak and AWS2 at ST Radar) in Nainital, covering the period from 2017 to 2025. Through rigorous statistical summaries, detailed visualizations, and advanced machine learning techniques, the study investigates:

- Trends and anomalies in temperature and rainfall at daily, monthly, and annual scales
- The occurrence and characteristics of extreme weather events
- Comparative insights across years, seasons, and stations
- The effectiveness of different machine learning models for temperature prediction

The findings provide valuable insights into the region's climatic stability and emerging variability, highlight the challenges of working with environmental datasets, and offer practical recommendations for monitoring and adapting to changing weather patterns. By integrating data-driven analysis with contextual understanding, this report aims to support local stakeholders, researchers, and policymakers in making informed decisions for Nainital's future resilience.

2. Project Outline

This project was executed in several structured phases to ensure a robust and user-friendly platform for meteorological data analysis and prediction. The main workflow is as follows:

2.1 Data Acquisition and Cleaning:

Meteorological data was provided for two weather stations in Nainital:

- AWS 1 (Manora Peak): Data from 2017 to March 2025
- AWS 2 (ST Radar): Data from 2021 to March 2025

  The raw data consists of high-frequency recordings, with entries for multiple parameters every 15 minutes throughout each day. The cleaning and preprocessing steps—detailed later in this report—were crucial for transforming the raw data into a structured, analysis-ready format.

2.2 Database Design and Setup:

The cleaned data was imported into a MySQL relational database. Special attention was given to schema design to facilitate flexible and efficient querying of temporal variability, allowing users to analyze patterns at daily, monthly, and yearly levels. Database views were created to aggregate the data at different time scales (daily and monthly), making it easier for both the backend logic and the end user to interact with summarized weather patterns.

2.3 Web Application Development:

Using Python and Streamlit, an interactive web application was developed and deployed via Render. The web app offers:

- Selection between AWS 1 and AWS 2 dashboards
- Querying and visualization of weather parameters (temperature, rainfall, etc.)
- User-friendly interfaces for exploring and comparing temporal trends
- Customizable data export (CSV, PDF reports, charts)

2.4 Predictive Modeling Module:

The application features a dedicated prediction interface where users can select different machine learning algorithms, train models on historical data, compare predicted and actual values, and generate forecasts for upcoming years—all within the web app.

2.5 Reporting and Export Features:

Users can easily export their queried results and visualizations, generate downloadable reports, and save data tables or charts for further offline analysis.

Intended Audience:

The platform is designed for researchers, students, meteorologists, and general users interested in exploring, analyzing, and forecasting meteorological variability using multi-year, high-frequency weather station data.

Note: Details on project scale, data cleaning, and the database schema will be provided in subsequent sections.

3. Objectives

- Empower users—including researchers, students, meteorologists, and the general public—to easily access, explore, and analyze multi-year weather data from two key stations in Nainital.
- Simplify access to high-frequency meteorological records through a unified, intuitive web platform, removing technical barriers to rich data exploration.
- Enable insightful analysis by providing tools for users to query, visualize, and compare weather patterns (temperature, rainfall, extremes) across different time periods and between AWS 1 and AWS 2.
- Facilitate discovery of trends shaping Nainital's meteorological landscape, supporting data-driven insights into local climate variability and change.
- Support predictive capabilities by integrating machine learning models that allow users to forecast future weather parameters and evaluate prediction performance.
- Promote transparency and reproducibility by allowing users to export data, charts, and reports for further offline analysis or collaborative research.
- Transform raw, complex datasets into actionable knowledge through robust data structuring, aggregation, and user-friendly interfaces.

These objectives serve to bridge the gap between raw meteorological data and meaningful insights, fostering a deeper understanding of weather dynamics in the Nainital region.

4. Data Collection & Preprocessing

4.1 Data Sources

Meteorological data was collected from two automated weather stations in Nainital:

- AWS 1 (Manora Peak): Parameters include station name, date, time, GPS, battery voltage, hourly and daily rain, soil moisture, temperature (average, max, min), wind speed/direction, pressure, humidity, and sun duration.
- AWS 2 (ST Radar): Parameters include station name, date, time, temperature (average, max, min), humidity, pressure, wind speed/direction, daily rainfall, solar radiation, UV-A, UV-B, battery voltage, and peripherals.

4.2 Raw Data Characteristics

- Formats: Excel and CSV files
- Frequency: 15-minute intervals for each parameter, resulting in high-frequency, multi-row daily data.

4.3 Quirks and Inconsistencies

- Missing values and empty columns
- Outliers and sensor errors in key parameters(e.g., impossible temperature readings due to sensor errors)
- Inconsistent units and column naming
- Non-uniform or incorrect date formats
- Data overlap and duplicate records across years (e.g., some files for 2017 containing data from early 2018)
- "Daily Rain" column often miscalculated in raw data and needed recomputation

4.4 Cleaning Steps

- Removed empty or irrelevant columns (e.g., latitude/longitude not needed for analysis)
- Standardized all date formats to YYYY_MM_DD
- Recalculated the "Daily Rain" column via SQL aggregation to correct cumulative errors
- Standardized column names and units across all files and both weather stations
- Created daily and monthly aggregated "view" tables from the 15-minute interval data for streamlined querying and visualization
- Added rainfall classification categories based on IMD standards using SQL logic (see below)
- Removed data overlap and duplicate records across years
- Ensured data consistency and completeness across the combined dataset

4.5 Derived Features and Aggregations

- Daily Aggregation: Created SQL views to summarize 15-minute raw data into daily averages/totals for temperature, rainfall, etc.
- Monthly Aggregation: Similar logic applied to produce monthly summary tables/views.
- Rainfall Classification: Used IMD guidelines to categorize daily rainfall using a SQL view.

Rainfall Category Definitions (as per IMD): source

Rain Category	Daily Rainfall (mm)
Very light rain (VLR)	0.1 – 2.4
Light rain (LR)	2.5 – 7.5
Moderate rain (MR)	7.6 – 35.5
Rather heavy rain (RHR)	35.6 – 64.4

Heavy rain (HR)	64.5 – 124.4
Very heavy rain (VHR)	124.5 – 244.4
Extremely heavy rain (EHR)	> 244.4
No rain	0 or missing

Sample SQL Query for Rainfall Classification

```
CREATE VIEW rainfall_classification AS
SELECT
  station_name,
  date,
  MAX(rain_daily_mm) AS total_rain_mm,
  CASE
    WHEN MAX(rain_daily_mm) BETWEEN 0.1 AND 2.4 THEN 'Very light rain (VLR)'
    WHEN MAX(rain_daily_mm) BETWEEN 2.5 AND 7.5 THEN 'Light rain (LR)'
    WHEN MAX(rain_daily_mm) BETWEEN 7.6 AND 35.5 THEN 'Moderate rain (MR)'
    WHEN MAX(rain_daily_mm) BETWEEN 35.6 AND 64.4 THEN 'Rather heavy rain (RHR)'
    WHEN MAX(rain_daily_mm) BETWEEN 64.5 AND 124.4 THEN 'Heavy rain (HR)'
    WHEN MAX(rain_daily_mm) BETWEEN 124.5 AND 244.4 THEN 'Very heavy rain (VHR)'
    WHEN MAX(rain_daily_mm) > 244.4 THEN 'Extremely heavy rain (EHR)'
    ELSE 'No rain'
  END AS rain_category
FROM data2_2021
GROUP BY station_name, date
ORDER BY station_name, date;
```

4.6 Sample: Raw Data vs Cleaned Data

Original dataset with errors and missing fields prior to data cleaning procedures. (4.a , 4.b)

Humidity(%)	Rain Daily(mm)	ST-5cm(Deg C)	Sun duration(Min)	LT(Deg C)	LW(Min)	Lat	Long	Alt	ST-20cm(Deg C)	ST3(Deg C)	ST4(Deg C)	ST5(Deg C)	ST6(Deg C)	ST7(Deg C)	ST8(Deg C)	ST9(Deg C)
25	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
26	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
27	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
27	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
29	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
30	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
33	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
34	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
34	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
35	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
35	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
37	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
39	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
40	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
40	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
40	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
41	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
42	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
41	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
43	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
44	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
45	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
45	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
46	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
48	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
48	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
49	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
50	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
50	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
51	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
51	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							
52	0	-40	0	-10	0	29.3585N	79.4578E	1945	-40							

Figure 4.a - Raw Data before Data Cleaning

Station ID	Station Name	Date	Time	GPS	Battery(V)	Rain Hourly(mm)	SM(m3/m3)	Temp(Deg C)	Temp Max(Deg C)	Temp Min(Deg C)	WS(m/s)	WD(Deg)	Pressure(hPa)
A0A769E4	NAINITAL	31/12/2019	18:30:00	L	12.2	0	0	4.8	4.9	4.4	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	18:45:00	L	12.2	0	0	4.7	4.9	4.4	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	19:00:00	L	12.2	0	0	4.6	5	4.4	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	19:15:00	L	12.2	0	0	4.8	4.9	4.5	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	19:30:00	L	12.2	0	0	4.7	4.9	4.5	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	19:45:00	L	12.2	0	0	4.8	4.9	4.5	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	20:00:00	L	12.2	0	0	4.7	4.9	4.5	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	20:15:00	L	12.2	0	0	4.8	4.9	4.7	0	359	809.1
A0A769E4	NAINITAL	31/12/2019	20:30:00	L	12.2	0	0	4.8	5	4.7	0	359	809.2
A0A769E4	NAINITAL	31/12/2019	20:45:00	L	12.2	0	0	4.8	5	4.7	0	359	808.9
A0A769E4	NAINITAL	31/12/2019	21:00:00	L	12.2	0	0	4.9	5	4.7	0	359	808.9
A0A769E4	NAINITAL	31/12/2019	21:15:00	L	12.2	0	0	4.9	4.9	4.6	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	21:30:00	L	12.2	0	0	4.7	4.9	4.6	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	21:45:00	L	12.2	0	0	4.7	5	4.6	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	22:00:00	L	12.2	0	0	4.7	5	4.6	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	22:15:00	L	12.2	0	0	4.8	4.9	4.5	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	22:30:00	L	12.2	0	0	4.5	4.9	4.5	0	359	808.8
A0A769E4	NAINITAL	31/12/2019	22:45:00	L	12.2	0	0	4.6	4.9	4.5	0	359	808.9
A0A769E4	NAINITAL	31/12/2019	23:00:00	L	12.1	0	0	4.8	4.9	4.5	0	359	808.8
A0A769E4	NAINITAL	31/12/2019	23:15:00	L	12.1	0	0	4.5	4.8	4.3	0	359	808.7
A0A769E4	NAINITAL	31/12/2019	23:30:00	L	12.1	0	0	4.3	4.8	4.2	0	359	808.9
A0A769E4	NAINITAL	31/12/2019	23:45:00	L	12.1	0	0	4.4	4.8	4.2	0	359	809
A0A769E4	NAINITAL	1/1/2020	0:00:00	L	12.1	0	0	4.3	4.8	4.2	0	359	809.1
A0A769E4	NAINITAL	1/1/2020	0:15:00	L	12.1	0	0	4.4	4.4	4.3	0	359	809.3

Figure 4.b - Raw Data before Data Cleaning

Processed data following cleaning, showing consistent and complete records. (4.c , 4.d)

Maximum Temperature	Minimum Temperature	Wind Speed	Wind Degree	Pressure	Humidity	Sun Duration
4.7	4.2	2.2	100	803.7	52	0
4.2	4.2	1.6	115	803.8	53	0
4.2	4	1.2	98	803.9	54	0
4.2	4	1.6	107	804	55	0
4.2	4	2.2	75	804.2	53	0
4.2	4.1	1.7	142	804.3	55	0
4.2	3.9	1	99	804.4	57	0
4.2	3.7	1.7	102	804.5	58	0
4.4	3.7	2	121	804.6	57	0
5.8	4.5	2.2	121	804.8	54	0
6.7	4.5	1.6	124	805.2	53	0
7.5	4.5	2.4	126	805.3	53	0
8.1	4.5	2.3	133	805.4	53	0
9.2	8.2	2.7	127	805.6	53	0
9.6	8.2	2.6	133	805.8	55	0
10.1	8.2	2.7	134	806.2	58	0
10.4	8.2	2.5	136	806.3	58	0
10.9	10.4	2.2	141	806.5	58	0
11.3	10.4	3.2	137	806.5	59	0
11.3	10.4	2.2	141	806.5	61	0
11.9	10.4	1.9	155	806.6	59	0
12.5	12.2	1.4	147	806.6	57	0
12.5	11.8	1	133	806.5	57	0
12.5	11.7	2.2	145	806.4	57	0
12.8	12.2	1.1	158	806.2	59	0
12.8	10.9	1.1	148	806.1	64	0
12.8	9.3	1.1	159	805.9	72	0
12.8	8.4	1.6	138	805.9	75	0
8.3	8	1.2	114	805.8	80	0
9	7.9	1	165	805.4	80	0

Figure 4.c - Data after Data Cleaning

```
mysql> select * from data_2018 limit 30;
```

station_name	date	time	gps	battery_v	Hourly Rain	Daily Rain	sm_m3_m3	Temperature
NAINITAL	2018-01-01	5:30:00	L	12.2	0	0	0	4.2
NAINITAL	2018-01-01	5:45:00	L	12.2	0	0	0	4.2
NAINITAL	2018-01-01	6:00:00	L	12.2	0	0	0	4.1
NAINITAL	2018-01-01	6:15:00	L	12.2	0	0	0	4.1
NAINITAL	2018-01-01	6:30:00	L	12.2	0	0	0	4.1
NAINITAL	2018-01-01	6:45:00	L	12.2	0	0	0	4.1
NAINITAL	2018-01-01	7:00:00	L	12.2	0	0	0	3.9
NAINITAL	2018-01-01	7:15:00	L	12.2	0	0	0	3.7
NAINITAL	2018-01-01	7:30:00	L	12.7	0	0	0	4.4
NAINITAL	2018-01-01	7:45:00	L	12.7	0	0	0	5.8
NAINITAL	2018-01-01	8:00:00	L	12.7	0	0	0	6.7
NAINITAL	2018-01-01	8:15:00	L	12.7	0	0	0	7.5
NAINITAL	2018-01-01	8:30:00	L	13.9	0	0	0	8.1
NAINITAL	2018-01-01	8:45:00	L	13.9	0	0	0	9.2
NAINITAL	2018-01-01	9:00:00	L	13.9	0	0	0	9.6
NAINITAL	2018-01-01	9:15:00	L	13.9	0	0	0	9.9
NAINITAL	2018-01-01	9:30:00	L	14.1	0	0	0	10.4
NAINITAL	2018-01-01	9:45:00	L	14.1	0	0	0	10.9
NAINITAL	2018-01-01	10:00:00	L	14.1	0	0	0	11.3
NAINITAL	2018-01-01	10:15:00	L	14.1	0	0	0	11.2
NAINITAL	2018-01-01	10:30:00	L	14.2	0	0	0	11
NAINITAL	2018-01-01	11:00:00	L	14.2	0	0	0	12.2
NAINITAL	2018-01-01	11:15:00	L	14.2	0	0	0	12.3
NAINITAL	2018-01-01	11:30:00	L	14.2	0	0	0	12.4
NAINITAL	2018-01-01	11:45:00	L	14.2	0	0	0	12.2
NAINITAL	2018-01-01	12:00:00	L	14.2	0	0	0	10.9
NAINITAL	2018-01-01	12:15:00	L	14.2	0	0	0	9.8
NAINITAL	2018-01-01	12:30:00	L	14.1	0	0	0	8.4
NAINITAL	2018-01-01	12:45:00	L	14.1	0	0	0	8
NAINITAL	2018-01-01	13:00:00	L	14.1	0	0	0	9

Figure 4.d - Data after Data Cleaning

5. Database Schema and Design

5.1 Overview

The project uses two MySQL databases—Naini_Data (for AWS 1) and Naini_Data2 (for AWS 2)—to store, aggregate, and analyze multi-year, high-frequency meteorological data. Both databases employ a year-wise table structure for raw data and provide daily and monthly aggregation views/tables for efficient temporal analysis. Rainfall classification views—using IMD standards—are also present in both databases.

5.2 Table Diagram

Naini_Data (AWS 1)

```
+-------------------+
|   data_YYYY       | (2017-2025)
+-------------------+
| station_name      |
| date              |
| time              |
| gps               |
| battery_v         |
| Hourly Rain       |
| Daily Rain        |
| sm_m3_m3          |
| Temperature       |
| Maximum Temperature |
| Minimum Temperature |
| Wind Speed        |
| Wind Degree       |
| Pressure          |
| Humidity          |
| Sun Duration      |
+-------------------+
```

5.3 Table Structures

Naini_Data (AWS 1) – Raw Data Fields (Example: data_2022):

- station_name: Name of the weather station (text, up to 100 characters)
- date: Date of observation (date format)
- time: Time of observation (text, up to 10 characters)
- gps: GPS coordinates (text, up to 50 characters)
- battery_v: Battery voltage (floating point number)

- Hourly Rain: Rainfall measured over an hour (floating point number)
- Daily Rain: Total rainfall for the day (floating point number)
- sm_m3_m3: Soil moisture content (floating point number)
- Temperature: Measured temperature (floating point number)
- Maximum Temperature: Highest temperature for the day (floating point number)
- Minimum Temperature: Lowest temperature for the day (floating point number)
- Wind Speed: Wind speed (floating point number)
- Wind Degree: Wind direction in degrees (floating point number)
- Pressure: Atmospheric pressure (floating point number)
- Humidity: Relative humidity (floating point number)
- Sun Duration: Duration of sunshine (floating point number)

Naini_Data2 (AWS 2) – Raw Data Fields (Example: year_2023):

- station_name: Name of the weather station (text, up to 50 characters)
- date: Date of observation (date format)
- time: Time of observation (text, up to 10 characters)
- temp_deg_c: Measured air temperature in degrees Celsius (floating point number)
- temp_max_deg_c: Maximum air temperature for the day in degrees Celsius (floating point number)
- temp_min_deg_c: Minimum air temperature for the day in degrees Celsius (floating point number)
- humidity_percent: Relative humidity as a percentage (floating point number)
- pressure_hpa: Atmospheric pressure in hectopascals (floating point number)
- ws_ms: Wind speed in meters per second (floating point number)
- wd_deg: Wind direction in degrees (floating point number)
- rain_daily_mm: Total daily rainfall in millimeters (floating point number)
- solar_radiation: Solar radiation (floating point number)
- uv_a: UV-A radiation (floating point number)
- uv_b: UV-B radiation (floating point number)
- battery_voltage: Battery voltage (floating point number)
- peripherals: Status or readings from connected peripherals (floating point number)

5.4 Aggregation and Classification Views (Columns Names)

Naini_Data (AWS 1)

- daily_YYYY_view:
  station_name, date, max_temp, min_temp, avg_temp, total_hourly_rain, total_daily_rain, avg_battery_v, avg_sm_m3_m3, avg_wind_speed, avg_wind_degree, avg_pressure, avg_humidity, avg_sun_duration

- monthly_YYYY:
  station_name, year, month_number, month_name, max_temp, min_temp, avg_temp, total_monthly_rain_hourly, total_monthly_rain_daily, avg_battery_v, avg_sm_m3_m3, avg_wind_speed, avg_wind_degree, avg_pressure, avg_humidity, avg_sun_duration
- rainfall_classification_view:
  station_name, date, rain_daily_mm, rainfall_category, year

Naini_Data2 (AWS 2)

- daily_data2_YYYY:
  station_name, date, avg_temp_deg_c, max_temp_deg_c, min_temp_deg_c, avg_humidity_percent, avg_pressure_hpa, avg_wind_speed_ms, avg_wind_direction_deg, avg_rain_daily_mm, avg_solar_radiation, avg_uv_a, avg_uv_b, avg_battery_voltage, avg_peripherals
- monthly_data2_YYYY:
  station_name, year, month, avg_temp_deg_c, max_temp_deg_c, min_temp_deg_c, avg_humidity_percent, avg_pressure_hpa, avg_wind_speed_ms, avg_wind_direction_deg, total_rain_mm, avg_solar_radiation, avg_uv_a, avg_uv_b, avg_battery_voltage, avg_peripherals
- rainfall_classification:
  station_name, date, total_rain_mm, rainfall_category, year

Note : Rainfall_classification- as per definitions mentioned in section 4.5

5.5 Design Choices & Highlights

- Year-wise Table Structure: Each year's data is stored in a separate table for both raw and aggregated data, facilitating both granular and summarized analysis.
- Aggregation Views & Tables: Daily and monthly views/tables enable efficient, flexible querying for temporal variability.
- Rainfall Classification: IMD guidelines are implemented as SQL views for immediate categorization of rainfall events.
- Normalization: Column names and units are standardized within each AWS database, and mapping is handled in the web app for cross-station comparisons.
- Scalability: The design supports large volumes of high-frequency data (potentially hundreds of thousands of rows per year per station).

This schema ensures robust storage, efficient querying, and a foundation for advanced analytics and visualization in the web app.

6. Web Application Structure

6.1 Technology Stack Flowchart

```
MySQL (Local)
  ▶ Used for: data storage, cleaning, querying
  │
Python in VS Code
  ├──▶ Wrote scripts for data processing
  ├──▶ Built app using:
  │     ├── Streamlit (web UI)
  │     ├── mysql-connector-python + SQLAlchemy (DB integration)
  │     ├── Pandas (data handling)
  │     ├── Plotly (visualizations)
  │     └── Scikit-learn (ML models)
  │
Local Testing
  └──▶ Ran Streamlit app locally using `python app.py`: Verified data
flow, UI, and ML output
  │
Railway (Cloud MySQL Hosting)
  └──▶ Migrated DB to cloud for global access
  │
Render (App Deployment)
  └──▶ Final app deployed and hosted publicly
```

6.2 Application Structure & Navigation

- Authentication: Simple login using hardcoded credentials in Streamlit session state.
- Sidebar Navigation:
  - Menu for Login, Get Credentials, Dashboard selection (AWS1, AWS2, Prediction Model)
  - Sidebar controls for each dashboard (filtering, feature toggles, report generation)
- Main Content Area: Renders the selected dashboard/module in Streamlit's main panel.

6.3 Main Modules/Pages

1. Login and Credentials (6.a)

- Login with predefined credentials, stored in-memory.
- "Get Credentials" page provides username/password for the user.

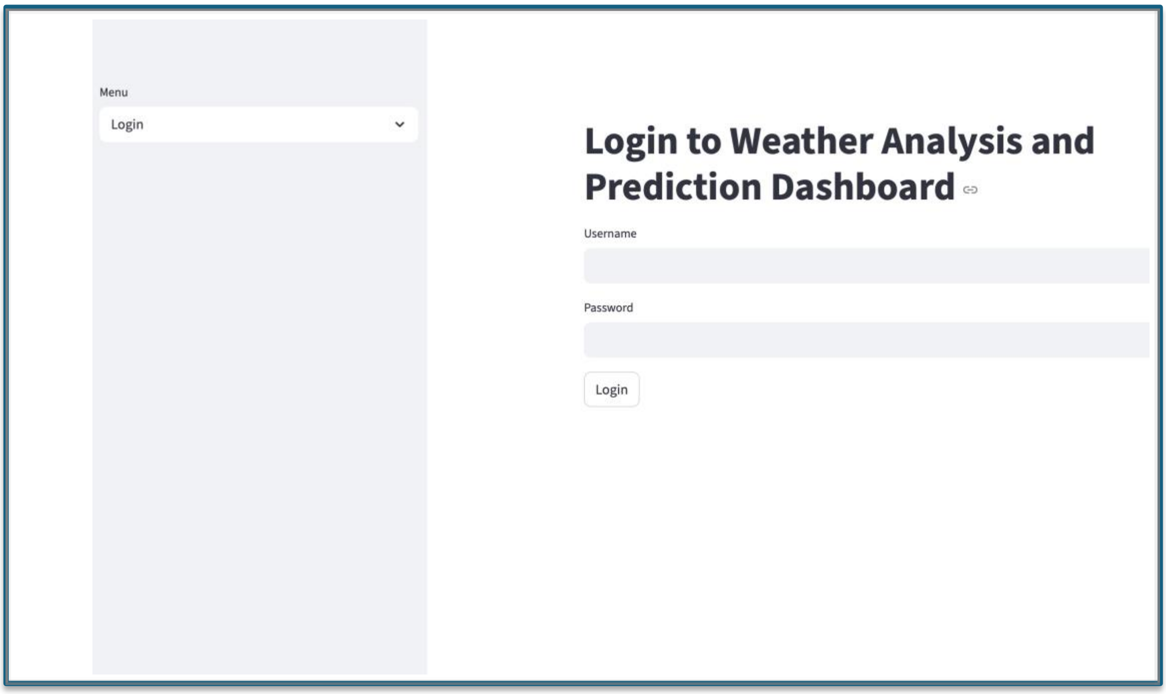


Figure 6.a The Login page for the Web App where the user enters their credentials

2. Weather Dashboards (AWS1 & AWS2) (6.b, 6.c)

- Tabs:
  - Temperature Data Viewer: Daily/yearly temperature tables and charts; option to include extremes.
  - Rainfall Data Viewer: Custom date range, yearly/monthly rainfall, extremes (heavy/very heavy events), downloadable tables/charts.
  - Trend Comparison: Compare temp/rainfall trends between two selected year groups.
  - Custom Data Explorer: Select any DB table, columns, filters, run queries, download as CSV.
- Key Features:
  - Dynamic data access based on user input (year, station, parameter, filter).

- Caching for efficient repeated queries.
- Download buttons for data (CSV) and full tab reports (PDF).
- Monsoon dates reference table.
- Responsive, multi-tab UI with Plotly charts.

Figure 6.b Tabs for AWS2

Figure 6.c Tabs for AWS2

3. Prediction Model Dashboard

- Model Trainer for Weather Data:
    - Train on any combination of years (2017–2025) and predict for any year (including future years).
    - Select target variable (temp, rain, wind, etc.) and features.
    - Choose model: Random Forest, Gradient Boosting, Linear Regression, SVR, LightGBM.
    - UI for hyperparameter tuning.
    - Train/test split is year-based; predictions evaluated on the selected target year.

- Shows metrics (R^2, MAE, MSE), time series and scatter plots (actual vs predicted), and explanations.
- Export results and charts as PDF or Word report.
- Download actual vs predicted data as CSV.

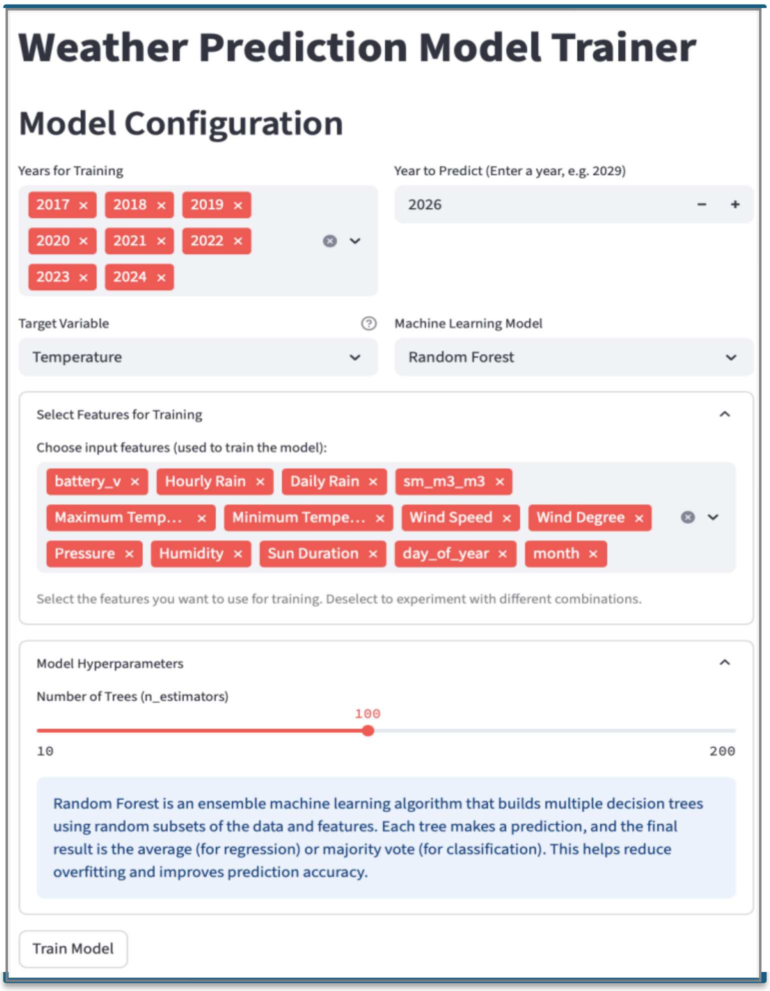


Figure 6.d Layout for Prediction Model

6.4 Backend Logic & Data Access

- Database Connections: Each module now connects to its respective MySQL database hosted on Railway.com via SQLAlchemy; credentials are configured for the Railway environment.
- Caching: Streamlit's @st.cache_data and @st.cache_resource decorators optimize performance for repeated queries and loaded resources.

7. SQL Queries and App Functionalities

This section is a comprehensive, user-friendly guide to all features of the AWS 1 and AWS 2 dashboards. It explains how to use each dashboard tab, what you can do, and how to get the most out of the app's interactive data exploration and export features.

7.1 Getting Started

Login

- On launching the application, you are greeted by a login screen.
- Enter the provided username and password (see the "Get Credentials" tab if you need them).
- Upon successful login, you will access the dashboard selection menu.

7.2 Dashboard Overview

After logging in, you can choose between:

- AWS1 (Manora Peak) Dashboard
- AWS2 (ST Radar) Dashboard
- Prediction Model

AWS1 and AWS2 dashboards are structurally similar, but each draws data from a different Automatic Weather Station (AWS).

7.3 Common Dashboard Layout

Each AWS dashboard is organized into four main tabs, accessible along the top:

1. Temperature Data Viewer
2. Rainfall Data Viewer
3. Temperature & Rainfall Trend Comparison
4. Custom Data Explorer

Sidebar controls are available for each tab, providing options to filter data, select years, and generate reports.

7.4 Detailed Tab Guide

1. Temperature Data Viewer

Purpose: Explore daily and yearly temperature data for the selected AWS.

A. Temperature Extremes Across Years:

- View a summary table and line chart of highest and lowest yearly temperatures.(7.a)
- Toggle inclusion of this summary via the sidebar.

SQL Query Example – Yearly Extremes:

```
-- AWS1
SELECT
  YEAR(date) AS year,
  MAX(max_temp) AS highest_max_temperature,
  MIN(min_temp) AS lowest_min_temperature
FROM (
  SELECT date, max_temp, min_temp, station_name FROM daily_{year}_view WHERE max_temp <=
35 AND min_temp >= -3
  -- repeated for each year with UNION ALL
) AS combined
WHERE station_name = %s
GROUP BY year
ORDER BY year;
```

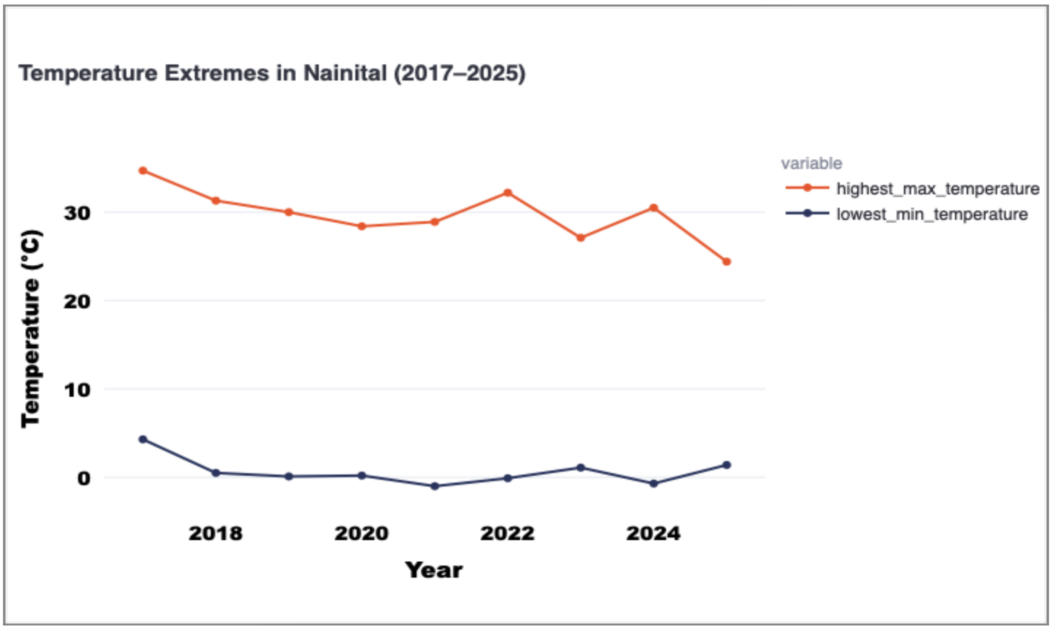


Figure 7.a Query 1 of the Temperature Data Viewer Tab

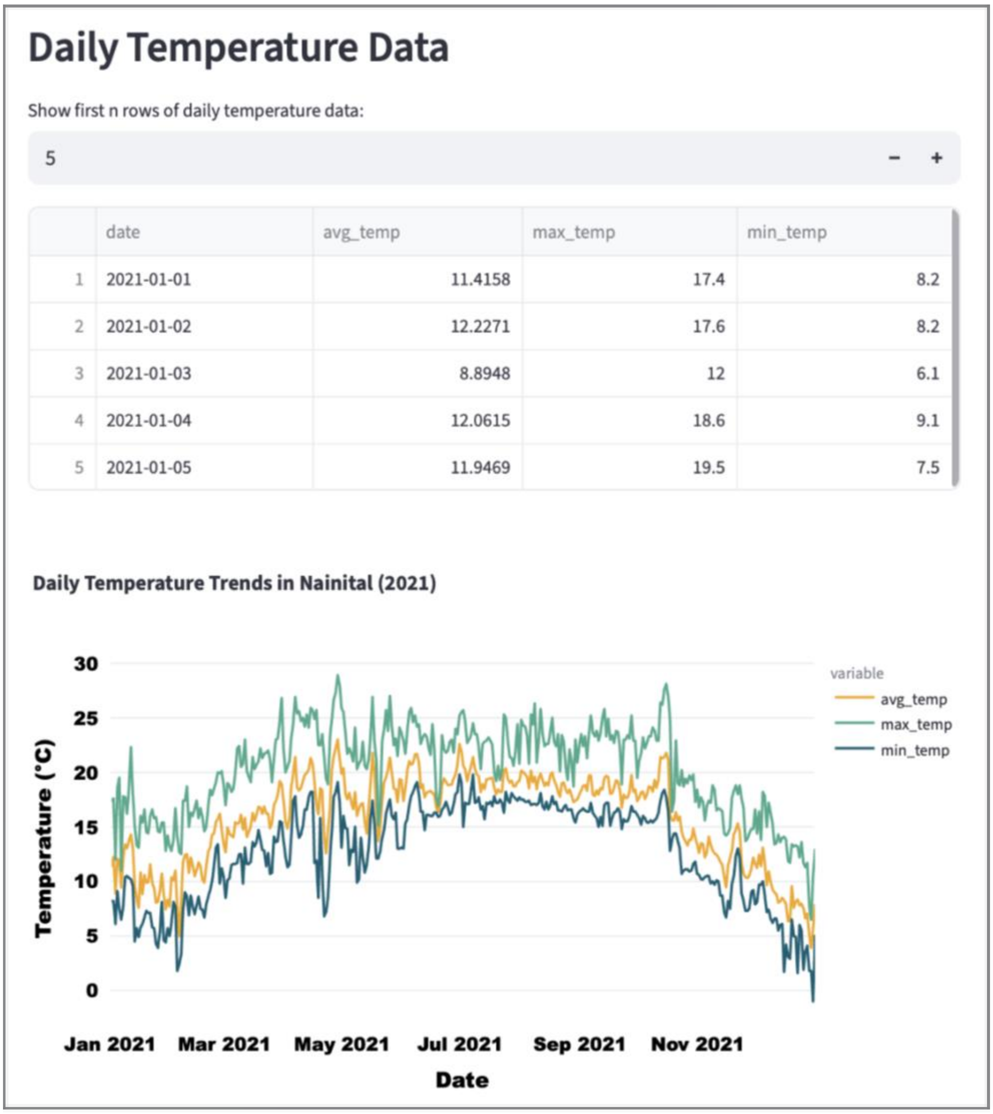
B. Daily Temperature Data:

- Select a year to examine day-by-day temperature values (average, minimum, maximum).

- Limit the number of rows displayed for focus or performance. (7.b)
- Download the entire dataset as a CSV file.

SQL Query Example – Daily Data:

```
-- AWS1
SELECT
  date,
  avg_temp,
  max_temp,
  min_temp
FROM daily_{year}_view
WHERE station_name = %s AND max_temp <= 35 AND min_temp >= -3 ORDER BY date;
```



	date	avg_temp	max_temp	min_temp
1	2021-01-01	11.4158	17.4	8.2
2	2021-01-02	12.2271	17.6	8.2
3	2021-01-03	8.8948	12	6.1
4	2021-01-04	12.0615	18.6	9.1
5	2021-01-05	11.9469	19.5	7.5

Figure 7.b Query 2 of the Temperature Data Viewer Tab

- Export Options:
  - Generate a PDF report of the temperature data and charts directly from the sidebar.

- Sidebar Controls:
    - Include/exclude extremes and daily data.
    - Generate/download tab-specific PDF reports.

2. Rainfall Data Viewer

Purpose: Analyze rainfall patterns, totals, and extreme events.

A. Rainfall Insights for Custom Date Range:

- Pick a year, then select any date range within that year.
- Instantly see total rainfall, a daily rainfall chart, and a summary table for your chosen period. (7.c)
- Reference table shows official monsoon dates, with source links for each year (7.d).

SQL Query Example – Date Range Rainfall:

```
SELECT date, total_hourly_rain
FROM daily_{year}_view
WHERE station_name = %s
AND date BETWEEN %s AND %s
ORDER BY date;
```

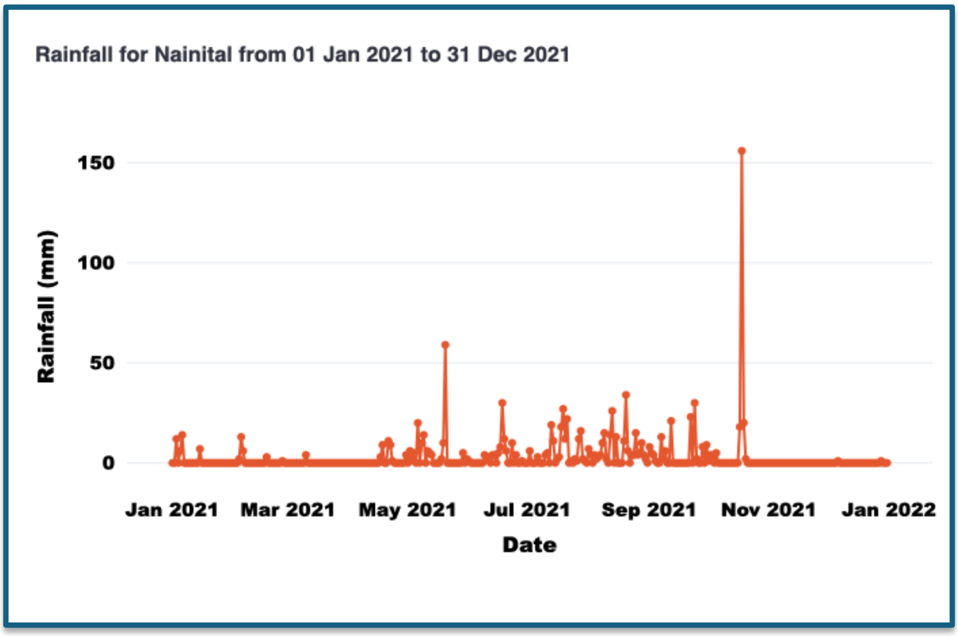


Figure 7.c Query 1 of the Rainfall Data Viewer Tab

Rainfall Insights for Custom Date Range

Reference: Monsoon arrival and withdrawal dates for Nainital (2017–2025):

Year	Dates	Source
2017		-
2018	1 Jun- 1 Oct	Source
2019	24 Jun–10 Oct	Source
2020	23 Jun–2 Oct	Source
2021	11 Jul–8 Oct	Source
2022		-
2023	24 Jun–30 Sep	Source
2024	29 Jun–4 Oct	Source
2025	20 Jun–(withdrawal not available)	Source

Figure 7.d Reference Table in the Rainfall Data Viewer Tab

B. Rainfall Distribution for a Selected Year:

Select any year to view monthly rainfall totals, displayed in a bar chart and table(7.e).

SQL Query Example – Monthly Distribution:

```
SELECT month_number, month_name, total_monthly_rain_hourly
FROM monthly_{year}
WHERE station_name = :station;
```

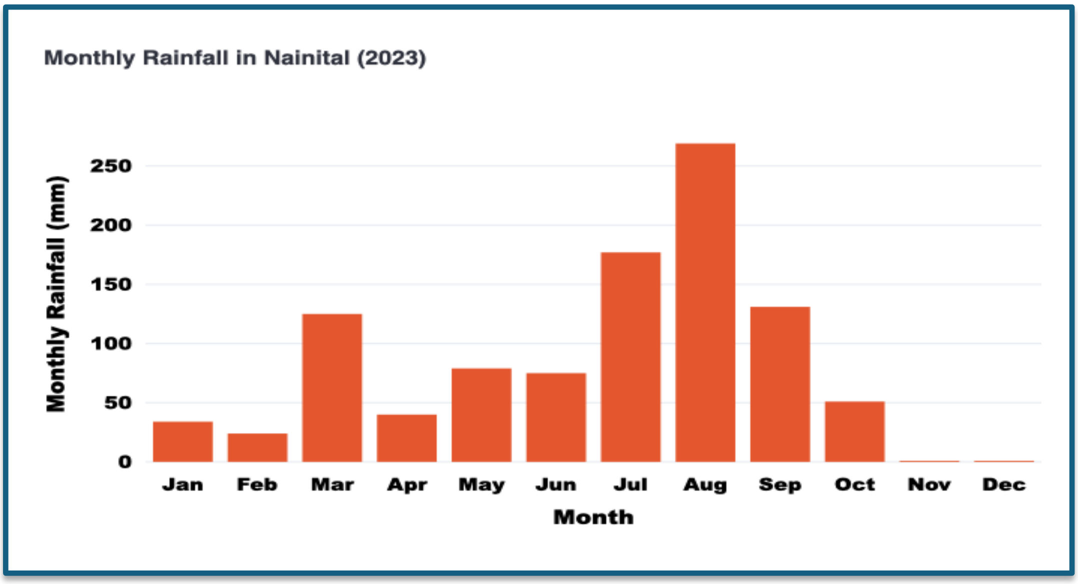


Figure 7.e Query 2 of the Rainfall Data Viewer Tab

C. Rainfall Comparison Across Years:

- Choose multiple years to compare monthly rainfall patterns across them. (7.f)
- Results appear in a multi-line chart and comparison table.

SQL Query Example – Multi-Year Comparison:

```
SELECT year, month_number, month_name, SUM(total_monthly_rain_hourly) AS monthly_rainfall
FROM (
  SELECT {year} as year, month_number, month_name, station_name, total_monthly_rain_hourly
FROM monthly_{year}
  -- repeated for each year with UNION ALL
) AS all_months
WHERE station_name = :station
GROUP BY year, month_number, month_name
ORDER BY year, month_number;
```

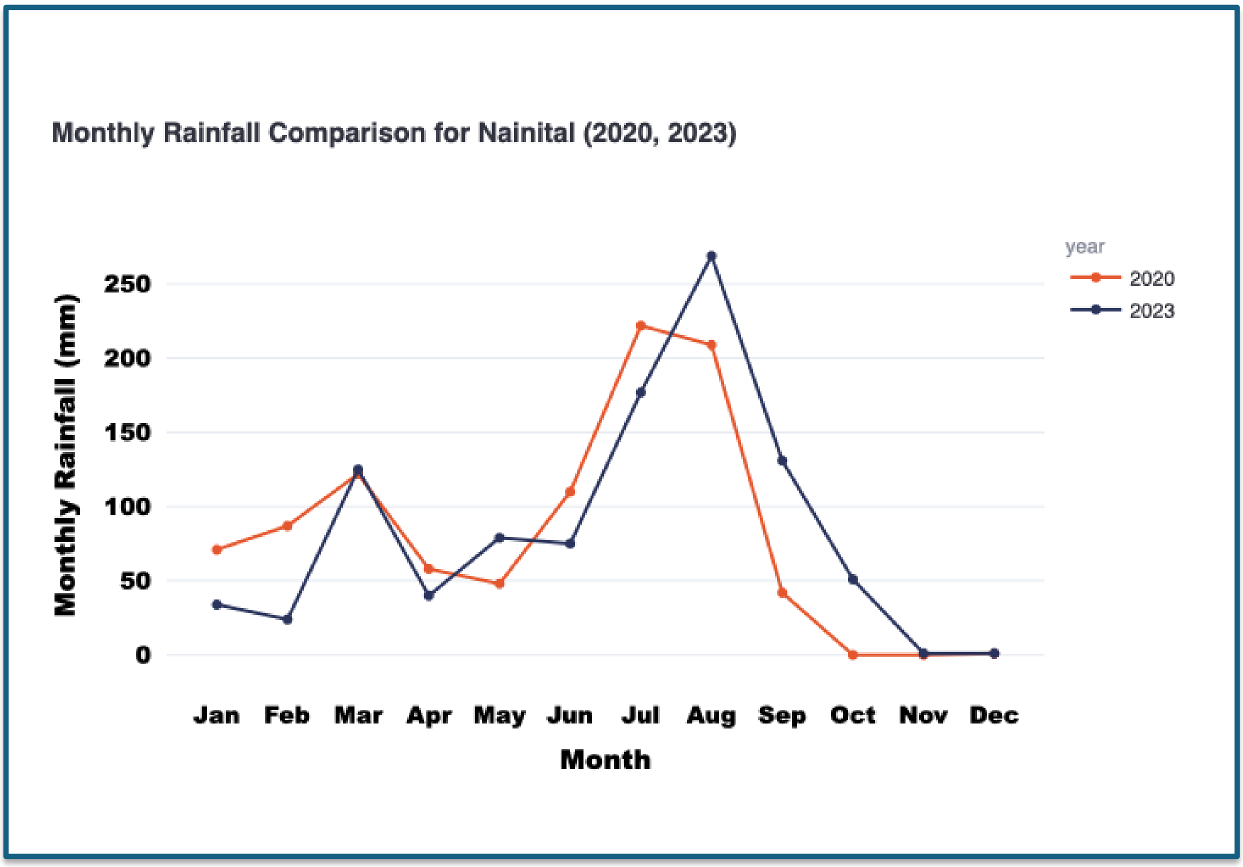


Image 7.f Query 3 of the Rainfall Data Viewer Tab

D. Extreme Rainfall Events:

- Filter and list days with "Rather Heavy," "Heavy," "Very Heavy," or "Extremely Heavy" rainfall. (7.g). A reference chart is provided for IMD rainfall category definitions.

SQL Query Example – Extreme Events:

```
SELECT DISTINCT date, station_name, rain_daily_mm, rainfall_category
FROM rainfall_classification_view
WHERE rainfall_category IN (%s, ...)  AND year IN (%s, ...)  AND station_name = %s
ORDER BY date;
```

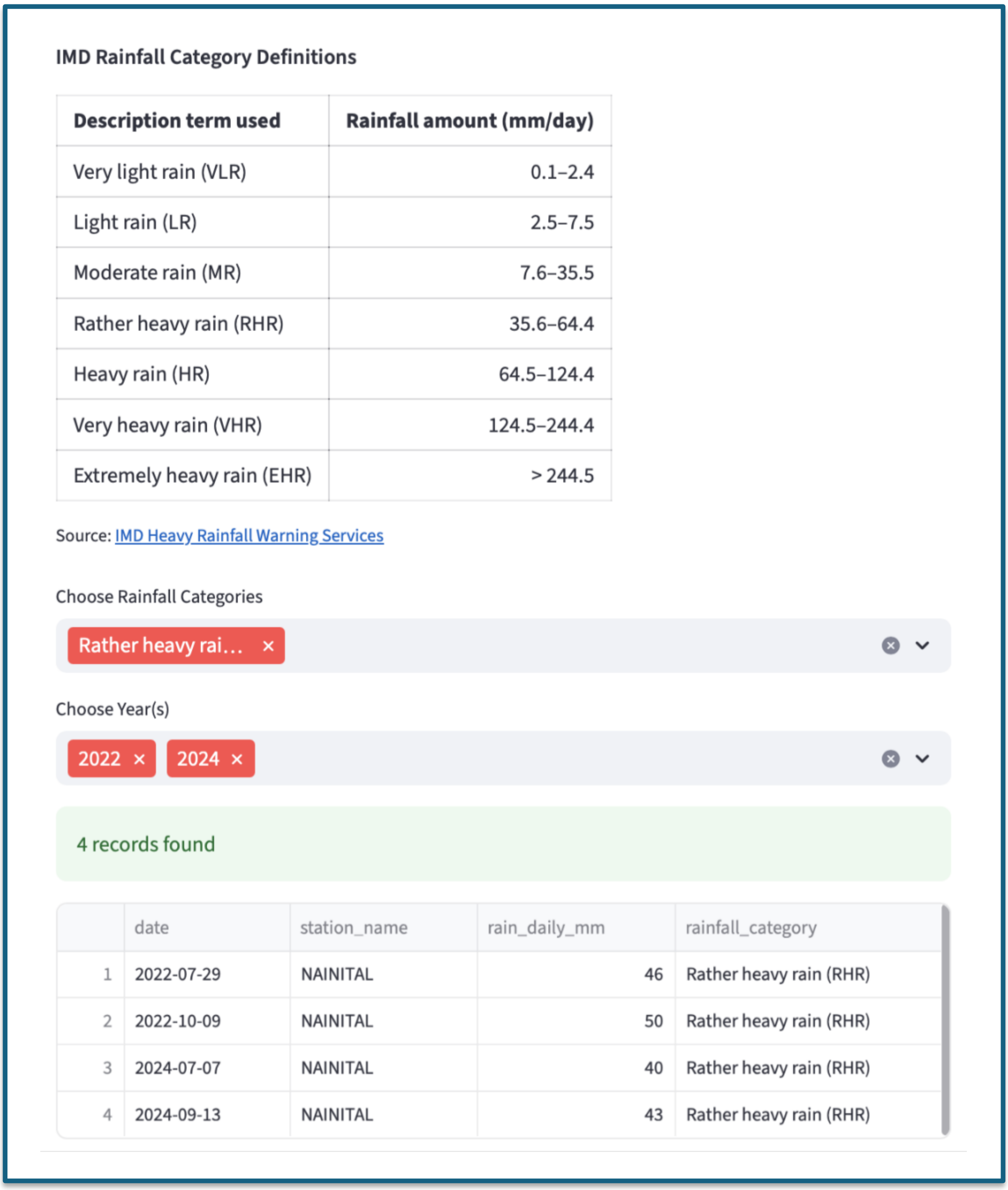


Figure 7.g Query 4 of the Rainfall Data Viewer Tab

3. Temperature & Rainfall Trend Comparison

Purpose:
Visually compare temperature and rainfall trends between two groups of years.

- Year Group Selection:
  - Select two separate groups of years for comparison.
  - For each group, choose two or more years.
  ‘
- Temperature Trends:
  - View and compare average temperature trends (as a line chart and data table) for both year groups. (7.h)

SQL Query Example – Average Temperature by Year Group:

```
SELECT date, AVG(avg_temp) as avg_temp
FROM (
  SELECT date, avg_temp FROM daily_{year}_view WHERE station_name = %s AND avg_temp <=
35 AND avg_temp >= -3
  -- repeated for each selected year with UNION ALL
) AS combined
GROUP BY date ORDER BY date;
```

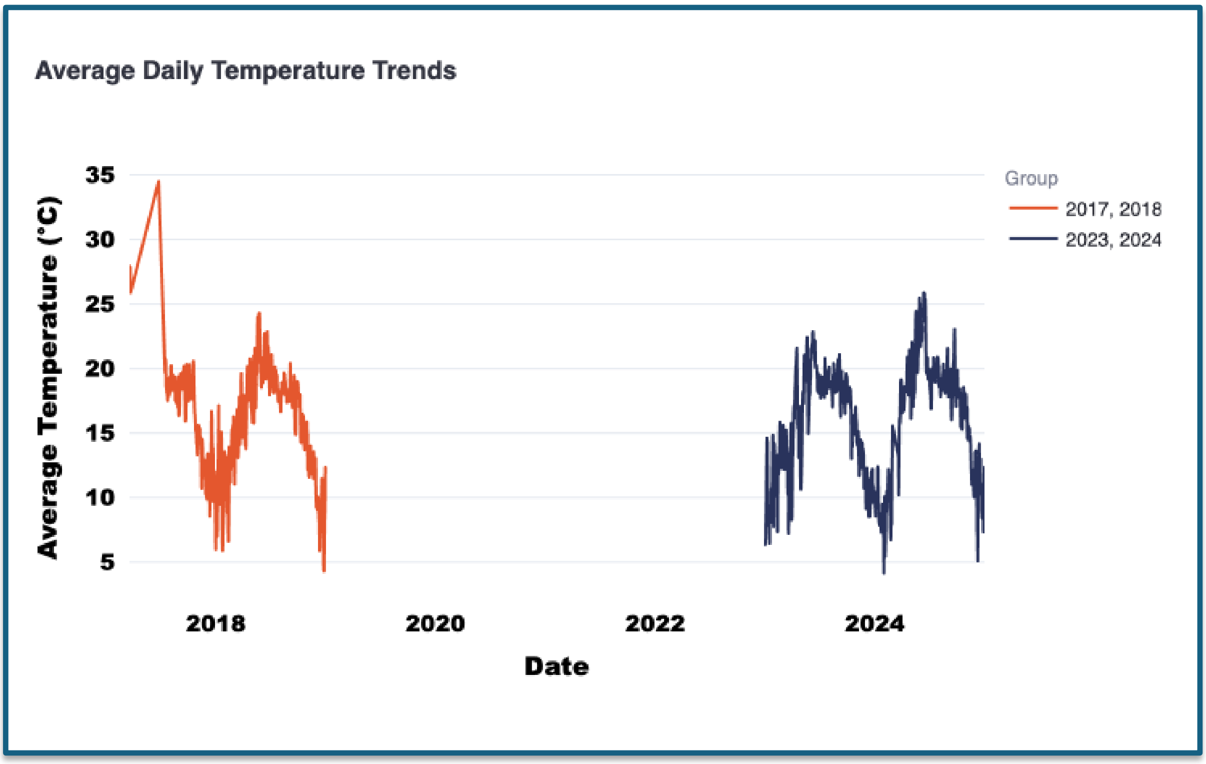


Figure 7.h Query 1 for the Temperature and Rainfall Comparison Tab

- Rainfall Trends:
    - View and compare monthly rainfall trends across the same year groups.(7.i)

SQL Query Example – Monthly Rainfall by Year Group:

```
SELECT month_number, month_name, SUM(total_monthly_rain_hourly) as rainfall
FROM (
  SELECT month_number, month_name, total_monthly_rain_hourly FROM monthly_{year} WHERE
station_name = %s
  -- repeated for each selected year with UNION ALL
) AS combined
GROUP BY month_number, month_name ORDER BY month_number;
```

- Export Options:
    - Download comparison tables as CSV.
    - Generate a PDF report for this tab's analyses.

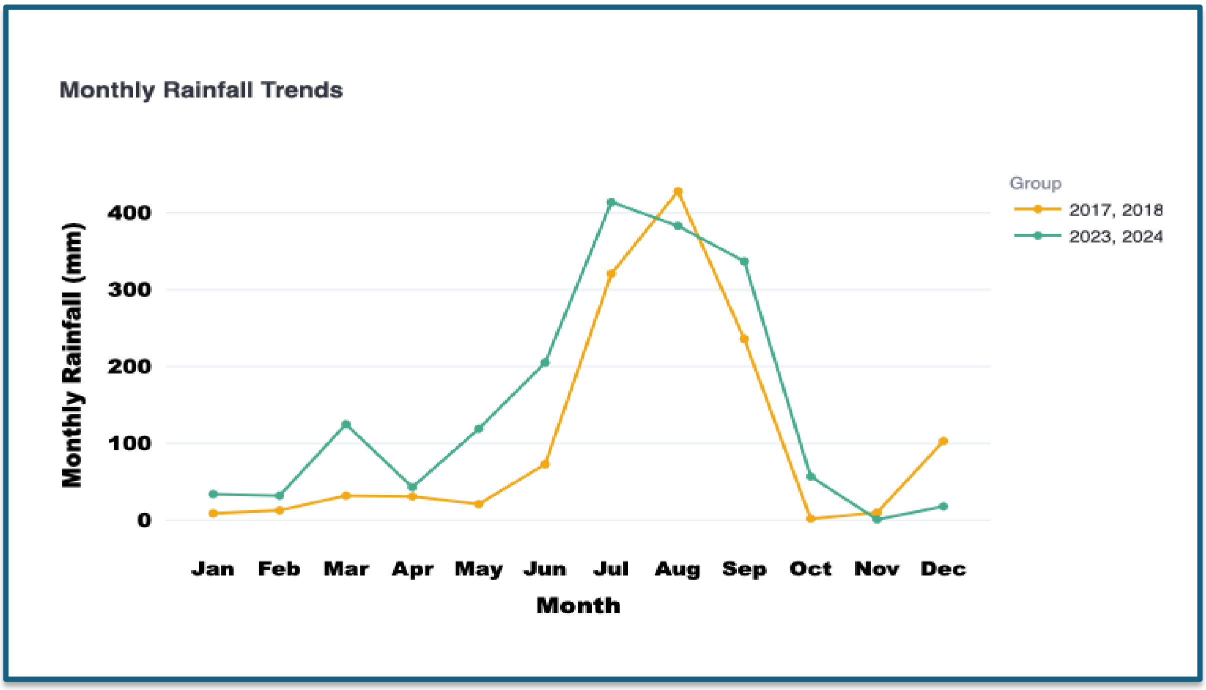


Figure 7.i Query 2 for the Temperature and Rainfall Comparison Tab

Sidebar Controls:

- Toggle inclusion of temperature and rainfall trend comparisons.
- Generate/download tab-specific PDF reports.

4. Custom Data ExplorerPurpose: Empower advanced users to access any table, column, or filtered dataset from the AWS database—no coding required.

- Table & Column Selection: Choose any available database table and select one or more columns to view.
- Filtering: Use quick preset filters (such as dates or rainfall/temperature thresholds) or write your own custom SQL WHERE clause for precise data extraction.
- Row Limiting: Set a maximum number of rows to display for performance and usability.
- Data Output & Download: Results are shown in a data table. Instantly downloadable results as a CSV file for your own analysis. (7.j)

SQL Query Example – Custom Explorer:

```
SELECT col1, col2, ...
FROM {selected_table}
WHERE {custom_where_clause} -- optional, user-defined or preset LIMIT {user_limit};
```

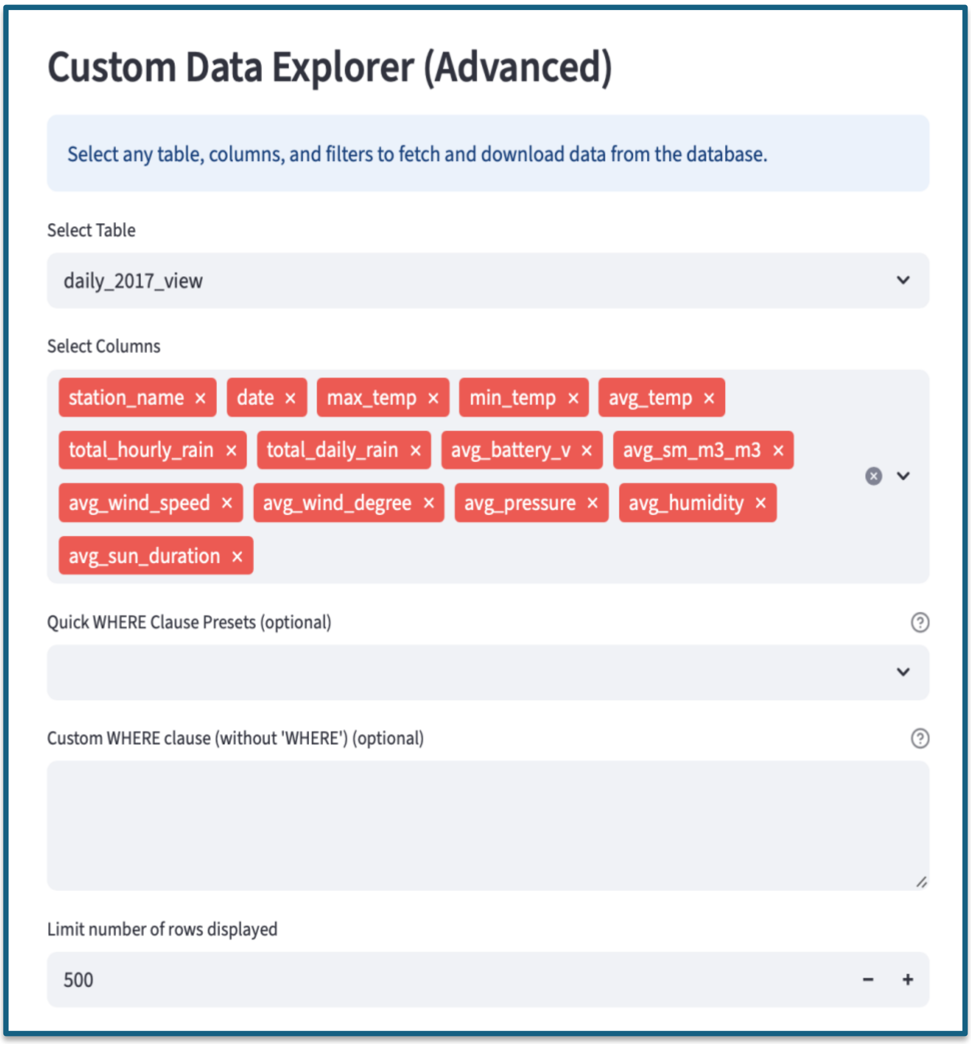


Figure 7.j Query 2 for the Temperature and Rainfall Comparison Tab

7.5 Sidebar Features & Exports (7.k)

- Consistent Sidebar Controls:
  - Each tab has sidebar options for "include/exclude" features, data selection, and export.
  - Every tab offers PDF report generation; some tabs allow CSV download of displayed data.
- Monsoon Reference Table:
  - For each AWS, a reference table displays monsoon arrival and withdrawal dates with sources.

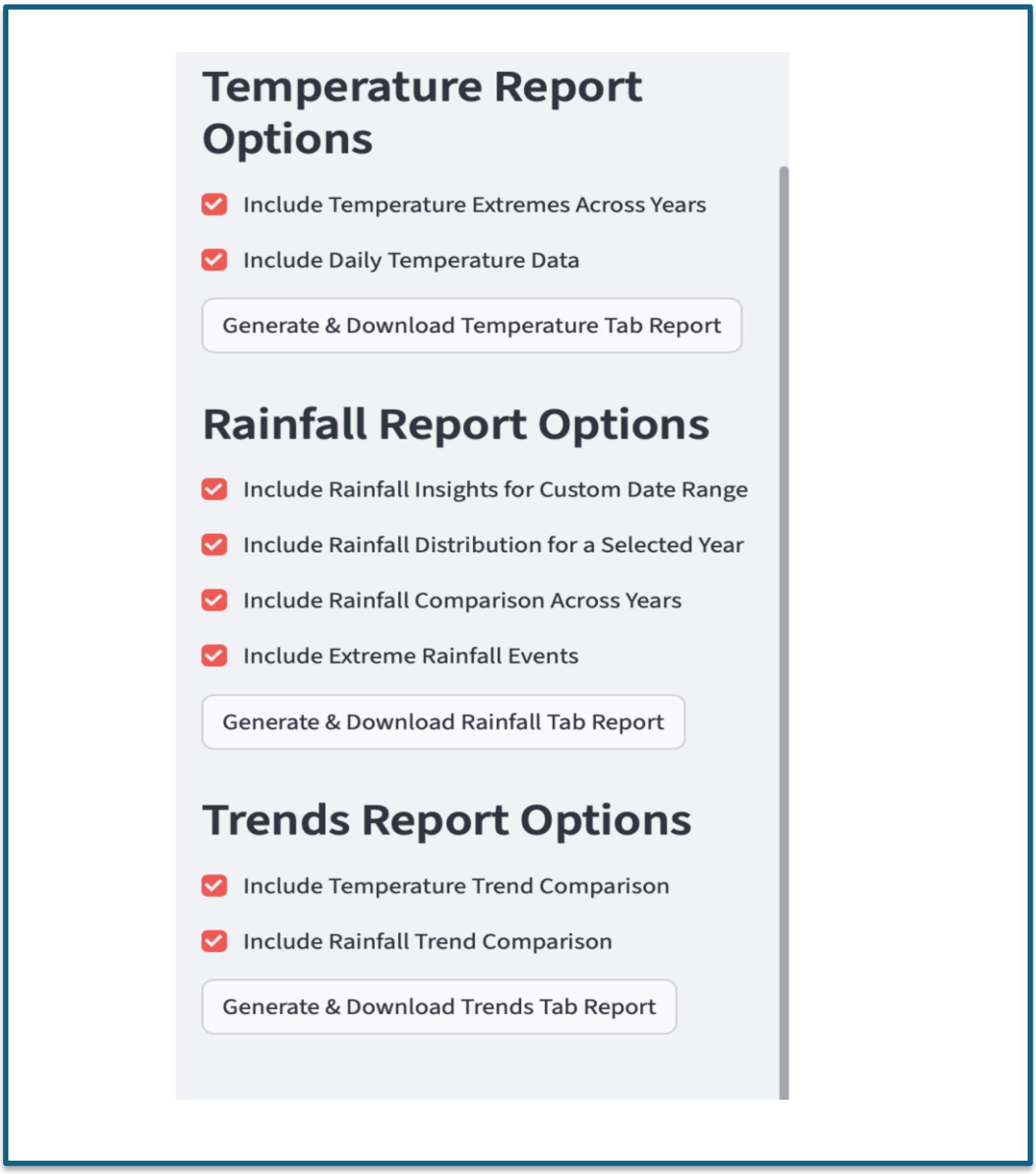


Figure 7.k Sidebar Features for Export

7.6 How AWS1 and AWS2 Dashboards Differ

- Data Source:
    - AWS1 uses data from the Manora Peak weather station.
    - AWS2 uses data from the ST Radar Station.
- Database Details:
    - Both dashboards are powered by MySQL databases, but their underlying tables and column names may differ slightly.
- Interface & Features:
    - All core features, tabs, and export options are consistent between AWS1 and AWS2, ensuring a uniform user experience.

7.7 Notes on Machine Learning Model

The application also features a Prediction Model tab, where you can train and evaluate machine learning models for weather forecasting.
All details, instructions, and user guidance for the ML module are provided in the next section.

The intuitive interface, rich analytics, and customized export options are designed for both casual users and advanced data explorers.

8. Machine Learning Model

8.1 Overview

The Weather Prediction Model Trainer lets you:

- Select any years of weather data for model training.
- Predict any variable (temperature, rainfall, wind, etc.) for any year (even future years).
- Choose the ML model, its features, and fine-tune its settings.
- Instantly see prediction accuracy, download data, and export beautiful PDF/Word reports.

8.2 Step-by-Step Guide

1. Model Configuration

a. Choose Training Years

- Select one or more years from the provided list. These years ʼdata will be used to "teach" the model.

b. Set Year to Predict

- Enter the year for which you want to forecast weather variables.
- You can choose a year within or beyond the range of training data.
- Warning: If you pick a year much earlier or later than your training data, the app will warn you that predictions may be unreliable (because the model is extrapolating).

c. Select Target Variable

- Pick what you want to predict (e.g., "Daily Rain", "Temperature", "Wind Speed", etc.).
- Only variables present in your data will be available.

d. Choose Model Type

- Options include:

- Random Forest: Powerful and good for non-linear data.
    - Linear Regression: Simple, fast, best for linear relationships.
    - Gradient Boosting: High accuracy, can capture complex patterns.
    - Support Vector Regression: Robust, good for small/medium datasets.
    - LightGBM: Fast, memory-efficient, handles big datasets.
- When you pick a model, a brief explanation appears.

e. Select Features

- Choose which features (columns) from your dataset the model should learn from.
- By default, all available features are selected. You can deselect any to experiment.

f. Adjust Model Hyperparameters

- For some models (e.g., Random Forest, Gradient Boosting, LightGBM, SVR), you can tune settings like the number of trees, learning rate, or regularization.
- The sidebar shows a clear explanation of each parameter and model.

2. Train the Model

- Click "Train Model" when satisfied with your configuration.
- The app will:
  - Load the selected years 'data from the database.
  - Train the chosen ML model on your selected features and target.
  - Automatically test the model by predicting the target variable for the year you selected.

SQL Note:
Data is loaded using queries like:

```
SELECT * FROM data_{year}
```

where {year} is each selected year for training and the year for prediction.

3. Results & Evaluation

After training, results appear immediately:

a. Model Used & Settings

- Shows your selected model, its hyperparameters, and the features used.

b. Model Explanation

- A short, readable description of how your chosen ML model works.

c. Charts

- Actual vs Predicted (Time Series):
    - Line chart comparing real vs predicted values over time.
    - Lets you visually judge how well the model tracks the actual data.

- Actual vs Predicted (Scatter):
    - Scatter plot comparing actual values (x-axis) to predicted values (y-axis).
    - The closer the points are to the diagonal, the better the predictions.

d. Scores & Metrics

- R^2 (Coefficient of Determination):
    - Measures how well the predictions fit actual data.
    - 1.0 means a perfect fit, 0 means the model is no better than guessing the average.
- MAE (Mean Absolute Error):
    - The average difference between predicted and actual values (lower is better).
- MSE (Mean Squared Error):
    - The average squared difference (sensitive to large errors; lower is better).
- Each metric is explained in plain English below the results.

e. Data Tables & Download

- Actual vs Predicted (Time Series):
    - Table listing dates, actual, and predicted values.
    - Downloadable as CSV.
- Actual vs Predicted (Scatter):
    - Table listing actual and predicted pairs.
    - Downloadable as CSV.

4. Exporting Results

- Use the “Export Results” section to generate a complete report.
- Choose between PDF and Word (.docx) formats.
- Reports include:
    - Your configuration (years, variables, features, model, parameters)
    - Model explanation
    - All metrics and scores
    - Charts (embedded as images)

- Click the download button to save your report.

5. Example Workflow (8.a, 8.b)

Suppose you want to predict Daily Rain for 2026 using data from 2017–2025:

1. Select training years: 2017–2025
2. Set year to predict: 2026
3. Target variable: Daily Rain
4. Model: Random Forest
5. Features: All default
6. Tune the number of trees (e.g., 50)
7. Train Model
8. Review results: (where data for a known year is available)
    - Check R^2, MAE, and MSE to judge accuracy.
    - Review charts to see prediction quality.
    - Download CSVs or export report as PDF/Word.

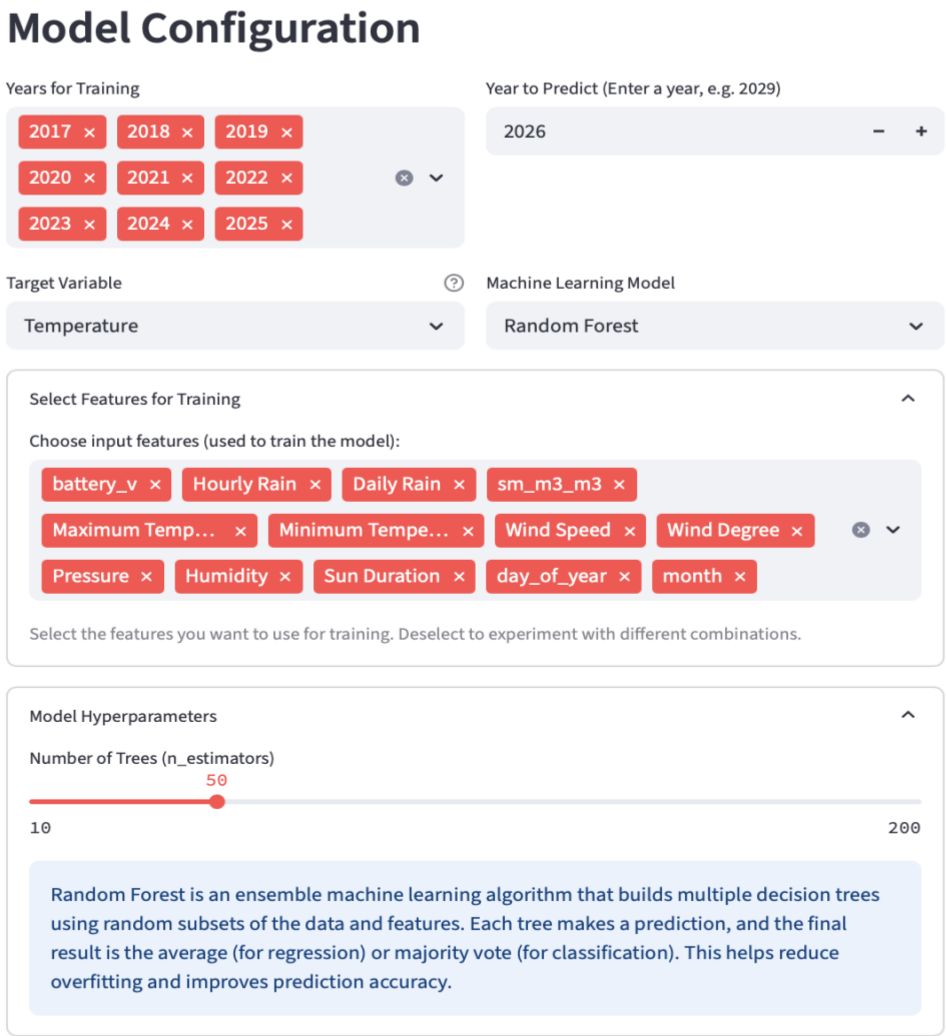
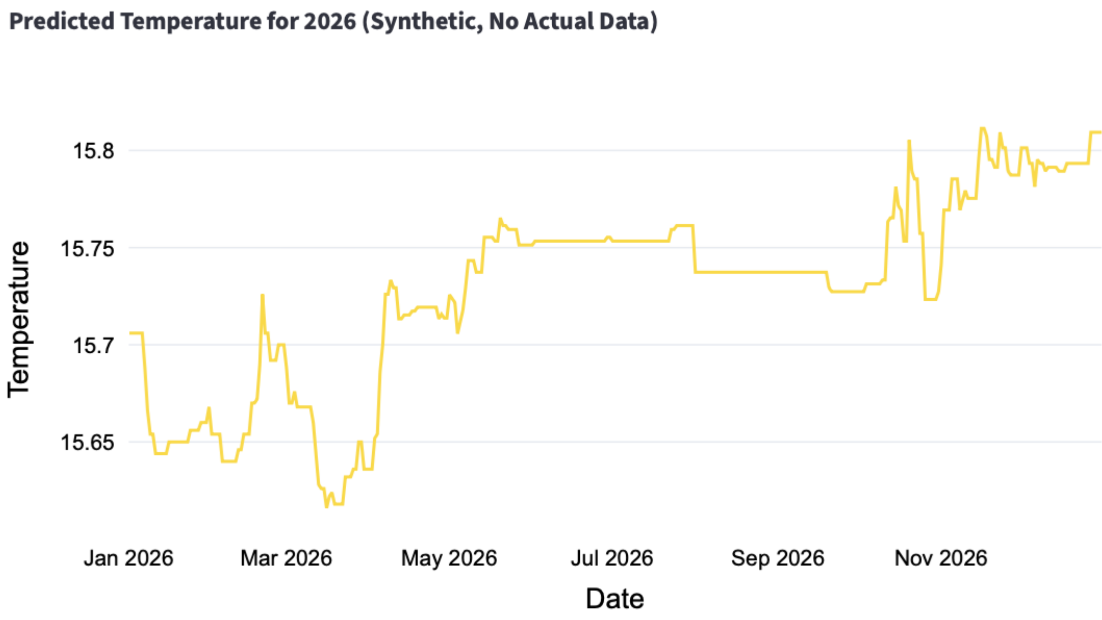


Image 8.a Example Workflow of the Prediction Model



Predicted Values (Time Series, Future Year)

	date	predicted
0	2026-01-01 00:00:00	15.706
1	2026-01-02 00:00:00	15.706
2	2026-01-03 00:00:00	15.706
3	2026-01-04 00:00:00	15.706
4	2026-01-05 00:00:00	15.706
5	2026-01-06 00:00:00	15.706
6	2026-01-07 00:00:00	15.688
7	2026-01-08 00:00:00	15.666
8	2026-01-09 00:00:00	15.654
9	2026-01-10 00:00:00	15.654

Figure 8.b Results obtained from the example workflow of the prediction Model

8.3 Rationale for Model Selection

The selection of machine learning models available in the Prediction Model module is intentional and rooted in both practical applications and the nature of weather data. Below is an explanation of why each model was included:

- Random Forest
  Chosen for its robustness and ability to capture complex, non-linear relationships in weather data. Weather patterns often involve interactions between variables that are not strictly linear. Random Forests are less prone to overfitting compared to individual decision trees and can handle missing values and noisy data, which are common in environmental datasets.
- Linear Regression
  Included as a baseline model due to its speed, simplicity, and interpretability. It is effective when relationships between predictors and the target variable are approximately linear. Linear Regression also serves as a reference point for comparing the performance of more complex models.
- Gradient Boosting
  Added for its high predictive power and ability to model subtle, non-linear effects. Gradient Boosting methods sequentially correct errors from previous trees, which often leads to superior accuracy in tabular datasets. This is valuable when predicting weather phenomena that may have intricate dependencies.
- Support Vector Regression (SVR)
  Selected for its robustness to outliers and flexibility in modeling both linear and non-linear relationships via kernel functions. SVR is particularly useful for smaller or medium-sized datasets and offers a different bias-variance tradeoff compared to tree-based models.
- LightGBM
  Included for its efficiency and scalability. LightGBM is a state-of-the-art gradient boosting framework that is optimized for speed and memory usage, making it suitable for large weather datasets and rapid experimentation. Its ability to handle categorical variables and missing data is an added benefit.

Summary:
These models were chosen to cover a spectrum of machine learning approaches—simple to complex, linear to non-linear, fast to highly accurate—offering users the flexibility to experiment and identify the best fit for their specific weather prediction tasks. Users are encouraged to try multiple models and compare their performance metrics (R^2, MAE, MSE) to make data-driven choices.

8.4 Tips & Best Practices

- Feature Selection:
  Experiment by removing or adding features to see how prediction quality changes.
- Model Choice:
  Try different models—sometimes a simpler one (like Linear Regression) works best.
- Extrapolation Warnings:
  Predicting far outside your training years may not be reliable; use caution.
- Exporting:
  Save your configuration and results for sharing or further analysis.

The Prediction Model tab provides a full ML workflow—configuration, training, evaluation, and reporting—entirely in your browser, with no coding required. All metrics and outputs are explained, so you can confidently interpret and use the forecasts produced.

9. Data Analysis & Findings

This section presents a comprehensive synthesis of the key findings from the analysis of weather data for Nainital (2017–2025), with a focus on temperature and rainfall trends, extreme weather events, and year-over-year comparative patterns. The insights are derived from both statistical summaries and a variety of comparative visualizations .

9.1 Average Daily Temperature Trends (2018–2024)

Consistent Seasonality:

The average daily temperature displays a well-defined seasonal pattern, peaking in summer (May–June) and dipping in winter (December–January) across all years. This underlines the persistence of a classic Himalayan mid-altitude climate regime.

Inter-annual Variability:

- 2018–2020: This period features more abrupt and frequent daily temperature fluctuations (orange line in visuals), suggesting higher short-term variability. This could be attributed to more variable weather or possible instrument variability during those years.
- 2022–2024: The temperature curves (blue line) become smoother and transitions between seasons are more gradual, hinting at either improved sensor calibration or genuinely more stable atmospheric conditions.

Temperature Range and Extremes:

Across all observed years, daily temperatures have remained broadly within 5°C to 25°C. No instances of extreme temperature anomalies were detected, indicating considerable climatic stability and absence of recent heatwaves or cold snaps. However, this apparent stability at the daily scale might mask subtler shifts or anomalies that become evident only through deeper monthly or yearly analyses. Looking at longer-term or aggregated data can reveal patterns—such as gradual warming/cooling, or unusual seasonal swings that are not visible in daily data.

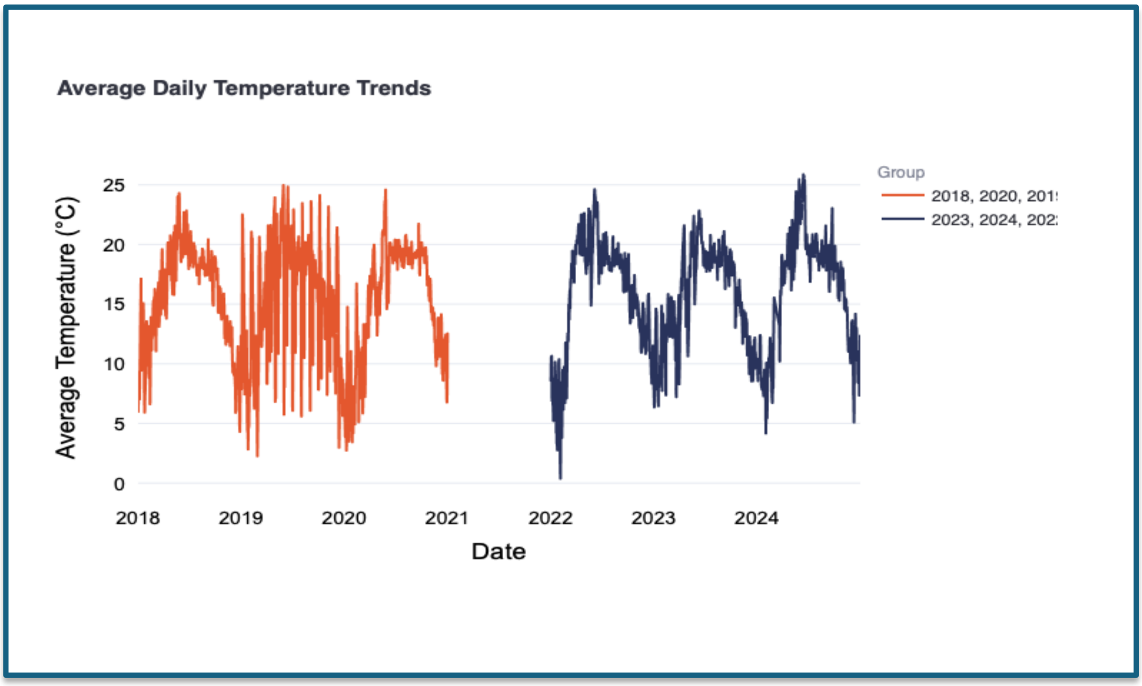


Figure 9.a Plot for Average Daily Temperature Trends

9.2 Monthly Rainfall Trends and Monsoonal Patterns:

- Classic Monsoonal Cycle: Rainfall strongly follows the established monsoon pattern, with the majority of precipitation recorded between June and September. Post-monsoon (October–December) and pre-monsoon (January–May) periods are generally dry.

- Comparative Trends:
    - 2023–2024: These years show a marked peak in July (over 650 mm), the highest in the dataset, while earlier years like 2018–2020 peaked at around 620 mm in August, indicating some shift in the timing and intensity of monsoon maxima.
    - Delayed Onset: In 2023–2024, rainfall in March–April was notably lower, suggesting a delayed monsoon onset compared to previous years.
    - Stable Withdrawal: Across all years, rainfall rapidly declines after September, confirming a consistent and sharp end to the monsoon season.
- Annual and Intra-annual Variability:
    - 2021 vs 2023: 2023 exhibits a smoother and more consistent rainfall curve, especially from March to October, reflecting stable monsoonal behavior. 2021, in contrast, features sharp monthly fluctuations, including unusual spikes in May (~150 mm) and October (~215 mm), which may be linked to isolated convective events or late-season storms.

- October Outlier: The exceptionally high rainfall in October 2021 is anomalous and may be indicative of a delayed monsoon retreat or post-monsoon disturbances

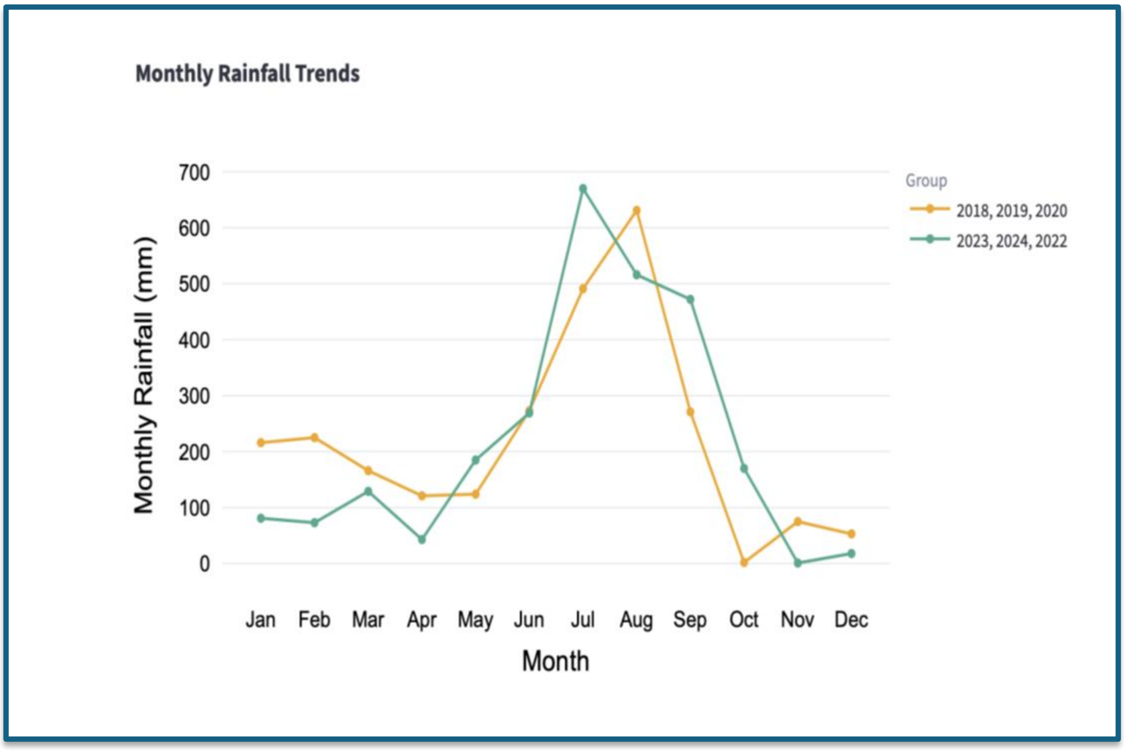


Figure 9.b Monthly Rainfall Trends and Monsoonal Pattern

9.3 Year-on-Year Rainfall Comparison: 2021 vs 2023

- Pre-monsoon Dryness:
  Both years had minimal rainfall in the early months, but April 2022 was particularly dry (0 mm) compared to 37 mm in 2021. May rainfall in 2023 (66 mm) was less than half of 2021 (150 mm), suggesting a drier pre-monsoon in 2022.
- Monsoon Peak:
  July 2023 recorded the maximum rainfall (256 mm) for both years, indicating a sharp, early monsoon peak. In contrast, 2021's rainfall was more evenly distributed across the monsoon months, with a notably high late peak in October (210 mm).

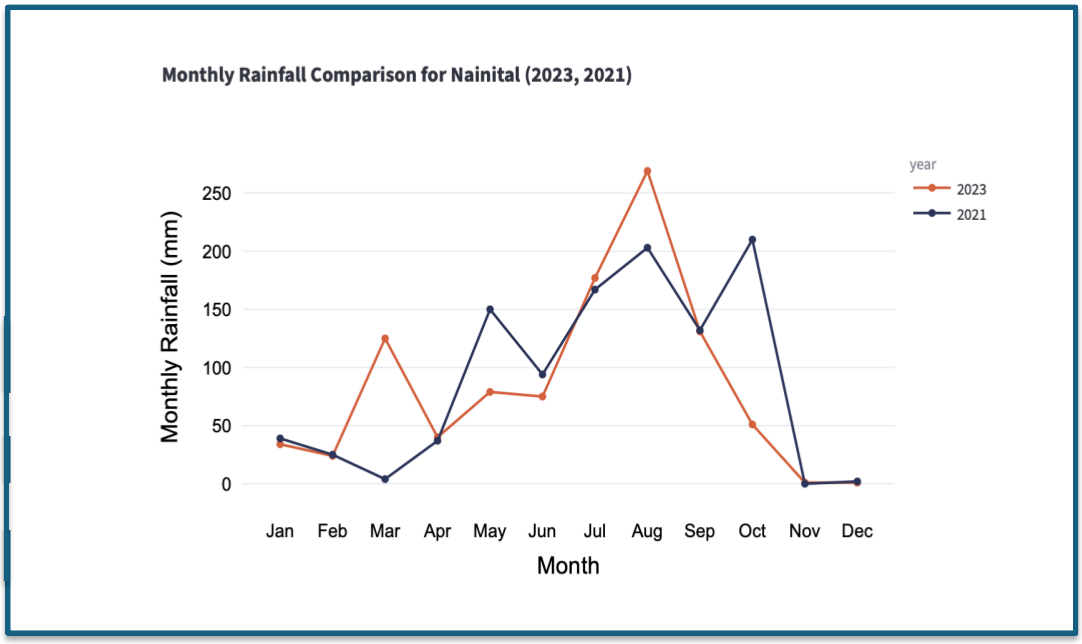


Figure 9.c Yearly Rainfall Trends and Monsoonal Pattern

- Post-monsoon Consistency:
  Both years saw negligible rainfall in November–December, supporting the finding of a sharp monsoon withdrawal.
- Total Annual Rainfall:
  While the total annual rainfall was slightly higher in 2021, 2023's rainfall was more concentrated, with fewer but more intense rainy months.
- Implications:
  The shift in monsoon timing and intensity has direct consequences for agriculture, water resource management, and flood risk. The 2023 pattern may increase flood potential, while 2021's extended wet season could support longer cropping cycles

9.4 Extreme Rainfall Events (2018–2024)

- IMD Classification Reference:
  According to the India Meteorological Department (IMD):
  - Very Heavy Rainfall (VHR): 124.5–244.4 mm/day.
  - Extremely Heavy Rainfall (EHR): >244.5 mm/day.
- Recorded Extreme Event:
  - Date: October 18, 2021
  - Location: Nainital
  - Rainfall: 156 mm (VHR category)

Figure 9.d Extreme Rainfall Events (2018–2024)

- Rarity and Timing:
  Only one VHR event was recorded in the 2018–2024 period, making such extremes rare in Nainital. This event occurred post-monsoon, aligning with the late rainfall peak in 2021.

- Risk and Impact:

  The rarity of such events underscores regional climatic stability. However, their occurrence, especially late in the season, raises the risk for landslides, urban flooding, and challenges for emergency management.

9.5 Temperature Extremes and Shifting Baselines (2017–2025)

- Maximum Temperatures:
  - Highest Recorded: 34.7°C in 2017.
  - Lowest Maximum: 24.4°C in 2025, a decline of over 10°C from the peak.
  - Declining Trend: There is a discernible downward trend in maximum summer temperatures, especially from 2022 to 2025, possibly due to increased cloud cover or rainfall.
- Minimum Temperatures:
  - Coldest Year: 2021, with a minimum of -1.0°C.
  - Recent Warming: By 2025, the minimum temperature increased to 1.4°C, indicating milder winter conditions.

- Trend: After some sharp lows in 2020–2022, minimum temperatures are now trending upward.

- Interpretation:
  The narrowing range between summer highs and winter lows points to a moderation of temperature extremes. This could be linked to regional climate changes, increased rainfall/cloud cover, or natural variability.
- Potential Impacts:
  - Cooler summers may reduce heat-related stress on local populations.
  - Warmer winters may affect frost-sensitive crops, ecological cycles, and biodiversity.

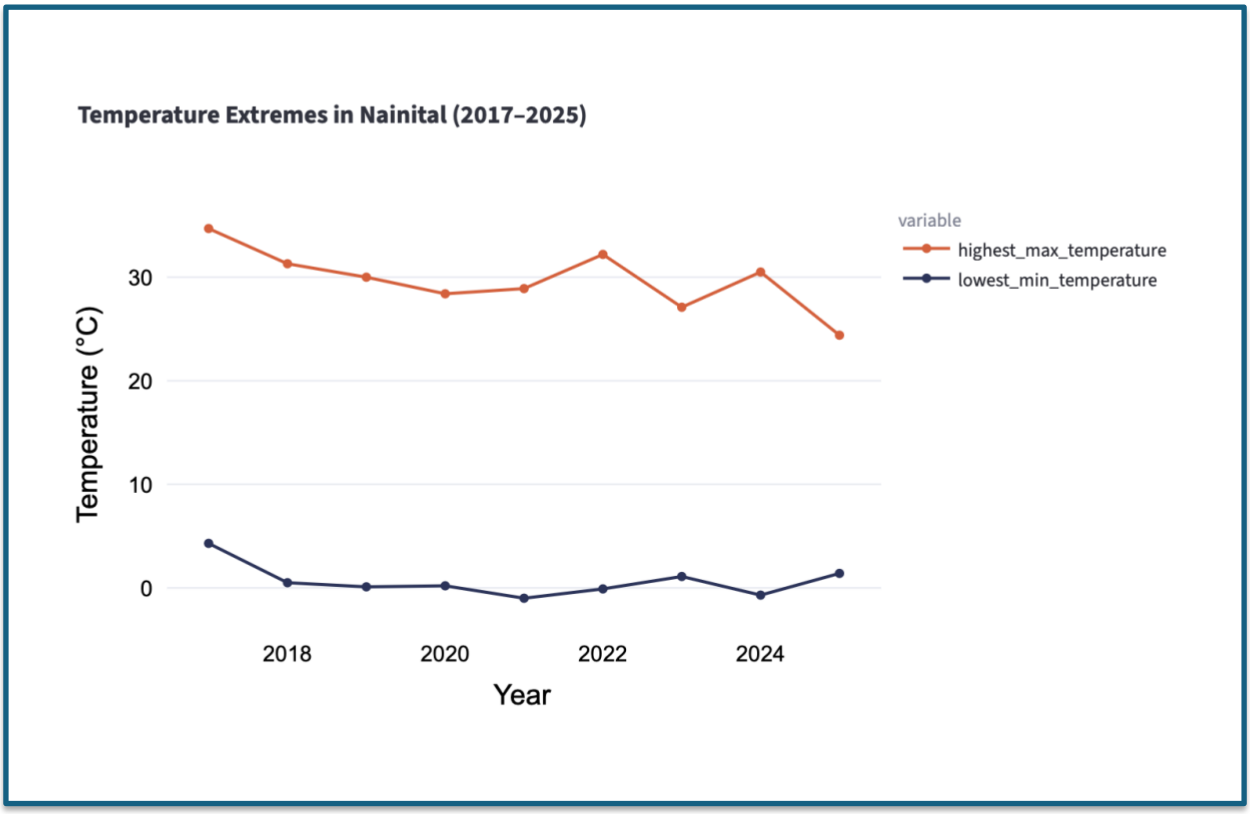


Figure 9.e Temperature Extremes and Shifting Baselines (2017–2025)

9.6 Daily Temperature Comparison: 2019 vs 2023

Volatility vs. Stability:

Metric	2019	2023
Max Daily Temp	Peaked at 30°C (Jan)	Stayed below 25°C
Min Daily Temp	Dipped to ~2°C	Mostly above 2°C
Avg Temp	More fluctuations	Smoother mid-year peak

2019 saw greater temperature extremes and variability, while 2023 was characterized by more moderate, stable temperatures. This supports the broader trend toward temperature moderation post-2017.

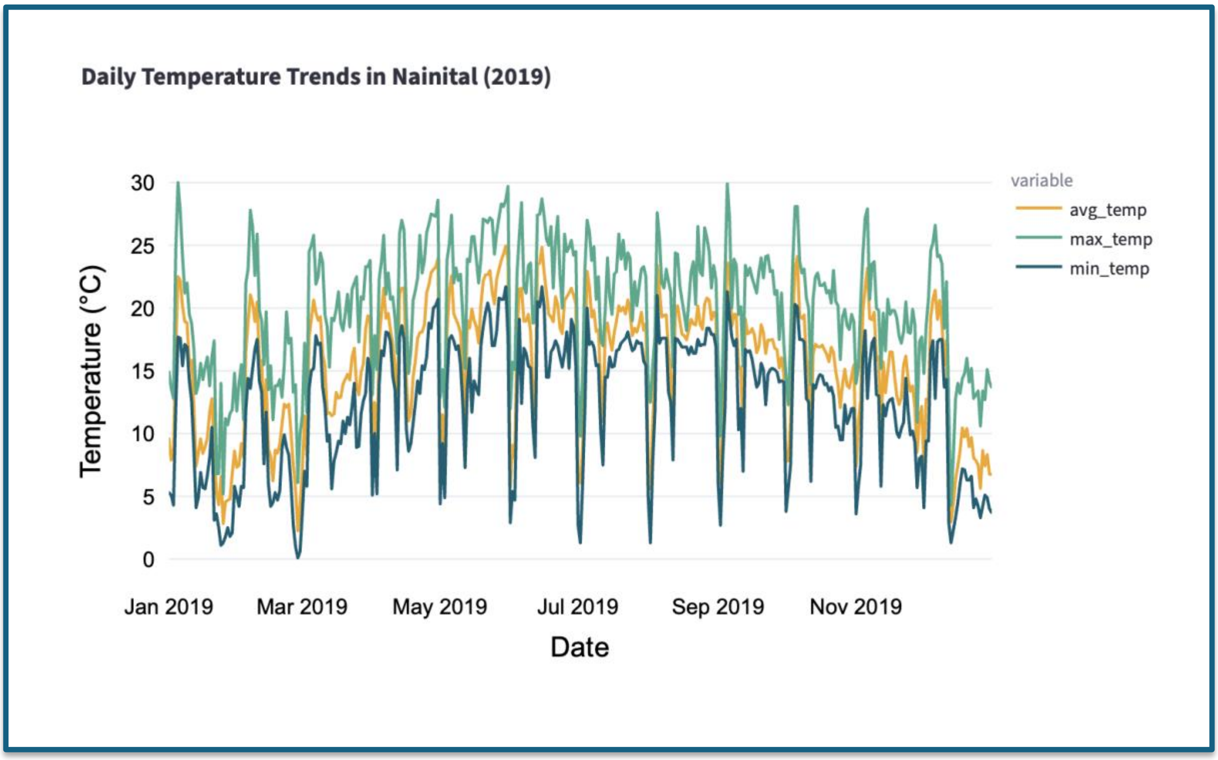


Figure 9.f Daily Temperature Comparison: 2019 vs 2023

9.7 Monthly Rainfall Anomalies and Distribution: 2019 vs 2021

Rainfall in 2019

- Characteristics:
    - High rainfall in January (~135 mm) and February (~125 mm) unusually wet winter.
    - Very low rainfall in July (~30 mm) indicating a critical monsoon failure month.
    - Single peak in August (~160 mm) followed by a dip in September (~115 mm).
    - Scattered rainfall in November and December (moderate but inconsistent).
- Assessment: Poor Distribution
    - Highly erratic distribution: Winter months were wetter than the early monsoon months.
    - Monsoon rainfall (June–Sept) was uneven, with a drastic drop in July, which is typically the wettest month.
    - Such irregularity can negatively affect agricultural planning, water storage and supply predictability, and ecosystem response.

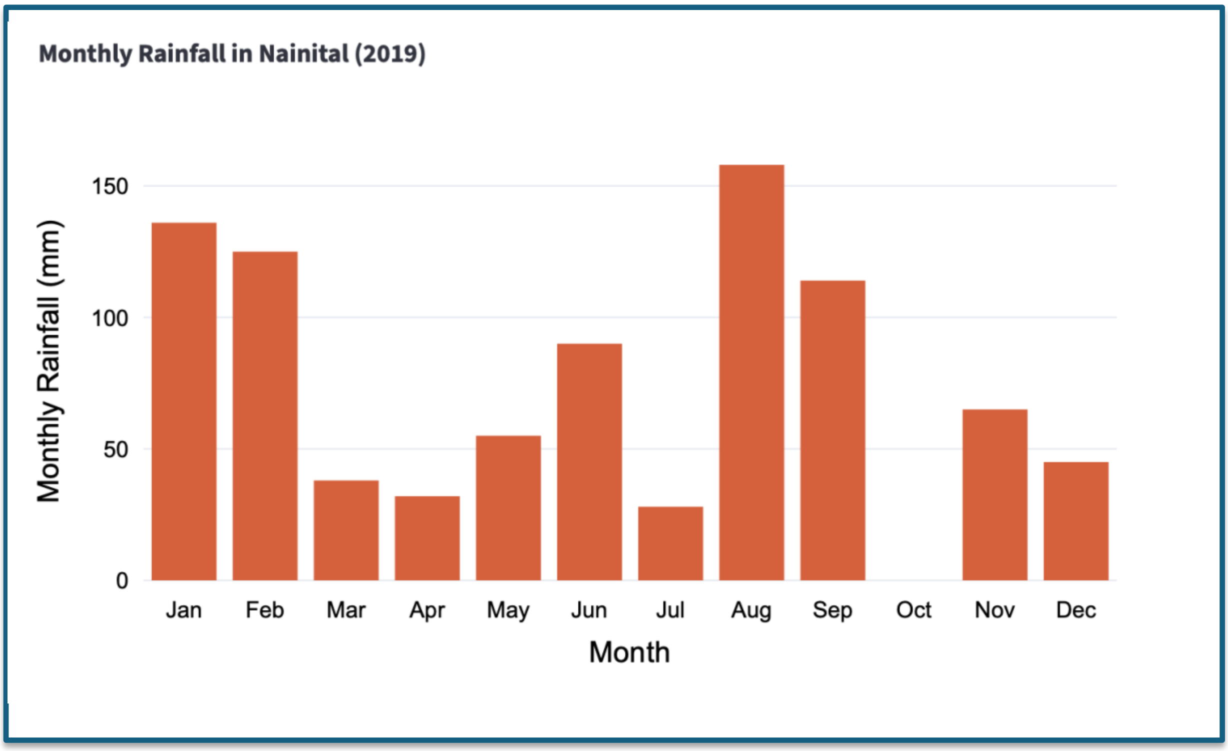


Figure 9.g Monthly Rainfall Distribution 2019

Rainfall in 2021

- Characteristics:
    - Dry January–April, with almost no rainfall in March.
    - Sharp increase in May (~150 mm) and sustained high rainfall through:
- July (~170 mm)
- August (~205 mm)
- October (~210 mm)
    - Tapering begins in November, with December being nearly dry.
- Assessment: Good Distribution
    - Clear, strong monsoon season from May to October.
    - Well-distributed rainfall during the growing season, supporting agriculture.
    - Minimal winter rainfall, which is climatologically expected and beneficial to maintain seasonal balance.

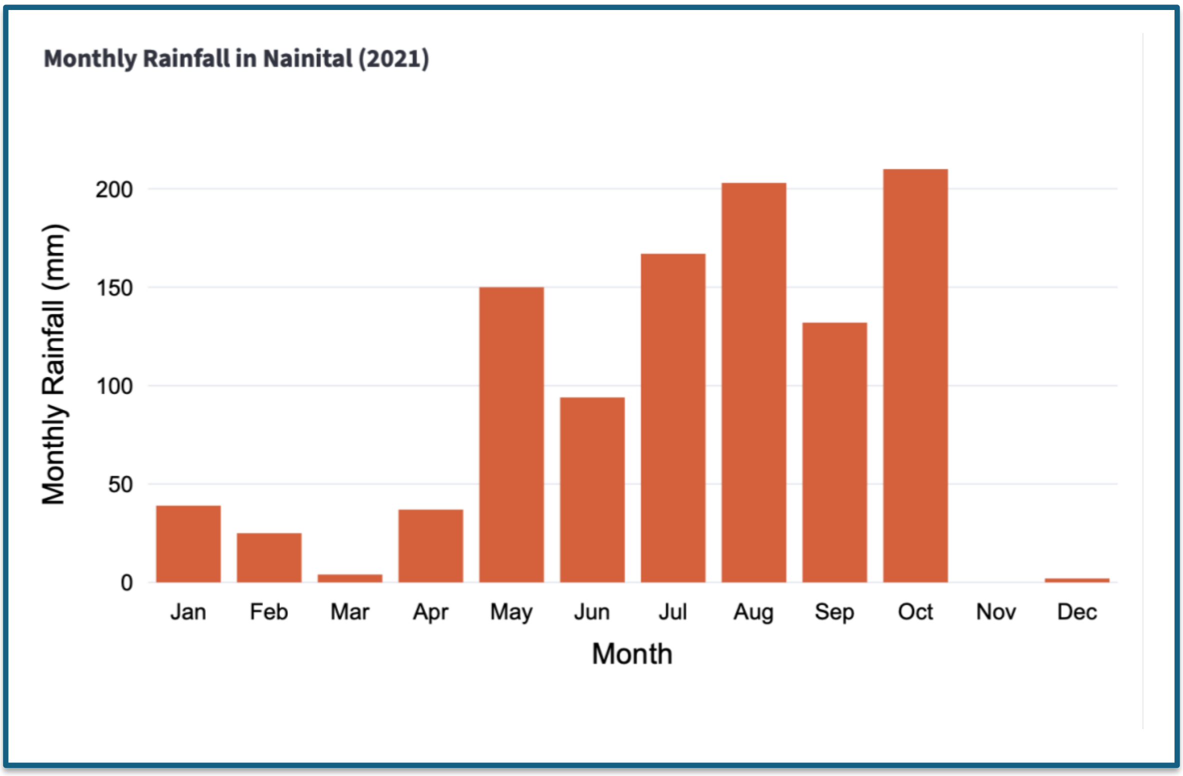


Figure 9.h Monthly Rainfall Distribution 2021

9.8 Machine Learning Model Evaluation for Temperature Prediction

To assess the effectiveness of different machine learning algorithms for temperature prediction in Nainital, two models were evaluated: Random Forest and Linear Regression. Both models were trained on historical data from 2017–2019 and used to predict daily average temperature for the year 2020, using identical input features to ensure a fair comparison (hourly rain, daily rain, maximum/minimum temperatures, and humidity).

Model Configurations

- Random Forest:
    - 500 trees (n_estimators=50)
    - Input features: hourly rain, daily rain, max/min temperatures
    - Target Variable: Temperature
- Linear Regression:
    - Same input features and training years as Random Forest

Performance Comparison

- Random Forest:
    - $R^2 = 0.9866$
    - MAE = 0.3335
    - MSE = 0.2839
- Linear Regression:
    - Displayed lower predictive accuracy (precise metrics may be included as available)
    - The scatter plot for this model showed a wider spread from the ideal y=x line, and the time series plot exhibited larger deviations between actual and predicted values.

Interpretation & Insights

- The Random Forest model was highly effective in capturing non-linear relationships within the weather data, achieving robust and accurate temperature predictions.
- Linear Regression, while simpler and more interpretable, could not match the accuracy of Random Forest, especially during periods of rapid temperature change or extreme events.

- Visualizations supported these results: Random Forest predictions closely tracked actual values with minimal deviation, while Linear Regression predictions were less precise, particularly during seasonal transitions or anomalous periods.

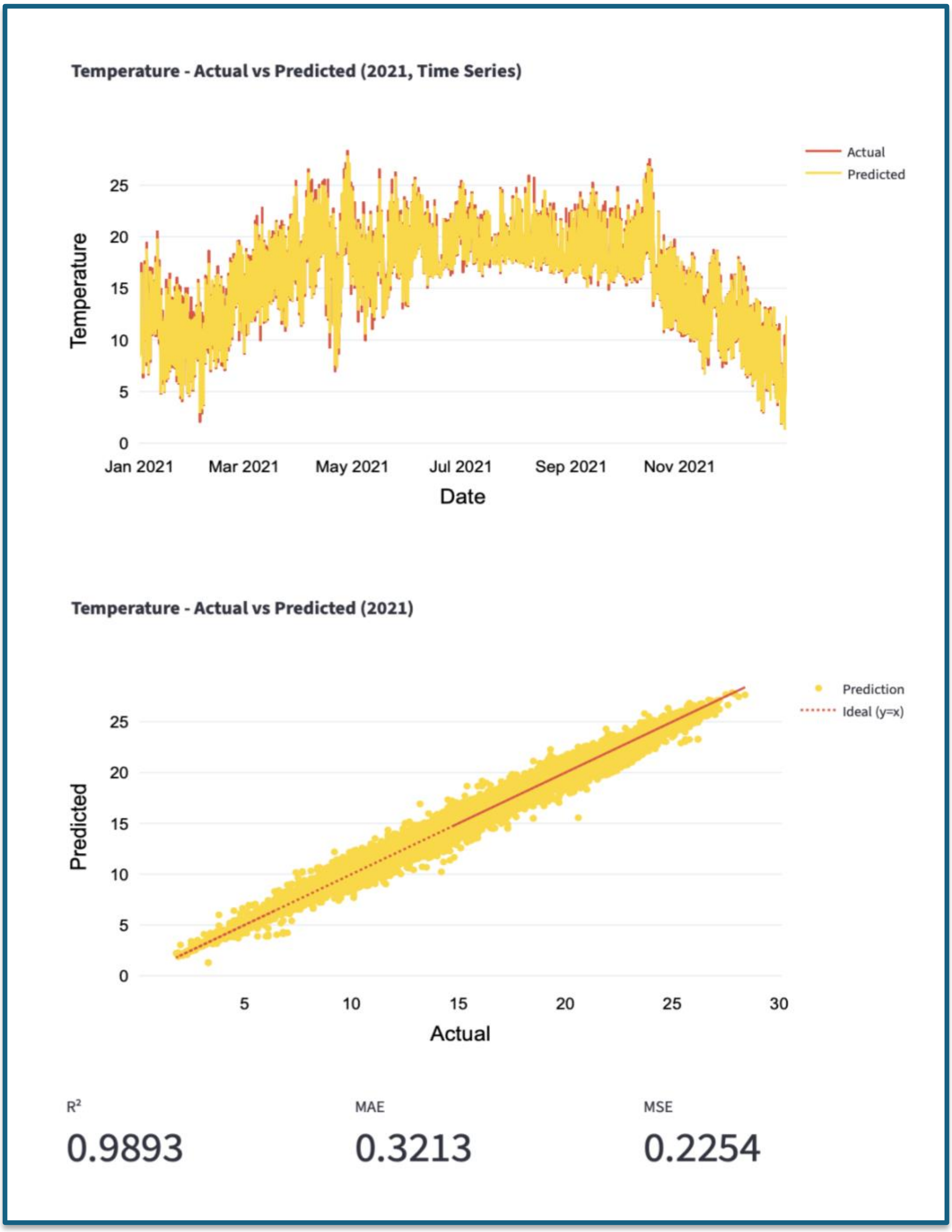


Figure 9.h Random Forest Regressor showing the best fir for the Data set

Implications

- Non-linear, ensemble approaches like Random Forest are strongly recommended for temperature forecasting in this region, especially when accuracy during variable conditions is critical.
- Simpler models may still provide value for baseline comparison or when interpretability is paramount, but users should be aware of their limitations in capturing complex environmental dynamics.

9.9 Synthesized Key Insights and Implications

- Changing Rainfall Patterns:
    - The monsoon is peaking earlier in recent years (notably in July 2022) and receding sharply thereafter.
    - Years like 2021 feature late monsoon peaks and erratic monthly rainfall, while years like 2023–2024 show greater stability.
- Rare but Impactful Extremes:
    - Only one Very Heavy Rainfall event was observed over eight years, emphasizing the rarity but potential impact of such extremes.
- Temperature Moderation:
    - Both maximum and minimum temperature trends point to a narrowing temperature range, indicative of climatic moderation or increased rainfall/cloudiness.
- Climatic Stability and Early Warning Signs:
    - Fewer extreme events and less volatility in recent years could signal a phase of short-term stability. However, shifting monsoon patterns, delayed onsets, and warming winters may be early indicators of broader climate variability.
- Machine Learning Insights:
    - Advanced models such as Random Forest significantly outperform simpler linear models for temperature prediction, underlining the importance of capturing non-linear relationships in environmental data.
- Practical Applications:
    - These findings are crucial for local planning—informing agriculture, water management, disaster preparedness, and climate adaptation strategies.

In summary, the data reveals a region in transition: while Nainital's climate remains generally stable, there are emerging signals of changing monsoonal dynamics, moderating temperature extremes, and the need for robust, non-linear predictive models. Continued monitoring and flexible resource management will be essential to adapt to these evolving patterns.

10. Challenges & Learnings

- Data Quality and Completeness:
  The raw AWS datasets often contained missing values, outliers, and inconsistencies in formats or units. Dealing with sensor errors, gaps in time series, and ensuring data validity required extensive cleaning, imputation, and cross-checking.
- Heterogeneity Across Years and Stations:
  Data from different years and between AWS1 and AWS2 sometimes varied due to changes in instrumentation, calibration standards, or recording protocols. This made direct comparisons and integrated analysis more complex.
- Detection and Treatment of Outliers:
  Isolating genuine extreme weather events from erroneous readings demanded careful analysis. It was at times challenging to distinguish between rare climatic events and sensor malfunctions, with both potentially distorting statistical summaries or model training.
- Short Temporal Coverage:
  The limited span of available data (approx. 7–8 years) constrained the ability to draw robust conclusions about long-term climate trends and made it difficult to separate natural variability from emerging patterns.
- Model Selection and Hyperparameter Tuning:
  Determining the most suitable machine learning models and their configurations required iterative experimentation. Balancing predictive accuracy, interpretability, and computational feasibility was a continual challenge.
- Interpretability of Machine Learning Outputs:
- While advanced models (e.g., Random Forest) improved predictive performance, explaining their predictions and extracting actionable insights demanded additional effort, particularly for audiences less familiar with such techniques.
- Visualisation and Communication:
  Effectively communicating findings—especially subtle trends, anomalies, or implications for local stakeholders—required thoughtful design of visualizations and narrative clarity to bridge technical and non-technical audiences.

11. Conclusion

This study presents a thorough examination of temperature and rainfall dynamics in Nainital using multi-year data from two automated weather stations (AWS1 and AWS2). The analysis reveals the following overarching conclusions:

- Climatic Stability with Emerging Variability:

  While the region demonstrates notable climatic stability—particularly in daily temperature ranges and the rarity of extreme rainfall events—there are clear signals of emerging variability. These include shifts in monsoon onset and retreat, irregular rainfall distributions in certain years, and a gradual moderation of temperature extremes.
- Monsoon Shifts and Rainfall Distribution:

  The timing and intensity of the monsoon are changing, with some years experiencing earlier peaks and others showing late-season rainfall surges. Rainfall distribution is not always optimal for agriculture or water management, as evidenced by years like 2019 (poor monsoon) and 2021 (well-distributed rainfall).
- Temperature Moderation:

  There is a discernible trend toward cooler maximum summer temperatures and milder winter lows, resulting in a narrowing annual temperature range. This could be due to increased cloud cover, changing rainfall patterns, or broader climatic factors.
- Extreme Events Remain Rare but Impactful:

  The occurrence of only a single Very Heavy Rainfall event (as per IMD criteria) over several years underscores the rarity of such extremes in Nainital. However, even isolated events have significant implications for disaster management and local infastructure.
- Advantages of Advanced Predictive Models:

  Machine learning experiments demonstrate that non-linear ensemble approaches, such as Random Forests, substantially outperform simpler linear models for temperature prediction. This highlights the value of integrating advanced analytics into local climate monitoring and forecasting.

- Need for Continued Monitoring and Adaptation:

  The observed patterns, especially shifting monsoon behavior and temperature moderation—underscore the importance of ongoing data collection, flexible water and agriculture management strategies, and adaptive planning for climate resilience.

12. Future Scope

- Enhance Long-Term Data Collection:
- Expanding the temporal scope and quality control of AWS datasets will improve the robustness of trend detection and future projections.
- Integrate Advanced Analytics:
- Continued use and development of machine learning and statistical models will enhance predictive capabilities and support better decision-making for local stakeholders.
- Promote Climate-Aware Resource Management:
  Water management, agriculture, and infrastructure planning should be informed by emerging climate patterns to mitigate risks from irregular rainfall and temperature shifts.
- Encourage Collaboration and Data Sharing:
  Facilitating collaboration between meteorological agencies, researchers, and local authorities will foster comprehensive and actionable climate insights for the region.

In summary, while Nainital's climate remains relatively stable, subtle but significant changes are underway. Proactive monitoring, advanced analytics, and adaptive management will be key to safeguarding the region's environmental and socioeconomic well-being in the face of evolving climatic trends.

13. Appendix

13.1 AWS Data Field Descriptions

AWS 1 (Naini_Data) – Example Fields

- station_name: Name of the weather station.
- date: Date of observation.
- time: Time of observation.
- gps: GPS coordinates of station.
- battery_v: Battery voltage.
- Hourly Rain: Rainfall recorded in the last hour (mm).
- Daily Rain: Total rainfall for the day (mm).
- sm_m3_m3: Soil moisture content (m^3/m^3).
- Temperature: Measured air temperature (°C).
- Maximum Temperature: Highest temperature recorded for the day (°C).
- Minimum Temperature: Lowest temperature recorded for the day (°C).
- Wind Speed: Wind speed (m/s or km/h as per station).
- Wind Degree: Wind direction (degrees).
- Pressure: Atmospheric pressure (hPa).
- Humidity: Relative humidity (%).
- Sun Duration: Duration of sunshine (hours).

AWS 2 (Naini_Data2) – Example Fields

- station_name: Name of the weather station.
- date: Date of observation.
- time: Time of observation.
- temp_deg_c: Measured air temperature (°C).
- temp_max_deg_c: Maximum daily temperature (°C).
- temp_min_deg_c: Minimum daily temperature (°C).
- humidity_percent: Relative humidity (%).
- pressure_hpa: Atmospheric pressure (hPa).
- ws_ms: Wind speed (m/s).
- wd_deg: Wind direction (degrees).
- rain_daily_mm: Total daily rainfall (mm).
- solar_radiation: Solar radiation (W/m^2 or as recorded).

- uv_a: UV-A radiation level.
- uv_b: UV-B radiation level.
- battery_voltage: Battery voltage (V).
- peripherals: Status or reading from additional sensors.

13.2 Sample Data Records

Sample AWS 1 Record:

station_name: Nainital_AWS1
date: 2022-07-15
time: 12:00
gps: 29.392, 79.458
battery_v: 12.4
Hourly Rain: 2.3
Daily Rain: 15.7
sm_m3_m3: 0.22
Temperature: 18.5
Maximum Temperature: 21.0
Minimum Temperature: 16.2
Wind Speed: 3.8
Wind Degree: 95
Pressure: 847.2
Humidity: 87
Sun Duration: 5.6

Sample AWS 2 Record:

station_name: Nainital_AWS2
date: 2023-08-10
time: 14:00
temp_deg_c: 19.2
temp_max_deg_c: 22.3
temp_min_deg_c: 16.8
humidity_percent: 80
pressure_hpa: 846.5
ws_ms: 2.5
wd_deg: 120
rain_daily_mm: 12.4
solar_radiation: 680
uv_a: 5.1
uv_b: 0.9
battery_voltage: 12.6
peripherals: 1

13.3 Abbreviations and Acronyms

- AWS: Automated Weather Station
- IMD: India Meteorological Department
- MAE: Mean Absolute Error
- MSE: Mean Squared Error
- R^2: Coefficient of Determination
- hPa: Hectopascal (unit of pressure)
- mm: Millimeter (unit of rainfall)
- °C: Degrees Celsius

13.4 Additional Resources

- Detailed code, processed datasets, and visualization scripts are available upon request.
- For further information on AWS sensor specifications, refer to manufacturer documentation or the IMD technical reports.

```
MySQL (Local)
  ▶ Used for: data storage, cleaning, querying
  |
Python in VS Code
  ├──▶ Wrote scripts for data processing
  ├──▶ Built app using:
  |     ├── Streamlit (web UI)
  |     ├── mysql-connector-python + SQLAlchemy (DB integration)
  |     ├── Pandas (data handling)
  |     ├── Plotly (visualizations)
  |    └── Scikit-learn (ML models)
  |
Local Testing
  └──▶ Ran Streamlit app locally using `python app.py` : Verified data flow, UI, and ML
output
  |
Railway (Cloud MySQL Hosting)
  └──▶ Migrated DB to cloud for global access
  |
Render (App Deployment)
  └──▶ Final app deployed and hosted publicly
```