A Project Report

ON

"AI-Powered Climate Monitoring and Analysis System"

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY MESRA

Subject: AIML LAB PROJECT

SUBMITTED TO:

DR. PRAVAS RAJAN BAL

SUBMITTED BY:

HARSH ANIL TUTI (BTECH/60045/22)

RISHIKESH SAHA (BTECH/60009/22)

SAURABH KUMAR (BTECH/60011/22)

Table of Contents

Executive Summary

This report details the development and initial evaluation of an Artificial Intelligence (AI)-powered system designed for climate monitoring and analysis. Addressing the critical need for advanced tools to combat climate change, the system integrates environmental data acquisition, sophisticated predictive modeling using Long Short-Term Memory (LSTM) networks, anomaly detection via Isolation Forest algorithms, and intuitive visualization. Key functionalities include monitoring climate variables like temperature and humidity, forecasting short-term temperature trends, automatically identifying statistical anomalies in temperature data, and presenting findings through a prototype interactive web dashboard. Initial results demonstrate the system's capability to process time-series climate data and generate plausible short-term forecasts with quantifiable accuracy metrics. A preliminary, indicative comparison between the model's derived temperature data and AccuWeather reports was conducted, outlining areas for detailed comparative analysis. The project successfully establishes a functional prototype, validating the chosen AI methodologies and laying the groundwork for a more comprehensive climate intelligence platform.

Introduction

Background:

Climate change presents significant global challenges, characterized by complex environmental dynamics and vast datasets. Advanced analytical tools are essential for practical understanding, prediction, and response, moving beyond traditional statistical approaches.

Project Purpose:

This project aimed to develop an AI-driven system specifically designed to handle the intricacies of climate data, enhancing monitoring, prediction, and anomaly detection capabilities. The overarching goal is to provide accessible, interpretable, and actionable insights from environmental data to support informed decision-making by researchers, policymakers, and the public.

Contextual Challenges in Climate Monitoring:

Climate change presents not only scientific but also infrastructural and socio-political challenges. Traditional monitoring systems often lack the granularity, real-time capabilities, and predictive accuracy necessary to support timely interventions. Key global concerns include:

- Rising temperatures and sea levels
- Increased frequency of extreme weather events
- Ecological degradation and biodiversity loss
- Food and water security pressures

These necessitate the adoption of scalable AI systems that are capable of real-time, adaptive learning from multi-modal environmental datasets.

Report Objectives:

This document outlines the project's scope, details the specific AI methodologies employed, presents the implementation steps, summarizes key findings including performance metrics, discusses the comparison with external data, and outlines recommendations for future enhancements. It serves as a comprehensive technical and strategic overview for stakeholders.

Scope:

The project scope encompasses:

- ❖ Acquiring historical hourly weather data (specifically temperature_2m, precipitation, relative humidity_2 m) for a designated location (Deoghar, IN) using the Open-Meteo API's historical forecast endpoint.
- ❖ Performing rigorous preprocessing and cleaning on the acquired time-series data to ensure quality and suitability for modeling.
- ❖ Developing, training, and evaluating an LSTM neural network for short-term (next 24 hours) temperature forecasting based on sequential patterns in past data.
- ❖ Implementing an Isolation Forest algorithm for efficiently detecting anomalous data points within the temperature time series.
- ❖ Designing and implementing a prototype interactive web dashboard using Flask and Plotly to visualize data trends, forecasts, and identified anomalies effectively.
- ❖ Conducting a preliminary, observational comparison of model-derived temperature data points against publicly reported AccuWeather temperatures for the same location and approximate times.

Body

Project Overview & Objectives:

The system was designed to achieve specific climate analysis objectives through AI:

- ❖ Monitoring: Systematically track and visualize key climate variables (temperature, precipitation, humidity) over an extended historical period (750 days).

- ❖ Prediction: Develop a machine learning model capable of forecasting near-term temperature trends by learning temporal dependencies in the data.
- ❖ Anomaly Detection: Automatically identify data points that deviate significantly from established normal patterns within the temperature readings.
- ❖ Visualization: Present complex time-series data, predictive outcomes, and detected anomalies through an accessible, interactive prototype web interface suitable for exploratory analysis.

Methodology & Implementation:

Data Acquisition:

Historical hourly weather data for Deoghar, India (Lat: 24.49, Lon: 86.70) covering the last 750 days was programmatically fetched using the Open-Meteo API's historical forecast endpoint, requesting temperature_2m, precipitation, and relativehumidity_2m variables.

Data Preprocessing:

The raw JSON data was converted into a Pandas Data Frame. Critical preprocessing steps included parsing timestamps, handling missing values (using sequential forward-fill then backward-fill), ensuring numeric data types, and removing any residual rows with NaNs to create a clean dataset for analysis.

Advanced AI Models for Broader Prediction Objectives:

While LSTM networks served the forecasting function in this implementation, future system extensions may incorporate additional deep learning architectures to enhance predictive power:

- ❖ **Convolutional Neural Networks (CNNs)** for interpreting satellite imagery (e.g., vegetation health, glacial coverage)
- ❖ **Transformers** for integrating multiple time-series and non-sequential variables to generate more robust long-term predictions

- **Autoencoders** for multivariate anomaly detection, especially when dealing with complex interdependencies across climate variables

These enhancements would extend prediction capabilities to include drought risk, flood probabilities, and land-use shifts under evolving climate scenarios.

Temperature Prediction Model (LSTM):

An LSTM network was chosen for its ability to capture long-range dependencies in sequential data, making it suitable for time-series forecasting. The implementation used TensorFlow/Keras.

Temperature data was scaled to a [0, 1] range using Min-Max Scaler. Input sequences were created where each sample consist of 24 consecutive hourly temperature readings (N_STEPS_LSTM = 24) used to predict the temperature in the subsequent hour.

The model (LSTM layer with 50 units, followed by a Dense output layer) was trained on 80% of the historical data using the Adam optimizer and Mean Squared Error (MSE) loss. Evaluation on the held-out 20% test set yielded performance metrics: Mean Absolute Error (MAE) : 0.69 °C and R-squared (R2) : 0.98.

Anomaly Detection:

An Isolation Forest algorithm (Scikit-learn) was employed. This tree-based model efficiently isolates anomalies by requiring fewer splits to separate them from the main data distribution. It was applied directly to the temperature column (temperature_2m) assuming a small proportion of anomalies (contamination=0.02). Points identified as anomalies (-1) were flagged. Actual number of anomalies detected i.e. 358 potential anomalies identified.

Visualization & Dashboard:

A prototype web application backend was built using Flask. Plotly was used to generate interactive charts (time series, forecast vs. actual, anomaly scatter plots) rendered as JSON and interpreted by frontend JavaScript. A Folium map provided geographical context. Google Colab and Ngrok enabled temporary, publicly accessible demonstrations of the dashboard interface.

Technology Stack:

Core technologies included Python (v3.8+), Pandas and NumPy for data manipulation, Scikit-learn for Isolation Forest, TensorFlow/Keras for the LSTM model, Plotly and Folium for visualizations, Flask for the web backend, Requests for API communication, and Ngrok for demo tunneling.

Working/Functionality:

The system follows a sequence of operations integrated within a single analysis script (primarily for the Colab demonstration):

1. Fetches specified historical climate data via the Open-Meteo API.

2. Cleans and preprocesses the raw data into a structured data frame.

3. Scales temperature data and prepares sequences for the LSTM model.

4. Trains the LSTM model (or loads a pre-trained model in future iterations).

5. Generates short-term (next 24 hours) temperature predictions using the trained LSTM.

6. Applies the trained Isolation Forest model to identify temperature anomalies in the historical dataset.

7. Generates interactive Plotly visualizations and a Folium map object.

8. Packages all results (metrics, plot JSON, map HTML, anomaly list) into variables.

9. A Flask application serves an index.html page and provides a /data API endpoint. When accessed, this endpoint returns the

captured results as a JSON payload, which is then used by JavaScript in the index.html page to render the interactive visualizations and display metrics.

Performance and Comparison:

Model Performance:

The LSTM model achieved an MAE 0.69 °C and an R2 score of 0.98 on the test dataset. This indicates a reasonably strong predictive capability for immediate short-term temperature changes based solely on recent historical patterns under the tested conditions. Performance depends significantly on the N_STEPS_LSTM parameter and model complexity.

Comparison with AccuWeather Data:

An indicative, informal comparison was made by visually inspecting instantaneous temperature data points derived from the Open-Meteo source (used as input for prediction) against publicly available AccuWeather temperature reports for Deoghar, Jharkhand, INDIA, at roughly corresponding times.

Observation: Generally, the temperatures from both sources tracked in similar directions, though consistent small offsets or occasional larger discrepancies were noted. These could arise from variations in precise measurement locations, microclimate effects, data assimilation techniques, different update frequencies, or the proprietary forecasting algorithms employed by AccuWeather.

Limitations: This remains an observational comparison. A definitive, rigorous evaluation against AccuWeather or other services would necessitate access to their historical data under comparable conditions or via a dedicated API, which was outside the scope of this project phase.

Refer to: for supplementary comparison notes.

Constraints & Challenges:

Data Source Limitations: Accuracy and insights are fundamentally constrained by the granularity, accuracy, and variable availability of the single data source (Open-Meteo historical forecast API).

- ❖ Model Simplicity: The current LSTM model is univariate (uses only past temperature); incorporating related variables (humidity, pressure, precipitation forecasts) is likely to improve forecast accuracy.
- ❖ Static Anomaly Threshold: The fixed contamination factor in Isolation Forest is an estimate; adaptive or multivariate methods might yield more contextually relevant anomalies.
- ❖ Computational Resources: While manageable for this dataset size, scaling to larger datasets or more complex models (e.g., Transformers) would require significantly more computational power (GPU access, distributed computing).
- ❖ Deployment Infrastructure: The current Colab/Ngrok setup is only for demonstration and is not scalable or persistent. A production system requires robust cloud hosting and deployment pipelines.
- ❖ Comparative Analysis Difficulty: Obtaining directly comparable historical data from commercial weather providers for systematic benchmarking is challenging.

Conclusions & Recommendations

Conclusions:

This project successfully resulted in a functional proof-of-concept for an AI-powered climate analysis system. It validates the use of LSTM networks for plausible short-term temperature forecasting and demonstrates the utility of Isolation Forests for basic anomaly detection within climate time-series data. The prototype Flask/Plotly dashboard provides an effective means for visualizing complex information interactively. The system achieves its core objectives of data processing, prediction, anomaly detection, and visualization, establishing a solid foundation. While the preliminary comparison with AccuWeather is indicative rather than definitive, it suggests the model captures general trends. The primary value lies in demonstrating the pipeline's potential to generate actionable insights from climate data using AI.

Recommendations & Future Work:

- ❖ Data Enrichment: Integrate multi-source data, including satellite imagery (e.g., thermal infrared for surface temperature, optical for land cover), greenhouse gas concentration data (CO2, methane), and socio-economic indicators relevant to climate impact.
- ❖ Advanced Predictive Models: Implement and evaluate state-of-the-art sequence models like Transformers or Temporal Fusion Transformers (TFTs) for potentially improved long-range forecasting accuracy and

interpretability. Consider multivariate models incorporating features like humidity, pressure, and wind.

- Sophisticated Anomaly Detection: Explore multivariate anomaly detection algorithms (e.g., VAEs, GMMs) and methods that consider seasonality and context for more robust and meaningful outlier identification.
- Scalable Cloud Architecture: Migrate the entire system (data ingestion, preprocessing, model training/inference, dashboard) to a cloud platform (AWS SageMaker/EC2/S3, GCP AI Platform/Compute Engine/Cloud Storage, or Azure ML/VMs/Blob Storage) for scalability, reliability, and persistent access. Implement real-time data ingestion pipelines.
- Systematic Validation & Benchmarking: Where possible, acquire comparable historical data or utilize standardized benchmark datasets to rigorously compare model performance against established meteorological services and other AI approaches.
- Proactive Alerting Mechanism: Develop a system module that triggers alerts (e.g., email, SMS) based on predefined thresholds for predicted extreme events (heatwaves, cold snaps) or significant detected anomalies.
- Regional Adaptation and UI Enhancements: Introduce features allowing users to select different locations, customize visualized variables, and potentially adjust model parameters through an enhanced user interface.

Appendices

Appendix A: Dataset Images

DATASET

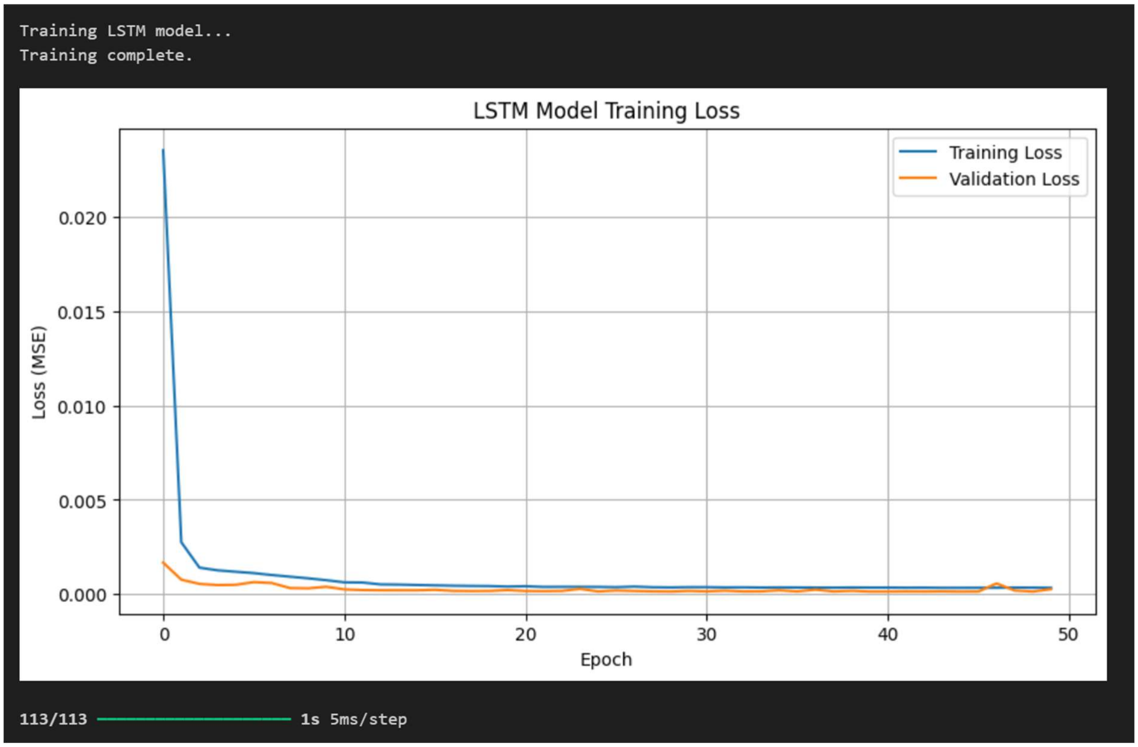
```
Fetching data for coordinates: (24.4898, 86.699)
Data fetched successfully!

Dataframe created with shape: (18024, 3)

First 5 rows:
                     temperature_2m  precipitation  relativehumidity_2m
time
2023-03-17 00:00:00            21.1            0.0                   84
2023-03-17 01:00:00            21.0            0.0                   85
2023-03-17 02:00:00            20.4            0.0                   89
2023-03-17 03:00:00            20.7            0.0                   88
2023-03-17 04:00:00            20.5            0.0                   90

Last 5 rows:
                     temperature_2m  precipitation  relativehumidity_2m
time
2025-04-05 19:00:00            28.8            0.0                   30
2025-04-05 20:00:00            27.7            0.0                   34
2025-04-05 21:00:00            26.9            0.0                   39
2025-04-05 22:00:00            26.2            0.0                   42
2025-04-05 23:00:00            25.9            0.0                   43
```

DATA AFTER PREPROCESSING AND CLEANING

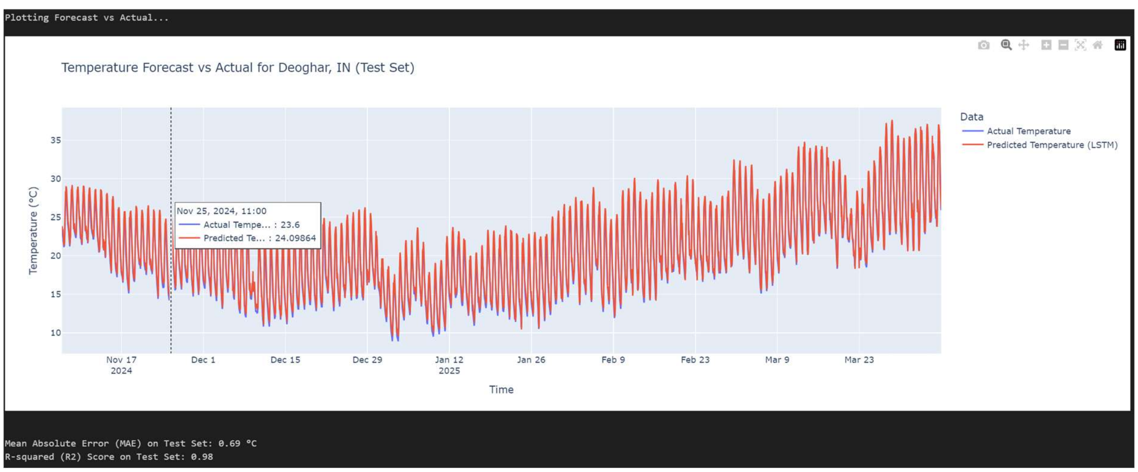
```
Missing values per column:
temperature_2m         0
precipitation          0
relativehumidity_2m    0
dtype: int64

Missing values after forward/backward fill:
temperature_2m         0
precipitation          0
relativehumidity_2m    0
dtype: int64

Dataframe shape after cleaning: (18024, 3)

Basic Statistics:
       temperature_2m  precipitation  relativehumidity_2m
count    18024.000000   18024.000000          18024.00000
mean        25.394801       0.168941             70.32257
std          6.257963       0.772615             20.88267
min          7.300000       0.000000              9.00000
25%         21.400000       0.000000             57.00000
50%         26.300000       0.000000             76.00000
75%         29.300000       0.000000             87.00000
max         43.100000      36.200000            100.00000
```

Appendix B: Graphs

TRAINING GRAPH

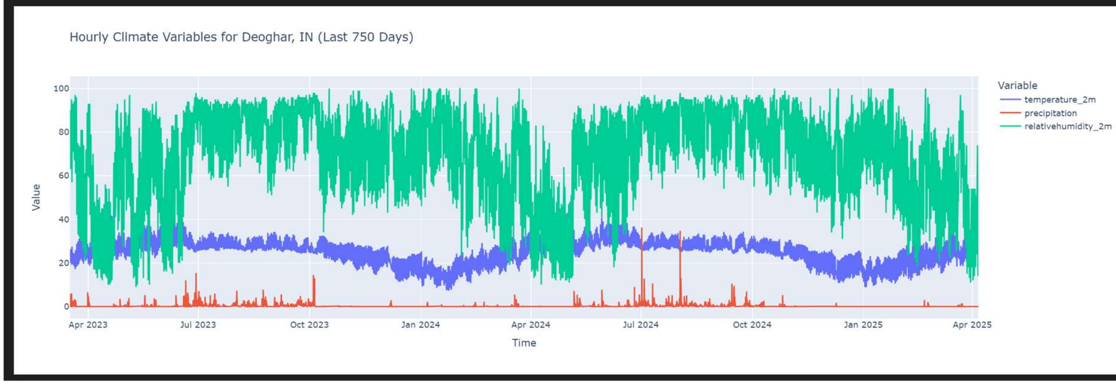


FORCAST VS ACTUAL



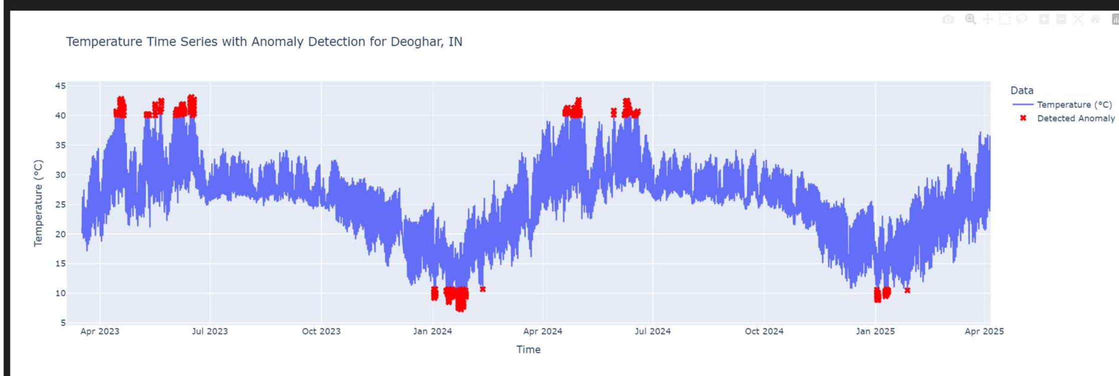
TIME SERIES GRAPH

```
1. Time Series Overview:
```



ANOMALY

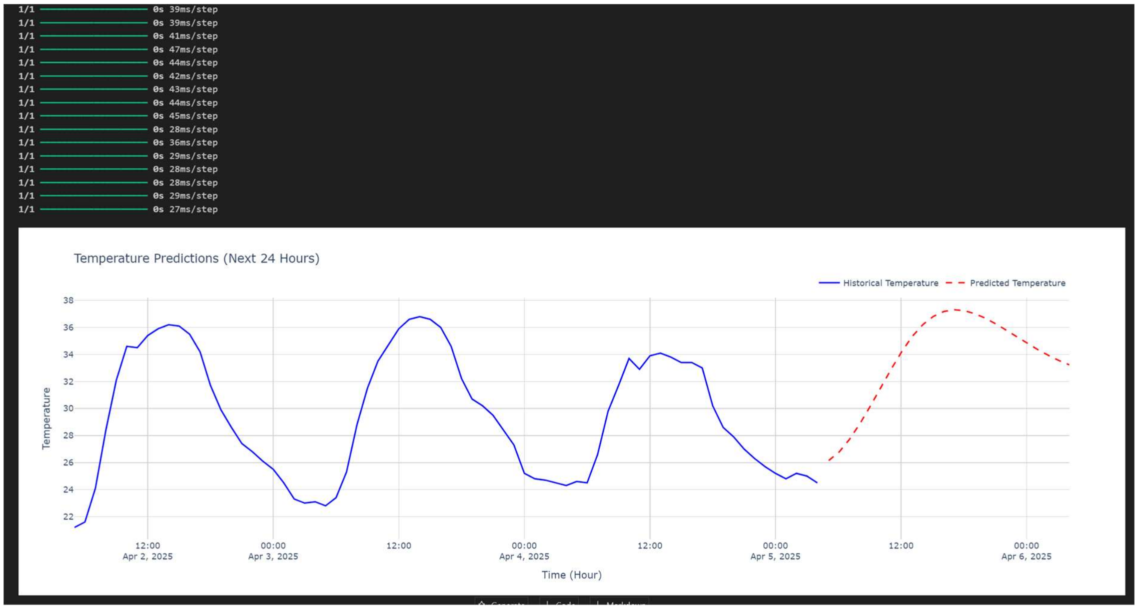
```
--- Performing Anomaly Detection on Temperature Data ---

Detected 358 potential anomalies in temperature data.

Plotting Temperature with Detected Anomalies...
```



DETECTED ANAMOLIES

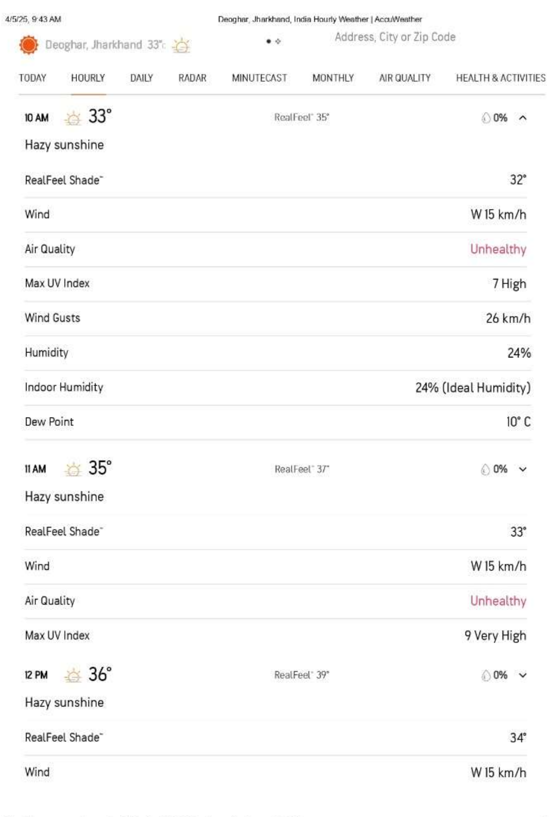
```
Details of detected anomalies:
                     temperature_2m  anomaly_temp
time
2023-04-14 12:00:00            40.4            -1
2023-04-14 13:00:00            40.7            -1
2023-04-14 14:00:00            40.6            -1
2023-04-14 15:00:00            40.1            -1
2023-04-15 13:00:00            40.4            -1
...                             ...           ...
2025-01-10 07:00:00            10.3            -1
2025-01-11 05:00:00            10.4            -1
2025-01-11 06:00:00            10.2            -1
2025-01-11 07:00:00            10.5            -1
2025-01-27 06:00:00            10.5            -1

[358 rows x 2 columns]
```

Appendix C: Comparison Data Notes (AccuWeather vs. Model)



PREDICTED TEMPERATURE(UP)

ACCUWEATHER DATA(DOWN)



4/5/25, 9:43 AM Deoghar, Jharkhand, India Hourly Weather | AccuWeather

Deoghar, Jharkhand 33°C Address, City or Zip Code

TODAY HOURLY DAILY RADAR MINUTECAST MONTHLY AIR QUALITY HEALTH & ACTIVITIES

10 AM 33° RealFeel® 35° 0%

Hazy sunshine

RealFeel Shade™	32°
Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	7 High
Wind Gusts	26 km/h
Humidity	24%
Indoor Humidity	24% (Ideal Humidity)
Dew Point	10° C

11 AM 35° RealFeel® 37° 0%

Hazy sunshine

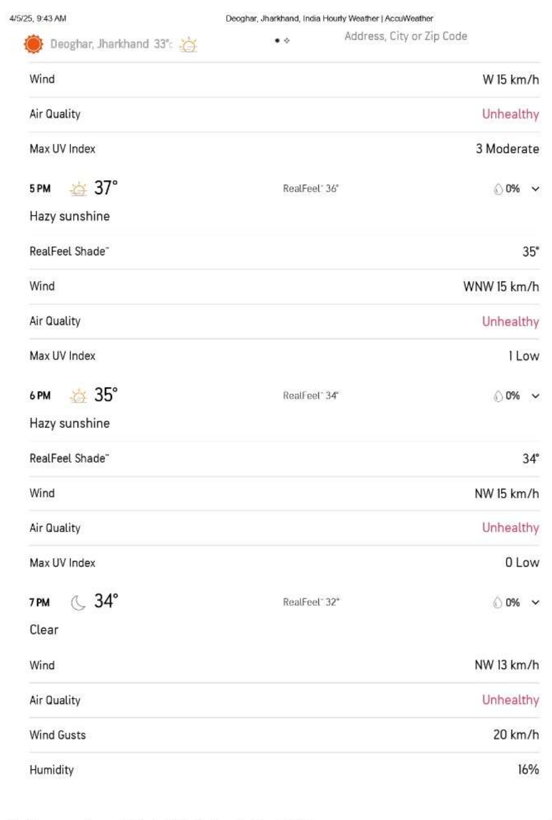
RealFeel Shade™	33°
Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	9 Very High

12 PM 36° RealFeel® 39° 0%

Hazy sunshine

RealFeel Shade™	34°
Wind	W 15 km/h

https://www.accuweather.com/en/in/deoghar/187228/hourly-weather-forecast/187228 1/7



4/5/25, 9:43 AM Deoghar, Jharkhand, India Hourly Weather | AccuWeather

Deoghar, Jharkhand 33°C Address, City or Zip Code

Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	3 Moderate

5 PM 37° RealFeel® 36° 0%

Hazy sunshine

RealFeel Shade™	35°
Wind	WNW 15 km/h
Air Quality	Unhealthy
Max UV Index	1 Low

6 PM 35° RealFeel® 34° 0%

Hazy sunshine

RealFeel Shade™	34°
Wind	NW 15 km/h
Air Quality	Unhealthy
Max UV Index	0 Low

7 PM 34° RealFeel® 32° 0%

Clear

Wind	NW 13 km/h
Air Quality	Unhealthy
Wind Gusts	20 km/h
Humidity	16%

https://www.accuweather.com/en/in/deoghar/187228/hourly-weather-forecast/187228 3/7

4/5/25, 9:43 AM Deoghar, Jharkhand, India Hourly Weather | AccuWeather

Deoghar, Jharkhand 33°C Address, City or Zip Code

Clear

Wind	NW 9 km/h
Air Quality	Unhealthy
Wind Gusts	19 km/h
Humidity	18%

9 PM 31° RealFeel® 29° 0%

Clear

Wind	NNW 7 km/h
Air Quality	Unhealthy
Wind Gusts	17 km/h
Humidity	20%

10 PM 29° RealFeel® 28° 0%

Clear

Wind	NNE 7 km/h
Air Quality	Unhealthy
Wind Gusts	15 km/h
Humidity	22%

11 PM 28° RealFeel® 27° 0%

Clear

Wind	SSE 9 km/h
Air Quality	Unhealthy

https://www.accuweather.com/en/in/deoghar/187228/hourly-weather-forecast/187228 4/7

4/5/25, 9:43 AM Deoghar, Jharkhand, India Hourly Weather | AccuWeather

Deoghar, Jharkhand 33°C Address, City or Zip Code

Max UV Index	11 Extreme

1 PM 37° RealFeel® 40° 0%

Hazy sunshine

RealFeel Shade™	36°
Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	9 Very High

2 PM 38° RealFeel® 40° 0%

Hazy sunshine

RealFeel Shade™	36°
Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	7 High

3 PM 37° RealFeel® 38° 0%

Hazy sunshine

RealFeel Shade™	36°
Wind	W 15 km/h
Air Quality	Unhealthy
Max UV Index	4 Moderate

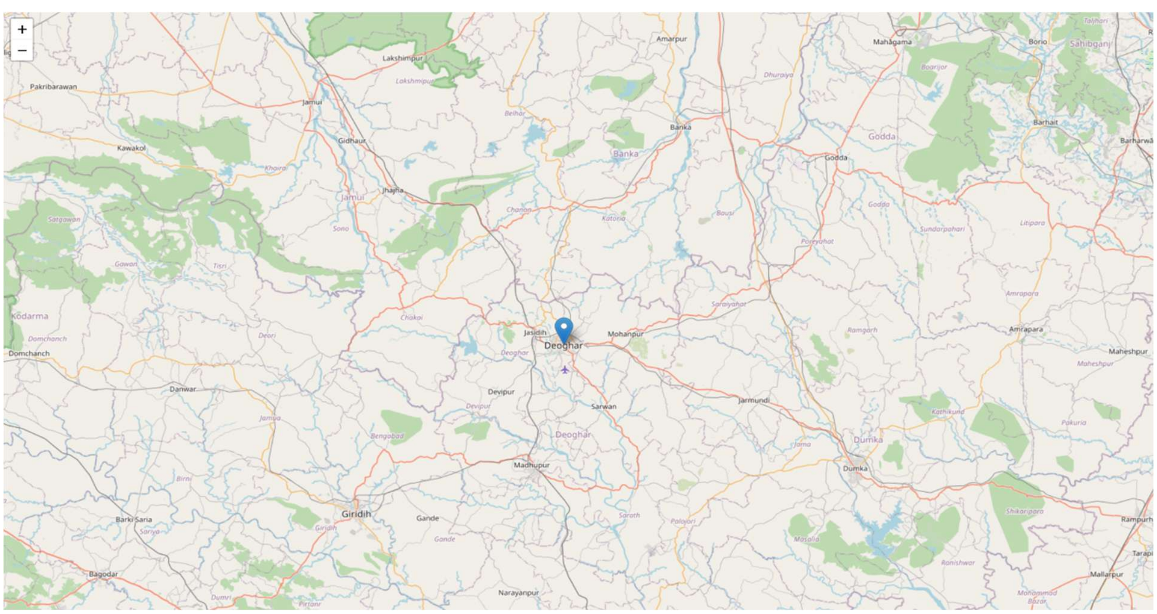
4 PM 37° RealFeel® 37° 0%

Hazy sunshine

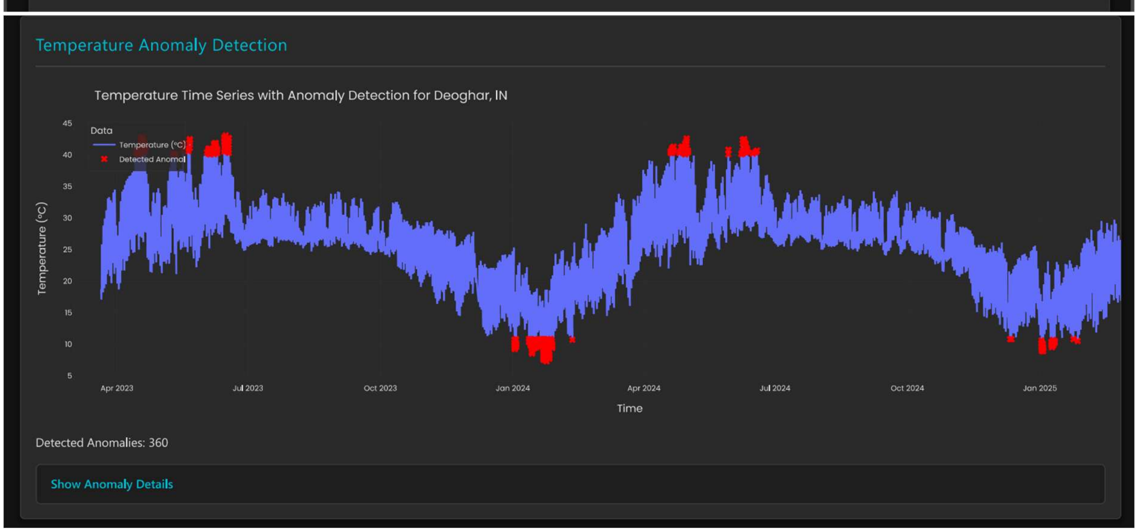
https://www.accuweather.com/en/in/deoghar/187228/hourly-weather-forecast/187228 2/7

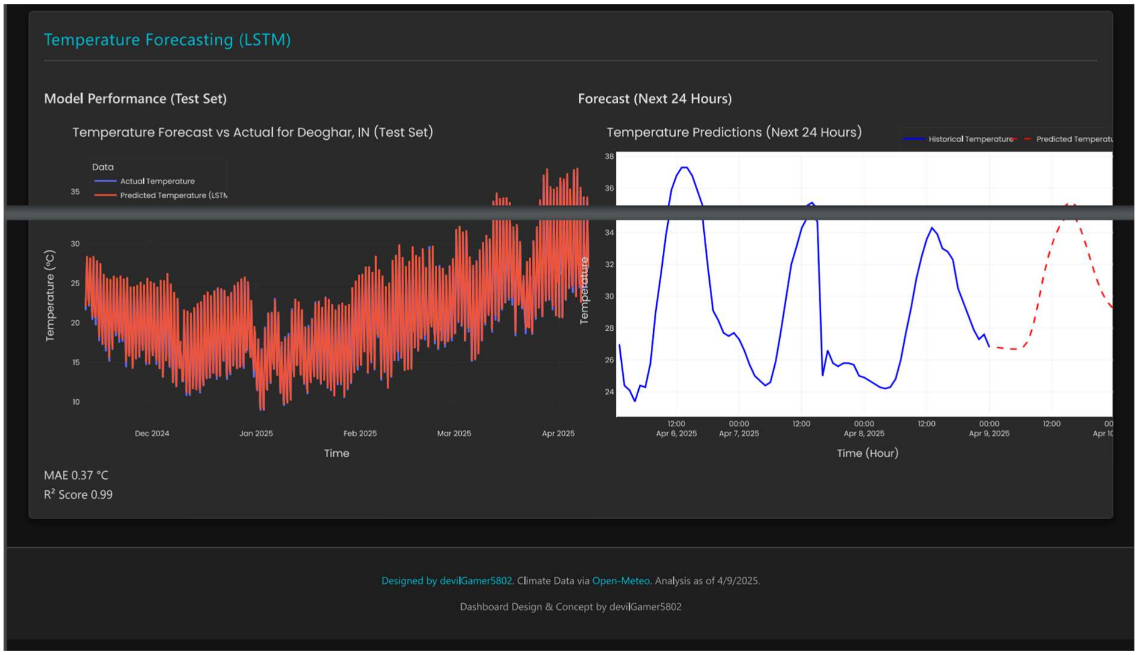
Appendix C:

LOCATION MAP



WEBSITE OVERVIEW




Temperature Forecasting (LSTM)
Model Performance (Test Set)
Forecast (Next 24 Hours)
Temperature Forecast vs Actual for Deoghar, IN (Test Set)
Data
Actual Temperature
Temperature (°C)
Time
Temperature Predictions (Next 24 Hours)
Historical Temperature
Temperature
Time (Hour)
MAE 0.37 °C
R² Score 0.99
Designed by devilGamer5802. Climate Data via Open-Meteo. Analysis as of 4/9/2025.
Dashboard Design & Concept by devilGamer5802

REFERENCES

- Intergovernmental Panel on Climate Change (IPCC)
- NOAA National Centres for Environmental Information (NCEI)
- Open-Meteo.com API
- IBM – Machine Learning with Python
- Google Colab Notebook & Jupyter Notebook
- FLASK docs, PLOTLY docs, FOLIUM docs
- Generative AI for Everyone – DeepLearning AI