Forecasting Daily Air Temperature in Doho, Uganda Using Machine Learning and Deep Learning

ACCURATE TEMPERATURE PREDICTIONS FOR AGRICULTURE, ENERGY, AND DISASTER PREPAREDNESS BY GROUP 6

Problem Statement & Justification

Problem:

Predict daily air temperature using historical weather data (2013–2016) in Doho, Uganda.

Accurate forecasts help farmers, energy providers, disaster management authorities, and urban planners.

Justification:

Weather data is sequential → temporal models like LSTM capture trends.

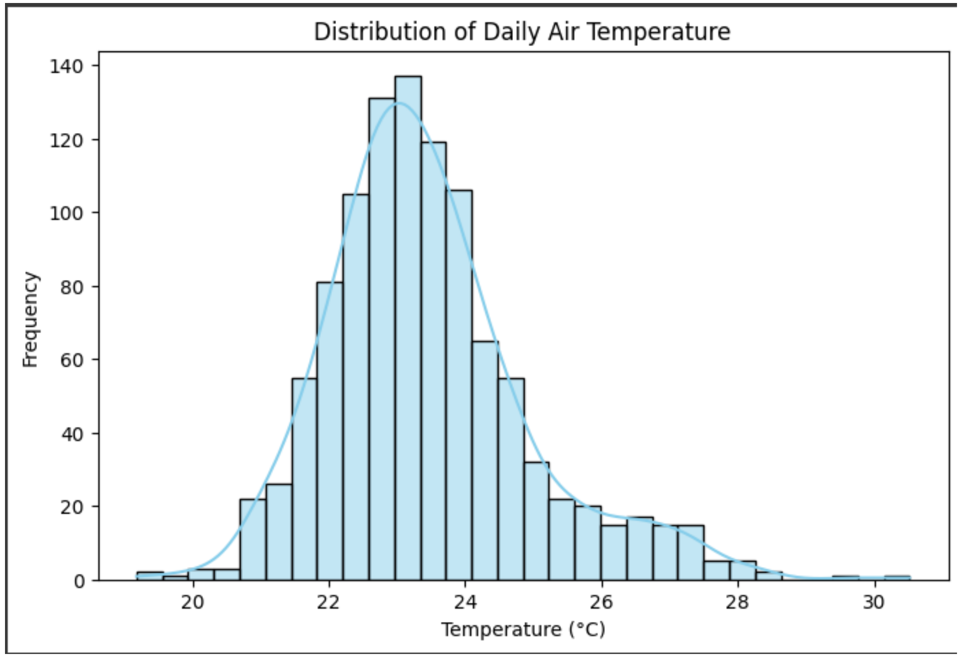
Feature engineering (lags, rolling stats, date-based features) improves model performance.

Supports data-driven decisions in agriculture, energy, and disaster preparedness.

Data & Feature Engineering

Dataset:

Daily observations from multiple stations (Temperature, Humidity, Min/Max Temp, Station ID).



Feature Engineering:

Lag variables: 1,2,3,7,14 days

Rolling statistics: 7-day mean & std

Date features: Month, Day-of-Year, Weekday

One-hot encoded station identifiers

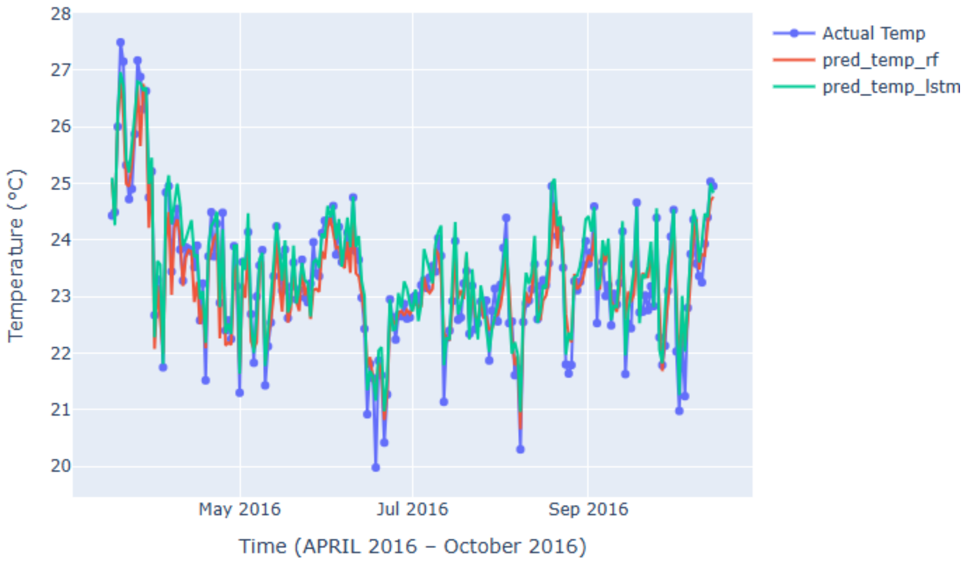
Modeling Approach & Metrics

Models Used:

Random Forest Regressor (RF) – classical ML

LSTM – deep learning for sequential data

Temperature: Actual vs Predicted



Evaluation Metrics:

Metric	Purpose
MSE	Penalizes large errors
MAE	Average absolute error
RMSE	Same units as temperature
R^2	Variance explained

Key Results & Interpretations

Model Performance:

Model	MSE	MAE	RMSE	R^2
Random Forest	0.176	0.332	0.419	0.872
LSTM	0.139	0.301	0.374	0.898

Observations:

LSTM outperforms RF across all metrics.

Captures temporal dependencies better → smoother, accurate forecasts.

Residuals centered around 0 → unbiased predictions.

Recommendations & Future Work

Recommendations for Stakeholders:

Farmers: Optimize irrigation, planting, harvesting.

Energy Providers: Plan load demand & infrastructure maintenance.

Disaster Management: Early warnings for heat/cold events.

Urban Planners & Policy Makers: Climate adaptation, urban heat planning.

Future Additions:

Include rainfall, wind speed, and humidity predictions.

Multi-station spatial modeling (Graph Neural Networks).

Ensemble modeling & real-time forecasts with dashboards/APIs.

Climate trend analysis for long-term planning.

PROJECT SUMMARY

The dataset available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UTVRUD

1. Project title
2. Problem statement and justification
3. Data description and derived features
4. Data cleaning
5. Data exploration and visualization (statistics, univariate, bivariate and multivariate analysis)
6. Modeling and evaluation

- Models used are random forest Regressor and long short term memory and their metrics (LSTM)
- Feature importance graph for random forest
- Temperature actual vs predicted graph
- Prediction error & anomaly detection graph
- Residual error distribution graph
- Metrics overview (**MSE,MAE,RMSE**, **R^2**)
- Recommendations
- future work