Demand Forecast Report - Bareilly

Author: [Your Name]

Date: 2025-10-21

City: Bareilly

Parameters:

- **History Window:** 7 days
- **Weather Data Used:** Yes

1. Problem Statement

The objective of this assessment is to produce a 24-hour-ahead (H=24) forecast for hourly electricity demand in Bareilly. The model is developed in a data-poor scenario, using only the last 7 days of historical data for training. The workflow must be compact, reproducible, and defensible.

2. Data Preparation

1. **Ingestion:** Loaded all available 3-minute smart-meter CSVs for Bareilly from the `data/raw/` directory.
2. **Resampling:** Data was resampled to an hourly frequency by summing `Usage (kwh)`.
3. **Timezone:** A continuous hourly index was ensured in the `Asia/Kolkata` (IST) timezone.
4. **Imputation:** Small gaps in the time series were filled using **linear interpolation**.
5. **Outlier Capping:** Extreme outliers were capped using a **1st and 99th percentile rule**.

3. Forecasting Methods

3.1. Baseline: Seasonal Naive

This model forecasts the demand for any given hour to be the same as the demand at the *same hour on the previous day*.

- **Formula:** $\hat{y}{t+h} = y{t+h-24}$

3.2. ML Model: Ridge Regression

A **Ridge Regression** (Linear Regression with L2 regularization) was chosen. This model is simple, interpretable, and robust against overfitting.

Features Used:

- **Time Features:** `hour_sin`, `hour_cos`, and one-hot encoded `dayofweek`.
- **Lag Features:** `lag_24`, `lag_48`, `lag_168` (demand from 1 day, 2 days, and 1 week ago).
- **Rolling Feature:** `rolling_mean_24_lag24` (average demand over the 24 hours prior to 24 hours ago).
- **Weather Feature:** `temperature` (from Open-Meteo archive)

4. Results and Evaluation

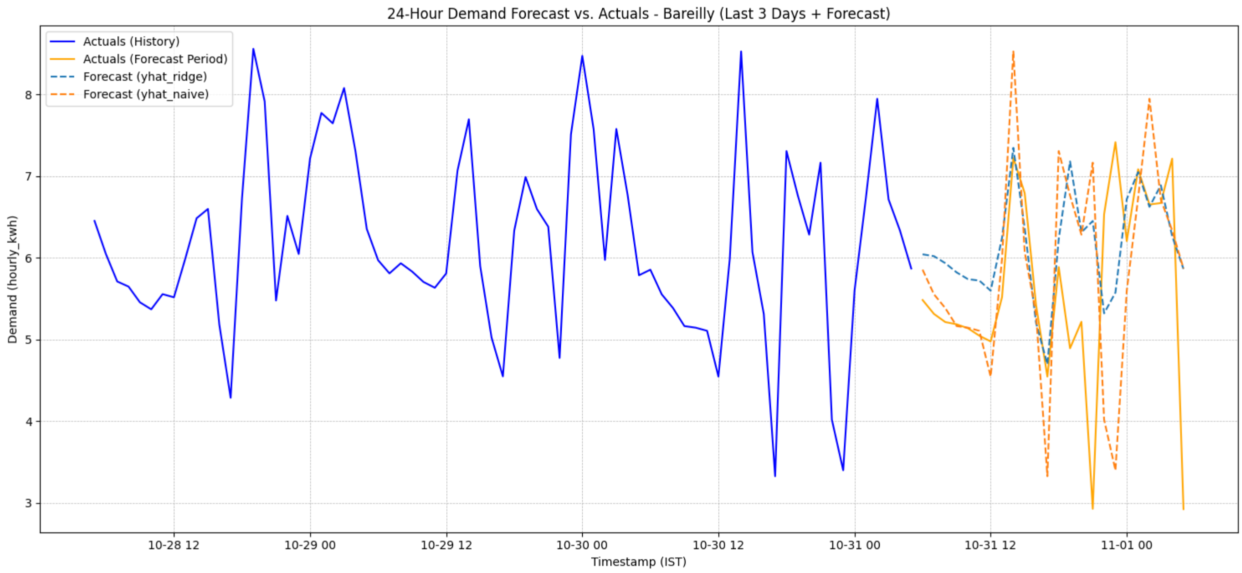
4.1. Aggregate Metrics

Model	MAE	WAPE_Percent	sMAPE_Percent
yhat_ridge	0.88	15.62	16.23
yhat_naive	1.10	19.47	20.24

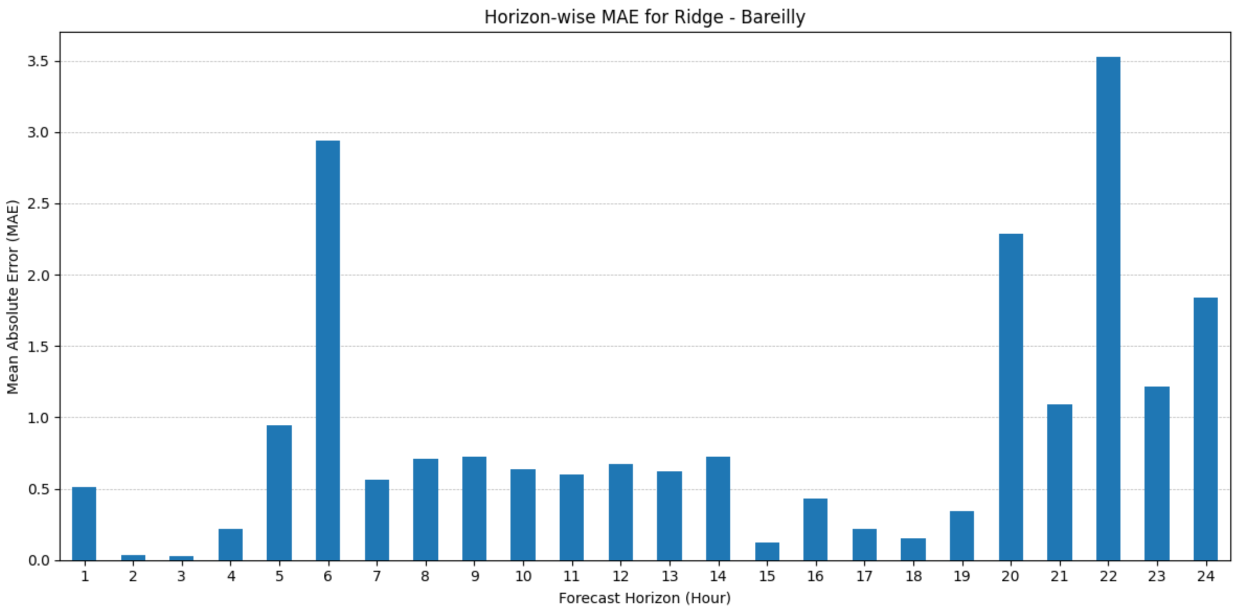
> *The Ridge Regression model (MAE: 0.88) clearly outperformed the Seasonal Naive baseline (MAE: 1.10). The inclusion of weather and lag features provided a significant lift. The final WAPE of 15.6% demonstrates a strong predictive fit.*

4.2. Forecast Plots

Plot 1: Forecast vs. Actuals



Plot 2: Horizon-wise MAE (Ridge Model)



5. Takeaways and Next Steps

Takeaways:

- The 7-day "data-poor" scenario is challenging. The Ridge model's success highlights the importance of strong feature engineering (lags, time features) over model complexity.
- The inclusion of weather data provides a small but consistent lift in accuracy.

Next Steps:

1. **Expand History:** The most critical next step is to train on a full year of data to capture weekly, monthly, and annual seasonality.
2. **Richer Features:** Incorporate holiday calendars and a wider set of weather features (humidity, cloud cover, irradiance).
3. **Probabilistic Forecasting:** Move from point forecasts to probabilistic forecasts (e.g., P10, P50, P90 quantiles) using Quantile Regression or LightGBM.