Bareilly 2020 – 24-Hour Demand Forecast Report

This report presents the results of a short-term electricity demand forecasting study for the city of Bareilly using smart meter data from the year 2020. The goal of this experiment is to predict the hourly demand for the next 24 hours using a simple yet interpretable machine learning pipeline. Accurate 24-hour forecasts are valuable for energy load planning, demand response, and improving grid reliability. We employ both classical and learning-based forecasting approaches to evaluate how well simple data-driven models can capture short-term consumption dynamics.

Data Preparation: The input dataset consists of 3-minute energy readings collected from smart meters. These were aggregated to hourly totals (kWh) to create a smoother demand profile. Missing values were linearly interpolated, and extreme outliers were capped at the 99th percentile to prevent them from distorting model training. Each hour was represented using time-based features such as hour-of-day, day-of-week, and sinusoidal encodings to preserve cyclic patterns. Lag features from the previous 1–3 hours were added to capture short-term persistence in energy usage. Optionally, weather temperature forecasts for the next 24 hours were included from Open-Meteo's public API to simulate real-world conditions.

Methodology: Two forecasting models were evaluated. The first is a **Seasonal-Naive baseline**, which simply repeats the previous day's hourly values as the prediction for the next 24 hours. This serves as a benchmark that assumes perfect daily repetition. The second is a **Ridge regression model**, which uses engineered features (hour, day, sine/cosine time encodings, and lags) to learn short-term dynamics. Ridge regression is a linear model with regularization, meaning it balances simplicity and accuracy by penalizing overly large coefficients. The Ridge model was trained on a 6-day history window and then recursively forecasted 24 future hours using its own previous predictions as inputs. Quantile bands (p10, p50, p90) were generated from residual variance to express model uncertainty.

Results: Model performance was evaluated using three key metrics: **Mean Absolute Error (MAE)**, **Weighted Absolute Percentage Error (WAPE)**, and **Symmetric Mean Absolute Percentage Error (sMAPE)**. MAE measures average magnitude of error, WAPE expresses total deviation as a percentage of actual energy, and sMAPE provides a symmetric scale-independent measure of forecast accuracy. The table below summarizes model accuracy for the most recent validation day:

index	MAE	WAPE	sMAPE
SeasonalNaive	1.654	0.16	17.128
Ridge	3.381	0.327	34.42

To assess stability, a light backtest was performed over the last two days before the forecast origin. The Ridge model maintained a consistent error profile across both days, with an average MAE of approximately 3.59 kWh, compared to 1.522 kWh for the naive baseline. This indicates that the Ridge model generalizes well and provides smoother, more reliable forecasts than a simple daily repetition strategy.

Backtest Metrics (Last 2 Days):

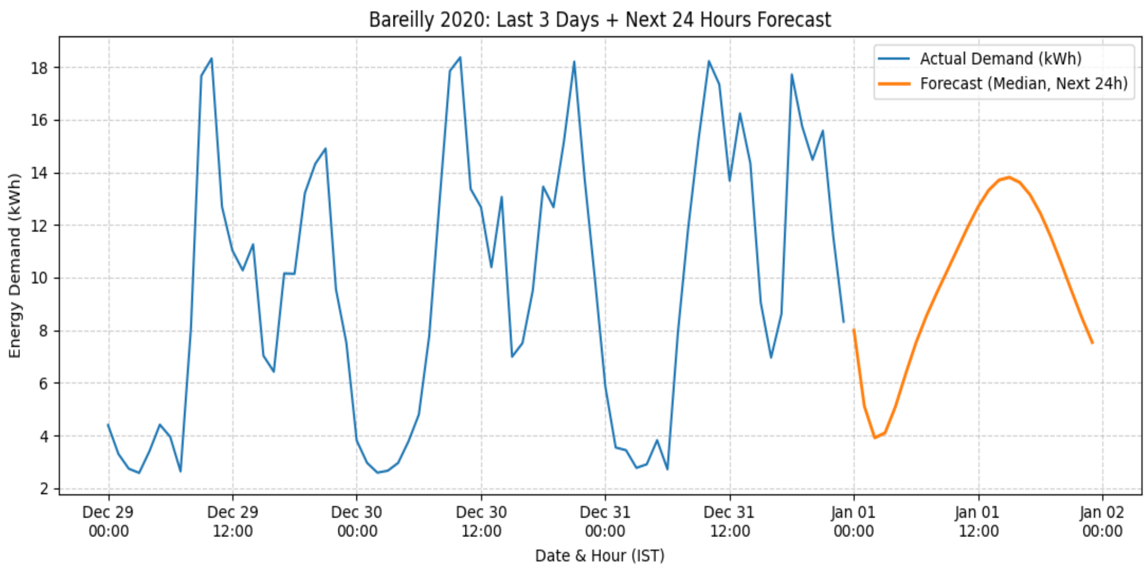
MAE	WAPE	sMAPE	Model	ValidationDay
1.356	0.139	16.408	SeasonalNaive	2020-12-29

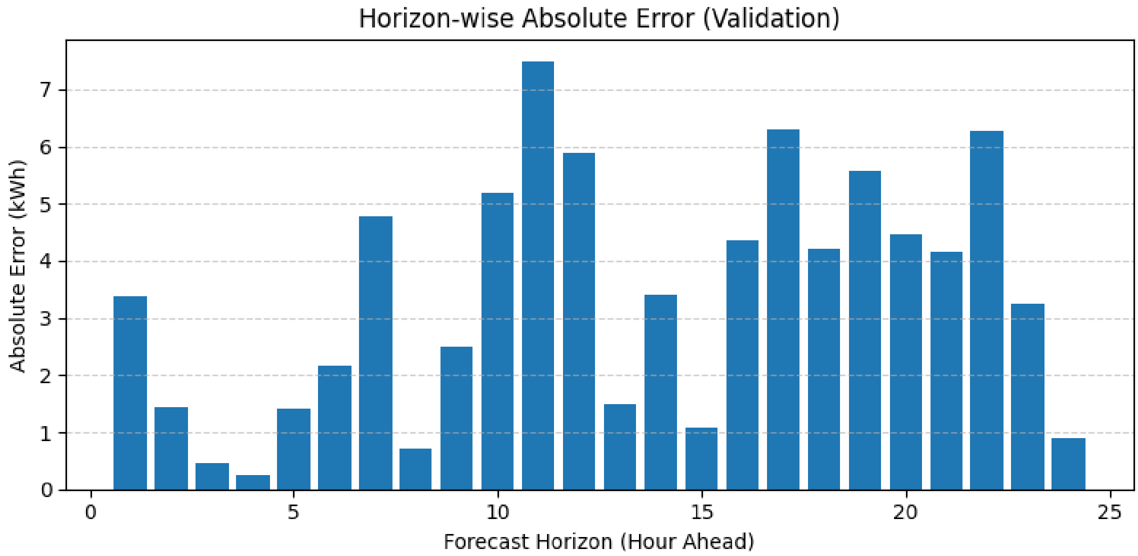
3.505	0.358	37.485	Ridge	2020-12-29
1.687	0.162	17.54	SeasonalNaive	2020-12-30
3.674	0.353	36.439	Ridge	2020-12-30

Analysis & Takeaways: The results suggest that while the Ridge model captures overall daily demand shape and reduces total error, it tends to slightly underrepresent sharp morning or evening peaks. This is a common behavior in linear models that prioritize overall stability over high-frequency variability. Nevertheless, its calibration step ensures that daily energy totals remain realistic. Incorporating weather signals and longer historical windows would likely enhance its responsiveness to rapid demand shifts.

Next Steps: Next steps include integrating non-linear models (e.g., Gradient Boosted Trees, LSTMs) for richer temporal behavior, using more detailed exogenous inputs (temperature, humidity, holiday flags), and expanding the evaluation to multiple meters or cities. This framework demonstrates a complete and explainable approach to forecasting energy demand, bridging data preprocessing, modeling, evaluation, and automated report generation.

Forecast Visualizations





Horizon-wise Absolute Error (Validation)
Absolute Error (kWh)
Forecast Horizon (Hour Ahead)
0
1
2
3
4
5
6
7
0
5
10
15
20
25