IIoT Time Series Forecasting

A06 AI-Driven Data Analytics with Nixtla

Adebanke Damola-Fashola | ITAI 3377 | March 2025

Introduction

Forecasting temperature data from IoT sensors.

Challenges: irregular timestamps, data quality.

Objective: Build a scalable forecasting pipeline.

Data Preparation & Preprocessing

Parsed timestamps and removed duplicates.

Resampled data to uniform frequency.

Visualized temperature trends.

Feature Engineering

- Extracted hour-of-day and rolling mean.
- Lag features to capture temporal dependencies.
- Improved model accuracy using exogenous variables.

Model Training & Evaluation

- Used Nixtla TimeGPT for forecasting.
- Included engineered features for training.
- Metrics: MAE of 3.72 and MSE 32.45 across 5 splits.

Generative Modeling

Built LSTM-based Variational Autoencoder (VAE).

Generated synthetic temperature data.

Combined real and synthetic data for retraining.

Impact of Synthetic Data

Improved model performance metrics.

Increased robustness to irregular patterns.

Supports better generalization in forecasting.

Key Insights

Robust Preprocessing Is Essential:
Careful timestamp parsing, duplicate aggregation, and resampling ensured clean and consistent time series data.

Feature Engineering Boosted Accuracy:
Rolling mean temperature and hour-of-day features captured both short-term and daily cyclic patterns.

Nixtla AutoML Simplified Deployment:
TimeGPT automated model selection and tuning, delivering competitive results with minimal manual effort.

Cross-Validation Ensured Robustness:
Rolling-origin cross-validation revealed performance stability over time, providing deeper temporal insights.

Generative Modeling Enhanced Dataset Diversity:
A Variational Autoencoder (VAE) generated synthetic temperature data, smoothing irregularities and improving forecast generalization.

Recommendations

Use Synthetic Data in Sparse Scenarios:
Augment datasets with VAE- or GAN-generated data to strengthen model robustness in low-volume environments.

Automate Feature Enrichment:
Develop pipelines to automatically extract features like day-of-week, seasonality, and rolling statistics.

Monitor Model Drift with Cross-Validation:
Implement rolling-origin validation regularly to detect changes in data behavior and maintain forecast accuracy.

Explore Other Generative Models:
Experiment with GANs or Diffusion Models for richer synthetic diversity, especially for multivariate time series.