Time Series Forecasting Using ARIMA: Detailed Project Report

Project Objectives

This project focuses on time series forecasting of Italian electricity load (ITloadnew) and solar generation (ITsolargeneration) using hourly data from 2016. The objective is to build a reliable ARIMA-based forecasting model, validate its performance using an 80/20 train-test split, and evaluate accuracy using RMSE.

Step 1: Data Loading and Initial Exploration

The dataset was loaded using pandas and contains UTC timestamps, electricity load, and solar generation values. Timestamps were converted to datetime format for time-based analysis. Initial visualizations revealed clear daily cycles in load and solar generation patterns.

Step 2: Missing Value Handling

ITloadnew contained 72 missing values, while solar generation had none. Forward-fill imputation was applied to preserve temporal continuity. Post-imputation checks confirmed no remaining missing values.

Step 3: Stationarity Testing

Augmented Dickey-Fuller (ADF) tests confirmed stationarity for both ITloadnew and ITsolargeneration (p-values < 0.05), indicating no differencing was required ($d = 0$).

Step 4: ACF/PACF Analysis

ACF and PACF plots were analyzed to determine model parameters. PACF cutoff after lag 2 and gradual ACF decay supported the selection of an ARIMA(2,0,2) model for load forecasting.

Step 5: Train-Test Split

The dataset was split into 80% training and 20% testing sets to evaluate performance on unseen data.

Step 6: Model Fitting

An ARIMA(2,0,2) model was fitted on the training data. The model converged successfully without requiring differencing due to stationarity.

Step 7: Forecasting and Evaluation

Forecasts were generated for the test period. Model performance was evaluated using RMSE, which was approximately 7715 MW, indicating strong predictive accuracy relative to the load scale.

Key Results

ADF p-value (Load): 3.84e-22
ADF p-value (Solar): 6.27e-07
ARIMA Order: (2,0,2)
Test RMSE: ~7715 MW

Business Impact

The developed ARIMA model provides reliable short-term electricity load forecasts, supporting grid planning, renewable energy integration, and cost optimization. The approach is scalable for real-time operational use.