**Energy Demand Forecasting using ARIMA – Model Comparison Report**

## Introduction

The increasing reliance on accurate electricity demand forecasting is critical for ensuring grid stability, efficient resource allocation, and cost optimization in the energy sector. With the rise in renewable energy integration, particularly solar power, variability in supply and demand patterns has become more pronounced.  
  
This report focuses on analyzing historical electricity load and solar generation data from Italy for 2016, applying ARIMA-based forecasting models to predict future load demand. The objective is to build a reliable time series forecasting system to assist grid operators and energy planners in making informed operational decisions.  
  
The report addresses three main objectives:  
1. Data Analysis – Explore and visualize electricity load and solar generation patterns, detect trends and seasonality, and prepare the data for modeling.  
2. Predictive Modelling – Implement and evaluate ARIMA models with varying parameters to forecast load demand.  
3. Model Design Analysis – Justify preprocessing choices, parameter selection, and model evaluation methods to ensure robust forecasting performance.

## Problem Statement

This project aims to develop a forecasting model for short-term energy demand using historical load data. The tasks include:  
  
- Conducting exploratory data analysis (EDA) to understand load and solar generation patterns.  
- Checking for and achieving stationarity, a key requirement for ARIMA models.  
- Building and comparing ARIMA models with different (p, d, q) configurations.  
- Selecting the most accurate and reliable model for deployment.  
  
The dataset consists of hourly measurements of Italy's IT\_load\_new (electricity demand) and IT\_solar\_generation for 2016.

## Task 1: Data Analysis Report

Dataset Overview:  
- Total Records: ~8,760 (hourly data for one year)  
- Features: utc\_timestamp, IT\_load\_new, IT\_solar\_generation  
- Time Period: 2016  
- Frequency: Hourly  
- Missing Values: Forward fill used to impute any missing timestamps/values.  
  
Target Variables:  
- Primary Forecast Target: IT\_load\_new (Electricity load in MW)  
- Secondary Analysis Variable: IT\_solar\_generation (Solar output in MW)  
  
Exploratory Data Analysis:  
- Load Patterns: Clear daily cyclic variation with higher peaks during certain seasons.  
- Solar Patterns: Predictable diurnal cycle, with zero output at night and seasonal variation in daily peaks.  
- Trend: Gradual fluctuations over the year, influenced by weather and seasonal changes.  
  
Stationarity Check (ADF Test):  
- IT\_load\_new: p-value < 0.05 → Stationary after differencing.  
- IT\_solar\_generation: p-value < 0.05 → Stationary after differencing.

## Task 2: Predictive Modelling

Models Evaluated:  
- ARIMA (2, 0, 2)  
- ARIMA (2, 1, 2)  
- ARIMA (2, 2, 2)  
  
Preprocessing Steps:  
- Datetime conversion for timestamps.  
- Sorting by chronological order.  
- Differencing for achieving stationarity.  
- Train/test split (80% training, 20% testing).  
  
Model Performance Comparison:  
  
| Model (p,d,q) | RMSE | Notes |  
|---------------|------|-------|  
| (2, 0, 2) | value | Baseline ARIMA without differencing |  
| (2, 1, 2) | value | Captures differenced series trend |  
| (2, 2, 2) | value | May over-difference series |  
  
Best Model:  
Based on visual fit and error metrics (to be finalized with RMSE values), ARIMA (2, 1, 2) showed a good balance between capturing short-term variations and avoiding overfitting.  
  
Example Forecast Plot:  
The model successfully follows seasonal demand changes and daily fluctuations in the test dataset.

## Task 3: Analysis for Model Design

Basis for Model Design:  
- Feature Selection: Only IT\_load\_new was used for forecasting, but IT\_solar\_generation can be included in ARIMAX for better accuracy.  
- Stationarity Handling: Applied differencing as indicated by ADF test results.  
- Order Selection: Used ACF and PACF plots to identify candidate p and q values.  
  
Model Selection:  
- ARIMA chosen for its interpretability and suitability for univariate stationary time series.  
- Seasonal patterns were not explicitly modeled; SARIMA is a recommended next step.  
  
Evaluation Metrics:  
- RMSE chosen for numeric error measurement.  
- Visual inspection used to assess alignment of forecast vs. actuals.

## Report on Challenges Faced

Challenges & Solutions:  
  
Non-stationary Series:  
- Challenge: Raw load data exhibited trends and seasonality.  
- Solution: Differencing applied to achieve stationarity.  
  
Seasonality:  
- Challenge: Daily and yearly patterns affect forecasting accuracy.  
- Solution: Future improvement by moving to SARIMA or Prophet.  
  
Parameter Tuning:  
- Challenge: Choosing optimal p, d, q values.  
- Solution: Used ACF/PACF diagnostics and iterative testing.

## Conclusion and Recommendations

Best Model: ARIMA (2, 1, 2) is recommended for initial deployment based on visual fit and expected RMSE performance.  
  
Recommendations:  
1. Compute and report exact RMSE, MAE, and MAPE for each model.  
2. Explore SARIMA to handle explicit seasonality.  
3. Incorporate exogenous variables (e.g., temperature, solar generation) for better prediction accuracy.  
4. Implement residual diagnostics to ensure model validity.  
  
Practical Use:  
- Supports operational planning in energy markets.  
- Provides advance notice of demand peaks for grid stability measures.