3803ICT
Big Data Analysis

Lab 04 – Data Analytics for Timeseries

Trimester 1 - 2022

Table of Contents

I. Basic Time-Series Analyses

Given the following historical data of exact sales numbers:

Month	Sales
1	5384
2	8081
3	10282
4	9156
5	6118
6	9139
7	12460
8	10717
9	7825
10	9693
11	15177
12	10990

1. **Visualize and interpret the pattern of this time-series**

2. **Predict future demand in month 13,14,15,16 without seasonality**

   ❖ Using moving average, exponential smoothing.

3. **Predict future demand in month 13,14,15,16 with seasonality**

   Divide the time-series into 3 cycles (Month 1-4, 5-8, 9-12), then
   ❖ Compute the average sales for each cycle
   ❖ Compute the seasonal indices for each month of each cycle
   ❖ Compute the seasonal indices for next cycle
   ❖ Use linear regression on the average sales of cycles to predict the average sale for next cycle
   ❖ Compute the seasonalize forecasts

4. **Evaluation: compare the above implemented methods**

   ❖ Compute forecast errors
   ❖ Make conclusions

II. Time-Series Analyses for Electricity Consumption

1. Electricity Consumption Exploration

Use the provided Jupyter notebook.

In this exercise, you will be asked to explore time series data in several ways.
You will be working with time series of Open Power System Data (OPSD) for Germany.

The data set includes daily electricity consumption, wind power production, and solar power production between 2006 and 2017.

- Date — The date (yyyy-mm-dd format)
- Consumption — Electricity consumption in GWh
- Wind — Wind power production in GWh
- Solar — Solar power production in GWh
- Wind+Solar — Sum of wind and solar power production in GWh

Tasks:
- Plotting and understanding a time series
- Change the scale
- Further exploration
- Fill missing values
- Series correlation

2. Electricity Consumption Prediction

Use the provided Jupyter notebook.

In this exercise, you will be asked to use insights from the previous exploration, and build forecasts for the electricity consumption.

You will be working with time series of Open Power System Data (OPSD) for Germany.

The data set includes daily electricity consumption, wind power production, and solar power production between 2006 and 2017.

- Date — The date (yyyy-mm-dd format)
- Consumption — Electricity consumption in GWh
- Wind — Wind power production in GWh
- Solar — Solar power production in GWh
- Wind+Solar — Sum of wind and solar power production in GWh

Tasks:
- Key concepts in time series
- Trends
- Seasonality
- Automatic decomposition
- MA process
- AR process
- ARMA process
- ARIMA process
- SARIMA process (OPTIONAL)