

Times series forecasting: Statistical models Stationary time series

Forecasting bandwidth usage allows data centers to better manage their computing resources. In the case where less bandwidth usage is expected, they can shut down some of their computing resources. This in turns reduces expenses and allows for maintenance. On the other hand, if bandwidth usage is expected to increase, they can dedicate the required resources to sustain the demand and ensure low latency, thus keeping their customers satisfied.

The goal is to develop a model to predict bandwidth usage for a large data center. Bandwidth is defined as the maximum rate of data that can be transferred. Its base unit is bits per second (bps). For this situation, there are 10,000 data points representing the hourly bandwidth usage starting in January 1, 2019. Here the bandwidth is measured in megabits per second (Mbps), which is equivalent to 106 bps

- 1. Start by loading the dataset "bandwidth.csv".
- 2. Visualize the time series.
- **3.** Verify if the time series is stationary.
- **4.** Based on your exploration, determine the most suitable model with the usual steps:
 - **a.** Apply the transformations to make the time series stationary
 - **b.** Determine the adequate model to predict the series by plotting the autocorrelation and partial autocorrelation functions
 - **c.** Find the optimal (**p** , **q**) parameters.
 - d. Fit the ARMA(p,q) model with the optimal parameters on the training set
 - **e.** Perform the residual analysis to validate if the model can be used to forecast.
- 5. Predict a full week of the bandwidth usage using the last week of points as test set.
- **6.** Compare the performance of the ARMA model with two baseline model: the mean method and the last known value method.
- **7.** Visualize all the forecasts in a single graph.
- 8. Compare the performance of all three models using the mean squared error metric.
- **9.** Reverse the forecast in order to bring the values to the same scale as the original data, by applying a cumulative sum.
- 10. Plot the forecasts on the original scale of the data with the actual values of the test set.