Kelas Artificial Intelligent - A
Kel. 6 (urutan ke 5)
Tanggal presentasi 3 Des 2018

1. Setiadi
2. Achmad Rafiqi Azka
3. Arga Bagus Priambodo

--

Guide ~ Predicting Process

1. Import library
   ```
   import warnings
   import itertools
   import numpy as np
   import matplotlib.pyplot as plt
   import pandas as pd
   import statsmodels.api as sm
   import matplotlib
   …
   ```
2. Load data from database
3. Sum sales and grouping by *order date*
4. Indexing
5. Resampling
6. Visualizing

   Plotting data into graph with *matplotlib* to analyze data sales visually.

   Some distinguishable patterns appear when we plot the data. The time-series has seasonality pattern, such as sales are always low at the beginning of the year and high at the end of the year. There is always a strong upward trend within any single year with a couple of low months in the mid of the year.

   We can also visualize our data using a method called time-series decomposition that allows us to decompose our time series into three distinct components: ***observed***, ***trend, seasonality, and residual***.

7. Time series forecasting with ARIMA

   We are going to apply one of the most commonly used method for time-series forecasting, known as ARIMA, which stands for Autoregressive Integrated Moving Average.

   Seasonal ARIMA models are usually denoted ARIMA*(p,d,q)(P,D,Q)m*, where *m* refers to the number of periods in each season, and the uppercase *P,D,Q* refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model.

   Ref.
   https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
   https://en.wikipedia.org/wiki/Box%E2%80%93Jenkins_method

   - Parameter Selection for the ARIMA Time Series Model

     a. When two out of the three terms are zeros, the model may be referred to based on the non-zero parameter, dropping "AR", "I" or "MA" from the acronym describing the model. For example, ARIMA (1,0,0) is AR(1), ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1).

     b. The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.
     https://en.wikipedia.org/wiki/Akaike_information_criterion

8. Validating forecast (MSE, RMSE)

   To help us understand the accuracy of our forecasts, we compare predicted sales to real sales of the time series with MSE, RMSE.

   - In statistics, the mean squared error (MSE) of an estimator measures the average of the squares of the errors — that is, the average squared difference between the estimated values and what is estimated. The MSE is a measure of the quality of an estimator—it is always non-negative, and the smaller the MSE, the closer we are to finding the line of best fit.

   - Root Mean Square Error (RMSE) tells us that our model was able to forecast the average daily furniture sales in the test set within 151.64 of the real sales. Our furniture daily sales range from around 400 to over 1200. In my opinion, this is a pretty good model so far.

9. Producing and visualizing forecasts

   Our model clearly captured furniture sales seasonality. As we forecast further out into the future, it is natural for us to become less confident in our values. This is reflected by the confidence intervals generated by our model, which grow larger as we move further out into the future.