07/04/2020

Executive Summary

Objective:

Time series Analysis and Forecasting: To design the model which can produce reliable predictions by analyzing and processing the time series data.

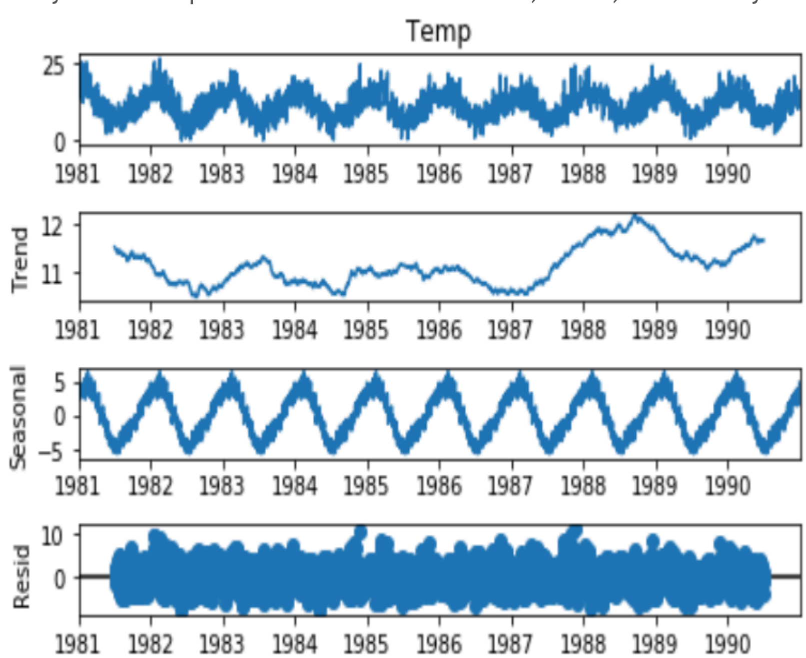
Problem Statement:

To predict the minimum temperature of one future period.

Problem Analysis:

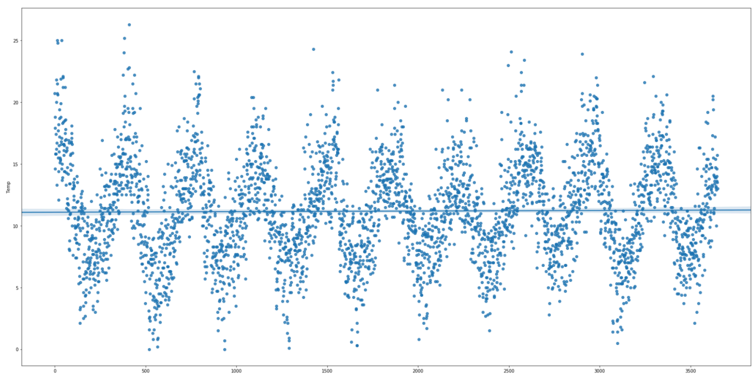
1. Dataset:
   Dataset used for addressing this problem is the minimum temperature of each day over 10 years of period.
   Dataset consists of two columns Date and Minimum Temperature.

2. Time Series Analysis:
   Analysis of Temperature values distribution, trend, seasonality.



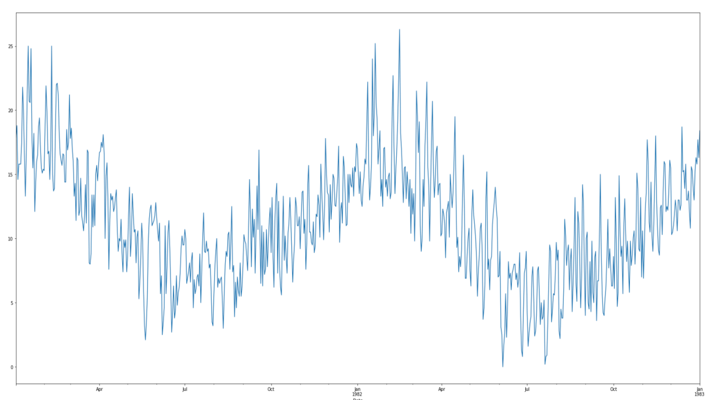
Trend:

Data shows the linear trend, slightly increasing over time.



Seasonality:

Data shows yearly seasonality as each year recurring similar pattern can be clearly seen. In the graph below, two years data is plotted. Both years, in months of June July lowest minimum temperature is recorded whereas in March and April highest value of minimum temperature is recorded which is summer season.



3. Model Building and analysis:

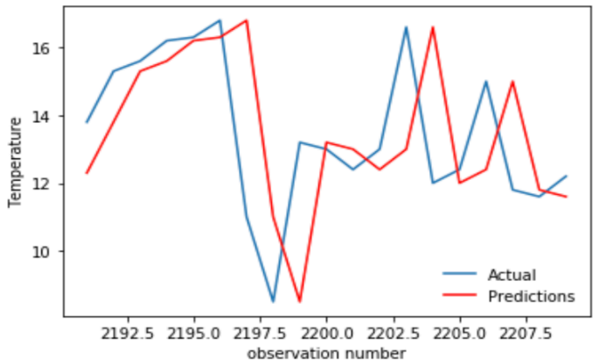
   To find the best model, several models are built and analyzed.

   a. Naïve Model:

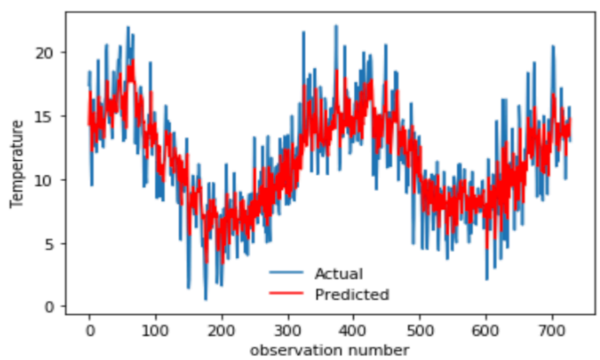
      Naïve Model considers today's temperature value as the forecasted value for tomorrow's temperature, which is naïve approach of forecasting.

      This model is considered as benchmark for evaluating the performance of several models.

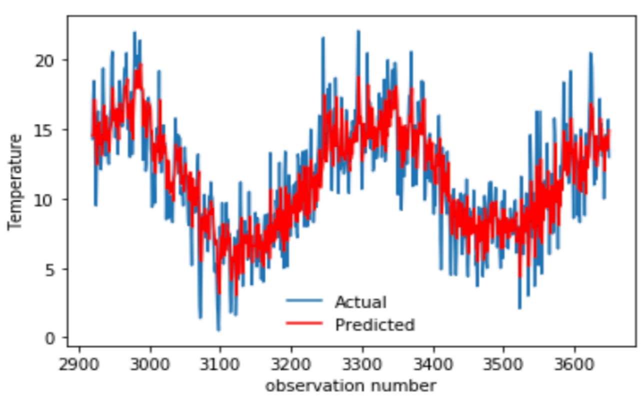
      Evaluation metric: Mean Absolute Error = 2.025



b. Autoregression Model:
   This model works very well on this dataset because dataset has linear trend.
   Evaluation Metric: Mean Absolute Error= 1.74

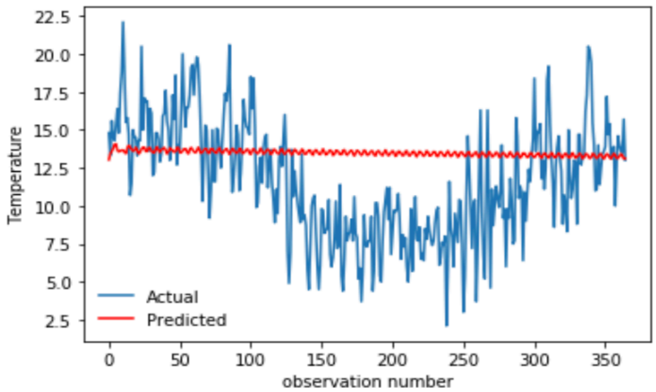


c. Moving Averages Model:
   Moving average model adds the predictions from naïve model and predictions of autoregression preformed on residuals of naïve model.
   This model performs very similar to autoregression model.
   Evaluation metric: Mean Absolute Error= 1.73

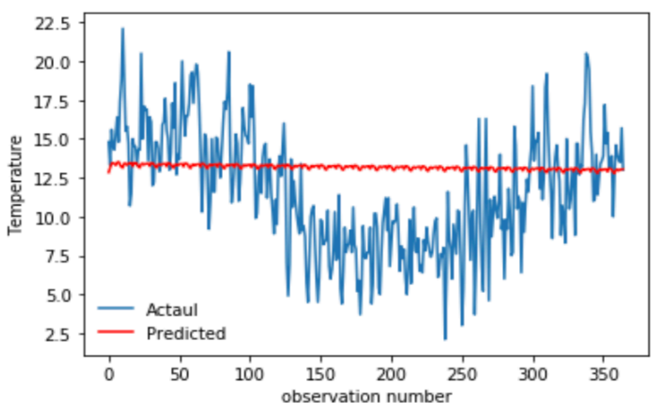


d. ARIMA (Autoregressive Integrated Moving Average)
   ARIMA does not perform very well on this dataset due to high seasonality in data. ARIMA does not deal with seasonality.

Evaluation Metric:
Mean Absolute Error= 3.46



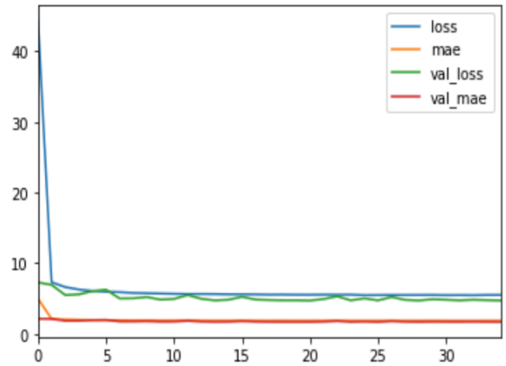
e. SARIMA (Seasonal Autoregressive Integrated Moving Average)
SARIMA, as well, does not perform very well as ARIMA though it handles seasonality.
Evaluation Matric:
Mean Absolute Error=3.38

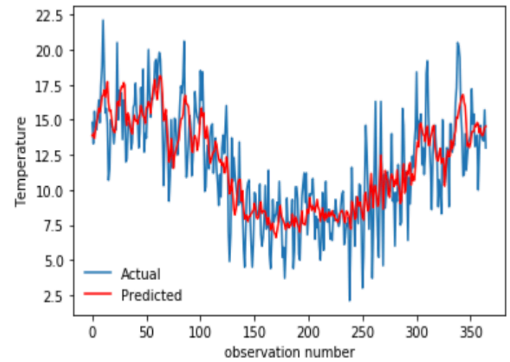


f. Artificial Neural Network
This model performs quite well.
New Seven features of type date-time features, lag features and window features are created using feature engineering.
Neural network model built has following architecture:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 30)                270
_________________________________________________________________
dense_1 (Dense)              (None, 30)                930
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 31
=================================================================
Total params: 1,231
Trainable params: 1,231
Non-trainable params: 0
```

Performance Evaluation: validation mean absolute error= 1.70





Result:

Final best preforming model can be Neural Network model. Additionally, Autoregression performs very well due to linear and steady trend in data.

Conclusion:

After analyzing the time series data, it is found that data shows yearly seasonality and linear steady trend.

On analysis of several models autoregression model and neural network model perform really well with approximate MAE equal to 1.5