

Beijing Air Pollution Time Series

By: Michael Rozenvasser

DATA

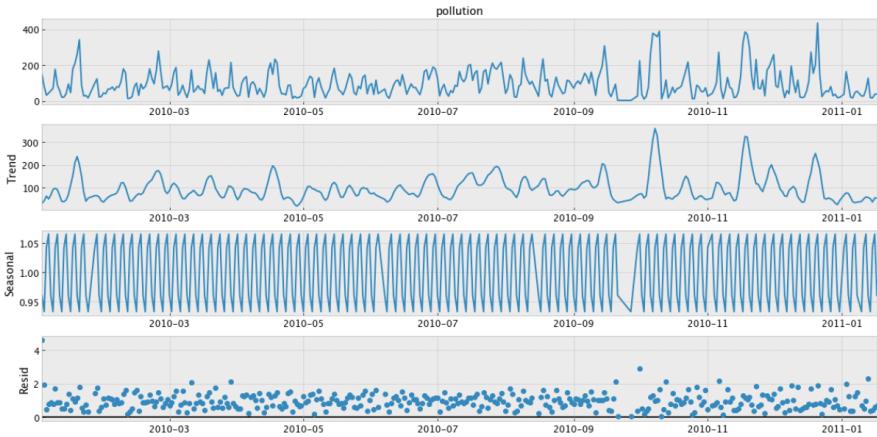
Beijing Air Quality Dataset

- **2010-2015**
- **About 1500 rows, 9 columns**
- **Changed column names when importing as well as turned hourly measurements of features into averages**
- **Dropped categorical feature named wind_dr**

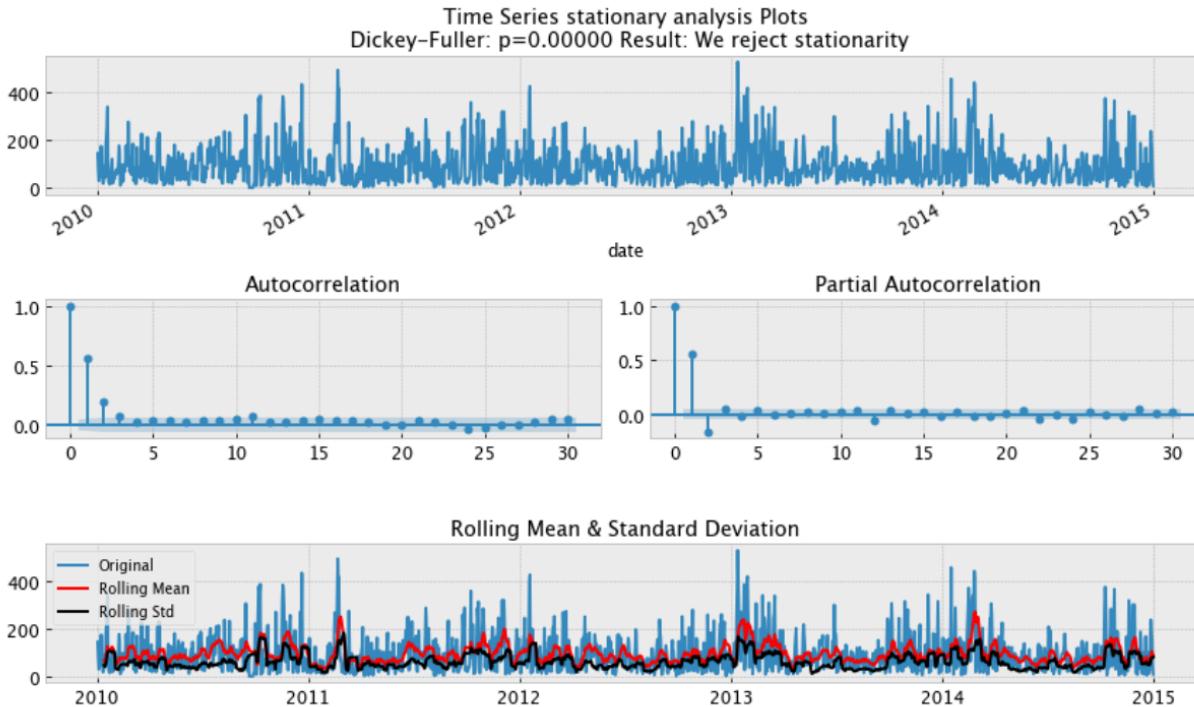
EDA

Focused on:

- **Level**
- **Trend**
- **Seasonality**
- **Noise Components**



EDA

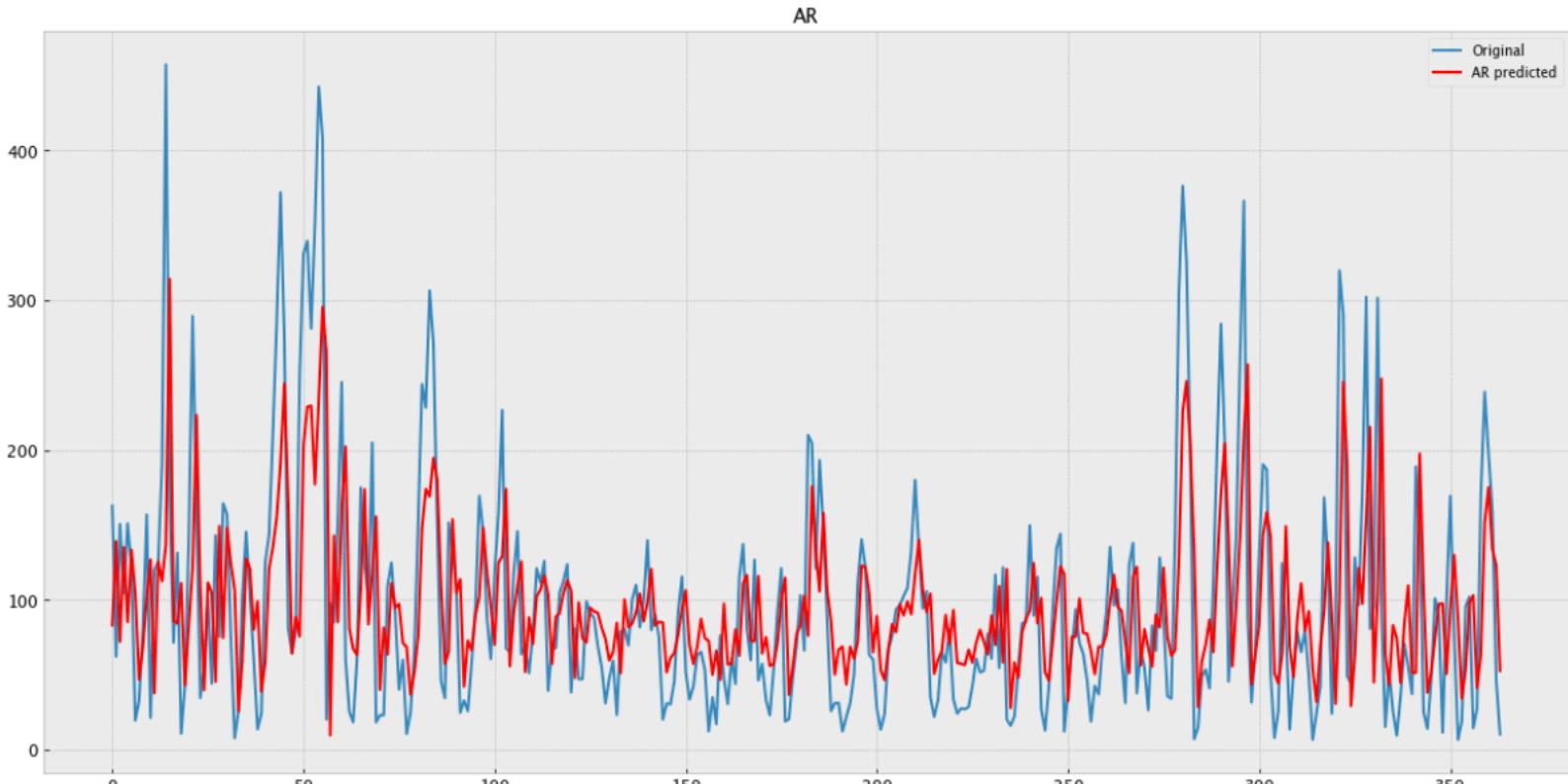


Results

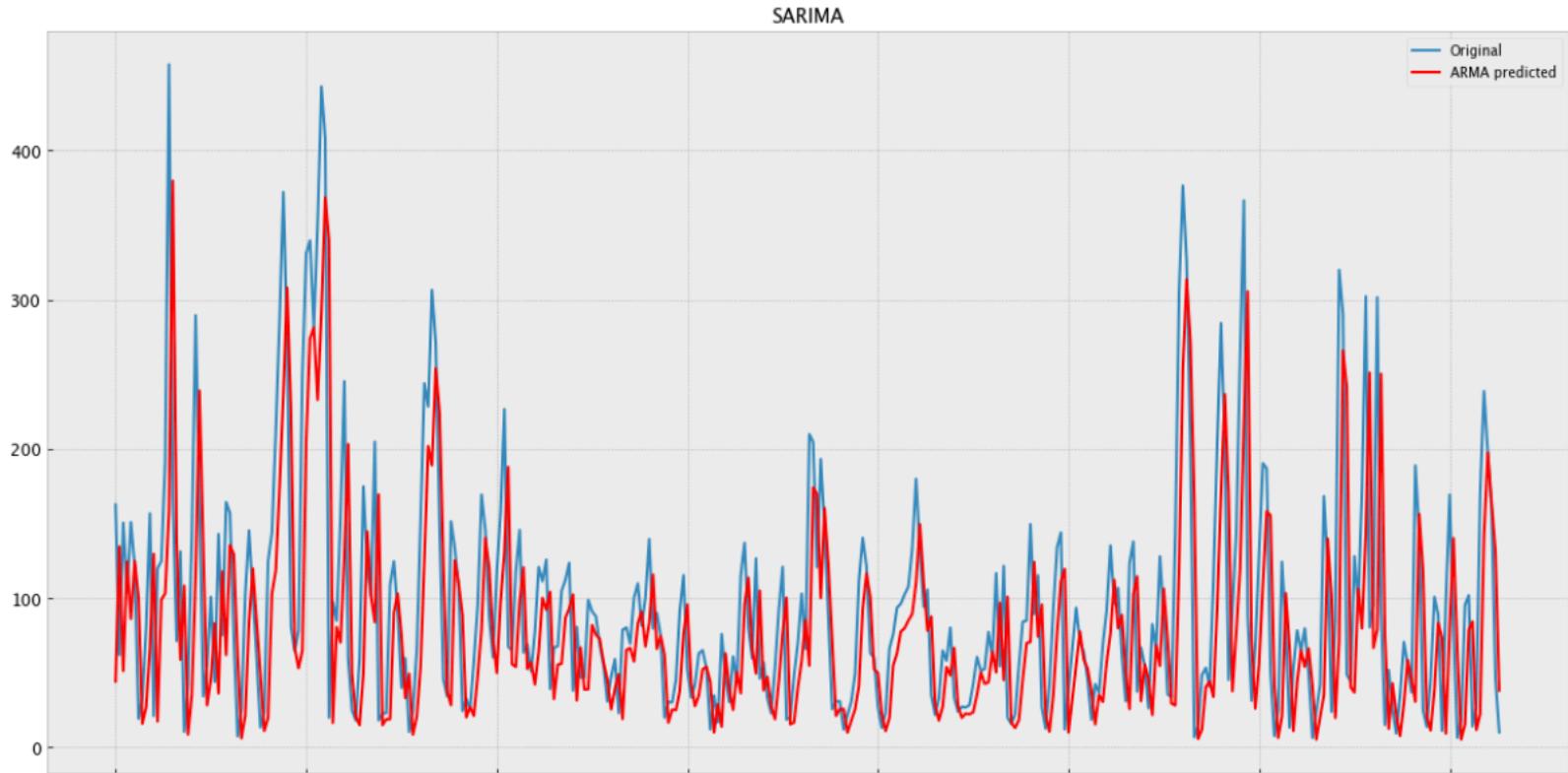
Models Tested:

- **Autoregression (AR)**
- **Autoregressive Moving Average(ARMA)**
- Autoregressive integrated moving average (ARIMA)
- Seasonal autoregressive integrated moving average (SARIMA)
- XGBoost
- Lightgbm
- Prophet
- DeepAR

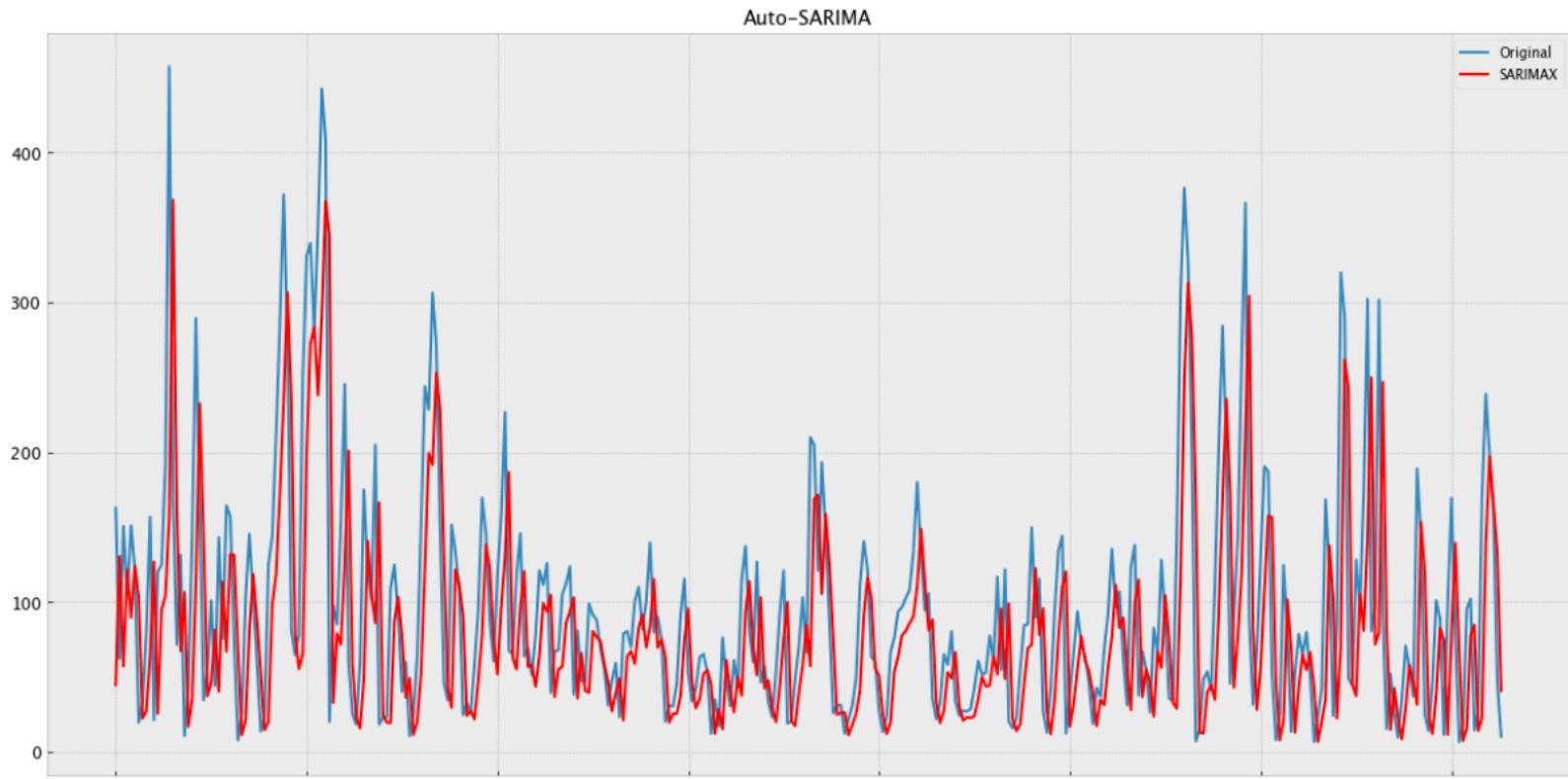
AR



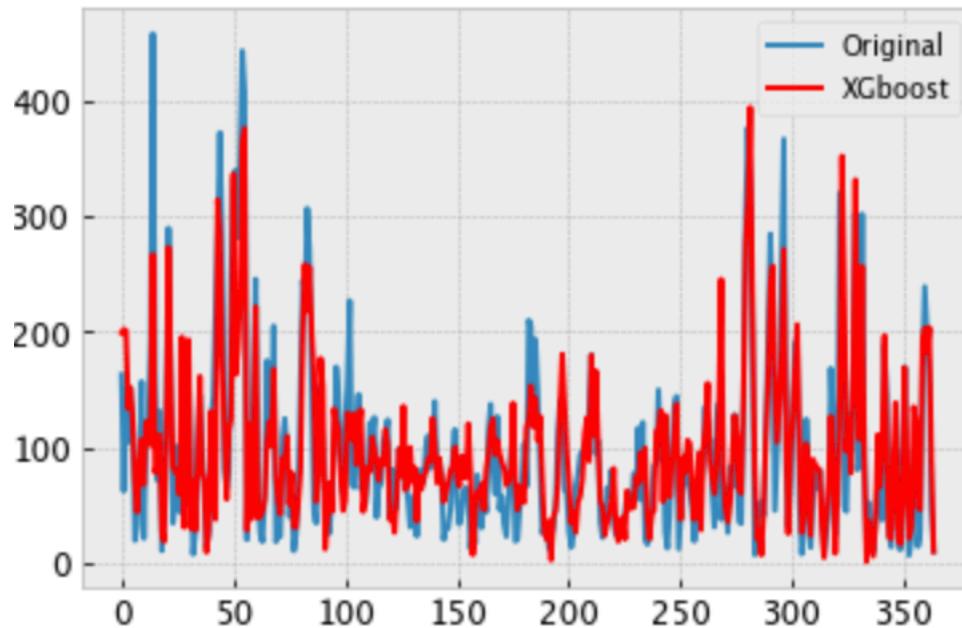
Seasonal Autoregressive Integrated Moving-Average (SARIMA)



Auto SARIMA



Linear Model - XGBOOST



Best Model

- ◎ Deep AR – The worst
 - MAE – 72
 - RMSE- 105

- ◎ Lightgbm
 - Mae - 30
 - RMSE - 43

Conclusion

- ◎ Most models that used extra features, performed better.
- ◎ Always perform extensive stationarity checks as it will help choose which models you want to use.
- ◎ Model called N-BEATS, Neural basis expansion analysis for interpretable time series forecasting