Predictive Maintenance using IoT Data

We are going to forecast the pollution level at a given point of time, given the pollution level of last N observations. We will be using a IOT simulated device with AWS Greengrass and collect device data into IoT core. We are planning to use IOT Analytics,S3 and Sagemaker for making predictions using trained model.

Dataset

https://archive.ics.uci.edu/ml/machine-learning-databases/00381/PRSA_data_2010.1.1-2014.12.31.csv

EDA:

1. Handling Missing Values – We dropped rows with null values for pollution
2. We created a date column combining hours/day/month
3. We plotted the correlation of pollution value with other independent variables and mentioned results in EDA notebook. Came up with optimal p value of 25 by plotting autocorrelation plot

Feature Engineering:

1. Data Normalization - Min-Max Scaler
2. Reshaping dataset

Model Training:

As we will be forecasting the pollution level based on last N timesteps, We experimented on two time series models. One is ARIMA and other is LSTM.

ARIMA:We built predictive ARIMA model iterating the test data and building the model again like rolling forecast with p=4,q=1 and d=0.

LSTM:We used LSTM and RNN with 20 epochs.

Model	Iterations/Epochs	RMSE
ARIMA	1(with optimal p value = 25)	135
ARIMA	Size of test data(Rolling forecast)	347
LSTM	20	442

Architecture/Deployment:

In Progress