Predicting Air Pollutant Concentrations in Beijing Using Regression Models

Md Rounak Jahan Raj
rounakjahanraj@gmail.com

Ephrem Alemu Mehammed
ephrem.alemu@aau.edu.et

TU Dortmund University

International Master of Advanced Methods in Particle Physics (IMAPP)

Problem

Challenge

Given various meteorological variables (temperature, pressure, etc), what are the predicted concentrations of air pollutants- PM10, SO_2 and NO_2 in Beijing?

Motivation

Air pollution is one of the most serious environmental challenges of our time. According to the World Health Organization, it contributes to around 8 million premature deaths every year. Moreover, 99% of the global population is exposed to air that does not meet WHO standards.

Dataset

Source	UC Irvine Machine Learning Repository
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
Information	Hourly data of 6 air pollutants + 6 meteorological variables from Beijing stations
Entries	420,768 rows x 17 columns
Important Features	"day", "hour", "TEMP", "PRES", "DEWP", "station", "WSPM"
Targets	"PM10", "SO_2" and "NO_2"
Previous Work	Prior work predicted "PM2.5"

Sequential DNN

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 24, 128)	89,088
dropout_2 (Dropout)	(None, 24, 128)	0
lstm_3 (LSTM)	(None, 64)	49,408
dropout_3 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 32)	2,080
dense_3 (Dense)	(None, 3)	99

Total params: 140,675 (549.51 KB)

Trainable params: 140,675 (549.51 KB)

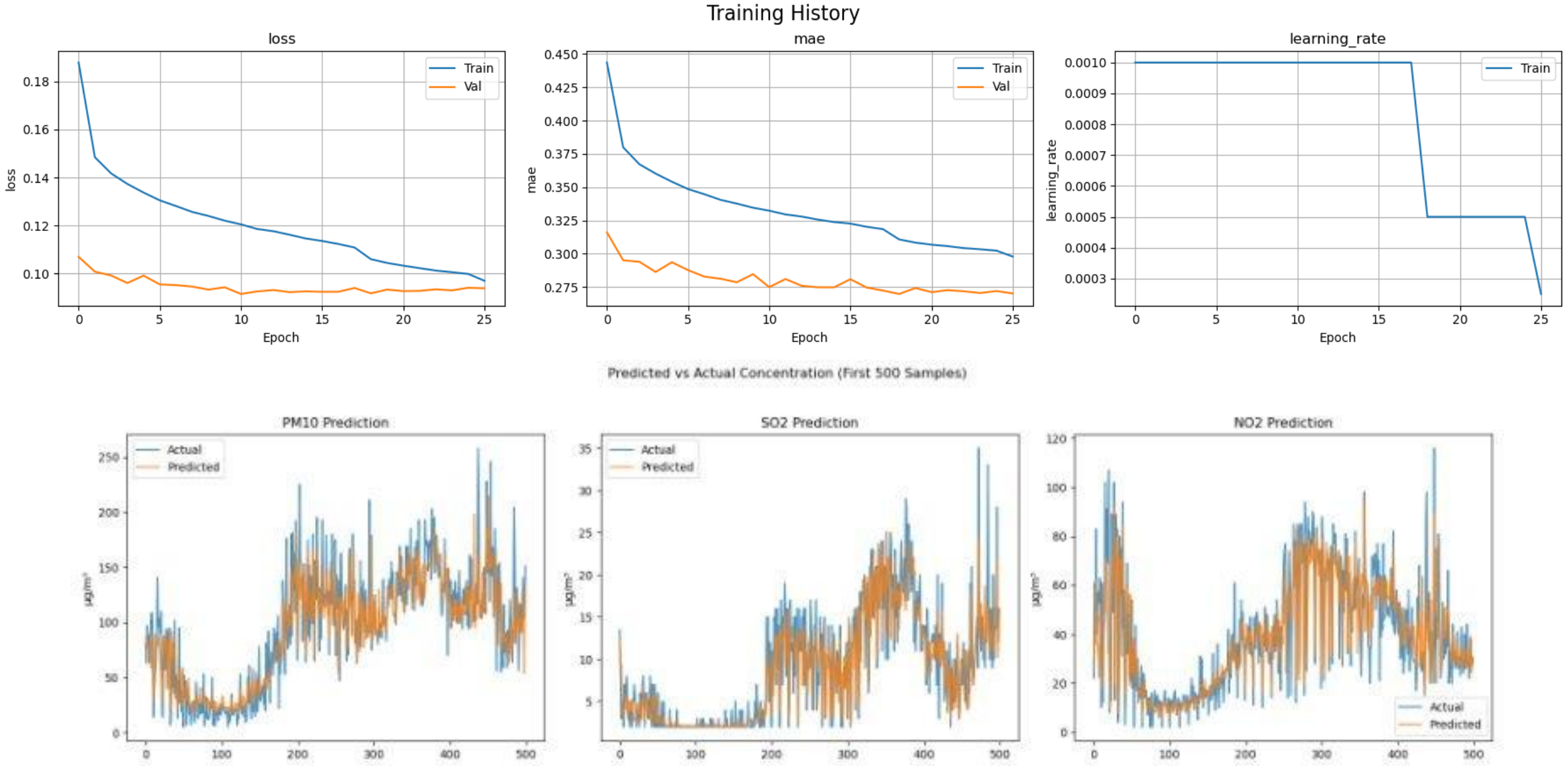
Non-trainable params: 0 (0.00 B)

Hyperparameter Tuning

To improve the performance of the Sequential DNN model:

- Number of units
- Optimizer
- Early Stopping
- Learning rate: Reduce LR On Plateau is used.
- Dropout & Batch Normalization to stabilize training
- Batch size
- Dropout regularization to avoid over fitting

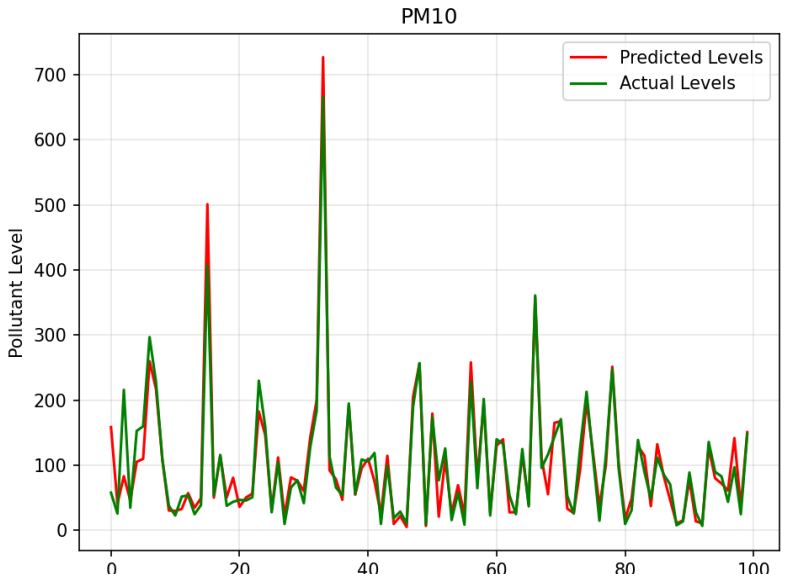
Results

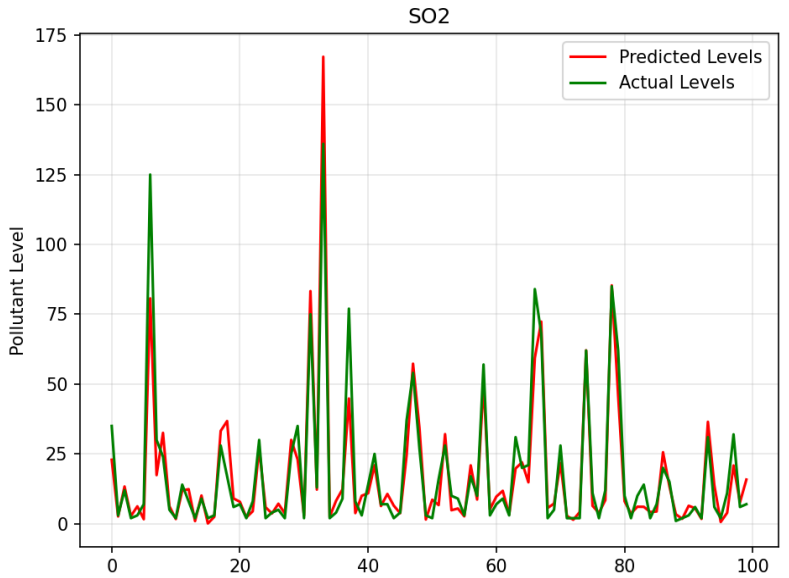

Training History
loss
Train
Val
loss
Epoch
mae
Train
Val
mae
Epoch
learning_rate
Train
learning_rate
Epoch
Predicted vs Actual Concentration (First 500 Samples)
PM10 Prediction
Actual
Predicted
µg/m³
SO2 Prediction
Actual
Predicted
µg/m³
NO2 Prediction
Actual
Predicted
µg/m³

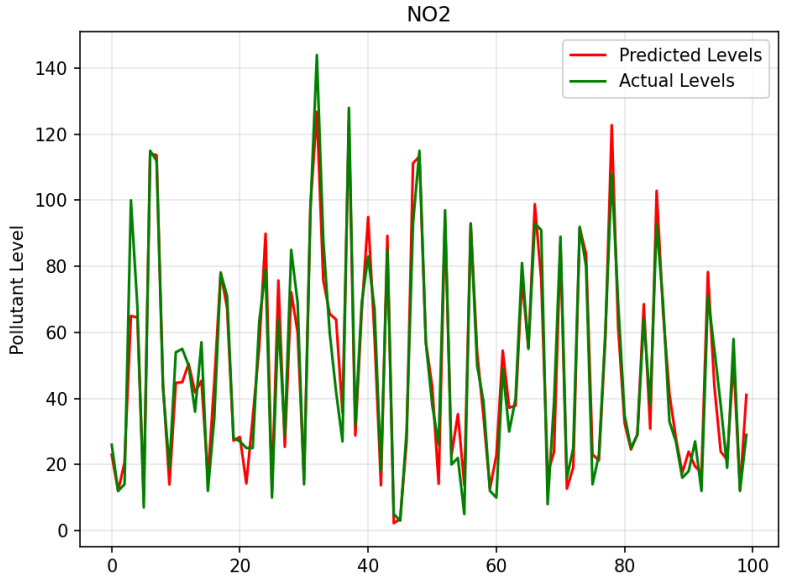
Results

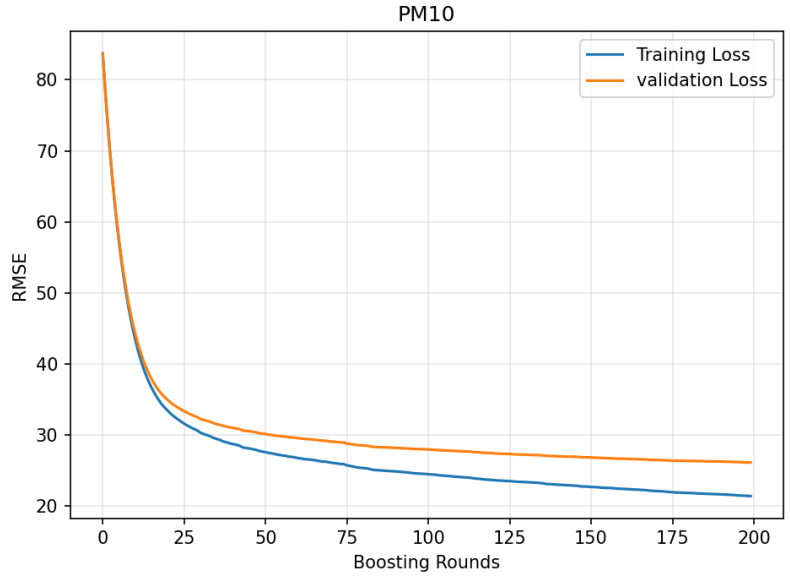
	MAE	RMSE	R^2
PM10	13.7	5.0	0.92
SO2	2.8	4.1	0.89
NO2	5.6	4.7	0.93

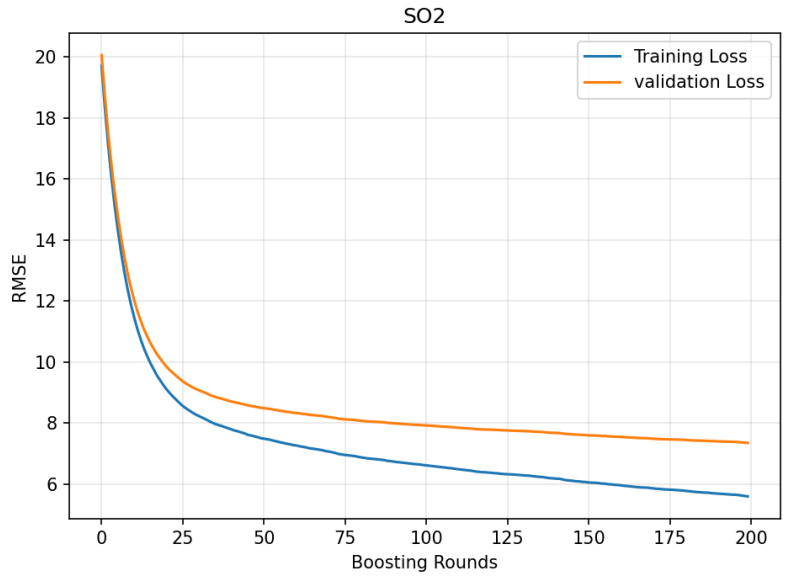
XGBRegressor

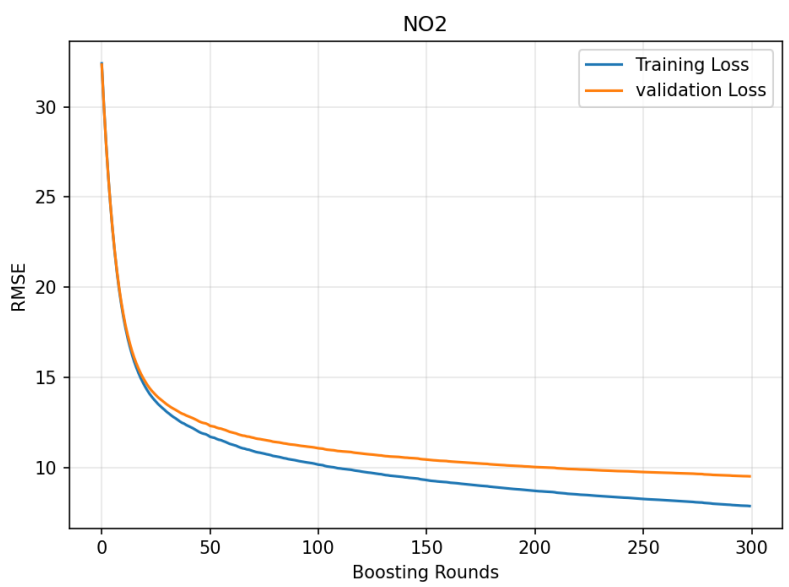












Conclusions

XGBRegressor:

	MAE	RMSE	R^2
PM10	15.76	3.97	0.91
SO2	3.99	2.00	0.89
NO2	6.77	2.60	0.93

Sequential DNN:

	MAE	RMSE	R^2
PM10	13.7	5.0	0.92
SO2	2.8	4.1	0.89
NO2	5.6	4.7	0.93

- **The performance of both models is comparable.**
- **Further work is required to improve the results.**