2702208575 \ MICHELLE NATHANIA \ DEEP LEARNING FINAL EXAM

air temperature forecasting with LSTM

OVERVIEW EDA PREPROCESSING MODELING EVALUATION

why air quality forecasting?

Forecasting future air temperature is crucial for many sectors, including public health and energy management.

	From Date	To Date	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)	NOx (ppb)	NH3 (ug/m3)	SO2 (ug/m3)	CO (mg/m3)	Ozone (ug/m3)
0	2017-09-05 14:00:00	2017-09-05 15:00:00	25.00	45.00	1.80	12.20	7.90	10.20	5.60	0.35	79.50
1	2017-09-05 15:00:00	2017-09-05 16:00:00	23.00	49.50	0.65	14.55	8.28	8.85	4.52	0.41	62.50
2	2017-09-05 16:00:00	2017-09-05 17:00:00	18.00	55.75	4.00	26.58	17.40	6.23	5.35	0.38	40.50
3	2017-09-05 17:00:00	2017-09-05 18:00:00	19.75	33.50	1.63	21.78	12.92	8.30	5.30	0.48	34.43
4	2017-09-05	2017-09-05	14.00	28.00	1.07	26.38	14.90	3.42	4.68	0.43	33.00

dataset

A rich air quality dataset spanning over **5 years**, containing **24 different features** including pollutants and meteorological data, recorded at **hourly** intervals.

objective

To build and evaluate several **LSTM models** to accurately predict the **air temperature** (AT) one hour into the future, based on data from the preceding five hours.

approach

Multivariate time series analysis

→

AT is influenced by many factors that provide richer context to the model

EXPLORATORY DATA ANALYSIS

first glance of the data

EDA is important to decide which preprocessing methods to use.

initial findings

- Substantial dataset, spanning from Sep 2017 - Mar 2023
- Data is skewed and has significant outliers

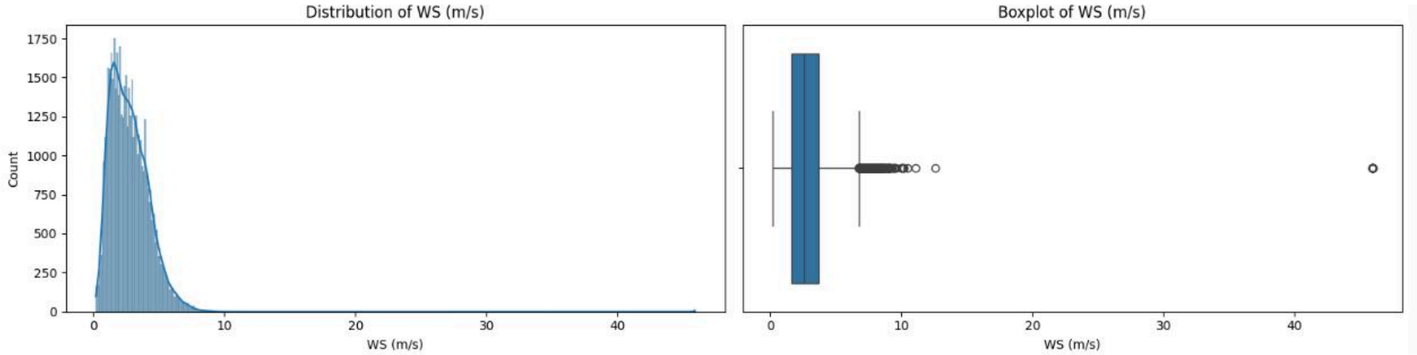
```
# of rows:  48802
# of columns:  24
```

Temp (degree C)	55.741568
Eth-Benzene (ug/m3)	50.846277
MP-Xylene (ug/m3)	19.560674
PM2.5 (ug/m3)	5.036679
CO (mg/m3)	4.948568

statistical summary

	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)
count	48802.000000	48802.000000	48802.000000
mean	34.597196	71.466791	2.506647
std	30.839854	48.413886	4.993837
min	0.250000	0.250000	0.030000
25%	11.750000	34.250000	1.450000
50%	24.500000	59.500000	1.900000
75%	51.000000	100.000000	2.600000
max	723.000000	835.500000	216.170000

numerical data viz.



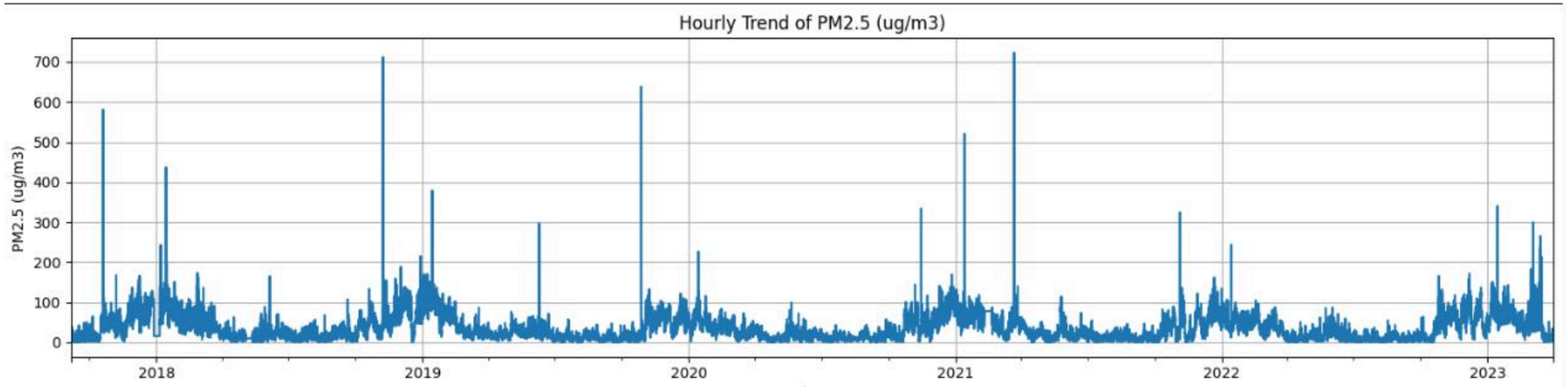
OVERVIEW | EDA | PREPROCESSING | MODELING | EVALUATION

EXPLORATORY DATA ANALYSIS

visualizing trends and patterns

identified patterns

strong annual seasonality

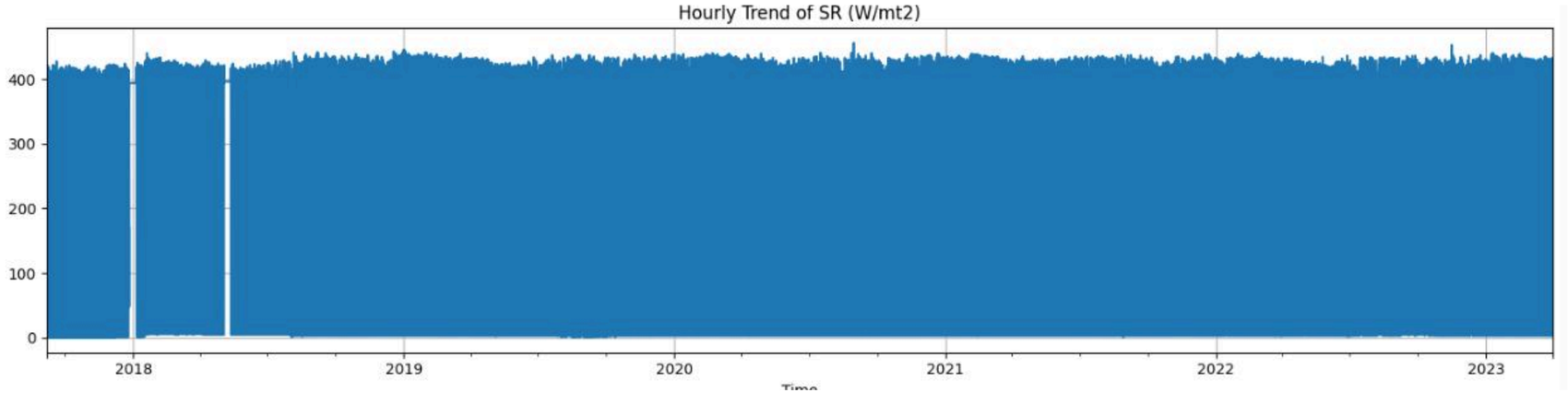


EXPLORATORY DATA ANALYSIS

visualizing trends and patterns

identified patterns

clear daily cycles

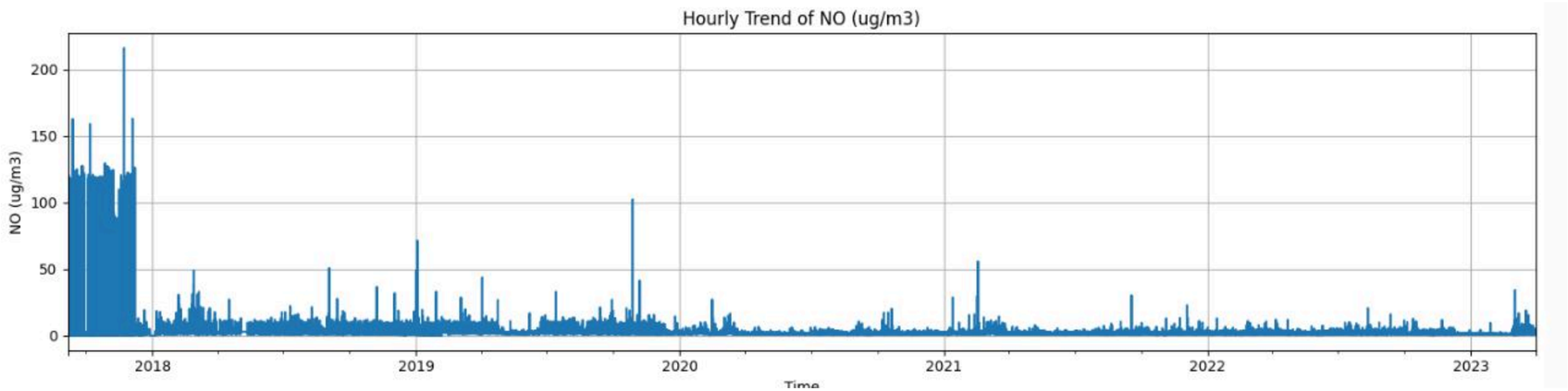


EXPLORATORY DATA ANALYSIS

visualizing trends and patterns

identified patterns

long term trends



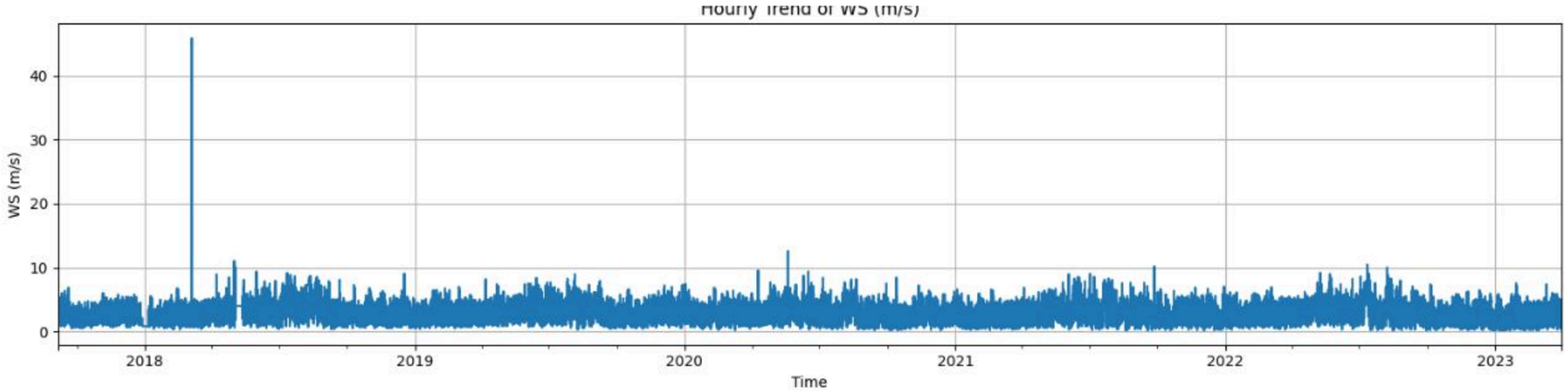
OVERVIEW | EDA | PREPROCESSING | MODELING | EVALUATION

EXPLORATORY DATA ANALYSIS

visualizing trends and patterns

identified patterns

extreme outliers



DATA PREPROCESSING

feature engineering & scaling

cyclical features

- P: LSTMs see numbers like 0 and 360 degree as far apart, when they are actually adjacent.
- S: Use sine and cosine functions.

splitting & scaling

- Split data 80/10/10 for train/val/test
- Scaled all features with RobustScaler

input sequence

- Make sliding window that LSTM understands
- input_length = 5
- output_length = 1

```
X_train shape: (39036, 5, 28) -> (Samples, Time Steps, Features)
y_train shape: (39036, 1, 1) -> (Samples, Prediction Time Steps, Target Features)

X_val shape:   (4875, 5, 28)
y_val shape:   (4875, 1, 1)

X_test shape:  (4876, 5, 28)
y_test shape:  (4876, 1, 1)
```

MODELING & TRAINING

model parameters

LSTM

Layer (type)	Output Shape	Param #
lstm_5 (LSTM)	(None, 10)	1,560
dense_5 (Dense)	(None, 1)	11

LSTM Modified

Layer (type)	Output Shape	Param #
lstm_6 (LSTM)	(None, 5, 128)	80,384
dropout_3 (Dropout)	(None, 5, 128)	0
lstm_7 (LSTM)	(None, 64)	49,408
dropout_4 (Dropout)	(None, 64)	0
dense_6 (Dense)	(None, 32)	2,080
dense_7 (Dense)	(None, 1)	33

LSTM Baseline Tuned

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 50)	15,800
dropout_1 (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 1)	51

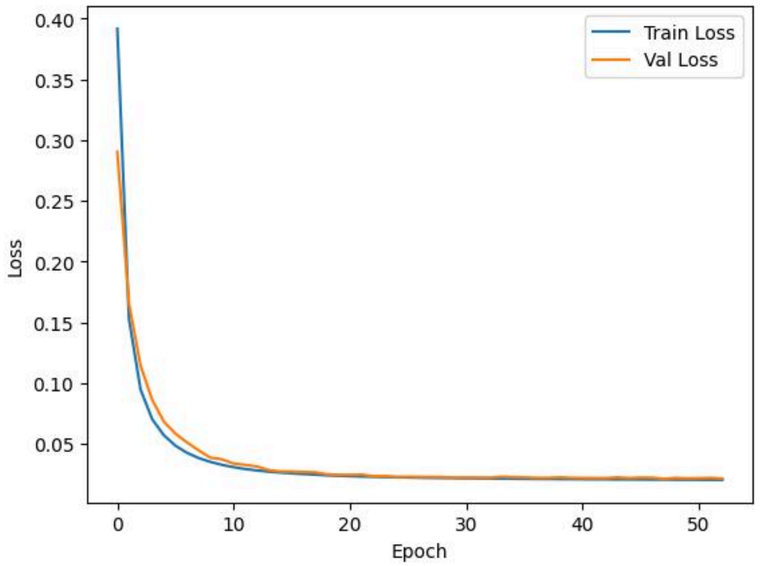
```
Trial 5 Complete [00h 01m 19s]
val_loss: 0.021622434258461

Best val_loss So Far: 0.018529120832681656
Total elapsed time: 00h 13m 06s
```

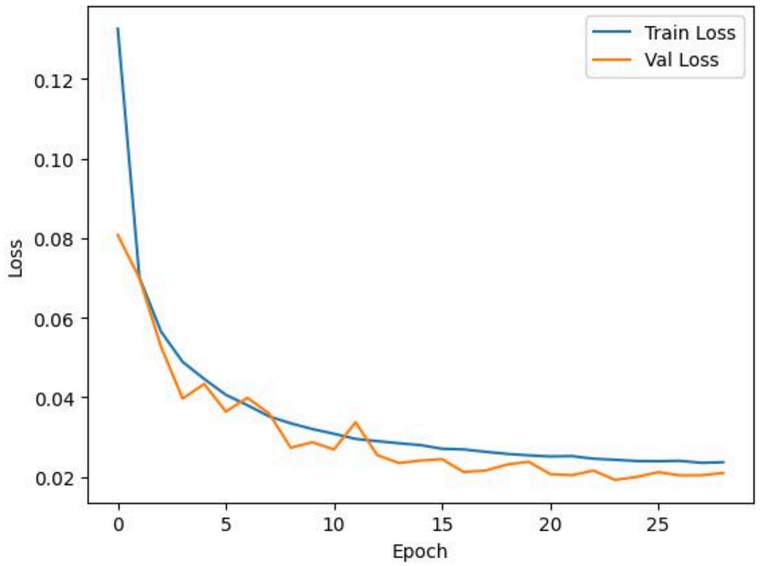
```
Optimal units in LSTM layer: 50
Optimal dropout rate: 0.15
Optimal learning rate: 0.001
```

loss plot

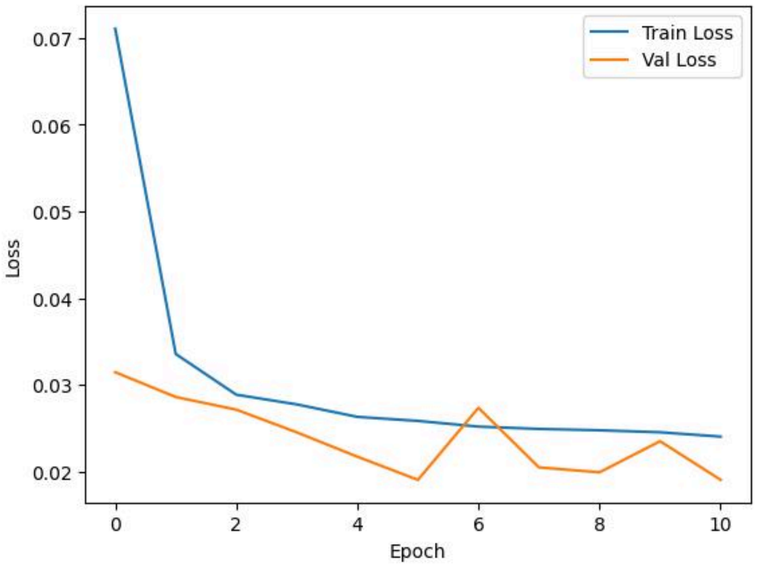
LSTM



LSTM Modified



LSTM Baseline Tuned



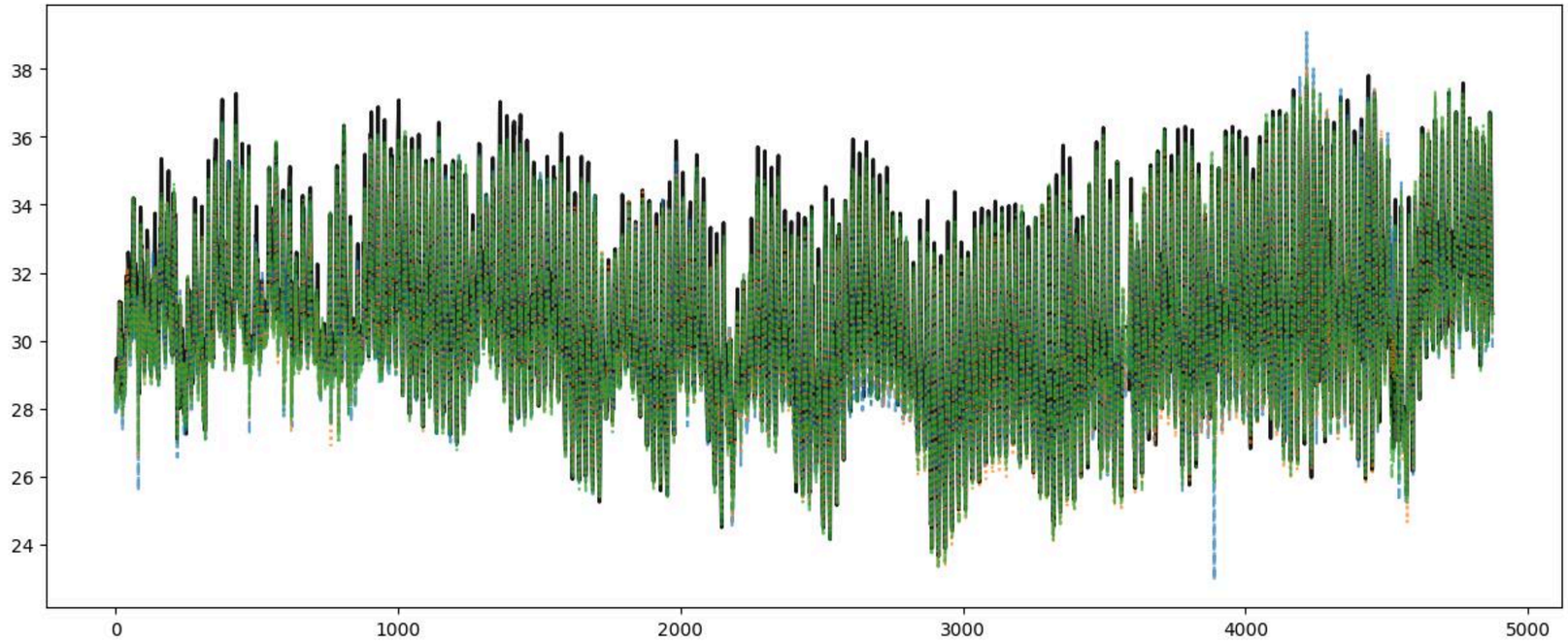
EVALUATION

evaluation metrics

Model	MAE	MSE	R-squared
Model 1 (LSTM)	0.737125	0.919030	0.869581
Model 2 (LSTM Modified)	0.701035	0.774790	0.890050
Model 3 (LSTM Baseline Tuned)	0.572590	0.542816	0.922969

All models performed well, with **Model 3 (LSTM Baseline Tuned)** being the **best** model.

verdict



- Treating this case as a **multivariate time series** is a correct approach.
- The preprocessing steps result in **good model performance**.
- Model doesn't require complex architecture, but tuning helps improve the performance significantly.

All models performed well, with **Model 3 (LSTM Baseline Tuned)** being the **best** model.