Chasing the WIND: Key Factors Driving Renewable Energy

Fall '24
DATASCI 207

Sonia Chakraborty
Jordan Lawhon
Tim Leong
Hunter Tonn
Mina Baghai

Background:

- Global Energy demand is expected to **increase** by 16-57% by 2050
  - Electric Vehicle Usage
  - Manufacturing/industrial
  - Data Centers
- To avoid the worsening impacts of climate change, we need to focus on **clean energy sources**

U.S. Energy Information Administration - EIA - independent statistics and analysis (2023) *EIA.*

Question

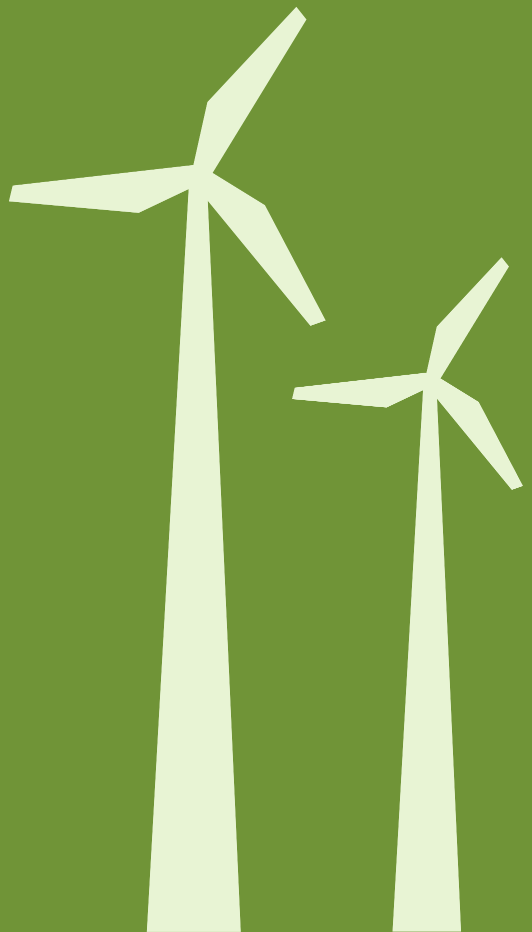
- ? How do atmospheric factors influence wind speed?
- ? How can we predict hourly wind speed to estimate energy production for utility providers throughout the year?

Sub question:

Can we forecast extreme weather conditions so that utility groups can take precautionary measures to avoid disruptions in energy production?

Data Source

- *Kaggle's Weather Long-Term Time Series Forecasting Dataset*
  - 2020 meteorological indicators measured at a Max Planck Institute weather station (located in Germany)
  - Contains atmospheric variables including air temperature, humidity, wind patterns, radiation, and precipitation **recorded every 10 minutes**
  - 20 variables. 52,560 data points per variable (365 days × 24 hours × 6 measurements per hour)
- Target Variable: **average hourly wind speed**



Data Preprocessing

Aggregation

Aggregated dataset to be on an hourly level. Took the average 10 minute variables per hour.

Lagged Features

Created lagged moving averages of features to incorporate the previous 6 hours of atmospheric data for predicting the next hour's wind speed.

Sine Cosine Transformation

Used sine and cosine transformations of timestamps to capture the cyclical nature of time, modeling daily and yearly patterns in weather data.

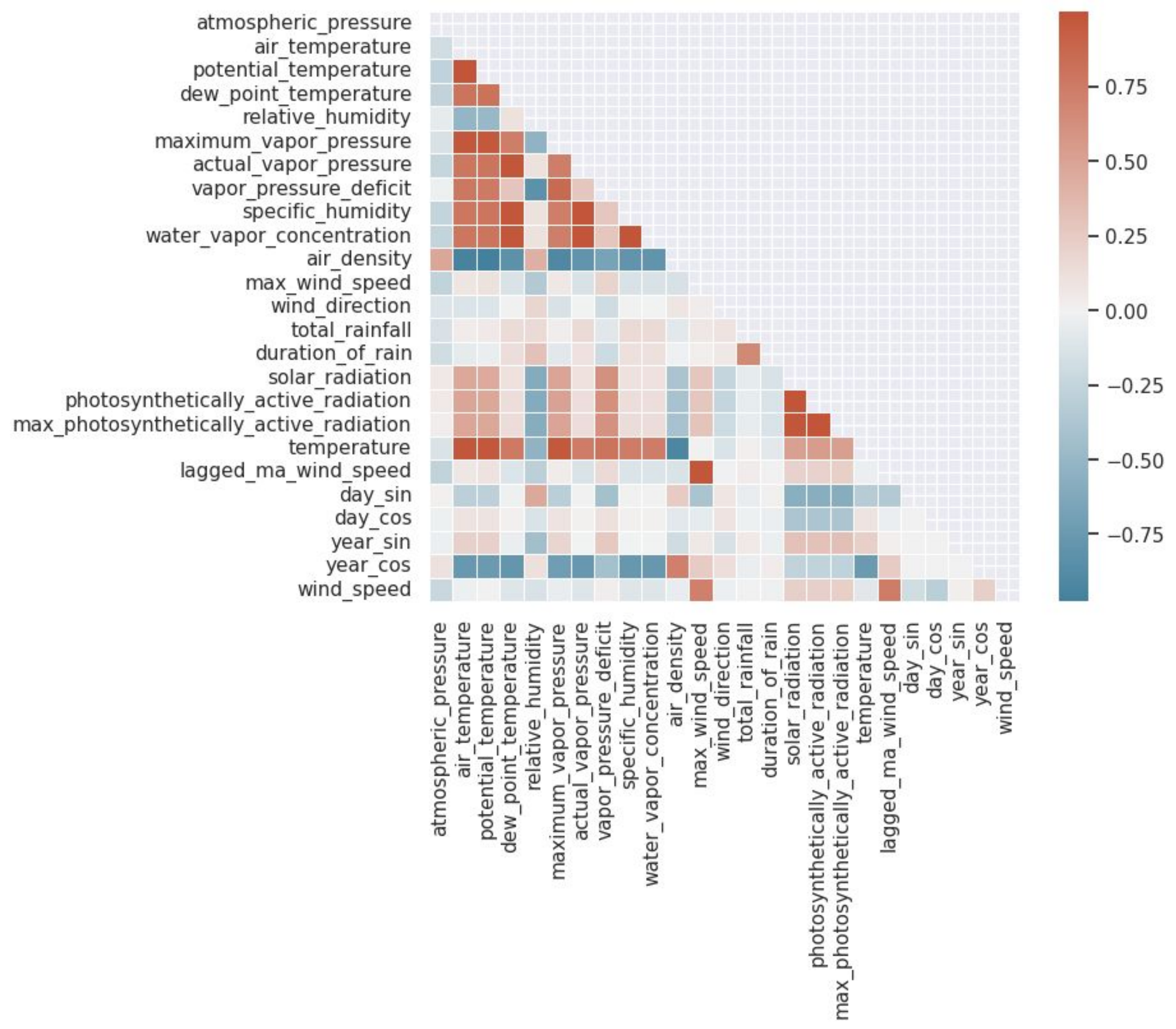
Train, Valid, Test Split

Time series split on the Train/Valid 23% in Test

Data was standardized based on training mean / std

Exploratory Data Analysis

Correlation Matrix



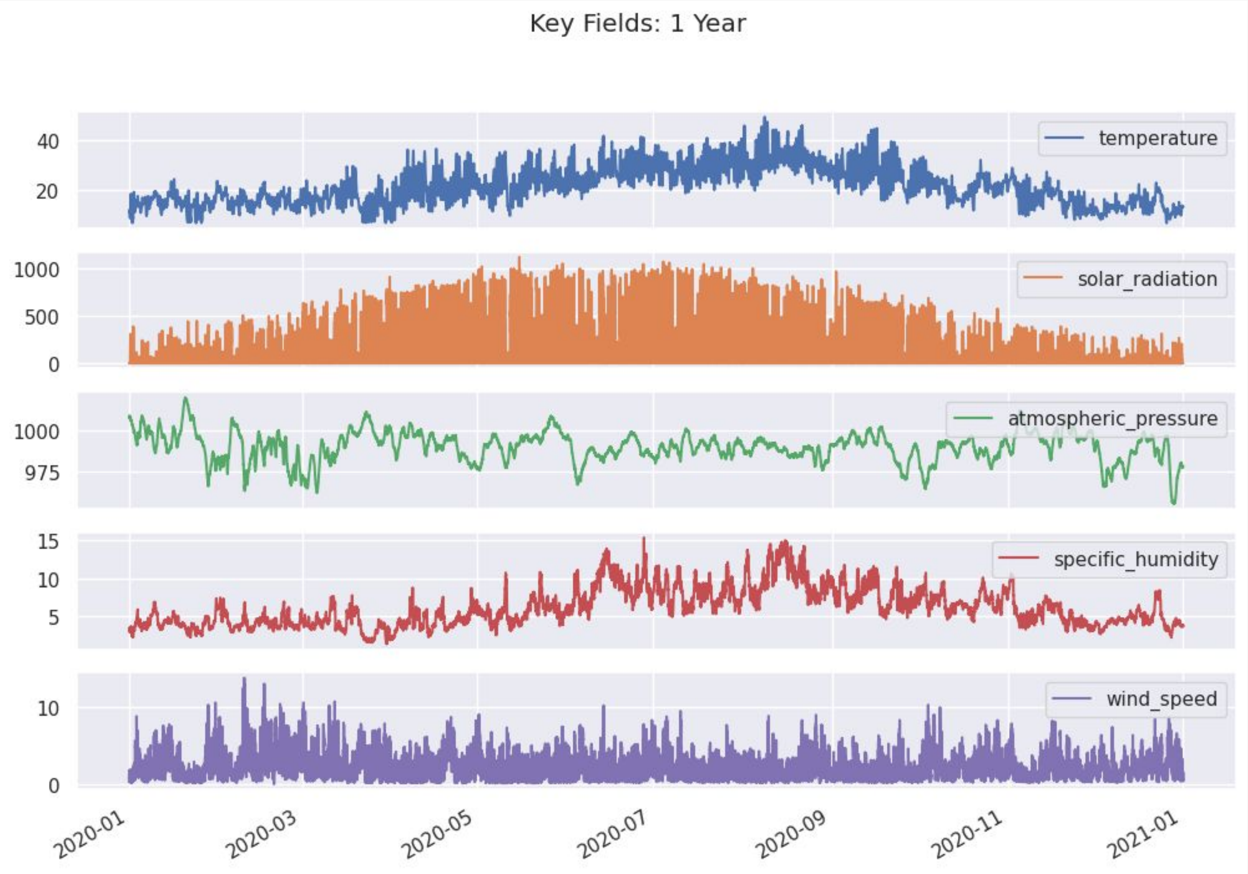
Positively correlated :

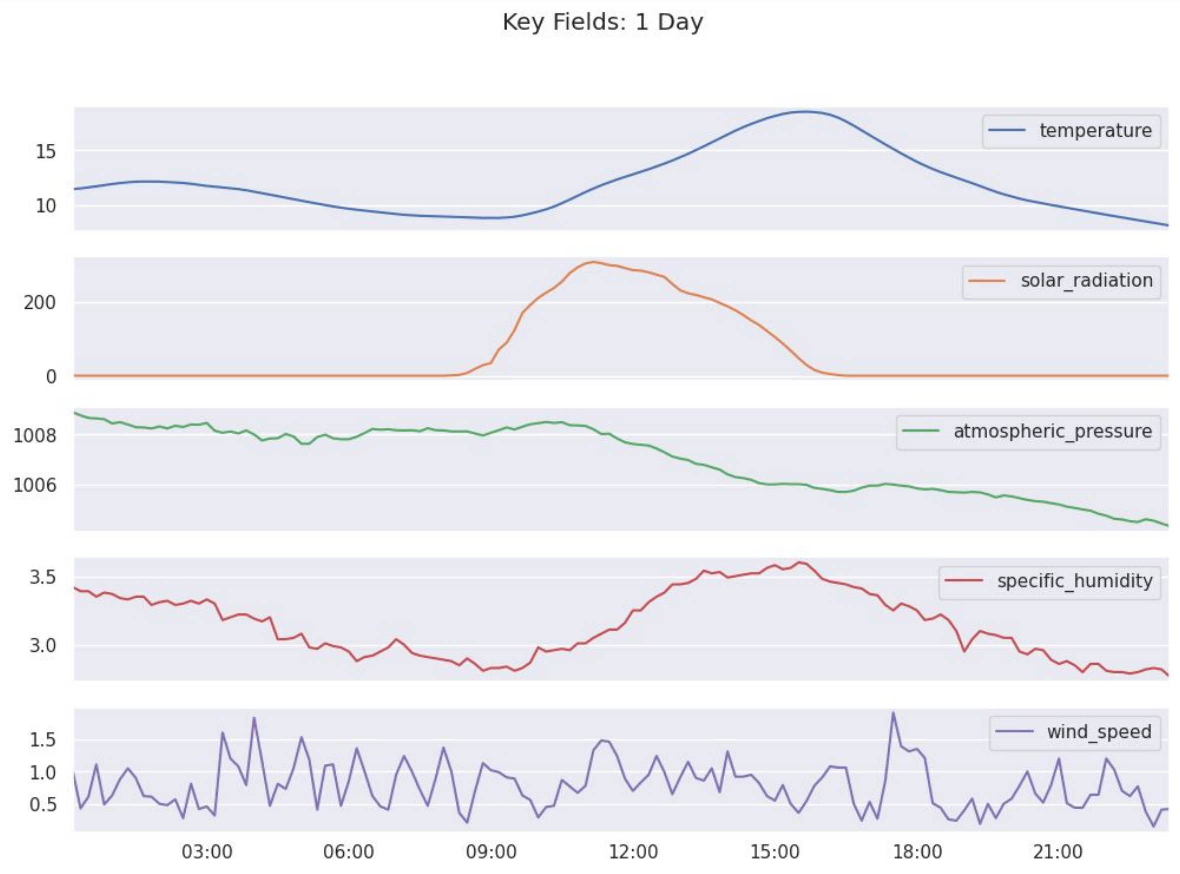
- Solar radiation,
- Lagged wind speed,
- Year cos

Negatively correlated:

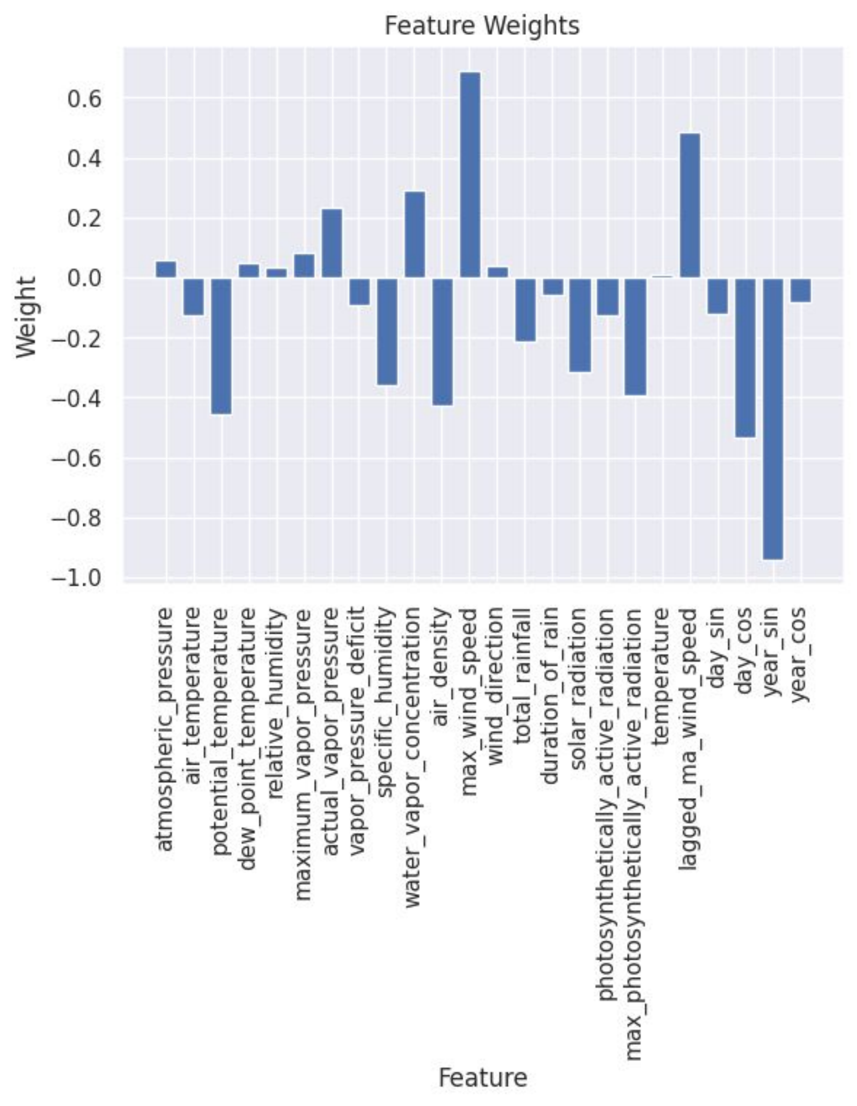
- Atmospheric pressure,
- Humidity,
- Dew point humidity,
- Day cos

Time Series Analysis





Feature Weights



Key Variables

- Potential temperature
- Actual vapor pressure
- Specific humidity
- Air density
- Max wind speed
- Solar radiation
- Max photosynthetically active radiation
- Lagged ma wind speed
- Day cosine
- Year sine

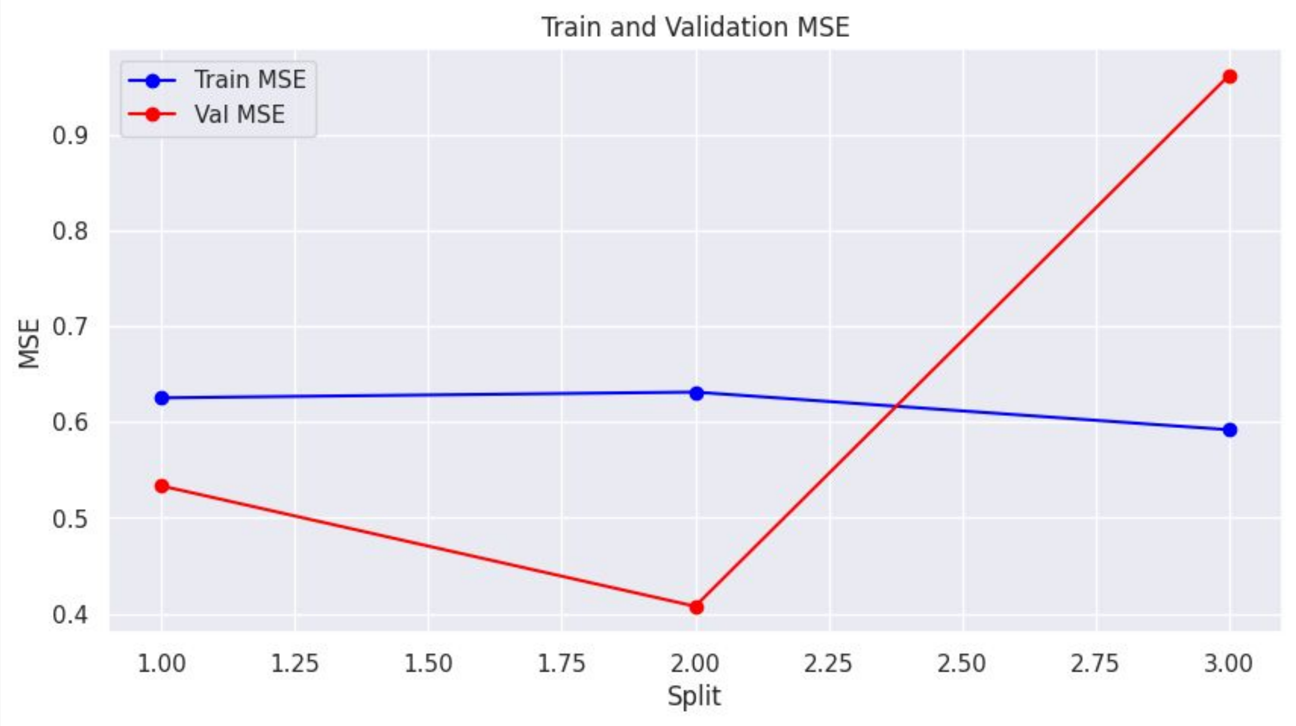
MODELING

Baseline Model

- Predict wind speed as the average wind speed from the training dataset
- Train MSE: 2.32
- Valid MSE: 2.31

Linear Regression Model

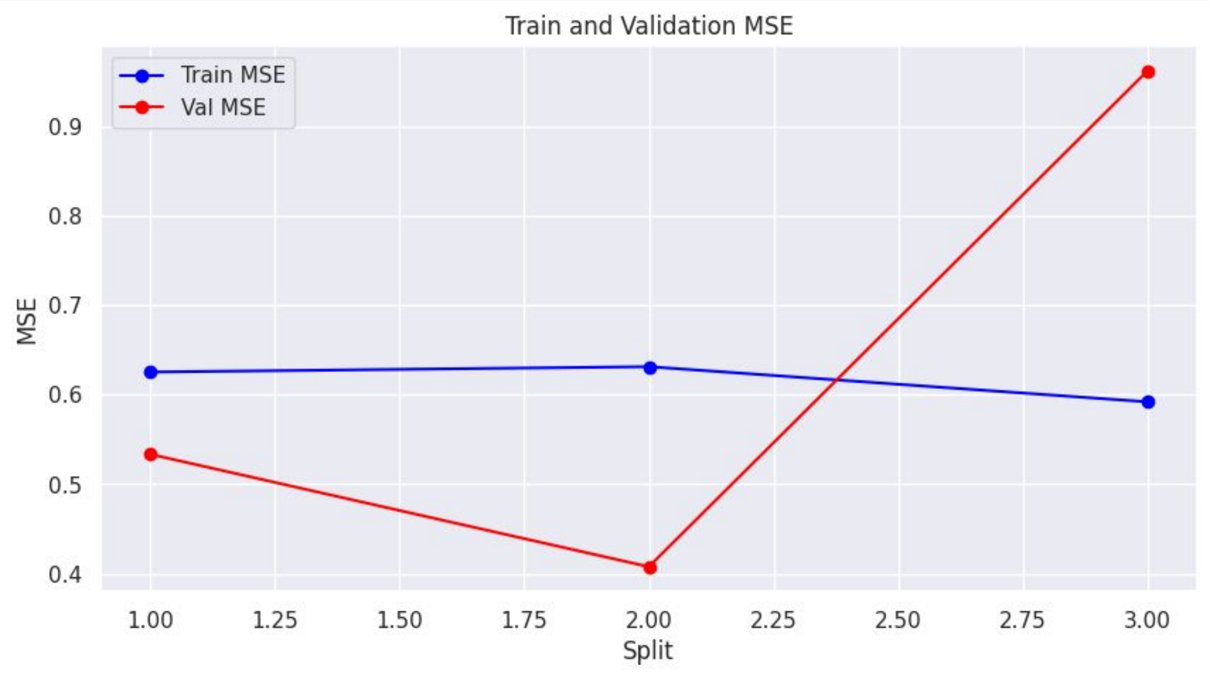
- Ridge Regression Model
  - TimeSeriesSplit cross validation
- Begins to overfit the data




```
Average Train MSE: 0.6162
Average Validation MSE: 0.6343
```

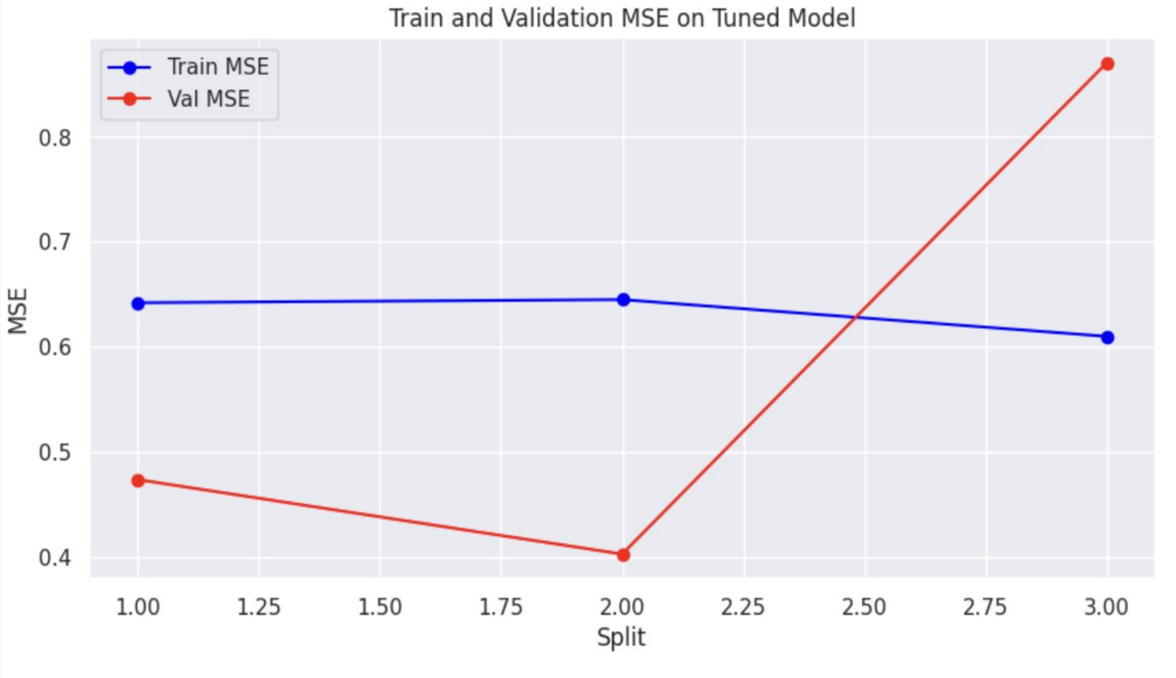
Linear Regression Model: Before And After Tuning

Alpha: 1.0




```
Average Train MSE: 0.6162
Average Validation MSE: 0.6343
```

Alpha: 21.2

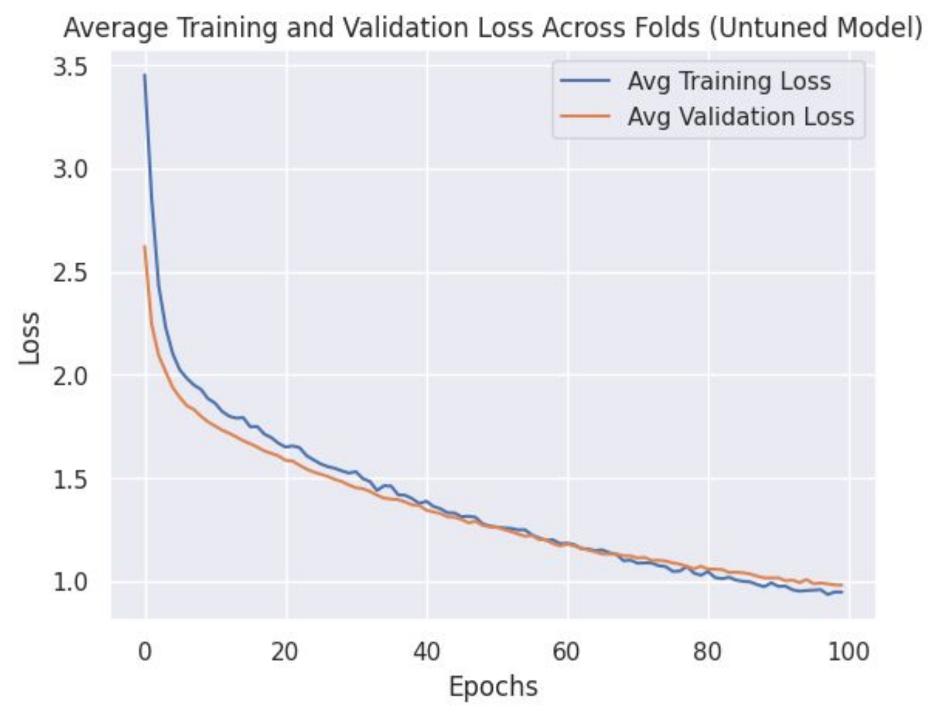



```
Average Train MSE: 0.6324
Average Validation MSE: 0.5823
```

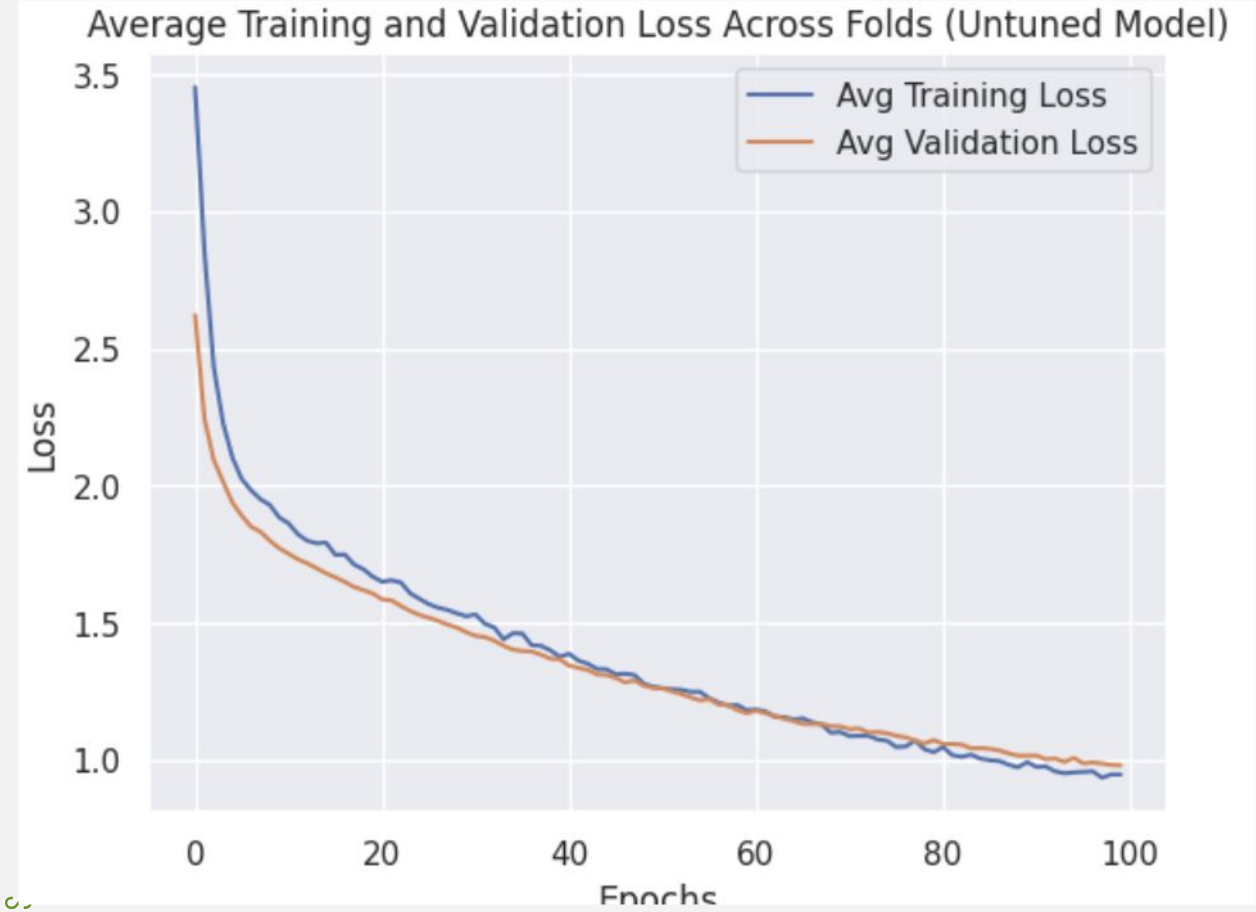
Neural Network

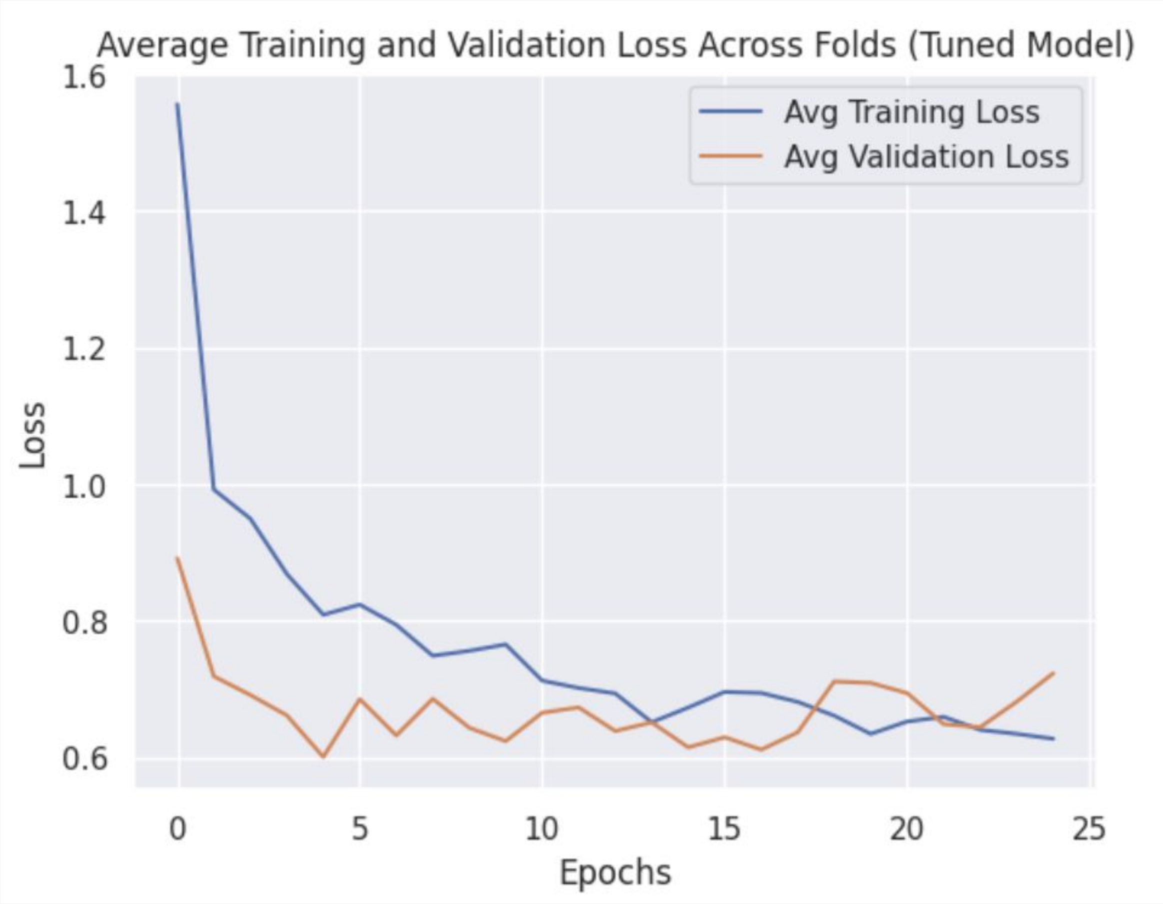
Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 24)	0
dense (Dense)	(None, 256)	6,400
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 128)	32,896
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 1)	65

Total params: 47,617 (186.00 KB)
Trainable params: 47,617 (186.00 KB)
Non-trainable params: 0 (0.00 B)



Neural Network: Before And After Tuning





Mean Validation Loss: 0.7496 ± 0.2820

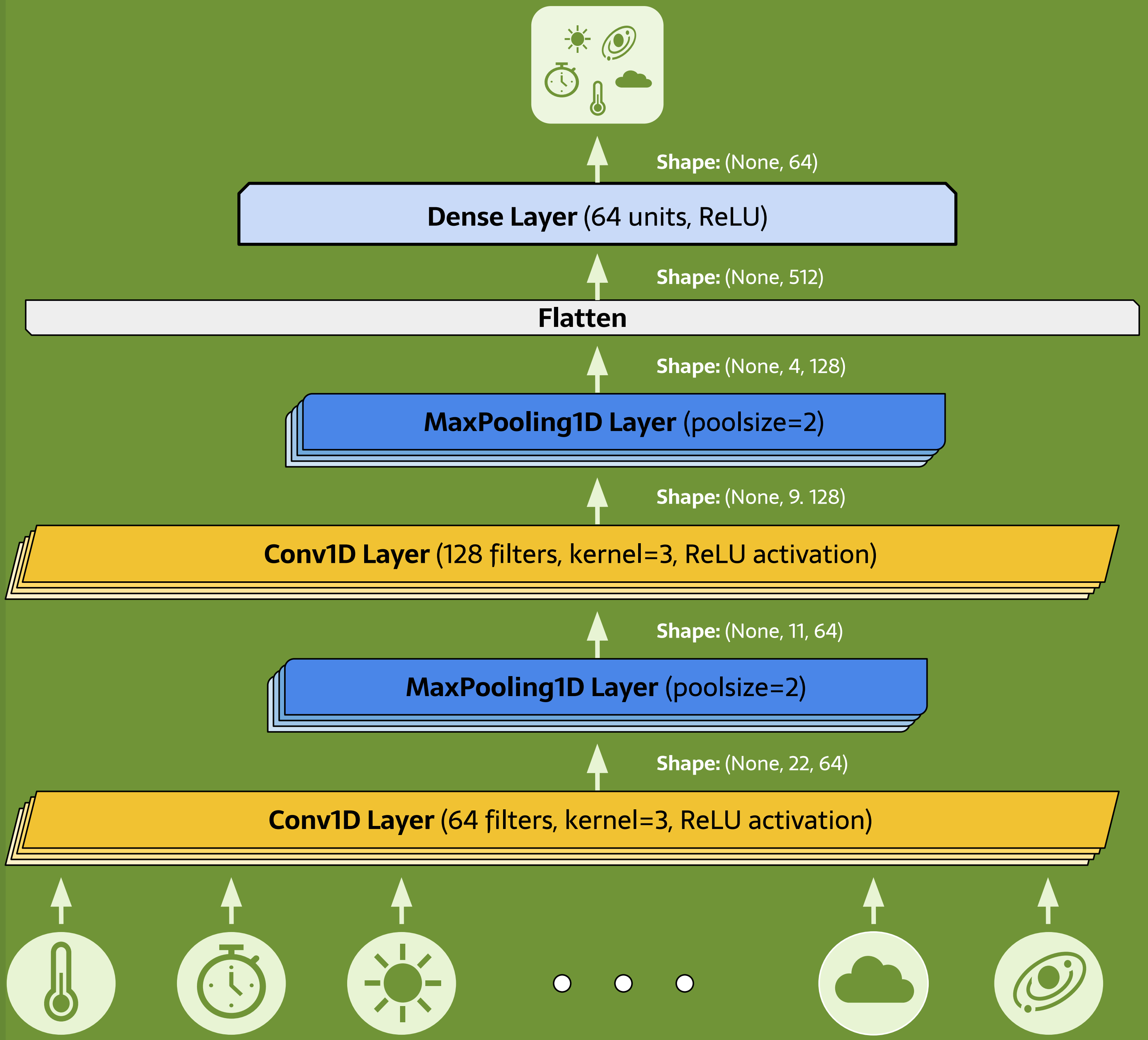
Mean Validation Loss: 0.5550 ± 0.1839

Neural Network: Before And After Tuning

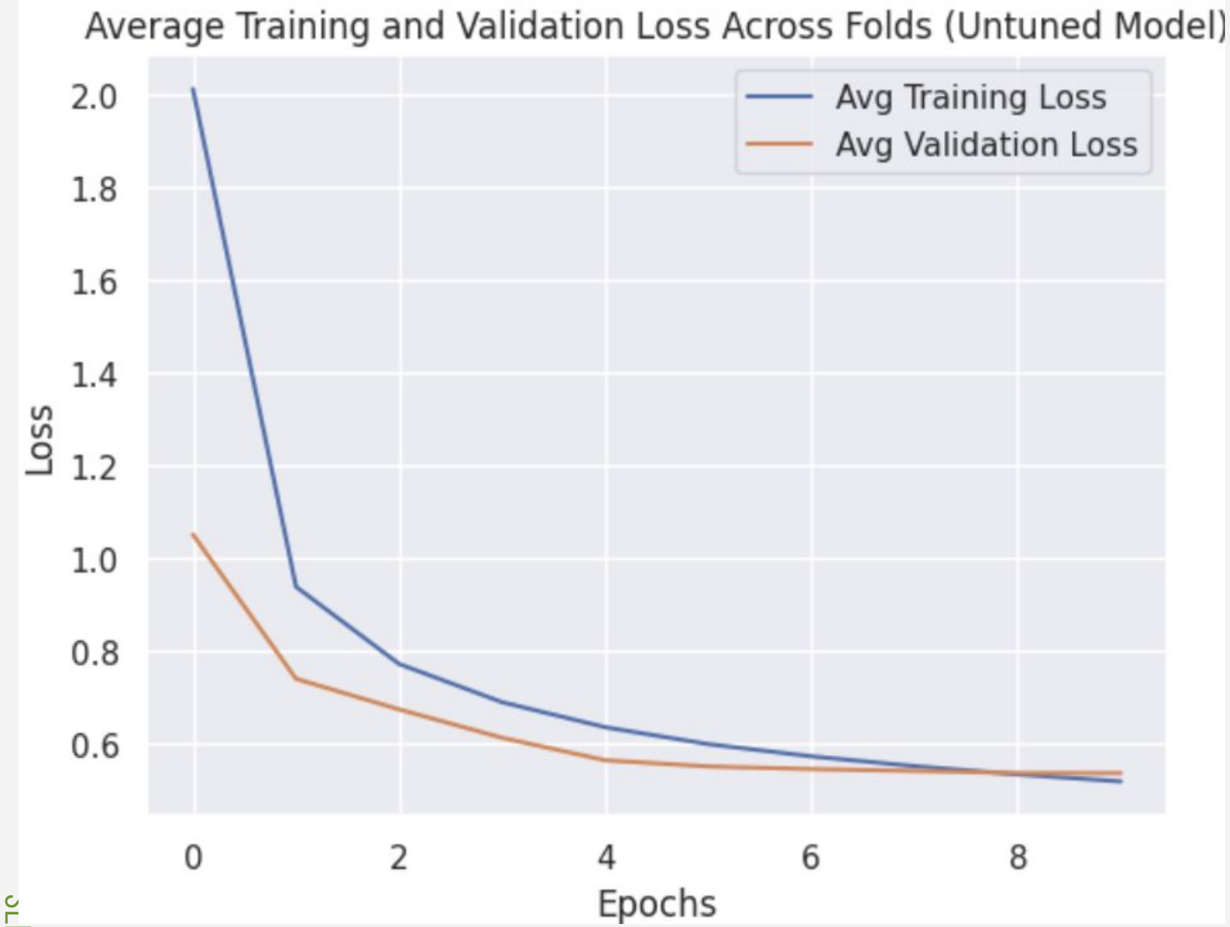
Hyperparameter	Hidden Layer	Untuned Value	Tuned Value
Units	1	256	160
L2 Regularization Penalty	1	0.01	0.0002
Dropout Rate	1	0.2	0.25
Units	2	128	96
L2 Regularization Penalty	2	0.01	0.0002
Dropout Rate	2	0.2	0.3500
Units	3	64	64
L2 Regularization Penalty	3	0.01	0.0007
Dropout Rate	3	0.2	0.25
Learning Rate	-	0.0001	0.008
Epochs	-	100	25

Convolutional Neural Network:

Model Architecture



CNN: Before And After Tuning



Mean Validation Loss: 0.5340 ± 0.2528

Mean Validation Loss: 0.4867 ± 0.1695

CNN: Before And After Tuning

Hyperparameter	Hidden Layer	Untuned Value	Tuned Value
Filters	1	64	40
Filters	2	128	10
Kernel Size	1	3	5
Kernel Size	2	3	4
Activation	1	relu	relu
Activation	2	relu	sigmoid
Pool Size	1	2	4
Pool Size	2	2	2
Epochs	-	10	100

Evaluation

CNN

```
63/63 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 2.6148
Test Loss: 2.4114246368408203
```

*Neural Network

```
63/63 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step - loss: 1.5572 - mean_squared_error: 1.4956
Test Loss: [1.2756891250610352, 1.2140346765518188]
```

PERFORMANCE COMPARISON

	Baseline Model	Linear Regression	Neural Network	CNN
Untuned Validation MSE	N/A	0.63	0.75	0.53
Tuned Validation MSE	2.31	0.58	0.55	0.49

Takeaways & Future Research

- Importance of Hyperparameter tuning to reduce and create constant validation loss
- Feature engineering was integral to identifying our most influential models for model performance
- New challenge of TimeSeries data!

In the future …

- Expand locations
- Get data for more than 1 year
- Splits and keeping temporality in data

Thanks!

Team Contributions

- **Mina Baghai:** data aggregation, feature engineering, baseline model, EDA
- **Tim Leong:** EDA, feature engineering, neural network model
- **Sonia Chakraborty:** Linear regression model, presentation slides
- **Hunter Tonn:** CNN model, presentation slides
- **Jordan Lawhon:** Model tuning, model evaluation, created GitHub repo