Solar Intensity Estimation using Time-Series weather data

Machine Learning Project

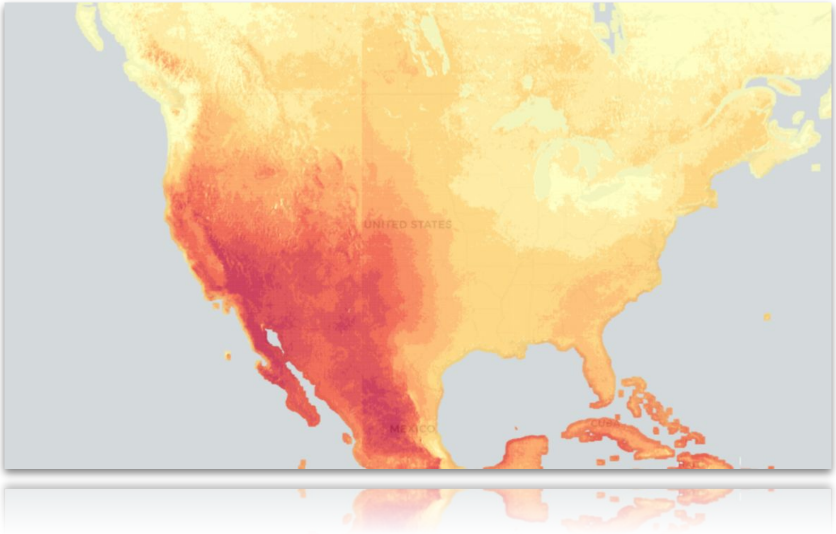
Naval Kumar Shukla
Kushal Juneja
Rishi Singhal
Udit Narang

Team Number - T3

INDRAPRASTHA INSTITUTE *of* INFORMATION TECHNOLOGY **DELHI**

Introduction

- <u>Goal</u> : Predict solar radiation using readily-available time series weather data.



Solar Power Intensity over North America [NSRDB]

Motivation

- Solar energy is inexhaustible energy resource.
- Solar panels only capture small amounts of energy from the sun.
- Highly variational nature of solar energy puts strain on conventional fossil fuel based energy sources.
- We aim to predict solar intensity 48 hours into the future using time-series weather data.

Solar Panel [Wikimedia Commons]

Literature Review

- Prediction models for smart homes for advanced planning of electricity consumption
- Compared Least Linear square regression and SVM considering the correlation between various weather observations

[1] N. Sharma, P. Sharma, D. Irwin and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), 2011, pp. 528-533, doi: 10.1109/SmartGridComm.2011.6102379.

Solar Power Plant [Wikimedia Commons]

Literature Review

- Forecasting renewable resources to benefit power systems
- Then the paper has focused on ANN. To select the most effective features they used - correlation & sensitivity analysis.
- After obtaining the best features, they have used different models like ANN, MLR & persistence forecasts model and compared their performance.
- The paper has also mentioned that removing night hours from the initial dataset has helped to slightly improve the model performance.

[2] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," 2015 North American Power Symposium (NAPS), 2015, pp. 1-5, doi: 10.1109/NAPS.2015.7335176.

Dataset Description

- Input Features

Weather Data

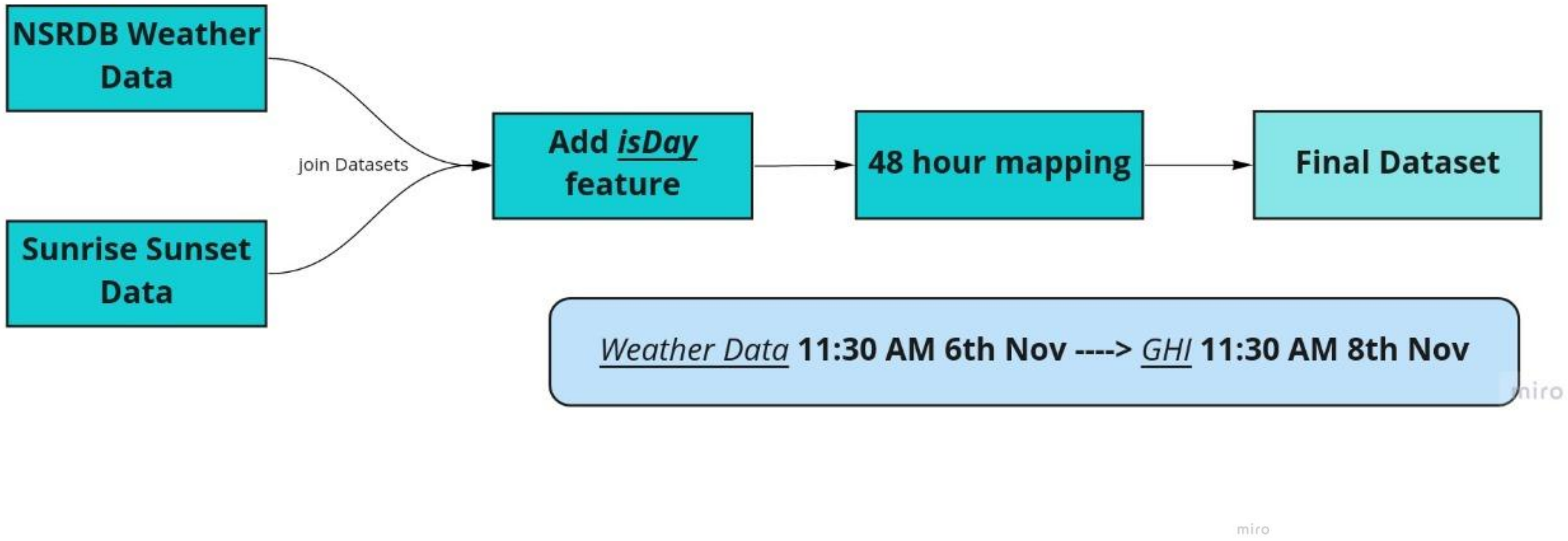
Month	Day	Hour
Minute	Temperature	Cloud Type
Fill Flag	Surface Albedo	Ozone
Pressure	Dew Point	Precipitable Water
Wind Direction	Wind Speed	Relative Humidity
Solar Zenith Angle		

Sunrise Sunset Data

Date	Sunrise Time	Sunset Time

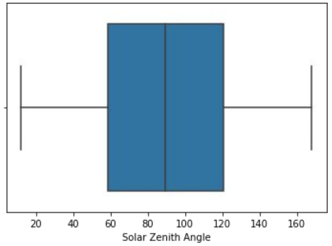
- Numerical Output : Solar Intensity ('GHI' in watts per square meter)

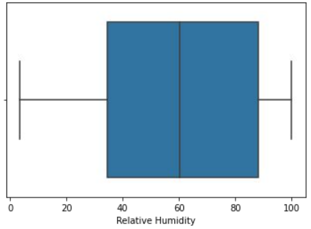
Dataset Preparation



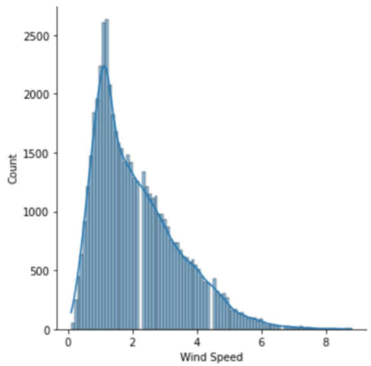
EDA

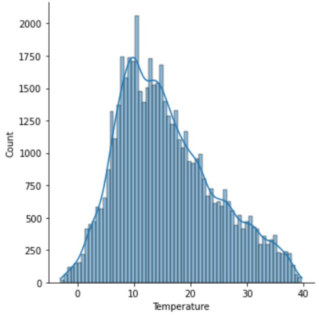
- Outlier detection using box plots





- Check skewness using histograms



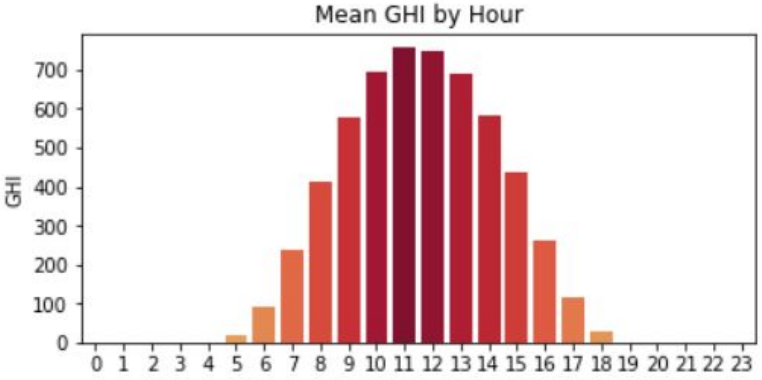


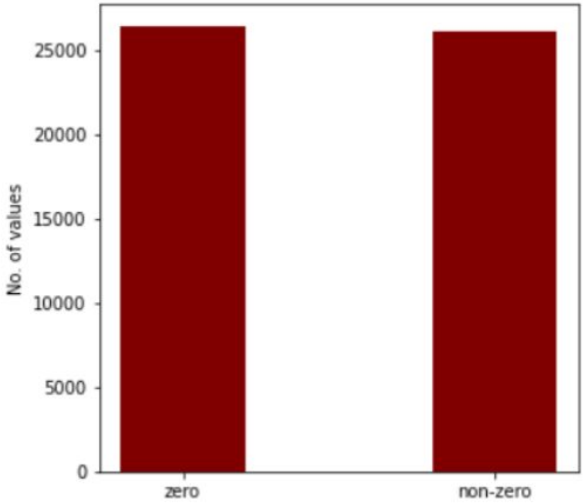
Skew Index for each feature

Feature	Skew Index
Month	-0.01
Day	0.07
Hour	0
Minute	0
Temperature	0.55
Cloud Type	1.30
Dew Point	-1.34
Fill Flag	3.96
Ozone	0.78
Relative Humidity	-0.091
Solar Zenith Angle	-0.00013
Surface Albedo	-0.71
Pressure	0.046
Precip. Water	0.62
Wind Direction	-0.73
Wind Speed	0.99
isDay	-0.038

EDA (contd.)

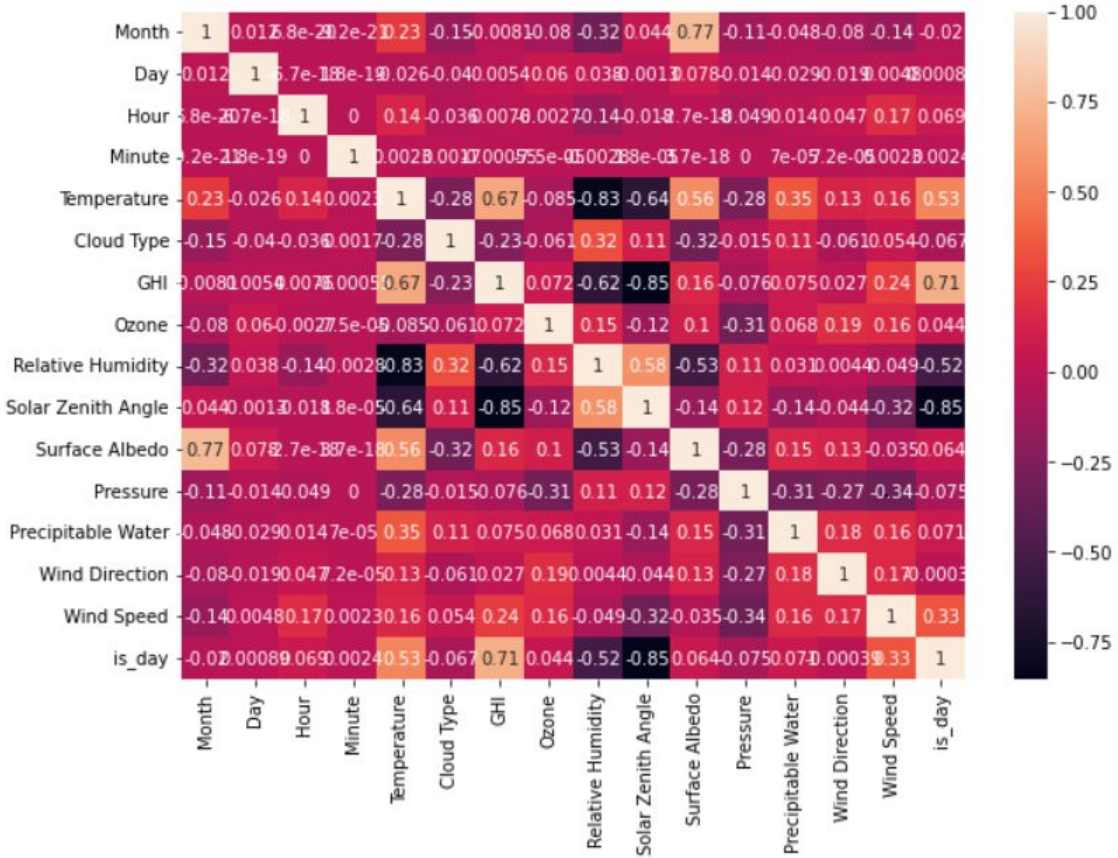
- Mean GHI by Hour & Mean GHI by month



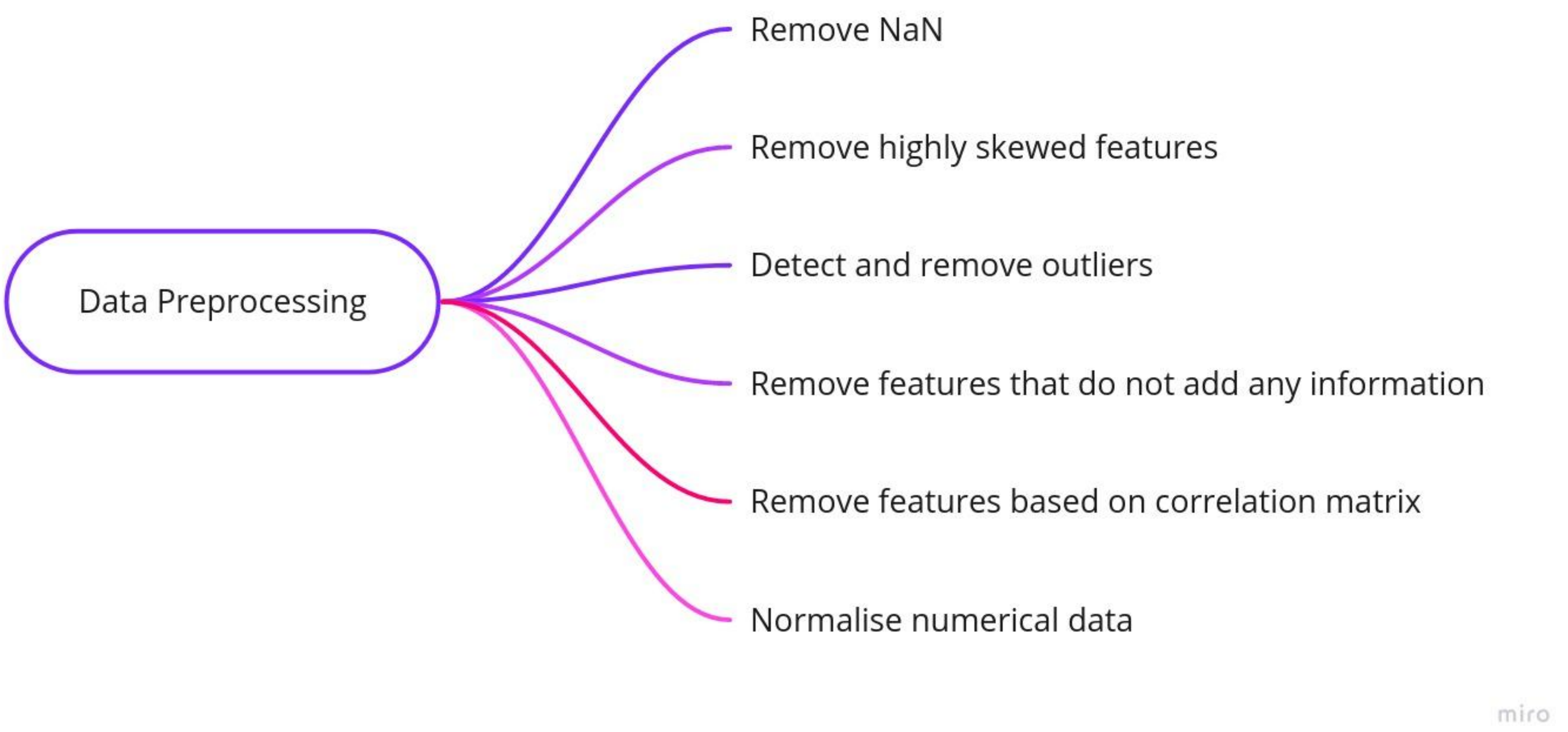


EDA (contd.)

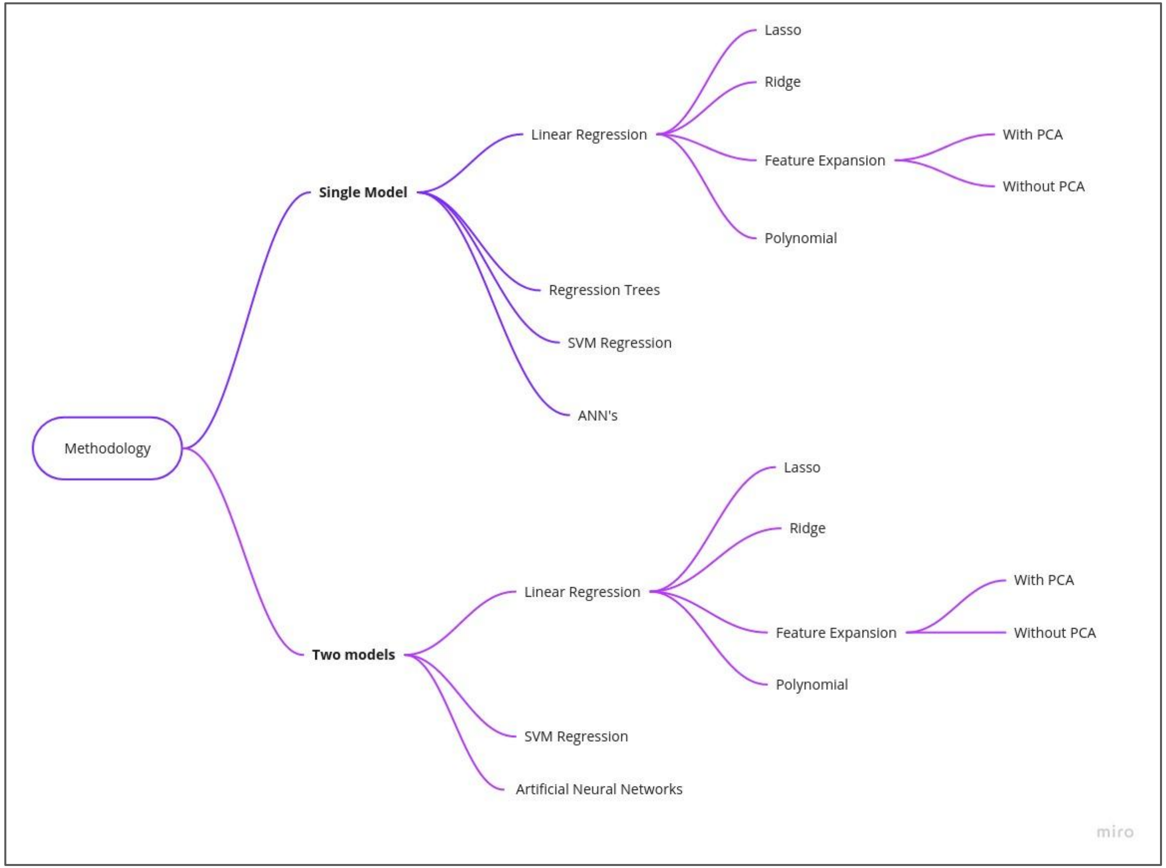
Correlation Matrix



Data Preprocessing



Methodology



Methodology 1

- Train data : 80 %
- Test data : 20 %
- Error : RMSE

Results & Analysis (Methodology 1)

Linear Regression

Baseline Model

Table 1: Single Model (Baseline)

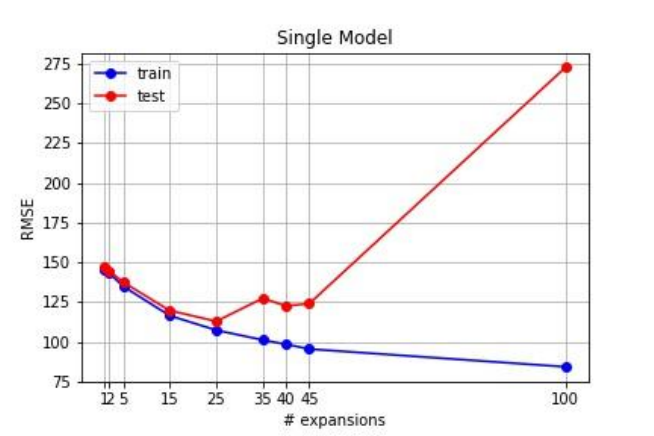
Model	Train RMSE	Test RMSE
Linear	151.554	152.019
Lasso	151.554	152.018
Ridge	151.554	152.017

Linear Regression

Feature Expansion w/o PCA

Best Performance:

25 Expansions



Linear Regression

Feature Expansion with PCA

Best Performance:

Expansions: 100, PCA Components: 250

Table 4: Single Model with PCA

Degree	PCA components	Train rmse	Test rmse
1	7	162.11	162.19
5	15	161.53	161.55
25	50	114.45	115.91
40	80	107.04	107.84
100	250	97.94	99.39

Polynomial Regression

Best Performance:

Degree: 4

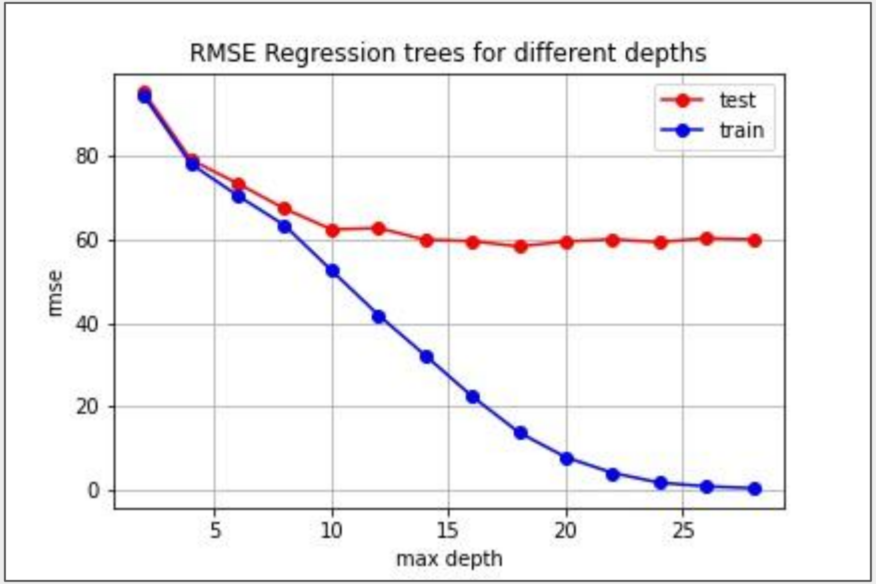
Table 7: Single Model

Polynomial Degree	Train RMSE	Test RMSE
2	78.95	79.96
3	69.49	70.53
4	51.98	55.99

Regression Trees

- **Best Performance:** max_depth = 10

Regression Tree	Train RMSE	Test RMSE
Without Pruning	0	59.9
Pre Pruning	52.6	62.4



SVM Regression

Best performance
kernel: rbf

Table 9: Single Model

Kernel	Train RMSE	Test RMSE
rbf	118.73	119.5
sigmoid	221.71	221.05
linear	153.62	153.97
polynomial	171.53	169.9

ANN's

Best Performance
Layers: 4
Layer Sizes: 14-64-32-1
All layers with LeakyRelu activation function

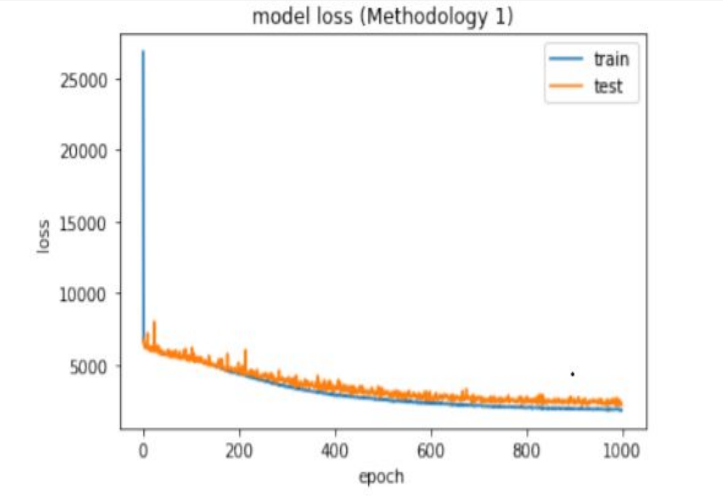


Table 11: Single Model(LeakyR means LeakyRelu)

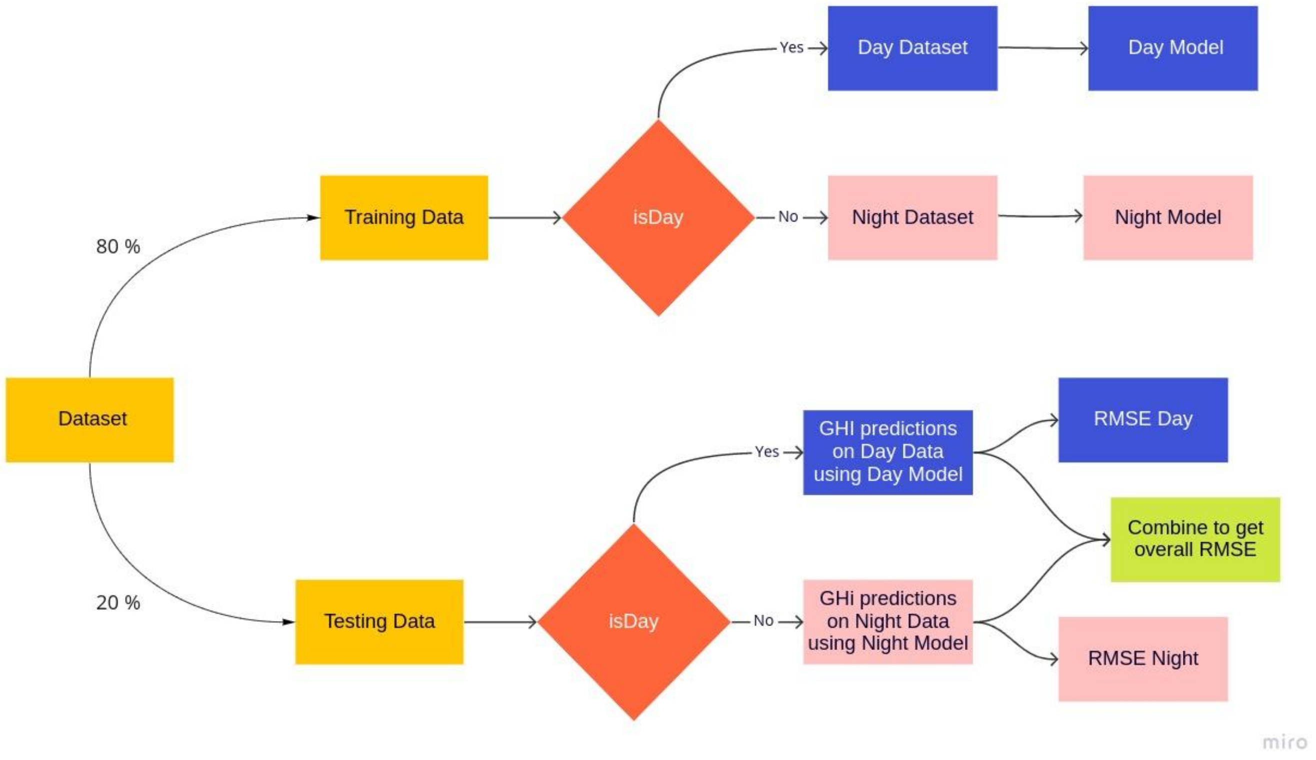
# Layers	Layer sizes	Activation Functions	Train RMSE	Test RMSE
4	14-64-32-1	LeakyR, LeakyR LeakyR, LeakyR	41.42	46.93
4	14-64-32-1	Relu, LeakyR LeakyR, LeakyR	54.21	56.69
4	14-64-32-1	Relu, Relu LeakyR, LeakyR	73.79	74.43

Drawbacks of Methodology 1

- So in this methodology we clearly see that as we are increasing the model complexity we are getting lower RMSE values on the train & the test set.
- But our aim is to get comparatively lower test RMSE values on less complex, simple ML models that we have used in approach 1.
- Thus to realise the above point, we have come up with an alternative approach which is discussed in the next section.

Methodology 2

Methodology 2 : What is the alternate approach ???



Results & Analysis (Methodology 2)

Linear Regression

Baseline Model

Table 1: Single Model (Baseline)

Model	Train RMSE	Test RMSE
Linear	151.554	152.019
Lasso	151.554	152.018
Ridge	151.554	152.017

Table 2: Day-Night Model (Baseline)

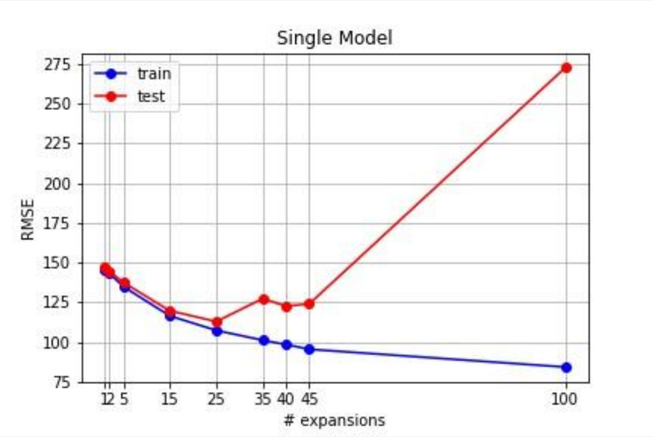
Model	Train RMSE	Test RMSE
Linear	81.135	82.182
Lasso	81.135	82.182
Ridge	81.137	82.178

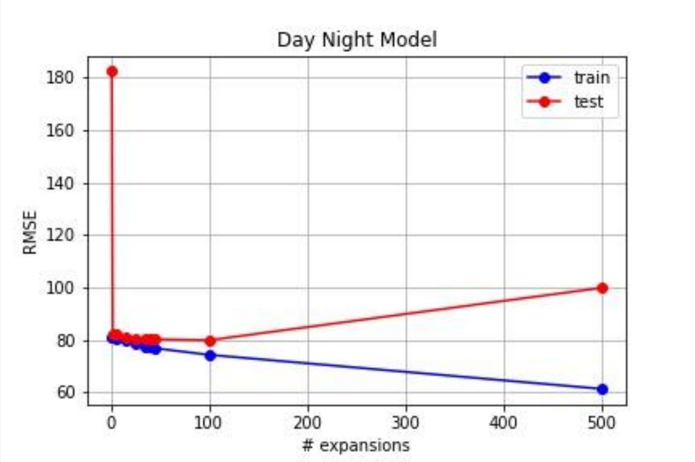
Linear Regression

Feature Expansion w/o PCA

Best Performance:

40 Expansions





Linear Regression

Feature Expansion with PCA

Best Performance:

Expansions: 100, PCA Components: 250

Table 4: Single Model with PCA

Degree	PCA components	Train rmse	Test rmse
1	7	162.11	162.19
5	15	161.53	161.55
25	50	114.45	115.91
40	80	107.04	107.84
100	250	97.94	99.39

Table 5: Day Night Model with PCA

Degree	PCA components	Train RMSE	Test RMSE
1	7	82.90	83.83
5	15	83.44	84.39
25	50	82.41	83.74
40	80	80.32	81.86
100	250	78.03	79.84

Polynomial Regression

Best Performance:

Degree: 3

Table 7: Single Model

Polynomial Degree	Train RMSE	Test RMSE
2	78.95	79.96
3	69.49	70.53
4	51.98	55.99

Table 8: Day Night Model

Polynomial Degree	Train RMSE	Test RMSE
2	75.66	76.81
3	63.36	65.58
4	45.11	55.63

SVM Regression

Best performance

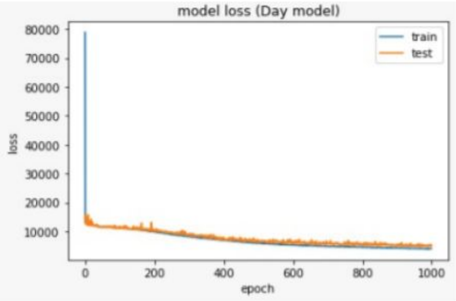
kernel: linear

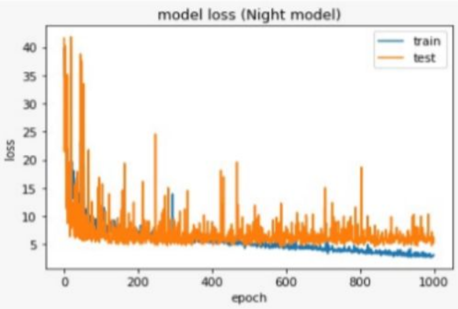
Table 9: Single Model

Kernel	Train RMSE	Test RMSE
rbf	118.73	119.5
sigmoid	221.71	221.05
linear	153.62	153.97
polynomial	171.53	169.9

Kernel	Train RMSE	Test RMSE
rbf	90.52	90.64
sigmoid	113.21	113.81
linear	85.63	87.71
polynomial	111.65	110.97

Table 10: Day Night Model





ANN’s

Best Performance

Layers: 4

Layer Sizes: 14-64-32-1

All layers with LeakyRelu activation function

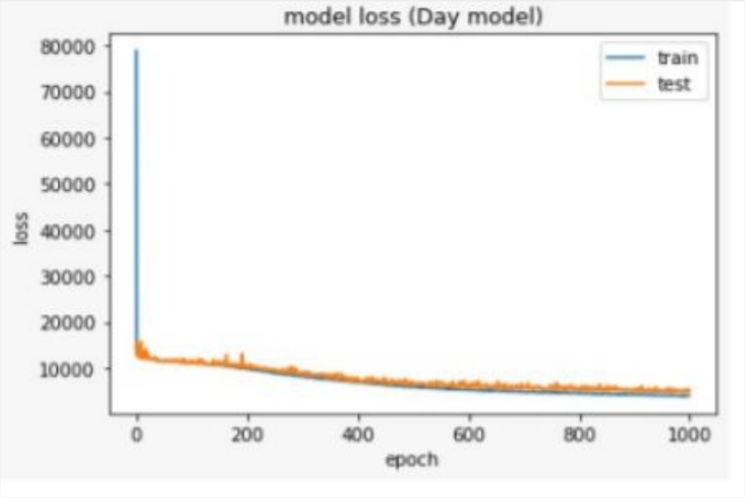
Table 11: Single Model(LeakyR means LeakyRelu)

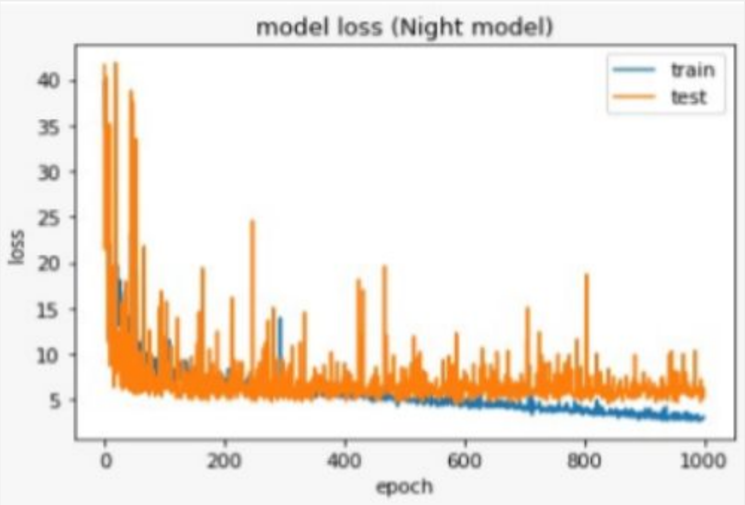
# Layers	Layer sizes	Activation Functions	Train RMSE	Test RMSE
4	14-64-32-1	LeakyR, LeakyR LeakyR, LeakyR	41.42	46.93
4	14-64-32-1	Relu, LeakyR LeakyR, LeakyR	54.21	56.69
4	14-64-32-1	Relu, Relu LeakyR, LeakyR	73.79	74.43

Table 12: Day Night Model(LeakyR means LeakyRelu)

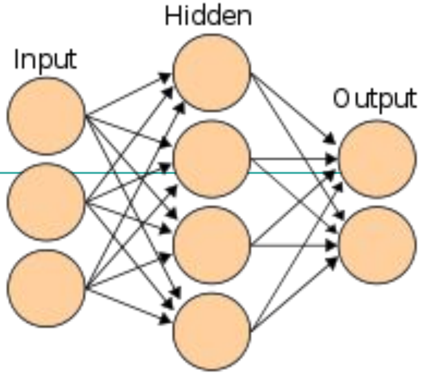
# Layers	Layer sizes	Activation Functions	Train RMSE	Test RMSE
4	14-64-32-1	LeakyR, LeakyR LeakyR, LeakyR	50.36	45.58
4	14-64-32-1	Relu, LeakyR LeakyR, LeakyR	59.95	53.33
4	14-64-32-1	Relu, Relu LeakyR, LeakyR	64.88	60.91

ANN Loss Plots





Conclusion



[wikimedia commons]

- ANN’s show best performance for both the type of methodologies
  - Drawbacks : High training time; Computationally expensive

- Also, we clearly saw that the single model scheme required complex ANN models to get sufficiently low train & test RMSE values.

- Thus, the Day-Night model scheme proved to be highly potent in achieving lower RMSE values using simpler models like LR, Regression Trees, SVM Regression etc, combined with feature expansion & PCA techniques, ,i.e, closer to the RMSE values that we got in method 1 with ANN models.

Future Work

Future work opportunities:

- Train dense neural networks
- Apply deep learning techniques such as LSTM's, recurrent neural networks etc.
- Apply clustering techniques
- Extend our predictions to more geographical locations

Member Contribution

Team Member	Contribution
Kushal Juneja	Data Preparation, EDA, Coming up with methodology II, Feature Expansion without PCA, Neural Networks, Regression Trees, Results Analysis, Report Making
Naval Kumar Shukla	Data pre-processing, Coming up with methodology II, Baseline LR (with Lasso and Ridge), Feature Expansion with PCA, Neural Networks, Results Analysis, Report Making
Rishi Singhal	EDA, Coming up with methodology II, Neural Networks, SVM Regression, Feature Expansion without pca, Polynomial Regression, Result Analysis, Report Making
Udit Narang	Data pre-processing, Coming up with methodology II, Baseline LR, Polynomial Regression, SVM Regression, Feature Expansion with PCA, Result Analysis, Report Making

Thank you