

Air Quality

This dataset was collected by scraping multiple sites using their permission. The aim of the project is to predict the air quality of the Bengaluru city based on the data collected from past years and to analyze whether how much of the population has grown since the Industrial revolution in India. Most urban cities suffer from the particulate matter pollution and Bengaluru being the tech city of India is no less polluted.

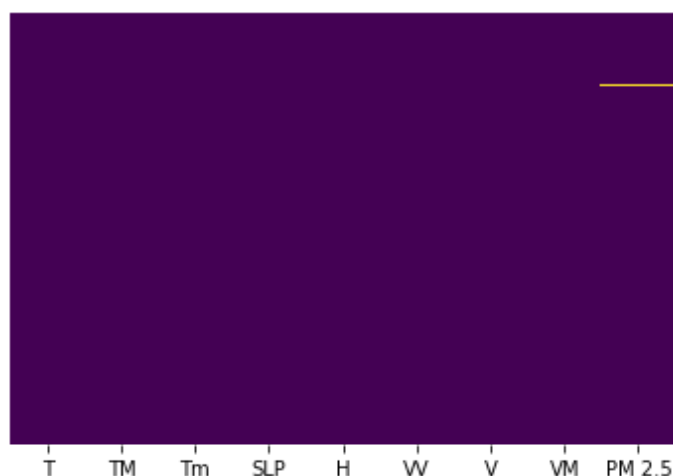
Using Neural networks along with hyperparameter tuning, aim is to predict the PM 2.5. Then from the best results the best candidate shall be chosen.

Data Summary:

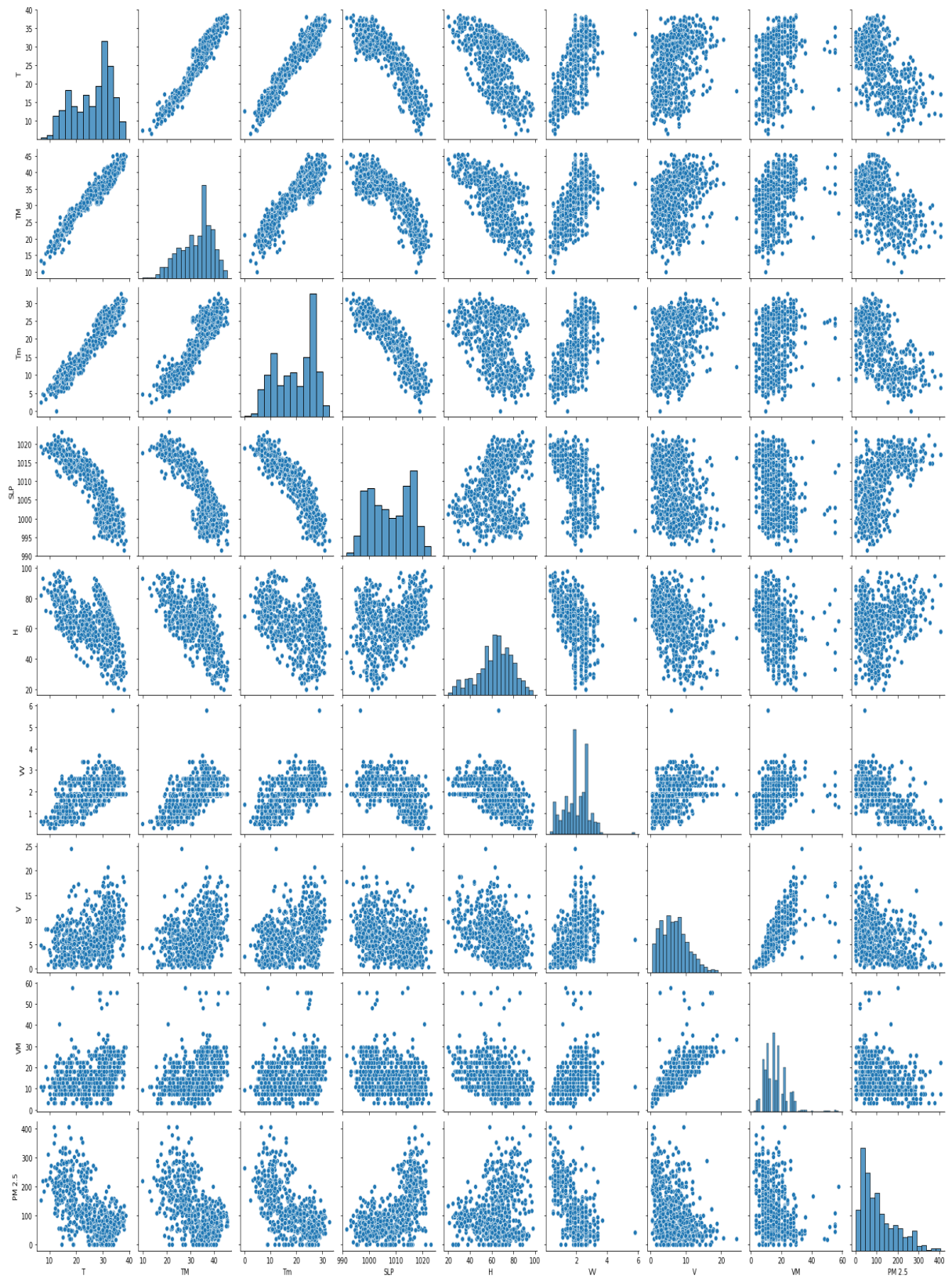
- 'T',
- 'TM',
- 'Tm',
- 'SLP',
- 'H',
- 'VV',
- 'V',
- 'VM',
- 'PM 2.5'

Data Exploration & Feature Engineering:

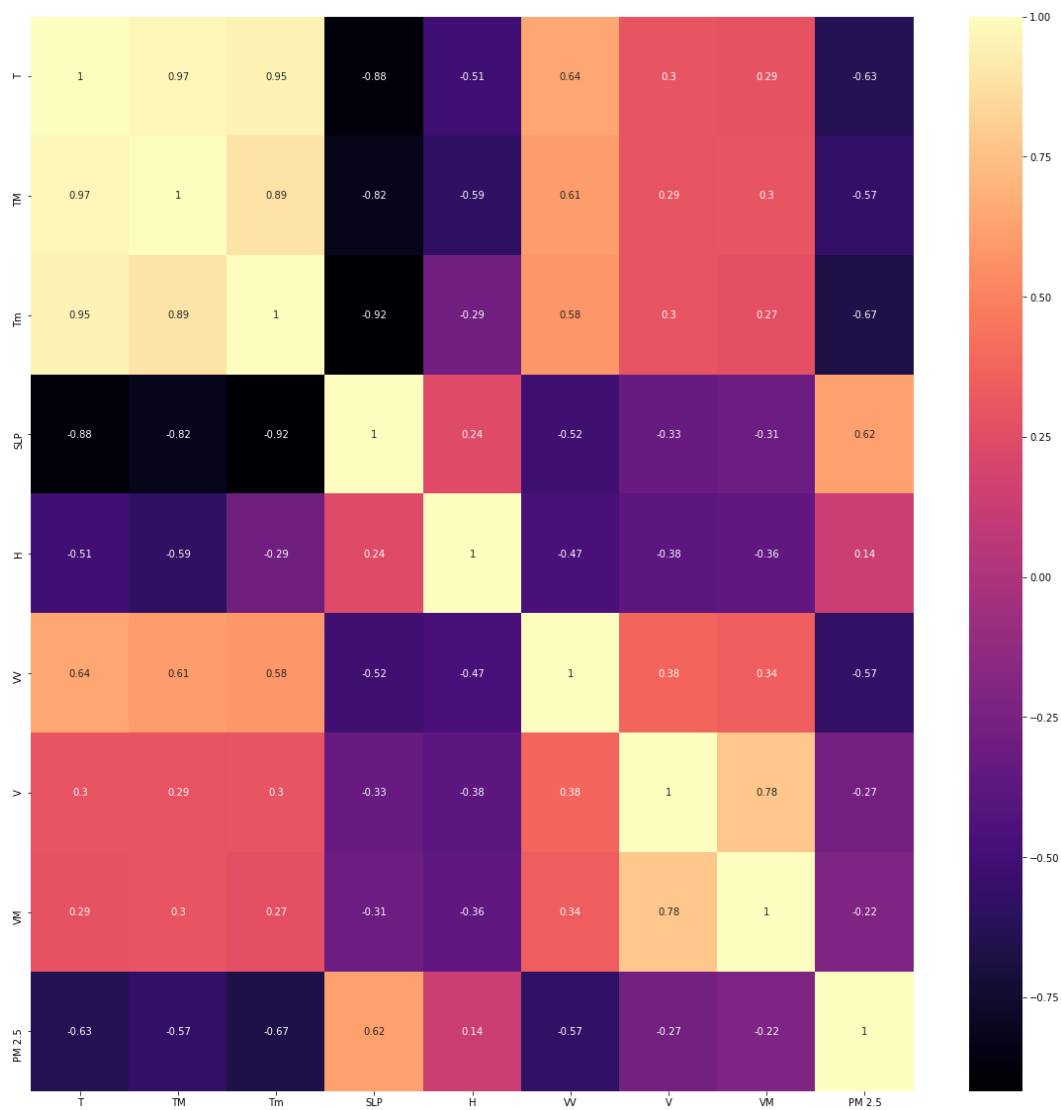
- While scraping and gathering data most of the null values had been dealt and the columns which were entirely null, were also dropped while collecting the data.
- Still there were two three rows in the target variable which had missing values, which were dropped.



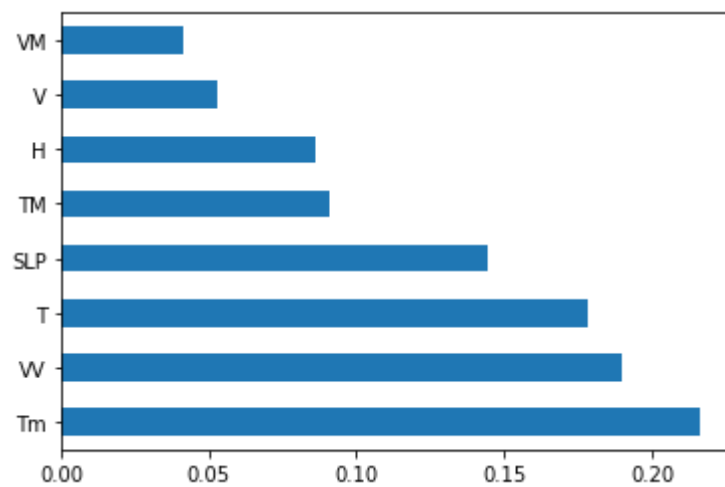
- As from the above it can be seen that most of the variables were not skewed and therefore there was no need for any changes,



- Columns like T, SLP, TM, are highly related and therefore the suitable columns were dropped, also keeping some multicollinearity in the dataset.



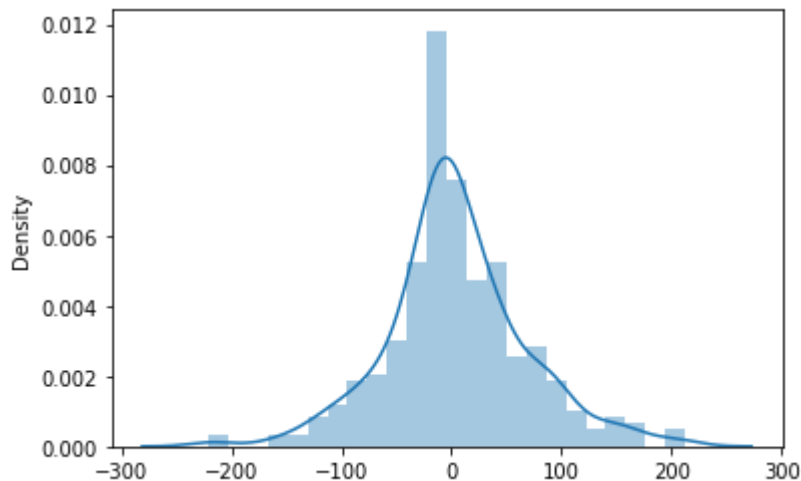
- Using the ExtraTreesRegressor feature importances were found which can be as seen from below:



Model Training and Testing:

- Neural network with different parameters were tested.

Conclusion:



The algorithms gave very closest Root-mean-squared error, however the difference in scores and errors are not significant, are actually almost identical. The best candidate based on rmse and score results would be neural network with softmax activation, therefore it is recommended to use this model as a final model that best fits the data in terms of accuracy.

Going Forth:

Can further try to optimize algorithms. To predict the PM 2.5 for an individual city, we could use other algorithms such as CNN, RNN or VCG etc, as well, even the non neural algorithms can be tested and tried whether they can give simple and efficient answers.