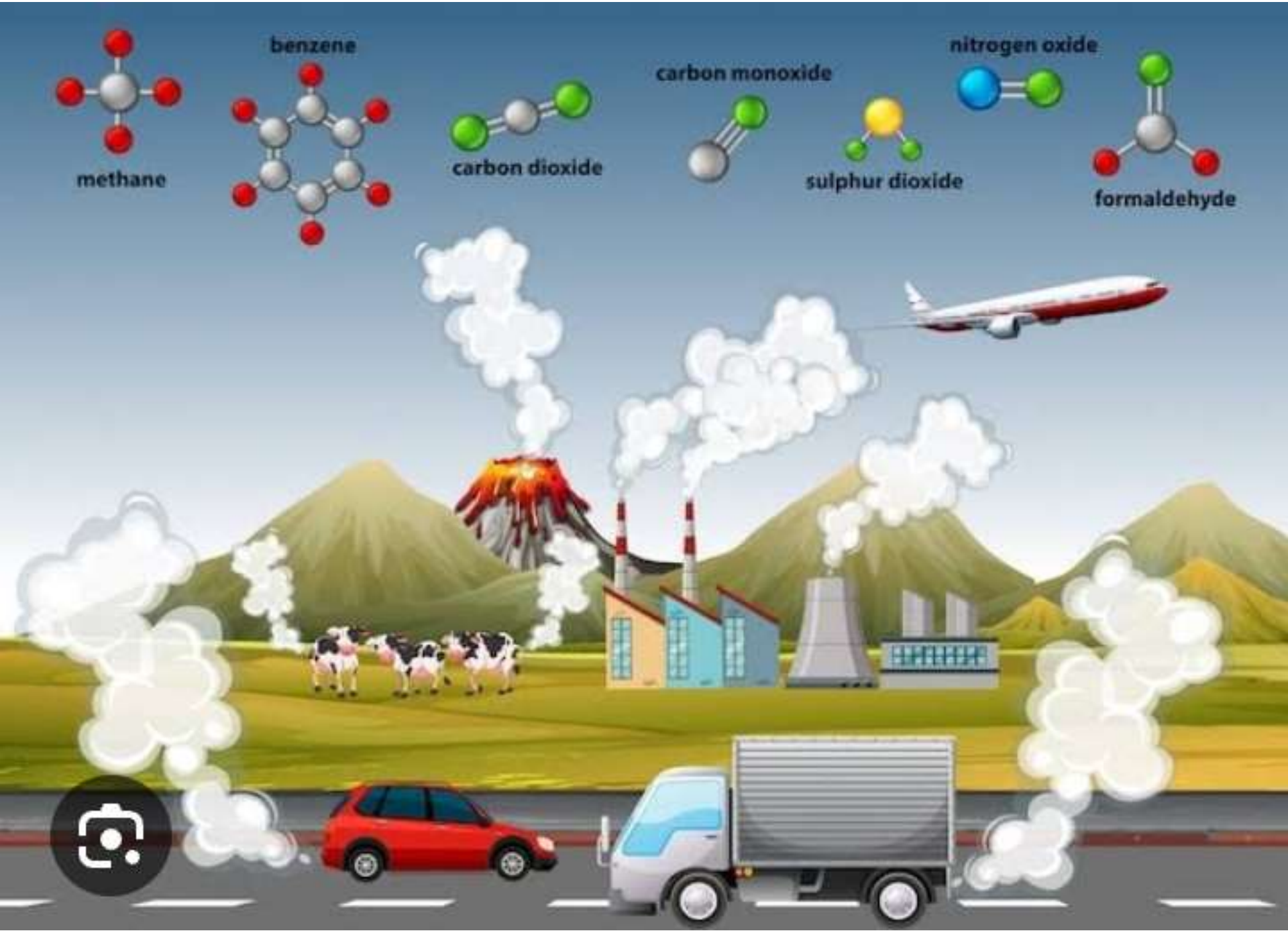
PHASE-2 PROJECT

Air quality analysis and prediction in TamilNadu

NAME:N.AGALYA.


methane
benzene
carbon dioxide
carbon monoxide
sulphur dioxide
nitrogen oxide
formaldehyde

Introduction:

. Tamil Nadu ,a state in South India experience Varying air quality levels due to industrial,Urban and natural factors.

.To perform air quality analysis and prediction several components are involved,Data collection, Data processing,Air quality index, Analysis and Modeling,Prediction and forecasting,visualization and reporting,public awareness and policy mitigation.

Linear regression:

- a linear regression model is a statistical approach used for modeling the relationship between a dependent variable (also known as the target or outcome) and one or more independent variables (predictors or features). The goal of linear regression is to find a linear relationship that best represents the data, making it a simple but fundamental tool for predictive modeling and understanding relationships within data.

Random forest:

- Random Forest is a popular machine learning algorithm that belongs to the ensemble learning category. It is used for both classification and regression tasks.
- **Gradient boosting:**
- Gradient boosting is a powerful machine learning technique used for both regression and classification tasks. It builds an ensemble of decision trees in a sequential manner, where each tree corrects the errors of its predecessor. Here's a brief overview of gradient boosting:

XG boosting:

- XGBoost, short for "Extreme Gradient Boosting," is a powerful machine learning algorithm that has gained popularity for its effectiveness in various predictive modeling tasks.
- **Reinforcement learning:**
- Reinforcement learning (RL) is a subfield of machine learning where agents learn to make decisions by interacting with an environment

Ensemble methods:

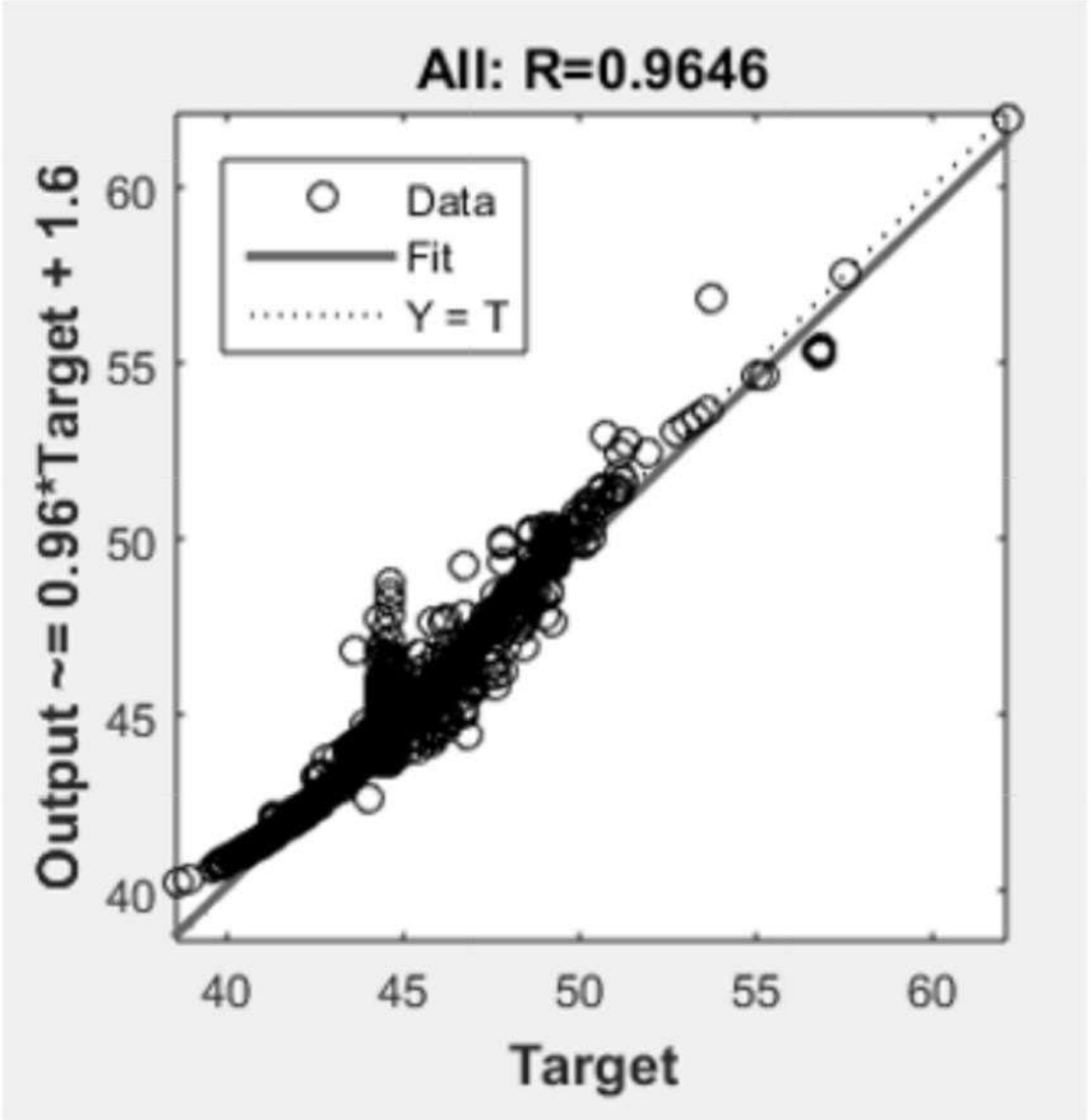
- Ensemble learning is a machine learning technique that combines the predictions of multiple models to improve overall predictive performance.
- Broadly this technique is classified into two
- Types namely
- Bagging
- Boosting.

Real-time Tamil Nadu
Most polluted city ranking

#	CITY	US AQI
1	Coimbatore	152
2	Ooty	107
3	Salem	105
4	Hosur	86
5	Namakkal	86
6	Vandalur	78
7	Tirunelveli	76
8	Arcot	74
9	Kattivakkam	72
10	Ariyalur	67

```
# Extracting Tamil Nadu state data alo
ne
tn = aq.query('state=="Tamil Nadu" ')
tn.sample(2)
```

	stn_code	sampling_date	state	location
355251	159.0	24/07/2012	Tamil Nadu	Chennai
353962	237.0	7/5/2011	Tamil Nadu	Coimbato


All: R=0.9646
Data
Fit
Y = T
Output ~= 0.96*Target + 1.6
Target
40
45
50
55
60

IV. CONCLUSION

In this paper, the prediction of $PM_{2.5}$ is done. Prediction is one of the application of artificial neural networks. The main objective of this research was to develop the model to predict of $PM_{2.5}$ in Alandur location based on data from monitoring stations. The developed model can be used as a decision making tool to create early warning about the pollution of air in the particular area. Based on the analysis, the model having PURELIN transfer function in the neural network structure produces the best performance in the prediction of air quality compared to the network structure that uses TANSIG transfer function based on the values of R and the prediction accuracy.