Supervised Classification Estimate towards Air Pollutant Quantification of Delhi and Udaipur

PriyanshaJain, Danish Paliwal, Aditya Maheshwari, Yogendra Singh Solanki

Abstract: ***The paper analyses air quality using supervised machine learning classifiers. The factorsconsidered for parameter selection towardsaffecting air quality are Benzene, BP(Barometer Pressure), PM_{10}(Particulate Matter), $PM_{2.5}$(Particulate Matter), RH(Relative Humidity), CO(Carbon Monoxide), NH_3(Ammonia), NO(Nitrogen Oxide), NO_2(Nitrogen Dioxide), NOx(Nitrogen Oxides), Ozone, SO2(Sulphur Dioxide).Curve fitting has been applied for analyzingpollutantsin air.***

Keywords:Air Quality, Particulate Matter, Nitogen Oxide, Carbon Monoxide, Sulphur Oxides.

I. INTRODUCTION

Air is ubiquitous and therefore it is quintessential to keep the air pollutant free for a healthy living. However, increase in pollutants is noxious. Yearly crop burning in northern India produces tons of crop residue evacuating smoke, Nitrogen Oxides, Sulphur Oxides, and PM to atmosphere. This leads to major haze and smog problems in winters especially in cities like Delhi. Emissions caused by vehicles leads to another source of air pollution [1]. Emissions by air pollutants by oxides of nitrogen, hydrocarbons and particulate matter leads to severe health effects like cancer, cardiovascular disease and respiratory illness[2].

Pollutants released in the atmosphere are:

Particulate Matter($PM_{2.5}$ and PM_{10}): Fine or tiny particles of solid or liquid released in gas. Naturally occurs through volcanoes, grassland and forest fires, living vegetation, sea spray and dust storms. Human activities like industrial processes, burning of fossil fuels and power plants generate aerosols.

Sulphur Dioxide (SO_2): Evacuated by industrial process and volcanoes. Its oxidation is a reason for acid rain [5].

Ammonia(NH_3): A type of gas having strong odor. It is hazardous and caustic.

Carbon Monoxide(CO): It is odorless, colorless and toxic gas. Produced throughincomplete combustion of fuel[3].Exhausts from vehicles are major source.

Ground Level Ozone (O_3): Formed by combining VOCs(Volatile Organic Compounds) and NOx (Nitrogen Oxides). Globally, this pollutant is at a high concern. High levels of ozone in atmosphere affect health of humans, problem in breathing, asthma aggravation and minimize functioning of lungs.

Exposure in high level of ozone can increase heart problems and daily impermanence [4].

Revised Manuscript Received on March 5, 2020.

***PriyanshaJain,** Techno India NJR Institute of Technology Udaipur, India. ***priyansha07jain@gmail.com**

Danish Paliwal, Techno India NJR Institute of Technology Udaipur, India

Aditya Maheshwari, Techno India NJR Institute of Technology Udaipur, India

Yogendra Singh Solanki, Techno India NJR Institute of Technology Udaipur, India

II. RESEARCH METHODOLOGY

An investigation has been done on air quality data of two cities of India that is Delhi and Udaipur. The data is fetched through 'Air Quality Monitoring Device'. To get the real time knowledge of the factors associated with air pollution a study was conducted on parameters such as Benzene, BP, PM10, PM2.5, RH, CO, NH3, NO, NO2, NOx, Ozone, SO_2.

The training dataset used is https://bit.ly/39QYmsN

The testing dataset is https://bit.ly/38MkZP5

Weka 3.8.3 tool is used to analyze the air quality.

III. REXPERIMENTAL RESULTS AND ANALYSIS

We have used various classifiers to classify data which includes RandomForest, LinearRegression, GaussianProcesses, SimpleLinearRegression, SMOreg and Random Sub Space. The results are as follows:

Table1: Training results using Weka 3.8.3

Classifiers	CC	MAE	RMSE	RRSE (%)	No. of Instances
Random Forest	0.9998	0.0036	0.0105	2.1033	728
Linear Regression	0.9993	0.0145	0.0187	3.7332	728
Gaussian Processes	0.9828	0.0748	0.094	18.809	728
Simple Linear Regression	0.9982	0.0259	0.0298	5.9527	728
SMOreg	0.9993	0.0143	0.019	3.797	728
Random Sub Space	0.9946	0.0404	0.0629	12.5717	728

CC: Correlation Coefficient, MAE: Mean Absolute Error, RMSE: Root Mean Squared Error, RRSE: Root Relative Squared Error

Table2: Training results using Weka 3.8.3

Classifiers	CC	MAE	RMSE	RRSE (%)	No. of Instances
Random Forest	0.9997	0.0045	0.0126	2.5166	60
Linear Regression	0.9996	0.0246	0.0273	5.4558	60
Gaussian Processes	0.9931	0.0921	0.1137	22.741	60
Simple Linear Regression	0.9999	0.027	0.0286	5.7197	60
SMOreg	0.9996	0.0241	0.0272	5.4457	60
Random Sub Space	0.9916	0.0462	0.0837	16.7376	60

CC: Correlation Coefficient, MAE: Mean Absolute Error, RMSE: Root Mean Squared Error, RRSE: Root Relative Squared Error

Table3: Deviation noted was:

Classifiers	CC	MAE	RMSE	RRSE (%)
Random Forest	0.0001	0.0009	0.0021	0.4133
Linear Regression	0.0003	0.0101	0.0086	1.7226
Gaussian Processes	0.0103	0.0173	0.0667	3.932
Simple Linear Regression	0.0017	0.0011	0.0012	0.233
SMOreg	0.0003	0.0098	0.0082	1.6487
Random Sub Space	0.003	0.0058	0.0208	4.1659

CC: Correlation Coefficient, MAE: Mean Absolute Error, RMSE: Root Mean Squared Error, RRSE: Root Relative Squared Error

The graphical representation of PM_{10}, $PM_{2.5}$, CO, SO_2, NH_3, NO and NO_2 is as follows:

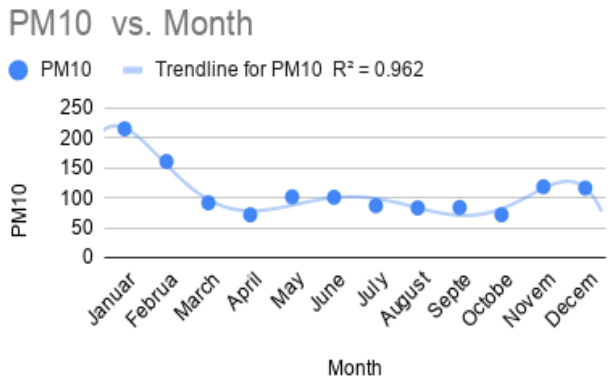

Fig1: Curve fitting (degree 6) of PM_{10} in Udaipur

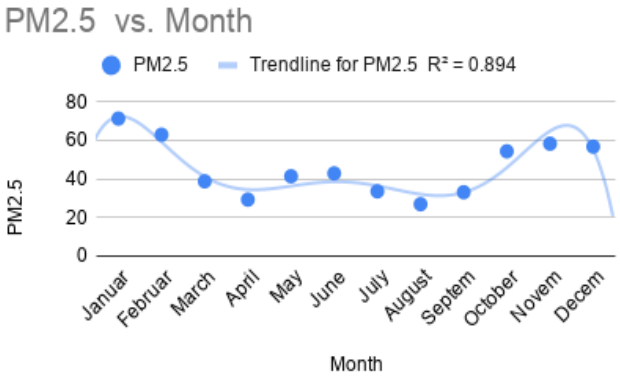

Fig2: Curve fitting (degree 6) of $PM_{2.5}$ in Udaipur

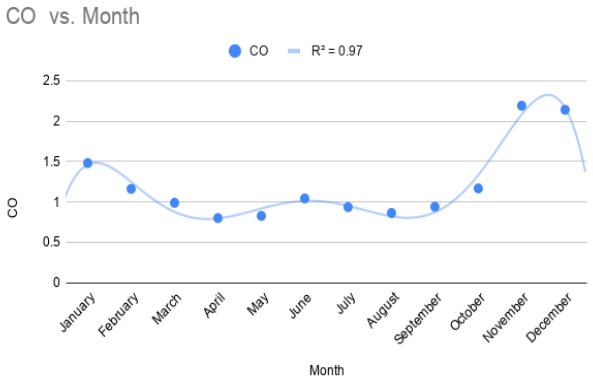

Fig3: Curve fitting(degree 6) of CO in Udaipur

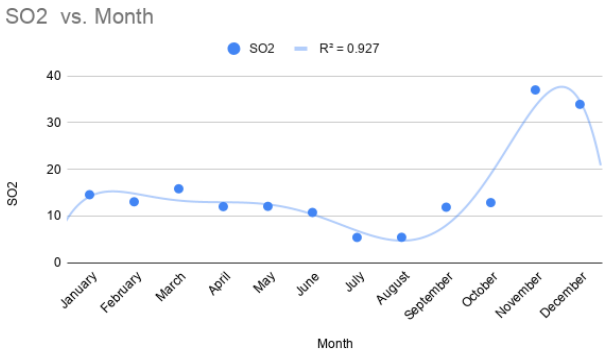

Fig4: Curve fitting (degree 6) of SO_2 in Udaipur

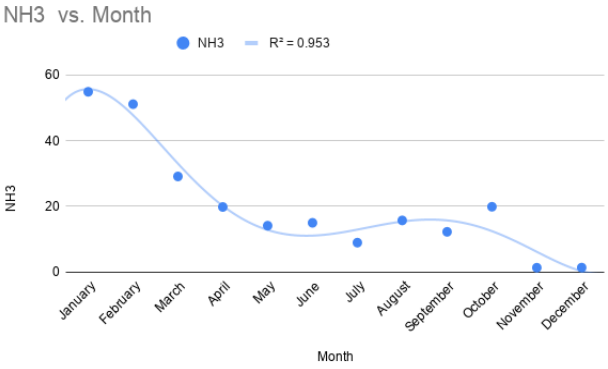

Fig5: Curve fitting(degree 6)of NH_3 in Udaipur

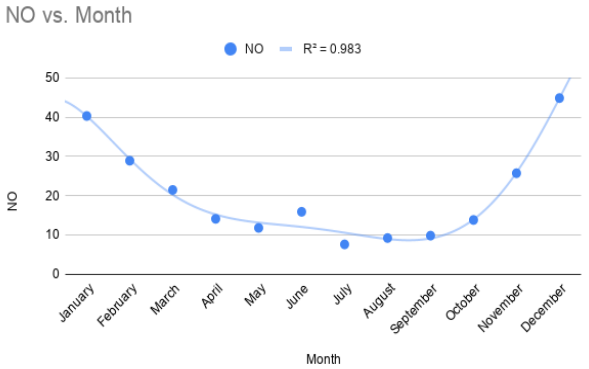

Fig 6: Curve fitting (degree 6) of NO in Udaipur

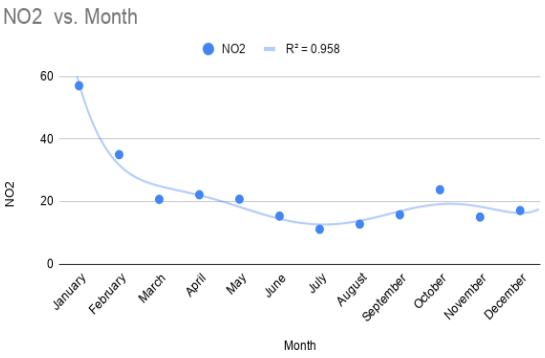

Fig7: Curve fitting(degree 6) of NO_2 in Udaipur

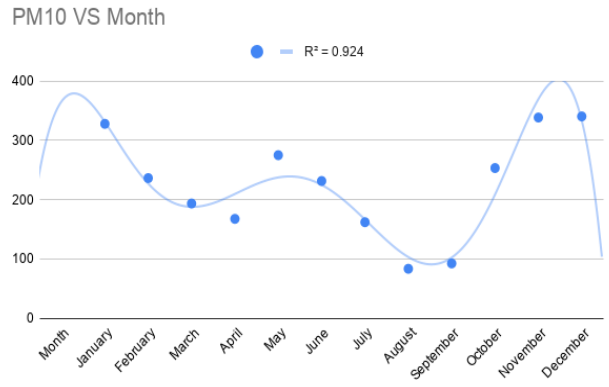

Fig8: Curve fitting(degree 6) of PM_{10} in Delhi

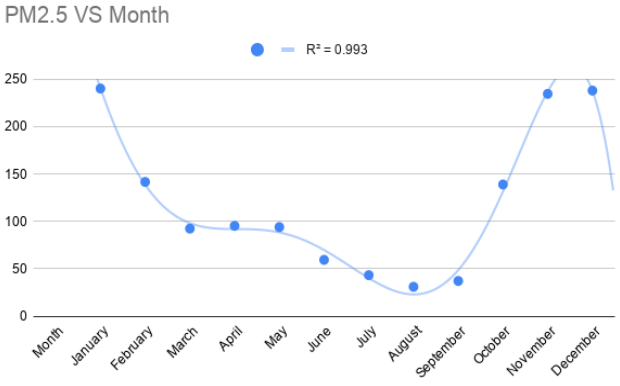

Fig9: Curve fitting (degree 6)of $PM_{2.5}$ in Delhi

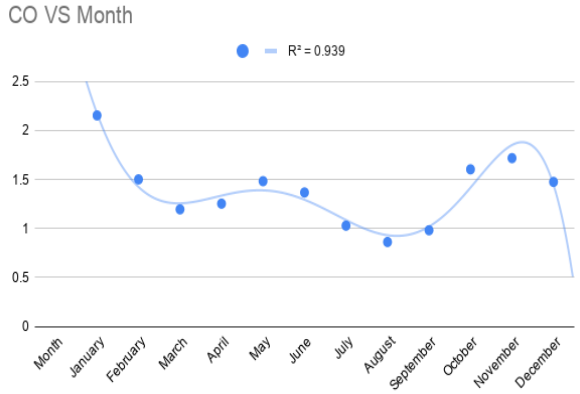

Fig10: Curve fitting(degree 6)of CO in Delhi

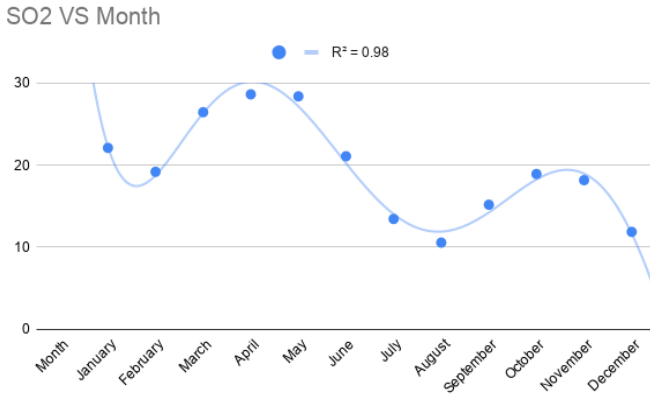

Fig11: Curve fitting(degree 6) of SO_2 in Delhi

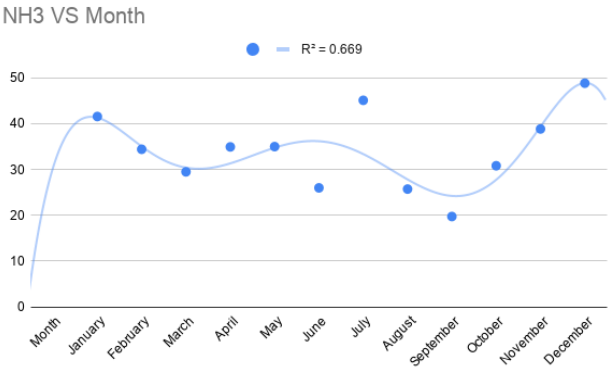

Fig12: Curve fitting(degree 6) of NH_3 in Delhi

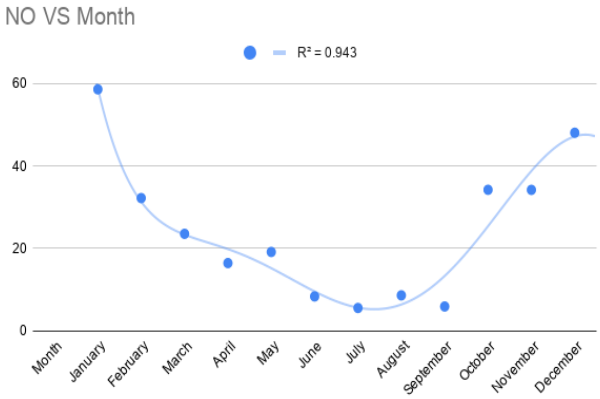

Fig13: Curve fitting(degree 6) of NO in Delhi

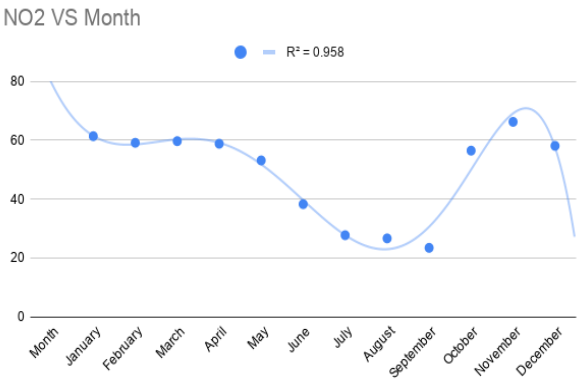

Fig14: Curve fitting(degree 6) of NO_2 in Delhi

The observationsare as follows:

1. **PM_{10}:**The concentration level in Udaipur city was high in the beginning of the year. Then it slip back in middle months. However, it augmented in the end of yearwhereas in Delhi, the concentration level was higher as compared to Udaipur in the beginning of year. It went through great fluctuations in middle years. However, it reached its epitome in November 2019, which was nearly 4 times as compared to Udaipur.
2. **$PM_{2.5}$:** The concentration level in Udaipur was $1/3^{rd}$ in the concentration level in Delhi in the beginning of the year. However, the level nadir in both the cities till August and from next month, the level

augmented and reached its epitome in November 2019.

3. **CO:** The concentration level was low in Udaipur as compared to Delhi in the beginning of the year.
4. Moreover, the levels in both the cities mitigate. It had a gradual fluctuation till August and from next month the level rose. In Udaipur, the highest level went in November 2019 however in Delhi it went in January 2019.
5. **SO_2:** The concentration level was low in Udaipur from the beginning of the year and had gradual fluctuations till August month and augmented in the following month. It reaches its peak in November month. However, in Delhi it started with high concentration as compared to Udaipur. It slips back in the next month but augmented in March 2019which was highest.
6. **NH_3:**The concentration level in Udaipur was highest in the beginning of the year. However, the level mitigates and had certain fluctuations in later months of the year. On the contrary, Delhi had certain fluctuations till September and was highest in December 2019.
7. **NO:**The concentration level in Udaipur was low as compared to the Delhi in the beginning of the year. Both cities level nadir till September month and then augmented in the following month. Both the cities had a highest level concentration in December 2019.
8. **NO_2:** The concentration level in Udaipur was highest in the year start. However, it just went down with certain fluctuations till the year end. On the contrary, it started with high concentration level but fell down till September 2019 and was highest in November 2019.

IV. CONCLUSION

The research finding is based on the simulation results after applying machine learning classifiers reflects that RandomForest generates optimum accuracy estimate (99.99%). The concentration level of Pm_{10}, PM $_{2.5}$, CO, SO_2, NO, NO_2 has been compared with respect to each region for Udaipur and Delhi.

REFERENCES

1. Dharmendra Kumar "Source, factors, and effect of air pollution on human lifestyle in India "
2. Vehicle emission (Green House gas emissions) greenvehicleguide.gov.au/pages/Information/VehicleEmissions
3. Financial Express "Delhi pollution : Carbon Monoxide among primary pollutants ; what is it and why it's a cause for alarm.
4. Balajeekarthik "Characteristics of the Ozone pollution and its health effects in India"
5. Shagun Kapil "India emits the most sulphur dioxide in the world"

AUTHOR PROFILE

Priyansha Jain is currently pursuing B.Tech. in computer science from Techno India NJR Institute of Technology. She is also pursuin
g graduation in Business Anlytics recognised by IBM.

Danish Paliwal is a final year student pursuing B. Tech in Information Technology, Techno India NJR Institute of Technology, Udaipur. She is also pursuing specialization in Business Analytics certified by IBM.

Aditya Maheshwari is working in the capacity of Project Lead- New Initiative & Research at Techno India NJR Institute of Technology Udaipur.
.

Yogendra Singh Solanki, is working as Asst. Professor, Department of Electronics and Communications, Techno India NJR Institute of Technology, Udaipur. He is practicing the IoT Devices, AI/ML and Data Science in Medical Analytics.

Prof DrPrasun Chakrabarti has received his PhD (Engg) from Jadavpur University in 2009. He is working as Executive Dean (Research and International Linkage) and Institute Distinguished Senior Chair Professor, Techno India NJR Institute of Technology.