Forecasting of Various Air Pollutant Parameters in Bangalore Using Naïve Bayesian

[1]Shivkumar M, [1]Sudhindra K R, [2]Pranesha T S, [3]Chate D M, [3]Beig G
[1]*Department of Electronics and Communication Engineering, BMS College of Engineering, Bengaluru-560019, India*

[2]*Department of Physics, BMS College of Engineering, Bengaluru-560019, India.*
Email: praneshats.phy@bmsce.ac.in
[3]*Indian Institute of Tropical Meteorology, Pune-411008, India* chate@tropmet.res.in

Abstract
Weather forecasting is considered to be of utmost important among various important sectors such as flood management and hydro-electricity generation. Although there are various numerical methods for weather forecasting but majority of them are reported to be Mechanistic computationally demanding due to their complexities. Therefore, it is necessary to develop and build models for accurately predicting the weather conditions which are faster as well as efficient in comparison to the prevalent meteorological models. The study has been undertaken to forecast various atmospheric parameters in the city of Bangalore using Naïve Bayes algorithms**.** The individual parameters analyzed in the study consisted of wind speed (WS), wind direction (WD), relative humidity (RH), solar radiation (SR), black carbon (BC), radiative forcing (RF), air temperature (AT), bar pressure (BP), PM_{10} and $PM_{2.5}$ of the Bangalore city collected from Air Quality Monitoring Station for a period of 5 years from January 2015 to May 2019. The study concluded that Naive Bayes is an easy and efficient classifier that is centered on Bayes theorem, is quite efficient in forecasting the various air pollution parameters of the city of Bangalore.
Keywords: *Forecasting, Air pollutant, Naïve Bayes*

1. Introduction

The

Air pollution is referred to as the release of dangerous air pollutants into the earth's atmosphere that can affect the health of humans and other creatures as well as lead to environmental damage (Ghorani-Azam, Riahi-Zanjani, & Balali-Mood, 2016). Various life threatening diseases such as cardiac disease, cancer, and respiratory problems may develop due to air pollution (Turner et al., 2011). Children in the age group of 0–5 years are the one that are mostly affected by the air pollution. They are reported to suffer from asthma, pneumonia and other respiratory diseases due to the exposure to air pollution (Hulin, Caillaud, & Annesi-Maesano, 2010). Out of all the vital elements required to sustain life, air is the most crucial factor for the survival of life on the planet. Due to increased economic and industrial development, there has been a massive change observed in the quality of air in all major cities of the world (Chow, 1995). The difficulty involved in monitoring the quality of air due to complex statistical calculations, it becomes hard for the administrators and policy makers to mark the level of safety in the air quality. The air quality of a certain area can be checked and improved by involving educated individuals having expertise in the regional as well as national air pollution related issues (Bao, Lu, & Shang, 2004). Various air pollutants such as NO2, black carbon (BC), SO_2, CO_2, particulate matters PM_{10} and $PM_{2.5}$ scientifically established to worsen the air quality. Among all of them, particulate matter $PM_{2.5}$ has been reported to be the most detrimental for which it is also referred to as slow poison. $PM_{2.5}$ is reported to have an aerodynamic diameter of $1\mu m<\varphi PM_{2.5}\leq 2.5\mu m$ (Davidson, Phalen, & Solomon, 2005; Li et al., 2018; Liu & Peng, 2018; Pope et al., 2018). Long term exposure to such harmful particulate air pollutant can have multiple health hazard leading to serious diseases such as lung cancer and cardiac disease that may lead to hospital admissions (Borja-Aburto, Castillejos, Gold, Bierzwinski, & Loomis, 1998; Schweitzer & Zhou, 2010; Xing, Xu, Shi, & Lian, 2016). Therefore, bringing down the level of $PM_{2.5}$ is a huge challenge to the policy makers and environmentalist which if done successfully will be considered to be a huge achievement for the pollution control boards. (Fattore et al., 2011), (Dunea et al., 2016), (Mehdipour & Memarianfard, 2017), and (Mehdipour, Stevenson, Memarianfard, & Sihag, 2018) concluded in their study that research on the reduction of $PM_{2.5}$ and other air pollutants are of great importance and necessity which have an positive correlation with the health impact among the people.

Various numerical methods have been developed and undertaken in the recent years that are focused on weather forecasting. In the recent times many statistical methods have been developed in order to process and analyze operational Numerical Weather Prediction so as to reduce errors associated with forecasting which needs significantly huge training data sets. There are many drawbacks associated with such methods one of which is the that such huge datasets are comparatively constrained by computation practicality (De Kock, Le, Tadross, &

Manuscript received March 5, 2024
Manuscript revised March 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.3.23

Potgeiter, 2008). Sample size is another important factor while performing forecasting as large sample size would allow to properly classify the dataset and variables with more number of cases

2. Measurement Site and Techniques

Bengaluru (120 58' N; 770 34' E; Average 900 MASL), is the capital city of Karnataka and IT sector of India which leads to increase in population. The whole dataset of Bangalore city (Karnataka) collected from Air Quality Monitoring Station for 5 year term from January 2015 to May 2019. The sample contains daily observations at various time intervals. The dataset has been checked and cleaned for observations containing either missing or null values following which the final dataset has been prepared and used for forecasting. The individual parameters analyzed in the study consisted of wind speed (WS), wind direction (WD), relative humidity (RH), black carbon (BC), radiative forcing (RF), solar radiation (SR), air temperature (AT), bar pressure (BP), PM_{10}, and $PM_{2.5}$.The dataset for the study is cleaned by omitting the missing values, duplicated and outlier values which are then filtered and made ready for the data mining process. The data is prepared for the study by transforming the final dataset into a csv format so as to make it ready for mining process. The AQMS is situated on the top of the science block working under the BMS College of Engineering, Bengaluru supported by Indian Institute of Tropical Meteorology, Pune, Govt. of India.

3. Bayesian Algorithm Naive

Bayesian algorithm is considered to be among one of the best and efficient is one of the most effective classification algorithms. This theorem was originally formulated by Reverend Thomas Bayesian. The Bayesian algorithm takes into consideration the probability of a hypothesis as a function of previous and new findings. The Bayesian algorithm takes into consideration how true the probability of a theory stands when introduced with new set of data and information. The theory has found its application in a diverse number of sector starting from marine biology to developing spam blocking algorithm for emails. There are many such features which make the Naive classifiers so desirable: Firstly, the Naïve classifier is easy to construct, which requires minimal knowledge among the background in comparison to general Bayesian networks that demands expertise in the relevant field in order to come out with fruitful solutions. Secondly, the naive based methods requires very little space and time intricacy (Dash & Cooper, 2002). Bayesian theorem is a method of calculating the hypothesis probability where Y is the event and X is the training dataset

$$p(Y \backslash X) = \frac{p(X \backslash Y)\, p(Y)}{p(X)}$$

This has been fund to have great applications in various fields. It is simple to estimate the probability of P(X | Y), P(Y), P(X) when the probability P(Y | X) is the most important and necessary. The theorem is very fundamental to Bayesian statistics that estimates the probability of any new findings based on data from earlier records (Chen, Yu, & GUAN, 2004; Feng, 2001).

4. Results

In this work the forecasting of the various weather parameters has been performed using Naïve Bayes algorithms using Python. The naïve forecasts of the various parameters in the study are as follows:

Figure 1 The prediction of the forecasting trend of AT (denoted in orange line).

It has been observed from the study that the forecasting for AT in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is °C and X-axis is date time. The maximum temperature predicted for during the forecasted period was observed to be 34 °C.

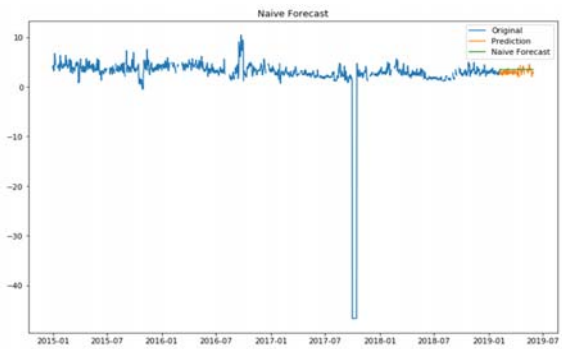


Figure 2 The prediction of the forecasting trend of BC (denoted in orange line).

It has been observed from the study that the forecasting for BC in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is µg/m³ and X-axis date time. The black carbon predicted for during the forecasted period was depicted in the figure 2.

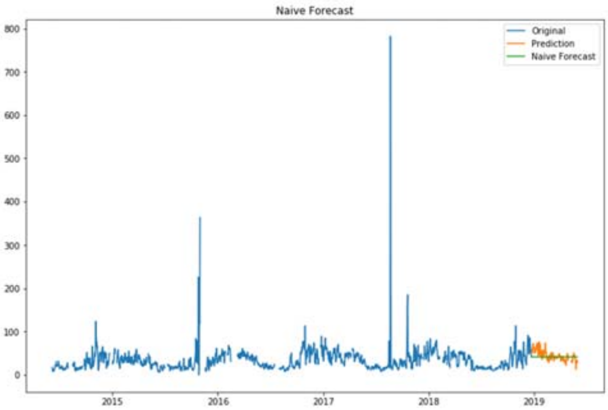


Figure 3 The prediction of the forecasting trend of $PM_{2.5}$ (denoted in orange line).

It has been observed from the study that the forecasting for P2.5 in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is µg/m³ and X-axis date time. The P2.5 predicted for during the forecasted period was depicted in the figure 3.

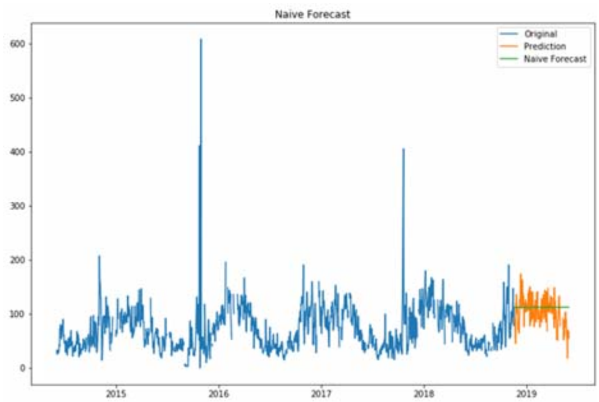


Figure 4 The prediction of the forecasting trend of PM_{10} (denoted in orange line).

It has been observed from the study that the forecasting for P10 in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is µg/m³ and X-axis date time. The P10 predicted for during the forecasted period was depicted in the figure 4 where it was found to get increased with time.

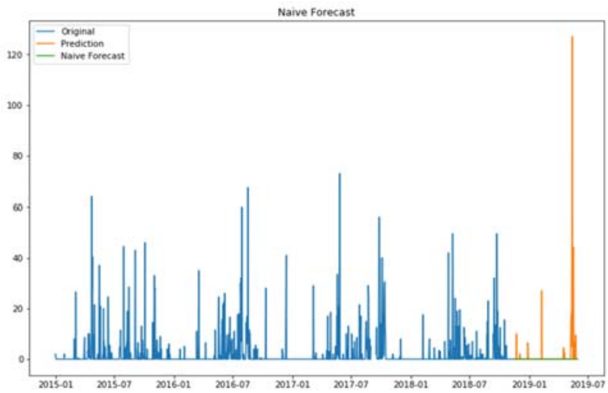
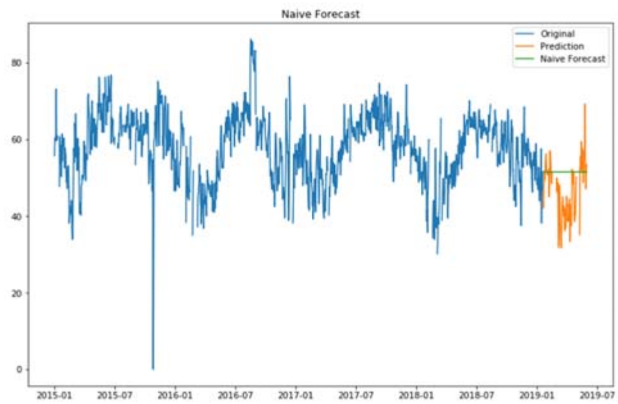


Figure 5 The prediction of the forecasting trend of RF (denoted in orange line).

It has been observed from the study that the forecasting for RF in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is mm and X-axis date time. The RF predicted for during the



forecasted period was depicted in the figure 5 where the maximum peak was observed to be above 120mm during the forecasted period.

Figure 6 The prediction of the forecasting trend of RH (denoted in orange line).

It has been observed from the study that the forecasting for RH in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is % and X-axis date time. The RH predicted for during the forecasted period was depicted in the figure 6 where the maximum peak was observed to be around 70% during the forecasted period. The forecasted RH was found to decrease initially followed by increased RH during the later part of the forecasted time period.

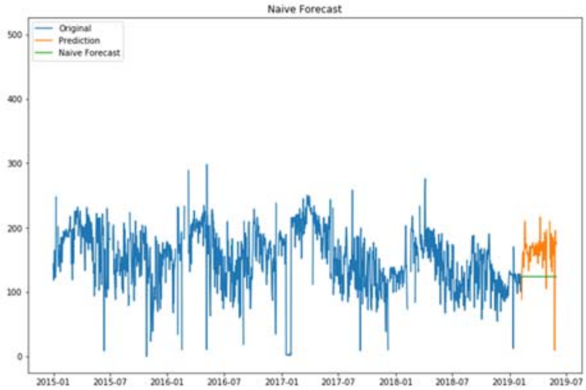


Figure 7 The prediction of the forecasting trend of SR (denoted in orange line).

It has been observed from the study that the forecasting for SR in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is W/m^2 and X-axis date time. The SR predicted for during the forecasted period was depicted in the figure 7 where the forecasted SR was found to increase initially followed by decrease SR during the later part of the forecasted time period.

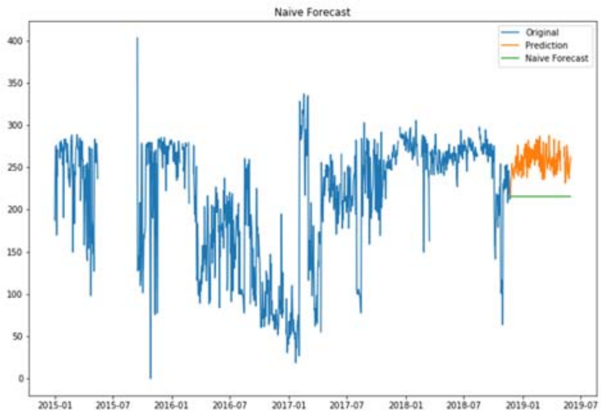


Figure 8 The prediction of the forecasting trend of WD (denoted in orange line).

It has been observed from the study that the forecasting for WD in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is Degree and X-axis date time. The WD predicted for during the forecasted period was depicted in the figure 8 where the forecasted WD was found to be around 220 to 280 degree approximately.

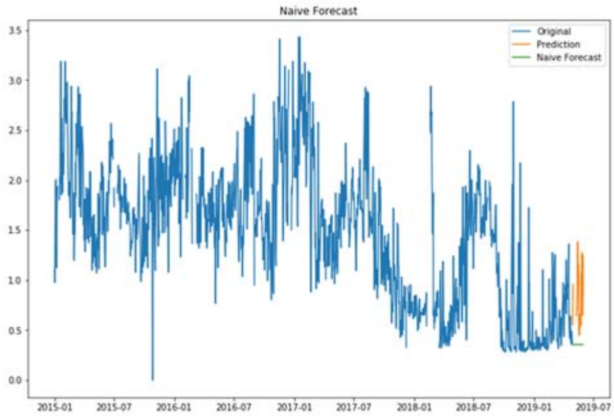


Figure 9 The prediction of the forecasting trend of WS (denoted in orange line).

It has been observed from the study that the forecasting for WS in the city of Bangalore follows the usual trend and varies around naïve prediction. The unit of Y-axis is m/S and X-axis date time. The WS predicted for during the forecasted period was depicted in the figure 9 where the forecasted WS was found to fluctuate around 0.4 to 1.3 m/S approximately.

4. Conclusion

The study performed forecasting on the various weather parameters using algorithms based on Naive Bayes. Naive Bayes is reported to be one of the most simple, powerful and efficient classifiers which uses the Bayes theorem, for performing forecasting calculations on the various weather parameters of the city of Bangalore.

Acknowledgment

The authors from BMS College of Engineering thank Prof. Ravi S. Nanjundiah, Director IITM, for choosing their college as a MAPAN station. They also thank the Principal of the college for encouragement.

References

1. Bao, C.-k., Lu, Y.-s., & Shang, J.-C. (2004). Framework and operational procedure for implementing strategic environmental assessment in China. *Environmental Impact Assessment Review, 24*(1), 27-46.
2. Borja-Aburto, V. H., Castillejos, M., Gold, D. R., Bierzwinski, S., & Loomis, D. (1998). Mortality and ambient fine particles in southwest Mexico City, 1993-

1995. *Environmental health perspectives, 106*(12), 849-855.

3. Chen, B., Yu, J., & GUAN, H.-w. (2004). Feasibility study of applications of data mining to weather forecast. *Applied Science and Technology, 31*(3), 48-50.
4. Chow, J. C. (1995). Measurement methods to determine compliance with ambient air quality standards for suspended particles. *Journal of the Air & Waste Management Association, 45*(5), 320-382.
5. Dash, D., & Cooper, G. F. (2002). *Exact model averaging with naive Bayesian classifiers.* Paper presented at the ICML.
6. Davidson, C. I., Phalen, R. F., & Solomon, P. A. (2005). Airborne particulate matter and human health: a review. *Aerosol Science and Technology, 39*(8), 737-749.
7. De Kock, M., Le, H., Tadross, M., & Potgeiter, A. (2008). Weather forecasting using dynamic bayesian networks: Citeseer.
8. Dunea, D., Iordache, S., Liu, H.-Y., Bøhler, T., Pohoata, A., & Radulescu, C. (2016). Quantifying the impact of PM 2.5 and associated heavy metals on respiratory health of children near metallurgical facilities. *Environmental Science and Pollution Research, 23*(15), 15395-15406.
9. Fattore, E., Paiano, V., Borgini, A., Tittarelli, A., Bertoldi, M., Crosignani, P., & Fanelli, R. (2011). Human health risk in relation to air quality in two municipalities in an industrialized area of Northern Italy. *Environmental research, 111*(8), 1321-1327.
10. Feng, L.-H. (2001). Application of neural network in weather forecast. *INFORMATION AND CONTROL-SHENYANG-, 30*(4), 365-367.
11. Ghorani-Azam, A., Riahi-Zanjani, B., & Balali-Mood, M. (2016). Effects of air pollution on human health and practical measures for prevention in Iran. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences, 21*.
12. Hulin, M., Caillaud, D., & Annesi-Maesano, I. (2010). Indoor air pollution and childhood asthma: variations between urban and rural areas. *Indoor air, 20*(6), 502-514.
13. Li, Q., Guo, Y., Song, J.-Y., Song, Y., Ma, J., & Wang, H.-J. (2018). Impact of long-term exposure to local PM 10 on children's blood pressure: a Chinese national cross-sectional study. *Air Quality, Atmosphere & Health, 11*(6), 705-713.
14. Liu, J. C., & Peng, R. D. (2018). Health effect of mixtures of ozone, nitrogen dioxide, and fine particulates in 85 US counties. *Air Quality, Atmosphere & Health, 11*(3), 311-324.
15. Mehdipour, V., & Memarianfard, M. (2017). Application of support vector machine and gene expression programming on tropospheric ozone prognosticating for Tehran metropolitan. *Civil Engineering Journal, 3*(8), 557-567.
16. Mehdipour, V., Stevenson, D. S., Memarianfard, M., & Sihag, P. (2018). Comparing different methods for statistical modeling of particulate matter in Tehran, Iran. *Air Quality, Atmosphere & Health, 11*(10), 1155-1165.
17. Pope, C. A., Ezzati, M., Cannon, J. B., Allen, R. T., Jerrett, M., & Burnett, R. T. (2018). Mortality risk and PM 2.5 air pollution in the USA: an analysis of a national prospective cohort. *Air Quality, Atmosphere & Health, 11*(3), 245-252.
18. Schweitzer, L., & Zhou, J. (2010). Neighborhood air quality, respiratory health, and vulnerable populations in compact and sprawled regions. *Journal of the American Planning Association, 76*(3), 363-371.
19. Turner, M. C., Krewski, D., Pope III, C. A., Chen, Y., Gapstur, S. M., & Thun, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American journal of respiratory and critical care medicine, 184*(12), 1374-1381.
20. Xing, Y.-F., Xu, Y.-H., Shi, M.-H., & Lian, Y.-X. (2016). The impact of PM2. 5 on the human respiratory system. *Journal of thoracic disease, 8*(1), E69.