Predictive Modelling of Air Quality Index at Various Locations in Vishakhapatnam City, India

Zeinab Hilal [1*], Dr. GVR Srinivasa Rao [2]

[1-2] Department of Civil Engineering, AU College of Engineering (A), Andhra University, Visakhapatnam-530003, Andhra Pradesh, India

* Corresponding Author: **Zeinab Hilal**

Article Info

E-ISSN: 3107-7099
Volume: 02
Issue: 04
Received: 07-05-2026
Accepted: 06-06-2026
Published: 04-07-2026
Page No: 09-15

Abstract

Air pollution is an important international challenge due to its increasing impacts both on public health and the environment. The Air Quality Index simplifies the complex data by transforming the parameters into a single numerical value. However, accurately estimating the AQI stays a difficult because of the complicated connections among the pollutants and weather factors. It is aimed in this study to develop predictive models for AQI using statistical modelling, with a concentration on evaluating the influence of weather variables. The data were gained from publicly available government monitoring stations, specifically the Andhra Pradesh Pollution Control Board, as well as, the Central Pollution Control Board in Visakhapatnam city. Data period was used for the years 2018–2024. The parameters used are: AQI, SO_2, NOx, PM_{10}, $PM_{2.5}$, NH_3, temperature, precipitation, and wind speed. At a particular air quality monitoring station, additional parameters such as NO, NO_2, CO, O_3, benzene, and toluene were included. Linear, polynomial, and exponential models are the models used for the statistical analysis. The results highlighted that the best model with the best effectiveness and fit is the polynomial model, which gives the best performance, the highest R^2 values (0.889, 0.937, 0.944, 0.900, 0.931, 0.703, and 0.662) across the studied stations were achieved. These findings can support improved air quality management and public health planning in Visakhapatnam city.

DOI: https://doi.org/10.54660/IJRCE.2026.2.4.09-15

Keywords: Air pollution, Air quality index, predictive models, Air quality management, Public health planning, Visakhapatnam, Andhra Pradesh

1. Introduction

Air pollution had become one of the most important problems that the environment and general health is facing, as it is caused by the growing urbanisation and industrialisation around the world (Ghorani-Azam *et al.*, 2016; Hilal & Rao, 2025) [7, 35]. Exposure to air pollutants causes both short-term and long-term health effects (Simu *et al.*, 2020; Mathew *et al.*, 2023) [28, 20]. Every year, obvious threats are caused to human health and the environment due to the air pollution, which costs in many lives lose (Manisalidis *et al.*, 2020) [19]. In order to control air pollution, it is important to understand the nature of the pollutants and their sources. Based on the pollutant's nature, the air pollutants could be classified into three categories: natural contaminants, aerosols, and gases and vapours. Natural fog, pollen, grains, and product of volcanic eruption are examples of natural contaminants, while, dust, smoke, mists, and fog, are examples of aerosols, and as an example for gases and vapours, Sulphur compounds, Nitrogen compounds, Oxygen compounds, could be included (Mitchell & Sweeney, 2018) [21]. Another classification based on the source of these pollutants is found, where the pollutants split into two types: primary, and secondary pollutants (Singh *et al.*, 2025; Hilal & Rao, 2025) [29, 35]. The pollutants that are directly released from certain processes, and remain in the same form tat they are released in, are called primary pollutants, such as vehicle exhaust emissions. In the other hand, the pollutants that are generated due to chemical reactions including primary pollutants, are called secondary pollutants, for example the photochemical smog which is due to chemical reactions between weather factors and some air pollutants (Simu *et al*., 2020; Singh *et al*., 2025) [28, 29]. Every day a huge number of pollutants, such as CO, CO_2, Particulate Matter (PM), NO_2, SO_2, O_3, NH_3, and Pb, which have a negative effect on the health of humans, animals, and plants, are released. These chemicals and pollutants are considered to be an important component of air pollution (Ghorani-Azam *et al*., 2016; Kumar & Pande, 2023) [7, 16].

The Air Quality Index (AQI) is designed to be a standard tool, aims to simplify the scientific information related to air pollution, and converts it into clearly and easily formula for public. Where this index is considered to be a useful tool to connect the complex quantitative data with the levels of potential health impact on human and environment (Beig. *et al*., 2010) [5]. It simplifies the complex data collected for various pollutants by transforming them into a single numerical value, a descriptive label, and a corresponding colour (Abas *et al*.,2024) [2]. According to the Andhra Pradesh Pollution Control Board, air quality refers to how suitable the air is for humans, plants, and animals to breathe. Poor air quality can harm health, the environment, the economy, and a city's livability, affecting both current and future generations. Outdoor air pollution is an important environmental health problem impacting people in low-, middle-, and high-income countries. A healthy environment is an essential aspect of the right, for both humans and other animals, to live, which leads to the importance of the quality of outdoor air for well-being.

To get better outcomes and to understand the air quality metric, the components of air pollution and their interactions in the atmosphere must be understood. Tiny airborne particles together form the fine particulate matter, which can cause harm to health, such as respiratory diseases (Harrison, 2020) [9]. These particles can be classified due to their size into two groups: $PM_{2.5}$ and PM_{10} (Kim *et al*., 2015) [13]. In addition to particulate matter, different gaseous pollutants such as nitrogen oxides (NO_X), sulfur dioxide (SO_2), carbon monoxide (CO), and VOCs also cause respiratory and environmental damage. Pollutants form from various human activities, such as daily, industrial, and commercial activities (Abas *et al*.,2024) [2].

In environmental studies, due to its complex interactions, machine learning becomes a very wide-ranging tool used for its analytical capabilities, especially in predicting air quality fields. Therefore, the handling and processing of large and heterogeneous dataset's ability of machine learning makes it proper for forecasting the air quality index, besides capturing changes over time and area (Kang *et al*., 2018) [11].

ML models are chosen based on data and work requirements, where there are many kinds of ML methods, such as linear regression, decision trees, random forests, support vector machines, and neural networks (Saravanan & Sujatha, 2018) [27]. The air quality index measures air pollution based on different parameters using historical and real-time data, for example, $PM_{2.5}$, PM_{10}, NO_2, and SO_2. Therefore, for protecting public health and helping policy decision makers, ML methods have been used to forecast AQI accurately (Kumar & Pande, 2023) [16].

Since human health and the environment have become very essential in recent years. Thus, to know the effects of what one breathes, a lot of studies have discussed their effects on the health. West *et al*. (2016) [33] concentrated on the negative effects of pollutants emitted from several sources into the air, especially in developing countries. As well as the researchers must have a better understanding of the sophisticated physical and chemical characteristics of air pollutants. Studies by Ravindiran *et al*. (2023) [26] and Natarajan *et al*. (2024) [22] forecast AQI using different machine learning models in Indian cities, and compare the effectiveness of these models. Different researchers use the SPSS program and apply several models to predict air quality based on historical and real-time data of different factors, capturing the relationship between AQI and its influencing parameters (Ma *et al*., 2020; Abdullah *et al*., 2020) [3].

This study tests machine learning models (e.g., multivariate linear, nonlinear regression) for predicting AQI levels using various data, including meteorological and pollutant datasets. The performance of each model is compared with the official, calculated, and published AQI values by the government agencies. The study aimed to establish a predictive model in order to forecast the air quality index in the study area (the city of Visakhapatnam) by using historical and real time data, involving the pollutant concentrations (e.g., PM2.5, NO_2, CO), meteorological factors (e.g., temperature, precipitation and wind speed), and official and government-calculated AQI values. For assessing the models' performance several metrics were used, and they are R-squared (R^2), Adjusted R-squared, Standard Error of Estimate, Significance of Predictors (p-values), and Multicollinearity Check.

2. Study Area and Sampling

Visakhapatnam (also known as Vizag) is considered to be the largest and most populous metropolitan city in the Andhra Pradesh state, it takes a place between the Eastern Ghats and the coast of the Bay of Bengal, which gives it a special geographical location that combines both mountainous and costal properties (Rani, 2016) [25]. It is considered to be the second biggest city on the east coast of India after Chennai and the fourth in South India (Rani, 2016) [25]. The city coordinates lie between 17.7041 N and 83.2977 E. The city's area is 682 km². The average elevation is 45 meters (Tanuja & Sreenivasulu, 2023) [32]. Visakhapatnam is situated in the Coastal Andhra Region. It has a tropical climate with hot, humid summers and moderate winters. The year is divided into the following seasons:

- **Winter:** includes of the months from December to February
- **Summer:** includes of the months from March to May
- **Monsoon:** includes of the months from June to September
- **Autumn:** includes of the months from October to November

By studying the city climate history, it is found that May is the hottest month of the city with average highs of 35°C and lows of 28°C. In contrast, December is the coldest month with average highs of 27°C and lows of 20°C.

The officially calculated and reported Air Quality Index (AQI) values along with its affecting parameters, which were used as sample data from 2018 to 2024, were collected from available air quality datasets provided by government monitoring stations, which are the Andhra Pradesh Pollution Control Board (APPCB) and the Central Pollution Control Board (CPCB) websites. The data were from six manual and one continuous monitoring stations, and the GVMC is the continuous monitoring station. While the manually monitored stations include:

- Industrial Estate, Autonagar
- Mindi Gudivada Appanna Kalyanamandapam
- Seethammadhar
- Gnanapuram
- Peda Gantyada
- Raitu Bazar

The data consisted of an average of 7,960 samples for each station, while the GVMC station consisted of 38,355 samples. The data was split into a 70:30 ratio as training and test data, respectively. The data was gathered from the government website and measured on an hourly basis. The data is converted into daily measurements in order to facilitate the calculation and analysis process by taking its average. The stations' variables are SO_2, NOx, PM_{10}, $PM_{2.5}$, NH_3, Temperature, Precipitation, and Wind Speed. Meanwhile, the GVMC station contained extra parameters besides what were listed earlier, are NO, NO_2, CO, O_3, Benzene, and Toluene. The study was conducted to detect a predictive Air Quality Index (AQI) equation established from gathered data, which was converted into daily averages, and then divided into a 70:30 ratio as training and testing sets.

According to the studies by Ma *et al*. (2020); Kumar & Pande, (2023) [16]; Ravindiran *et al*. (2023) [26]; Natarajan *et al*. (2024) [22], which used the same technique, the AQI values used in this study are officially calculated and reported by the relevant regulatory authorities, and are treated as reference values for evaluating the performance of the prediction models. This was done in order to achieve approximate with the highest possible accuracy.

3. Methodology

Out of SPSS statistical and machine learning characteristics, it was successfully used for the Air Quality Index prediction. In simple terms, SPSS (Statistical Package for the Social Sciences) can be defined as a statistical analysis, data mining, and management software tool. As well as, the program provides a simple interface that help users with minimal programming experience to perform different statistical tasks (Kaurav *et al*., 2020) [12]. SPSS supports data entry, preparation, visualization, and modelling, all these features make it suitable for all users whereas beginners or experienced researchers. According to that it is commolnly used in different fields, such as social science, business, and research (Keith *et al*., 2020; Kaurav *et al*., 2020) [12]. It also simplifies data preparation tasks like managing missing values by giving easy to apply features and commands (e.g., PM2.5, PM10, NO2). Different models, such as regression, decision trees, and neural network models, are supported to analyse relationships between air pollutants and weather variables (Keith *et al*., 2020).

Figure 1 spotlights the key steps included in the methodology.

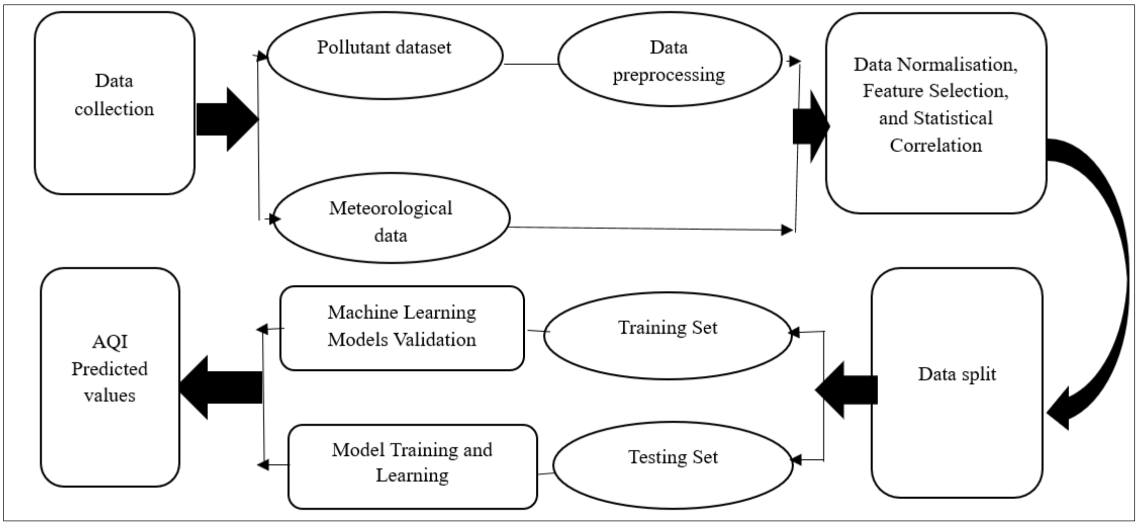

Fig 1: The Methodology Flow Chart

3.1. Multivariate Linear Regression

Regression analysis is one of the important analytical models used to study and estimate the relationship between variables. This analyse aims to modeling the relationship in order to explain the variables attitude and predict its future values. It is widely use over different fields, such as engineering, physical and chemical sciences (Douglas *et al*., 2012).

Multiple linear regression is an extend to simple linear regression, which aims to analyse the relationship between one dependent variable and multiple independent ones. This model is based on the basic principles of simple linear regression. But the main difference is to include more than one independent variables in the model, this allows for deeper understanding of the effect of each independent variable on the dependent variable, and controlling the impact of other variables (Song, 2018; Ma *et al*., 2020) [30].

The linear relationship between a single independent variable X and a dependent variable Y is described using a simple linear regression model, where the relationship is illustrated by a straight line (Douglas *et al*., 2012; Abdullah *et al*., 2020; Ma *et al*., 2020) [3]. The following equation represents the simple linear regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

- **Y:** Represents the value of dependent variable
- **X:** Represents the value of independent variable
- **β_0:** Represents the regression line intercept
- **β_1:** Represents the regression line slope
- **ε:** Represents the random error component, which accounts for the model failure of precisely fitting the data.

To extend simple linear regression, which uses a single independent variable, by incorporating multiple predictors, we get the Multiple Linear Regression (MLR) (Ma *et al*., 2020). The objective of it is to develop a mathematical model that captures the relationship between the dependent variable and the independent variables, enabling predictions of the dependent variable based on the values of the predictors (Song, 2018; Darman & Abd Hamid, 2024) [30, 6]. The general equation for multiple linear regression is provided below:

$$y = \beta_0 + \beta_1 x1 + \beta 2x2 + \cdots + \beta n\, xp + \varepsilon \quad (2)$$

In formula (2), x1, x2,.., xp represent independent variables. Here, x_1 represents for the value of SO2, x_2 for NOX, x_3 for PM10, x_4 for PM2.5, x_5 for NH3, x_6 for Temperature, x_7 for Precipitation, and x_8 for Wind Speed.

If there are N samples, it is more appropriate to deal with multiple regression models if they are conveyed in matrix notation (Douglas *et al*., 2012; Song, 2018) [30], as shown by formula (3):

N groups of samples where $(x_{i1}, x_{i2}, \ldots, x_{ip})$, (i =1, 2, …, n)

$$y = \begin{Bmatrix} y1 \\ y2 \\ \vdots \\ yn \end{Bmatrix}, x = \begin{Bmatrix} x11 & x12 & \ldots & x1p \\ x21 & x22 & \ldots & x2p \\ \ldots & \ldots & \ldots & \vdots \\ xn1 & xn2 & \ldots & xnp \end{Bmatrix}$$

$$\beta 0 = \begin{Bmatrix} \beta 1 \\ \beta 2 \\ \vdots \\ \beta p \end{Bmatrix}, \varepsilon = \begin{Bmatrix} \varepsilon 1 \\ \varepsilon 2 \\ \vdots \\ \varepsilon n \end{Bmatrix} \quad (3)$$

Therefore, the equation that represents the multivariate linear regression in a matrix form is as follows:

$$Y = X\beta + \varepsilon \quad (4)$$

In multiple linear regression (sometimes called multivariate linear regression when involving one dependent variable and multiple independent variables), the error term Ɛ must meet specific assumptions to ensure the regression results are valid and meaningful (Ma *et al*., 2020).

3.2. Polynomial Regression Models

Since the relationship between AQI and pollutants, along with meteorological factors, does not always follow a linear pattern, we should seek an alternative model to develop a better equation for predicting AQI. The regression analyses that include more than one independent variable are known as Multiple regression. Polynomial regression technique is one type of Multiple regression; in this kind, the dependent variable is regressed against different powers of the independent variables (Ostertagová, E. 2012) [23].

The linear regression method, which is expressed by the simple equation y = βx + Ɛ, works as a general frame for fitting relationships that are linear in the unknown parameters β. This includes the significant category of polynomial regression models (Douglas *et al*., 2012). For example, a quadratic polynomial model in a single variable demonstrates this approach.

$$y = \beta 0 + \beta 1x + \beta 2x2 + \varepsilon \quad (5)$$

Additionally, a quadratic polynomial model with two variables, expressed as:

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \beta 3x12 + \beta 4x22 + \varepsilon \quad (6)$$

Which is also a linear regression model. Polynomials are extensively applied in scenarios where the response shows a curvilinear pattern, as they can effectively model even complicated nonlinear relationships within relatively narrow ranges of the x variables (Douglas *et al*., 2012).

3.3. Exponential Regression Models

An exponential model assumes that the dependent variable (e.g., AQI) grows or decays at a rate proportional to its current value (Douglas *et al*., 2012). A typical form might be:

$$\text{AQI} = \beta 0 * e^{\beta 1x} + \varepsilon \quad (7)$$

- **AQI:** Dependent variable (Air Quality Index).
- **x:** Independent variable (e.g., $PM_{2.5}$, temperature).
- **β_0:** Base value
- **β_1:** Growth/decay rate.
- **ε:** Random error.

In SPSS, the formula should be written in a simple exponential model, for example: b_0 * EXP (b_1 * $PM_{2.5}$), where b_0 represents β_0 and b_1 represents β_1, and so on. In the parameter box, set the initial values for each parameter; this assists SPSS in starting the iteration process. After running the analysis, SPSS will estimate the parameters and provide output, including parameter estimates, standard errors, and fit statistics, so these results can be used to assess how much this model can fit the data and explain the relation between the variables. The analysis process would be rerun until a negative or positive variable estimations is found.

This analysis supports the use of an exponential model, in order to understand and explain the nonlinear relationship between the air quality index (AQI) and its impacting factors. The use of this model is suitable in the cases that the variables show nonlinear correlation, and this enable the researcher to represent the cumulative or accelerated changes in the AQI as a response to the changes in the meteorological variables, such as, temperature, humidity, wind speed, etc. (Douglas *et al*., 2012).

3.4. Model Comparison

The final stage of the methodology is the systematic comparison between the analysis models after completing the analysis and estimating the variables. This stage is aimed to detect the most accurate and efficient model to interpret the relationship between variables. The comparison depends on analysing a group of statistical indicators that used for assessing the models and their suitability for the data, which helped in choosing the best model among all models. The used indicators for comparing include the following:

First, R^2 shows the proportion of variance in AQI explained by the predictors (e.g., pollutants like $PM_{2.5}$, NO_2, etc.). Higher R^2 (closer to 1) indicates a better fit (Joseph *et al*.,2019). Secondly, the Standard Error of Estimate, which shows the average distance between the reference AQI values and the predicted values. A smaller standard error means more accurate prediction (Douglas *et al*., 2012). Thirdly, Significance of Predictors (p-values): Check it for each independent variable. If $p < 0.05$, the predictor is statistically significant (contributes meaningfully to AQI prediction)

(Pallant, 2020) [24]. Finally, the Multicollinearity Check (Variance Inflation Factor) should be less than 5 to be generally acceptable (Pallant, 2020) [24].

4. Results and Discussion

The predicted AQI values were compared with the official reference values using, calculated and published using standard statistical measures, to evaluate the predictive ability of the different models (Linear, Polynomial, and Exponential) used in this study, for predicting the Air Quality Index (AQI) over the monitoring stations in the city of Vishakhapatnam. For the illustrative and comparing purposes, Table 1 shows the equations of each model developed for each station.

Table 1: The model equations of each station

Station	Model	Model Equation
Industrial Estate, Autonagar	Linear	AQI = 9.628 + 0.832 × PM10 − 0.099 × NH3 + 0.086 × PM2.5 + 0.204 × NOX
	Polynomial	AQI = 5.385 + 1.023 × PM10 − 0.001 × PM10² + 0.080 × PM2.5 − 0.097 × NH3
	Exponential	AQI = 38.625 × EXP(0.003 × NOX + 0.008 × PM10 + 0.001 × PM2.5 − 0.001 × NH3)
Mindi Gudivada Appanna Kalyanamandapam	Linear	AQI = 7.428 + 0.901 × PM10 + 0.139 × PM2.5 − 0.094 × NH3
	Polynomial	AQI = −2.017 + 1.209 × PM10 − 0.002 × PM10² − 0.195 × PM2.5 + 0.003 × PM2.5² − 0.001 × NH3²
	Exponential	AQI = 59.382 × EXP(0.009 × PM10 − 0.002 × NH3 − 0.016 × Temperature)
Seethammadhara	Linear	AQI = 4.260 + 0.942 × PM10 − 0.456 × SO2 + 0.060 × PM2.5 − 0.040 × NH3 + 0.173 × NOX
	Polynomial	AQI = −4.111 − 0.342 × SO2 + 1.147 × PM10 + 0.060 × PM2.5 + 0.0001 × NH3² − 0.001 × PM10²
	Exponential	AQI = 56.150 × EXP(−0.012 × SO2 + 0.010 × PM10 − 0.015 × Temperature)
Gnanapuram	Linear	AQI = 19.230 + 0.772 × PM10 − 0.046 × NH3 + 0.132 × Wind Speed
	Polynomial	AQI = 4.042 + 1.118 × PM10 − 0.002 × PM10² − 0.046 × NH3 + 0.002 × Wind Speed²
	Exponential	AQI = 52.703 × EXP(0.006 × PM10 − 0.002 × Precipitation)
Peda Gantyada	Linear	AQI = 7.807 + 0.895 × PM10
	Polynomial	AQI = −31.441 − 0.131 × SO2 + 0.622 × NOX + 1.108 × PM10 + 0.034 × PM2.5 − 0.078 × NH3 + 2.026 × Temperature − 0.007 × Precipitation + 0.167 × Wind Speed − 0.015 × NOX² − 0.001 × PM10² − 0.040 × Temperature² − 0.002 × Wind Speed²
	Exponential	AQI = 45.073 × EXP(−0.009 × SO2 + 0.010 × PM10 − 0.006 × Temperature)
Raitu Bazar	Linear	AQI = 21.748 + 0.677 × PM10
	Polynomial	AQI = 17.573 + 0.674 × PM10 − 0.004 × PM2.5² + 0.291 × PM2.5
	Exponential	AQI = 35.141 × EXP(0.009 × PM10)
GVMC	Linear	AQI = 108.497 + 0.726 × PM2.5 + 0.363 × PM10 − 3.370 × Temperature + 0.194 × O3 − 0.215 × NO + 0.219 × NH3 + 0.407 × Wind Speed
	Polynomial	AQI = 675.631 + 0.705 × PM2.5 + 0.347 × PM10 − 43.759 × Temperature + 0.739 × Temperature² + 0.002 × O3² − 0.003 × NOX² + 0.476 × NH3 − 0.003 × NH3²
	Exponential	AQI = 164.970 × EXP(−0.003 × NO − 0.002 × NO2 + 0.002 × NH3 + 0.002 × O3 + 0.003 × PM2.5 + 0.004 × PM10 − 0.035 × Temperature)

The results of R-squared (R^2), Adjusted R-squared (Adj. R^2), Standard Error, significance levels (p-value), and VIF are the metrics used for the compression approach, are shown in Table 2.

Table 2: The comparison metrics of the models.

Station	Models	R^2	Adj. R^2	Std. Error	p-value	VIF
Industrial Estate, Autonagar	Linear	0.888	0.888	8.833	0.040	<5
	Polynomial	0.889	0.888	8.809	0.000	<5
	Exponential	0.853	-	-	-	-
Mindi Gudivada Appanna Kalyanamandapam	Linear	0.932	0.931	7.818	0.040	<5
	Polynomial	0.937	0.937	7.689	0.000	<5
	Exponential	0.853	-	-	-	-
Seethammadhara	Linear	0.939	0.939	5.261	0.000	<5
	Polynomial	0.944	0.943	5.073	0.000	<5
	Exponential	0.811	-	-	-	-
Gnanapuram	Linear	0.879	0.878	8.571	0.000	<5
	Polynomial	0.900	0.899	7.786	0.000	<5
	Exponential	0.742	-	-	-	-
Peda Gantyada	Linear	0.927	0.926	6.757	0.000	<5
	Polynomial	0.931	0.930	6.577	0.000	<5
	Exponential	0.855	-	-	-	-
Raitu Bazar	Linear	0.694	0.693	9.939	0.000	<5
	Polynomial	0.703	0.701	9.815	0.000	<5
	Exponential	0.683	-	-	-	-
GVMC	Linear	0.650	0.649	32.731	0.006	<5
	Polynomial	0.662	0.660	32.204	0.032	<5
	Exponential	0.629	-	-	-	-

Across the monitoring stations, and based on the comparison metrics, the polynomial model was found to give the best performance, holding the highest R^2 and lowest standard errors compared with Linear and Exponential models. In contrast, the Exponential model has the weakest predictive ability.

Seethammadhara station has the strongest fit with the values of $R^2 = 0.944$, Adj. $R^2 = 0.943$, and std. error = 5.073. While GVMC shows the weakest performance with the values of $R^2 = 0.662$, Adj. $R^2 = 0.660$, and std. error = 32.204. VIF values for all the models were below 5, which means there are no multicollinearity problems, and p-values for all stations and for all models were below 0.05, which means all the predictors are statistically significant and contribute meaningfully to AQI.

These results point to the suitability of the Polynomial model to predict AQI and explain its variations across the stations.

5. Conclusion and Recommendations

This study aimed to improve a predictive model in Visakhapatnam for the Air Quality Index (AQI) using historical and real-time data acquired from pollution control websites of the Andhra Pradesh Pollution Control Board (APPCB) and the Central Pollution Control Board (CPCB). The objectives were to identify the key indicators affecting air quality. The relationships and interactions between these parameters and the AQI were analysed.

The analysis was built using SPSS software, and the model performance was compared using statistical indicators such as R-squared (R^2), Adjusted R-squared, Standard Error of Estimate, significance levels (p-values), and multicollinearity tests. From the results gained earlier, the following conclusions were drawn.

The polynomial model was found to be the most effective model among the tested ones. This model detected that the relationship between AQI and its predictors is not completely linear and differs with seasonal changes, influenced by meteorological conditions. The selection of the best model was based on its ability to give a realistic and accurate estimation when compared to reference AQI values. The values of R^2 were as follows: (0.889, 0.937, 0.944, 0.900, 0.931, 0.703, and 0.662) respectively for the stations, while the adjusted R^2 values for the stations respectively were as follows: (0.888, 0.937, 0.943, 0.899, 0.930, 0.701, 0.660), which indicate a better fit.

6. Funding Declaration

There is no fund has received for any part of this research

References

1. Hair JF Jr, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 8th ed. Harlow: Pearson; 2019.
2. Abas M, Zeeshan M, Khalil MH, Sial SA, Yaqoob Z, Sial DG, *et al*. Navigating air quality assessment: A holistic review of air quality index metrics and assessment approaches. 2024.
3. Abdullah S, Napi NNLM, Ahmed AN, Mansor WNWM, Mansor AA, Ismail M, *et al*. Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia. Atmosphere. 2020;11(3):289. doi:10.3390/atmos11030289.
4. Attwal K, Singh A. Exploring SPSS statistics for data mining and statistical modeling. Int J Stat Appl Math. 2024;9(3):9-16. doi:10.22271/maths.2024.v9.i3a.1721.
5. Beig G, Ghude SD, Deshpande A. Scientific Evaluation of Air Quality Standards and Defining Air Quality Index for India. Pune: Indian Institute of Tropical Meteorology; 2010.
6. Darman H, Abd Hamid NZ. Performance comparison of haze prediction using chaos theory and multiple linear regression. J Qual Meas Anal. 2024;20(3):23-34. doi:10.17576/jqma.2003.2024.02.
7. Ghorani-Azam A, Riahi-Zanjani B, Balali-Mood M. Effects of air pollution on human health and practical measures for prevention in Iran. J Res Med Sci. 2016;21(1):65. doi:10.4103/1735-1995.189646.
8. Halsana S. Air quality prediction model using supervised machine learning algorithms. Int J Sci Res Comput Sci Eng Inf Technol. 2020;8:190-201. doi:10.32628/cseit206435.
9. Harrison RM. Airborne particulate matter. Philos Trans R Soc A. 2020;378(2183):20190319. doi:10.1098/rsta.2019.0319.
10. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. 5th ed. Hoboken (NJ): John Wiley & Sons; 2012.
11. Kang GK, Gao JZ, Chiao S, Lu S, Xie G. Air quality prediction: Big data and machine learning approaches. Int J Environ Sci Dev. 2018;9(1):8-16. doi:10.18178/ijesd.2018.9.1.1066.
12. Kaurav RPS, Gursoy D, Prakash M. Why do we need SPSS? In: An SPSS Guide for Tourism, Hospitality and Events Researchers. London: Routledge; 2020. p.20-34.
13. Kim KH, Kabir E, Kabir S. A review on the human health impact of airborne particulate matter. Environ Int. 2015;74:136-143. doi:10.1016/j.envint.2014.10.005.
14. Kim Y, Oh H. Comparison between multiple regression analysis, polynomial regression analysis, and an artificial neural network for tensile strength prediction of BFRP and GFRP. Materials. 2021;14(17):4861. doi:10.3390/ma14174861.
15. Kukreja H, Bharath N, Siddesh CS, Kuldeep S. An introduction to artificial neural network. Int J Adv Res Innov Ideas Educ. 2016;1(5):27-30.
16. Kumar K, Pande BP. Air pollution prediction with machine learning: A case study of Indian cities. Int J Environ Sci Technol. 2023;20(5):5333-5348. doi:10.1007/s13762-022-04241-5.
17. Ma L, Gao Y, Zhao C. Research on machine learning prediction of air quality index based on SPSS. In: 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA). Piscataway (NJ): IEEE; 2020. p.1-5. doi:10.1109/ICCNEA50255.2020.00011.
18. Madan T, Sagar S, Virmani D. Air quality prediction using machine learning algorithms—a review. In: 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). Piscataway (NJ): IEEE; 2020. p.140-145. doi:10.1109/ICACCCN51052.2020.9362912.
19. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: A review. Front Public Health. 2020;8:14. doi:10.3389/fpubh.2020.00014.
20. Mathew A, Gokul PR, Raja Shekar P, Arunab KS, Ghassan Abdo H, Almohamad H, *et al*. Air quality

analysis and PM2.5 modelling using machine learning techniques: A study of Hyderabad city in India. Cogent Eng. 2023;10(1):2243743. doi:10.1080/23311916.2023.2243743.

21. Mitchell R, Sweeney F. Air Pollution. Scientific e-Resources; 2018.
22. Natarajan SK, Shanmurthy P, Arockiam D, Balusamy B, Selvarajan S. Optimized machine learning model for air quality index prediction in major cities in India. Sci Rep. 2024;14(1):6795. doi:10.1038/s41598-024-54807-1.
23. Ostertagová E. Modelling using polynomial regression. Procedia Eng. 2012;48:500-506. doi:10.1016/j.proeng.2012.09.545.
24. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS. 7th ed. London: Routledge; 2020. doi:10.4324/9781003117452.
25. Rani KD. Sustainable urban transport planning strategies for land use pattern in Visakhapatnam city. Int J Eng Dev Res. 2016;4(2):114-119.
26. Ravindiran G, Hayder G, Kanagarathinam K, Alagumalai A, Sonne C. Air quality prediction by machine learning models: A predictive study on the Indian coastal city of Visakhapatnam. Chemosphere. 2023;338:139518. doi:10.1016/j.chemosphere.2023.139518.
27. Saravanan R, Sujatha P. A state of the art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. In: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). Piscataway (NJ): IEEE; 2018. p.945-949. doi:10.1109/ICCONS.2018.8663155.
28. Simu S, Turkar V, Martires R, Asolkar V, Monteiro S, Fernandes V, *et al*. Air pollution prediction using machine learning. In: 2020 IEEE Bombay Section Signature Conference (IBSSC). Piscataway (NJ): IEEE; 2020. p.231-236. doi:10.1109/IBSSC51096.2020.9332184.
29. Singh V, Srivastava RK, Bhatt AK. Major air pollutants. In: Battling Air and Water Pollution: Protecting Our Planet's Vital Resources. Singapore: Springer Nature; 2025. p.31-42. doi:10.1007/978-981-96-4375-2_3.
30. Song X. Urban population growth model based on SPSS regression analysis. In: 2018 13th International Conference on Computer Science & Education (ICCSE). Piscataway (NJ): IEEE; 2018. p.1-7. doi:10.1109/ICCSE.2018.8468788.
31. McCormick K, Salcedo J, Poh A. SPSS Statistics For Dummies. 4th ed. Hoboken (NJ): Wiley; 2020.
32. Tanuja S, Sreenivasulu V. Seawater intrusion using chemical indicators in the coastal aquifers East Godavari and Visakhapatnam districts. J Emerg Technol Innov Res. 2023;23(10):11.
33. West JJ, Cohen A, Dentener F, Brunekreef B, Zhu T, Armstrong B, *et al*. What we breathe impacts our health: Improving understanding of the link between air pollution and health. Environ Sci Technol. 2016;50(10):4895-4904. doi:10.1021/acs.est.5b03827.
34. Zakaria M, Mabrouka AS, Sarhan S. Artificial neural network: A brief overview. Neural Networks. 2014;1:2.
35. Hilal Z, Rao GS. A correlative study of air pollution and meteorological factors in Visakhapatnam City, India. Int J Adv Eng Manag. 2025;7(11):21-28. doi:10.35629/5252-07112128.

How to Cite This Article

Hilal Z, Rao GVS. Predictive modelling of Air Quality Index at various locations in Vishakhapatnam City, India. Int J Multidiscip Evol Res. 2026;2(4):9-15. doi:10.54660/IJRCE.2026.2.4.09-15.