LITERATURE SURVEY

* **Anikender Kumar, PramilaGoyal (2011)** presented the study that forecasts the daily AQI value for the city Delhi, India using previous record of AQI and meteorological parameters with the help of Principal Component Regression (PCR) and Multiple Linear Regression Techniques. They perform the prediction of daily AQI of the year 2006 using previous records of the year 2000-2005 and different equations. After that this predicted value then compared with observed value of AQI of 2006 for the seasons summer, Monsoon, Post Monsoon and winter using Multiple Linear Regression Technique [1]. Principal Component Analysis is used to find the collinearity among the independent variables. The Principal components were used in Multiple Linear Regression to eliminate collinearity among the predictor variables and also reduce the number of predictors [1]. The Principal Component Regression gives the better performance for predicting the AQI in winter season than any other seasons. In this study only meteorological parameters were considered or used while forecasting the future AQI but they have not considered the ambient air pollutants that may cause the adverse health effects. The Principal Component Regression performs better than any other season at forecasting the AQI throughout the winter. The ambient air contaminants that may have harmful effects on health were not taken into account while projecting the future AQI in this study; only meteorological parameters were employed.
* **Huixiang Liu (et al. 2019)** :- For the purposes of their study, Huixiang Liu (et al. 2019) chose two distinct cities: Beijing and an Italian city. Using two distinct publicly available datasets, they have predicted the Air Quality Index (AQI) for the city of Beijing and the concentration of NOx in an Italian city.The Beijing Municipal Environmental Centre [5] has released the first Dataset for the period of December 2013 to August 2018 with 1738 instances public. This dataset includes fields such the hourly averaged AQI and the concentrations of PM2.5, O3, SO2, PM10, and NO2 in Beijing. The second dataset, which contains 9358 cases, was compiled using data collected in Italian cities between March 2004 and February 2005. This dataset's features include hourly averaged CO, NOx, and NO2 concentrations, as well as non-methane hydrocarbons like benzene.However, because NOxprediction is one of the key predictors for assessing air quality, they mainly concentrated on it. For the prediction of AQI and NOx concentrations, they used Support Vector Regression (SVR) and Random Forest Regression (RFR) approaches. SVR performs better when forecasting AQI, but RFR performs better when predicting NOx concentration.
* **Ziyue Guan and Richard O. Sinnot (2018)** predicted the PM2.5 concentration using a variety of machine learning techniques. Data were gathered from the Environment Protection Agency's (EPA) official website for Melbourne, which offers information about PM2.5 air parameters, and they also gathered unofficial data from Airbeam, a mobile device designed to measure PM2.5 levels [8]. Artificial Neural Network (ANN), Linear Regression (LR), and Long Short Term Memory (LSTM) recurrent neural networks were utilised as machine learning methods for the PM2.5 prediction, although LSTM performed the best and predicted high PM2.5 values with a respectable level of accuracy.
* **HeidarMaleki (et al. 2019)** :- The hourly concentration values for the ambient air pollutants NO2, SO2, PM10, PM2.5, CO and O3 were forecasted by HeidarMaleki (et al. 2019) for the stations Naderi, Havashenasi, MohiteZist and Behdasht in Ahvaz, Iran, the most polluted city in the world. At the four air quality monitoring sites in Ahvaz mentioned above, they have additionally calculated and forecasted the Air Quality Index (AQI) and Air Quality Health Index (AQHI). In order to estimate the hourly concentration of air pollutants and the two air quality indices, AQI and AQHI, from August 2009 to August 2010, they employed an Artificial Neural Network (ANN) machine learning method. Meteorological parameters, air pollution concentrations, time, and date are among the inputs that go into ANN algorithms.
* **Aditya C. R. (et al.) (2018)** used computer algorithms to identify and predict the PM2.5 concentration level based on a dataset that included atmospheric conditions in a particular city. They additionally forecasted the PM2.5 concentration for a specific day [10]. First, they use the Logistic Regression method to categorise the air as being contaminated or not polluted, and then they use the Auto Regression algorithm to forecast the future value of PM2.5 based on historical data.
* **Nidhi Sharma (et al. 2018):-** The full data analysis of air pollutants from 2009 to 2017 was carried out by Nidhi Sharma (et al. 2018), who also made a critical assessment on the trend in air pollution from 2016 to 2017 in Delhi, India [14]. Sulfur Dioxide (SO2), Nitrogen Dioxide (NO2), Suspended Particulate Matter (PM), Ozone (O3), Carbon Monoxide (CO), and Benzene are among the pollutants for which they have forecasted future trends.The future values of the pollutants indicated previously have been forecast using data analytics time series regression forecasting based on historical data. The monitoring stations in Delhi at AnandVihar and Shadipur are being studied, per the outcomes of this study. The findings indicate that PM10 concentration levels have dramatically increased, and that Delhi's levels of NO2 and PM2.5 have also increased [14]. CO is anticipated to decrease by 0.169 mg/m3, whereas NO2 concentration is anticipated to rise by 16.77 mg/m3, Ozone is anticipated to increase by 6.11 mg/m3, Benzene is anticipated to decrease by 1.33 mg/m3, and SO2 is anticipated to rise by 1.24 mg/m3.
* **Mohamed Shakir and N.Rakesh (2018)** used the WEKA tool to estimate the impact of environmental factors like temperature, wind speed, and humidity on the air pollutants mentioned above. They also examined the proportion of different air pollutants (NO, NO2, CO, PM10, and SO2) with respect to time of day and day of the week. The Karnataka pollution control board's data was gathered. The study found that the concentration levels of air pollutants grow throughout working days, especially during the busiest times of the day, and decrease during weekends or holidays [15] by utilising the ZeroR algorithm in the WEKA tool. The study illustrates the relationship or correlations between environmental variables like temperature, wind speed, and humidity and air pollutants like NO, NO2, PM10, CO, and SO2 using simple K-means clustering methods.
* **KazemNaddaf (et al., 2012) :-** For analysing the effects of PM10, SO2, NO2, and O3 on human health in Tehran City, Iran, the country's most populous city, KazemNaddaf (et al., 2012) employed the AirQ programme recommended by the World Health Organization [16]. All-cause mortality, cardiovascular illnesses, and respiratory diseases were taken into consideration as health effects. According to the study's findings, PM10 had the most adverse health effects on Tehran City's population of 8,700,000 and was also responsible for an excess of total mortality of 2194 deaths out of 47284 in a year [16]. There have been an excess of 1458, 1050, and 819 occurrences of mortality related to SO2, NO2, and ozone, respectively. These findings demonstrate that air pollution in Tehran was a serious issue, and it is necessary for Tehran to lessen its negative health effects.
* **YusefOmidiKhaniabadi (et al.2016) :-** The primary goal of this study, is to determine the relationship or association between health effects, such as mortality rates from cardiovascular diseases, and air pollutants like PM10, NO2, and O3 for the Iranian city of Kermanshah between the years of 2014 and 2015. For this, they made use of the AirQ software suggested by WHO.Premature deaths from cardiovascular disorders account for 188 cases, 33 cases are associated to NO2, and 83 cases are related to O3[17].According to the study's findings, the mortality risk will increase by 1.066, 1.012, and 1.020 accordingly for every 10/m3 increase in PM10, NO2, and O3 concentration levels.
* **S.TikheShruti (et al. 2013) :-** For the purpose of forecasting future concentration levels of air pollutants such as oxides of sulphur (SOx), oxides of nitrogen (NOx), and respirable suspended particulate matter (RSPM) over the years 2005 to 2011, S.TikheShruti (et al. 2013) used two soft computing algorithms, artificial neural network (ANN), and genetic programming (GP). Pune, a city in Maharashtra, is ranked second among the most polluted cities in India. Based on hourly average data values of pollutants concentration spanning more than 7 years, they have constructed a total of six models (three of each algorithm ANN and GP). Amongst these two algorithms, GP algorithms performs better than ANN.
* **ArchontoulaChaloulakou (et al., 2003) :-** In order to anticipate the PM10 concentration over a two-year time period for the city of Athens, Greece, ArchontoulaChaloulakou (et al., 2003) used Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) methods. The training dataset contains two thirds of the records or cases that are available, and the remaining records or cases were equally divided into the validation and test set, so the dataset was partitioned into three unequal subsets before applying input to ANN [20]. In this study, ANN and MLR were compared, and the results show that ANN performs better than MLR. If an ANN is properly trained, this study claims that it will provide the required prediction answers or results.