Phase 4

Air quality analysis and Prediction in Chennai

Abstract: Air pollution level is increased due to industrial and commercial activities of major cities; the quality of air is being affected by emissions from industries and ever-increasing vehicular population. The study contains air pollutants such as Sulphur dioxide (SO2), Nitrogen (NO2), Respirable Suspended Particulate Matter (RSPM) from 2009-2018 per annum for locations in Chennai such as Anna Nagar, Adyar, Kilpauk, and T.Nagar. In this paper, using Support Vector Machine (SVM) algorithm the air pollution level can be predicted in the nearby future using R tool for statistics. This will be useful to take necessary preventive measure to reduce the air pollution level.

Keywords : Data Analytics, Prediction, Support Vector Machine, Air pollution.

I. INTRODUCTION

When the air is getting affected it will cause air pollution. Air pollution is harmful. It will also harm all types of living organisms. Both man-made activities and natural processes can produce air pollution. Air pollution majorly occurred due to the reason of air pollutants. The air pollutants are of two types such as primary, secondary and some minor air pollutants. The primary air pollutants are sulphur dioxide, nitrous oxide, types of particulate matter, carbon monoxide, carbon dioxide, volatile organic compounds, persistent free radicals, toxic metals, chlorofluorocarbons (CFC), ammonia, odours, and radioactive pollutants. The secondary air pollutants are particulates obtained from primary gaseous compounds, ground-level ozone, and peroxyacetyl nitrate. Minor air pollutants are hazardous air pollutant and persistent organic pollutant. Poor urban air quality and indoor air pollution are considered as the world's worst toxic problem in 2008 proposed by Blacksmith Institute. In 2014, according to the World Health Organisation (WHO) report, air pollution in 2012 caused the deaths of the circumference of 7 million people across the worldwide; this is an approximate calculation about by one from the International Energy Agency (IEA).

Chennai **formerly** called Madras located on the south-eastern coast of India in the north-eastern part of Tamil Nadu. As of 2017, there are more than 2.2 million households only 40 percent of the residents without owning a house. It is mostly clay, shale, and sandstone. It has 3 rivers and many lakes spread over the city. Due to reason, it lies on the thermal equator and cost that prevents an extreme variation in seasonal temperature so it has a tropical wet climate. Cyclones are unpredictable. Water supply treatment is managed by Chennai metro water supply and sewage by sewage board. It is divided into four broad regions: North, Central, South, and West. North is based on industrial area, South and West previously mostly residential, are fast becoming a commercial one. It has some major power station called North Chennai Thermal Power Station (NCTPS) appropriated around 25km from Chennai located near Ennore with a total equipped scope of 1,830 MW(2,450,00hp and this is one of the major power station in Tamil Nadu. The Chennai International Airport is the fourth busiest airport in India. It has eight mobile phone service companies operate 7 networks GSM networks including Airtel, Aircel, BSNL, Vodafone, Tata Docomo GSM, Idea, Reliance GSM and 3 CDMA networks including MTS, Reliance CDMA, Tata Docomo CDMA and internet connections such as 2G, 3G, and 4G mobile broadband are provided by few operators in the city. Analysing of SO2, NO2, and RSPM in the ambient air at four different selected regions namely Anna Nagar, Adyar, Kilpauk, T.Nagar of Chennai city from the period from 2009-2018.

1.2. Analysis

Ambient Air Quality Monitoring Programme at Chennai: Tamil Nadu Pollution Control Board(TNPCB) is operating four ambient air quality monitoring stations in Chennai under National Air Quality Monitoring Programme (NAMP) funded by Central Pollution Control Board such as Anna Nagar as a residential area which is originated in north-western of Chennai, T.Nagar as an affluent residential and commercial neighbourhood of Chennai, Adyar as premier residential area which is originated in southern suburb of Chennai, and Kilpauk as most preferred residential areas in Chennai The ambient air quality monitoring network measures a number of air pollutants at a number of locations in the country so as to meet objectives of the monitoring. Any air quality monitoring network thus involves selection of

pollutants, locations, and frequency. Network design also based on pollutant types in the atmosphere through various common sources, called common urban air pollutants, such as Sulphur dioxide, Oxides of Nitrogen, and Particulate Matter, Total Suspended Particulate Matter, Respirable Suspended Particulate Matter etc. The areas to be chosen primarily are such areas which represent high traffic density, industrial growth, human population and its distribution, emission source, public complaints if any and the land use pattern etc. Generally the basis of a network design pollution source and the pollutant present as most of the time

II.LITERATURE SURVEY

Udit et.al discussed that mitigation of poor air quality for health and Environmental reasons has been a primary focus for local governments and implement a real time project for air pollution detection and monitoring using wifi module. Munir Ahmad described the transformation and highly intelligent data mining approaches. Now allow organizations to collect, categorize and analyze user's reviews and comments from micro-blogging sites regarding their series and products. The SVM as well as provide them a baseline for future trends and comparisons.

Tony Ryan and Simon Armitagesummarised how nano poster absorbs pollution. Yves provided comprehensive description of the principles of the underlying algorithms and how they are applied to enhance prediction accuracy. YasinAkmAyturan et.al developed valid models to predict air pollution levels in order to determine future concentrators or to locate pollution sources. Wei-Zhen et.al Lu applied SVM to predict air pollutant levels in advancing time series based on the monitored RBF network in predicting air quality parameter with different time series and better generalization performance than the RBF model.Asha et .al discussed about different approaches and comparative analysis of data mining techniques like linear regression, back propagation and big data analytics like Map reducing, Geostatistical methods to predict urban air quality.

Wen-Jian Wang et.al proposed statistical learning theory that can be used for regression and series prediction. S.Gulia et.al discussed about urban air quality Management planning, Air Quality Control Region (AQCR) and Air Quality Models (AQMs) and their performance Evaluation. Rakesh Kumar yadav et.al stated that why PCA is utilized with numerous classifications and which one combination is ideal to use in future

III.METHODOLOGY

This paper contains the air pollution level data collected from selected regions of Chennai such as Anna Nagar, Adyar, Kilpauk, and T.Nagar from the period from 2009-2018. These data sets were collected from Tamil Nadu Pollution Control Board (TNPCB) online. These data sets of ten years from 2009-2018 is divided into two different data sets such as 2009-2013 as one data set and 2014-2018 as another. These data sets contain the values of air pollutants level such as Sulphur dioxide (SO2), Oxides of Nitrogen (NO2), and Respirable Suspended Particulate Matter (RSPM).

3.1. Air pollution analysis using Support Vector Machine (SVM) algorithm

By using a machine learning algorithm called Support Vector Machine (SVM), data prediction can be performed. Support vector machine algorithm is basically a classifier algorithm used mostly for finding the classification of data and for regression of data which means the prediction of data. The most objective of support vector algorithm is to find the best hyper plane between two classes. The hyper plane is used to separate the classes completely into two non-overlapping classes. SVM is used for functioning the linear data as well as non-linear data. In linear data, by using the hyper plane the model is easily classified but in most cases, the model does not classify correctly and this situation obtains in non-linear data. So for finding the hyper plane of non-linear data SVM uses kernel function. It is used to minimize the misclassifications and maximize the margin. While maximizing the margin this algorithm tries to manage the slack variable to zero. It does not minimize completely the misclassifications (NP-complete problem) but from the margin the sum of distances. SVM uses kernel trick used to find the kernel function to convert data into a large dimensional trait zone to make it likely to function the linear separation when the non-linear parameter does not base on the dimensionality of the trait zone.

3.2. Prediction works

To build the statistical graph of air pollutant level by using the R programming language, with the help of these air pollutant data sets. R is an open source tool and it is free to download and use. R language is derived from the S programming language which is trending in 2019 and even in future also because most of the statisticians using R as a user-friendly environment for build their statistical references. Based on this statistical reference the future data of air pollutants level such as SO2, NO2, and RSPM for selected regions of Chennai such as Adyar, Anna Nagar, Kilpauk, and T.Nagar with the help of machine learning algorithm namely Support Vector Machine(SVM). This future prediction data of the next five years will be used for taking some necessary preventive measure to reduce the pollution level of these selected four regions of Chennai. Fig 1 describes the process flow.
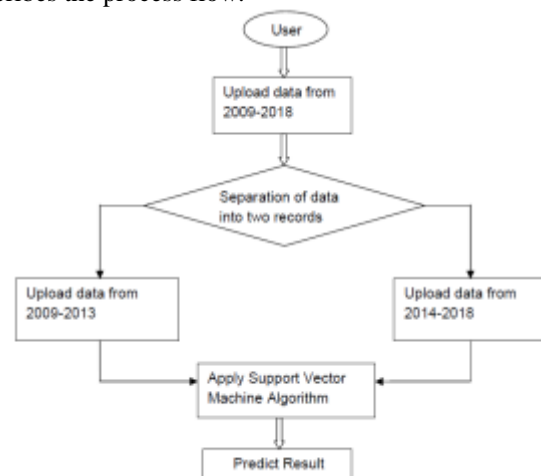
IV. RESULTS AND DISCUSSIONS

4.1. Variation in air pollutants

The air is polluted due to some air pollutants such as SO2, NO2, PM, RSPM, TSPM etc. These air pollutants are mainly occurred based on residential and commercial area. The NAMP monitor the air data for different locations in Chennai during the annual period of time. They monitor the air data for every month for each and every year and upload the data in online. They measure the data using microgram per meter cube (µg/m3). The data is based on all types of natural activities include meteorological (weather or climate-related data) of twenty-four hours for every year. Sometimes the data may be varied due to weather change.

Criteria for Sulphur dioxide (SO2)

Sulphur dioxide is the chemical compound which is highly toxic gas with pungent and irritating smell also a good reluctant. Based on volcanic activities So2 is produced such as the burning of smelting of mineral ores (aluminium, copper, zinc, lead, and iron) and fossil fuels (coal and oil) that contain sulphur.

Criteria for Nitrous oxide (NO2)

Nitrous oxide is produced from the reaction of nitrogen and oxygen gases in the air during combustion, especially at high temperatures. It heavily occurs due to motor vehicle traffic in large cities, the amount of NO2 entered into an air pollution atmosphere can be varied.

Criteria for Respirable Suspended Particulate Matter (RSPM)

Respirable Suspended Particulate is one the particulate matters of an aerodynamic diameter which is less than or equal to 10 micrometers, is also called as PM10. Combustion processes, vehicles, and industrial sources are the main reasons for particulate matter produce. If the particulate matter level will get increased at high level definitely it will damage our lungs as soon as possible.
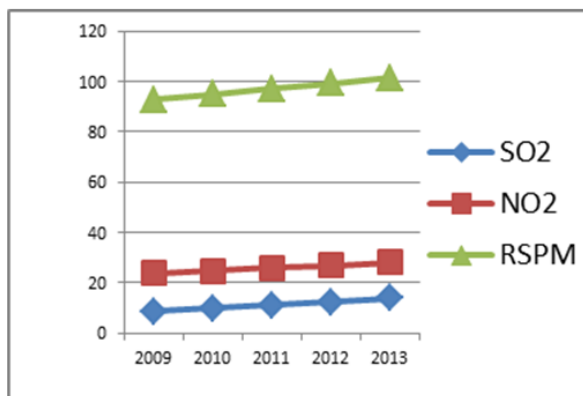
4.2. Statistical analysis

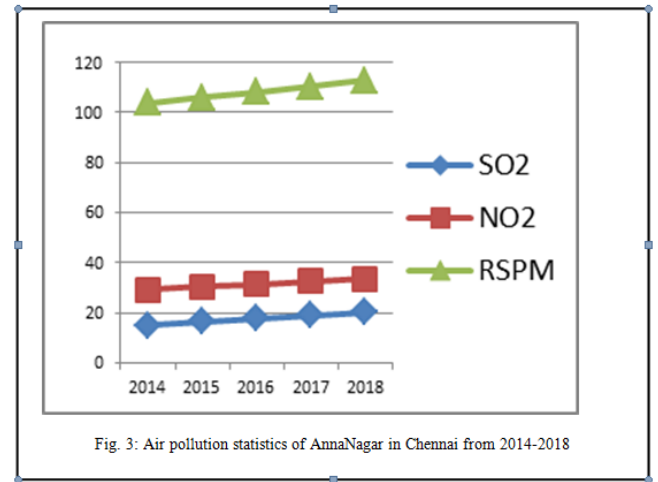Fig.2. Air pollution statistics of AnnaNagar in Chennai from 2009-2013

Fig. 3: Air pollution statistics of AnnaNagar in Chennai from 2014-2018

Fig.2 shows the statistical graph of air pollution level in Anna Nagar from 2009-2013 and Fig.3 shows the statistical graph of air pollution level in Anna Nagar from 2014-2018 from the above two graphs RSPM is highly occurred and then NO2 and SO2. This two graphs show that the air pollution level in increasing order and it represents the linear data.
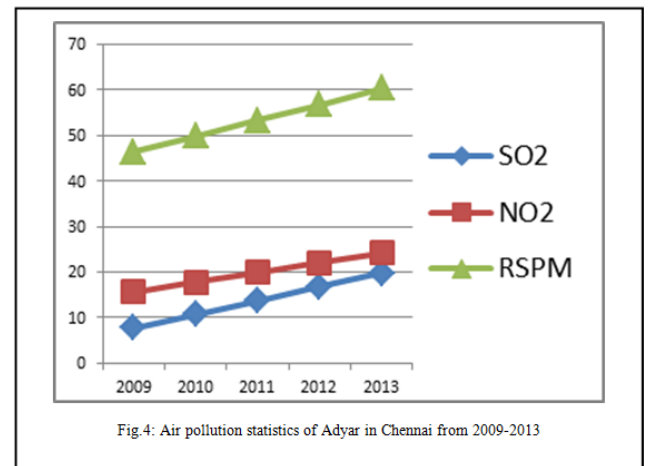

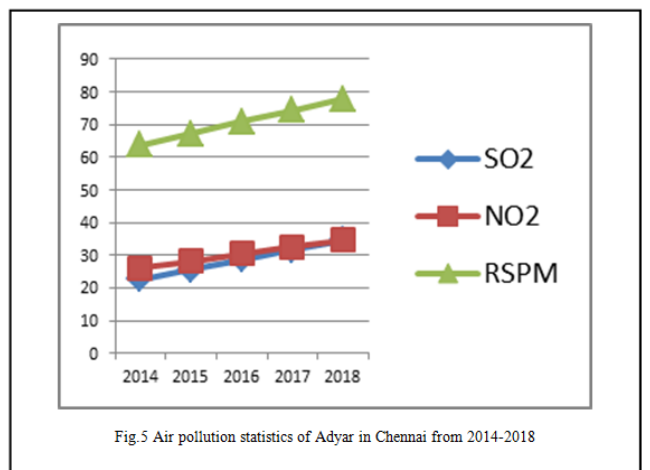Fig.4: Air pollution statistics of Adyar in Chennai from 2009-2013

Fig.5 Air pollution statistics of Adyar in Chennai from 2014-2018

Fig.4 shows the statistical graph of air pollution level in Adyar from 2009-2013 and Fig.5 shows the statistical graph of air pollution level in Adyar from 2014-2018 from the above two graphs RSPM is highly occurred and then NO2 and SO2. This two graphs show that the air pollution level in increasing order and it represents the linear data.

Fig.6: Air pollution statistics of Kilpauk in Chennai from 2009-2013

Fig 7: Air pollution statistics of Kilpauk in Chennai from 2014-2018
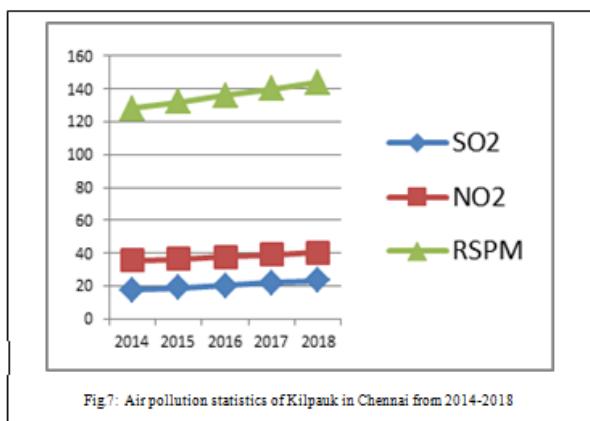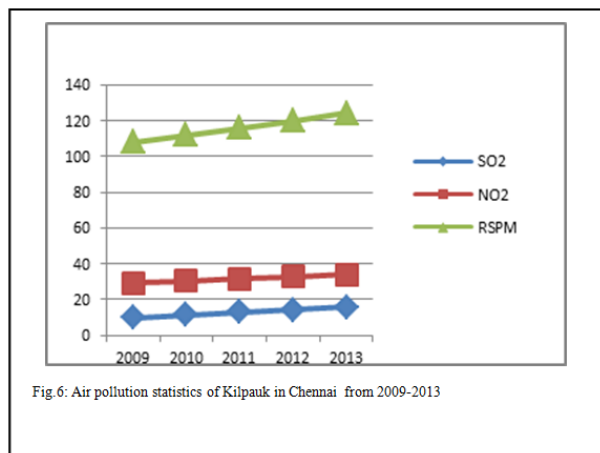
Fig.6 shows the statistical graph of air pollution level in Kilpauk from 2009-2013 and Fig.7 shows the statistical graph of air pollution level in Kilpauk from 2014-2018 from the above two graphs RSPM is highly occurred and then NO2 and SO2. This two graphs show that the air pollution level in increasing order and it represents the linear data.

Fig.8: Air pollution statistics of T.Nagar in Chennai from 2009-2013

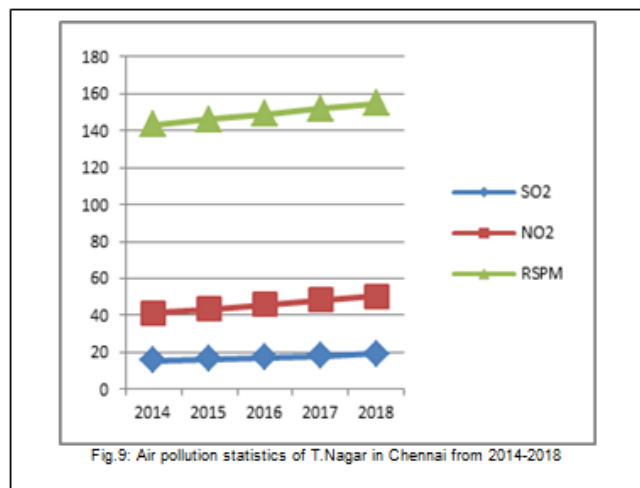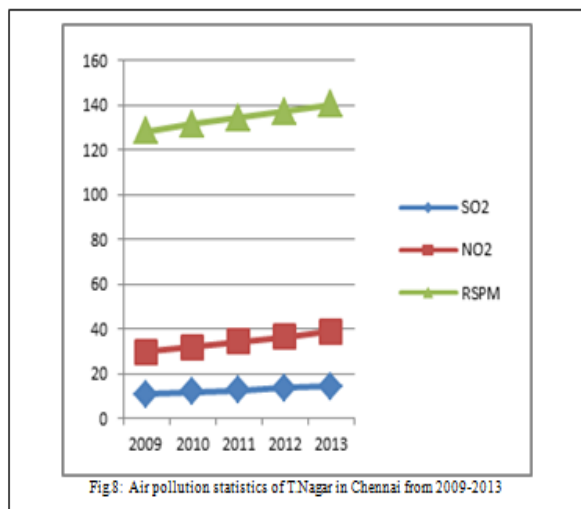Fig.9: Air pollution statistics of T.Nagar in Chennai from 2014-2018

Fig.8 shows the statistical graph of air pollution level in T.Nagar from 2009-2013 and Fig.9 shows the statistical graph of air pollution level in T.Nagar from 2014-2018 from the above two graphs RSPM is highly occurred and then NO2 and SO2. This two graphs show that the air pollution level in increasing order and it represents the linear data.
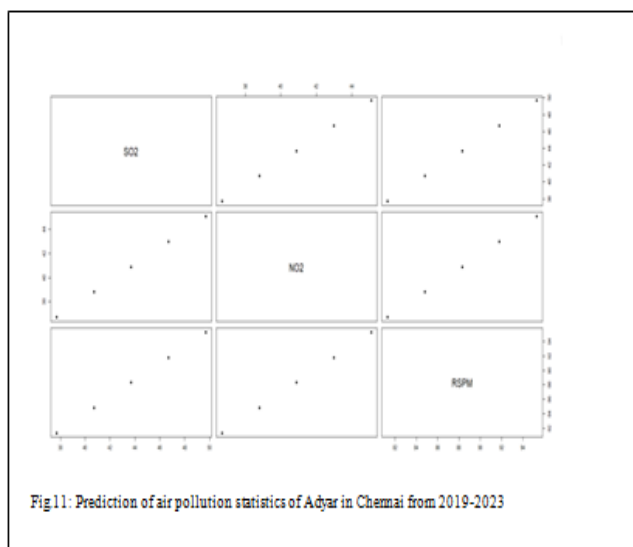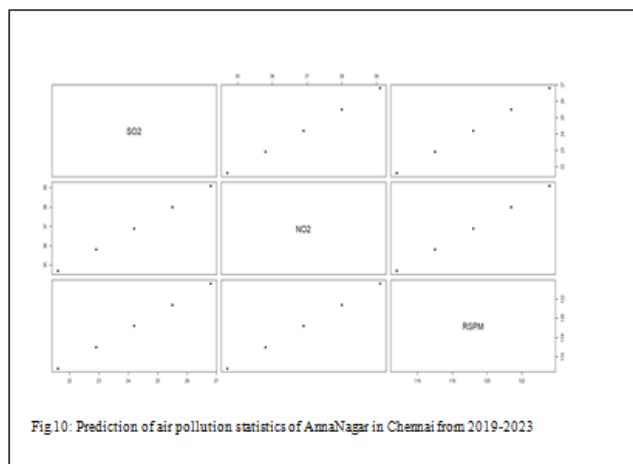

Fig.10: Prediction of air pollution statistics of AnnaNagar in Chennai from 2019-2023

Fig.11: Prediction of air pollution statistics of Adyar in Chennai from 2019-2023

Fig.10 and Fig.11 represents the prediction of air pollution statistics such as SO2, NO2, and RSPM of selected regions of Chennai from 2019-2023 using Support Vector Machine (SVM) algorithm. The prediction result also builds in linear model. The annual air pollution statistics up to future prediction of selected regions of Chennai such as Annanagar, Adyar, Kilpauk, and T.Nagar for air pollutants such as SO2,NO2, and RSPM from 2009-2023 are mentioned below. In this graphs, the x-axis referred as years range from 2009-2023 and y-axis referred as air pollutants level.
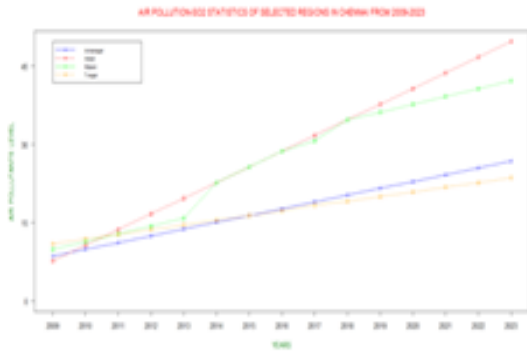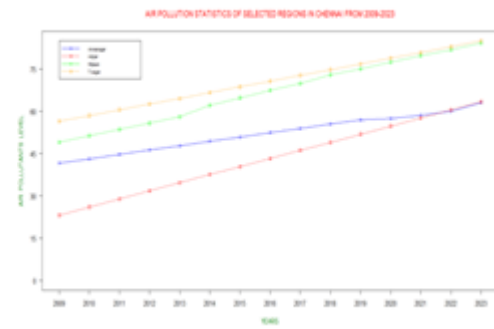

Fig.12: Annual SO2 statistics of selected regions in Chennai from 2009-2023
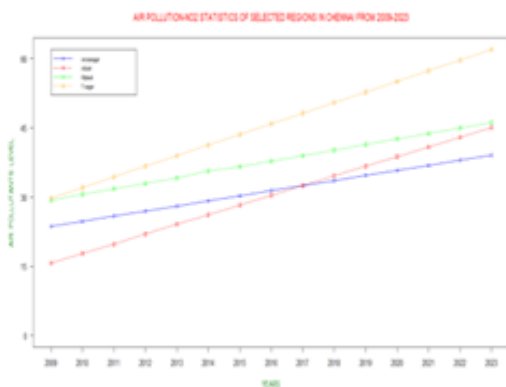
Fig.13: Annual NO2 statistics of selected regions in Chennai from 2009-2023
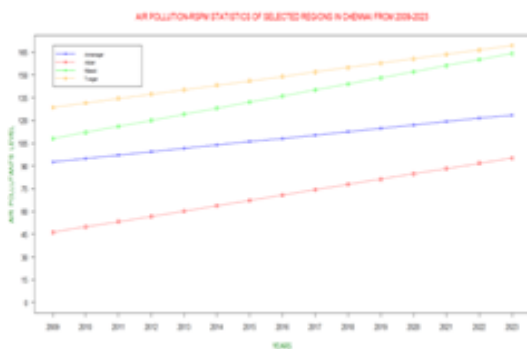
Fig.14: Annual RSPM statistics of selected regions in Chennai from 2009-2023

Fig.15: Annual air pollution statistics of selected regions in Chennai from 2009-2023

Fig.12, Fig.13, Fig.14 and Fig.15 represents annual air pollution statistics consists of average of SO2, NO2, and RSPM of selected regions of Chennai from 2009-2023. In prediction result, it shows T.Nagar is considered as highly polluted area, Kilpauk is considered as second polluted area, Annanagar is considered as third polluted area, and Adyar is considered as fourth polluted area. Burning mineral ores such as aluminium, copper, zinc, lead, and iron, burning of fossil fuels such as coal and oil, and weather data activity changes. These are the main reasons will consider SO2 is highly occurred. Motor vehicle traffic is the main reason for high accuracy of NO2. T.Nagar is a large area of public place, market areas, streets with schools and colleges, commercial streets and attracts a lot of pedestrian traffic. These are the main reasons will consider T.Nagar as highly polluted area among these selected four areas of Chennai.

4.3. Air Pollution Preventive Measure and Reduction Efforts

Most of the developed countries using the concept of land-use planning for getting better infrastructure and for their future purpose. This plan is used to divide their living areas into different blocks and mapped it clearly before itself. Properly dispose the e-waste and medical waste not in public place. Try to create a civic education on Vehicle pollution, try to reduce the combustion of fossil fuels like coal and oil, and change to pure the power origins such as hydral, solar, and wind powers will reduce half of the air pollution level. Titanium dioxide (TiO2) is a nano particle pollution eater. According to Prof Simon Armitage and Prof Tony Ryan implemented the idea successfully of reducing pollution sources by using Titanium dioxide. But it doesn't eat all the pollutants of air except NO2. This detail proof is mentioned in the below reference [12]. These are the possible methods to reduce air pollution in selected regions of Chennai in future.

V.FUTURE ENCHANCEMENTS

In 1981, the rules of Air act (Pollution Control and Prevention) the whole states of Tamil Nadu have been considered as an air pollution control area. In that, Chennai is one of the most polluted cities in Tamil Nadu under this act because it is considered as a highly growing residential and