International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Air Pollution Prediction in Smart Cities by using Machine Learning Techniques

K. Rajakumari, V. Priyanka

Abstract: ***The urban air pollution has an immediate effect on man health specifically in developing and mechanical countries. It can cause health issues such as cancer, cardiovascular diseases and high mortality rates. Continuous checking of contamination empowers the metropolitans to dissect the present traffic circumstance of the city and take their decision accordingly. Existing exploration has utilized diverse AI apparatuses for pollution forecast; notwithstanding, relative examination of these methods is regularly required to have a superior comprehension of their handling time for numerous datasets. In this work, we look at forecasting the air contamination by dealing with parameters of three different gases like SO_2,NO_2,O_3.This process involves to pre-processing the times series. However, pre-processing involves a similarity measure, we explore the use of Dynamic Time Warping (DTW),LSTM,ARIMA Model for time series prediction, K-means, Support Vector Regression is then used to classify the spatio-temporal pollution data of different areas over a period of 10 years.***

Keywords: Air pollution forecasting, Machine learning algorithms, pseudo code.

I.INTRODUCTION

Air contamination creates the threat of respiratory and coronary affliction in the population. This is a biggest health issue affecting whole population. Air contamination is by and by apparent as the single greatest common human hazard. disease brought about via air pollution like stroke, ischemic coronary illness, constant obstructive pneumonic disease, lungs malignant growth and intense lower respiratory infection. Worldwide, encompassing air contamination participates 6.7% of all deaths.3.7 million deaths were attribute able to AAP in 2012.Most of these deaths, about 88%, happen in lower and middle salary nations [1]. In Africa, it is evaluated that more than 800,000 deaths for each year are because of air contamination. In South Africa, outdoors contamination was assessed to cause3.7%of national mortality from cardiopulmonary infection and 5.1% of mortality owing to malignant growths of the trachea, bronchus and lungs in grown-ups more established that 30 years old in 2007. Indoor air contamination is additionally a significant issue with ±20% of South African family units presented to 10 air quality screen contraptions. At the point when all is said in done, each station simply has one screen gadget. It should be kept up at between times, in like manner there will be no yields for the station when the contraption is being kept up, recalibrated, or has various issues. Third, the sorts of urban air related information are diverse for the improvement of data acquirement developments. In any case, there isn't an inside and out recognized judgment to uncover the essential driver of the occasion and scattering of air tainting. In this way, it is hard to mention to that what sorts of data are the essential significant features for addition and figure, and the key factors for condition divisions to hinder and control air pollution. This paper is stirred to focus all of these troubles by utilizing the information available in the unlabeled information what's more, the spatio-transient data, and performing feature decision and connection assessment for the urban air related information. This is exacerbated by concerns of temperature rise, global warming and rise in sea levels. Several international organizations such as the Paris Climate Agreement initiated by the United Nations aim to reduce the effect of pollution. Moreover, nations are dedicated to developing techniques to reduce pollution as well as methods to reduce the usage of fossil fuels. The United Nations primarily stress that it is important to predict the carbon footprint of each nation. This is necessary to initiate policy decisions and regulations which need to be implemented by the respective country to curb the ill effects. This consequently stresses on the development of reliable methods to predict the pollution levels in a country over a period of time as well as to predict the levels of different pollutants for targeted solutions. Countries deploy many sensors to record different pollutant levels in urban areas as well as near industrial zones, but the main index used by governments to depict the pollution levels is the Air Quality Index. This is an important measure as it helps to determine the overall quality of air which consequently is used to determine the adverse health and climate effects which are caused to the environment. In this work, we present a spatiotemporal prediction model which could be highly effective in determining the AQI as well as individual pollutant levels over a period of time. To overcome the limitations posed by large-scale data and high variability we adopted a unique combination of approaches to predict, accurately, the different pollution levels as well as the AQI. We conducted several experiments using different models and determined a low cost-complexity combination of models.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Dr.K.Rajakumari, Assistant Professor, Department Of Computer Science And Engineering, Sri Shakthi Institute Of Engineering And Technology, Coimbatore, TamilNadu, India,raji1anju@gmail.com
Priyanka V, PG-Scholar, Department Of Computer Science And Engineering, Sri Shakthi Institute Of Engineering And Technology, Coimbatore, TamilNadu, India,priyasrishakthi@gmail.com

II.THEORY

Due to the varied topography and extent of industrialization in the city, predicting environment pollutant values help in foreseeing the effect and extent of pollution.

Several machine learning algorithms are used to forecast environmental facts. The data set provides information on the city, NO2,O3,SO2 levels for through 10 years. Relation between the pollutants to their geographical locations translates the problem into a classification issue. Compared to other methods, SVM is particularly useful since the data involves a time series and it is not suitable for non-linearly related. This method can also provide a better generalization error. In order to predict continuous values, however, leads to the use of a variation of SVM - SVR,LSTM,ARIMA

MODEL,K-means

clustering algorithms.

A. Data Set description

The data used for analysis has been obtained from the pollution control board department for 10 years. It contains statistical data such as mean, max and min for 3 kinds of pollutants namely, Nitrogen Dioxide, Sulphur Dioxide, Ozone. It also contains temporal data in the form of dates and the data was collected every 6 hours. Moreover, the dataset contains Geospatial data in the form of area code , address of area, necessary values of the data collection station.

B. Supervised learning

The supervised machine learning algorithms are those algorithms which needs outside help. To be classified and each branch represents a value that the node can take.The input dataset is separated into train and test dataset. The train dataset has output variable which should be predicted or grouped. All calculations take in an examples from the preparation data set and apply them to the test data set for forecast or classification [4].The work process of supervised machine learning algorithm is given in Fig.1.Three most acclaimed regulated supervised machine learning algorithms have been talked about here.1) Decision Tree: Decision trees are those kind of trees which gatherings traits by arranging them dependent on their qualities. Choice tree is utilized predominantly for arrangement reason. Each tree comprises of hubs and branches. Every hub speaks to characteristics in a gathering that will be arranged and each branch speaks to a worth that the hub can take.

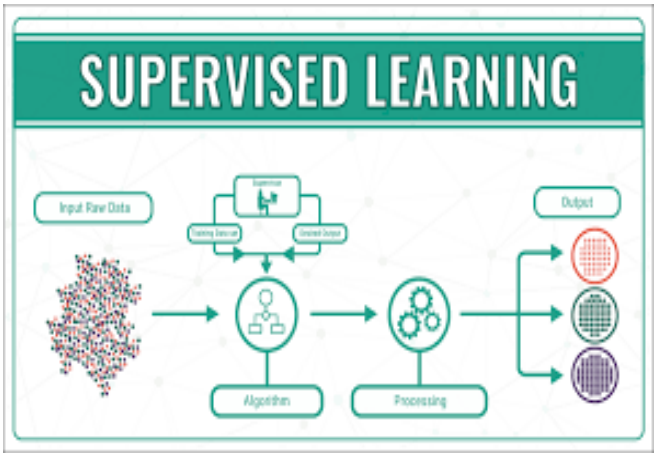

Fig.1.Supervised learning

i)Support Vector Machine:

Another most generally utilized best in class supervised learning algorithm is Support Vector Machine (SVM). It is principally utilized for order. SVM chips away at the rule of edge figuring. This essentially, makes edges between the classes. The edges are attracted such a style, that the separation between the AyonDey/(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3),2016, 1174-1179 www.ijcsit.com 1175 edge and the classes is greatest and subsequently, limiting the characterization blunder

ii) Support Vector Regression

The Support Vector Regression (SVR) uses vague principles from the SVM for gathering, with only a few minor contrasts. Above all else, on the grounds that .yield is an authentic number it ends up being difficult to predict the present information, which has boundless potential results. Because of backslide, an edge of resistance (epsilon) is set in estimate to the SVM which would have just mentioned from the issue. In any case, other than this reality, there is likewise an increasingly confounded explanation, the calculation is progressively confused along these lines to be taken in thought. In any case, the primary thought is consistently the equivalent: to limit blunder, individualizing the hyperplane which enlarges the edge, recollecting that bit of the error is persevered. The model created by SVR just depends upon a subset of the readiness data, considering the way that the cost limit with respect to building the model ignores any arrangement data that is close (inside an edge ε) to the model desire. Assume we are given preparing information {(x1, y1),...,(x, y)} ⊂ X × R, where X indicates the space of the information designs (for example X = Rd). In ε-SV relapse, we will likely discover a capacity f (x) that has all things considered ε deviation from the really acquired targets yi for all the preparation information, and simultaneously is as level as could reasonably be expected. At the end of the day, we couldn't care less about mistakes as long as they are not exactly ε, however won't acknowledge any deviation bigger than this.

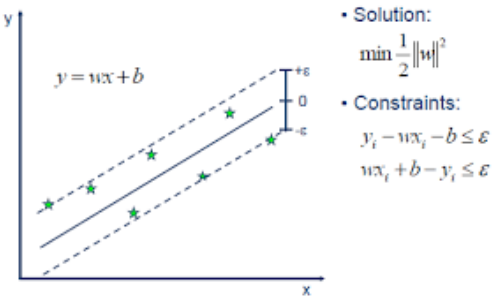

Fig. 2.Support Vector Regression

Fig 2: Hyperplane for SVR graph is shown here. Only the points outside the region covered by dotted lines contribute to the cost insofar, as the deviations are penalized in a linear fashion.

C.Unsupervised learning

Exactly when new data is displayed, it uses the as of late learned features to see the class of the information.It is for the most part utilized for prediction and clustering. A work process of unsupervised learning is given in Fig. 3.

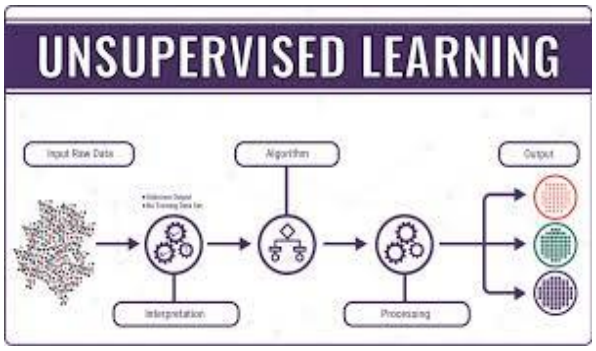

Fig. 3.Unsupervised Learning

i) K-means Clustering

Clustering or grouping is comes under unsupervised learning The things which has comparative qualities are placed in a similar cluster. This calculation is called k-means. Which is a least complex calculation and famous unaided unsupervised learning method. K-means calculation recognizes k number of centroids, and then dispenses each datum point to the nearest cluster, while keeping the centroids as little as could be allowed. Here the information bunched by month insightful and region astute to figure the contamination. technique that when starts, makes bunches consequently.The things which has comparable attributes are placed in a similar cluster. This algorithm is called k-means.K-means algorithm recognizes k number of centroids, and afterward apportions each datum point to the closest bunch, while keeping the centroids as little as could be expected under the circumstances. Here the data clustered by month wise and area wise to forecast the pollution.
K-MEANS(P,k)

Input: A dataset of points P={$p_1,\ldots,p_n$) a number of clusters K

Output:centres{$c_1,\ldots,c_k$}implicitly dividing P into k clusters

Choose k initial centres C={$c_1,\ldots,c_k$}

1. While stopping criterion has not been met
2. do //assignment step:
3. for i=1,…N //Loop initializing
4. do find closest centre $c_k \epsilon$Cto instance p_i
5. assign instance pi to set C_k
6. update step:
7. for i=1,…,k
8. do set c_i to be the centre of mass of all pointsin C_i

Pseudocode of K-Means clustering algorithm

D.SemiSupervised learning

Semi–supervised learning algorithm is a strategy which consolidates the intensity of both regulated and solo learning it will in general be natural item full in those zones of AI and data mining where the unlabelled data is starting at now present and getting the named data is a monotonous system [5]. There are numerous classes of semi-supervised learning [6]. Some of which are examined beneath:

i)Generative Models

Generative models are one of the most established semi-directed learning strategy expect a structure like p(x,y) = p(y)p(x|y) where p(x|y) is a blended conveyance for example Gaussian blend models. Inside the unlabeled information, the blended parts can be recognizable. One marked model for every part is sufficient to affirm the blend circulation.

ii) Self-Training

In self-training, a classifier is prepared with a part of named information. The classifier is then encouraged with unlabeled information. The unlabeled focuses and the anticipated names are included in the preparation set. This system is then rehashed further. Since the classifier is learning itself, thus the name self-preparing**.**

iii) Transductive SVM

Transductive help vector machine or TSVM is an expansion of SVM. In TSVM, the named and unlabeled information both are taken. It is utilized to name the unlabeled information so that the edge is most extreme between the named and unlabeled information. Finding a careful arrangement by TSVM is a NP-difficult issue.

E.ReinforcedLearning

Reinforcement learning is a sort of realizing which settles on choices dependent on which moves to make to such an extent that the result is increasingly positive. The student has no information which moves to make until it's been given a circumstance. The move which is made by the student may influence circumstances and their activities later on. reinforcement adapting exclusively relies upon two criteria: experimentation search and deferred result [7]. The common model [8] for reinforcement learningis portrayed in Fig 4.

Fig. 4.Reinforcement learning

In the figure, the specialist gets an information I, current state s, state change r and info work I from nature.
In light of these data sources, the operator produces a conduct B and makes a move a which creates a result. E. Perform Multitask learning which has a straightforward objective of helping different students to do better. When perform multiple tasks learning calculations are apply to undertaking, it recalls the methodology how it tackled the issue or how it scopes to the specific end. The calculation at that point utilizes these means to discover the arrangement of other comparable issue or errand. This encouraging of one calculation to another can likewise be named as inductive exchange system. In the event that the students share their involvement in one another, the students can adapt simultaneously instead of separately and can be a lot quicker [19]. F. Outfit Learning at the point when diverse individual understudies are merged to outline only a solitary understudy then that particular sort of learning is called assembling learning. The individual understudy may be Naïve Bayes, choice tree, neural system, and so on. Outfit learning is a hotly debated issue since 1990s. It has been seen that, an assortment of students is frequently better at making a specific showing as opposed to solitary students [9]. Two well known Ensemble learning procedures are given beneath [10]:

i) Boosting

Boosting is a system in troupe realizing which is used to decrease predisposition and fluctuation. Boosting makes an assortment of frail students and convert them to one in number student. A feeble student is a classifier which is hardly compared with real order. Then again, a strong student is a kind of classifier which is unequivocally corresponded with genuine classification [10].

ii) Bagging

Bagging or bootstrap accumulating is applied where the exactness and security of an machine learning ought to be expanded. It is relevant to classification and regression Packingadditionally reducesdifference and aides in dealing with overfitting [11]

F.ARIMA

Autoregressive Integrated Moving Average model is a present moment (at any rate 40 information focuses) time-arrangement forecast model which can be utilized basically for foreseeing information which has a low fluctuation or less anomalies and will in general follow a steady pattern. This model is generally appropriate for information which shows a significant level of regularity. On the off chance that an absence of regularity, there is a high possibility that the estimations related with the model won't be processed because of specific imperatives. We tried this model on our information with the NO2 mean, and time traits from all territories with slack estimation of 5 and got a RMSE of 2.006. While this worth is little, there is a noteworthy issue to be considered. For this, the information is the mean estimation of contamination level for a specific date over every one of the territories. For a specific date, each state will bring about a similar forecast.
Algorithm :Finding optimal ARIMA

```
1.procedureFINDOPTIMALARIMA
2.  aic  ←inf
3.  for p  ←0 to 3 do// Outer loop
4.   for d  ← 0 to 2 do  // inner loop
5.    for q  ← 0 to 3 do  //inner most loop
6.     model ← fit(arima  ( p,d,q,allow_drift ←
7.     True, allow_mean ← True),x)
8.     aic_curr← compute_AIC(model)
9.     if aic_curr<aicthen //condition checking
10.      model_opt  ←model  //assignment
11.       aic← aic_curr  //assignment
12. return model_opt  //return the value
```

Pseudocode of ARIMA Model

G. LSTM(RNN)

The basic role of LSTM is to adapt long haul conditions in the information and forestall the long haul reliance issue for example keeping the model from "recalling" the information over a significant stretch of time. This strategy is combined with Recurrent Neural Networks with a couple of changes in the Recurrent hubs of the RNN.
The fundamental change in the rehashing hub is that rather than one neural system layer, there are four layers which connect with one another. These four layers are themselves utilized in a "repetitive" way utilizing entryways to permit/forbid data through them. The four layers are: Three sigmoid layers and a tanh layer. We tried this model on our information and got a RMSE of 3.378 utilizing 30 shrouded layers and an age of 5. This underlying appraisal is a decent marker of the precision of the model.
Expanding the quantity of ages would additionally upgrade the RMSE yet because of absence of processing assets, the expectations related with it have not been determined. The diagram appeared underneath is an unpleasant gauge of how precise the model is over the information.The blue diagram speaks to the genuine purposes of the information while the orange chart shows the anticipated estimations of the model.

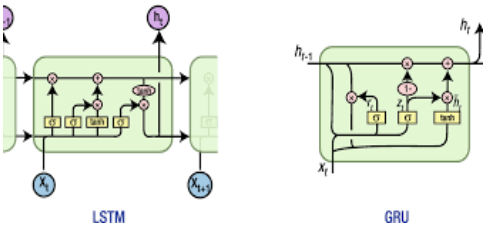

Fig. 5.Crux of LSTM

III.METHODOLOGY

There are two primary phases in the system:

1. Training phase: The training data set is used to trainthesystem and fits a model (line/curve) based on the algorithm chosen accordingly.

2. Testing phase: The system is provided with the inputs and is tested for its working. The accuracy is checked then and there. And therefore, the data that is used to train the model or test it, has to be taken appropriate.

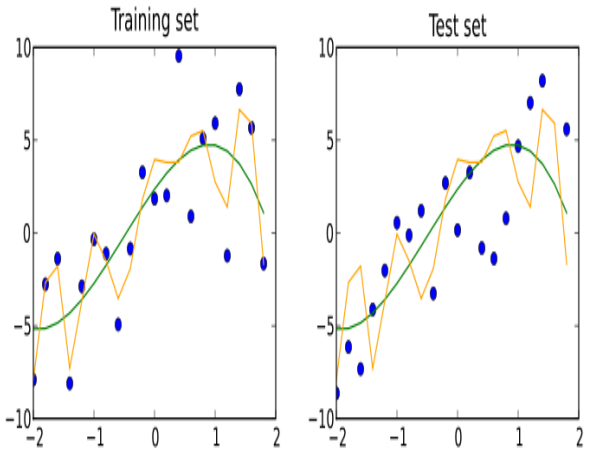

Fig.6.Phases of Machine learning

The framework is intended to distinguish and foresee contamination level and subsequently proper calculations must be utilized to do the two distinct assignments. Before the calculations are chosen for additional utilization, various calculations were looked at for its exactness. The appropriate one for the assignment was picked. The information on similitude between time arrangement is broadly utilized for discourse acknowledgment and mark acknowledgment. In our task, we utilize two bits of information - factors impacting contamination and regularity saw over 10years. Concerning these ideas we decide the closeness between time arrangement of numerous regions and the similitude between time arrangement of the 120 months in the years. If you don't mind note that we have worked to a great extent with NO2 information as this has been believed to be the reason for lung infections. Through inspecting papers, for example, [1], we comprehend that air contamination in one region is influenced by topology, atmosphere and mechanical exercises in contiguous territories. By only fitting a relapse model on the whole information. In this manner, to conquer such a hindrance impact, we dig into bunching air contamination arrangement of individual region. The subsequent bunches join those time arrangement that are comparative, for example they would be topologically, geologically and climatically comparative.

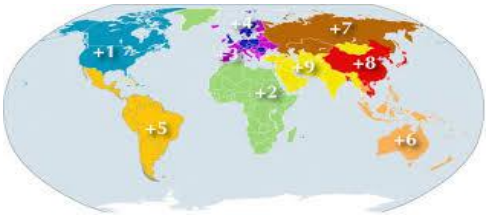

Fig.7. NO2 mean distribution generated from our dataset.

NO2 mean circulation created from our dataset. This viably manages the spatio-transient conduct of the information. While seeing the information gathered, we had additionally seen a regularity in the data]. This was supported when grouping did. This data would then be able to be fit onto relapse models, for example, ARIMA to foresee atmosphere based contamination statistics. merely utilizing Euclidean separation brings about no weight for stage moved time arrangement. For example: in the event that double cross arrangement are T0: 1213110 and T1: 8121311, at that point with Euclidean separation, the separation between the two is determined piecewise. (1 and 8), (2 and 1), (1 and 2) and so on. Nonetheless, as it is recognizable, the double cross arrangement vary just by one position. This doesn't make the arrangement as removed as Euclidean separation finishes up it to be. Dynamic Time Warping, a framework N*N is made with the squared good ways from one in the main, time arrangement to each point in the other time arrangement. With the above model, the network underneath is gotten. Each component is (t0 - t1)^2. With the assistance of this network, those separation components are picked to such an extent that the total of the arrangements is the base entirety. In the event that that is the situation, the featured components would be picked. Along these lines, the stage distinction between the two arrangement doesn't add to the separation. In any case, it is one-sided towards lessening the previously mentioned impact and looks at the last information purpose of a period arrangement to the first of another. Henceforth rules, for example, Boundary conditions - limiting the arrangement got from the framework to start and end at the slanting parts of the bargains, coherence conditions - confining the quantity of components contrasted with locate the most limited separation, monotonicity condition - ensuring the focuses thought about are dispersed in time from the last cycle.

Table- I: DTW Matrix

t_0/t_1	1	2	1	3	1	1	0
1	0	1	0	4	0	0	1
1	4	1	0	4	0	0	1
3	0	1	1	0	4	4	9

1	1	1	0	4	0	0	1
2	0	0	1	1	1	1	4
1	0	1	0	4	0	0	1
8	49	36	49	25	49	49	64

IV.IMPLEMENTATION

The first step for implementation is to pre-process the data available for training the model. The high dimensionality had an adverse impact on the model building.

The attributes included pollution level prediction for 3 different pollutants. Thus we trimmed down to one pollutant - NO2. The same model can be applied to different pollutant attributes as per requirement. Also, the data provided is for 10 years with hourly values for all the days. We trimmed this down to one value per day which will be mean of all values recorded for the day. This reduced the dimensionality of the time-series data and made it possible to train the model given our hardware and time constraint. For geographic attributes, the dataset contained information of 3 different gases. We generalized the data to the areas level to ensure that we can identify similarity in data patterns for nearby areas, hence reducing geographic dimensionality.

Our initial approach involved using Support Vector Regression (SVR). We implemented the Support Vector Machine algorithm and tried to extend it into SVR

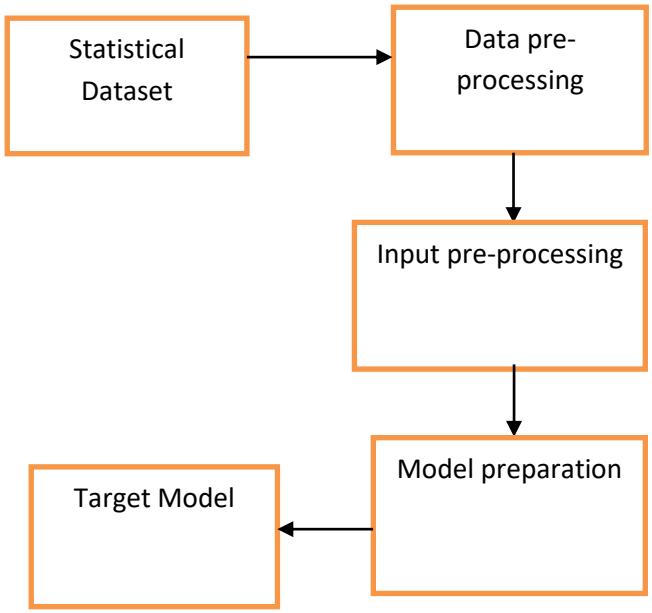

Fig. 8.Implementation flowof the project

The algorithm to generate possible hyperplanes, we used the approach same as SVM, but with the change that the hyperplane with maximum data points in given margin was selected. This hyperplane hence is the SVR for the data. Due to mathematical complexities that we could not get through with, we switched to the implemented method. Nonetheless, implementing SVR using python libraries, we were able to successfully validate that SVR is one of the correct ways to carry out the objective of the project. Post-pre-processing, the next logical step was to establish correlation among different levels of data in context of timeseries present in the data. Initial experiments were focused on discovering time-series' at the 'month' level of the data, for each day of the given month. The graph shown below is for a sample of the data. The graph clearly indicates the extreme variance and lack of correlation between different values. This consequently shows a definitive lack of seasonality for a given month. To resolve this issue, we analysed yearly time-series' of different states and established that there is noticeable correlation between time-series' of different areas. Thus, to increase intra-cluster correlation, the time-series' of different areas which are similar to each other are merged.

Time series analysis algorithm.

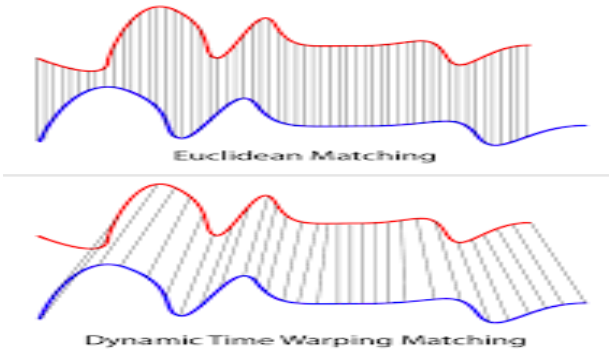

Fig. 9.DTW vs Euclidean

In the wake of getting the time arrangement information, we have to discover to decide the separation between two, time arrangement to shape bunches. Euclidean is one of the most widely recognized strategies used to decide the separation, however it isn't viable for time arrangement information. Consequently, we utilize weighted Dynamic Time Warping to ascertain the arrangement between any two given time arrangement. Dynamic time traveling finds the ideal nonlinear arrangement between two, time arrangement. To measure this outcome, we determined the arrangement between the time arrangement's of 2000 and 2001. In the wake of applying both the strategies .the outcomes are:

Euclidean distance = 125

Dynamic Time Warping (window size = 10) = 73

Table- II: Parameter choices

K-Means	Cluster-4 Iteration=10
DTW	Alignment=5 LB_Keogh Reach: 5
ARIMA	Lag order: 5 Degree of differencing: 1

Presently attempt to fit a regression line over constantly arrangement's in a given cluster.

This regression line can be utilized to anticipate a future time-arrangement for the states that zone related with the cluster. For our case, we have utilized Autoregressive Integrated Moving Average (ARIMA). Given our information area, we recover the group to which the area has a place and utilize the regression that was fit for the cluster to predict the pollution level.

V.RESULTS AND DISCUSSIONS

For forecasting air pollution level, we thought about different models. The most appropriate strategy according to our assessment is to group states with comparable conduct of pollution levels.

After fruitful execution of the undertaking, the outcomes demonstrated that K-Means clustering followed by ARIMA can be utilized for time series expectations. We see that pollution levels follow seasonal behaviour.Utilizing K-Means clustering, expresses that follow comparative contamination level conduct were cluster together.For calculating distance for clustering, we thought that Euclidean distance is not the better approach. Dynamic Time Warping is the better measures to calculate the alignment between two time-series. Actualizing DTW alongside LB-Keogh (lower bound DTW) affixes the DTW for the given huge dataset. At long last, ARIMA can be utilized to make time-series regression over the groups for a forecast. This furnishes with one time series regression line for each group. Along these lines, we have k clusters, each speaking a group of similar behaviour patterns, with each cluster fitted to one regression line each.It can, in the future, be extended to cover the time series data, with more computation power,It can be carried out at hourly basis. This provides a more drilled/scrutinized analysis of the time-series data. With a proper understanding of mathematics behing powerful machine learning algorithms will be utilized in future for better accuracy.

Table- III: Month wise clustering

Cluster	Months
1	July, August
2	January, December
3	February, March, October, November
4	April, May, June, September

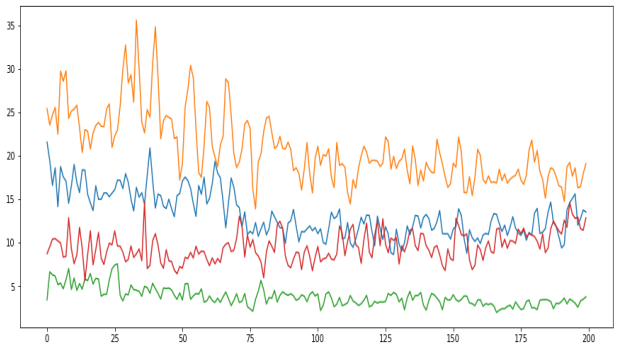

Fig.10. Cluster centroids of area-wise clustering

VI.CONCLUSION

In this research work ,the prediction of Air pollution particles is carried out by using various machine learning algorithms, and optimized it with best time series supporting algorithms like ARIMA MODEL.

REFERENCES

1. National Patterns in an environmental injustice inequality Air pollution in the united states LARA P. CLARK,DYLAN B.,MILLET.
2. World Health Organization,"Monitoring ambient air quality for health impact assessment," WHO Regional Office Eur., Copenhagen, Denmark, Tech. Rep. 85, 1999.
3. U. Gehring et al., "Traffic-related air pollution and the development of asthma and allergies during the first 8 years of life," Amer. J. Respiratory Critical Care Med., vol. 181,no. 6, pp. 596–603, 2010.
4. S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249-268
5. X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, Vol. 3, No. 1, Pages 1-130
6. X. Zhu, "Semi-Supervised Learning Literature Survey", Computer Sciences, University of Wisconsin-Madison, No. 1530, 2005
7. R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992
8. L. P. Kaelbing, M. L. Littman, A. W. Moore, "Reinforcement Learning: A Survey", Journal of Artificial Intelligence Research, 4, Page 237-285, 1996
9. D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, Pages 169-198, 1999
10. Z. H. Zhou, "Ensemble Learning", National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China.
11. https://en.wikipedia.org/wiki/Bootstrap_aggregating
12. World Health Organization, "WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfurdioxide,"WorldHealthOrg.Geneva,Switzerland,Tech.Rep.WHO/ SDE/PHE/OEH/06.02, 2005
13. O. A. Postolache, J. M. D. Pereira, and P. M. B. S. Girao, "Smart sensors network for air quality monitoring applications," IEEE Trans. Instrum. Meas., vol. 58, no. 9, pp. 3253–3262, Sep. 2009.
14. V. Sharma, S. Rai, A. Dev, "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10, October 2012
15. S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014
16. A. Kumar, H. Kim, and G. P. Hancke, "Environmental monitoring systems: A review," IEEE Sensors J., vol. 13, no. 4, pp. 1329–1339, Apr. 2013.
17. World Health Organization, "WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide," World Health Org., Geneva, Switzerland, Tech. Rep. WHO/SDE/PHE/OEH/06.02, 2005
18. O. A. Postolache, J. M. D. Pereira, and P. M. B. S. Girao, "Smart sensors network for air quality monitoring applications," IEEE Trans. Instrum. Meas., vol. 58, no. 9, pp. 3253–3262, Sep. 2009.
19. S.-C. Hu, Y.-C.Wang, C.-Y.Huang, and Y.-C. Tseng, "Measuring air quality in city areas by vehicular wireless sensor networks," J. Syst. Softw., vol. 84, no. 11, pp. 2005–2012, 2011
20. J.-Y. Kim, C.-H.Chu, and S.-M. Shin, "ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring," IEEE Sensors J., vol. 14, no. 12, pp. 4230–4244, Dec. 2014.

AUTHORS PROFILE

Priyanka V, is a citizen of India. She is a student of master's in computer science at the Department of Computer Science in Sri Shakthi institute of Engineering and technology TamilNadu. Her area of research includes machine learning, Artificial intelligence, Big Data, IOT, and algorithms.

Dr. K. Rajakumari, obtained her both Bachelor's and Master's degree in Information Technology from Anna University of Technology, Chennai. She also received Doctor of Philosophy in Cloud computing during the year 2017.She was a dynamic professor of Information Technology at SNS Institute of Technology After 12+ years of experience, she moved on to Sri Shakthi Institute of engineering and Technology, Coimbatore to take-up additional professional roles and responsibilities. She organized several workshops and seminars for the benefit of both students and professors as well. Moreover, She hosted both National and International conferences to enhance the knowledge base of the students. She holds rich hands-on experience in C & C++, Java, Python Programming and RDBMS, Software Engineering & Testing. She is also member of Institute of Electrical and Electronics Engineers (IEEE), Universal Association of Computer and Electronics Engineer (UACEE), International Association of Engineers (IAENG).