Convex-Hull & DBSCAN Clustering to Predict Future Weather

Ratul Dey
Computer Science & Engineering
Institute of Engineering & Management
ratul170292@gmail.com

Sanjay Chakraborty
Computer Science & Engineering
Institute of Engineering &Management
sanjay.chakraborty@iemcal.com

Abstract — **Machine learning methods are increasingly being used in conjunction with conventional meteorological observations in the synoptic analysis and conventional weather forecast to extract information of relevance for agriculture and food security of the human society in India. Density based clustering approach is incrementally used to predict the future weather conditions in this paper. One famous preprocessing approach, known as Convex-Hull is also used before fed the pollutant data into the clustering algorithm. This Convex-Hull method is strictly used to convert unstructured data into its corresponding structured form. These structured data is efficiently and effectively used by the DBSCAN clustering algorithm to form resultant clusters for weather derivatives. This forecasting database is totally based on the weather of Kolkata city in west Bengal and this forecasting methodology is developed to mitigating the impacts of air pollutions and launch focused modeling computations for prediction and forecasts of weather events. Here accuracy of this approach is also measured.**

Keywords— Clustering; Weather Forecasting; DBSCAN; Convex-Hull; Prediction.

I. INTRODUCTION

Weather Forecasting is an approach which is used to forecast the weather based on the previous or current weather conditions, for a particular region and particular time period, using some science, algorithm and technology. In this paper we focus on weather prediction using air pollutant data by Convex-Hull and DB-Scan clustering approach. Traditionally weather is dependent upon various factors like wind flow, rain fall, air molecules, day and night time ratio, cold wave etc. This paper focuses on weather prediction using Air Molecules and partially wind flow. At the initial stage we divides our actual database (12 months) into 4 parts, which give some opportunity to divide our climate span according to day and night time ratio and wind flow. According to previous data we use Convex-Hull technique to create structural data from them. In the next step those structured data are passed through the famous DBSCAN clustering technique to create a set of clusters of similar data after satisfying some certain conditions [1,2]. These clusters give some specific Results which will give the prediction for future weather. DBSCAN approach has two specific parameters such as min-points, and threshold distance. So we can create clusters according to the threshold value and this is the self-learning phase of our technique. In the very same stage clusters are fully determined by the nature's effect in weather. For prediction stage, the proposed algorithm chooses previous data on that particular day for the last few years and then uses priority based protocol to predict the future weather condition. Forecast day to day weather is difficult to all over the world. Here we choose a region (temperate) Kolkata as a domain for sake of simplicity. Here we consider only the effect of the sea wave comes from 'Bay of Bengal' and also the effect of the traditional cold wave comes from Himalayan mountain in the winter season.

The following sections are organized as related work, proposed methodology, benefits of the model, illustrative example, result & analysis and conclusion.

II. RELATED WORK

This paper focuses on giving different ways to forecast the weather in some certain regions. Here we present some papers which focuses on the similar work. Weather forecasting using data mining is the knowledge related self learning predictive approach. The Weather forecasting concept of clustering will be decided as the batch of data having same result. Increment clustering like Incremental DBSCAN approach is chosen as cluster and new data also implement on that procedure. In DBSCAN approach here midpoint, threshold distance, min number of point to create a perfect cluster. In the case of incremental part here new data chosen as the shortest distance of the cluster head. [3] Convex-Hull approach is defined as the data base to create the structural data, so we include low data set as well as high data set both.[4] In the K-Means approach cluster decides as the minimum distance of those data and create the cluster [5,6,7]. R Packages for the probabilistic weather forecasting using "Bayesian Model Averaging" which implement (GOP) Geo-statistical Output Perturbation Method. [8] The new forecasts data archival system called TIGGE at NCAR, where TIGGE is a part of the incremental THORPEX program [9].

978-1-4799-6908-1/15/$31.00 ©2015 IEEE

III. PROPOSED MODEL

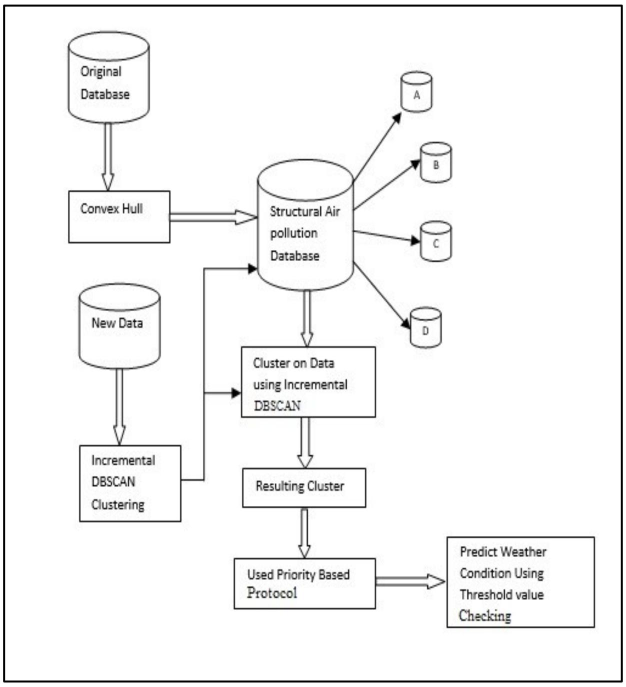

Fig.1. Flow Diagram of Proposed Work

✡ Proposed Algorithm

Step 1:- Collect air pollutant data (CO, NO_2, 0_3, PM_{10}, SO_2) in every hour and store them into the Database.

Step 2:- Use Convex-Hull to get structural data. It includes all extreme data and then store them into 'structural Air pollutant Database'.

Step 3:- Structural air pollutant database splits into 4 parts on the basis of wind flow and day night time ratio.

Step 4:- Apply DBSCAN Clustering to create clusters using structural data.

Step 5:- If any new data is inserted into the database then use incremental DBSCAN clustering [5] to accommodate with that new data.

Step 6:- Finally, find the resulting clusters.

Step 7:- Then the priority based protocol is used on those resulting clusters to give the weather prediction on the basis of last three years.

Step 8:- From the final result, the probable weather conditions and also a temperature range can be predicted.

Initially we collect different air pollutant data like CO, NO_2, O_3, $PM_{10,}$ SO_2 etc. for each and every hour. Then this data are passed through the Convex-Hull technique to get structural air pollutant data from the hourly collected data (Table1.). This structural data are divided into four different databases according to the wind flow and day night ratio. Now this split databases are passed through the DBSCAN clustering technique to create the clusters for every air pollutant data. If some new data are added then use incremental DBSCAN clustering. From the resulting clusters we can determine the weather characteristics of air molecules data. After analysing the last three years data, we can predict weather derivatives by using priority based protocol, (1/3-α), (α), (1/3)]. As per statistical record [11,12], it can be analysed that weather conditions use to similar with 2nd last year data. So we will choose the priority protocol (1/3-α) as the last year, choose (α) as 2nd last year (highest priority) and choose (1/3) as 3rd last year. Prediction result has been calculated on the basis of last three years [13], where α is a constant variable.

IV. BENEFITS OF THE MODEL

This paper used Convex-Hull to get structural format of air molecules data, so we found a structural database from a large dataset in one step. Now we can create cluster using DBSCAN which give a particular radios, min value, Threshold Distance, in case of K-Mean algorithm we found only min value and threshold distance, so the cluster centre chosen by manually which is completely unethical. Here we use incremental DBSCAN, if some new data want to store in the previous database then check fast the nearest outer cluster data that can create a new cluster with certain criteria of min threshold value and min point, if not satisfy DBSCAN condition then choose the nearest cluster header and again check criteria of DBSCAN [5,6]. If satisfy then include on that particular cluster or treated as outer cluster. For weather prediction we use priority based protocol, which give some statistical value of previous year data to predict upcoming weather condition. Now this protocol can predict per day weather forecast and also predict long term nature of weather.

V. ILLUSTRATION WITH EXAMPLE

It has been chosen "February 2014" as a toy example. Dynamic environment prediction using previous data is really tough job. In weather prediction there are different causes which effect on weather factors like rain fall, air pollutants, wind flow etc. In this paper, we discuss only the effect of Air Molecules. For perfect prediction we collect data in hourly basis on a particular Zone. Trigger analysis depends on number of observed points, so if the observation points are less, then this protocol gives erroneous result. In case of DBcluster, It also depends on the number of observed points in learning phase [2]. R has a steep learning curve which use graphical user interfaces (GUI) to encompass point and-click interactions, but they generally do not have the polish of the commercial offerings. So the quality of some packages is less than the perfect. After using the DBcluster technique, it takes all the extreme values so that it gives all the possibilities in the dynamic environment efficiently [4].

Table1: Hourly data

Date	Time	CO	NO_2	O_3	Pm_{10}	SO_2
01/02/14	00:00	1.78	121.4	2.95	203.25	10.21
01/02/14	01:00	1.80	76.65	13.34	204.23	7.39
01/02/14	02:00	1.94	50.18	21.26	169.38	5.48
01/02/14	03:00	1.15	35.80	20.44	157.6	4.71
01/02/14	04:00	0.62	32.58	22.46	147.05	4.38
01/02/14	05:00	1.02	37.03	20.89	155.2	5.61
01/02/14	06:00	0.65	41.85	11.70	140.8	4.66
01/02/14	07:00	0.54	42.25	17.13	157.9	4.33
01/02/14	08:00	0.01	41.15	21.85	157.25	4.15
01/02/14	09:00	2.43	37.10	25.49	148.3	6.01

At the beginning stage, we collect air pollutant data in hourly basis. Database has been collected from West Bengal Pollution Control Board [14], Kolkata which is situated in Victoria Memorial air base station. At the first step collect the last 6 years data hourly basis on that particular zone (table1). Then apply Convex-Hull which takes values from the outer region of the graph:-

01/02/14:-
1.780000,1.800000,1.940000,1.150000,0.620000,1.020000,0.650000,0.540000,0.010000,2.430000,11.390000,13.520000,5.320000,0.950000,0.640000,0.790000,1.030000,0.920000,0.660000,0.780000,1.230000,1.520000,1.720000,2.210000

Upper Hull : (0.0,1.78)(11.0,13.52)(23.0,2.21)
Avg = 5.836666666666666

Lower Hull :
(0.0,1.78)(4.0,0.62)(8.0,0.01)(18.0,0.66)(19.0,0.78)(22.0,1.72)(23.0,2.21)
Avg = 1.1114285714285714

Complete Convex-Hull :
(0.0,1.78)(4.0,0.62)(8.0,0.01)(18.0,0.66)(19.0,0.78)(22.0,1.72)(23.0,2.21)(11.0,13.52)

Avg = 2.6624999999999996, Convex-Hull value can be calculated as,

$$\text{Convex-Hull value} = \frac{\text{Sum of outer region point}}{\text{Total numbers of outer region point}}$$

Now this data are passed through the Convex-Hull approach that gives daily basis data (table 2) from that hourly data.

Table 2: Using Convex-Hull Day wise data collected

Date	Temp	CO	NO_2	O_3	Pm_{10}	SO_2
01/02/14	26	2.66	64.707	32.2	150.38	24.10
02/02/14	26	0.92	80.007	25.79	165.67	4.231
03/02/14	27	1.93	140.57	28.69	219.47	5.44
04/02/14	28	1.74	153.55	47.81	184.91	9.27
05/02/14	30	1.35	79.95	31.82	153.46	9.59

Then the hourly data are passed through the Convex-Hull to find a structural format of air molecules database. It gives all possibilities of data range in the hourly data set. So here we get day wise air pollutant data. The entire year has been divided into four parts according to weather condition and wind flow. Now database has been divided into Dec, Jan, Feb, as database1, Mar, Apr, May as database2, June, July, Aug as database 3 and Sep, Oct, Nov as database 4.

Table 3: Parameter Table

Concentration	CO	NO_2	O_3	Pm_{10}	SO_2
Low	temp low	No Effect	No Effect	No Effect	No Effect
Normal	temp Normal	Dry	No Effect	Dust	Dry
High	temp High	Fog, Dry	Humid High	smog, dust, fog	Smog, dry
Extreme High	temp Extreme High	Fog, Dry	Humid High	smog, dust, fog	Smog, dry

Every air pollutant has specific effect on weather. According to the observation of last 6 years, table3 describes the effect of various air pollutant molecules on weather.

CO:- It is a colourless, odourless, non-toxic greenhouse gas. Increase of Co_2 gives warmer climate in earth surface. Increase of CO_2 reflects the weather as smoggy humid and hot. $Co_2 + 2H^+ + 2e^- \rightarrow Co + H_2O$ can create carbon monoxides, which equally affect on the weather. Due to greenhouse effect the temperature of the environment is increased. [15,16]

NO_2:- It is shown as brown haze dome or plum downwind of cities. It also create dry atmosphere, and create smog. Nitrogen filled tires will change when the temperature changes, just as it does with air filled tires, because nitrogen and oxygen respond to changes in ambient temperature in a similar manner. Air, oxygen and nitrogen will all behave exactly the same in terms of pressure change for each 10 degrees of temperature change. [15,16]

O_3:- Effects on our weather might be small, the ozone concentrated in the stratosphere is crucial to a hospitable climate. Generally Ozone stays in stratosphere. It reacts with other air molecules, in presence of sun light. It is created when oxygen molecules (O_2) are split by ultraviolet radiation into two separate oxygen atoms (O). These oxygen atoms can collide with oxygen molecules to form ozone (O_3). It split like $CO+O_3 \rightarrow CO_2+O_2$ or $NO+O_3 \rightarrow NO_2+O_2$. If number of O3 molecules decries by reaction then *UV* ray hit the earth surface, and increase the temperature. [16]

SO_2:- Sulphur dioxide (SO2) is a colourless gas, belonging to the family of gases called sulphur oxides (SO_2). It reacts on the surface of a variety of airborne solid particles. The most important man-made sources of sulphur dioxide are fossil fuel combustion, smelting, manufacture of sulphuric acid, conversion of wood pulp to paper, incineration of refuse and production of elemental sulphur. It is responsible for the Acid Rain. In presence of oxygen and water it forms Sulphuric Acid (H_2SO_4). SO_2 also form smog (smoke, fog), which create visibility problem. [15,16]

PM_{10}:- Particulate matters are solid, liquid, gas. In large number of PM_{10} can increase temperature. It creates dust, smoke, fog etc. [15]

This protocol is based on weather forecasting by the analysis of the air molecules. So, at learning phase we collect air molecules data for each and every hour from automatic station at Victoria Memorial, Kolkata (22° 34' N, 88° 24' E) under the West Bengal Pollution Control Board and also we collect all weather conditions on that particular day. We chose the effect of five necessary air pollutant data as CO, NO_2, O_3, PM_{10}, SO_2 and also collect the temperature along with other weather parameters. This hourly air molecules data has been calculated through the Convex-Hull technique to get structural data. It gives day wise values of every air pollutant and other weather conditions as well. According to the wind flow and the day night ratio, we divide total database into four parts DB1, DB2, DB3, DB4, respectively. (December, January, February), (March, April, May), (June, July, August), (September, October, November) These daily air pollutant data values are calculated by the DBSCAN technique to create cluster according to the individual database. Now DBcluster found few cluster for each and every air molecules on the individual database. If some new data is inserted to the existing database then it checks the shortest distance of nearest cluster centre, if it satisfy the conditions (minpts and threshold distance) of DBSCAN then it include as the cluster member. If new data does not satisfy the cluster criteria then it will be called outer layer.

From the graphical representation identify the cluster on perticular air pollution database here all the graph's ' X' axix define time and 'Y' axix define air pollutent data value.

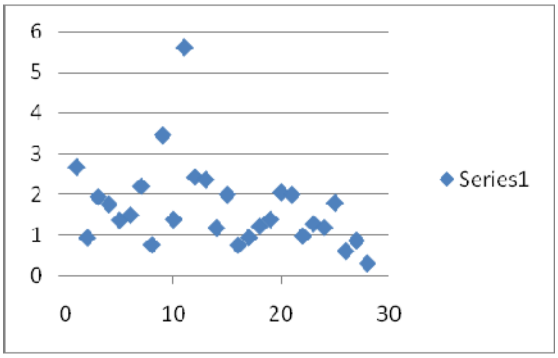

Fig. 2: CO Cluster Data for Month Feb 2014

RESULT FOR 'CO'
Minimum points 5
Threshold Distance .25
Cluster head 1 is 2.0, 0.926
Cluster head 2 is 3.0, 1.931

Table 4: Cluster of CO

Cluster 1	Cluster 2
2.0,0.926	3.0,1.931
8.0,0.752	4.0,1.746
14.0,1.166	15.0,1.987
16.0,0.743	20.0,2.043
17.0,0.933	21.0,1.988
22.0,0.972	25.0,1.782
27.0,0.857	7.0,2.194
26.0,0.597	
5.0,1.359	
10.0,1.376	
18.0,1.208	
19.0,1.378	
23.0,1.268	
24.0,1.178	
6.0,1.48	

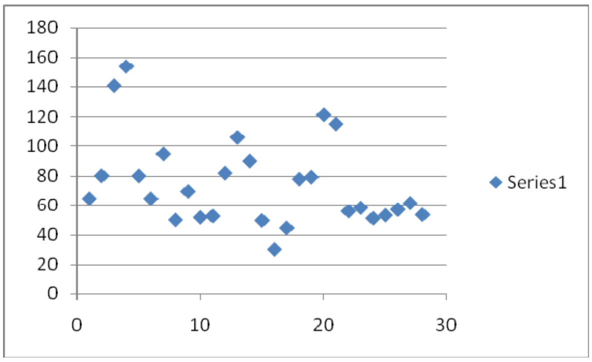

Fig. 3: NO_2 Cluster Data for Month Feb 2014

RESULT FOR 'NO_2'
Minimum points 5
Threshold Distance 5
Cluster head 1 is 2.0, 80.007
Cluster head 2 is 8.0, 50.113
Cluster head 3 is 26.0, 61.581

Table 5: Cluster of NO_2

Cluster 1	Cluster 2	Cluster 3
2.0,80.007	8.0,50.113	1.0,64.708
5.0,79.95	10.0,51.798	6.0,64.695
12.0,81.787	11.0,52.706	23.0,58.392
18.0,77.755	15.0,49.723	25.0,57.216
19.0,79.112	24.0,51.22	26.0,61.581
	27.0,53.652	
	22.0,56.031	
	25.0,57.216	
	23.0,58.392	

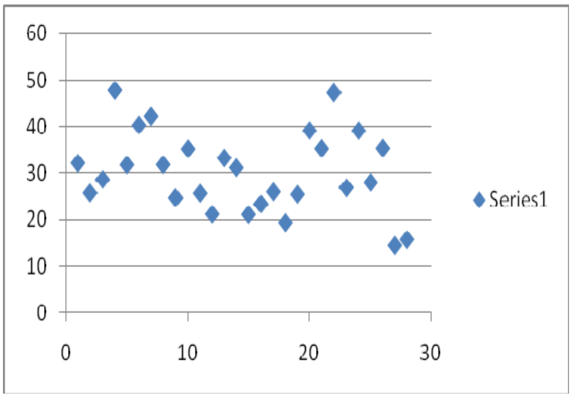

Fig. 4: O_3 Cluster Data for Month Feb 2014

RESULT FOR ‘O_3’
Enter the minimum points 5
Enter the Threshold Distance 3
Cluster head 1 is 1.0, 32.2
sub head 1 is 13.0, 33.298
Cluster head 2 is 2.0, 25.796
sub head 1 is 3.0, 28.698
sub head 2 is 9.0, 24.765
sub head 3 is 11.0, 25.721
sub head 4 is 16.0, 23.427

Table 6: Cluster of O_3

Cluster 1	Cluster 2
1.0,32.2	2.0,25.796
5.0,31.829	3.0,28.698
8.0,31.873	9.0,24.765
13.0,33.298	11.0,25.721
14.0,31.218	16.0,23.427
10.0,35.239	17.0,26.032
21.0,35.331	19.0,25.513
26.0,35.386	23.0,26.921
	25.0,28.048
	14.0,31.218
	12.0,21.237
	15.0,21.168

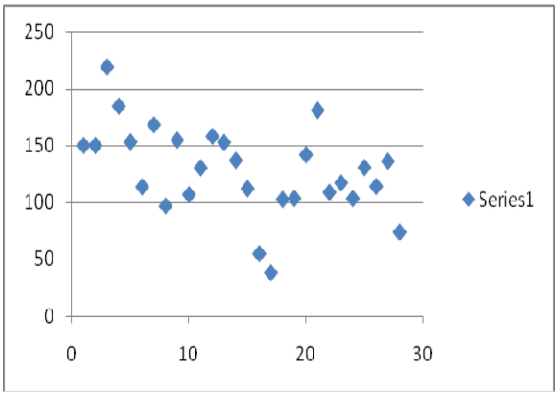

Fig. 5: PM_{10} Cluster Data for Month Feb 2014

RESULT FOR ‘PM_{10}’
Minimum points 5
Threshold Distance 5
Cluster head 1 is 1.0, 150.389
Cluster head 2 is 6.0, 113.869
Cluster head 3 is 10.0, 107.07

Table 7: Cluster of PM_{10}

Cluster 1	Cluster 2	Cluster 3
1.0,150.389	6.0,113.869	10.0,107.07
2.0,150.389	15.0,112.337	18.0,102.877
5.0,153.462	22.0,109.101	19.0,103.748
9.0,155.026	23.0,117.311	22.0,109.101
13.0,153.015	26.0,114.27	24.0,103.654
12.0,158.348		

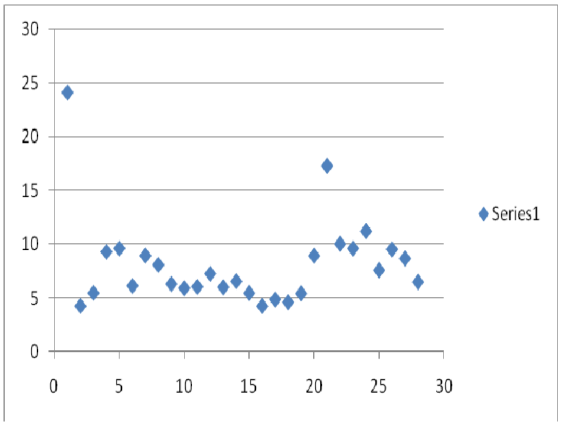

Fig. 6: SO_2 Cluster Data for Month Feb 2014

RESULT FOR ‘SO_2’
Enter the minimum points 5
Enter the Threshold Distance .5
Cluster head 1 is 4.0, 9.274
Cluster head 2 is 6.0, 6.112
sub head 1 is 9.0, 6.285
sub head 2 is 10.0, 5.896

Table 8: Cluster of SO_2

Cluster 1	Cluster 2
4.0,9.274	6.0,6.112
5.0,9.593	9.0,6.285
7.0,8.922	10.0,5.896
20.0,8.914	11.0,6.02
23.0,9.585	13.0,5.982
	14.0,6.537
	28.0,6.475
	3.0,5.443
	15.0,5.433
	19.0,5.397

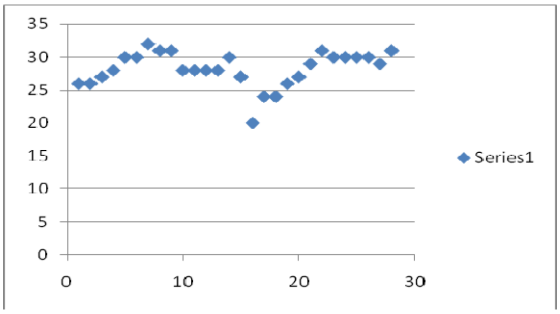

Fig. 7: Temparature Cluster Data for Month Feb 2014

RESULT FOR TEMP
Enter the minimum points 5
Enter the Threshold Distance .5
Cluster head 1 is 4.0, 28.0
Cluster head 2 is 5.0, 30.0
sub head 1 is 6.0, 30.0
sub head 2 is 14.0, 30.0

Table 9: Cluster of Temp

Cluster 1	Cluster 2
4.0,28.0	5.0,30.0
10.0,28.0	6.0,30.0
11.0,28.0	14.0,30.0
12.0,28.0	23.0,30.0
13.0,28.0	24.0,30.0
	25.0,30.0
	26.0,30.0

Table 10: Clustering Table for Air Pollutant data

Date	CO	Cluster	NO_2	Cluster	O_3	Cluster	SO_2	Cluster	PM_{10}	Cluster
01/02/14	2.6625	3	64.707	1	32.2	2	24.10889	Outer Cluster	150.3886	2
02/02/14	0.92625	1	80.007	1	25.7964	2	4.231429	1	150.3886	2
03/02/14	1.93142	2	140.572	Outer Cluster	28.69767	2	5.442857	1	219.479	3
04/02/14	1.74625	2	153.55	Outer Cluster	47.8184	3	9.274286	2	184.915	2
05/02/14	1.35857	1	79.95	1	31.82862	2	9.593333	2	153.462	2
06/02/14	1.48	1	64.695	1	40.35248	3	6.111667	2	113.868	2
07/02/14	2.19444	2	94.5722	2	42.22493	3	8.921818	2	168.51	3

After using DBSCAN we have identified the number of clusters for each air pollutant data, here we have defined the number of clusters with respect to different air molecules on sub-databases. Generally within the same cluster weather conditions are same. Now for prediction purpose of given date we have chosen last few years air molecules data on that particular date, then use priority based protocol for the prediction purpose.

VI. RESULT AND DISCUSSION

Table 11: Resultant Table

Date	Predict Weather condition	Predict Temp. Range	Actual Temp.	HIT
01/02/15	Normal Temp, Smog	24-27	24	Y
02/02/15	Normal Temp, Smog, Fog	24-27	25	Y
03/02/15	Hot, Dry	24-27	26	Y
04/02/15	Hot, Humid	26-30	29	Y
05/02/15	Cloudy, Humid, Fog	26-30	28	Y

The final resultant table is given above where based on nature of clusters and result of priority based protocol a temperature range has been calculated. The comparison between the predicted range and actual temperature is also given in the table 11. The traditional method is to measure accuracy of the result on behalf of given threshold. The equitable threat score (ETS) [7] measure the prediction of the area of the participation of any given threshold with respect to random number.

$$ETS = \frac{H - CH}{\text{Total number of outer region } F+O-H-CH}$$

Biased score is the ratio of the forecast area (F) and observed area (O).

$$Bias = \frac{F}{O}$$

So the expected number of hit in a random forecast (CH) can be expressed as,

$$CH = \frac{F \times O}{NUM}$$

ETS: Equitable threat score
F: Number forecast point
O: Number of observed point
H: Number of hit
CH: Expected number of hit

$$ETS= \frac{272-292}{5+1076-272-292} = -\frac{20}{536} = 0.0373$$

(Choose positive value)

$$Bias= \frac{362}{1076} = 0.33$$

$$CH= \frac{366+1076}{5} = \frac{1460}{5} = 292$$

$$Accuracy = \frac{Number\ of\ matched\ records}{Total\ number\ of\ records} \times 100$$

$$= \frac{272}{366} \times 100$$

$$= 74.5\% \text{ (approx)}$$

Initially we use this approach using Convex Hull and K-means cluster technique. But we faced some problems. Firstly Convex Hull get structural data then K-means gives cluster of the data set. After a certain period we analysis some problem like when choosing cluster centre manually, so it does not give the perfect result [18,19]. Using manual approach, it takes a cluster range so there have some cut points for a particular cluster. If some points slightly high or low than the cut point then it goes to the next cluster, but practically weather behaviour is not like that. Then we switch this protocol by the DBSCAN approach, which have the special concept of minimum point and the threshold distance which are used to identify the cluster properly.

VII. CONCLUSION

In this paper, a new technique is established to predict weather of upcoming days by the help of incremental DBSCAN clustering algorithm and Convex-Hull technique. This technique is suitable for the dynamic databases where the climate data are changed frequently. In this paper the accuracy of this proposed technique is calculated based on their corresponding hit and miss times. In future, other incremental clustering algorithms can be used to predict the weather and can compare them with each other to detect which algorithm among them provide better accuracy. In this paper, we mainly give our focus to predict some certain factors of weather, like temperature, humidity etc. But in future this protocol can also be used to predict rain fall, storm etc. all these important issues of weather.

REFERENCES

[1] R. Dey, S. Chakraborty, L. Dey, "Weather Forecasting Using Convex Hull & K-Mean Technique- An Approach" In Proc. of *Elsevier*, *IEMCON* August 2014, pp.288-295.

[2]V. Kedia, V. Thummala, K. Karlapalem, "Time Series Forecasting through Clustering A Case Study," *COMAD*, 2005. pp.183-191.

[3]Md. K. Salimullah, Md. K. Rahman, Md. M. Islam, "The fundamental formulas for vertices of convex hull," American Journal of Engineering Research (AJER), vol.02, no.07, pp-62-65, 2013.

[4]S. Santhosh Baboo and I.Kadar Shereef, "An Efficient Weather Forecasting System using Artificial Neural Network," *International Journal of Environmental Science and Development,* vol. 1, no. 4, October 2010.

[5]S. Chakraborty and N. K. Nagwani, "Analysis and Study of Incremental K-Means Clustering Algorithm," in proc. of *International conference on high performance architecture and grid computing (HPAGC)* sponsored by Springer Germany, Punjab (India), 2011.

[6]S. Chakraborty and N. K. Nagwani, "Performance evaluation of incremental K-means clustering algorithm," *IFRSA International Journal of Data Warehousing & Mining (IIJDWM),* vol.1,pp-54-59, 2011.

[7]S. Chakraborty, N. K. Nagwani, L. Dey, "Weather Forecasting using Incremental K-Means clustering," *CiiT International Journal of Data Mining & Knowledge Engineering*, May 2012.

[8]C. Fraley, A. Raftery, T. Gneiting, M. Sloughter and V. Berrocal , "Probabilistic Weayher Forecasting in R," *The R Journal* vol.3, no.1, june 2011.

[9]T. N. Krishnamurti, A. D. Sagadevan, A. Chakraborty, A. K. Mishra, and A. Simon, "Improving Multimodel Weather Forecast of Monsoon Rain Over China Using FSU Super ensemble", *Advances In Atmospheric Sciences,* vol. 26, no. 5, 2009, pp. 813–839.

[10]S. Young, I. Arel, Thomas P. Karnowski, Derek Rose "A Fast and Stable Incremental Clustering Algorithm," † *International University of Tennesee* 2010.

[11]D. Schiopu, E. Georgiana Petre, C. Negoita "Weather Forecasting using SPSS Statistical Methods," vol.LXI, no. 1/2009.

[12]Sean D. Campbeel and Francis X. Diebold "Weather Forecasting for Weather Derivatives," *American Statistical Association Journal of the American Statistical Association*, vol. 100, no. 469, 2005.

[13]L. Zhou, Z. Wang, Y. Wang, H. Yang and S. Huang "R/S Analysis on Anhui Climate Change in Recent 56 Years," in proc. of *Information Technology and Applications (ITA), 2013 International Conference,* November 2013, pp. 483 –485.

[14]West Bengal Pollution Control Board. www.wbpcb.gov.in.

[15] Air pollution and Climate change", *Science for Environment Policy, European Commission*, issue, November-2010.

[16] H. Eerens, M. Amann, Jelle G. van Minnen, "European Topic Centre on Air And Climate Change (Etc/Acc)", *European Topic Centre on Air and Climate Change (ETC/ACC),* September 2002.

[17]K. Mumtaz, and K. Duraiswamy, "A Novel Density based improved k-means Clustering Algorithm–Dbkmeans," *IJCSE*, vol. 02, no. 02, pp.213-218, 2010.

[18]S. Chakraborty, N.K.Nagwani, "Analysis and Study of Incremental DBSCAN Clustering Algorithm", International Journal of Enterprise Computing and Business Systems, Vol. 1, Issue 2 July 2011.

[19] Sanjay Chakraborty, N.K.Nagwani and Lopamudra Dey, "Performance Comparison of Incremental K-means and Incremental DBSCAN Algorithms" *International Journal of Computer Applications (0975 – 8887),* Vol. 27, No.11, August 2011.

[20] Edward E. Lorenz, "An Experiment in Nonlinear Statistical Weather Forecasting" National Center for Atmospheric Research, 2 Boulder, Colo. 80307, pp.590-602, february 1977.

[21]V. J. BERROCAL, A. E. RAFTERY, T. Gneiting and R. C. Steed, "Probabilistic Weather Forecasting for Winter Road Maintenance," *American Statistical Association Journal of the American Statistical Association* June 2010, vol. 105, no. 490.

[22] MathType software application, Design Science, MathType 6.9 (2013).