Dynamic Modeling Technique for Weather Prediction

Jyotismita Goswami
Department of Computer Science & IT and Engineering
Assam Don Bosco University
Azara,Guwahati-15,Kamrup(Assam),India
jyotizgoswami09@gmail.com

Alok Choudhury
Department of Computer Science & IT and Engineering
Assam Don Bosco University
Azara,Guwahati-15,Kamrup(Assam),India
alokchoudhury002@gmail.com

Abstract: --

For an agro-based country like India, the impact of weather is to a large extent. The prevailing weather conditions turns out to be major affecting factor for the proper development of the country. Therefore it is of utmost importance to know accurately the fore -coming weather conditions so that we can take better decisions to tackle problems. A variety of predictions models have already been developed, backed up with different methods like soft computing, case based reasoning, feature extraction methods etc. A number of researches have been done or being done using the above mentioned approaches for forecasting. This paper discusses the different prediction models available till now along with the comparison of their methodologies applied, their key findings making its study worthwhile for a better start for the generation of a new weather prediction model along with the description of our new methodology for weather prediction using dynamic climate models.

Keywords-- Data Mining, Prediction, Data Mining Techniques, REGCM Model.

I. INTRODUCTION

The term weather refer to the existing scenario of air on earth prevailing at a certain place and time for a short period where as Climate defines the state of existence of the same for a longer period. Researchers are busy applying the latest technologies in order to increase the efficiency of prediction. Today, weather forecasts are made by referring the past records collect by remote sensing satellites for several years and make the prediction by discovering the pattern of atmospheric change. The ever changing nature of the atmosphere demands a high and sophisticated computation to arrive at accurate results .It turns out to be an essential activity nowadays as most sectors including agriculture sectors, industries, aviation etc. are getting highly dependent upon it.

Weather forecasting is all about predicting the pattern of the changing atmosphere.Weather data are obtained by ground observations, observations from ships and aircraft, Doppler radar, and satellites. This information is then sent to meteorological centers where the data are collected, analyzed, and made into a variety of charts, maps, and graphs. Modern high-speed computers performs data analysis also predicts the upcoming change in the next hours.

Traditional weather forecasting approaches can be divided into 2 main divisions:

(a) The empirical approach.

(b) The dynamical approach.

The first approach makes the prediction by looking for similar cases in the database matching with the input , often referred to by meteorologists as analogue forecasting. This approach is useful for predicting local-scale weather when recorded cases are plentiful.

The second employs mathematical equations for forecasting. Meteorologists designs atmospheric models by approximating the atmosphere parameters by mathematical simulations to describe the atmospheric change. The equations are then programmed into a computer and weather data are fed as input . The computer does all the necessary computations ,evaluates the pattern change and lastly outputs the result in the form of graphs or rules etc. This approach proves useful for modeling large-scale weather phenomena .

II. LITERATURE REVIEW:

A. Basic definitions:

In order to develop a prediction model we need to have an idea of the basic technical terms related with it.some of the definitions are given below:

A model is basically a program ,or mathematically an equation written to show the variation of output against the corresponding input.It abstracts all the underlying techniques applied, displaying the change of pattern with respective output. Now the question is why we need a prediction model. As we know that climate affects our daya to day activities to a large extent, so in order to have an idea of the conditions ahead of time facilitates us to make the required decisions .The factors playing a major role to be taken care of before developing a new forecast model are mentioned below:

i. **The dataset.:** The selection a dataset is a prime factor .It can be attained from different sources :For Eg: historical data ,satellite data colleted by high resolution sensors mounted on them ,or it can be station data collected on a daily basis by instruments like rain gauge,data logger etc.

ii. **The Variables**: The number of variables used in the input training set affects the performance of the model. dimensionality reduction can be used for removing unwanted attributes.

iii. **Data Analysis**: Since the atmospheric parameters are inter-related so care has to be maintained while preparation of training and testing sets. As the output is evaluated against the training applied to the model, so if training goes wrong, it can pose serious affects in prediction results.

iv. **Period**: Choosing a forecast period is of utmost importance ,describing whether the prediction will be short term, medium term or long term forecast.

Coming to the implementation phase ,it can be carried out in different platforms such as linux ,Macintosh, Windows etc , although linux is given due consideration for its command driven feature.

B. Data mining softwares

There are a variety of data mining softwares available to deal with the high dimensional weather data namely WEKA, TANAGRA, ORANGE, RAPIDMINER, DBMiner etc.

C. Algorithms Used for Weather Prediction

i K- Nearest Neighbour algorithm :K-NN algorithm[19] one of the most popular and simplest clustering tool used in scientific and industrial applications. The K-NN works: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbours. If **k = 1**, then the object is simply assigned to the class of its nearest neighbour. It is sensitive to the local data structure and can efficiently compute the decision boundary.

ii Decision Tree: A Decision Tree is a predictive model ,flow-chart-like tree structure usually used for classification. Each branch represents an outcome of the test. Leaf nodes represent class distribution. The decision tree structure provides an explicit set of ―if-then rules making the results easy comprehension [20].The partitioning of input space is based on a sequence of tests where ach interior node tests the value of some input variable, and the branches from the node are labelled with the possible results of the test. The leaf nodes represent the class to which it falls. The classification of a specific input instance is thus performed by starting at the root node and, depending on the results of the tests, following the appropriate branches until a leaf node is reached .

iii Naïve Bayes**:** Bayesian classification [21] is a kind of the statistical classification based on Bayes probability theory. Naïve Bayes is a tree, containing a root node, with leaf nodes each representing an attribute variable. It descripts the properties of the object to be classified. It states that the presence or absence of a particular feature is unrelated to the presence of absence of any other feature, given the class variable. Sometimes parameter estimation uses the method of maximum likelihood also requires a small amount of training data to estimate the parameters required for classification.

iv Artificial Neural Network**:** ANN is an information processing paradigm inspired by the way biological nervous system composed of a huge number of highly interconnected processing elements (neurons) working together to solve the problem.ANN analyze the data through a learning process making them suitable for weather forecasting. They provide a methodology for solving non-linear problems with the advantage of extracting the relationship between inputs and outputs of a process. The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations.

v Fuzzy logic :A fuzzy logic model[22] is also known as a fuzzy inference system or fuzzy controller is composed of two functional components. One is the knowledge base, which contains a number of fuzzy if–then rules and a database to define the membership functions of the fuzzy sets used in the fuzzy rules. Based on this knowledge base, the second component is the fuzzy reasoning or decision-making unit to perform the inference operations on the rules. Two operations are performed for fuzzy logic modeling.

When data are ready, a fuzzification operation is processed to compare the input variables with the membership functions on the premise part to obtain the membership values of each linguistic fuzzy set. These membership values from the premise part are combined through a min operator to get firing strength (weight) of each rule in order to generate a qualified consequent (either fuzzy or crisp) of each rule depending on this firing strength. Then the second operation is the defuzzification to aggregate the qualified consequents to produce a crisp output . FL is very useful in modelling complex and imprecise systems, and fuzzy set theory is a powerful tool and its applications have rapidly increased with establishing its utility in numerous areas of the scientific world.

D. Related work

A good number of work has been done in this area.The following papers gives an overview of the techniques or models developed till date ,helping us to gain the concepts required before stepping into a good research.

Ibrahim Adeyanju, [1] developed a prediction model rather a system named CBR-METEO, applying the case based reasoning technique.It makes forecasts ,based on the match between previous records in the database and input. The system is built using jCOLIBRI an existing CBR framework providing generic modules for each component making the process easier. The architecture has 4 components: Retrieve, Reuse, Revise and Retain.

In 2010 ,M.Kannan et al.[2] developed another model for short term rainfall forecasting using regression method The training has been carried out for the three months (winter season) rainfall data of Tamil Nadu , for a five years dataset. Multiple linear regressions and Karl Pearson correlation have been used to predict the rainfall using the previous year's data from the specific time period as input.

In 2010 Ch.Jyosthna Devi et al. in [3],applied soft computing approach in prediction of temperature.A neural network-based algorithm was developed using Back Propagation Neural Network (BPN) technique which can best approximate a large class of functions. A 3-layered neural network is designed and trained with the existing data In 2007 , [5] Paras et al. developed a soft computing model by applying Feed forward ANN along with back propagation to predict the future weather conditions. During the training process, the network adaptively changes its synaptic weights to reduce the total system error within specified tolerance.The statistical indicators used as input features for the model include, Moving Average (MA**)**, Exponential Moving Average (EMA), Oscillator (OSC), Rate of Change (ROC), Moments($\mu 3$). In 2010,Karsten Steinhaeuser, Nitesh V. Chawla, and Auroop R. Ganguly, [4] carried out research based on extraction of ocean climate indices from historical data using traditional clustering methods in addition to network based clusters.

Regression works by estimating the output value based on the input parameters. It is widely used to solve classification problems and forecasting. S. Kotsiantis et al.in 2007 [6] tried to explore the performance of different data mining techniques in estimating minimum, maximum and average temperature . In addition to this another research was carried out in 2010 [9] , a hybrid technique combining the features of both regression fusion and dynamic selection for daily temperature prediction.

A new enhanced method of self organizing data mining was proposed by Godfrey C. Onwubolu et al. in 2007 [7] called enhanced Group Method of Data Handling (e-GMDH) for weather forecasting. It combines the best of both statistics and Neural Networks features along with the additional principle of induction. Meghali A. Kalyankar in 2013, [12] applied k-means to predict the climate of a region using historical weather data.Following the same purpose , Folorunsho Olaiya and Adesesan Barnabas Adeyemo in 2012, [13] tried to predict a number of weather parameters using data mining algorithms ,ANN and Decision Tree.

III. PROPOSED WORK FOR PREDICTION

In modern times the prediction scenario has changed a bit. We can use artificial intelligence comprising of neural networks, rough sets in order to obtain a more realistic value close to the actual values.However we here we are going to propose a new idea of predicting the weather parameters, introducing the dynamical prediction modeling.For this purpose we will be dealing with RegCM model.

Before going through the implementation methodology let us have a brief idea about the basics:

A. Climate

The term Climatology basically refers to the study of climate and also scientifically defined as weather conditions averaged over a period of time. This modern field of study is regarded as a branch of the atmospheric sciences and a subfield of physical geography, which is one of the Earth sciences.

B. Satellite data:

Satellite data refers to the information about our Universe measured by the instruments aboard by satellites and transferred to the ground. Thus we can obtain the data regarding the climate from these satellites after which much processing is required to arrive at proper results. The 3 types of satellite available are as follows:

i Infrared Data: Infra-red satellite data consists data regarding the temperature of the cloud tops (or ground) and does not rely on visible light for illumination, so are available 24 hours a day. Colder temperatures are shown in brighter shades of white and warmer temperatures in darker shades of grey.

ii Visible Data : Visible satellite data consists data about what the sky (and ground) "look" like at any one time making this data useful during daylight hours. Thicker cloud is shown as brighter white in these images, regardless of its height and thin cloud as paler grey shades.

iii Water Vapour Data: Water vapour data show moisture in the atmosphere that may or may not be in the form of clouds.

NetCDF is a set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.Since we are dealing with high dimensional satellite data so Netcdf since it is platform as well as machine independent .The conversion procedure from ASCII to text files is given by Fig 1.

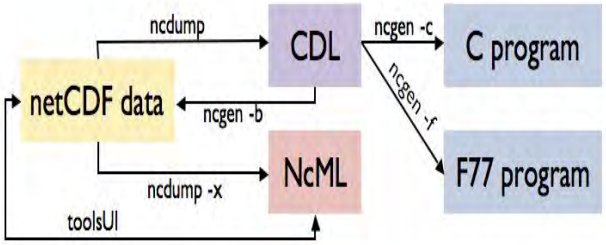

Fig 1: Netcdf Utilities

Weather forecasting deals with the methodologies providing timely and expected weather forecasts which is highly crucial for agriculture based countries. Its origin took place in about 19th century, when the great American meteorologist Cleveland Abbe concluded from his experiments that meteorology is essentially the application of hydrodynamics and thermodynamics to the atmosphere. Climate models, both global and regional, are the primary tools that aid in our understanding of the many processes that govern the climate system. These models use differential equations ,conservation law,formulated based on the factors governing the physical behavior of the atmosphere ,dividing the Earth into a 3D grid coordinate system. The interaction among these variables (wind components , surface pressure, temperature, mixture of cloud water , ice ,snow etc) with the adjacent grid cells help in calculation of future atmospheric conditions.

C. REGCM Model

RegCM is an open source Regional Climate Model(Limited Area Model),o riginally developed by Giorgi et al. and then modified, improved and discussed by Giorgi and Mearns . It uses the ***Downscaling*** method for getting a clear high-resolution weather information (for Eg giving a better representation of the underlying topography at a scale of 50 km or even less) compared to that relatively coarse-resolution information by global climate models (GCMs).

The model architecture is given by Fig(2).

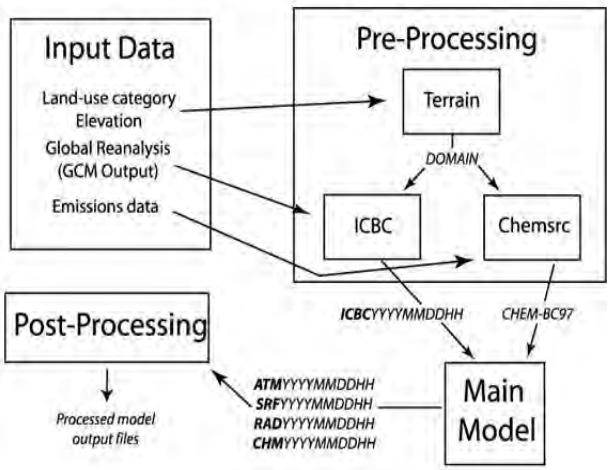

Fig2: REGCM Model Architecture

Model Components: The Terrain file is used for creating the domain file consisting localized topography , land use databases, projection information and land sea mask. The SST for the model is created using sst program containing the sea surface temperature to be used in generating the icbc for the model and lastly the ICBC files created using icbc program, contains surface pressure, temperature, horizontal wind components and time resolution for the input file.

After successfully running the model it generates 4 files in the output directory.

i. ATM –Contains atmosphere status of the model.
ii. SRF-Contains surface diagnostic variables.
iii. RAD –Contains radiation information.

The weather paprameters are represented in grid structure(Fig 3) by the RegCM model.The co-oredinate system helps in easy handling of data as well as we can get hold of specific region data differing at every 1^0 latitude and longitude.

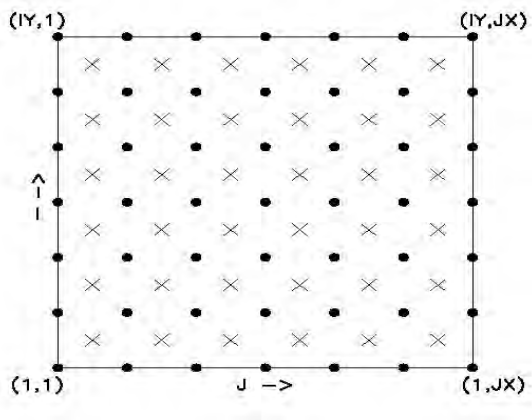

Fig 3:Grid Representation

D. Software Requirements.

1. Unix or Linux OS.
2. FORTRAN 90/95 compiler, python language interpreter.
3. Make utility (GNUmake).
4. NetCDF library.
5. MPI (for parallel shared memory).
6. Graphics (GrADS, FERRET, NCL,) for visualization.

The data required for the proper working of the RegCM model is given by the Fig(4)

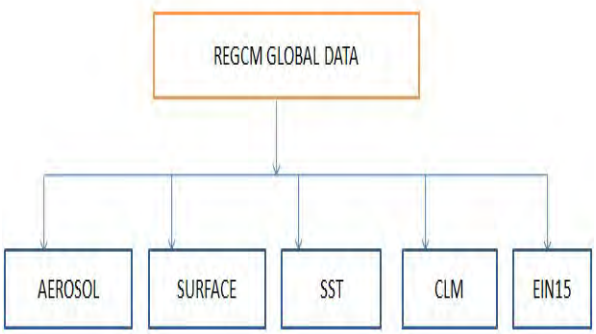

Fig 4: Data Required for Running the Model

E. Working Methodology

We are going to have a short term prediction for a period of (2-3) days.So we start with the model installation followed by its configuration in our sytem having all the necessary requirements and we run the experiments using dual processor,Open MP ,Grads ,Netcdf software packages. Proceeding towards implementation first we need to localize the model on a world region. This is done with the terrain program which reads the database of the entire globe. Next we will be dealing with the task of creating the sea surface temperature for a global dataset followed by the creation of ICBC (Initial Conditions and Boundary Conditions) for the model

itself.When all the above work is done we can proceed to our task of launching our first test of model simulation. Now for the prediction purpose we will be having first a simulation of weather parameters. We will be providing the historic data as as the boundary conditions and today weather parameters values as initial conditions.The model will build all the necessary files for the particular dataset.In the present work the multi-source (satellite and reanalysis weather and climate data) in Netcdf format are used to analyse the data.Since our region of study is India so we need to locate India on the grid (Fig 5)by giving the proper latitude and longitude values.Moreover if we need to select a particular small area domain so we can do this with an user defined algorithm discussed below.

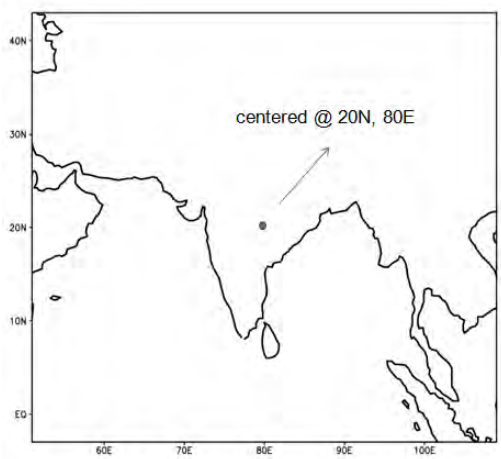

Fig 5: Map Domain

Algorithm for generating new file (user defined format).

i. Select the domain range from user for a particular variable of selected file.
ii. The values of required parameters are extracted and these are written to another file of same or different format.
iii. The header information and the new modified file are formatted to match the input file format.
iv. The formatted file is encrypted using the corresponding utility to get the output file in the same or different format.

The prediction results obtained from the model are validated with the historic data collected over a region for a certain period.We evaluated the results taking only a single parameter , rainfall with the yearly dataset for the year(2005) collected from CSIR-MMACS ,Bangalore. However it can be extended for high number of weather parameters. Fig 6 gives a snapshot of simulation result taking rainfall as the weather parameter in India.

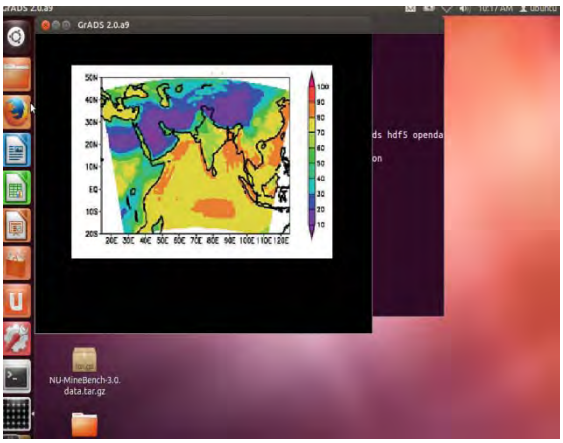

Fig 6: Simulation of Rainfall Results

We also carried out validation of our model output against real world historic data given by (Fig 7) obtained from the Meteorological Dept (IMD) of India for rainfall observation for the monsoon months .

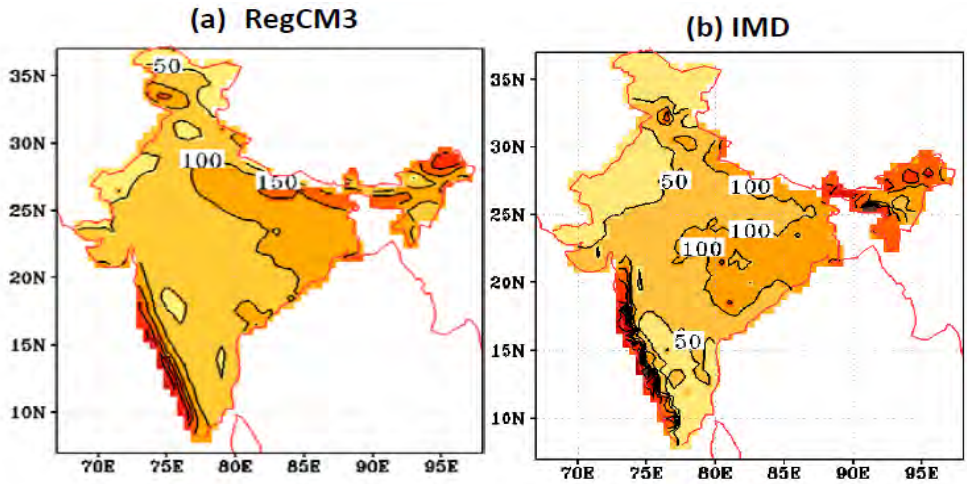

Fig 7: Model validation with the real India Meteorological Dept (IMD) rainfall observation Monsoon (Average of June+July+Aug+Sept) Rainfall

We found the optimized technique which will yield the best results taking the minimum time. We carried out extensive analysis on different projections based on India, giving the latitude and longitude as required. Mainly 3 types of projections namely LAMCON, POLSTR, NORMER. These projections are carried out to evaluate the model performance in different nodes in order to find out the optimized need for simulation followed by sensitivity studies using these projections to find the better one in simulating the rainfall at regional scale . These projections are nothing but the phases which yield the maximum information about prediction.These projections are like For Eg: When we take photo from different angles and end up selecting the one giving more information about the objectWe found out that LAMCON projection is the best and thus it can be used for minimizing the prediction job.Thus we ended up getting better results compared to other methods using this dynamic method of weather prediction.In addition to our work of prediction we have also developed our method of clustering the different regions according to the intensity of rainfall ie.which are the regions having high rainfall and which having less.The algorithm proposed by us for this technique is discussed below

Algorithm

i. Model Inputs: Input file (of different formats).
ii. Convert input format to ASCII.
iii. Enter the required Domain of Interest (like for Karnataka, 12N-16N, 70-74E).
iv. Enter the period of analysis (Like Daily, Weekly, Monthly or Annual).
v. Analysis with different statistics Like Mean, SD, Anomaly, Correlation etc.
vi. System outputs : output file in the ASCII format and or (required).
vii. Visualization: Image and Graphical representation.

Finally the overall overview or rather the structure of our proposal is given by (Fig 8)

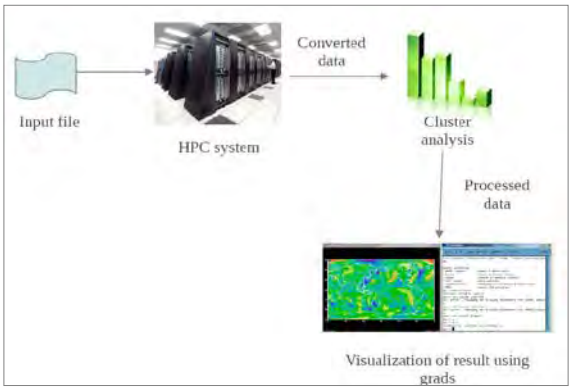

Fig 8: Extraction and Visualization of Output Data

F. Applications of Regional Climate Modeling

i Model development and validation with a high resolution for smaller areas.
ii Used in process studies such as topographic effects, regional hydrolic budgets.
iii Climate change studies
iv Paleoclimate studies(Climate effects of aerosols).
v Seasonal prediction
vi Impact studies

IV. CONCLUSION

After going through all the above works done so far in this field we can have the idea that fields like computer networks, various techniques like regression, classification algorithms ,clustering etc have made our prediction job quite a easy task. The results vary according to the techniques applied in the model, some outperforming others. The interrelationship among the weather attributes makes prediction a difficult task. But with the development of mathematical models its no longer a complex procedure. Our results give a higher accuracy than compared to others.Thus we have the idea that dynamic models play a very important role in prediction of different weather parameters.

REFERENCES

[1] Ibrahim Adeyanju, ”Generating Weather Forecast Texts with Case Based Reasoning”, International journal of computer applications (0975 – 8887) volume 45– no.10, may 2012.
[2] M.Kannan, S.Prabhakaran, P.Ramachandran“Rainfall Forecasting Using Data Mining Technique”, International journal of engineering and technology, vol.2 (6), 397-401,nov2010.
[3] Ch.Jyosthna Devi , B.Syam Prasad Reddy, K.Vagdhan,Kumar,B, Musala Reddy, N.Raja, “ANN Approach for Weather Prediction Using Back Propagation”, ISSN:2231-5381,Vol-23,2010
[4] Karsten Steinhaeuser, Nitesh V. Chawla, and Auroop R. Ganguly, “Comparing Predictive Power in Climate Data: Clustering Matters”, 2010.
[5] Paras, Sanjay Mathur, Avinash Kumar, and Mahesh Chandra “A Feature Based Neural Network Model for Weather Forecasting”, World Academy of Science Engineering and Technology 10, 2007
[6] S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, K. Menagia, “Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values”,International Journal of Mathematical, Physical and Engineering Sciences, July 22, 2007.
[7] Godfrey C. Onwubolu, Petr Buryan, Sitaram Garimella, Visagaperuman Ramachandran,Viti Buadromo and Ajith Abraham, “Self-Organizing Data Mining for Weather Forecasting” ,IADIS European Conference Data Ming, 2007.
[8] Badhiye S. S, Wakode B. V, Chatur P. N. Aanalysis of Temperature and Humidity Data for Future Value Prediction”, (ijcsit) International Journal of Computer Science and Information Technologies, Vol. 3 (1), 2012, 3012-3014.
[9] S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, K. Menagias, “A Hybrid Data Mining Technique for Estimating Mean Daily Temperature Values”, 2010.
[10] Zahoor Jan, M. Abrar, Shariq Bashir, and Anwar M. Mirza, “Seasonal to Inter-Annual Climate Prediction Using Data Mining knn Technique”,2009.
[11] Karsten Steinhaeuser, Nitesh V. Chawla ,Auroop ,R.Ganguly “An Exploration of Climate Data Using Complex Networks”,2009.
[12] Meghali A. Kalyankar, Prof. S. J. Alaspurkar, “ Data Mining Technique to Analyse the Metrological Data”, volume 3, issue 2, February 2013 ISSN: 2277 128X.
[13] Folorunsho Olaiya, Adesesan Barnabas Adeyemo,“Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, University of Ibadan, Ibadan, Nigeria, Information Engineering and Electronic Business, 2012, 1, 51-59.
[14] Sarah N. Kohail, Alaa M. El-Halees, “Implementation of Data Mining Techniques for Meteorological Data Analysis”, The Islamic University of Gaza, International Journal of Information and Communication Technology Research,Volume 1 No. 3, July 2011 ISSN-2223-4985 .
[15] Y.Radhika & M.Shashi, “Atmospheric Temperature prediction using Support Vector Machine”, International Journal of Computer Theory & Engineering, Vol.1.No.1.April 2009 1793-8209.
[16] Mohsen Hayati & Zahra Mohebi, Temperature Forecasting Based on Neural Network Approch World Applied Science Journal 2(6) 613-620, 2007,ISSN: 818-4952 ©IDOSI Publications 2007 .
[17] Montgomery, D. C., and Lynwood, A. J., 1996: Forecasting and Time Series Analysis, McGraw-Hill.
[18] T. Hall, Precipitation forecasting using a neural network, Weather and Forecasting, Boston, 14, Iss. 3; 338–346, Jun 1999.
[19] Berkhin P, 2002, CA, Tech. Rep, "Survey of clustering data mining techniques, Accrue Software, San Jose".
[20] Folorunsho Olaiya and Adesesan Barnabas Adeyemo, February 2012, “Application of Data Mining Techniques in Weather Prediction and Climate Change Studies”.
[21] McCallum A, Nigam K, 1998, AAAI-98 workshop on learning for text categorization, “A comparison of event models for naive bayes text classification”, 752: 41-48.
[22] Somia A. Asklany, Khaled Elhelow, I.K. Youssef, M. Abd El-wahab,”Rainfall events prediction using rulebased fuzzy inference system”, February 2011.