Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model

Ganesh P. Gaikwad, Prof. V. B. Nikam
Department of Computer Engineering and Information Technology, VJTI, Mumbai

ABSTRACT

Indian Meteorological Department (IMD) has progressively expanded its infrastructure for meteorological observations, communications, forecasting and weather services and it has concurrently contributed to scientific growth. Rainfall Prediction is the application of science and technology to predict the state of the atmosphere for a given location. Meteorological data mining is a form of data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Weather is one of the meteorological data that is rich by important knowledge. In this paper we study the different rainfall prediction models like Weather research and forecasting, Seasonal climate forecasting, Global data forecasting and General data mining rainfall prediction model.

Keywords: **Data Mining, Forecasting, GDFS, HPCS, WRF, SFS.**

1. Introduction

The Weather Research and Forecasting (WRF) model is a numerical weather prediction (NWP) and atmospheric simulation system designed for both research and operational applications. While the Global Forecast System (GFS) is a global numerical weather prediction system containing a global computer model and variational analysis run by NOAA.

Data mining is the process of extracting or mining knowledge from large amount of data. In other words Data mining is the efficient discovery of valuable, non-obvious information from a large collection of data. It extracts hidden predictive information from large databases, is a powerful new technology with great potential to help in analysis of data and for decision making. Data mining functionalities are used to specify the kind of patterns to be found in general data mining tasks. In general data mining tasks can be classified into two categories: descriptive and predictive. Descriptive mining characterize the general properties of the data in the database. Predictive mining tasks perform inference on the current data in order to make predictions.

The increasing availability of climate data during the last decades (observational records, radar and satellite maps, proxy data, etc.) makes it important to find effective and accurate tools to analyze and extract hidden knowledge from this huge data.

Meteorological data mining is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Useful knowledge can play important role in understanding the climate variability and climate prediction. This understanding can be used to support many important sectors that are affected by climate like agriculture, water resources and tourism. To make an accurate prediction is one of the major challenges facing meteorologist all over the world.

2. Forecasting

Description or calculation of what will probably happen in future

2.1 Types of Forecasting

The weather forecasts are divided into the following categories

Now casting: Now Casting in which the details about the current weather and forecasts up to a few hours ahead are given

Short range forecasts(1 to 3 days): Short range forecasts in which the weather (mainly rainfall) in each successive 24 hrs. Intervals may be predicted up to 3 days.

Medium range forecasts (4 to 10 days): Medium range forecasts Average weather conditions and the weather on each day may be prescribed with progressively lesser details and accuracy than that for short range forecasts.

Long range /Extended Range forecasts (more than 10 days to a season): There is no rigid definition for Long Range Forecasting, which may range from a monthly to a seasonal forecast.

3. Rainfall Prediction Models: A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. There are two approaches to predict rainfall. They are Empirical method and dynamical methods.

3.1 General Forecasting Model

Making a weather forecast involves five steps: observation, collection and transformation of data, plotting of weather data, analysis of data and extrapolation to find the future state of the atmosphere, and prediction of particular variables.

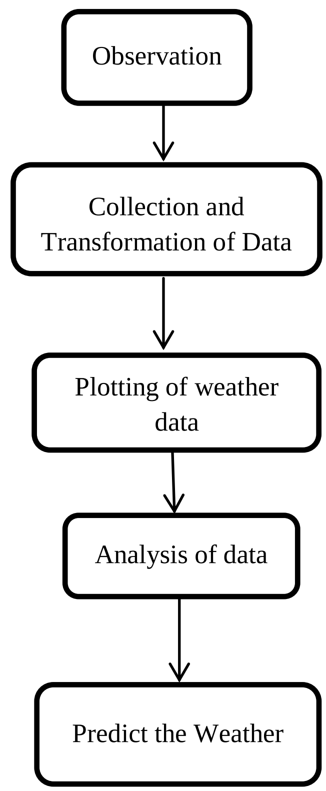

Fig1. General Forecasting Model

3.2 Dynamical Model

In dynamical approach, predictions are generated by physical models based on systems of equations that predict the

evolution of the global climate system in response to initial atmospheric conditions.
The Dynamical approaches are implemented using numerical rainfall forecasting method.

3.2.1 Weather Research and Forecasting Model

The Weather Research and Forecasting (WRF) model is a numerical weather prediction (NWP) and atmospheric simulation system designed for both research and operational applications. The development of WRF has been a multi-agency effort to build a next-generation forecast model and data assimilation system to advance the understanding and prediction of weather and accelerate the transfer of research advances into operations. The geogrid defines model domains and interpolates static geographical data to the grids. ungrib extracts meteorological fields from GRIB-formatted files. The metgrid horizontally interpolates the meteorological fields extracted by ungrib to the model grids defined by geogrid.

Each of the WPS programs reads parameters from a common namelist file, as shown in the figure.This namelist file has separate namelist records for each of the programs and a shared namelist record, which defines parameters that are used by more than oneWPS program.

The ungrib program reads GRIB files, degribs the data, and writes the data in a simple format, called the intermediate format.GRIB (GRIdded Binary or General Regularly-distributed Information in Binary form) is a mathematically concise data format commonly used in meteorology to store historical and forecast weather data.

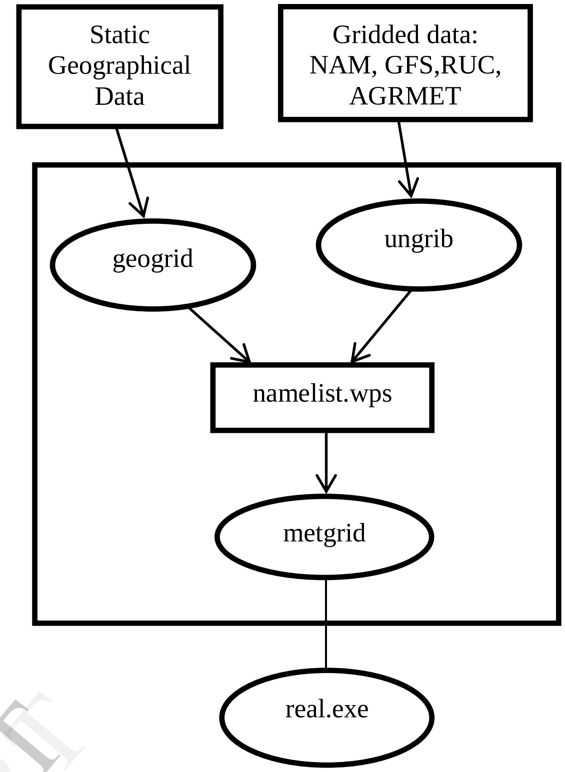

Fig2. WRF Preprocessing System

3.2.1.1 WRF Software Architecture

The first step consists of discomposing the execution ofthe model in independent tasks.
Each task is implementedin an independent Python script

prepreprocess.py: This script processes tasks beforethe model execution.
geogrid.py: Responsible for executing the GEO-GRID module of the WRF model.
ungrib.py: Responsible for executing the UNGRIBmodule.
metgrid.py: Responsible for executing the MET-GRID module.
real.py: Responsible for executing the REAL module.
wrf.py: Responsible for executing the WRF module.
The output that the WRF model produces is in netCDF format, Unidata. The graphic representations are generated using output in order to visualize results. These graphics can

ISSN: 2278-0181
Vol. 2 Issue 7, July - 2013

be generated by using an additional script that can be included in the tasks workflow.

Wfmanager.py script: This script is responsible for coordinating the entiresequential sending process of tasks. In orderto monitorthe beginning and end of each task, wfmanager.py uses the log file that the SGE job scheduler generates with the resultof the execution of each job. The first step consists of defining a workflow that includes all of the tasks. In order to define this workflow,a file in XML format is used. This XML filefollows a few rules and contains a series of entities:

Work-flow entity: The work-flow entity should contain only one series oftask entities that define each task. Workflow supportstwo attributes. The date attribute contains the date andForecast start-time, and the forecast attribute that indicates the number of forecast hours from the start time.

Task entity: The task entity contains the definition of the task. Asequence of elements in this entity defines the work-flowentity. Each task entity should contain an element for eachone of the following entities:

ID entity: Assigns a name for a task.

Script entity: This indicates the script path that thetask executes.

Paramlist entity: Contains the list of parametersthat each script needs to carry out a task. Each script contains a different number of parameters

real.py script: This script is responsible for the execution of the REALmodule of the WRF model. It uses information obtainedfrom the METGRID module, and, as such, has to be executed afterwards. One only has to execute the real.exeprogram using mpirun, indicating the number of nodes.

wrf.py script: The wrf.py script is responsible for the execution of theWRF module. It uses information obtained from REALmodule, and, as such, has to be executed afterwards. Oneonly has to execute the wrf.exe program, using the mpirun,indicating the number of nodes.

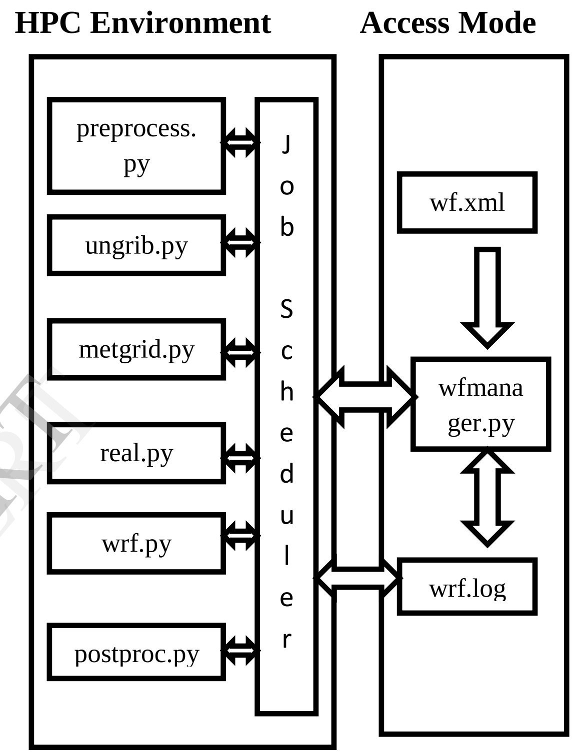

Fig3. Software Architecture

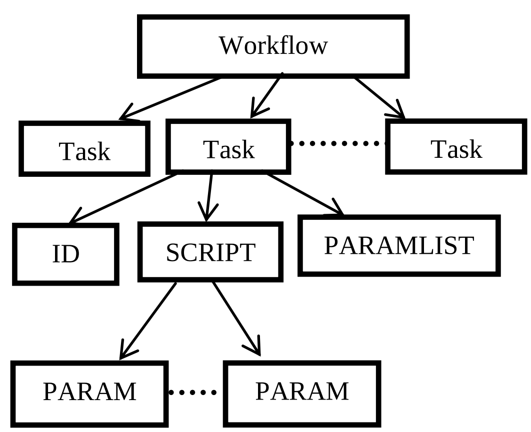

Fig4. Work-flow Hierarchical Structure

3.2.2 Seasonal Climate Forecasting

The CGCM is run by the BoM out for 9 months every day. Forecast products are generated from dynamical model output using data analysis software. The resulting derived forecast products are persisted in self-describing files with additional metadata to support the clients that deliver the outlooks. Forecast data is exposed via a data server. Scheduled processes access and reformat the data for SCOPIC (Seasonal Climate Outlooks for Pacific Island Countries) access. Custom web services use the data server's interface to the forecast data to provide maps, data, and line plots. The Pacific Adaptation Strategy Assistance Program (PASAP) Portal consumes the outputs of the custom web services, and displays model based outlooks as overlays on dynamical maps and standard plots.

The high predictability of seasonal climate in the tropical Pacific provides opportunities for using seasonal forecasts to improve the resilience of climate sensitive sectors throughout the region. Since 2004 the Pacific Island-Climate Prediction Project (PI-CPP) managed by the Australian Bureau of Meteorology (BoM) has built seasonal prediction capabilities within National Meteorological Services (NMS) of Pacific Island countries through the development and provision of decision support software and training. The software, SCOPIC (Seasonal Climate Outlooks for Pacific Island Countries) uses a statistical approach to generate seasonal outlooks based on discriminant analysis using relationships between local predict and variables.

3.2.3 Global Data Forecasting System The Global Forecast System (GFS) is a global numerical weather prediction system containing a global computer model.This mathematical model is run four times a day and produces forecast up to 16 days in advance.It is widely accepted that beyond 7 days the forecast isvery general and not very accurate.The main purpose of the GDPFS shall be to prepare and make available to Members in the most cost effective way meteorological analyses and forecasting products.

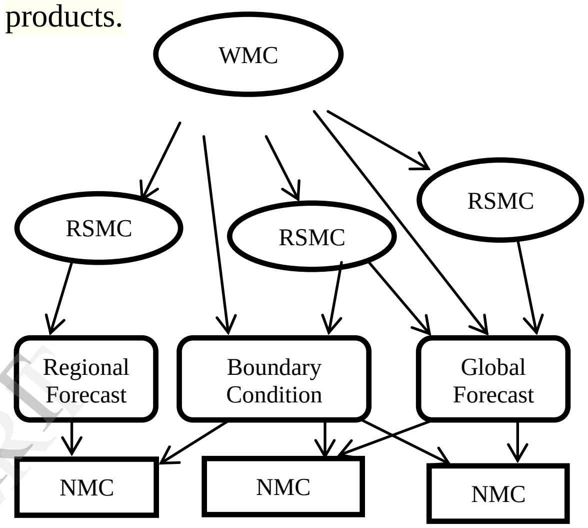

Fig5. World Wide Network for Data

Functions of GDPFS

- Real-time functions of the GDPFS shall include: Pre-processing of data e.g. retrieval, quality control, sorting of data stored in a database for use in preparing output products
- Preparation of forecasting products (fields of basic and derived atmospheric parameters) with up-to global coverage.
- Preparation of specialized products such as limited area very-fine mesh short, medium, extended and long range forecasts, regional climatewatches, and environmental

ISSN: 2278-0181
Vol. 2 Issue 7, July - 2013

quality monitoring and other purposes.

- Monitoring of observational data quality
- Post-processing of NWP data using workstation and PC-based systems with a view to producing tailored value added products and generation of weather and climate forecasts directly from model output.
- Preparation of special products for climate-related diagnosis (e.g. 10-day or 30-day means, summaries, frequencies, anomalies and historical reference climatologies) on a global or regional scale
- Maintenance of a continuously-updated catalogue of data and products stored in the system
- Exchange between GDPFS Centre's of ad hoc information via distributed databases.

3.2.4 Implementation of Global Forecast System (GFS)

A new Global Forecast System (GFS) has been implemented at Northern Hemisphere Analysis Center of IMD on High Power Computing Systems (HPCS). The new GFS is running in experimental real-time mode since 15th January 2010. This new higher resolution global forecast model. The GFS at IMD Delhi involves 4 steps as given below:

Steps 1 - Data Decoding and Quality Control: First step of the forecast system is data decoding. It runs 48 times in a day on half-hourly basis, as soon as GTS data files are updated at regional telecom hub (RTH) of global telecom system (GTS) at IMD New Delhi.

Steps 2 – Preprocessing of data (PREPBUFR): Runs 4 times a day at 0000, 0600, 1200 & 1800 UTC.

Step 3 - Global Data Assimilation (GDAS) cycle: The Global Data Assimilation cycle runs 4 times a day (00, 06, 12 and 18 UTC). The assimilation system is a global 3-dimensional variational technique, based on NCEP's Grid Point Statistical Interpolation (GSI) scheme, which is the next generation of Spectral Statistical Interpolation (SSI).

Step 4 – Forecast Integration for 7 days: The analysis and forecast for 7 days is performed using the HPCS installed in IMD Delhi. One GDAS cycle and seven day forecast (168 hour) run takes about 30 minutes.

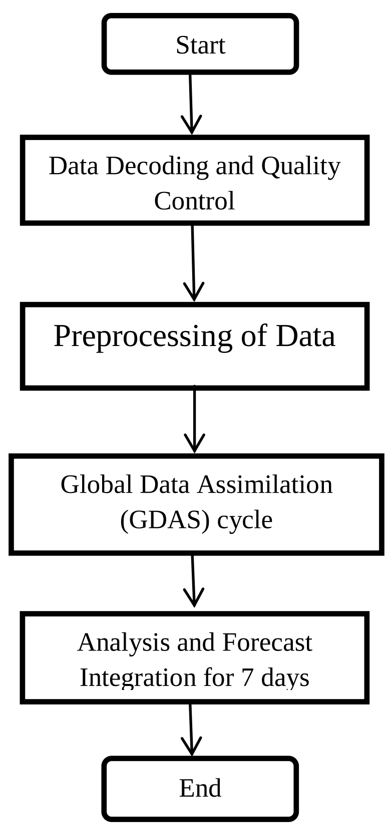

Fig6. Flow Chart of Global Forecast System

4.HPCS use at the Meterological Centre High Performance Computing System (HPCS) with peak speed 14.2 Tera Flop was commissioned in IMD New Delhi.The High end servers at 12 different locationsacross the country(Pune; Regional Met. Centers Delhi, Kolkata, Chennai, Mumbai, Guwahati and Nagpur; Met. Centers Ahmedabad, Bangalore, Chandigarh, Bhubaneswar and Hyderabad) are installed.

a. Computing Racks with peak Power: Peak Speed 14. 4 Tera FLOPS, 28 Nodes: POWER-6, 4.7 GHz Processors &128 Giga Bytes Memory per Node.

b. Storage: 300 Tera Bytes (100 TB online and 200 TB near online), **Archival:**200 Tera Bytes

c. Operating Environment: IBM-AIX 5.3 with Parallel Computation Support

d. Network Bandwidth: 10 Gbps for Switching (Clustering)

e. Computing Power:4 High End Servers with a total Computing Power (134 GF x 4) = 536 G FLOFS,8 Racks for Storage, 1 Rack of Robotic Tape Library

f. Computer System: (a) Altix- 350 (b) 0rigin 200 and (c) IBM P5/595 (64 processors).

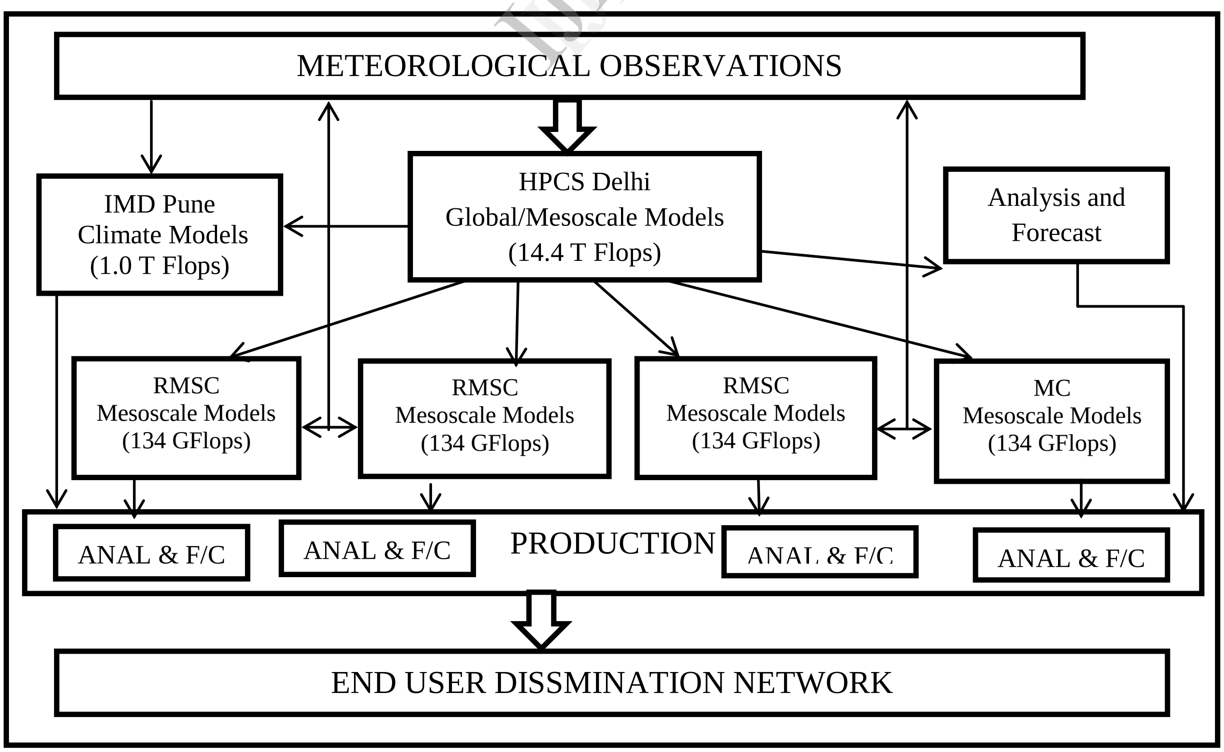

Fig7.High Performance Computing System (HPCS) Data Flow Diagram

5. Proposed Methodology

5.1 General Data Mining Rainfall Prediction Model:

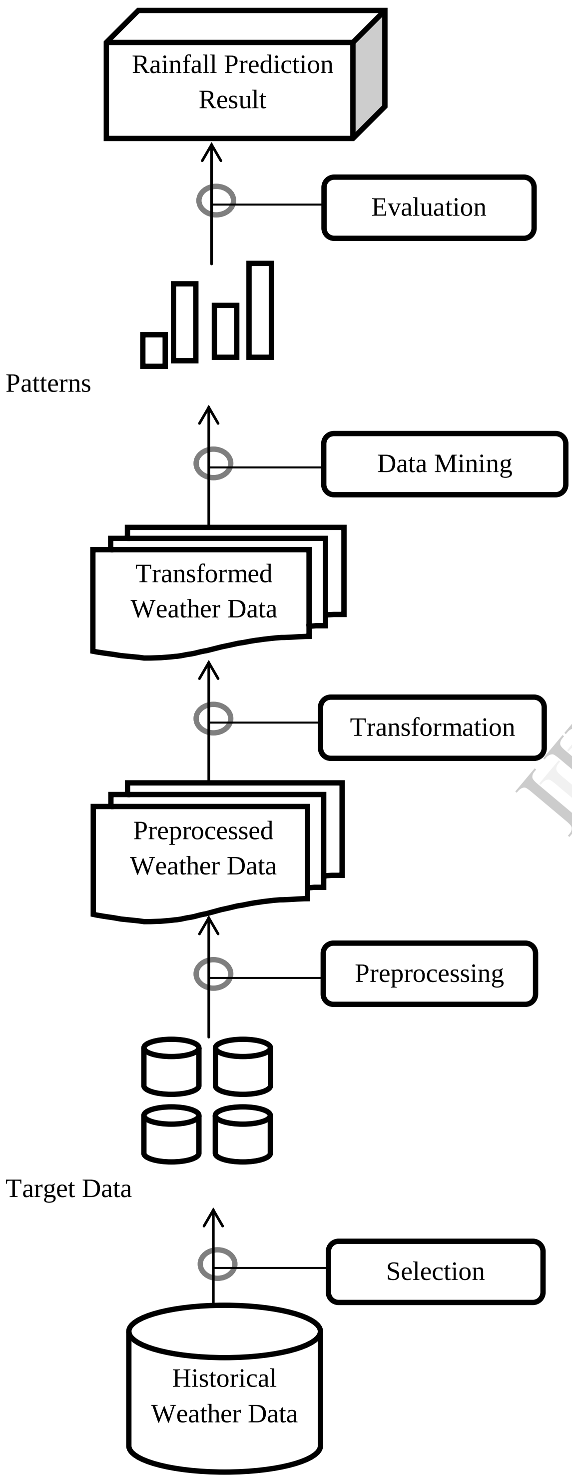

Fig8. General Data MiningModel for Rainfall Prediction

In general data mining prediction model first we collect the historical weather data. Data set were collected from Indian Meterological Department Pune. The collected data consist of different features include daily dew point temperature (Celsius), relative humidity, wind speed (KM/H), Station level pressure, Mean sea level, wind speed, pressure and rainfall observation.Creating a target data set selecting a data set or focusing on a subset of variables or data samples on which discovery is to be performed.

Then important step in the data mining is data preprocessing. One of the challenges that face the knowledge discovery process in meteorological data is poor data quality. For this reason we try to prepare our data carefully to obtain accurate and correct results. First we choose the most related attributes to our mining task. For this purpose we neglect the wind direction. Then we remove the missing value records. In our data we have little missing, because we are working with weather data.Then finding useful features to represent the data depending on the goal of the task.

After preprocessing and transforming the weather data choosing the data mining task i.e. classification, regression and decision tree. Then applying different data mining techniques i.e. K-NN, Naïve Bayesian, Multiple Regression and ID3 on weather data set and makes the rainfall prediction i.e. Rainfall Category or No Rainfall Category.

Conclusion

In this paper we study the different numerical weather prediction model and general data mining techniques for rainfall prediction. Data mining tasks provide a very useful and accurate knowledge in a form of rules, models, and visual graphs. This knowledge can be used to obtain useful prediction and support the decision making for different sectors. So we used different data mining techniques on Meterological data set to predict the rainfall on thebasis of previous year (Historical) Weather data set. This study will help us for rainfall prediction.

References

[1] Dale Barker, Xiang-Yu Huang, Zhiquan Liu, Tom Auligné, Xin Zhang, Steven Rugg, Raji Ajjaji, Al Bourgeois, John Bray, Yongsh eng Chen, Meral Demirtas, Yong-Run Guo, Tom Henderson, Wei Huang, Hui-Chuan Lin, John Mich alakes, Syed Rizvi, and XiaoyanZhang "The Weather Research and Forecasting model's community variational/ensemble data assimilation system WRFDA."

[2] Andrew Charles1, David McClymont2, and Roald de Wit1, David Jones1 A "Software architecture for seasonal climate forecasts in the tropical Pacific Australian Bureau of Meteorology", DHM Environmental Software Engineering Pty. Ltd. 19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011.

[3] "Manual on theGlobal Data-processing andForecasting System" World Meterological Organization. Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)

[4] Folorunsho Olaiya Department of Computer & Information Systems, Achievers University, Owo, Nigeria Adesesan Barnabas Adeyemo University of Ibadan, Ibadan, Nigeria "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies"I.J.

[5]http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html.

[6] A.M. Guerrero-Higueras, E. Garca-Ortega and J.L. SanchezAtmospheric Physics Group, University of Leon, Leon, SpainJ. LorenzanaFoundation of Supercomputing Center of Castile and Leon, Leon, SpainV. MatellanDpt. Mechanical, IT, and Aerospace Engineering, University of Leon, Leon, Spain "Schedule WRF model executions in parallel computing environments using Python"

[7]Andrew Kusiak, Member, IEEE, Xiupeng Wei, Anoop Prakash Verma, Student Member, IEEE, and Evan Roz "Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach" IEEE Transactions On Geoscienceand Remote Sensing 1.

[8] Umesh Kumar Pandey S. Pal VBS Purvanchal University, Jaunpur Umesh Kumar Pandey et al," Data Mining: Aprediction of performer or underperformer using classification" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol 2 (2), 2011.

[9] http://en.wikipedia.org/wiki/Weather_Forecasting

[10] INDIA METEOROLOGICAL DEPARTMENT (IMD)Ministry of Earth Sciences (MoES)Government of IndiaNew Delhi

[11] Weather Research and Forecasting (WRF) Model Performance Research and ProfilingOctober 2008.

[12] Gilad Shainer1, Tong Liu1, John Michalakes2, Jacob Liberman3,Jeff Layton3, Onur Celebioglu3, Scot A. Schultz4, Joshua Mora4, David Cownie4 Mellanox Technologies Dell, Inc. 2National Center for Atmospheric Research Advanced Micro Devices (AMD) "Weather Research and Forecast (WRF) ModelPerformanceand Profiling Analysis on Advanced MulticoreHPC Clusters".

[13] Yongjian Fu Department of Computer Science University of Missouri Rolla" Data Mining: Tasks, Techniques and Applications".

[14] Annual Joint WMO Technical Progress Report on the Global Data processing and Forecasting System (GDPFS) including Numerical Weather Prediction (NWP) Research ActivitiesMarch 2010.