National Conference on Advancements in Computer Science and Engineering
In association with International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Rainstorm Prediction System

Ms. Harshitha H[1], Ms. Pooja Kumari[1], Ms. Simran Agarwal[1]

[1]Department of Computer Science, New Horizon College of Engineering, Outer Ring Road, Panattur post, Kadubeesanahalli, Bengaluru, Karnataka, India

ABSTRACT

Rainstorm is a devastating disaster that usually occurs during rainy seasons at Himalayan regions. The recent floods in the _Kedarnath' area, Uttarakhand are a classic example of flash floods in the Mandakini River due to cloudburst that devastated the country by killing thousands of people besides livestock. The traditional methods used for cloudburst prediction are weather forecasting, data mining techniques for weather prediction by modelling meteorological data, laser beam atmospheric extinction measurements from manned and unmanned aerospace vehicles. These techniques are more expensive and time consuming along with uncertainty of accurate prediction. The proposed method in this paper is Arduino based cloudburst predetermination system with real time calculation of rainfall intensity.

The rainfall prediction is done with the use of machine learning in minimal costs. The complete weather forecasting setup is flexible enough to be installed anywhere and make weather predictions without much historical experience. We used different machine learning algorithm to check the accuracy of rainfall prediction.

Keywords— Aerospace, Disaster, Forecasting; Intensity, Prediction, Arduino.

I. INTRODUCTION

Cloudbursts are extreme weather events in which an area registers more than 100 mm rain in just one hour. So, create a prediction system that will be able to forecast extreme weather events such as cloud bursts at least two days in advance.

Early prediction and warning of such severe local weather systems is crucial to mitigate societal impact arising from the accompanying flash floods.

Farmers can know when to plant or harvest their crops.

People can choose where and when to take their holidays to take advantages of good weather Surfers known when large waves are expected Regions can be evacuated if hurricanes or floods are expected Aircraft and shipping rely heavily on accurate weather forecasting

This architecture is compared with other previous proposals and it demonstrates an improvement on the ability to predict the accumulated daily precipitation for the next day.

II. MOTIVATION BEHIND THE RESEARCH

Today's world the enormous data sets required and inherent unpredictability of the Earth's atmosphere makes predicting future events very tricky indeed. Current computer models are required to make judgments of several large-scale phenomena. These

include things like how the Sun heats the Earth's atmosphere; how pressure differences are affect wind patterns and how water-changing phases (ice to water to vapor) affect energy flow through the atmosphere.
One example is the Numerical Weather Prediction (NWP). This model studies and analyses vast data sets from satellites and other sensors to provide short term weather forecasts and long term climate predictions.
Complexity of time series data for various cases of weather patterns with more number of parameters can result in quantitative estimation
This system also helps many people to save their lives before knowing the weather in advance.

III. LITERATURE SURVEY

Literature survey is the most important step in the software development process (4). Machine learning technology adapts and responds to data, learning over time to better answer search queries. This helps provide better accuracy in search results at a speed beyond human capabilities.

Cloudburst Predetermination System Arpit Tiwari 1 , S K Verma2 (Department of CSE, GBPEC, Pauri, India)

IOSR Journal of Computer Engineering (IOSR-JCE) Cloudburst is a devastating disaster that usually occurs during rainy seasons at Himalayan regions. The recent floods in the _Kedarnath' area, Uttarakhand are a classic example of flash floods in the Mandakini River due to cloudburst that devastated the country by killing thousands of people besides livestock. These techniques are more expensive and time consuming along with uncertainty of accurate prediction.

- **A neural network-based local rainfall forecast system using meteorological data on the Internet:**

A case study using data from the Japan Meteorological Agency
In this study, we evolve and test a native rainfall (precipitation) prediction system based on artificial neural networks (ANNs). Our arrangement can involuntarily obtain meteorological statistics used for rainfall forecast from the Web. Meteorological data from apparatus fixed at a local point is also shared among end users in our system.

- **A neural network-based local rainfall forecast system using meteorological data on the Internet:**

In this study, we evolve and test a native rainfall (precipitation) prediction system based on artificial neural networks (ANNs). Our arrangement can involuntarily obtain meteorological statistics used for rainfall forecast from the Web.
Meteorological data from apparatus fixed at a local point is also shared among end users in our system. The ultimate goal of the work was the practical usage of —big data‖ on the Internet. Also, the splitting of data amongst end users for precise rainfall prediction.

IV. COMPARATIVE ANALYSIS

In this paper the comparative analysis mainly done based on the temperature, Atmosphere, humidity. Based on the above criteria the below section shows the some of the aspects that were less/not found in the previously existed system.
The existing system produces certain disadvantages;

- It is extremely difficult to accurately predict the rainstormin two or three days(5).
- The traditional methods used for cloudburst prediction are weather forecasting, data mining techniques for weather prediction by modelling meteorological data, laser beam atmospheric extinction measurements from manned and unmanned aerospace vehicles.(5).
- These techniques are more expensive and time consuming along with uncertainty of accurate prediction.
- The back-propagation-feed advance neural network can be utilized in numerous applications such as character identification, weather and financial forecast, face detection etc(5).

- The present paper presents a mechanism of prediction of cloudburst by detecting the formation of cumulonimbus type cloud, using the brightness temperature (TB) (5).
- The proposed system removes some of the drawbacks of existing system such as;

The proposed system provides the prediction in 2 or 3 days advance.

- Unlike traditional system,which predicts the rainstormin 5 to 6 days. Documents with quality rich textual content will be included in the top results.
- This system also helps many people to save their lives before knowing the weather in advance.
- Most of the documents that are produced in the results are most relevant to the user‘s request.
- It consumes very less amount of time to be implemented unlike other techniques that consume a lot of time to process very huge database and further finding patterns of hidden knowledge in order to produce predictions.

V. PROPOSED SYSTEM

This system will predict rainstorm based on parameters such as temperature, humidity and wind. This system is a web application with effective graphical user interface .System will take this parameter and will predict cloud burst by comparing the present and previous data in database. System will calculate cloud burst based on these data, therefore this prediction will prove reliable.

This also uses the ARDUNO and RAIN GAUGE to get the precipitation of the rain and then it will get the output to the machine leaning algorithms.

Here we are implementing on the basis of machine learning with its algorithms which are suitable for forecasting the rainfall in particular. Depending upon the various factors given above.

We are implementing through APIs of dataset and there we are taking the data for which state we are predicting the cloudburst.

It consumes very less amount of time to be implemented unlike other techniques that consume a lot of time to process very huge database and further finding patterns of hidden knowledge in order to produce predictions.

- Descriptions of data to be entered into the system
- Descriptions of operations performed by each screen
- Descriptions of work-flows performed by the system
- Descriptions of systemreports or other outputs
- Who can enter the data into the system
- How the systemmeets applicable regulatory requirements

A Support Vector Machine (SVM)

This a computer algorithm that learns by example to find the best function of classifier hyperplane to separate the two classes in the Input space.The SVM analyzed two kinds of data, i.e. linearly and non-linearly separable data .The example of linearly separated data is shown in fig. below. Best hyperplane between two classes can be found by measuring the hyperplanemargin and find out the maximum points

$$f(x) = w\,x + b \quad (1)\ T$$

Where x refers to a training pattern, w is referred to as the weight vector and b as the bias term

Artificial Neural Networks:

Artificial neural networks (ANN) have existed in various forms since the 1940s (McCulloch and Pitts, 1943; Good fellow et al., 2016), but have received renewed interest in recent years (Good fellow et al., 2016). An ANN is a collection of neurons, which are small computational units that superficially mimic the way neurons work in nature. A single neuron is simply a weighted sum of a set of inputs, plus a bias, with an applied activation function, Fig. 1 (left). A non-linear activation function fact(•) is important for

success in applying ANNs, otherwise the resulting model output is simply a linear combination of the inputs. The equation for a single neuron can be written as: yk = fact(b+xiwi) (1) The power of ANNs comes from connecting many neurons together in a network. The simplest network structure is a feed forward network, as shown in Figure 1 (right). Neurons are connected in simple layered structures where the inputs of each neuron are connected to all the outputs of the previous layer. If we describe the inputs xi and weights wi in matrix form, we can write a whole layer of neurons as: y = fact(W x) (2) where the bias is included as w0 = b by adding an artificial constant input x0 = 1,. A feed forward ANN is built by connecting multiple layers together. The inputs to the network are connected to the inputs of the first hidden layer. The first hidden layer can then be connected to more hidden layers. The last hidden layer connects to the output layer. The output of the ANN is given by this output layer. We can then write a single non-linear matrix equation for the whole network. An example equation for an ANN with three hidden layers is:

y out = f out act (Wout f (2) act (W(2) f (1) act (W(1) x
))) (3)
]

Equation (3) shows that an ANN is simply a non-linear matrix equation with a large number of coefficients. Each W(j) matrix can be large, thus allowing the ANN model to fit complex non-linear systems.

Random forest:

Random forest is a kind of machine-supervised learning algorithm on understanding the ensemble. Ensemble training is indeed a form of erudition whereby you multiply combine various kinds of algorithms or just a similar procedure to create a more efficient model of forecasting. This algorithm combines many algorithms of a similar kind i.e. multiple decision trees, resulting in forest of trees, hence the term "Random Forest". This algorithm may be used for tasks of regression as well as classification. The initial steps involved in applying the random forest algorithm are selecting M arbitrary archives from the dataset, constructing a decision tree largely dependent on those M archives, choosing the amount of trees we want to see in the algorithm and repeating steps one and two. For classification issues, the group to which the new record belongs is predicted by each forest tree. At last, the new record was allocated to the division where the majority vote is received. The algorithm of random forests is also not bias, as there have been various trees and then each tree is proficient on a data subset. The random forest algorithm relies on the influence of "the audience;" hence, the algorithm's common bias reduces.

Polynomial regression

Like linear regression, uses the relationship between the variables x and y to find the best way to draw a line through the data points. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x). Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function E(y | x) is linear in the unknown parameters that are estimated from the data.

Logistic regression

This is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).Mathematically, a logistic regression model predicts P(Y=1) as a function of X. It is one of the simplest ML algorithms that can be used for

various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

The advantages of proposed systemare;

- Rapid and abrupt climate changes have remarkably increased the importance of a weather app.
- Today, weather apps can provide accurate forecasts to enable users to make informed decisions.
- Integration of advanced features and technological progress have made it easy to know the sudden changes in the weather conditions along with real - time alerts.
- Since ages, humans want to know about weather conditions in advance. In the current age of the Internet and mobility, weather apps make it possible to predict weather conditions on the move. Summing up, your weather mobile app with necessary features can make the users prepared for any weather-related

VI. ARCHITECTURE OF THE SYSTEM

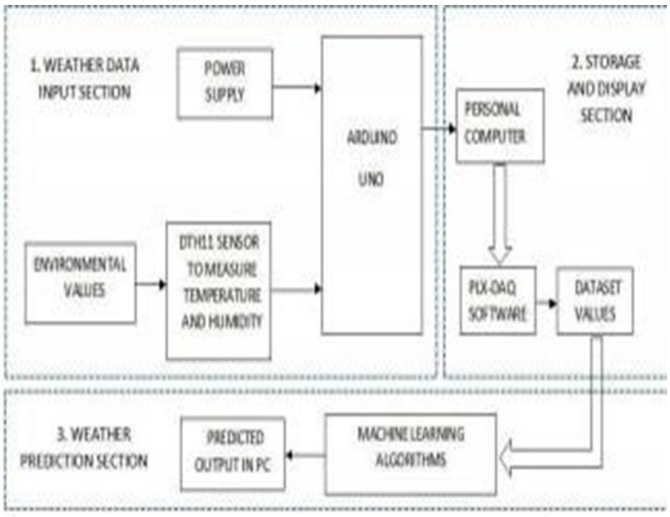

Fig: 1 Architecture of the System

The above architecture diagram shows us the estimated overall architecture of the whole system

A. Weather Data Input section

In this section, a combined sensor to measure both temperature and humidity is used. The sensor works well to send the real time data continuously to the interfaced microcontroller. The controller is powered with a voltage supply after rectification done in a DC adapter. The function of this section is to collect the readings of the weather data. The measured values are ready for further processing based on the requirement.

B. Storage and Display Section

In this section, the output is then given to the Personal computer and then displayed in an Excel file. This is done by means of specific software to interface the Arduino outputmeasured values to excel file. The final sheet of excel will be the dataset file available for the machine learning prediction.

C. Weather/Rainstorm Prediction Section

This section includes the setup for the dataset being given as input to machine learning models to accurately predict the rainfall based on various algorithms. The algorithms will make the prediction based on various approaches and different sets of data. The predicted percentage value is displayed on the Personal computer and can be compared with previous historical weather input

VII. IMPLEMENTATION

The flowchart diagram below clearly describes the working of the system.

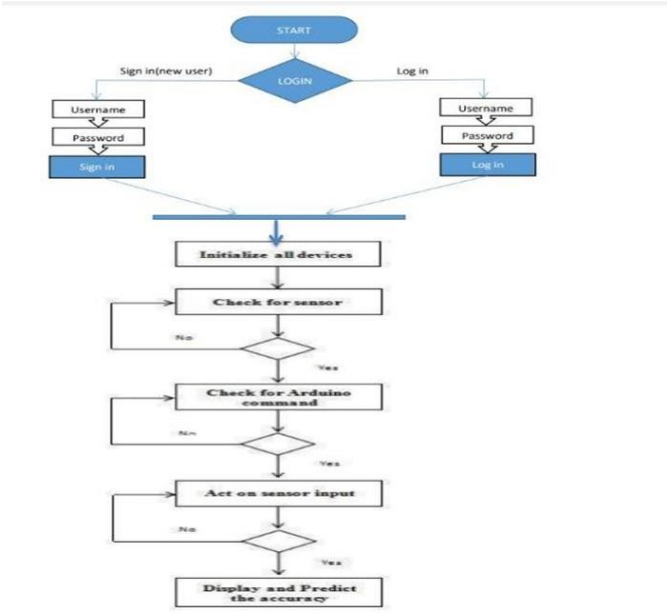

Fig: 2 Flowcharts

So, to have a better understanding of the dataset and for better comparison, first preprocess should be done. Before going for the prediction, preprocess can be done. It is representation of the dataset in form of graph. It eases the process of comparison and along with that it also gives a better understanding of the dataset present. Dataset should be split in two parts, the first part deals with training the algorithm used and the rest part used to predict the amount of rainfall. Rainfall is predicted only with the algorithm with more accuracy. The algorithm used should undergo training before it does prediction. So, in this part of the system, the training is been done. This is done with the above mentioned approaches. This step gives a proper idea of which algorithm is more accurate among the all. Then the remaining dataset (which is not used in training) is being used and rainfall prediction is been done. This part is also done in both the approaches. Finally, after the all the process is completed, the result is received in form of graph and table which shows the future rainfall and the accuracy of the algorithm. The accuracy is received in the form of Metrics and excel sheet. In Metrics along with the accuracy different types of errors are also shown and the same is represented in the excel sheet. After all, at last the predicted value is stored in excel sheet and is received.

A. Errors calculated

The accuracy of the approaches is being calculated against the types of errors that can produce negative effect on the algorithm. These errors can affect the algorithm‘s accuracy and hence are being calculated. The types of errors that is being calculated are MAE, MSE, RMSE and RSQUARED. MAE calculates all the absolute errors and then finds the mean value for all. It first calculated the mean of all the dataset present, then subtracts the mean value with each data individually and add all the resultant value and finally divides it with the total number of dataset present. MAE = (1) Next error is MSE. It is almost similar to mean absolute error. MSE = (2) The only difference is, instead of adding the resultant (subtracted value of mean with each dataset), it finds the square of it and add them. RMSE error is being calculated by subtracting all the predicted and actual values with each other, finding all the squares of it and adding all the squared value the total value that we will receive is stored. The stored value then further divided by total values present. The resultant value is squared rooted. (4) (X= regression error (sum squared) Y = total error (sum squared)) the above errors are being calculated by subtracting the division value of sum squared regression error and sum squared total error with value one.

B. Front end

This screen will pop up in the screen as soon as the project starts. The user needs to give the login for using the application. If he/she doesn't have they can register with the new email id and password and can login so they can use the application successfully.

C. Anaconda

Anaconda is an open source software used to run python codes.

Jupiter notebook, spyder

The errors values and the accuracy value are popped up in anaconda with all algorithms. In anaconda, while the project is being executed, the dataset which is being used is shown there. So, while execution of the code, the user can cross verify it and can stop the processing instant on finding that the wrong dataset is being used.

D. Visualization

This option deals with representing the dataset in form of graphs. Different types of graphs are being produced after execution of this process. Below are the different Metrics and graphs:

Fig 1 explains the Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them.

Fig 3

Fig 4

The accuracy score achieved using random forest to compare with all the accuracy score in project in fig 2

Fig 5

The accuracy score of the rainstorm is achieved by decision tree algorithmwhich is in fig 3.

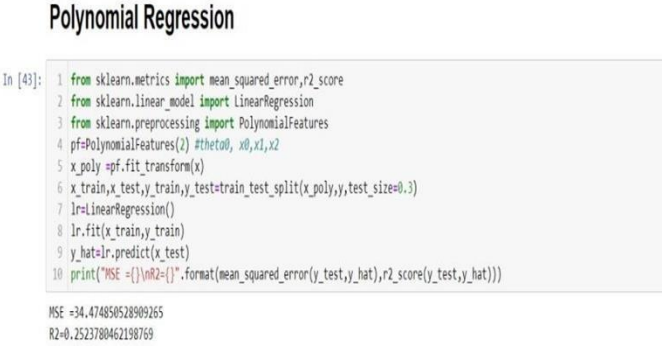

Fig 6

The fig 4 explains the prepossessing of the data using polynomial regression.

Output Snapshots:

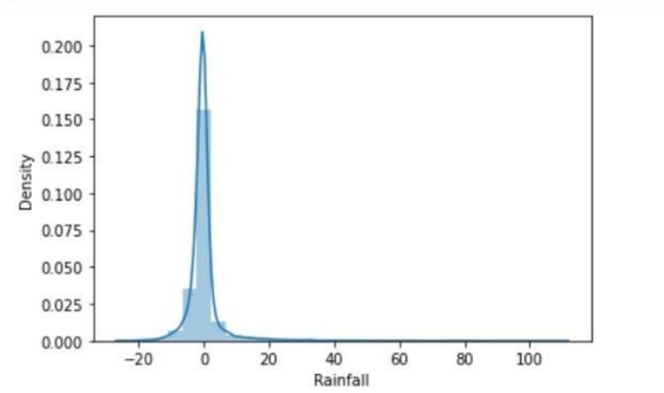

Fig 7: Density with rainfall

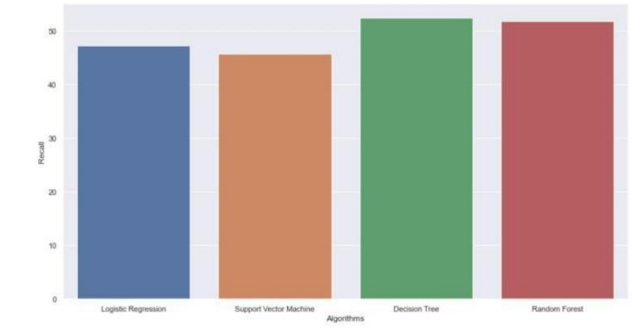

Fig 8:Recall of algorithms

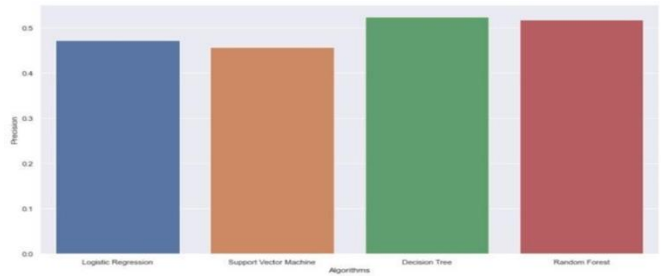

Fig 9: Accuracy of algorithms

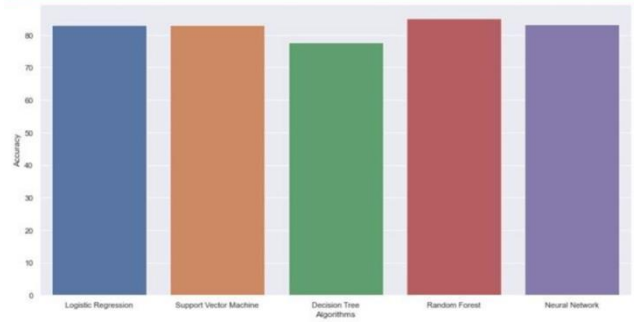

Fig 10 :Precision of algorithms

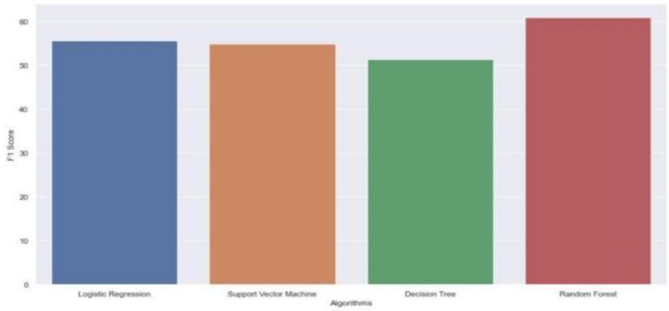

Fig 11: F1 score of algorithms

VIII. PROPOSED METHODOLOGY

This system proposes an advanced answer for the weather monitoring system using IoT to make its real time data easily accessible over a very wide range. Predicting rainfall is one of the tougher task but using appropriate parameters and classifying themcan help in predicting rainfall.

Classifying the data using SVM algorithm makes it easier to predict the rainfall. The system deals with monitoring weather and climate changes like Temperature, Humidity, Wind speed, Carbon monoxide level in the air, Light intensity, UV radiations, Soil moisture and after analysing and classifying the parameters, rainfall percentage is predicted. The result will be received in the form of graphs and excel sheets. For preprocess , all the result will be received in the form of different graphs and for machine learning and neural network , the accuracy will be received in the form of Metrics as well as excel sheet and accordingly the predicted value will be received in the form of excel sheet which will contain two columns ID and predicted value. IDs will be same as that of in the datasheet. To get for which region prediction is being done, IDs should be matched with the IDs present in dataset.

A. Ardiuno Uno

It is an open-source physical computing stage based on a easy micro-controller board, and a development environment for writing software for the board. Arduino can be used to develop interactive objects, taking inputs from a variety of switches and or sensors, controlling a variety of lights, motors, and other physical outputs.

Arduino is a single-board microcontroller, intended to make the application of interactive objects or environments more accessible.

Fig 12: Ardiuno Uno Board

B. Cloud Platform (WEB APP)

Cloud platform is designed to store and process Internet of Things (IoT) data. This platform is built to take massive volumes of data generated by devices, sensors, applications, websites and initiate actions for real time responses.

Temperature and humidity sensor The DHT11 is a fundamental, ultra-low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor

and a thermistor to calculate the neighboring air, and spits out a digital signal on the data pin (no analog input pins needed)

IX. CONCLUSION

The conclusion is drawn that a very reliable, efficient and less cost product has been developed which can make life more comfortable and securable. This project does not require any hard installations and can be easily installed in old installations. So, it is easily compatible with old systems. Since the project is prediction based and thus doesn't require any extra cost of installing software. Our project will prove to be efficient for many weather monitoring stations, as it will help them in having an alternative solution when there is any absence of the high cost weather monitoring setup. The rainfall prediction is done with the use of machine learning in minimal costs. The complete weather forecasting setup is flexible enough to be installed anywhere and make weather predictions without much historical experience.

We used different machine learning algorithm to check the accuracy of rainfall prediction. We have compared SVM, Random Forest, Navie Bayes and MLP (Multilayer perceptron) classifiers. From the above figure 3 we can conclude that Random forest is the Machine learning algorithm which is suitable for rainfall prediction in India.

Currently machine learning used in no. of industries. As the data increases the complexity of that data will increase and for that we are using machine for the better understanding of that data. In Rainstorm predictions its pretty helpful with good accuracy score and in rainfall also its gives pretty good predictions. In future we are planning to increase our work in Crop prediction and Flood prediction with the rainstorm prediction

X. REFERENCES

[1]. http://www.arduino.cc

[2]. Data mining techniques for weather prediction: a review, International Journal on Recent and Innovation Trends in Computing and Communication, Volume:2, Issue:8ISSN:2321-8169 2184 – 2189

[3]. Kedarnath flash floods: a hydrological and hydraulic simulation study, Research Communications

[4]. National Remote Sensing Centre, Indian Space Research Organisation, Balanagar, Hyderabad

[5]. Arduino, G., Reggiani, P. and Todini, E., Recent advances in flood forecasting and flood risk assessment. Hydrol. Earth Syst. Sci., 2005, 9(4), 280–284.

[6]. Srinivasan, J., Predicting and managing extreme rainfall. Curr. Sci., 2013, 105(1), 7–8.

[7]. IMD, A preliminary report on heavy rainfall over Uttarakhand during 16–18 June 2013. India Meteorological Department, Ministry of Earth Sciences; July 2013; http://imd.gov.in/doc/ uttrakhand_report_04_09_2013.pdf 9 [8]. Auroop R Ganguly, and Karsten Teinhaeuser, —Data Mining for Climate Change and Impacts,‖ IEEE International Conference on Data Mining, 2008.

[8]. Badhiye S. S., Wakode B. V., Chatur P. N. —Analysis of Temperature and Humidity Data for Future value prediction‖, IJCSIT Vol. 3 (1), 2012

[9]. http://en.wikipedia.org/wiki/Weather_forecastin g#Modern_method

[10]. V. K. Dabhi and S. Chaudhary, —Hybrid Wavelet-Postfix- GP model for rainfall prediction of Anand region of India,‖ Advances in Artificial Intell., pp. 1-11, 2014.

[11]. Pinky Saikia Dutta, Hitesh Tahbilder, —Prediction Of Rainfall Using Data mining Technique Over Assam‖, Indian Journal of Computer Science and Engineering (IJCSE),