International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

An Implementation of Data Mining Technique for Weather Forecasting

Rakshanda Dharmik[1], Anjali Pandey[1], Shraddha Hude[1], Sneha Mankar[1], Prof. Manoj S. Chaudhari[2]
[1]BE Scholars, Department of Information Technology Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, India
[2]Assistant Professor, Department of Information Technology Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, India

ABSTRACT

Weather forecasting is a critical application in meteorology and has been a standout amongst the most logically and mechanically difficult issues the world over. In this paper, we research the utilization of data mining strategies in forecasting most extreme temperature and rainfall. Weather prediction approaches are tested by complex weather wonders with restricted perceptions and past data. Weather wonders have numerous parameters that are difficult to identify and measure. Expanding improvement on correspondence systems empowered weather forecast master systems to coordinate and offer assets and along these lines hybrid system has risen. Despite the fact that these upgrades on weather forecast, these master systems can't be completely dependable since weather forecast is primary issue.

Keywords: Weather Forecast, Data Mining, Classification, Supervised Learning, Implementation, Performance Study.

I. INTRODUCTION

As of late various new strategies and the new innovations are showed up. Among them the machine learning and data mining applications are delivering their significant commitments. In the different applications, for example, online business suggestion, securities exchange prediction, spam sifting and others are produced with the assistance of data mining procedures. The data mining methods are offered to dissect the chronicled data and set up their encounters. This experience or the learning is utilized to recognize the comparative sort of data for characterization assignment, or for making prediction and estimation. In this proposed work the machine learning based order and prediction is contemplated in detail.

Likewise of that utilizing the data mining strategy another prediction show is readied utilizing the hybrid system of machine learning. The proposed method is a weather forecasting model; the proposed weather forecasting procedure is breaking down the chronicled weather data and the concerned properties. These traits are expected focuses, dampness and others, these qualities are in charge of Clemet or weather conditions. In this manner utilizing these qualities investigation the weather forecast is performed. For assessment of the characteristics and recognizable proof of comparable examples the proposed method incorporates both the managed learning and unsupervised learning strategy. Along these lines the proposed method is a hybrid strategy of learning and prediction.

As talked about already data mining methods are useful for different sorts of example extraction, target acknowledgment and prediction. These extricated important examples frame the data helps for basic leadership, business insight and other innovative

errand. Hence a rich study on the diverse data mining applications all the more particularly the prescient data displaying is performed. In light of the perceptions the data is more successful parameter by which the execution of any indicator is depends hence the whole commitment is characterized in three noteworthy spaces. To start with the data pre-handling, in this stage the sort of data is dissected and the key qualities are perceived. To do this the information data is first assessed and the uproarious substance and the missing data are dealt with. This approach enhances the nature of data for learning and example recognizable proof.

In this undertaking we utilized data mining calculation we utilized Decision tree calculation for grouping weather parameters, for example, most extreme temperature, least temperature, rainfall, evaporation and wind speed as far as the month and year.

II. RELATED WORK

M. A. Kalyankar and S. J. Alaspurkar [2] utilized data mining methods to procure weather data and locate the shrouded designs inside the substantial dataset to exchange the recovered data into usable information for grouping and prediction of weather condition. Data mining process is connected to remove information from Gaza city weather dataset. This learning can be utilized to acquire helpful predictions and bolster the basic leadership process. Dynamic data mining strategies are required to manufacture, that can learn powerfully to coordinate the idea of quickly variable weather nature and sudden occasions.

P. Hemalatha [3] executed data mining strategies for managing the way of the boats amid cruising. Worldwide Positioning System is utilized for distinguishing the region in which the ship is as of now exploring. The qualities of weather data incorporates atmosphere, stickiness, temperature, stormy. The weather report of the territory followed is contrasted and the current database. The examined dataset is given to the choice tree calculation, C4.5 and ID3. The choice got with respect to the weather condition is told to the ship and the way is picked appropriately. A nearby collaboration between the factual and computational groups gives cooperative energy in data investigation. Hardly any nonstop credits should be modified as ID3 can't straightforwardly manage the constant reaches.

A.R.W.M.M.S.C.B. Amarakoon [1] proposed a system that uses the authentic weather data and applies the data-mining calculation "K-Nearest Neighbor (KNN)" for grouping of these chronicled data into a particular time traverse. The k closest time ranges is then additionally taken to anticipate the weather of Sri Lanka. The everyday weather data is gathered for finish one year. It creates exact outcomes inside a sensible time for a considerable length of time ahead of time. It is inferred that KNN is valuable to dynamic data, the data that progressions or updates quickly and gives better execution when contrasted with alternate procedures. Coordinating component choice strategies can even give more precise outcomes.

S.S. De [5] utilized ANN to forecast the Max. And Min. Temperature for Monsoon month. The temperature of June, July and August has been anticipated with the assistance of January to May temperature. The data of three months of 1901 to 2003 is utilized. The ANN display produced here is a solitary concealed layer demonstrate with 2 hubs at shrouded layer. After 500epochs the outcome is approves. The Max. Mistake showed up is 5%.

Mohsen Hayati and Zahra Mohebi [4] uses ANN for one day ahead prediction of temperature. They utilized MLP to prepare and test ten years (1996-2006) meteorological data. For exactness of prediction they split data into four seasons and after that for every season one system is displayed. Two arbitrary concealed days in each season are chosen to

test the execution. The mistake in result shifts between 0 to 2 MSE.

This article [7], presents a correlation of two sub testing nonparametric techniques for outlining calculations to forecast time arrangement from the combined month to month rainfall. Both methodologies depend on simulated bolster forward neural systems (ANNs).

This paper [6] examines the utilization of weather outfit predictions in the use of ANNs to stack forecasting for lead times from one to ten days ahead.

A fluffy k-NN weather prediction system can enhance the strategy of tirelessness climatology utilizing immediate, proficient, master like correlation of over a significant time span weather cases. Persevering climatology is a simple forecasting system for perceiving for short range weather prediction [9].

The indicative meteorology gives a premise to understanding the hole amongst forecasters and specialists. It should give some establishment for symptomatic meteorology and it isn't a weight from which forecasters ought to be mitigated [10]. Rather, it is a fundamental part of logical forecasting.

Examinations the fluffy weather forecasts, which are figured in the system and used to forecast contamination fixations and to explore the viability of forecasting contamination focuses, [8] putting the reliance between specific qualities, depicting the weather forecast all together and demonstrating the pertinent fluffy numbers in air contamination forecasting.

III. IMPLEMENTATION APPROACH

To foresee the weather by numerical means, meteorologists have created environmental models that rough the air by utilizing scientific conditions to portray how climatic temperature, weight, and dampness will change after some time. The conditions are customized into a PC and data on the present environmental conditions are sustained into the PC. The PC unravels the conditions to decide how the distinctive environmental factors will change throughout the following couple of minutes.

The regulated calculations are chipping away at marked data. Such sort of data more often than not found in organized arrangement. This organized data has some pre-characterized class marks that are speaking to the results of the mix of qualities. Consequently utilizing the preparation data the calculation learned on pre-characterized designs. In this exhibited work the Bhopal weather forecasting data for most recent five years are utilized. That preparation dataset contains the distinctive weather characteristics perceptions and the class marks as the weather conditions.

The Methodology of the system comprise of following fundamental stages:

A. Classification:

Classification is an exemplary data mining procedure in light of machine learning. Fundamentally, arrangement is utilized to characterize everything in an arrangement of data into one of a predefined set of classes or gatherings. Order strategy makes utilization of numerical strategies, for example, choice trees, direct programming, neural system, and insights. In characterization, we build up the product that can figure out how to arrange the data things into gatherings. For instance, we can apply arrangement in the application that "given all records of representatives who left the organization, foresee who will most likely leave the organization in a future period." For this situation, we partition the records of workers into two gatherings that named "leave" and "remain". And after that we can ask our data mining programming to characterize the representatives into isolated gatherings.

For achieving the desired result we have opted for C4.5 Classifier which a classification mechanism based on Decision Tree. The Algorithm of C4.5 Classifier is as follows:

Algorithm:

Step 1: Read trained data instances.

Step 2: Calculate Overall entropy

$$Entropy(t) = -\sum_{j} p(j|t)\log_2 p(j|t)$$

Step 3: Calculate entropy of every attribute

$$M_i = Entropy(N_i) = -\sum_{j} p(j|N_i)\log p(j|N_i)$$

Step 4: Calculate information gain of every attribute

$$Gain_{split} = Entropy_{(p)} - (\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i))$$

Step 5: Build Tree

Step 6: build prune tree

The Basic Flow of the system is as shown in Figure 1.

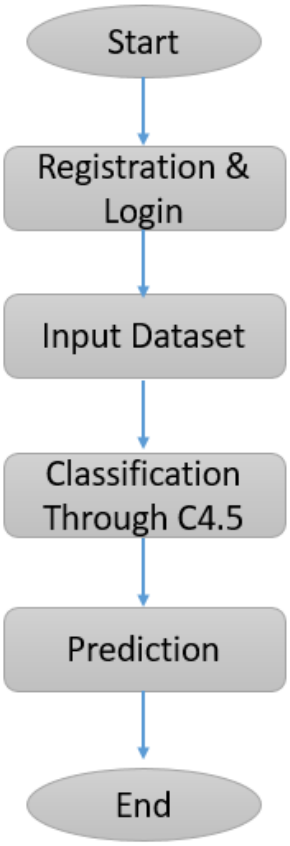

Figure 1. System Flow Chart

B. Prediction:

The prediction, as its name inferred, is one of a data mining strategies that finds the connection between free factors and connection amongst needy and autonomous factors. For example, the prediction investigation strategy can be utilized as a part of the deal to foresee benefit for the future on the off chance that we consider the deal is an autonomous variable, benefit could be a needy variable. At that point in view of the authentic deal and benefit data, we can draw a fitted relapse bend that is utilized revenue driven prediction.

IV. IMPLEMENTATION OUTPUT

Implementation of the system is shown in following figures

Figure 2. Home Screen

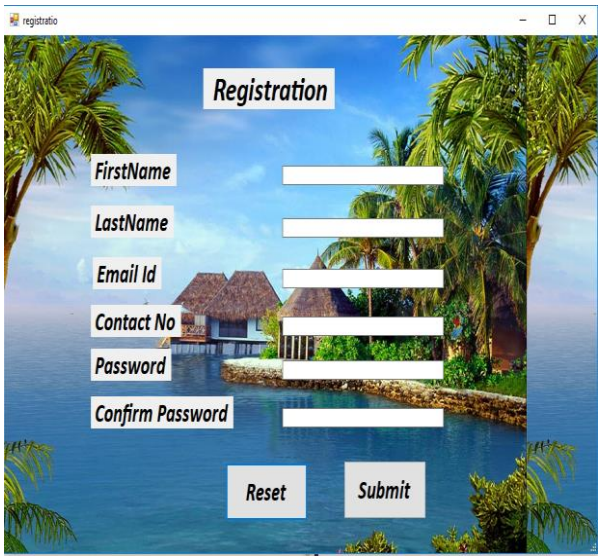

Figure 3. Registration

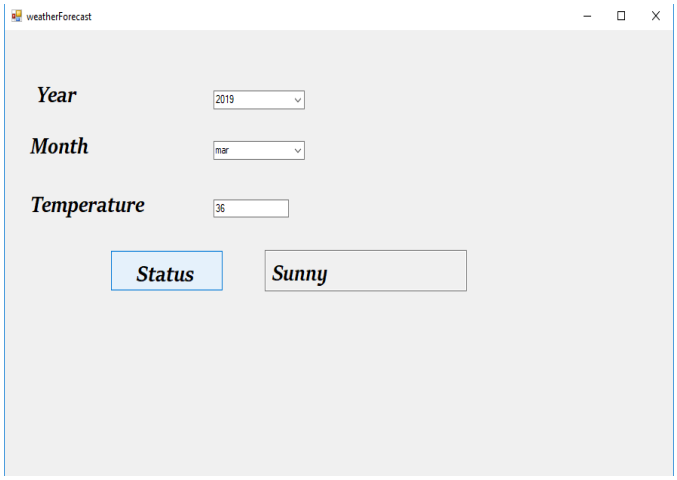

Figure 4. Weather Prediction

V. CONCLUSIONS

The proposed work is expected to discover the answer for exact weather data displaying and prediction utilizing the authentic data. In this way the data mining method is examined for growing such sort of data show. The data mining methods investigate the data of some pre-characterized example and concentrate the noteworthy on the data. Utilizing the separated examples from the data the model takes preparing and arranged for characterizing or foreseeing the comparable examples of related class marks. The characterization of the data bolstered by directed strategy of data mining. The key issue in weather prediction is to achieve the relationship among the class names and the properties which are utilized for foreseeing the weather conditions. In this way utilizing the accessible precise procedures another data display is produced for weather forecasting.

VI. REFERENCES

[1]. A.R.W.M.M.S.C.B. Amarakoon, "Effectiveness of Using Data Mining for Predicting Climate Change in Sri Lanka", 2010.

[2]. Meghali A. Kalyankar, S. J. Alaspurkar, " Data Mining Technique to Analyse the Metrological Data", International Journal of Advanced Research in Computer Science and Software Engineering 3(2), 114-118, February – 2013.

[3]. P.Hemalatha, "Implementation of Data Mining Techniques for Weather Report Guidance for Ships Using Global Positioning System", International Journal Of Computational Engineering Research Vol. 3 Issue. 3 , march 2013.

[4]. 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.

[5]. S.S. De, University of Kolkata, Artificial Neural Network Based Prediction of Max. & Min.Temperature in the Summer-Monsoon month over India, Applied Physics Research, Vol.1, No.2,Nov-2009.

[6]. Kit Yan Chan,'' Neural-Network-Based Models for Short-Term Traffic Flow Forecasting Using a Hybrid Exponential Smoothing and Leven berg–Marquardt Algorithm", IEEE trans on intelligent transportation system, VOL. 13, NO. 2, pp.644-646,JUNE 2012.

[7]. Stephen Dunne and Bidisha Ghosh," Weather Adaptive Traffic Prediction Using Neuro wavelet Models", IEEE trans on intelligent transportation system, VOL. 14, NO. 1, pp.370, MARCH 2013.

[8]. D. Domanskam. Wojtylak, "Fuzzy Weather Forecast In Forecasting Pollution Concentrations".

[9]. Bjarne K. Hansen And Denis Riordan,"Weather Prediction Using Case-based Reasoning And Fuzzy Set .

[10]. Charles A. Doswell Iii And Robert A. Maddox, "The Role Of Diagnosis In Weather Forecasting" 11th Conf.Weather Forecasting And Analysis June 1986