

## Data Mining Techniques in Weather Data Analysis

Jyostnarani Tripathy<sup>1</sup>, Jagannath Ray<sup>2</sup>, Mohini Prasad Mishra<sup>3</sup>

<sup>1,2</sup>Associate Professor, Department of Computer Science Engineering, Gandhi Institute For Technology (GIFT), Bhubaneswar

<sup>3</sup> Assistant Professor, Department of Computer Science Engineering, Gandhi Engineering College, Bhubaneswar

---

**Summary:** Weather analysis has been playing its vital role in meteorology and become one of the most challengeable problems both scientifically and technologically all over the world from the last century. This study carries historical weather data collected locally at Faisalabad city, Pakistan that was analyzed for useful knowledge by applying data mining techniques. Data includes ten years' period [2007-2016]. It had been tried to extract useful practical knowledge of weather data on monthly based historical analysis. Analysis and investigation was done using data mining techniques by examining changing patterns of weather parameters which includes maximum temperature, minimum temperature, wind speed and rainfall. After preprocessing of data and outlier analysis, K-means clustering algorithm and Decision Tree algorithm were applied. Two clusters were generated by using K-means Clustering algorithm with lowest and highest of mean parameters. Whereas in decision tree algorithm, a model was developed for modeling meteorological data and it was used to train an algorithm known as the classifier. 10-fold cross validation used to generate trees. The result obtained with smallest error (33%) was selected on test data set. While for the number of rules generated of the given tree was selected with minimum error of 25%. The results showed that for the given enough set data, these techniques can be used for weather analysis and climate change studies.

**Key words:** Data Mining, K Mean Clustering, Decision Trees, Weather Data Analysis

---

### I. Introduction

In present era weather forecasting and analysis has become a challenging problem around the world from the last century. The reason behind are the two main factors: Firstly, it is useful for many human activities like agriculture sector, tourism and natural disaster prevention. Secondly, due to various technological advances like the growth of computational power and ongoing improvements in measuring systems.

All over the world, major challenges faced by meteorologist are the accuracy of weather analysis and its prediction. On the other hand, researchers had tried to predict different meteorological parameters by utilizing different data mining techniques. While some of these techniques are more precise than others. Over the past few decades the availability of climate data has been increased. Such sources of climate data like observational records, understudy data, etc. makes it more important to find tools with higher accuracy rate to analyze different patterns from massive data. Therefore, meteorological data mining is a form of mining which is concerned with finding hidden patterns inside massive data available. So, the information extracted can be transformed into practical knowledge. This knowledge plays a vital role to understand the climate change and prediction. Having Knowledge of meteorological data is the key for variety of application to perform analysis and prediction of rainfall and it also does good job for prediction of temperature, humidity and irrigation system.

In this research, we have gathered useful knowledge on historical weather data that was collected locally at Faisalabad city. The data comprise ten year of period. While the records obtained include maximum temperature, minimum temperature, wind speed and rainfall observation. After data pre-processing we applied the outlier analysis, clustering algorithm and classification techniques. After utilizing these techniques and algorithm we have represented and described the importance of meteorological field by extracted knowledge.

Data mining objectives is to provide accurate knowledge in the form of useful rules, techniques, visual graphs and models for the weather parameters over the datasets. This knowledge can be used to support the decision-making for various sectors. The goals for data analysis are those which involve weather variations that affect our daily runtime changes in min and max temperature, humidity level, rainfall chances and speed of wind. This knowledge can be utilized to support many important areas which are affected by climate change includes Agriculture, Water Resources, Vegetation and Tourism. Studies shows that human society is affected in different ways by weather affects. For example, water resources are the main sources of irrigation in production of agriculture crops and the amount of rain is one of them that affects the crops abruptly due to climate change. It is also directly related to the different human activities. Moreover, poor growth and low quality is due to negative effects of weather resulting in failure of high production. Hence, changes in weather conditions are risky.

## **II. Literature Review**

Many scholars have made efforts to implement different mining techniques in the areas of meteorological data based on weather data analysis and prediction. Meteorology data mining has been successfully employed in the field of developing important forecasting applications.

M. Viswambari, 2014; surveys the various techniques implemented in data mining to predict weather. Data mining uses various technologies to forecast weather for predict wind pressure, rainfall, humidity, etc. Classification in data mining differentiates the parameters to view the clear information. By looking at the survey provided by Divya Chauhan, 2014; it provides views of different literatures of some algorithms implemented by various researchers to utilize different data mining techniques for Predicting Weather. In this field the work done by different researchers is shown in tabular form where it has been reviewed and compared. Decision tree and k-means clustering algorithm seems to be good at predating weather with higher accuracy than the other techniques of data mining. Some researchers have tried to make the dynamic prediction. The paper of Jyotismita Goswami, 2014; discusses various models for prediction that are applied and compared with their methodologies which are available till now. Their crucial findings marked this study very valuable for a better starting point to generate a new weather prediction model with new description of methodology for predicting weather by using different models of dynamic change in climate.

Sarah N. Kohail, 2011; “tried to extract useful knowledge from daily weather historical data collected locally at Gaza city. All data mining techniques are applied and describe extracted knowledge importance in the meteorological field, used for prediction and decision making”. Zahoor Jan1, 2008; developed a system for prediction weather that utilizes the historical data of an area (rainfall, temperature, wind speed etc.) and applied the algorithm of data mining i.e. “K-Nearest Neighbor (KNN)” to classify this historical data within this specific time span. The “K-Nearest Neighbor (KNN)” then uses these time spans to predict the weather accurately. These experiments demonstrate that the system is generating accurate results inside reasonable time frame for months to come. Meghali A. Kalyankar, 2013; where k-means clustering is implemented for predicting the change in climate of a regional area using historical data of the weather.

## **III. Materials and Methods**

### **Sample Dataset**

In this research article, daily historical weather data for ten years (2007 to 2016) was used in analysis. The data was collected from metrological station located at Faisalabad 33.40 North and 73.80 East. Following procedure was implemented includes, data cleaning, data selection, conversion of data and data mining.

### **Data Cleaning**

At this phase, a reliable data model was setup for handling missing data, finding and removing duplicate data means misleading data. Finally, the procedure of cleaning takes place which successfully converts data into a suitable form for mining.

### **Data Assortment**

At this phase, analysis of relevant data was decided and retrieved from the dataset. Meteorological data set with attributes, their type and description is presented in Table 1, while analysis of the numeric values is also shown in Table 2.

### **Data Conversion**

That is also known as the data association. This is selected form of data into a suitable data mining stage. Save the data files in comma-separated by value (CSV) format of file and data set was standardized to reduce the data scaling.

### **Data Mining Phase**

This phase has divided into the three more stages. At individually stage, the algorithm for analyzing meteorological data sets is implemented. Then test methods are used in this study which is the percentage split of the data set for training, cross validation and testing of the remaining percentage. Subsequently, this recognizes the knowledge representation of interesting patterns.

### **Methodology**

This article was taken different steps, using a different method in each step, with high precision temperature, wind speed and rainfall parameters values of weather data and displays the analytics power of data mining technology point

### A. Data collection

Data collection is a main integral part of implementing mining techniques, for this challenge, a thermometer, barometer, hygrometer, anemometer and data recording systems was used. Data recording system provides weather data to excel in tabular form. "Data record based on a digital processor which is used by the built-in sensor or an external instrument and sensors associated with position data of the time of the electronic device can automatically collect and records data of 24 hours. This was the main and most important benefits of data recorder. It was used to collect weather data from local stations at Faisalabad to a devoted lab PC, then copy the transferred weather data to an Excel spreadsheet and recorded on daily basis along with monthly basis to identify data.

### B. Data pre-processing

The data preprocessing is the next step of data mining after collection of data. Challenges in temperature, rainfall and wind speed data; knowledge discovery process is facing poor data quality. Thus, the data is pre-processed to remove noise and unwanted data. Pretreatment means concentrating the removal of other unwanted variables from the data, while the data preprocessing includes these steps:

1. Data scrubbing: it's the stage where noise and irrelevant data is removed. Data cleaning procedures are implemented to fill out missing values and to eliminate noise in recognizing outliers and to correct data irregularities
2. Data integration: it's recognized as the data conversion; in this stage, the suitable form of data is converted for the procedure of data mining by reduction of data and construction of attributes.

### Discovery of Knowledge

For the acquisition of knowledge, the different data mining techniques are implemented that are Outlier-Analysis, Clustering and Classification in WEKA (Data Mining Software).

### C. Analysis of Result

The values under analysis such as temperature, rainfall and wind speed were analyzed depending on the mining results.

### 3.1 Proposed approach

In this paper, different data mining techniques are applied. Firstly, the K-means clustering algorithm was applied on the given data set which was then altered into appropriate form from unstructured data format after the stage of preprocessing. Secondly, the decision tree (J48 algorithm) was applied. Where 70% of data was taken as training data and remaining 30% was testing data. The model of training methodology was shown in figure 3.

**Figure 3:** Training methodology of the model

## IV. Results and Discussion

There was used the Weka environment that has 4 applications which are Explorer, Knowledge Flow, Experimenter and Simple CLI. The collected data-set was changed in a file of extension ".arff" and loaded in the environment of Weka. First, the attribute reduction was used for data preprocessing then the simple k-means clustering and j48 algorithm were used.

### 4.1 K-means clustering

=== Run information ===

Scheme:weka.clusterers.SimpleKMeans"weka.core.EuclideanDistance -R first-last"

Relation: Weather Data

Instances: 132

Attributes: 6

- Year
- Month
- TMAX
- TMIN
- RAIN
- WIND

## V. Conclusions and Recommendations

In this paper, k-means clustering and decision tree building process were implementation; both are the most common data mining techniques tried to highlight the method that the stored data about past measures can

be used for the future ones. Here, j48 (decision tree algorithm) was tried to create decision-trees & rules for the classification of parameters of weather such as minimum temperature, maximum temperature, precipitation and wind-speed per months and years. Experimental trends about sufficient data over-time was analyzed and the significant deviations was identified that showing the change in climate patterns. Future work can include expanded database with other important weather parameters and include using this weather information in agriculture sector reform with cutting edge technologies.

### **References**

- [1]. A. R. Chaudhari, D. P. Rana, & R. G. Mehta. (2013). Data Mining with Meteorological Data. *International Journal of Advanced Computer Research*, Volume 3(Issue 11), Pages 5.
- [2]. Adeyemo, A. (2013). Soft Computing Techniques for Weather and Climate Change Studies. *African Journal of Computing & ICT*(Volume 6), Pages 14.
- [3]. Agboola A.H., Gabriel A. J., & Aliyu E.O., Alese B.K. (2013). Development of a Fuzzy Logic Based Rainfall Prediction Model. *International Journal of Engineering and Technology*, Volume 3, Pages 9.
- [4]. Badhiye S. S., Wakode B. V., & Chatur P. N. (2012). Analysis of Temperature and Humidity Data for Future value prediction. *International Journal of Computer Science and Information Technologies*, Volume 3, Pages 3.
- [5]. Juraj Bartoka, Ondrej Habalab, Peter Bednarc\*, Martin Gazaka, & Ladislav Hluchýb. (2012). *Data Mining and Integration for Predicting Significant Meteorological Phenomena*. Elsevier, Pages 10.
- [6]. Jyotismita Goswami, & Alok Choudhury. (2014). Dynamic Modeling Technique for Weather Prediction. *International Journal of Computer Science & Engineering Technology*, Volume 5, Pages 8.

Jyostnarani Tripathy "Data Mining Techniques in Weather Data Analysis" *International Journal of Engineering Science Invention (IJESI)*, Vol. 05, No. 11, 2016, PP 122-125.