ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 12, pp. 22128-22131, December, 2017

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

WEATHER DATA ANALYSIS AND PREDICTION USING DATA MINING TOOLS AND JUPYTER NOTEBOOK

Rishabh Taneja*

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai

DOI: http://dx.doi.org/10.24327/ijrsr.2017.0812.1206

ARTICLE INFO

Article History:

Received 06th September, 2017
Received in revised form 14th October, 2017
Accepted 23rd November, 2017
Published online 28th December, 2017

Key Words:

Forecasting of weather, data mining and its techniques, Big Data Analytics.

ABSTRACT

Weather forecasting has been an interesting, quite important and an integral part of the meteorological department. This paper details our analysis of the Weather Database, which contains information such as weather conditions of different regions in the world. We gather a series of relative information from the pre-existing data and draw relationships using various data mining techniques. In particular, we concentrate on factors relevant to the weather conditions of the locations, the relationship between weather conditions, time of the year, maximum temp, minimum temperature, rainfall, and humidity. Then we create a model according to the factors and making use of the database and train the classifiers. We tested these algorithms and compared the results and finally concluded which gave the best results. The paper also reports on the techniques used, giving their implementation and usefulness.

INTRODUCTION

Forecasting of weather is specifically determining or predicting the weather using different algorithms and finding out the pattern from the graph obtained. It is basically an application of various strategies in science and technology to predict the weather condition for the future date. Forecasting has been challenging task for quite a while for the meteorological department because of the unpredictable natural calamities. Weather data is collected using the ground station and aerial data and is found in different formats thus avoiding the possibilities of concurrency or loss of data while converting the format. The collection of quantitative data about the atmosphere and meteorological data is analyzed and thus change in pattern is observed. The disorderly nature of the atmosphere, high level of computational power to solve the problems and errors involved in preprocessing stage leads to inaccuracy in forecasting the weather. Pattern recognition is the most important aspect of weather forecasting. To avoid or lessen this inaccuracy in forecasting, different data mining techniques [1] are used.

For an effective forecasting; it is needed to identify the correlation between the attributes of weather, which indirectly have a role in the weather changes. So, this article proposes an effective weather prediction model by considering various attributes together with their correlations together with data mining techniques.

Also, productively utilizing the restricted measure of water below the effect of the worldwide environmental difference or to creatively give sufficient time for the surge with dry season cautioning, there is a need to look for a propelled demonstrating procedure for enhancing stream-flow anticipating on a transient premise. The information mining model, the crude information is always the principle outcome. Now, we will gather and construct a data-set which contains atmosphere related qualities in Chicago over a past filled with the years. The data-set incorporate credits identified with waterfall, temperature, and so on which will be accumulated by an atmosphere along with water space specialists. In next stage so as to construct an atmosphere and climate classifier, we need to prepare the model by utilizing real information. Subsequent to preparing and assessing the model, it will be utilized for future estimates [5].

Atmosphere demonstrating may incorporate concentrate the accompanying characteristics:

1. Authentic climate records

Corresponding author:* **Rishabh Taneja
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai

2. Day by day precipitation and highest and the lowest temperatures
3. Sprinkling and the parameters of temperature with respect to time, area, and tallness.
4. Temp of the air, relative mugginess, soil dampness, temp of the soil, and so on.

With a specific end goal to apply information mining methods on atmosphere determining, a few preprocessing methods ought to be used to enhance precision and take out exceptions. Cases of a few of the included advances may include:

1. Information ascending is a vital advance after which the models are detailed and created. Every info factors were institutionalized by deducting the mean and isolated by the standard- deviation. This information would produce an arrangement of ordinary irregular factors which signify '0' and on the other hand standard deviation as '1'.
2. Planning regarding contributions for the alignment and also the confirmation of models
3. A few accessible information mining systems can be connected relying upon reasonableness and exactness. Cases incorporate Neural Networks, KNN, Genetic programming, and so forth.

Data Mining

Information mining [2] is considered as science and innovation of investigating information in the request to find undetermined examples. Related improvements that are, hypothetical, and innovative advancements are added to our insight into the air climate conditions. Climate forecast is a critical objective of the barometrical research. Consequently changes in the condition of climate are very dangerous for humans. It influences human culture in possible conceivable way. Climate is forecasted by utilizing the information accumulated by satellite. Different parameters related to climate like the temperature and precipitation is anticipated utilizing a picture taken by satellite to get the patterns. The factors characterizing climate conditions changes persistently with time, forecast model is created either measurably or maybe by utilizing some different means such as a choice tree, counterfeit neural-network systems, relapse, grouping methods of information mining. Climate forecast [3] is a type of information mining which is worried about finding covered up designs inside to great extent accessible meteorological information.

The securing was considered as a standout amongst the most critical stages of information investigation. The information must be gathered physically so the amount was likewise little. So the choices depended on constrained data. Be that as it may, now, gathering information has turned out to be simpler, what's more, putting away it has turned out to be economical. Lamentably, as the measure of data expands, it winds up noticeably harder comprehending it. Information mining involves extensive significance and needs for the availability and wealth of this data in the database. Information Mining can be characterized as the way toward removing valuable data and learning from an expansive measure of unstructured and organized information, which is likewise a compelling method for finding learning. Information-mining is a method related to adapting to the exponential development which includes information and data. Information-mining filter throughout the huge databases looking for an intriguing example and connections among occurrences. In hone, information mining gives many apparatuses by which huge measure of information can be broken down consequently. The procedure of information mining follows the sequence: pre-processing, examination, information trading.

Different information mining methods [4] are Regression, Prediction, Association, Outlier Detection, Clustering and Classification. The expectation finds a connection between free factors and connection amongst reliant and free factors. There are different calculations of arrangement and expectation.

Climate Forecasting

The different strategies utilized as a part of the forecast of climate are:

***Synoptic climate expectation*:** It is the customary approach to climate expectation. Succinctly alludes to the perception of distinctive climate components inside the particular time of perception. So as to monitor the evolving climate, a meteorological focus readies a progression of brief diagrams each day, which frames the exceptionally fundamental of Climate figures. It includes immense gathering and examination of observational information got from a great many climate stations.

***Numeral climate expectation*:** It involves utilization of the energy of the PC to anticipate the climate. Complex PC programs all are kept running on the super-computers which givesforecasts on numerous environmental parameters. One defect is that the conditions utilized are not exact. On the off chance that the underlying state of the climate isn't totally known, the forecast will not be totally exact.

***Statistical climate forecast*:** They are utilized alongside the numerical techniques. It utilizes the previous records related to climate information to remove the redundancy. The primary reason for existing is that to discover the viewpoints of climate which are great markers without bounds occasions. Just the general climate can be anticipated along these lines.

Environmental change [7] is a very critical change in the factual appropriation of climate designs over periods. It might be an adjustment in normal climate conditions or maybe the dispersion of occasions than normal (e.g., more or, then again less outrageous climate occasions). The term is here and there used to allude particularly to environmental change caused by human action, rather than changes in the atmosphere that may have come about as a major aspect of Earth's normal forms. Environmental change today is synonymous with anthropogenic a dangerous atmospheric deviation. Proof for climatic change is taken from an assortment of sources that can be utilized to reproduce past atmospheres.

Steps followed

Collection of data: The information used in this was gathered from Chicago, source-Kaggle. The total time period was of 288 months, that is, February1990 to October 2012. The following strategies were concluded at this phase of the exploration: Cleaning of data, Selection of data, Transformation of data and Mining of Information.

Cleaning of Data: Here, a constant configuration of the information was shown was produced which totally dealt with ambiguous information, finding repeated information, and finally getting rid of awful information. Then, the cleaned information was altered into a design reasonable for mining.

Selection of Data: After the data cleaning stage, it's time for selection of important and relevant data for the next stage that is data transformation. Here appropriate data is chosen and processed. The selected data is selected in such a manner that it has a complete set of information required for best results. The data with missing information is discarded.

Transformation of Data: It is the arrangement where the chosen information is transformed inframes proper mining of the information. The record was spared in Comma Isolated Esteem (CVS) documented form.

Max Temp	Mean Ter	Min Temp	Dew Poin	MeanDew	Min Dew	Max Hum	Mean Hu	Min Hum	Max Sea	Mean Se	Min Sea
7	4	2	5	3	2	100	95	76	1010	1008	1004
7	3	0	6	3	0	100	92	71	1007	1003	997
5	3	2	5	1	-1	100	85	70	1005	999	996
7	3	-1	-2	-3	-4	86	63	49	1012	1010	1005
2	0	-1	2	0	-3	100	95	86	1012	1008	1005
7	3	1	2	-1	-3	100	82	57	1014	1010	1008
2	0	-2	1	-1	-3	100	93	75	1016	1014	1009
8	4	1	7	4	1	100	96	87	1015	1005	1003
12	10	8	8	3	0	100	65	44	1015	1008	1003
13	8	3	8	5	1	93	83	63	1021	1018	1016
16	10	5	10	7	5	100	90	67	1024	1022	1021
17	10	3	8	6	3	100	84	55	1024	1022	1020
15	9	4	9	4	-1	100	75	38	1022	1021	1019
11	6	0	4	2	0	100	81	50	1022	1020	1019
11	8	5	7	5	4	100	83	62	1021	1019	1018
12	9	6	8	7	6	100	89	72	1019	1017	1015
13	9	5	9	5	1	100	78	47	1022	1018	1015
6	3	2	6	4	2	100	98	93	1022	1019	1015
9	7	6	7	6	5	100	97	87	1015	1011	1004
9	7	5	9	7	5	100	100	100	1008	1005	1003
10	8	6	7	6	5	100	89	71	1013	1011	1009
13	8	3	8	5	3	100	85	58	1018	1016	1015
12	10	9	10	9	9	100	96	82	1018	1014	1013
13	10	7	9	8	7	100	90	67	1025	1021	1018
12	9	7	9	7	6	100	89	76	1025	1024	1023

Fig 1 Sample Data Set

Jupyter Notebook

The Jupyter Notebook [6] is very popular and highly used because of it being an open-source web application that enables you to make records that contain account content, conditions, representations and live code. Utilizations include information cleaning and change, numerical recreation, measurable displaying, information perception, machine learning, and significantly more.

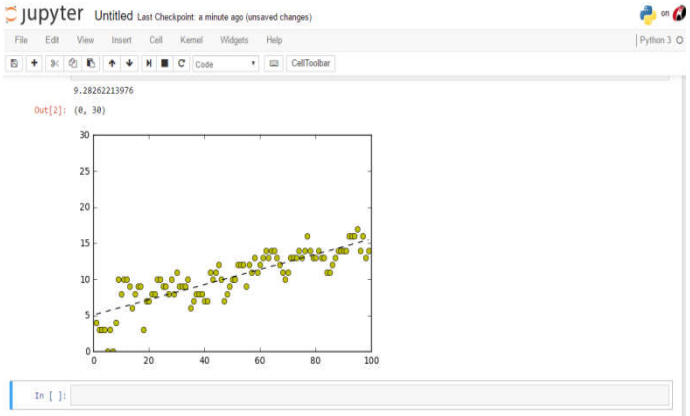

Fig 2 Scatter plot using Jupyter Notebook

RESULTS AND GRAPHS

Various machine learning techniques have been used to plot graphs for weather vs Day using different criteria.

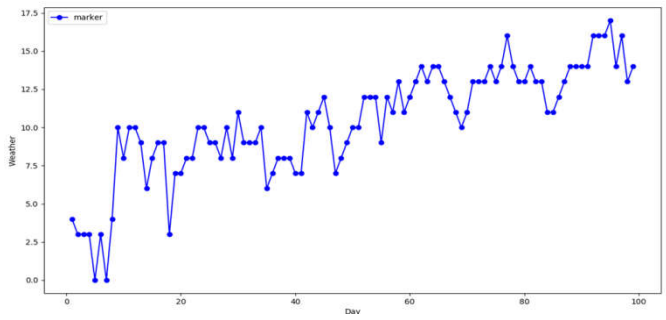

Fig 3 Weather vs Day graph using jupyter notebook

Finally, these graphs are analysed and conclusions are made. Following are the graphs which represent the change in temperature, min. wind speed, humidity, sea level pressure, visibility, cloud cover.

Fig 4 Graphs for all the factors considered.

FUTURE SCOPE AND CONCLUSION

The task can be additionally reached out to provide an all-around definite connection between climate estimating and appalling occasions which will help the client to take careful steps already. This will turn out to be useful in sparing assets and human lives.

This paper discussed various factors involved in weather prediction and performed various machine learning algorithms to find the best and accurate results. It discussed all the necessary factors required to predict the weather. It differed from various other results and was much accurate because of the consideration of only important factors rather than involving every factor which doesn't have any major impact on the outcome. This proposed model can be used for forecasting various other things as well and can be applied to different fields by just applying some minor manipulation in the data and algorithm according to the requirement.

References

1. Bharati M. Ramageri , Data Mining techniques and applications, *Indian Journal of Computer Science and Engineering* Vol. 1 No. 4 301-305
2. Raj Kumar, Dr. Rajesh Verma, Classification Algorithms for Data Mining: A Survey, *International Journal of Innovations in Engineering and Technology* (IJIET),Vol. 1 Issue 2 August 2012,ISSN: 2319 - 1058
3. Ayham Omary, Ayham Omary,Ahmed Zghoul, Ahmad Banihani, and Izzat Alsmadi, An Interactive Predictive

System for Weather Forecasting, IEEE Conference Publications, 2012
4. Data-Mining-Based System for Prediction of Water Chemistry Faults, Andrew Kusiak, and Shital Shah, Ieee Transactions on Industrial Electronics, VOL. 53, NO. 2, APRIL 2006.
5. Yuko Tachibana and Mikihiko Ohnari, Prediction Model of Hourly Water Consumption in Water Purification Plant through Categorical Approach, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics
6. http://jupyter.org, Jupyter Notebook sessions.
7. Krystyna Michałowska; Ewa Głowienka; Beata Hejmanowska2017 Baltic Geodetic Congress (BGC Geomatics), Remote Sensing Methods in the Study of the Impact of Long-Term Process of Sulphur Mining on Environmental Changes of the Carpathian Foreland, IEEE Conference Publications

How to cite this article:

Rishabh Taneja.2017, Weather Data Analysis and Prediction Using Data Mining Tools And Jupyter Notebook. *Int J Recent Sci Res.* 8(12), pp. 22128-2131. DOI: http://dx.doi.org/10.24327/ijrsr.2017.0812.1206

****