FINDING RELATION BETWEEN PARAMETERS OF WEATHER DATA USING LINEAR REGRESSION METHOD

L. Sunitha[1], M. Balraju[2], J. Sasikiran[3], B. Anil Kumar[4]

[1]*JNTU, Hyderabad, Telangana*
[2]*Krishna Murthy Institute of Technology and Engineering, Hyderabad*
[3]*Farah Institute of Technology, Chevella, Telangana*
[4]*Farah Institute of Technology, Chevella, Telangana*

Abstract

Weather data is a form of time series data. Using the data mining techniques, we can estimate the trend of the time series data and some interesting patterns within the weather data. We are collected data from blue planet met club is a weather station at DAV BDL public school bhanur, which is installed by center for Science and Technology secunderabad and devices sponsored by Department of Science & Technology (DST), Government of India, *New Delhi. The data consists of four years' period [2012-2015]. In this paper we are trying to obtain the relation between the temperature, dew point and humidity knowledge from daily data collected. After collecting applying mining technique regression method for finding relationship between the parameters. Weather data is divided into clusters based on similarity and applied regression method within the relationship between the temperature and humidity. We are using data Analysis tool in Excel for regression and very useful and accurate knowledge in a form of visual graphs.*

Keywords: *data mining, linear regression, Data Analysis, parameters.*

----------***----------

1. INTRODUCTION

Atmospheric conditions change constantly. For forecasting we are collected four times a Day. Collecting so much data allows them to track the changes and make a more accurate to forecast. It also helps scientists who study weather data long period of time to better understand climate change. All parameters have to be considered in order to make an accurate prediction. Weather data parameters include temperature, relative humidity, wind direction, Air pressure or atmospheric pressure; wind speed and rain fall alone when trying to make a forecast. We have to understand each measurement and what it could mean to the larger weather picture. When we keep them all together to get an absolute forecast. Following parameters are considered in weather data. All the readings are noted four times a day for every three hours at 8.30 AM, 11.30Am, 2.30PM and 5.50 PM.

1. Temperature: Temperature is a measure of the air's hotness or coldness. Here the temperature recorded in Celsius.
2. Humidity: Amount of water vapor in the air is called Humidity.
3. Dew Point: Dew point saturated point, it is near to low temperature.
4. Wind speed: Wind speed related to differences in air pressure.
5. Atmospheric Pressure: Atmospheric pressure is a weight of air in atmosphere above us.
6. Rain: Rain gauges generally measure the rain fall, it is sometimes reported as inches or centimeters.

2. RELATED WORKS

Data mining methods have been applied successfully to build a very important application in the field of meteorology like predicting abnormal events like hurricanes, storms and river flood prediction [1][5]. Another contribution to detect severe events using data mining is by [4] and [6]. Peters at el. [6] used the volumetric radar data to detect storm events and classify them into various types: hail, heavy rain, tornadoes, and wind. Using data mining in weather application is not limited to prediction, but it also extend to participate in many important fields like water resource management [2] and air pollution management [3]. Mining techniques also can be applied to various types of data like weather images and radar maps extend to characteristic features extracted from this weather images can be used to represent various weather patterns [7].

3. PROPOSED WORK

i. Data Collection: To achieve this study, we use manually recorded the daily observations for four years' period [2011-2015] recorded for weather station at DAV BDL public school. The obtained record includes the daily minimum, maximum and average of humidity (%), temperature (Celsius), dew point, wind speed (KM/H), wind direction, air pressure and rainfall observation.

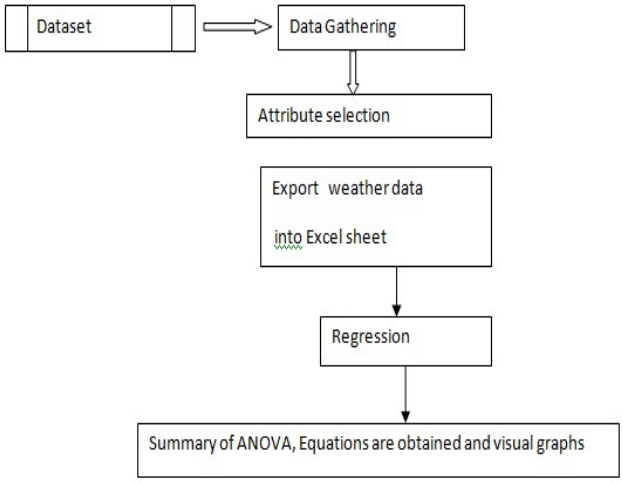

Fig 1. overview of proposed work.

ii. Attribute selection: All the parameter of weather data sets real time data sets, like temperature, humidity, air pressure, wind speed and wind direction. And dew point

iii. Linear regression: Data analysis regression method for finding relation between parameters and results are stored in visual graphs.

4. Regression steps

1. Read n
2. Assign sumx=0, sumy=0, sumsq=0, s sumxy=0
3. For i=1 to n do
4. Read x[i], y[i]
5. sumx =sum +x
6. sumxsq = sumsq +x*x
7. sumy =sumy +y
8. sumxy= sumxy +x*y

End for

9. denom=n*sumsq-sumx*sumx
10. a0 = (sumy *sumxsq –sumx*sumxy)/denom
11. a1 = (n*sumxy-sumx*sumy)/denom
12. write a1, a0

4. EXPERIMENTAL RESULTS

Relative humidity of 100% indicates the dew point is equal to the current temperature and that the air is maximally saturated with water. When the moisture content remains constant and temperature increases, relative humidity decreases. Observed relation in year, next divided data based on similarity into four clusters, cluster1: January, February and march, cluster2: April May and June, cluster3:July,August and September, cluster4:October, November and December.

4.1 Relation between Temperature and Dew Point

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.297382
R Square	0.088436
Adjusted R Square	0.085925
Standard Error	3.304544
Observations	365

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	384.567	384.567	35.21672	6.89E-09
Residual	363	3963.964	10.92001		
Total	364	4348.532			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	27.36992	1.028469	26.61231	2.58E-87	25.34742	29.39243	25.34742	29.39243
dewpoint	0.295158	0.049737	5.934368	6.89E-09	0.197349	0.392967	0.197349	0.392967

Y=a+bx, here y=temp a=27.36992, b=0.295158 , x=dew point

Temp =27.36992+0.295158*dew point

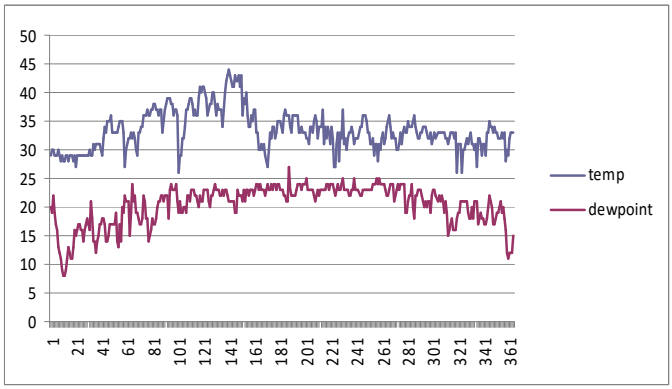

Fig 2. Graph between temp and dew

4.2 Relation between Dew Point and Humidity

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.407686
R Square	0.166208
Adjusted R	0.163911
Standard E	3.184246
Observati	365

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	733.6906	733.6906	72.36018	4.76E-16
Residual	363	3680.611	10.13942		
Total	364	4414.301			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	12.3124	0.963352	12.78079	3.85E-31	10.41795	14.20685	10.41795	14.20685
humidity	0.106507	0.012521	8.506479	4.76E-16	0.081885	0.131129	0.081885	0.131129

Relation between dew point and humidity

Dew point =12.3124 +0.106507*humidity

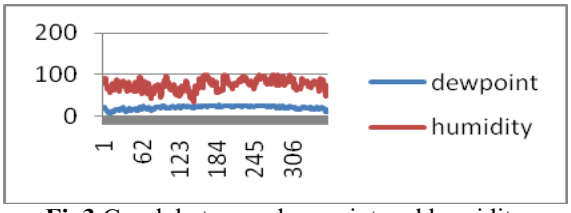

Fig3. Graph between dew point and humidity

4.3 Relation between Temperature and Humidity

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.535344
R Square	0.286593
Adjusted I	0.284628
Standard I	2.923389
Observati	365

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	1246.26	1246.26	145.8262	1.87E-28
Residual	363	3102.271	8.546202		
Total	364	4348.532			

	Coefficient	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	43.90553	0.884433	49.64255	7E-164	42.16627	45.64479	42.16627	45.64479
humidity	-0.13881	0.011495	-12.0759	1.87E-28	-0.16142	-0.11621	-0.16142	-0.11621

Relation between temp humidity

Temp=43.90553-0.13881*humidity

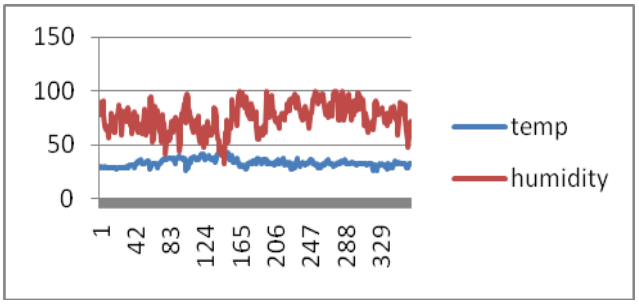

Fig4. Graph between temp humidity

4.4 Relationship between Temperature and Humidity in Cluster Wise

Cluster1: January to March
Temp=41.64 - 0.13792 * humidity

Cluster 2: April to June
Temp=51.90419 - 0.21913 * humidity

Cluster 3: July to September
Temp=43.016 - 0.122 * humidity

Cluster 4: October to December
Temp=43.016 - 0.122 * humidity

5. CONCLUSION AND FUTURE WORK

In this paper we applied knowledge discovery process to extract knowledge from weather dataset. Data collected from weather station we walk through all knowledge discovery process and applied many data mining technique, linear regression. Weather data consists of attributes temperature, pressure, humidity and dew point, wind speed rain fall and wind direction. we are find a relation between various parameter using regression. For data analysis and experimental results, we are used Data Analysis and data mining method regression and graphs from Excel tool provide a very useful knowledge in a form of visual graphs. This knowledge can be obtained helpful to prediction and support the decision making in study of climate change, pollution management, and disaster management sectors. Complete year classified into four clusters based on similarity or seasonal change after that applied regression method over the clustered data. In future we can build a system an automatic and accurate prediction. We find sudden events from Weather data is Dynamic nature rapidly change and historical data over long period.

ACKNOWLEDGEMENT

I would like to express my cordial thanks to Sri. CA. Basha Mohiuddin, Chairman, Smt. Rizwana Begum-Secretary and Sri. Touseef Ahmed-Vice Chairman, Dr.M. Anwarullah, Principal and Dr J. Sasikiran Dean Academics - Farah Institute of Technology, Hyderabad for providing moral support, encouragement and advanced research facilities acknowledge my sincere and profound gratitude to my guide, Dr M. BalRaju for his valuable guidance. And I would like to thank Mr. Sheshadri Naidu Principal and blue planet met club in charge Krystapher of DAV BDL pubic School for providing weather data, invaluable suggestions encouragement

REFERENCES

[1]. Bartok J., Habala O., Bednar P., Gazak M., and Hluch L., "Data mining and integration for predicting significant meteorological phenomena," Procedia Computer Science, p.37 – 46. 2010.
[2]. Jan Z., Abrar M., Bashir S., Mirza A., "Seasonal to Inter-Annual Climate Prediction Using Data Mining KNN Technique," Wireless Networks, Information Processing and Systems Communications in Computer and Information Science, pp. 40-51.2009.
[3]. Li S., and Shue L., "Data mining to aid policy making in air pollution management," Expert Systems with Applications, vol. 27, pp. 331-340, 2004.
[4]. Li X., Plale N., Vijayakumar R., Ramachandran S., Graves H., "Conover. Real-time storm detection and weather forecast activation through data mining and events processing, "To appear Earth Science Informatics, H.A. Babaie, Ed.,Springer. 2008.
[5]. Mohammadi K., Eslami H. R., Kahawita R.,"Parameter estimation of an ARMA model for river flow forecasting using goal programming. "Journal of Hydrology, 331, 293–299. 2006.

[6]. Peters J., Suraj Z., Shan S., Ramanna S., Pedrycz W., Pizzi N., "Classification of meteorological volumetric radar data using rough set methods," Pattern Recognition Letters, pp.911–920. 2003
[7]. Siddiqui, K.J. and Nugen, S.M., "Knowledge based system for weather information processing and forecasting," Geoscience and Remote Sensing Symposium, pp.1099-1101. 27- 31 May 1996.