Pak. j. eng. technol. sci.
Volume 6, No 1, 2016, pg 81-93
ISSN: 2222-9930 print
ISSN: 2224-2333 online

Learning from Experience in Context of Climate Prediction in Pakistan using Data Mining Techniques Avidence from Computer Research n Pakistan

Saadia Karim

Graduate Researcher, saadia.karim@iobm.edu.pk

ABSTRACT

The purpose of this paper is to analyze the climate changes in Pakistan, identify issues related to weather disasters and to revisit weather prediction approaches. The proposed approach is based on different algorithms and their comparisons with reference to past 5years (2010 - 2015) data on 12 attributes. A flow diagram is given that identifies steps included in the process. Results are obtained using WEKA 3.7.13 (latest version 2015). The KNN algorithm and memory-based reasoning algorithm shows the accuracy of predicting weather forecasts. The BPANN algorithm is used to analyze the data set along with KNN and memory-based reasoning algorithms. Decision tree shows the accuracy of predicting weather forecasts. The KNN is used with Bayesian approach in this research. Attributes used in this research shows significant relationship while many of those work as independent variables. Since, for weather prediction these attributes are very important, we used variant factors based on time and date. The KNN algorithm using Bayesian classifier provides accurate results compare with memory-based reasoning of Decision Tree and BPANN trainlm and trainbr.

Keywords:

K-nearest Neighbor (KNN), Memory-Based Reasoning, Data Mining, Hourly time, Weather prediction.

1. INTRODUCTION

The weather prediction is important for the human begin to survive on the Earth. Climate change is the function of knowledge, skill and expertise to predict the condition of the weather conditions for a present and future time in a given location. The atmospheric pressure and changes in the sun's position cause the change in weather like the lowest atmospheric pressure, the average wind speed, the highest wind speed. Numerous scientists have applied ANN, M-KNN, KNN and many Memory-Based Algorithms to predict climate in the world. The data related to weather conditions and climate is collected for certain periods. This means that forecasts become less precise as the dissimilarity in present time and as well as time for which the prediction is being made increases. The Intergovernmental Panel on Climate Change (IPCC) after a systematic analysis on global shell hotness on Earth concluded that it will boost from 2 to 6°F by year 2100, and the associated increase in sea level will be 15 to 95 cm. As per the IPCC evaluation description, the authentication of predicted impacts of climate change is gradually growing. There is also a proof of risk arising from climate change causing less availability of freshwater which is probable to turn down particularly in huge river basins, most glaciers melting, and rainfall unpredictability in northern areas of Pakistan. This means that billions of people in Pakistan will be affected by the 2050s. Human health totally depended on the climate effects in weather and air. The increase of heat waves from sun cause many skin problems, particularly in those areas where the people cannot adapt with warmness of temperature. The climate change effects in Pakistan direct to increases the probable spread of diseases; including dengue, yellow fever, and malaria, by extending the variety of organisms such as insects that bear these diseases in moderate region.

My aim is to determine a system that uses for ensembles and model to minimize the error and pick the most accurate outcome for climate change in Pakistan. Several steps are involved in predicting the appropriate result using KNN (K-Nearest Neighbor) algorithm and memory-based reasoning algorithm: Decision Tree, Bayesian, and Lazy Classifiers. These are:

- facts set(atmospheric pressure, dew point, temperature, wind speed and direction, humidity),
- facts incorporation and investigation,
- arithmetical weather prediction,
- reproduction output

2. PROBLEM STATEMENT

In Pakistan, the weather is very important because we donate less than the overall world donates in the Greenhouse Gas(GHG) [18][19] productions, which results in many negative effects in our climate like; the increased melting of Himalayan glaciers, loss of fresh water, a predictable decrease in crop yields in Pakistan which cause the hunger and food security risk. The collision of climate change make worse the inequalities of resource use and make stronger social factors leading to volatility, disagreements, dislocation of people and transformation in migration patterns. The hourly base prediction of climate in modern developing world is difficult because the hour rate of weather changes is increasing. Hence my aim is to propose an efficient system that can measure the Pakistan's climate using meteorology data attributes and different algorithms and finding the appropriate algorithm that can help in giving less error and maximizing accurate results using KNN (K- Nearest Neighbor) algorithm and memory-based Modelling of system.

3. LITERATURE REVIEW

[1] In this paper the researchers define the two factors why weather forecasting is major scientific and challenging problem. The researcher used Decision Tree and Artificial Neural Network algorithms with the meteorological data composed between years 2000 to 2009 from the city of Ibadan, Nigeria. The complete algorithm comparison is made which developed a predictive Neural Network model for prediction and the goal of this model is based on comparison of the actual weather data on predicted time periods. The consequences demonstrate that by giving massive data we have to use Data Mining techniques for predicting the weather for climate change learning. [2] In this paper the researchers define that the analysis on the climate change and global warming is a very serious and effective challenge. The Global warming and climate change effects are dealt by the scientific community. Such cooperation, based on sharing of knowledge and expertise among the scientific community, is need worthy to tackle the issues related to global warming and climate change. Pakistan, in pursuit of solutions regarding climate change, has also given a serious attention by making an apex body "The Task Force on Climate Change". The said has been established by the Prime Minister, which assembles quarterly and semi-annually to discuss the issues related to climate change and their ramifications. These meeting are chaired by the Prime Minister himself. Its mandate includes organization of climate change based research, assemblage of research groups, identification of the knowledge gaps, and assessment of the capacity and compilation of unified research results for the guidance of planners and policy makers. [3] In this paper the researchers define the collision of seasonal to inter-annual climate forecast on civilization, commerce, farming and all features of human life. In the last few years the achievement ratio of weather prediction has increased. Climate attributes are used to predict the system with Sea Surface Temperature (SST) being a major factor. In order to get complete experimental accurate results, the mathematical and statistical models are used. Researchers developed a system which analyses the historical data of a region with attributes: (rain, wind speed,

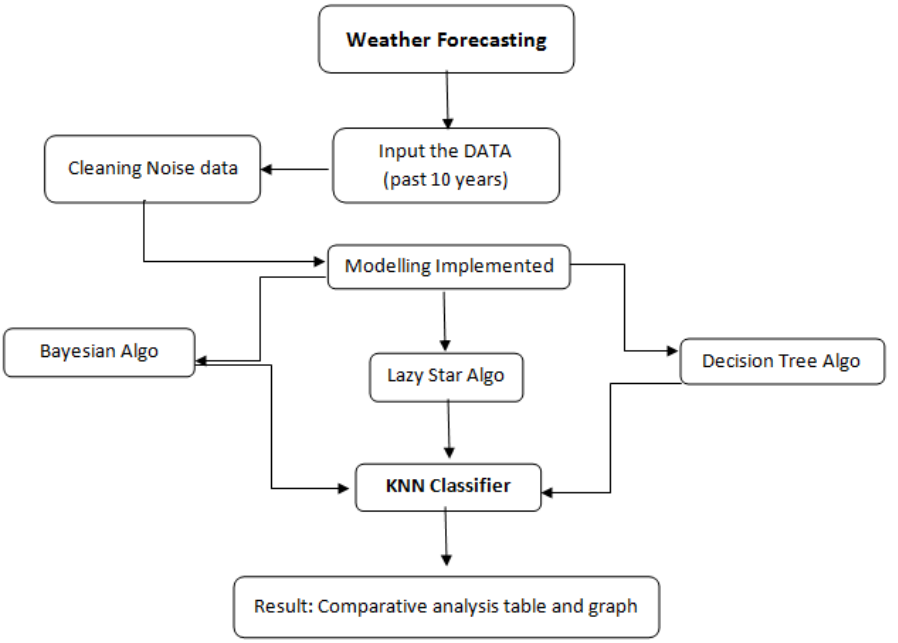
dew point, temperature, etc.), through applying Data-Mining Algorithm "**K-Nearest Neighbour (KNN**)" for categorization of chronological data into a specific time duration. The research shows that the system produces efficient results within the time. [4] Forecasting is to see ahead the incidents that underscore real occasion generation. Weather prediction is a science and technology submission of prediction of atmospheric circumstances. In the last few decades, weather is highly considered and technically demanding difficulty around globe. We come across hearing of different disaster caused by disorder in the weather pattern behaviour. The working of weather forecasting is based on different time periods. Several models convey forecast of weather from genuine time to yearly period. In this paper, we build up a classification that brings out climate prediction using previous state of weather having attribute (Temperature, Date, Hour, Time, Wind Speed **(WS), Dew** Point **(DP),** Wind Direction **(WD),** Relative Humidity, Wind Chill **(WC)** and Standard Pressure **(SP).** A variety of data mining techniques are employed for calculation of weather forecasting including K-Nearest Neighbour, Decision Trees and naive Bayes. Decision Tree has reach fairly promising performance among all the algorithms. [5] In this paper the researchers state the weather to be solitary measure that has an effect on the human beings lives in each aspect. Consequently, forecast of weather occurrence is of most important concern for human civilization to keep away from or reduce the obliteration of climate vulnerability. Weather forecasting is multifaceted due to missing values and noise in dataset. Frequent efforts are made to make prediction as precise as possible, but the complexity of noise affects the accuracy of forecasted result. In this editorial, a complete forecasting procedure that makes use of the customized version of kNN classification is suggested. The key feature of the MkNN is its robustness towards the noise hence producing more accurate outcome. The projected comprehensive technique is proficient to predict 17 special weather attributes. The data was collected from the National Climatic Data Centre, and used in this research for analyses which was carried out by comparing the WP-MkNN and prediction models. The

experimental result shows the technique which is proposed in this research is one of the most accurate in noisy dataset.

4. METHODOLOGY

In my research, the hourly base prediction of climate change in Pakistan using Data Mining classifiers: K-Nearest Neighbour method, Memory-Based Reasoning: Decision Tree, Bayesian, and Lazy Classifiers which Increasing access to high quality information about the impacts of climate change, Improving technological responses by setting in place early warning systems and information systems to enhance disaster preparedness. The data set of at least 5 years (2010 - 2015) of different attribute which impact the climate is taken from metrological department. The algorithm is compared and verified the impact and performance ratio. The Memory-based Reasoning Algorithms are used for comparing between them and KNN algorithm. The tool which is used**:** WEKA 3.7.13 (latest 2015 version). The selected dataset have missing and noisy in it. The collected data convert into consistent and proper format by using supervised Attribute Selection filter in WEKA 3.7.13.

5. BLOCK DIAGRAM:



Then the attributes are selected and the classifiers are implemented like:

Figure 01: Shows the different classifiers used with KNN.

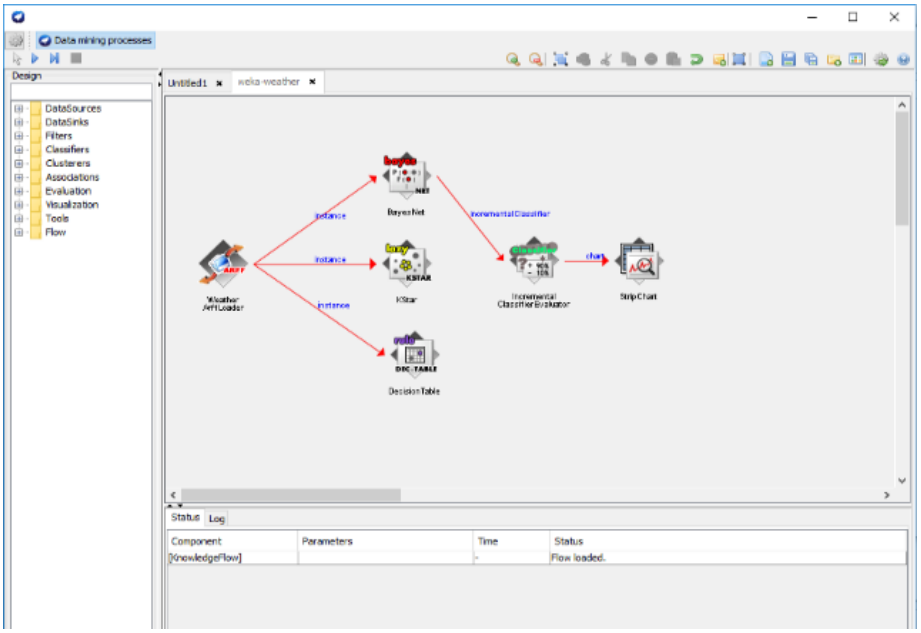


Table 1: Shows 01 Jan, 2014 data and variables with their units.

Month	Date	Time (PKT)	Temp.	Dew Point	Humidity	Pressure	Visibility	Wind Dir	Wind Speed	Precept
January	1	12:25 AM	8.0 °C	-11.0 °C	25%	1022 hPa	6.0 km	NNE	11.1 km/h / 3.1 m/s	N/A
January	1	12:55 AM	8.0 °C	-11.0 °C	25%	1022 hPa	6.0 km	NNE	11.1 km/h / 3.1 m/s	N/A
January	1	1:25 AM	8.0 °C	-11.0 °C	25%	1022 hPa	6.0 km	NNE	9.3 km/h / 2.6 m/s	N/A
January	1	1:55 AM	8.0 °C	-11.0 °C	25%	1021 hPa	6.0 km	NNE	14.8 km/h / 4.1 m/s	N/A
January	1	2:00 AM	10 °C	-7 °C	17%	1022 hPa	4 km	NE	7.4 km/h /	-
January	1	2:55 AM	8.0 °C	-11.0	25%	1021 hPa	6.0 km	NNE	9.3 km/h / 2.6 m/s	N/A

				°C						
January	1	3:25 AM	8.0 °C	-11.0 °C	25%	1021 hPa	6.0 km	NNE	9.3 km/h / 2.6 m/s	N/A

Now, in my proposed model the data is converted into consistent and algorithm the KNN doesn't learn an explicit mapping proper format for data mining purpose. The experiment is performing on over 80000 + records.

Table 2

	KNN	Decision Tree	BPANN – trainlm	BPANN – trainbr
Acc%	83.23%	80.62%	77.2%	81.05%
Sen	0.832	0.826	0.123962	0.862599
FP Rate	0.017	0.826	-	-
MCC	0.996	0.992	585642.46682e-0	169780.71479e-0

Accuracy %, Sensitivity, FP Rate (False Positive Rate), MCC Mathew Correlation Coefficient

6. CONCLUSION / FUTURE WORK

The goal of this paper is develop a method that can perform a robust computation model for the climate change impact in Pakistan as major problems. This paper can figure out the best Algorithm and steps to analyse the data set and give the high performance feature to be taken in advance. The system become a knowledge base system which have past data using as experience to work efficiently and provide more accurate data. The KNN algorithm proves that it is

more efficient algorithm using with Bayesian and lazy classifiers. In this paper we have achieved overall accuracy of 83.23%, sensitivity of 0.8323, FP Rate of 0.017 and MCC of 0.996 and the compare result of Decision tree overall accuracy of 80.62%, sensitivity of 0.826, FP Rate of 0.826 and MCC of 0.992 and BPANN trainlm overall accuracy of 77.2%, sensitivity of 0.123962 and MCC of 585642.46682e-0 and BPANN trainbr overall accuracy of 81.05%, sensitivity of 0.862599 and MCC of 169780.71479e-0.

References

[1] F. Olaiya and A. B. Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," Int. J. Inf. Eng. Electron. Bus., vol. 4, no. 1, pp. 51–59, 2012.

[2] G. Rasul, "An Analysis of Knowledge Gaps in Climate Change Research Profile of the organizations working on Climate Change," Pakistan J. Meteorol., vol. 7, no. 13, pp. 1–9, 2010.

[3] Z. Jan, M. Abrar, S. Bashir, and A. M. Mirza, "Seasonal to Inter-annual Climate Prediction Using Data Mining KNN Technique," Springer-Verlag Berlin Heidelb., pp. 40–51, 2008.

[4] Z. U. Khan and M. Hayat, "Hourly Based Climate Prediction Using Data Mining Techniques by Comprising Entity Demean Algorithm," Middle-East J. Sci. Res., vol. 21, no. 8, pp. 1295–1300, 2014.

[5] M. Abrar, A. Tze, H. Sim, D. Shah, S. Khusro, S. Lecturer, and C. Author, "Weather prediction using classification," vol. 26, no. 5, pp. 2217–2223, 2014.

[6] J. Tucker, M. Daoud, N. Oates, R. Few, D. Conway, S. Mtisi, and S. Matheson, "Social vulnerability in three high-poverty climate change hot spots: What does the climate change literature tell us?," Reg. Environ. Chang., vol. 15, no. 5, pp. 783–800, 2015.

[7] S. Ng, N. E. Basta, and B. J. Cowling, "Association between temperature , humidity and ebolavirus disease outbreaks in Africa , 1976 to 2014," Eurosurveillance, vol. 19, no. 35, pp. 1–11, 2014.

[8] M. G. Lawrence, "The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications," Bull. Am. Meteorol. Soc., vol. 86, no. 2, pp. 225–233, 2005.

[9] G. D. la Torre-Gea, G. M. Soto-Zarazua, R. G. Guevara-Gonzalez, and E. Rico-Garcia, "Bayesian networks for defining relationships among climate factors," Int. J. Phys. Sci., vol. 6, no. 18, pp. 4412–4418, 2011.

[10] M. Hasan, A. Baig, and G. Rasul, "Diagnosis of the impact of deep depressional activity in northern Arabian sea over Karachi during monsoon Abstract: Introduction :," vol. 5, no. 9, 2008.

[11] M. S. Gadiwala and F. Burke, "Climate Change and Precipitation in Pakistan -A Meteorological Prospect," vol. 4, no. 2, pp. 10–15, 2013.

[12] V. Drycap and H. D. M. Dm, "Dewpoint in Compressed Air Frequently asked questions :," pp. 2–4.

[13] V. Castán Broto and H. Bulkeley, "A survey of urban climate change experiments in 100 cities," Glob. Environ. Chang., vol. 23, no. 1, pp. 92–102, 2013.

[14] I. Ahmad, D. Tang, T. Wang, M. Wang, and B. Wagan, "Precipitation Trends over Time Using Mann-Kendall and Spearman's rho Tests in Swat River Basin, Pakistan," Adv. Meteorol., vol. 2015, 2015.

[15] "The Perception of Local Community about the Effects of Climate ....," no. August 2015, 2014.

[16] Z. Yasmeen, "Inversion Layer and its Environmental Impact over Karachi," vol. 7, no. 14, pp. 53–62, 1979.

[17] G. Warming, "Homogeneous climate regions in Pakistan Ijaz Hussain * Juergen Pilz and Gunter Spoeck," vol. 3, pp. 55–66, 2011.

[18] Janjua, P. Z., Samad, G., & Khan, N. U. (2010). Impact of Climate Change on Wheat Production A Case Study of Pakistan, 1–26.

[19] Malik, W., Shahid, H., Zafar, R., Uddin, Z., Wazir, Z., Anwar, Z., … Ali, S. S. (2012). Role of Pakistan in Global Climate Change through Greenhouse Gas Emissions (GHGs). Research Journal of Environmental and Earth Sciences, 4(11), 996–1001.