AN IMPLEMENTATION OF REGRESSION MODEL IN BIG DATA PROCESSING FOR WEATHER STATION DATA

Renukadevi G[1] Selvakumar K[2] Tamilarasan S[3] Venkatakrishnan S[4]

[1] Research Scholar, Department of Computer Science & Engineering, Annamalai University, TN, India.
[2] Professor,Department of Information Technology, Annamalai University, TN, India.
[3] Department of Computer Science, Bharathiar University, TN, India.
[4]Assitant Professor,Department of Computer Science, Annamalai University, TN, India

Abstract
The continuous monitoring of weather is useful to predict the upcoming weather based on the previous data. The Internet of Things is included in weather monitoring to get the data from anywhere in the world. The big data processing is the part of Internet of Things to analyze the weather station data to prepare the report for the future prediction. The regression model is the type of big data processing used to predict the future data based on the previous data. In this paper, the regression model is used for the big data processing to prepare the report as well as to predict the upcoming weather with more accuracy. The data stored in the SQL Server can be retrieved for preparing the report. The sensor DHT 22 is used to measure the Temperature and Humidity and ESP8266 is used to process the sensor data and post the same to the cloud data base.

***Key Words*:** Big data Processing, Regression Model, Weather Station data, REST API, SQL Server

Introduction

In big data processing, there are many algorithms available to process the huge data. The weather station data is used to find the weather conditions as well as prediction. To process these types of data, the regression model is suitable, and the prediction based on the available data can be easily implemented using this regression model.tha data stored in the SQL Server can be retrieved using the authorized access, and process those data using the algorithm developed for the data prediction. The data from weather station is very huge, because it is updated on every 30 seconds.

Proposed Methodology

The DHT 22 is the Digital Temperature and Humidity sensor with higher temperature measurement range and higher operating stability, which is mainly used for military applications. It is the resistive type sensor which is having Thermisor (NTC) and the Piezo Resistor to measure the temperature and humidity in wide ranges. The accuracy of the DHT 22 is more when compared to the DHT11. So, it is most preferred in the important applications. The temperature measurement ranges from -40’c to +125’c and the Humidity measurement ranges from 0 to 100% RH. The stability is high as well as the reliability is also high during measurements. Fig.1. shows the DHT 22 Sensor and its Pin out, Fig.2. shows the Block Diagram of Proposed Model (Stage1) and Fig.3. shows the connection diagram or Interfacing ESP8266 and DHT 22 Sensor with LED Indication. The ESP8266 is used to get the internet connectivity to post the data in the cloud database. The data from cloud canbe stored in the permanent SQL server for further processing.

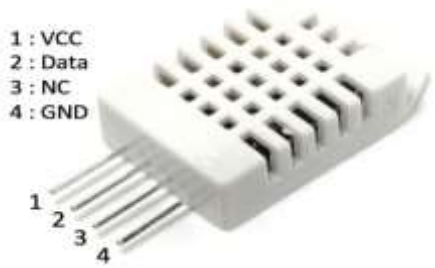

Fig.1. DHT 22 Sensor and its Pin out

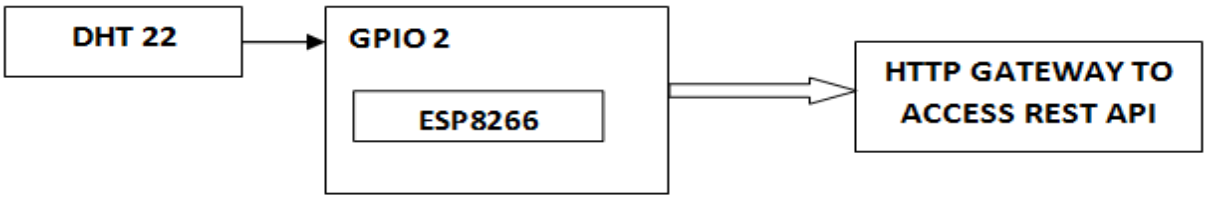

Fig.2. Block Diagram of the Proposed Model (Stage1)

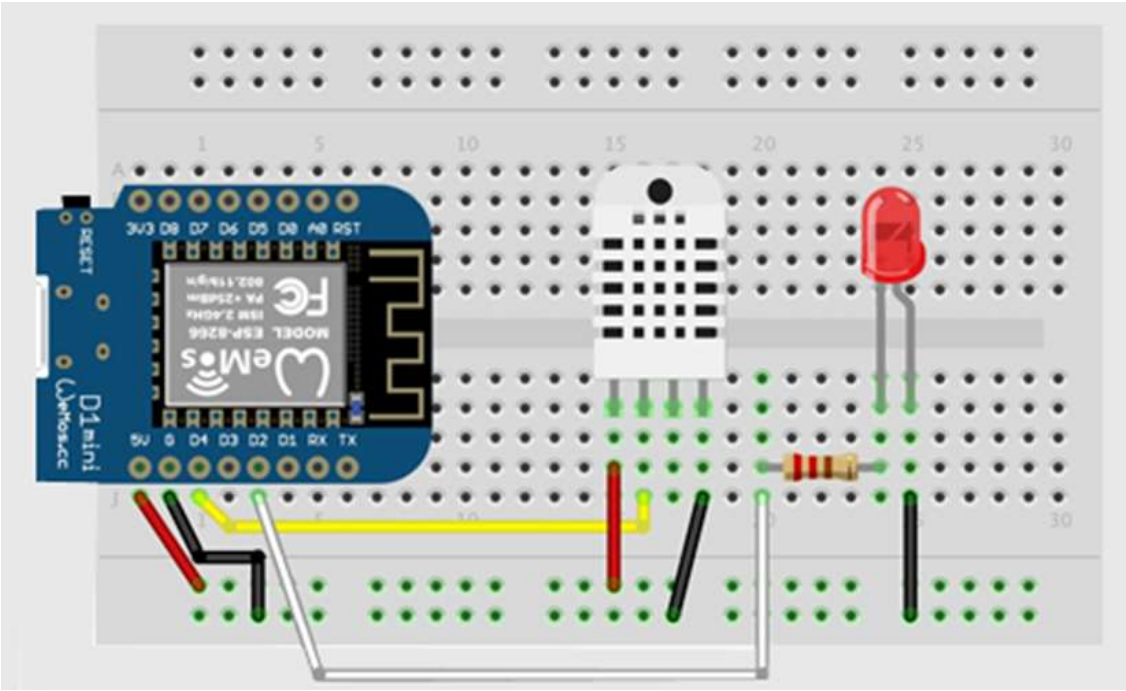

Fig.3.a. Interfacing ESP8266 platform and DHT 22 Sensor with LED Indication

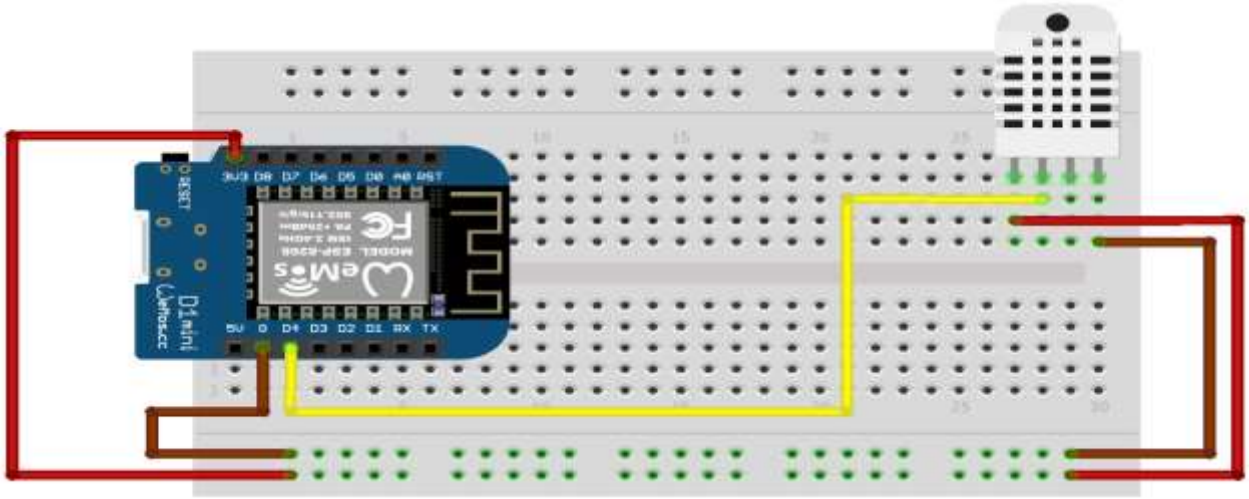

Fig.3.b. Interfacing ESP8266 platform and DHT 22 Sensor without LED Indication

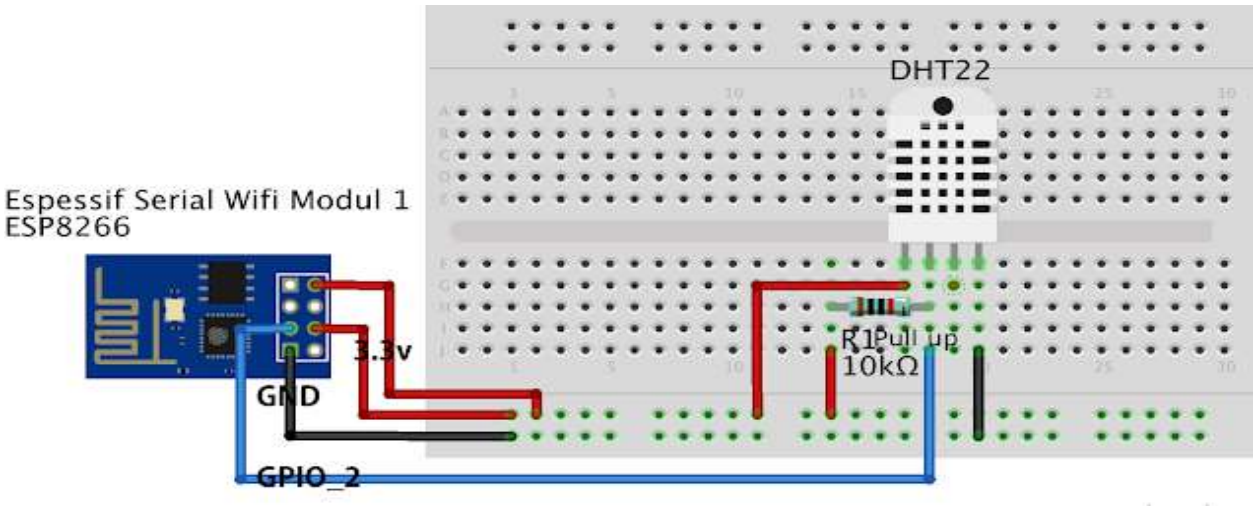

Fig.3.c. Direct Interfacing of ESP8266 and DHT 22 Sensor with Pull up resistor

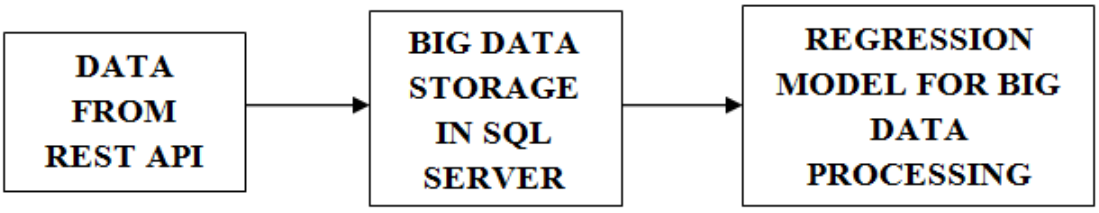

Fig.4. Block Diagram of the Proposed Model (Stage2)

The data stored in SQL server is processed using Regression Model, and the weather is predicted based on the analysis.

Results and Discussions

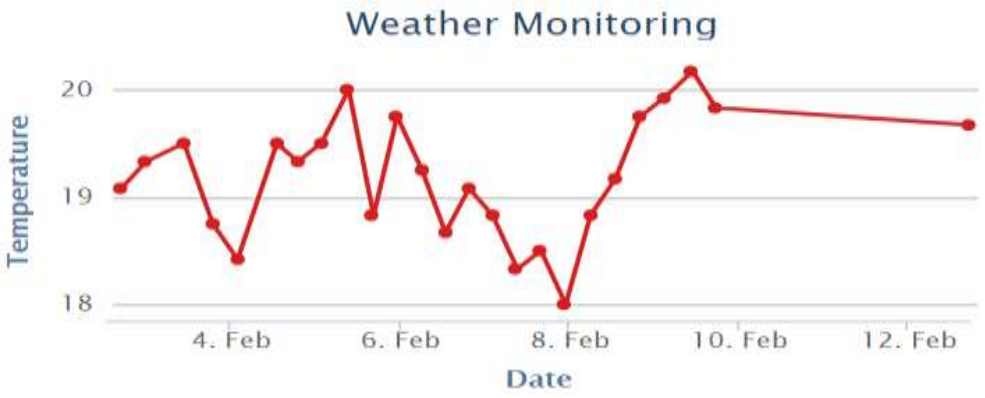

Fig.5. Sample Temperature Readings during Feb, 2020

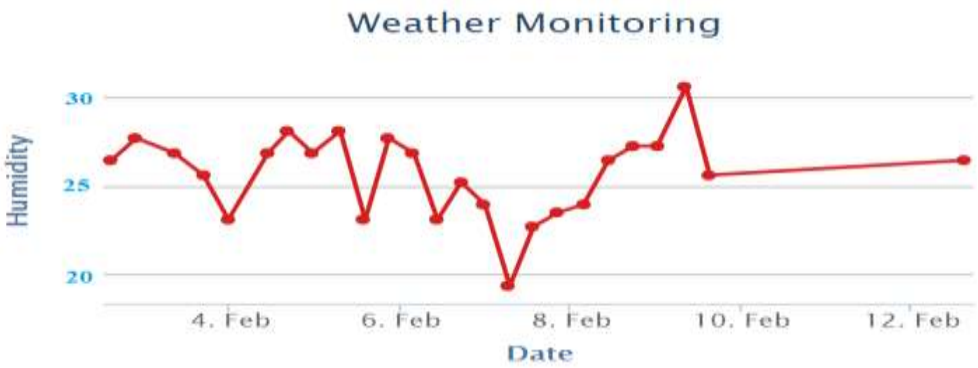

Fig.6. Sample Humidity Readings during Feb, 2020

Table 1.Weather Readings of Feb, 2020 with Status

Date	Temperature(*C)	Humidity(%RH)	Status
3rd Feb 2020	19.2	26.5	Rainy
4th Feb 2020	16.9	23.0	Cloudy
5th Feb 2020	19.5	27.5	Rainy
6th Feb 2020	19.8	26.5	Partly Sunny
7th Feb 2020	18.9	22.8	Cloudy
8th Feb 2020	18.0	24.0	Cloudy
9th Feb 2020	19.9	27.5	Partly Sunny
10th Feb 2020	19.8	26.0	Partly Sunny
11th Feb 2020	19.7	26.5	Cloudy
12th Feb 2020	19.6	26.7	Rainy
13th Feb 2020	19.6	26.8	Rainy
Average	**19.17**	**25.8**	**Partly Cloudy**

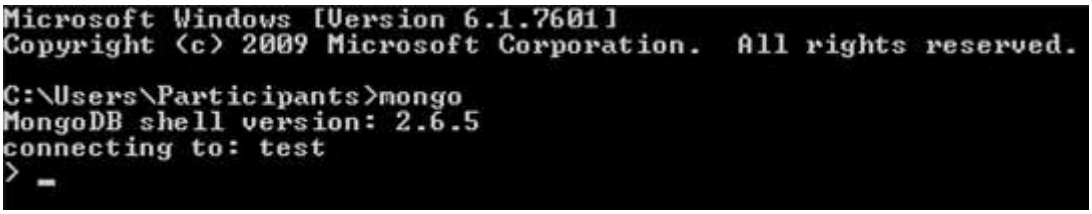

Fig. 7. Connecting to SQL Server from REST API using test command

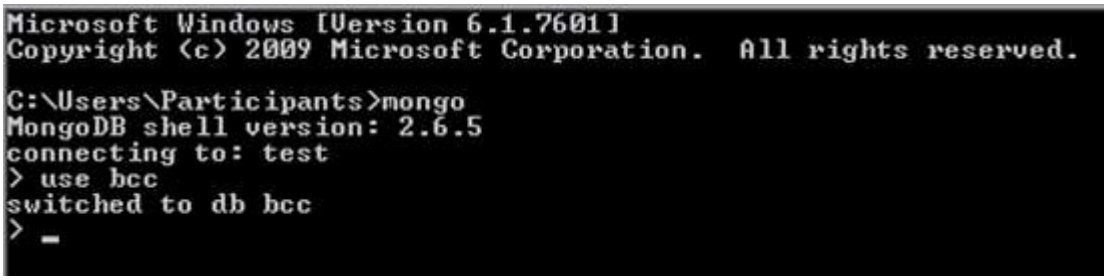

Fig. 8. Switching to SQL Server database

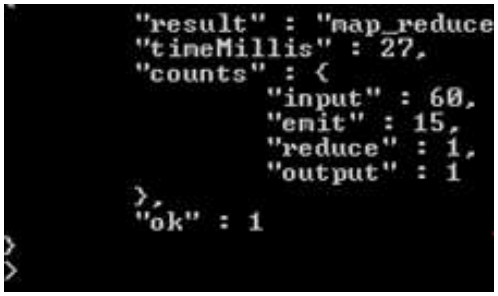

Fig. 9. Consolidated Output of the Regression Model

Conclusions

Thus**,** the regression model is used for the big data processing to prepare the report as well as to predict the upcoming weather with more accuracy. The measured data can be stored permanently in the SQL Server. The data stored in the SQL Server can be retrieved for preparing the report.

References

[1]. Narayanan, S., Samuel, P. & Chacko, M. Improving prediction with enhanced Distributed Memory-based Resilient Dataset Filter. J Big Data 7, 13 (2020).
[2]. Wasserman S et.al, Social network analysis: methods and applications, Cambridge University Press; 1994.
[3]. Saheb, T., et.al,. Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. J Big Data 7, 12 (2020).
[4]. Madden S. Intel Lab Data (2004). http://db.csail.mit.edu/labdata/labdata.html. Accessed 08 May 2019.
[5]. Teh, H.Y., Kempa-Liehr, A.W. & Wang, K.I. Sensor data quality: a systematic review. J Big Data 7, 11 (2020).
[6]. Weng J, et.al, editors. Twitter rank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on Web search and data mining; New York: ACM; 2010.
[7]. Abu-Salih, B., et al. Time-aware domain-based social influence prediction. J Big Data 7, 10 (2020).
[8]. Lai X-a. Segmentation study on enterprise customers based on data mining technology. In: 2009 first international workshop on database technology and applications. New York: IEEE; 2009. p. 247–50.
[9]. Alkhayrat, M., et.al, A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. J Big Data 7, 9 (2020).
[10]. Hsu YF, et.al, Self-aware workload forecasting in data center power prediction. In: 18th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID); 2018.
[11]. Taheri, S., Goudarzi et.al, Learning-based power prediction for geo-distributed Data Centers: weather parameter analysis. J Big Data 7, 8 (2020).
[12]. Abbas A, et.al, Text analytics to support sense-making in social media: a language-action perspective. MIS Q. 2018;42(2):427–64.
[13]. Rizk, A., Elragal, A. Data science: developing theoretical contributions in information systems via text analytics. J Big Data 7, 7 (2020).
[14]. Beck G, et.al, A Distributed and approximated nearest neighbors algorithm for an efficient large scale mean shift clustering. arXiv Prepr arXiv190203833. 2019.
[15]. Bakhthemmat, A., et.al, Decreasing the execution time of reducers by revising clustering based on the futuristic greedy approach. J Big Data 7, 6 (2020).
[16]. Darmawan H, et.al, Anomaly detection based on control-flow pattern of parallel business processes. TELKOMNIKA. 2018;16:2808–15.
[17]. Sarno, R., Sinaga, F. & Sungkono, K.R. Anomaly detection in business processes using process mining and fuzzy association rule learning. J Big Data 7, 5 (2020).
[18]. Li Y, et.al, Understanding and analyzing java reflection. ACM Trans Software Eng Methodol. 2019;28(2):7:1–50.
[19]. Mahapatra, T., Prehofer, C. Graphical Flow-based Spark Programming. J Big Data 7, 4 (2020).
[20]. Livshits B et.al, Reflection analysis for java. In: Proceedings of the third Asian conference on programming languages and systems, APLAS'05. Berlin: Springer; 2005, pp. 139–60.