

A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence

J.V. Anand
Department of Electronics and Communication Engineering,
PACE Institute of Technology and Sciences,
Ongole, India

Abstract: The paper emphasis on an instinctual and efficacious method to forecast and analyze the condition of different atmospheric determinants all over the world. The difficulty in the extant remedies or the apparatus is its incapability to provide a comprehensive information regarding the evaluation of the attributes of the atmosphere. The proposed methodology in the paper gathers the actual information about the atmospheric attributes such as the water, air, the forest and the tree cover etc. from the government bases and processes the collective information. The methodology does the extrication transformation load over the original collective data's that are in its raw format. The converted information sets are imported into the database to develop a dash boards with the multiple information's displayed on it. This allows to have an evaluated data about the various atmospheric factors. To forecast the deteriorations and the conditions of the atmospheric attributes the methodology proffered utilizes the Fuzzy C means clustering, R-studio, and the ARIMA frame work. The dash board assists the NLP enabling the end users to post their queries as input and get back the desired output. The developed deterioration forecasting and evaluation can be used in the evaluation of the conditions of atmospheric attributes for the different countries in the world.

Keywords: Atmospheric Deterioration, Data Mining, Business Intelligence, Evaluation, Forecasting

1. Introduction

The human over population and the over consumption has caused worse impacts over the atmospheric attributes resulting in global warming and the environmental degradation like ocean acidification, mass extinction and loss in biodiversity leading to ecological collapse . The human over population has even caused over exploitation, deforestation and pollution. The ecosystem today has turned to be highly toxicant as the result of the atrocities caused by humans. So there is necessity to constantly improvise the conditions of all the atmospheric attributes by evaluating and fore casting.

So the paper puts forth the method that is reliable, extendable and a proficient business intelligence for examining the conditions of the atmospheric factors such as the water, air, pollution level, and the

particulars of forest etc. for tracking changes in the atmospheric attributes to bring in a positive impact on the factors of the atmosphere. The methodology does the extrication transformation load (E-tl) over the original collective data's that are in its raw format. The converted information sets are imported into the database and the Microsoft power-BI to develop a dash boards with the multiple information's displayed on it. This allows to have an evaluated data about the various atmospheric factors.

To forecast the deteriorations and the conditions of the atmospheric attributes the methodology proffered utilizes the Fuzzy C means clustering, R-studio, and the ARIMA frame work. The dash board assists the NLP enabling the end users to post their queries as input and get back the desired output. Further the developed deterioration forecasting and evaluation can be used in the evaluation of the conditions of atmospheric attributes for the different countries in the world.

The dash board set off engages the business intelligence methods that includes the data mining and ware housing to examine the atmospheric conditions. The remaining paper is organized with the related works in 2, the proposed design in 3. The results and discussion in 4 and conclusion and the future work in 5.

2. Related Works

The evaluation of the air condition prevailing in the atmosphere based on different types of toxicant likes nitrogen dioxide, sulfur-dioxide, carbon-monoxide, ozone or the trioxide. , was conducted by the authors Gore et al [1] the information set was comprised with the "air quality index" values. The work carried out in the paper applies the J48 decision tree and the naive Bayes algorithm to forecast the health of the environment. The distinguishing centered on the "air quality index" is observed under the types good, worse and very worse. The methodology using the decision tree allowed to have 91.9% of accurate forecasting and the accuracy of the navies Bayes was observed to be 86%.

The author Taneja, et al [2] performed the evaluation using the data mining to observe the air pollution level Delhi, and forecasted the future pollution level of the state. The data mining methods such as the MLP and the linear regression were used in the paper to measure the pollutants level along with the particulate matter. Yuheng, S., K et al [3] to obtain the appropriate water conditions employed the "adaptive weighted fusion method" and assessed the quality of the water using the BPNN- back propagation neural network. The method put forth created a software to track and examine the quality of the water, the outcomes of the validation enabled to have a 93.34% accurate assessment of water quality.

Kingsy et al [4] utilized the improved K-means clustering to examine the pollution in the air and estimated the co-efficient of correlation using the tracked pollutant data set. The coefficient estimated, determines the “air quality index” to calculate the level of the pollution in a particular location. Souza et al [5], put forth the “data mining in local urban management to reveal the consequences resulting from the air pollution and its interference with the public health in three cities of the metropolitan region of the Curitiba, Brazil.”

To reveal the impacts of the climatic changes as a result of human over exploitation of resources the Dhore, et al [6] in his paper utilizes the data mining to forecast the weather changes. Yan et al [7] proposed a “extreme machine learning procedure the DSO to determine the quality of the water based on four parameters such as Ph, CODMn, DO, and NH3-N Kaur et al [8] performed a "A Comprehensive Review on Air Pollution Detection Using Data Mining Techniques." Wang et al [9] performed "air quality data analysis and forecasting platform based on big data."

Smys, S., et al [10] has conducted the "Big Data Business Analytics as a Strategic Asset for Health Care Industry." Joseph, S. Iwin Thanakumar et al [11] discusses the uses of the data mining algorithms in the intelligent computing system. Raj, Jennifer S. et al [12] elaborates the “failures of the AI that has lead a trail to the emergence of the computational tools, that leads to the increase of the new regime that is the computational intelligence” J. Vijitha Ananthi et al [13] proffered the automated greenhouse using the sensors to limit emission of significant gases such as the CO₂, CH₄, N₂O into the atmosphere.

3. Proposed Design

From the related works it was known there are no apparatus that provide a comprehensive evaluation and the forecasting of the collective attributes of the atmosphere. So the proposed design develops a proficient and an extendable apparatus to examine the attributes such as the water, air, toxicant level and the tree as well as the forest cover and produces a summary with regard to the data collected. The data collected is significant for creating a forecasting frame work that forecasts the quality of the atmosphere, enabling one to make decision based on the outcomes.

3.1 Architectural Overview

The architecture is segregated into two stages as Unit 1 and Unit 2. The unit 1 examines the huge set of atmospheric information collected from the government sources and delivers the suggestions and as well as forecasts the various atmospheric impacts of various areas. The unit 2 is the extension of the unit 1

incorporating the devices to measure the pollution level and have a real time examination of various atmospheric attributes. The proposed architecture is laid out in the figure.1

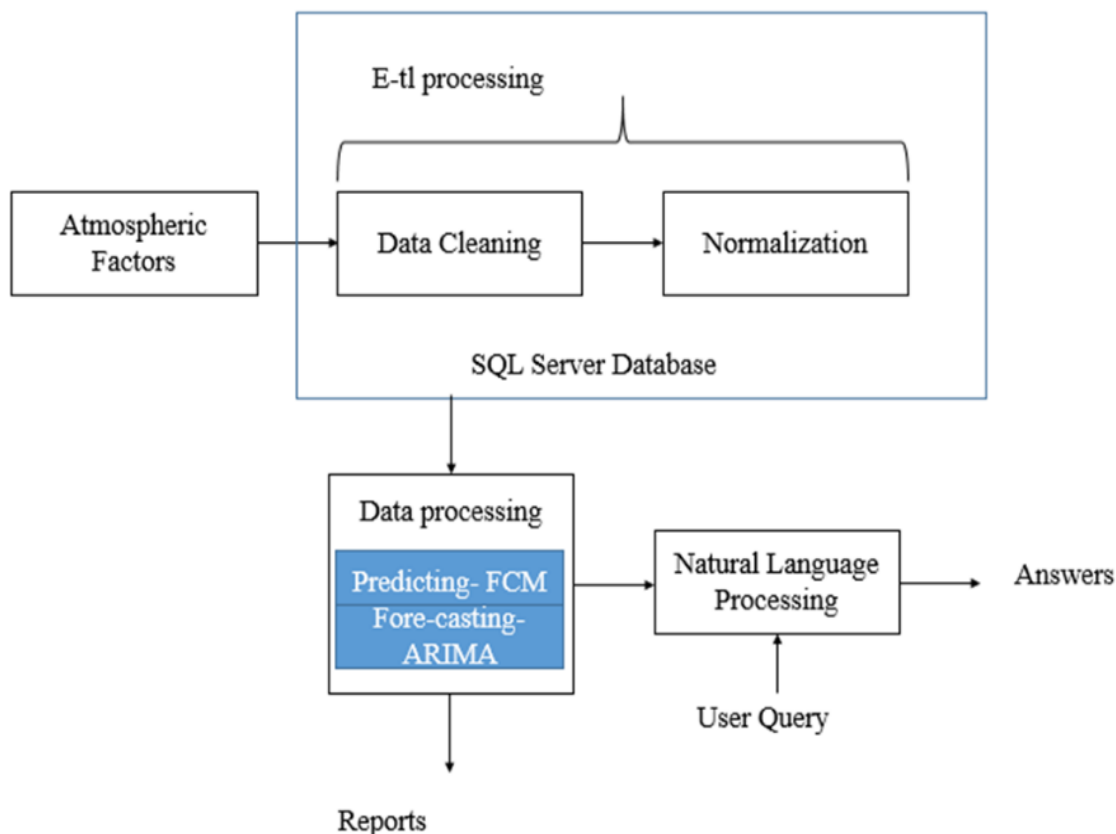


Figure.1 Proposed Architecture.

The actual information set about the condition of the air, water and the particulars of the forest is obtained from the official site and the E-tl is applied to the original information set , the information that is converted analyzed and preserved in the SQL database. The evaluation of the data is done engaging the R-studio and the Fuzzy C Means and forecasts using the ARIMA frame work. The analyzed results are represented in form dash board using the Microsoft Power-BI that assists the NLP, so the queries entered into it in text form are answered with the necessary outputs

3.2 Implementation

The implementation procedure starts with the E-tl process that takes the actual data from the official sites and converts it into meaningful useable data. The atmospheric attributes in the official sites are present in tabulated form. The data collected are mixed to have the physical and the logical model with the normalization done using the “1NF, 2NF, 3NF” to minimize the repetitions. The Microsoft SQL is utilized to prepare and convert the data. The question of the users on the connected data base is computed without difficulties and are easily written. Looking after the entire database is easily done in the “SQL server.” The data prepared and converted are loaded to the business intelligence tool. The atmospheric attributes utilized in the analysis are (i) Water for which the data are gathered from “national water monitoring program” from various tracking stations located across the country. The minimum level and the maximum level and the current level of the Ph, conductivity, dissolved oxygen, BOD, nitrate, fecal coliform and the total coliform are gathered. (ii) Air for which the maximum, minimum and the current level of the five major toxics such as the suspended particulate matter, particulate matter and the Respirable SPM, SO₂ and the NO₂ were collected form the “central board of pollution” likewise the (iii) Forest cover (open forest, very and moderately dense forest and shrubs etc.) (iv) Tree cover were also collected from the forest department. The figure.2 depicts the details in the database

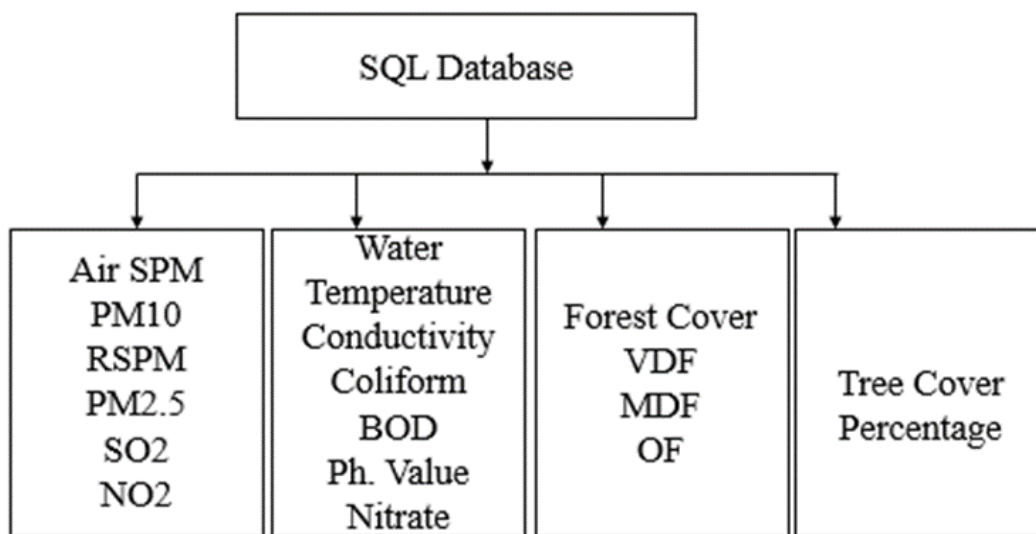


Figure.2 Database Contents

The Microsoft power business intelligence displays the data gathered into a dash carrying multiple information’s about the atmospheric attributes. The power-Business Intelligence is a “cloud based analytics” that provides a view of the analytical particulars of the business that is important. It assists

interaction with regard to the questions posted and displays the results in the dashboard over the Web. The query interaction in the power-BI is done by the NLP that take the input in the text form preprocess it to eliminate the stop words and does the tokenization as well as stemming. The keywords are extricated using the device that takes the critical role of information retrieval. The key words are related to the data and the answers are revealed on the dash board. The data’s closely resembling the key word are segregate and the closely resembling words are grouped to identify the appropriate one employing the fuzzy C means. The complete setup is designed using the R studio an integrated development environment.

4. Results and Discussion

The over view of the outcomes on testing the proposed design with the real world data and the queries is observed by performing two procedures of validation. (i) E-tl testing, (ii) “cross-plat form testing” the procedures in the conveying the information from the stages of extrication to loading may end up in few manual errors that may lead to worse information’s, the tables with multitudes of data are verified to make sure the contents integrity. For this purpose the “addition of values and the combinations of the multiple columns in the table are conducted.” The figure.3 below is the outcome of the proposed design in form of dash board.

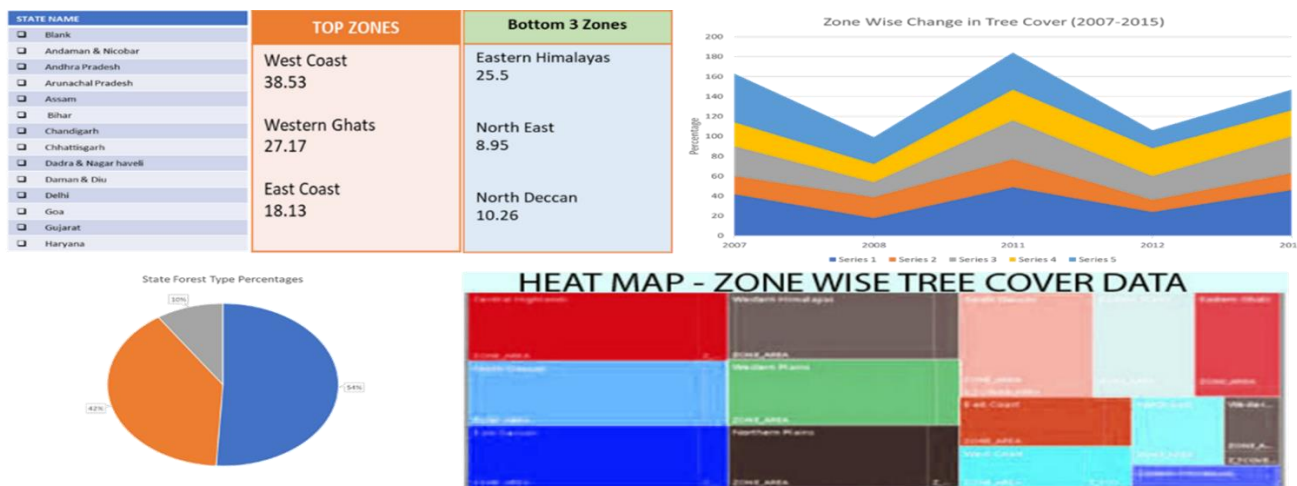


Figure.3 Proposed Outcome in form of Dash Board for Forest Cover

The validating setup was developed using various systems and the dash board was publicized with the assistance of the Power-Business Intelligence in a desktop, with the processor Intel pro, and 16 GB RAM ,

1TB HDD and 128 GB SDD, the swiftness and the genuineness of the dashboard was also checked. Alterations were made to the produced results on the dashboard and then put out. The dash board on the figure .3 was validated with the information gathered.

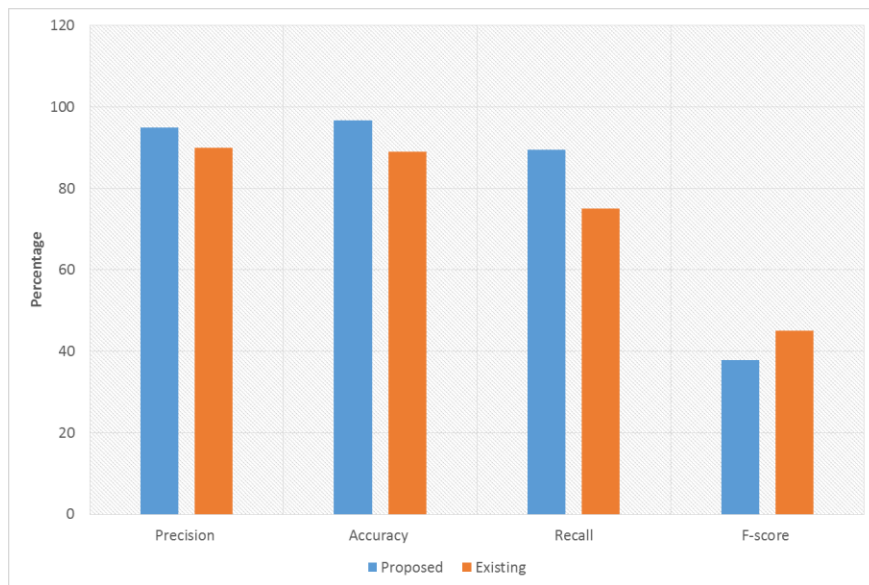


Figure.4 Accuracy, Precision, Recall, F-score

The Figure.5 shows the accuracy, precision, recall, and the F-score of the proposed design enumerated using the TPR, FPR, TNR and FNR observed. The results observed were compared with the existing methods in [1-5] proved the result observed through the proposed method is genuine and accurate.

5. Conclusion

The paper aimed in developing a collective atmospheric deterioration forecasting and evaluation system to provide a comprehensive results that included the complete details of the quality of air, water and as well as the particulars of the forest and the tree cover. The information were gathered from multiple official sites and transformed into meaningful insights and conveyed into the Microsoft Power-BI to display the data in the form dash board with multiple information's. The interaction in the Power-BI was handled using the NLP and the Fuzzy C Means to display the board according to the user's request. This cloud based system has no limitation so it could be extended worldwide. The results based on the accuracy and the precision

and F-score proves the genuiness of the data displayed in the dash board and as well as the authenticity of the system.

References

- [1] Gore, Ranjana Waman, and Deepa S. Deshpande. "An approach for classification of health risks based on air quality levels." In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, pp. 58-61. IEEE, 2017.
- [2] Taneja, Shweta, Nidhi Sharma, Kettun Oberoi, and Yash Navoria. "Predicting trends in air pollution in Delhi using data mining." In *2016 1st India International Conference on Information Processing (IICIP)*, pp. 1-6. IEEE, 2016.
- [3] Yuheng, S., K. Yingchun, and C. Jing. "Ecological water quality assessment system based on BP neural network." In *Information Technology and Mechatronics Engineering Conference (ITOEC), IEEE*. 2017.
- [4] Kingsy, Grace R., R. Manimegalai, Devasena MS Geetha, S. Rajathi, K. Usha, and Baseria N. Raabiathul. "Air pollution analysis using enhanced K-Means clustering algorithm for real time sensor data." In *2016 IEEE Region 10 Conference (TENCON)*, pp. 1945-1949. IEEE, 2016.
- [5] Souza, Fabio T., and William S. Rabelo. "A data mining approach to study the air pollution induced by urban phenomena and the association with respiratory diseases." In *2015 11th International Conference on Natural Computation (ICNC)*, pp. 1045-1050. IEEE, 2015.
- [6] Dhore, Aishwarya, Anagha Byakude, Bhagyashri Sonar, and Mansi Waste. "Weather prediction using the data mining techniques." *Int. Res. J. Eng. Technol.(IRJET)* 4, no. 5 (2017): 2562-2565.
- [7] Yan, Hongmei, Yueming Liu, Xiaohong Han, and Yangyang Shi. "An evaluation model of water quality based on DSA-ELM method." In *2017 16th International Conference on Optical Communications and Networks (ICOON)*, pp. 1-3. IEEE, 2017.
- [8] Kaur, Pawandeep. "A Comprehensive Review on Air Pollution Detection Using Data Mining Techniques." *Training* 6, no. 09 (2019).
- [9] Wang, Jinghan, Jinnan Zhang, XueGuang Yuan, Yu Tang, Hongyu Hao, Yong Zuo, Zebin Tan et al. "Air quality data analysis and forecasting platform based on big data." In *2019 Chinese Automation Congress (CAC)*, pp. 2042-2046. IEEE, 2019.
- [10] Smys, S., and C. V. Joe. "Big Data Business Analytics as a Strategic Asset for Health Care Industry." *Journal of ISMAC* 1, no. 02 (2019): 92-100.
- [11] Joseph, S. Iwin Thanakumar, and Iwin Thanakumar. "Survey of data mining algorithm's for intelligent computing system." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 01 (2019): 14-24.
- [12] Raj, Jennifer S. "A Comprehensive Survey on the Computational Intelligence Techniques and Its Applications." *Journal of ISMAC* 1, no. 03 (2019): 147-159.

- [13] Raj, Jennifer S., and J. Vijitha Ananthi. "Automation Using Iot in Greenhouse Environment." *Journal of Information Technology* 1, no. 01 (2019): 38-47.

Authors Biography

J.V. Anand works as professor in the Department of Electronics and Communication Engineering, at PACE Institute of Technology and Sciences, in Ongole, India. His research area are Big Data Analytics, Industrial System and Collective Intelligence on all fields of Computing and Communication Technologies and mainly focuses on computational aspects of Networks, Cloud Computing, Parallel Computing, Distributed Computing, Service Computing, Software Evolution, Business Process Computing, Internet Computing and Human Computer Interactions