Framework to Study Impact of Environmental Parameters on Human Health Using GIS and Data Mining

Swati Vitkar[1], Dr. Rekha Singhal[2]

[1]Research Scholar, IT Deparment,JJTU

Jhunjhunu, Rajasthan, India

[1]swativitkar@gmail.com

[2]Innovation labs

com TCS Mumbai, India

[2]rekha.singhal@tcs

Abstract—**Health is of major concern among people. There are various factors which could affect human health. In this paper we propose a framework to analyze various environmental parameters like air and water pollutants and to understand their impact on human health. The study is confined to understand the diseases which may be caused by these parameters. The framework makes use of data mining techniques for the analysis of the data collected from different sources. The results are projected in maps using GIS technology. A case study is taken up to demonstrate the working of the framework.**

Keywords— **Air and Water Monitoring System, pollutants, GIS, Data mining, Human Health.**

I. INTRODUCTION

Understanding and finding out the health status of a society /population is an extremely complex concept, even with supportive epidemiological studies. One of the main reasons for this complexity lays in the multi-factorial nature of health related exposure vectors. Some of these factors are nutrition, psychological and social conditions and the overall indoor and outdoor environmental conditions of the population under study. Extensive literature exists on the effects of these factors on human health based on which the exposure response functions have been evaluated.

The above argument implies that promoting health of people or improving their quality of life goes hand-in-hand with their interaction with immediate environment. In turn, improvement of environment (indoor and outdoor) leads to prevention of environment-related diseases and thus good health of a populace. In the long run, good health of a population has a direct effect on the economy and educational status of a society. This brings out the importance of a clean environment with investments in public health and science education as part of good governance.

Under this study is an attempt to assess the quality of environment by monitoring the air and water parameters in specific sites in Navi Mumbai area. The study aims to assess the pollution-related health impairment / risks among residents in and around the study sites through personal interviews and survey.

These health problems are mainly in the realm of respiratory, gastro-intestinal. Apart from this vector borne diseases could partly be attributed to the unhygienic environment with stagnant water pools and open dumping of garbage in public localities that promotes the proliferation of mosquitoes, rodents and stray dogs.

Data mining techniques were implemented to develop a Spatial Information and Knowledge Database on the water quality of 7 zones of Navi Mumbai. The analysis of results drawn at various stages of work was to be projected in maps. For this, Geographical Information System (GIS) was found to be the most appropriate tool for projecting the spatial distribution of various air and water quality parameters on the maps. In this way, the water quality problems could be identified and could be correlated with the activities taking place to interpret the reasons for deterioration of its quality.

This study will help society and sensitize people about their surrounding environment. Further it will help in implementing the data mining for assessment of environmental impact of pollutants. It will also help in predicting the future trend of data as an outcome depending on the previous assessment details.

The framework suggested entitles us to design a model that is scalable, flexible, cost effective, and secure. It will work successfully for similar data or can be modified as per the requirements. It aims to suggest that data mining technique is the optimal solution to design and develop a user friendly interface that will be easy to access and will provide the necessary functionality. The survey and analysis of pollution pattern on human beings using GIS can become further useful for achieving sustainable management of these water resources and air pollutants which is the main aim of this project.

A) Beneficiaries:
- Health Practitioners
- Policy makers (CIDCO,NMMC)
- Socialists
- NGO's working for environment
- Any environment conscious individual

B) Outline of the document:

Section I of this paper covers introduction. Section II provides details of the literature reviewed for the purpose of the study. Section III details the Research Methodology adopted for this project. Section IV describes the health analysis framework. In Section V the Case study is discussed followed by the conclusion and limitations of this case study are discussed in section VI and VII of the paper respectively.

II. LITERATURE REVIEW

Health is an integrated state of well-being and requires the fulfillment of basic needs - such as water and nutrition – and therefore, the ideal set up of a health system would be funded and staffed by the State and be linked to a large array of community based organized groups and NGOs. ARAVALI(Association for Rural advancement through Voluntary action and Local Involvement), is an NGO working for rural health development, began formal operations in 1997, the proprietary report published by ARVALI was undertaken for our study. The report reveals that ARAVALI realized that there is a need to build their perspective on health from boarder livelihood framework (where the health expenditure seems to be maximum from a family). They also felt a necessity to enter into the field of health very concretely as the NRHM framework provides a platform for partnerships to broaden. Some of their initiatives include assessment of NGOs for Rajasthan State AIDS Control society under the target intervention project on HIV/AIDS, facilitating discussions on health in the GO-NGO-PRI forums in Dholpur, Baran and Jhalawar districts of Rajasthan, providing institutional and to some extent technical trainings to NGOs who are working on health issues on various components of maternal and child health. This NGO is working for human health but not for air and water pollutants and its effects on human health.

Reference [22] strongly advocates to strengthen, integrated primary health care systems, and rebuild efficient ones where they are very weak or inexistent. Clearly, national macroeconomic mechanisms, which bring together responsibility for health and finance, have a responsibility to undertake. The task is bringing health back on top of the agenda, nationally and globally.

Reference [18] explains mobile communication and it is suggested to improve the hospital services to the patients and their relatives. They have integrated mobile communication to Hospital Information System (HIS) to automate several of health services. In this paper they discussed several applications of mobile HIS (m-HIS) and proposed architecture to in build mobile communication in HIS at a hospital. They have also presented an emulation of prototype development to achieve the same using open source tools.

Reference [19] explains that by using GPRS enabled mobile phone and based on his location the nearest health center will be shown on the map.

NMMC Environment Status Reports from the year 2008 were studied; the whole data about Navi Mumbai with map and Air, Water, Waste Management, data is given. Also population, Health victims only suffering with malaria are published. It is found that its just a compiled information at one place.

Health And environment Action Based Learning published by HBCSE, Mankhurd with joint venture of NSS cell of University of Mumbai was studied. Air, Water, Waste, Green cover Data is collected in scientific way and further it is analysed in chemical labs. Physical properties are also studied and proper conclusions are drawn. Also health surveys were conducted to correlate health with pollution and conclusion is drawn. The remedial measures are also suggested. It is found that such a huge data should be managed somewhere to keep it up-to-date and data should be retrieved fast and queries should be replied.

The growing volume of data usually creates an interesting challenge for the need of data analysis tool that discover regularities in these data. Data mining has emerged as disciplines that contribute tools for data analysis, discovery of hidden knowledge's and autonomous decision making in many application domains. So they have incorporated the data mining in achieving attention for our basic necessitates such as water and air. Data mining is extracting patterns from data, it is becoming an increasingly important tool transform these data into information that is commonly applied in the wide range of profiling practices, such as banking, marketing, fraud detection and discovery as in [1]. The field of data mining is concerned with findings new pattern in large amounts of data. Widely used in business, it has scarce application to environment related study. Data mining can be applied to environment for e.g.: to find out which pollutant is likely to make largerdifference. Here we are interested in mining air and water parameters from environment perspective. The goal of their project is to define how to make data possible to mine, to identify which data mining techniques are useful and understand how to discover and present patterns that are in favor for the learners, practitioners concerned for the environment as in [2].

The process of tracking and mining such as air and water parameters in order to enhance the condition of air and water is relatively recent but they are already number of studies doing so and researchers are trying to merge their new ideas. Simple statistics, queries or visualization techniques are useful to give environmentalist an overall view of the current status of

environment. E.g. Environment concerned people use the parameters and various values to control the air and water. Data mining is analytic process design to discover data in search of consistent patterns and/or systematic relationships between variables and then to validate the findings by applying the detected pattern to new subsets of data as in [3].

Many application domains have employed data mining or GIS techniques, but not both, to promote their business.

In health care domain, Mitchell as in [4] described several prototypical uses of data mining, including an expert system able to predict women at high risk of requiring an emergency C-section. Merck-Medco Managed Care, a pharmaceutical insurance and prescription mail-order unit of Merck, used data mining to help uncover less expensive but equally effective drug treatments for certain types of diseases or patients as in [5]. Earlier work by Dr. B.S. Mahajan was carried out through "Health And Environment Action Based Learning" (HEAL) project undertaken by Homi Bhabha Centre for Science Education (HBCSE), a branch of Tata Institute of Fundamental Research (TIFR), Mumbai. They have collected air, water and health data for 5 zones of Navi Mumbai and published a proprietary report. But as they have done this project at environmental perspective but not applied any technology.

Reference [20] presents the use of GIS to integrate and display human air pollution exposure data in the Southern California Children's Health Study (CHS). The CHS is a 12-year prospective epidemiological study designed to improve the understanding of the potential chronic health effects of ambient air pollutants, such as ozone, particulate matter, nitrogen dioxide, and acids. The study was conducted over 5000 students are followed longitudinally from ages 10 to 18 to assess respiratory symptoms and pulmonary function by a research team from University of Southern California's Keck School of Medicine. Extensive analyses are conducted of associations between the student's health outcomes and measured ambient air pollution levels, modeled human air pollution exposure, and other surrogates for exposure such as traffic density. Evaluation of the children's exposure to pollutants emitted by mobile sources is a particularly important component of the study and one for which GIS is essential. This paper presents a discussion of how GIS is being used for this project and the development of a geo database, which integrates the locations of children's residences and schools with modeled and measured pollutant exposure data.

The existing model can cater to handle similar type of data. The primary &secondary data as mentioned below will be collected from various sources. It will then be tabulated, coded and analyzed using various GIS and data mining techniques as per requirement. Based on the analysis, appropriate predictions

and conclusions will be drawn and suggestions will be arrived at.

III. RESEARCH METHODOLOGY

This paper is based on combination of explorative and experimental research. The explorative part involves exploring the possibility of data mining and GIS technology to design a cost effective, scalable, secure, user friendly model which will help to assess environmental (i.e. air and water pollutants) impact on human health. The experimental part involves exploring the possibility of implementing data mining regression analysis tool to achieve the objective. This concept will be applied to design environmental impact assessment model and study the benefits of implementing this system.

Secondary data for Air and water pollution is collected from NMMC environmental labs for last 7 years. Also environmental status reports for past few years were studied to get related data. Health data is collected from NMMC health department and NMMC Environmental Status Report (ESR). Also Maharashtra Pollution Control Board data is also collected to study its trend, and find its impact on human health.
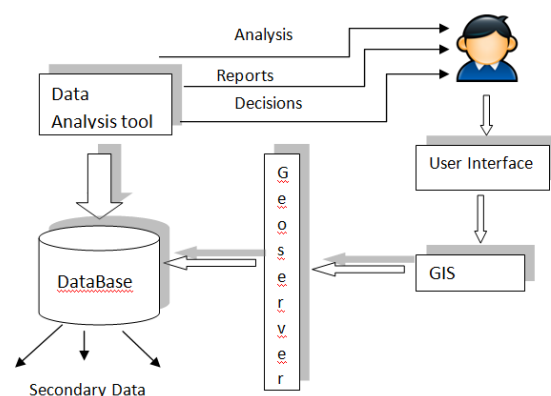
IV. HEALTH ANALYSIS FRAMEWORK

Fig.1: Health Analysis Framework

1)   Data Base: This Database will store the secondary data. Three different databases will be created water pollutants data, Air pollutants data and third database will be created for storing the rules which will help our framework to predict the accurate names of the diseases caused due to air and water pollutants.

2)   Geo Server: GeoServer is a feature rich, standards compliant open source server that will help us to connect our information to the geospatial web.

3)   GIS: GIS extends spatial data capabilities to give the user results in the form of maps, graphs, 3-D analysis. Etc. GIS

will help up giving reports in the form of graph will be easily understandable by the layman also.GIS can also do analyse (buffer analysis, neighborhood analysis) of the data.

Data mining analysis tool can be used to predict the air and water pollution pattern and gives us the forecasted values and also help us to find its effect on human health.Which will further enable local governing body to take preventive measures.

V. CASE STUDY

The present study was confined to NaviMumbai(19^0 2' 2.20" N, 73^0 0' 43.71"E). This city lies across the Thane creek, north east of Mumbai and flanked by the Thane creek waters on its west, south-west and north-west contours.

Navi Mumbai was developed as a planed city by City and Development Corporation of Maharashtra (CIDCO). There are 6 nodes and approximately 30-50 sectors in each node.

In several nodes the city has fresh water lakes (ponds). The Navi Mumbai Municipal Corporation (NMMC) and CIDCO have also impounded creek water creating a series of water holding ponds.

The present case study was confined to 7 different zones of Navi Mumbai which are: Nerul, Vashi, Belapur, Airoli, Ghansoli, Turbhe, Koparkhairne. Each zone with more than two water bodies and air stations. After aollecting the air and water pollutants data from these different zones at uniform interval of time, we applied data mining techniques for fining the missing data values and studied the trend of the data also predicted its future trend. By storing the rules in another database we could able to find out the diseases caused by these air and water pollutants from the same area. Also we are able to find out the upcoming number of health victims due to these reasons.

VI. CONCLUSION

We have deployed GIS and data mining to analyze the data concerning environment, and got a series of results. In the future, we will apply these techniques to analyze the data collected from New Mumbai area to study its pollution pattern and future trend and the impact of this air and water pollution on human health.

To support indicators for environmentally sustainable development, there is a real need for research to enhance the development of technologies, which contribute to the maintenance of environmental quality (water, air, soil). The first step of such a research program consists in collecting and analysing data to provide useful tools for environmental monitoring and forecasting. Such tools would be also helpful for pollution prevention and compliance with environmental laws. Furthermore, if properly managed, they can be applied in

environmental protection, for public information and lower operational costs in industry.

Based on such a principle our research project aims at providing tools for the analysis of these databases to retrieve environmental data. The proposal objective is to develop adaptive methodology and tools to tackle this specific problem, starting from previous work in the fields of artificial intelligence learning theory, statistics, geostatistics and time series analysis.

The scientific method used in this project consists in starting by solving real environmental problems (case studies in water pollution analysis, air quality forecasting) and then to compare results and methodology to generalise.

VII. LIMITATIONS

The lack of timely availability of reliable and authentic environmental data has been a major bottle neck in achieving the full benefits of the said model. The environment being a multi-disciplinary subject, a multitude of agencies is involved in collection of environmental data. However, there is no single organization in India which tracks the data available amongst these agencies and makes it available in one place, in a form and manner required by practitioners in the field of environmental impact assessment in India. Further, the environmental data is not available in value added forms that can enhance the quality of the Environmental Impact Assessment. This in turn adversely affects the time and efforts required for conducting the environmental impact assessments by project proponents and also timely environmental clearances by the regulators.

REFERENCES

[1]   Sk. AlthafHussainBasha Asst. Professor, School of Computing Department ―A Comparative Analysis of Prediction Techniques for Predicting - Graduate Rate of University‖

[2] Educational Data Mining: a Case Study Agathe MERCERON* and Kalina YACEF.

[3] Fang-YieLeu, Tai-ShiangWANG : DATA ANALYSIS USING GIS AND DATA MINING ―

[4] T.M. Mitchell, "Does Machine Learning Really Work?" AI Magazine, vol.18, no.3, 1997, pp.1120.

[5] V. McCarthy, "Strike It Rich, "Datamation, vol.43, no.2, 1997, pp.44-50.

[6] H. Baltazar, "NBA Coaches' Latest Weapon: Data Mining," PC Week, March 2000, pp.69-69.

[7] Dr.B.S.Mahajan,Ms.Suma Nair, "Protocol guide Health and Environment: Action based Learning", HomiBhabha Centre for Science Education

[8] Dr.B.S.Mahajan, Ms. Suma Nair, "Technical Report Health and Environment : Action based Learning", HomiBhabha Centre for Science Education.

[9] Silberschatz, Korth, Sudarshan, "Database System Concepts"

[10] Ramkrishanan,Gehrke, "Database Management Systems" 3rd Edition

[11] Chang," Introduction to GIS" Tata McGraw Hill

[12] I. H.Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2nd edition, 2005.

[13] Data mining techniques and concepts Han and Kamber : 2nd Edition.

[14] G. K. Gupta, "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.

[15] F. Karimipour1, M. R. Delavar1* and M. Kinaie2, "Water Quality Management Using GIS Data Mining**",** Journal of Environmental Informatics 5 (2) 61-72 (2005).

[16] F. Karimipour, M.R. Delavar* and H. Rezayan, "Neighborhood Analysis in Water Pollution Estimation", Environmental Informatics Archives, volume 3 (2005), 232 – 238

[17] Vaniker Y.V, Patel M.N., Kharod K.P., Matieda I.C.,: "Telephone exchange Information and Planning System – A Case Study of Vastrapur Telephone Exchange" GIS@development – July –August 1999. Vol III Issue 4

[18]   Dr. Rekha Singhal**,** Shreya Bokare,"Mobile Health Information System"

[19] Vijayalakshmi R, Conceptual Framework for Ubiquitous Epidemic Control by integration of GPS mobile and GIS

[20]   Tami H. Funk and Frederick W. Lurmann, Sonoma Technology, Inc., Petaluma, CA. "Using GIS to Investigate Children's Exposure to Air Pollution".

[21]   Swati Vitkar, "Conceptual Model for Emergency Medical care

[22]   http://www.who.int/macrohealth/action/en/ngo_paper_sri_lanka.pdf retrieved on 18th Jan 2012 at 19:20 hrs IST.

[23]   Environmental Status Report 2008-09,2009-10,2010-11, NMMC.