International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Analysis and Prediction of Child Mortality in India

Rushikesh Gawande[1], Shreiya Indulkar[2], Himani Keswani[3], Malvika Khatri[4], Pallavi Saindane[5]

[1,2,3,4,5]*Department of Computer Engineering, VES Institute of Technology, Mumbai - India*

Abstract - *Reducing the mortality rate and increasing health awareness is one of the aims of the planners of development and decision makers. For the policies to be made in the correct direction, it is necessary to analyse the mortality rates and the reasons behind the variation. The aim of this study is to extract patterns in rate of child mortality based on socio economic attributes such as: birth rate, fertility rate, mother literacy and undernourishment. Also, detailed study of diseases such as malaria, diarrhoea etc. pertaining in infants and their contribution towards child mortality is performed. The analysis of PHCs available to fulfil the requirements is necessary for maintaining good health and reducing mortality rates. Geospatial analysis of exact locations of PHCs can be further used to formulate need of more PHCs and predicting their locations.*

***Key Words*: child mortality, mortality rates, PHC, machine learning, geospatial analysis**

1. INTRODUCTION

According to Sustainable development goals research, 17 thousand fewer children die each day than in 1990, but more than 5 million children still die before they turn 5 each year. Reducing the mortality rate and increasing health and mortality issues is one of the aims of the planners of development and decision makers. In this way, extracting parameters and their effect on rate of mortality is very important. For ensuring proper enforcement of the policies, it is necessary to obtain frequent and accurate results of the areas. Traditional methods like census and various surveys usually consumes many days or even months to analyse the areas. Thus, we aim at performing exhaustive analysis and presenting our findings in form of graphical representation along with designing a predictive model for the same.

2. LITERATURE SURVEY

The study by AK Singha, et al [1] shows that infant and child survival depend on a host of socioeconomic, environmental, and contextual factors. The distribution of infant and child mortality and their determinants vary across genders, socioeconomic groups, and geographical regions. The author A Singh et al [2] states that poverty and malnutrition exacerbate the risk of infants and children to various infectious diseases like diarrhoea and pneumonia, and heighten the probability of death, particularly among children with low birth weight. The studies have documented that poor economic status of household, low female literacy, poor nutritional status of mother, young age at marriage of mother, large family size, low autonomy of women, and inadequate access to health care services typically lead to disproportionately higher risk for the health status of mothers and their children. According to S Khare et al [3] malnutrition is one of the global health problems especially in the area of child survival. In developing countries, malnutrition is one big problem which is directly or indirectly responsible for half of all deaths worldwide among children under the age of five. According to G Toscano et al [4] parental education, income per capita and health service indicators are the three most important determinants of child mortality. Author V Suriyakala et al [5] concluded that socio-economic factors like fertility rate, national income, women in labour force, expenditure on healthcare and female literacy rates influence the infant mortality rates. The study by C Lahariya, et al [6] states that Neonatal conditions (33%), pneumonia (22%), diarrhea(14%) are the leading causes of infant deaths in India. It addresses the issue regarding neonatal infant mortality that is due to inadequate access to basic medical care during pregnancy and after delivery which is a major cause of infant mortality in India and offered an approach for using data mining in classifying mortality rate related to accidents in children under five.

3. DRAWBACK OF EXISTING SYSTEM

Elimination of child mortality has been the concern since many decades and greatest in countries like India. There are many organizations to help reduce child mortality and many policies implemented to help the nation overcome child mortality, but one of the main problems is finding the impact of various attributes on mortality. The variation of child mortality with respect to different areas is not provided so major actions can be taken to improve the living environment. Also, there is no exhaustive study of diseases which are prevalent in children under the age of five.

4. METHODOLOGY

Data collection and data set preparation: The data collected was based on child mortality according to various attributes.

Developing algorithms: We applied ML Algorithms for analysing mortality rate over areas in a period of time and various affecting attributes. A machine learning model was developed for mortality rate prediction based on various attributes. Algorithms used are: Linear Regression, Random Forest, and Decision Tree.

Training and experimentation on datasets: Training prediction model using ML algorithm. The model was trained on training set and the results were verified.

Analysis on real life scenario: The model was tested for accuracy. The model was then tested on testing dataset and the results were verified using data available.

5. HARDWARE & SOFTWARE REQUIREMENTS

Hardware Requirements:
PC with following specifications-
RAM: 4GB or more
Hard Disk: 1TB approx.
Processor: Intel Core i5 or higher
Operating System: Windows
Software Requirements:
Python based Computer Vision and Machine Learning libraries are exploited for the development and experimentation of the project. Tools such as Anaconda with Spyder IDE are utilized for this process. Power BI, a Microsoft software is used for geospatial analysis.

6. ANALYSIS AND RESULTS

6.1 Analysis of Mortality

We have studied the trend of the different variables affecting mortality and presented it graphically. The analysis led to the conclusion of the variation in the following factors which affects the mortality:

1. Birth Rate
2. Below Poverty Line Index
3. Fertility Rate
4. Mother Literacy
5. Female Primary Education
6. Undernourishment

6.2 Analysis of Diseases affecting Mortality

The data about neonatal and post-neonatal deaths occurred due to malaria and diarrhoea was studied along with the total child mortality in the same year. We concluded from the analysis that diarrhoea, malaria and pneumonia contribute the most to child mortality. There has been steep decrease in number of deaths due to these diseases. The malaria incidence rate has fallen by an estimated 37 per cent and the mortality rates by 58 per cent. Yet the present number is too significant.

6.3 Geospatial Analysis

We did the gathering, display, and manipulation of spatial and historical data. Described it explicitly in terms of geographic coordinates or locations identifier as they are applied to geographic models. We filtered out relevant from irrelevant data and applied it to conceptualize and visualize the order hidden within the apparent disorder of geographically sorted data. Doing so allowed us to provide more accurate trend analysis, modelling and predictions. The average number of PHC and sub centres in each state are represented and is cross referenced with the average mortality in that particular region. The actual locations for all the existing PHCs for the state of Maharashtra are pointed. The child mortality in those geographical regions is also mapped alongside.

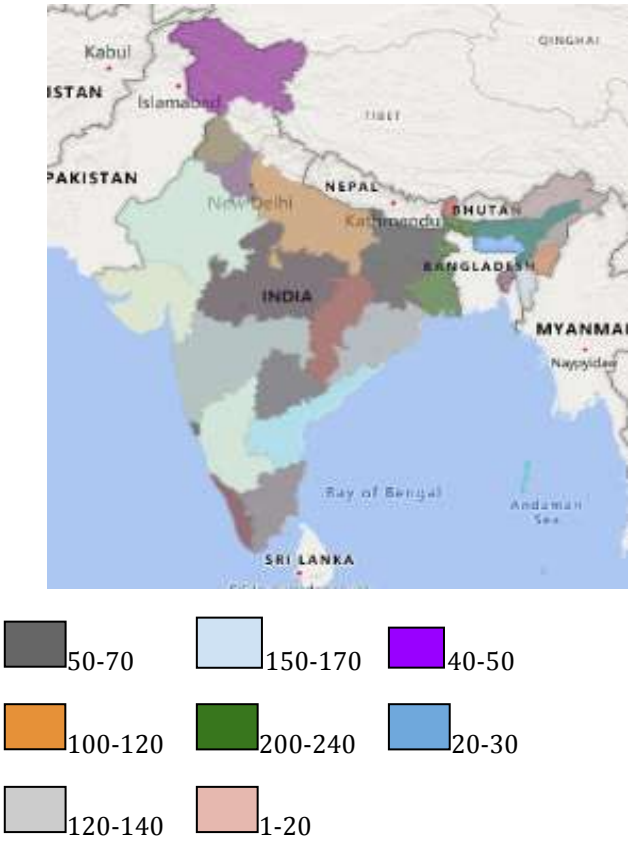

Fig 1 - The average number of sub centres along with their PHCs for all geographic regions of India

Fig -2 The actual locations of all the PHC's in the state of Maharashtra are pinpointed.

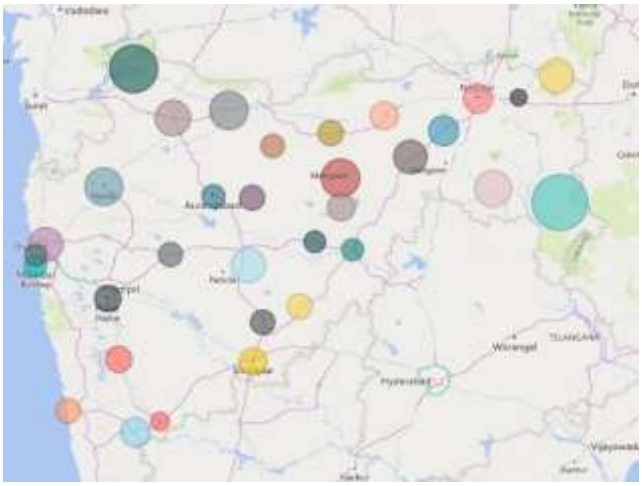

Fig 3 - Indicators for the rate of child mortality in various districts of Maharashtra where the size of the indicators is directly proportional to the extent of mortality in that particular district.

6.4 Prediction of Mortality

We trained our model based on parameters using three different algorithms. Of the total data collected, eighty percent was used as training data and the remaining twenty percent as testing data. The algorithms used were:

1) Multiple Linear Regression: We trained the model when we had eight independent variables and a dependent variable i.e. total number of deaths for children under five, by fitting a linear equation to observed data.

2) Random Forest: Random forest is a supervised learning algorithm. It builds multiple decision trees and merges them for a more stable prediction. It helps to measure the relative importance of the features on the prediction.

3) Gradient Boosting: Gradient boosting is a machine learning algorithm which uses regression and classification. It accumulates weak prediction models, mostly decision tree. It builds the model in a particular stage-wise fashion like some other boosting methods do

Table - 1: Comparison of algorithms

Algorithm	Mean Absolute Error	Mean Square Error	RMS Error	Accuracy
Multiple Linear Regression	74.26	8767.06	93.63	95.81%
Random Forest	72.83	8596.72	92.72	96.09%
Gradient Boosting	93.23	9943.48	99.72	95.04%

From Table 1 it is clear that for the considered scenario, random forest yielded the best results as compared to the multiple linear regression and gradient boosting algorithm.

7. CONCLUSION

Our analysis was able to anticipate the rate of mortality based on various attributes. Also, the trends in mortality from 2000 till 2016 is analysed. Even though the mortality has decreased over these years, still the present toll is around 25,00,000. Also, the diseases of diarrhoea, malaria and pneumonia which largely affect the child mortality were studied and their contribution was analysed. The geospatial analysis was performed involving the location of PHCs and the mortality for those regions. This data analysis will be further useful in formulating the ideal number of PHCs required in order to reduce the mortality and ensure good health and facilities. This analysis and prediction can be used by various government healthcare organizations, policy makers and Health Ministry.

8. REFERENCES

[1] AK Singha, D Phukan, S Bhasin and R Santhanam,"Application of Machine Learning in Analysis of Infant Mortality and its factors", http://www.researchgate.net/publication, Published in January 2013 accessed on September 2018 .

[2] A Singh, PK Pathak, RK Chauhan and W Pan, "Infant and Child Mortality in India in the Last Two Decades: A Geospatial Analysis", https://journals.plos.org/plosone, Published in November 2011 accessed on September 2018.

[3] S Khare, S Kavyashree, D Gupta and A Jyotishi, "Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data", http://www.sciencedirect.com/science, Published in August 2017 accessed on October 2018.

[4] G Toscano and G Hossain, "Predicting the Effect of Parental Education and Income on Infant Mortality Through Statistical Learning", https://ieeexplore.ieee.org, Published in May 2018.

[5] V Suriyakala, M.G.Deepika, D Gupta and J Amalendu, "Factors Affecting Infant Mortality Rate In India:An Analysis Of Indian States", https://link.springer.com, Published in September 2018

[6] C Lahariya, VK Paul "Burden, Differentials and Causes of Child Deaths in India" , Published in Indian J Pediatr (2010).

[7] LM Trandafir, OE Frasinariu, C Corciovă, LV Boiculese and M Moscalu, "Prevention of Cardiovascular Risk Factors in Childhood Obesity" The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017.

[8] M Hossein, Z Ehghaghi, H Memarzadeh and B Zibanejad "Applying Data Mining In Medical Data With focus on mortality related to accident in children" IEEE International Multitopic Conference in 2008

[9] G.Richards ,VJ Rayward Smith, PH Sonksen, S Carey and C Weng, "Data mining for indicators of early mortality in a database of clinical records" Elsevier, Artificial intelligence in medicine 22 (2001)

[10] L Liu, HL Johnson, S Cousens, J Perin,S Scott, JE Lawn and I Rudan, "Global, regional, and national causes of child mortality:an updated systematic analysis for 2010 with time trends since 2000" for the Child Health Epidemiology Reference Group.

[11] https://data.unicef.org/resources/resource-type/datasets/

[12] https://knoema.com/atlas/India/topics/Health/Nutrition/Prevalence-of-undernourishment

[13] https://knoema.com/atlas/India/topics/Education/Primary-Education/Female-students-in-primary-education

[14] https://knoema.com/atlas/India/Fertility