

## Regional Health Survey and Prediction Of Epidemics Using K-Nearest Neighbour Algorithm

Saleekha. K

Dept. of Computer Science, Farook College, Calicut University, City, India

e-mail: saleekhamammed@gmail.com

**Abstract**— Predicting the chance of spreading epidemics in a particular area can be done based on the regional health survey data. An android application is used for data collection which can be installed in the ASHA worker's Smart phone predict the chance of spreading epidemics in a particular area based on the regional health survey data collected by ASHA workers under a particular PHC. An android application is used for data collection which can be installed in the ASHA worker's Smart phone. Using these application, they can collect data from each home. The collecting data would be the factors that may affect the spread of epidemics in that area. Thus, the entries in the application is in such a manner. Using these application, they can collect data from each home. The collecting data would be the factors that may affect the spread of epidemics in that area. K-Nearest Neighbour is executed on the dataset to get the chance of spreading epidemic.

**Keywords**—Epidemic, Health centre, KNN

### I. INTRODUCTION

Many research have been conducted to predict the disease outbreaks in a region using different machine learning algorithm. This research is focused on how a data mining algorithm can efficiently implemented for the prediction of epidemics. Data collection is done through health survey under the control of PHC (Primary Health Center) s in that location, which are held by ASHA (Accredited Social Health Activist) workers.

Data mining algorithm like Bayesian classification, KNN, Artificial Neural Network algorithm, Decision tree are studied by the many scholars for prediction of diseases like diabetes, heart disease, etc.... In this research by collecting health information from each home of particular area by ASHA workers through an Android application and data is stored in the server located in the health Centre. The Administrator perform the chance spreading epidemics in an area using K-Nearest Neighbor algorithm.

Since many algorithms like Decision Tree gives the result as YES or NO, KNN is more suitable for prediction and implementation. Because it is based on the similarity of the features. Even though no occurrence of exact match in the data set may can produce the result.

Rest of the paper is organized as follows, Section I contains the introduction of the research work, Section II contain the previous research works that have been conducted in the related areas, Section III methodology with Data Flow Diagram, Section IV describes results and discussion,

Section V contain the conclusion and future scope of the research.

### II. RELATED WORK

Pi Guo et al published a research article on Developing Dengue forecast model using machine learning, a case study in China. Here, a dengue search index was constructed for developing the predictive models. SVR (Support Vector Regression method) is used as candidate models to predict dengue incidence [1].

Danqing Xu et al conducted a study on Predicting Epidemic Tendency through Search Behavior Analysis. Here, with web user behavior data, considers the problem of predicting the tendency of HFMD (Hand-Foot and Mouth Disease). During winter and summer CDC reports the total number of HFMD infected cases every week to monitor the HFMD occurrences. These cases are collected weekly from clinical laboratories and hospitals for a specific period. Anonymous search logs were gathered for same period. Then developed an online prediction system using SVR (support Vector Regression) method [2].

Sidharth Satish and Smitha.G. R have conducted a study on Epidemic Disease detection and forecasting. In this study they developed a technique to collect large amounts of symptomatic data from various sources and model it. Then developed generalized epidemic prediction approach based on current activity of epidemic and finally attempt to forecast

future activity and spread patterns of the outbreak based on the pattern viewed from the Bayesian model [3].

Anand A. chaudari and S P Akarte conducted a research on Fuzzy and Data mining-based Disease prediction using K-NN algorithm. This paper investigates applying K-NN to help in the diagnosis of disease especially heart disease. Here first locate the training dataset. The Euclidian distance measure is used to calculate how close each member of the training set. Here the research is focused on heart disease prediction and its risk factor using KNN [4].

Hakizimana Leopold et al have conducted a study on Survey and analysis of classification and regression data mining techniques for disease outbreak prediction in datasets. In this paper review and analysis has been conducted on different classification and regression prediction model. They studied the advantages, Disadvantages and applications of classification techniques like K-NN, Decision tree, Support vector machine, neural network and Bayesian. They suggest to use hybrid technologies for classification and prediction [5].

### III. METHODOLOGY

Here an Android application is developed to collect health data in a region. This application is used by the ASHA workers, who are assigned to different region under a health center. The data collected through this application is stored in the server located in the health center, which is managed by the administrator and he/she executes the prediction algorithm.

Data Flow Diagram for the System is given in *Figure 1*

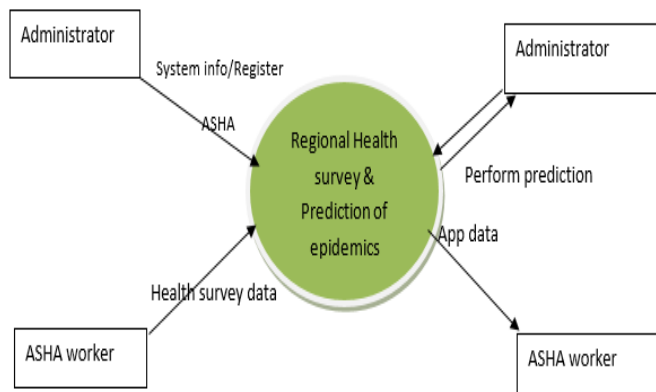


Figure 1. DFD level 0

Since the algorithm is executed by the administrator, he/she gives entries to the questions that are asked in the health survey conducted by the ASHA workers in each home and

based on the similarity of the features KNN predict the result. The data set must be rich enough to get the accurate result. The data collection should be based on the features that may affect the spread of epidemics.

### IV. RESULTS AND DISCUSSION

The chance of spreading epidemic in an area is predicted using KNN algorithm. The output will be different for different seasons. Thus, the result is stored in the database for further study. The output may be as low, very low, medium, high, very high.

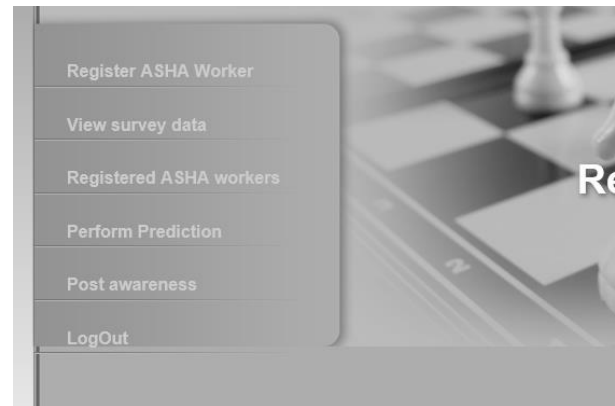


Figure 2 Administrator home page

Figure 3 prediction page

Predicting the chance of epidemics with some features is given in *Figure3*. Administrator functions are given in *Figure2*.

Based on the output necessary suggestions can be given to the health authorities and government can take appropriate actions.

## V. CONCLUSION and Future Scope

Using KNN algorithm it is easy to predict the chance of epidemic in an area based on the health survey data. Using the similarity of the feature set this algorithm predict the result.

This work is for a region under a Primary Health Centre in a locality. The features considered for the health survey may vary for different area of our country. So, we want to develop such an efficient application for data collection. It can also check the whether the vaccination has any effect on chance spreading diseases.

## REFERENCES

- [1] Pi Guo, Tao Liu, Qin Zhang, Li Wang, Jianpeng Xiao, Qingying Zhang, Ganfeng Luo, Zhihao Li, Jianfeng He, Yonghui Zhang and Wenjun Ma, "Developing a dengue forecast model using machine learning: A case study in China", *PLoS Neglected Tropical Diseases*, 2017 October, 11(10):e0005973. doi:10.1371/journal.pntd.0005973. eCollection 2017 Oct
- [2] Danqing Xn, Yinqun Liu, Min Zhang, Shaoping Ma, Anqi Cui and Liyun Ru, "Predicting Epidemic Tendency through Search Behavior Analysis", *International Joint Conference on Artificial Intelligence*, vol-3, 2011S
- [3] Sidharth Satish and Smitha.G.R, "Epidemic Disease Detection and Forecasting; A survey", *International Journal of Advance Research, Ideas and Innovations in Technology*, vol-3, ISSN-2454-132X, 2017.
- [4] Anand. A. Chaudari and Prof. S.P. Akarte, "Fuzzy and data mining based Disease prediction using K-NN Algorithm", *International Journal In Engineering and Technology*, vol-3, pp.2319-1058, April 2014
- [5] Hakizimana Leopord, Dr. Wilson Kiruto Cheruiyot and Dr. Stephen Kimani, "A survey and Analysis on Classification and Regression Data mining techniques for Diseases Outbreak Prediction in Datasets", *International Journal of Engineering and Science*, vol-5, pp-01-11, 2016
- [6] Dhanya P. Varghese and Tintu .P.B "A Survey on Health Data using Data Mining Techniques" *International Research Journal of Engineering and Technology*, vol.02, Oct 2015
- [7] K.Navini, Dr.M.B.Mukesh Krishnan, Lavanya, Dr.A.Shanthini, "Mobile health based smart hybrid epidemic surveillance system to support epidemic control programme in public health informatics", *IEEE, INSPEC* no-17261157
- [8] Abd Elrazek Abd Elrazek, "How can data mining improve health care", *Applied Mathematics and Information science*, vol. 11, pp. 585– 588, March 2017.