Determination of Hygienity Status of the City using Data Mining Technique

X. Jermiya,
Student, M.Tech(IT),
Dr.Sivanthi Aditanar College Of Engineering,
Tiruchendur, Tamilnadu, India

Abstract: **Nowadays, many places are unhygienic. Unhygienity is the biggest issue that makes the health diseases. All the people should aware of the unhygienity problems. Some action should be taken to avoid the diseases. All the actions are based on each city hygienity problems. The problems are received from the people through the feedback. Each person feedback is a measuring tool which is used to know hygienity status of the particular city. The people can also write their city problem which is making the health diseases to people. This project analyses the hygienic status of the various places by receiving feedback from the interested people who wants to improve their city hygienic level and thereby helps the Government and health care officials to promote their activities regarding hygienity. Various kind of Classification techniques are used and metrics measure the performance. Based on the metrics best classification technique is identified. This kind of identifications are may be used to improve the measurement**.

Keywords: Classification, Data mining, Decision tree, Feedback, Hygiene, Support vector machine.

I. INTRODUCTION

Hygiene is a set actions conducted to maintain the things clean and prevent the people from diseases. Hygiene is an important factor for ever to maintain. There are many obstacles which make the place as unhygienic. To make the place hygienic, finding the obstacles is very important. Unhygienic things spread the diseases and it affects the health of the people. Unhygienity affect the standard of each city. In the earliest existing system, there is no analysis technique used to identify the hygienic obstacles. Thus new idea has been emerged which makes use of different kind of data mining techniques in order to effectively analyze the current scenario regarding hygienity of various places of Tamilnadu. This project analyses the hygienic status of the various places by receiving feedback from the interested people and thereby helps the Government and health care officials to promote their activities regarding hygienity.

The individual person's feedback is a measuring tool which is being used to know the status of the hygienity of particular place. Asking person's feedback is the opportunity to express the individual's opinion and to raise their issue for the consideration of government or any health care organization. Feedback is not only for addressing the hygienic issue and also to enhance the facility which has to be undertaken by the government.

Feedback contains the questionnaires about the environmental hygienity which include the detail of water, air pollution and health hygiene. The authorized person can be allowed to answer the questionnaires after the verification of person's identity. The person is allowed to give feedback about the place where he/she currently lives. As it is possible to receive more than one feedback for a particular place, the project concentrates on the analysis of the feedbacks, which in turn reports the status of hygienity. Different data mining tools such as classification tools are used for classifying the feedback data and it builds the classifier model.

The performance of various classifier models are studied based on different metrics. The best classifier model will be identified at the end. The results of the classifier model will be very useful to take enhancement activities.

II. RELATED WORK

The biggest challenges in nowadays are how to utilize the large amount of data, how to manage and improve the quality of large data. Data mining technique is very well for managing the large data and which is used for extracting the hidden information from the large dataset. Data mining technique used in several areas. For example banking sector, fraud detection, healthcare organization, higher learning institution and telecommunication *A. M. Abaidullah et al.*[1] describes the way how to gather the feedback from the student to measure the instructor performance in higher leaning institution.

Asanbe M.O.et al.[4] was proposed neural network and decision tree techniques and also described why these techniques have been selected. Important reason behind the selection is their performance in various domains. Both techniques applied in industry, business, science and education with good performance.

Three classifier models (ID3, C4.5 and MLP) was used. The ID3 decision tree was the fastest but not well in classification. Neural network work well in classification. But it was very slow. Here the instructor performance evaluated based on their work experience and rank. *T..Balasubramanian et al.[5]* this paper explores the idea how the data mining technique can be used in health care organization. Questions are created based on the scientist opinion about the fluoride water and also make the people aware of fluoride water.

III.PROPOSED METHODOLOGY

A. *Problem Description*:

There is no one use these of kind of ideas to receive the problem of each region people about the hygienity status. In the existing system, there is no one use any analysis technique for improving the hygienity status by receiving the feedback.

Disadvantages:

In existing system, the hygienity obstacles are very difficult to find. The water pollution, air pollution, unhealthy food and home hygienity are considered as a unhygienity obstacles.

In proposed system, a set of questions are created for the people to interact about the city hygienity level. Interested people will respond to that list of question. This kind of analysis will be taken for each region to improve the human living standard. Each person can address the hygienic issue to health care organization.

Advantages:

It is very useful to monitor and improve the status of each city in Tamilnadu or over India. People to be more concerned about the health. It is more efficient because of the classification technique.
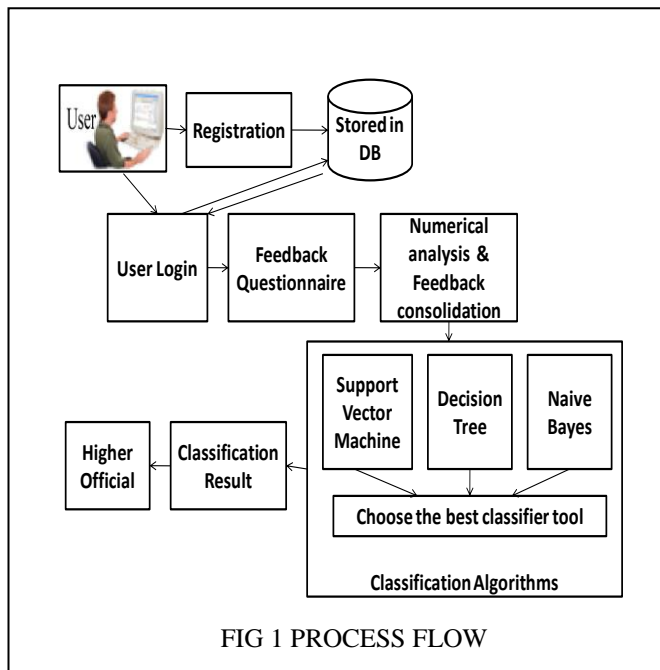
B. *Proposed System Design*

FIG 1 PROCESS FLOW

The above figure shows the process flow of this paper. Here the first step is registration of user or people and it will be verified once they login. User registration is the process of registering themselves to report the hygienity level of their city. Registered people details are maintained to keep secret. Once the users enter the ID in the login page, it will be verified. If the login person enter wrong ID they will not allowed to report the problems. If the person is a valid user then only he/she can allow to the feedback questionnaire. Then feedbacks are received from the people

and then all the feedbacks are consolidated. Different data mining techniques applied and their performances are measured by the metric to identify the best classification technique. The metrics such as accuracy, precision, recall and specificity are measured. Finally, the output or statement about the city hygienity level will be identified and viewed by the health organization.

1. REGISTRATION:

Registration is the starting process done by the people who want to improve their city hygienity level. All the registered people details are stored for verification. In Registration process, the user should enter their personal details like name, password, email id, contact number and their district.

2. USER LOGIN:

Once the people or user registered, the detail will be stored in Database for verification. After the registration process, a user Id will be generated automatically. If the user knows about the Id and password, then only they can login in to the application.

3. FEEDBACK QUESTIONNAIRE

Here the response values of each question in the form {1,2,3,4,5}. 1 represent the worst level of hygienity status, 3 represent the average level of hygienity status and 5 represent the highest level of hygienity status of the city.

TABLE I.DETAILS OF THE VARIABLES

Variable	Description
Q1	Do you have a good and fresh fruits and vegetables in your place?
Q2	Mosquitoes is there in your place?
Q3	Anybody affected by a mosquito?
Q4	Are you having a toilet room in your home?
Q5	Are you cleaning the place around your home?
Q6	The corporation water is pure in your place?
Q7	Everybody clean their places?
Q8	Polythene covers used by the people?
Q9	Factories are affecting the people health?
Q10	Hospitals and dispensaries are available in your place?
Q11	Are you spraying the insecticides on walls?
Q12	Are you throwing the garbage in dustbins?
Q13	Are you using the dustbins in your street?
Q14	Are you covered the dustbins?
Q15	Corporation water affect anyone health?
Q16	Have a tree in road sides?
Q17	Are you plant a tree in your place?
Q18	Every people aware of hygiene?
Q19	Noise pollution is becoming a very critical issue in your place?

Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTICCT - 2017 Conference Proceedings

4. NUMERICAL ANALYSIS AND CONSOLIDATION OF FEEDBACK

After the person replied the feedback, the feedback can be classified based on the answer. Same answer for each question is grouped to calculate the approximate results. For each the question people may answer from 1 to 5. Same region people answer will be grouped to represent their city clean status. Like that all the people feedback for each region is consolidated to know each city hygienity level. All the feedbacks are analyzed by the classification technique.

4. CLASSIFICATION TECHNIQUES

Different kind of classification technique is applied to analyze the people feedback. Classification technique assigns the collection of item to the target category. It aims to predict the target class. The Classification technique always analyze the input and produce the accurate result. Here Decision tree, Support vector machine and Naïve bayes classification techniques are chosen to analyze the input data.

a. Decision Tree algorithm:

Decision tree represented in a form of tree structure and it consist of root node, branch and leaf node. In decision tree, each node represents the test on attribute and each branch represents the outcome of the test and leaf node represents the class label. Decision tree is a binary tree which consists of 0 and 1.
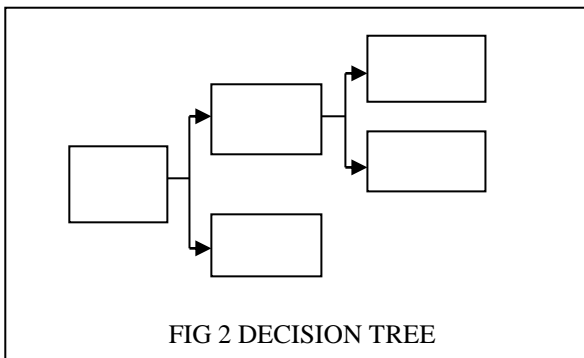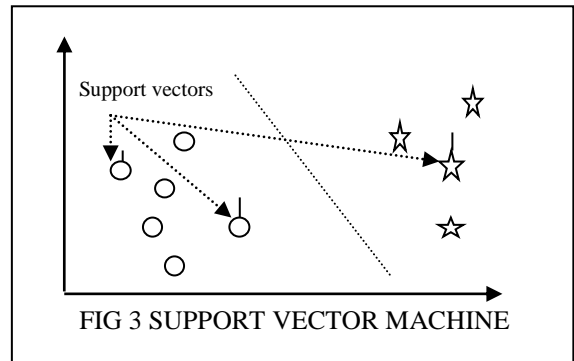

FIG 2 DECISION TREE

Decision tree is very easy to understand and interpret.

b. Support Vector Machine:

Support vector machine (SVM) can support both the classification and regression problems. But it mostly suitable for classification problems. In n-dimensional space, the data plot as a point. Then the classification will be performed by finding the hyper plane which separates the two different classes perfectly. SVM also uses the margin space to define the perfect hyper plane if two or three hyper plane is available.

FIG 3 SUPPORT VECTOR MACHINE

Support vector machine is much capable for smaller data and produce the best result. SVM has many features one of these is, it can ignore the outlier if it present and it has a kernel function. Kernel function can transform the low dimensional input space into high dimensional space.

c. Naïve Bayes:

Naïve Bayes classifier works based on bayes theorem with an assumption. It actually assumes the particular feature of the class that is unrelated to any other feature. Bayesian has high accuracy and speed.

5. CHOOSE THE BEST CLASSIFIER TOOL

Metrics are used to choose the best classifier tool. All the classification performances are compared using precision and recall metrics. Here precision and recall metrics are used measure the classification performance.

Precision measures the correctness rate. Precision is defined as the number of true positives divided by the total number of elements labeled as belonging to the positive class. Recall measure the True positive rate. Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class

6. CONSOLIDATED OUTPUT:

Data mining is the biggest area that applied in many applications. Here it applied to health care organization to define the hygienity status. Finally, the obstacles which affected the people health are identified and the enhancement activities will be taken by the higher officials. This kind of analysis is very useful to the government and health care organization to take the respective activities for making the hygienic city. It will improve the city standard and also improve the human living standard. This indicates the effectiveness of the data mining.

IV. EXPERIMENTAL RESULT AND ANALYSIS:

All these above classifiers are simply good working. SVM, Decision tree and Naïve Bayes these three classification algorithms are produce more effective result. Naïve Bayes has high accuracy and speed.

Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTICCT - 2017 Conference Proceedings

V.CONCLUSION

Data mining technique is utilized to analyze the questionnaires in the basis of people responds. Data mining techniques are applied in health care organization to improve the standard of human life and also used for solving the problem of hygienity. Health care organization and government both concentrate on finding a health hazards which makes the people to sick. Data mining is very applicable to use many areas.

VI. REFERENCES

[1] A. M. Abaidullah, N. Ahmed, and E. Ali, ''Identifying hidden patterns in students' feedback through cluster analysis,'' *Int. J. Comput. Theory Eng.*, vol. 7, no. 1, pp. 16–20, 2015.

[2] N. Delavari, S. Phon-Amnuaisuk, and M. R. Beikzadeh, ''Data mining application in higher learning institutions,'' *Inform. Edu.-Int. J.*, vol. 7 , no. 1, pp. 31–54, 2007.

[3] V. Kumar and A. Chadha, ''An empirical study of the applications of data mining techniques in higher education,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 3, pp. 80–84, 2011.

[4] Asanbe M.O., Osofisan A.O. and William W.F. " Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique"International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 10 – No.7, March 2016 – www.ijais.org.

[5] T.Balasubramanian and R.Umarani "Clustering as a Data Mining Technique in Health Hazards of High levels of Fluoride in Potable Water" *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No.2, 2012*

[6] V. Kumar and A. Chadha, ``An empirical study of the applications of datamining techniques in higher education," *Int. J. Adv. Comput. Sci. Appl.*,vol. 2, no. 3, pp. 80_84, 2011.

[7] Aranuwa Felix Ola and Prof. SellapanPallaniappan "A data mining model for evaluation of instructors'performance in higher institutions of learning usingmachine learning algorithms"*International Journal of Conceptions on Computing and Information TechnologyVol. 1, Issue 2, Dec' 2013; ISSN: 2345 – 9808.*

[8] *RandaKh. Hemaid1 and Alaa M. El-Halees "Improving Teacher Performance using DataMining"*International Journal of Advanced Research in Computer and Communication Engineering *Vol. 4, Issue 2, February 2015.*

[9] Chandrani Singh & Arpita Gopal, "Performance Analysis of Faculty using Data Mining Techniques", International Journal of Computer Science and Application Issue 2010, ISSN 0974-0767, 2010.

[10] Sona MARD˙IKYAN, Bertan BADUR "Analyzing Teaching Performance of Instructors Using Data Mining Techniques" Informatics in Education, 2011, Vol. 10, No. 2, 245–257.