Health Advisory Using Data Mining

[1]Prof. Prerna S. Kulkarni, [2]Pooja S. Todkar, [3]Muthupriyanka A. Nadar, [4]Sagar Mane
[1]Asst. Professor, [2,3,4]BE Student, Computer Engg. Dept. SSJCET, Asangaon, India.
[1]prernakulkarni09@gmail.com, [2]poojavk21@gmail.com, [3]nadarpriya49@gmail.com,
[4]sagarmane1039@gmail.com

Abstract : Health Advisory, as the name suggests, is a desktop application. According to the user requirements it advises him/her to gain weight, maintain weight or reduce weight. Classification, Clustering and Association are the three algorithms used to obtain the best possible solution. It is an application that user can handle easily to access information about health related issues(being under-weight or over-weight). The application creates awareness about various health related issues and allows the user to access proper expert's advice.

Keywords — classification , clustering, association algorithms, data mining, fitness, health, advice.

1. INTRODUCTION

Data Mining is one of the significant topics in the field of knowledge extraction from data set. It addresses the major concerns of efficiency, accuracy and usability. The proposed system uses this same conception of data mining to categorize various health related problems and detect if the person is over-weight or under-weight[1].

1.1 CLASSIFICATION ALGORITHM

Systems that build classifiers are one of the frequently used tools in data mining. Such systems take as input a collection of cases, each belonging to one of a small number of classes and described by its values for a fixed set of attributes, and output a classifier that can accurately predict the class to which a new case belongs. These notes describe C4.5 , a inheritor of CLS and ID3. Like CLS and ID3, C4.5 creates classifiers conveyed as decision trees, but it can also build classifiers in more coherent rule set form. We will sketch the algorithms active in C4.5, focus some changes in its successor C5.0, and conclude with a couple of open study issues training classifier (algorithm). Classifier is expected to give ideal decision tree depending on the entropy of each attribute. Further decision tree will be verified using unlabelled dataset and hence we will be able to get parameters for confusion matrix.

Given a set S of cases, C4.5 first grows an initial tree using the divide-and-conquer algorithms follows:

• If all the cases in S belong to the same class or S is small, the tree is a leaf labeled with the most recurrent class in S.

• Otherwise, choose a test based on a single attribute with two or more results. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets *S*1, *S*2, . . . according to the outcome for each case, and apply the same process recursively to each subset[2].

1.2 CLUSTERING ALGORITHM

Clustering groups data illustrations into subsets in such a manner that similar illustrations are grouped together, while different illustrations belong to different groups. The illustrations are thereby organized into an efficient representation that exemplifies the population being sampled. Formally, the clustering structure is symbolized as a set of subsets $C = C1; : : : ;Ck$ of S, such that: $S = Ski=1$ Ci and $Ci \setminus Cj = ;$ for $i \neq j$. Subsequently, any illustration in S belongs to exactly one and only one subset. Clustering

of objects is as prehistoric as the human need for describing the noticeable characteristics of men and objects and recognizing them with a type. Therefore, it clinches various scientific disciplines: from mathematics and statistics to biology and genetics, each of which uses diverse terms to describe the topologies formed using this analysis. From biological "taxonomies", to medical "syndromes" and genetic "genotypes" to manufacturing "group technology"— the problem is identical: forming categories of entities and conveying individuals to the proper groups within it[3].

1.3 ASSOCIATION ALGORITHM

The copiousness of data produces the appearance of a new field named data mining. Data collected in large databases become raw material for these knowledge discovery techniques and mining tools for "gold" were necessary. The current expert system machineries, which typically rely on users or domain experts to manually input knowledge into knowledge bases. This technique contains errors, and it is extremely time-consuming and costly. Data mining tools which perform data analysis may expose important data patterns, subsidizing greatly to business stratagems, knowledge bases, and scientific and medical research [4]

1.4 LITERATURE SURVEY

SR. NO	*ALGORITHM*	*Accurate*	*Scalable*	*Interpretable*	*Usable*	*Robust*	*Versatile*
1.	***Classical (LR,LDA)***	Neutral	Good	Good	Good	Neutral	Neutral
2.	***Visualization***	Good	Bad	Good	Good	Neutral	Bad
3.	***Decision trees***	Bad	Good	Good	Good	Good	Good
4.	***K-Means***	Neutral	Good	Good	Good	Bad	Neutral
5.	***Apriori association***	Good	Neutral	Good	Good	Bad	Good

Table 1: Literature Survey

1.5 AIMS AND OBJECTIVES:

The entire procedure of data mining is a repetitive implementation of following three steps:

- Collect the data required for mining
- Do the mining process for the data
- Get the analogous outcome and then express the outcome in certain way

If the mining outcomes are not pleased with what we want to, the three steps will endure to be carried out until mining outcomes are corresponding with target outcomes.

- A transitory explanation of the health exposures that people are unaware about.
- Better diagnosis leads to access to ultimate resolution.

The ultimate resolution:

- Access to Expert's guidance and his/her take on the problem.
- Information about the expert's locality. (nearest)

2. STAGES IN PROPOSED SYSTEM

2.1 PROPOSED SYSTEM

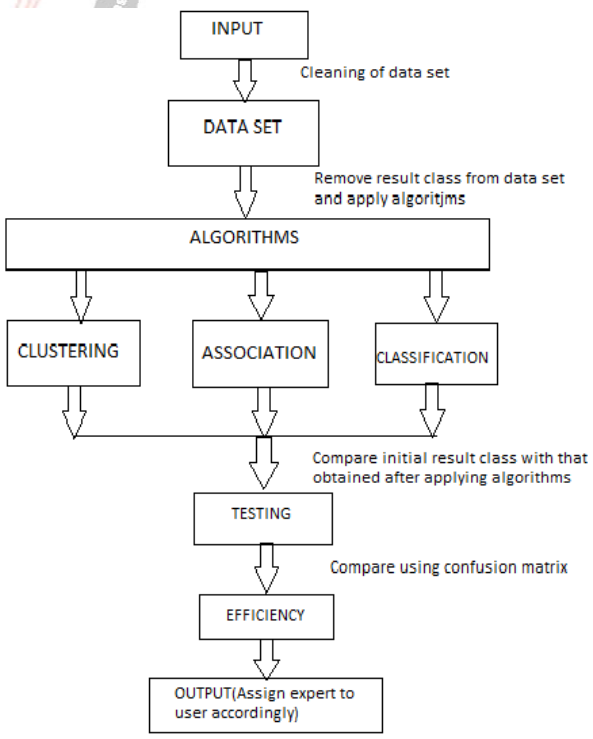


Fig.1. System architecture diagram of the proposed system

The proposed structure can be divided in three diverse layers. In the initial layer the user enters all the desired data and a dataset is attained. This layer also accomplishes as the front end of the proposed structure.

In the second layer the algorithms play a fundamental role. The system includes clustering, association and classification algorithms.[3,4] .

This layer can be well-thought-out as the decision making layer of the proposed structure as it leads the user to the expert guidance. The clustering algorithm used in the proposed structure is k-means and the attained outcome are clusters. The clusters which are obtained are associated with each other using association patterns.

Finally, classification is applied to these patterns and they are directed to the corresponding expert guidance using a classification algorithm like ID3.

The final layer is the testing part which is executed using a tool like WEKA. The confusion matrix which is achieved in WEKA is used verify the efficiency and accuracy of the structure.

2.2 METHODOLOGY

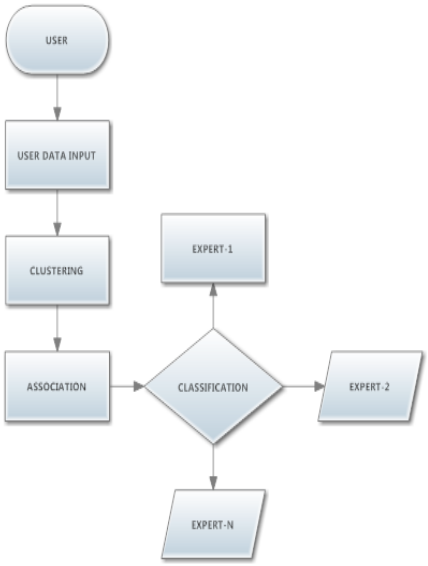


Fig. 2. Methodology of the proposed system exemplifying steps involved in the complete process, presenting the flow of control and overall functioning of the system.

The proposed system meritoriously uses a combination of clustering, classification and association algorithms to efficiently convey the preeminent possible advice from the expert to the user's issue.

The functioning of the proposed system can be divided into four phases.

In the initial phase, the user is requested to enter his constraint in the system i.e. the user specifies if he/she wants to lose weight, maintain weight or gain weight. Along with that, the user will also be requested to enter certain information(like weight, height, age, gender, body type etc) which will be used by the system to direct the user to the optimal solution to his/her problem.

The system considers different significant aspects of the user's daily lifestyle and ensures that all these factors are assimilated in the decision making task.[3]

These factors have been selected based on the research done and keeping accuracy and efficiency in mind.

The factors considered by the system are as follows:

1. Weight
2. Height
3. Gender
4. Body Type
5. Drinking
6. Smoking
7. Medical Condition
8. Sleeping hours
9. Physical activity

BMI is calculated based on weight and height provided by the user. Once the data is entered, the user submits it.

In the next phase, the provided information is assembled to form different clusters based on the data provided by user using a clustering algorithm like k-means. Clusters are formed based on these sub-groups of the mentioned factors.

These factors are divided into sub-groups as follows:

1. Body Type: Ectomorph, Mesomorph, Endomorph
2. BMI: Underweight, Normal, Overweight, Obese
3. Smoking: High, Medium, Low
4. Drinking: High, Medium, Low
5. Health Condition: Depends on the user

6. Sex: Male, Female
7. Sleeping hours: 5-5.99, 6-6.99, 7-7.99, 8-8.99, >=9
8. Physical Activity: High, Medium, Low

Physical activity: High, Medium or Low Based on the above specified sub-groups the user's information is accordingly assembled to form the corresponding clusters.

In the third phase, we will use association rules[4] on the clusters by which patterns are detected.

In the final phase, these patterns are finally directed to get the best expert guidance for the corresponding pattern using a classification algorithm like ID3. The final output obtained from the system is a decision tree.

2.3 DESIGN DETAILS

i. E-R DIAGRAM

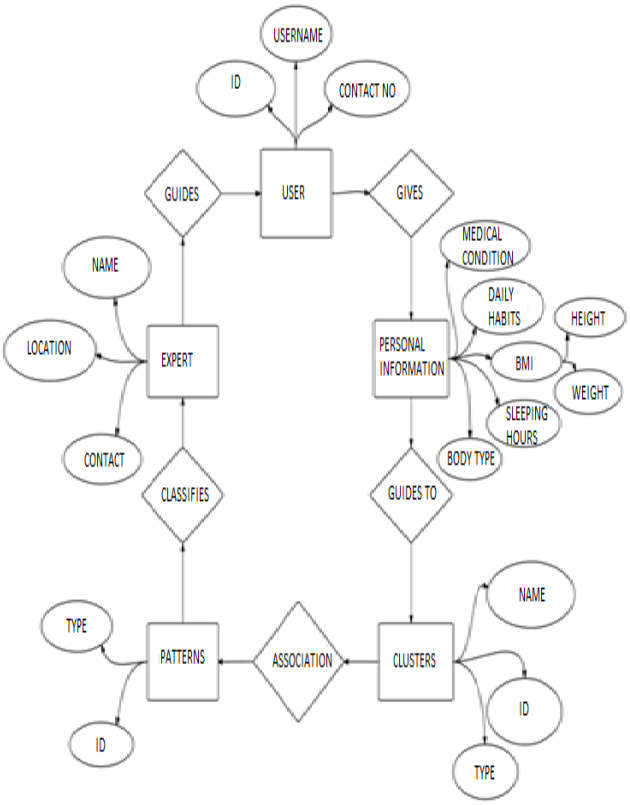


Fig.3. E-R Diagram of the proposed system showing relationship between entities involved in the system.

ii. ACTIVITY DIAGRAM

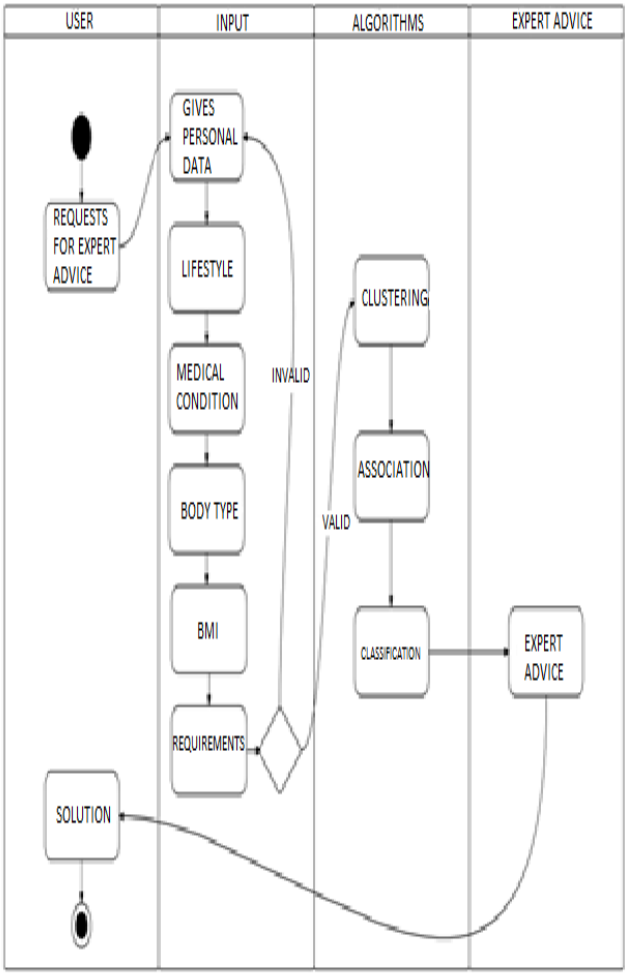
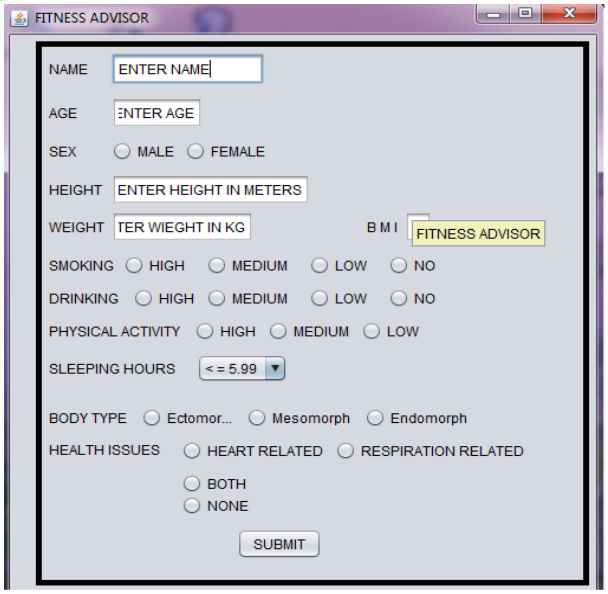


Fig.4. Activity Diagram of the proposed system illustrating the order of activities followed in the system.

2.4 FRONT END



III. CONCLUSION

The following can be determined from the above study.

BMI in itself is not accurate enough for analyzing the weight associated health harms hence other factors such as the user's lifestyle must be taken into account while analysis is done.

The proposed system can efficiently work as a consistent data mining unit since the decision making process is mechanized.

If the number of experts is more it would directly lead to a rise in the number of advices available for the system to choose from for the user. This will train the system well making it less partial and thus refining the efficiency of the system which in turn will lead to a better analysis and better advice.

The system can be organized at any health centre and can be utilized to direct any customer to the best expert available for tackling the customer's issue.

REFERENCES

[1] F. Lu and K. Turner, "Improving adolescent fitness attitudes with a mobile fitness game to combat obesity in youth," in Under review to:IEEE 5th International Consumer Electronics Society Games InnovationConference, IEEE, 2013.

[2] Fabio Buttussi, Luca Chittaro, "Smarter Phones for Healthier Lifestyles: An Adaptive Fitness Game," IEEE Pervasive Computing, vol. 9, no. 4, pp. 51-57, 2010.

[3] 2012 2nd International Conference on Management and Artificial Intelligence IPEDR Vol.35 (2012) © (2012) IACSIT Press, Singapore Parameter Identification and Selection for Childhood Obesity Prediction Using Data Mining MuhamadHariz B. Muhamad Adnan, Wahidah Husain and Nur`Aini Abdul Rashid School of Computer Sciences, UniversitiSains Malaysia, 11800 USM, Penang, Malaysia

[4] S. Marshall, S. Biddle, T. Gorely, N. Cameron, and I. Murdey, "Relationships between media use, body fatness and physical activity in children and youth: a meta-analysis," Journal of Obesity, vol. 28, pp. 1238–1246, 2004.

[5] F. Lu and K. Turner, "Improving adolescent fitness attitudes with a mobile fitness game to combat obesity in youth," in Under review to:IEEE 5th International Consumer Electronics Society Games InnovationConference, IEEE, 2013.

[6] L. Graves, N. Ridgers, K. Williams, G. Stratton, G. Atikinson, and N. Cable, "The physiological cost and enjoyment of wii fit in adolescents, young adults and older adults," Journal of physical activity &health, vol. 7, pp. 393–401, 2010.

[7] Pew Internet & American Life Project, "Teens, video games and civics," tech. rep., Washington D.C., 2008.

[8] L. Graves, G. Stratton, N. D. Ridgers, and N. T. Cable, "Energy expenditure in adolescents playing new generation computer games," British Journal of Sports Medicine, vol. 42, pp. 592–594, 2008.

[9] L. Leininger, Comparing perception of exertion and enjoyment between exergaming and treadmill exercise. Fresno, California: California StateUniversity, 2007.

[10] D. L. Graf, L. V. Pratt, C. N. Hester, and K. R. Short, "Playing active video games increases energy expenditure in children," Pediatrics, vol. 124, no. 2, pp. 534–540, 2009.

[11] L. Gorgu, G. O'Hare, and M. O'Grady, "Towards mobile collaborative exergaming," in Advances in Human-oriented and PersonalizedMechanisms, Technologies, and Services, 2009. CENTRIC'09. SecondInternational Conference on, pp. 61–64, IEEE, 2009.

