Data Mining of Epidemical Diseases Using Clustering (The Case of Tropical Disease Hospital Omdurman)

Hoyam Omer Ali Abdallah[1, 2], Saif Eldin Fattoh Osman[2], Tariq Abdul Karim[4]

[1]Department of Information Technology, Saudi Electronic University, Jazan , Saudi Arabia

[2] Department of Information Technology, Emirates College for Science & Technology, Khartoum, Sudan

[4]Department of Computer Science and Information Technology, College of Computer Science, University of ALneelain , Khartoum, Sudan

Abstract: *This research aims to work the system works in the division of patient data to a group of varieties depending on the involvement of similar properties, it also helps the recipient to understand the natural structure of the data sets. This system helped the community in the fight against diseases, helped hospitals to know more areas vulnerable to disease epidemiology and therefore the hospital to know the cause of the spread of this disease in this region and follow the procedures necessary guidance, health and other environment to combat Alammeradouhl problems and increase the predictive ability in achieving the goals and to reach a more precise knowledge.*

Keywords: DataMining, Data, Clustering, patients, Algorithm

1. Introduction

The large number of existing data and stored in databases and increase the spread of huge storage warehouses, known as ((data ware houses led to the emergence of a question many researchers on how to use them. It became necessary to find ways and means of techniques to extract information and knowledge from such accumulated data and use problem-solving and decision-making using a modern computer applications, which are considered modern technology made the computer think as a man thinks and does as does the man is known as artificial intelligence came the idea detection and exploration on this data clever ways to help solve problems and make decisions. It became necessary to find ways and means of techniques to extract information and knowledge from such accumulated data and use problem-solving and decision-making using a modern computer applications. [1,2].

The large number of existing data and stored in databases, for patients made it difficult to take advantage of this data, which in turn led to other problems side can be identified as the difficulty of identifying patients and the difficulty of identifying the number of patients and know the results of the analysis difficult to identify the areas most vulnerable to the disease

The research goals, assessing the validity of the results of the clustering algorithms build future predictions and explore the behavior and trends, which allows to take the right decisions and take timely and detach patient data into clusters of similar and finding the right number of bunches Portal for explanations of observed phenomena and prove hypotheses and get relations group hidden or unexpected. The importance of this research to explore the knowledge and useful new solutions in the future and the liquidation of the large amount of data to reflect slightly for the entire data and the ability to explore and focus on the most important

information in databases and improve the diagnosis and outcome of health care.

2. Methodology

The use of Java language for system design algorithm and generate dynamic reports. Where it has been linked to the database system my sql for data storage, where it was determined the value that is excavated to build it and then apply the algorithm of k means on the data stored in the database or Excel files, as the system works to assemble infected a particular disease in groups (clusters) based on similarity in the region and based on the age and the aim of a high-level exploration of what is happening within the database

In this search Used algorithm (K - means),it is One of the simplest ways, and it has the following general form:
1) Determine the Number of parts (k cluster)
2) Randomly select *'k'* cluster centers.
3) Calculate the distance between each data point and cluster centers.
4) add each element of the remaining items to the part that its the cluster center is minimum of all the cluster centers, and Used Euclidean distance to measure the dimension
5) Select new centers of parts that have been formed.
6) Regroup items about the new centers, and then define the new centers.
7) Repeats the above steps till settle the distribution, which do not change centers.

3. Results and Discussion

Increase the especially in poor areas of epidemic diseases due to the conditions of hygiene. And children less vulnerable to disease epidemic seriously As for hepatitis Most patients are travelers to developing the country and

who were not exposed by the hepatitis A virus, increases the likelihood of infection when to have sex or drug use with the disease and those with the holders. More than half of patients with the same viral hepatitis C males, the vast majority of the sample of patients with hepatitis married. Epidemic diseases for 5% of births transmitted by their mothers follows not carriers of the virus, but for infants and children, the percentage of the onset of symptoms are often less. In proportion ranging from 5-10% of the patients who continue to have the virus as a chronic disease and pregnant women are most vulnerable and in particular to disease epidemiology and be death ratio have a much higher, perhaps up to 20% compared to less than 1% when Alachrin.hat considered persons between 15-40 years old are more prone to infection and the likelihood of death from infection is less than 0.5%
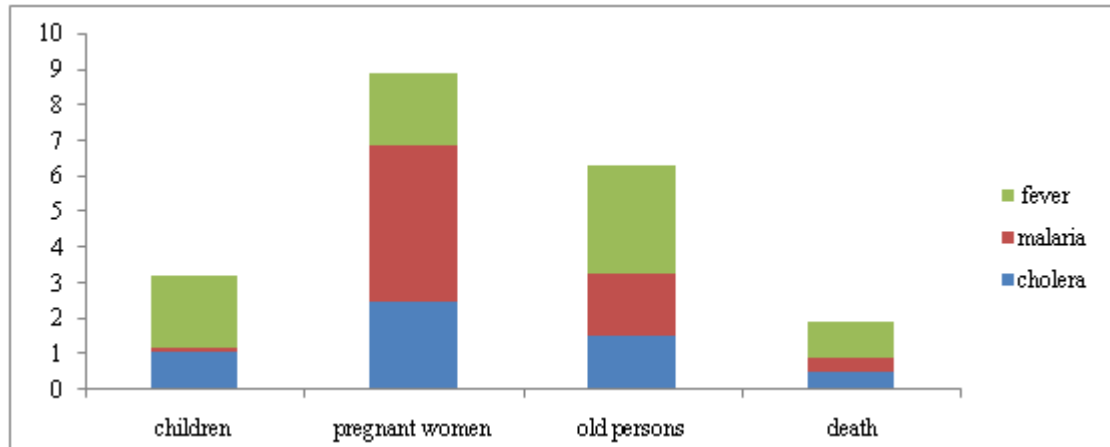

Figure 1: Dynamic Report

4. Conclusion

Featuring our time (the era of the Internet and the digital economy) large-scale spread of the data until It became impossible for analysts to extract meaningful information only by resorting to the traditional approaches to the analysis of the preliminary data.

The study also explained that the presence of large amounts of data stored in databases and data stores led to an increased need to develop tools characterized by force to analyze the data and extract information and knowledge of them, a technique designed to extract knowledge from the massive amounts of data, it is a modern technology has established itself firmly in the era of information and use it provides to companies and organizations in all areas the ability to explore and focus on the most important information in the databases[3,4,5].

Through this study was to identify patients with epidemic diseases and the number, age and scattered areas where these diseases.

5. Acknowledgments

References

[1] Ala'din Alwan. Noncommunicable diseases: a major challenge to public health in the region. Eastern Mediterranean Health Journal. 1997; 3(1): 6-16.

[2] Al Riyami AA and Afifi M. Clustering of cardiovascular risk factors among Omani adults. Eastern Mediterranean Health Journal. 2003; 9(5/6): 893-903.

[3] J. Han and M. Kamber, Data Mining, Concepts and Techniques. Morgan Kaufman, 2001.

[4] M. Bouguessa, S. Wang, and Q. Jiang, "A K-Means-Based Algorithm for Projective Clustering," Proc. 18th IEEE Int'l Conf. Pattern Recognition (ICPR '06), pp. 888-891, 2006.

[5] C.H. Cheng, A.W. Fu, and Y. Zhang, "Entropy-Based Subspace Clustering for Mining Numerical Data," Proc. ACM SIGMOD '99, pp. 84-93, 1999.