

## **Prediction of Diabetes Mell-EH-Tiss Using Unsupervised Learning Method**

**Sonal Arora and Kundan Munjal**

*Student, Computer Science and Technology,  
Lovely Professional University, Phagwara  
Assistant Professor, Computer Science and Technology,  
Lovely Professional University, Phagwara*

### **Abstract**

In the medical environment, data is “information rich” but “knowledge poor” because the data which is available is not “mined” as it cannot discover the hidden patterns and information which is used for reaching at the effective decision. Also, there is a lack of effective tools for analysis which can discover the hidden correlation and trends in data. Data mining have innumerable applications in field of business and scientific domains. Precious knowledge can be determined by the application of data mining techniques in the healthcare systems. Diabetes is a disease which has attacked most of the people around the world. Almost all people are reporting their problems of diabetes whether occurred by hereditary or any other problem. People around the world are using glucometer devices for testing their diabetes unknown of the fact that their results vary if another test is conducted after the first test. This leads to the error in detection of glucose with  $\pm 10$  mg/dl. The data which is hidden can be mined and extracted so as to obtain the fruitful results out of that data. The main aim of this research study is to extract the hidden knowledge and patterns from the diabetes dataset and preprocessing of that data takes place. K-means Clustering algorithm to obtain the accuracy of the prediction is applied and also the new algorithm will be designed which will be the improved version of K-means to increase the accuracy and see the effect on the prediction of diabetes. This research study is very beneficial for the doctors and patients. It will provide the user with a user friendly environment without the need of doctor or any hospital staff and will help to mine the data using new technique being designed. These algorithms being designed will define the effectiveness and efficiency of the method used to predict the diabetes mellitus.

**Index Terms**— Disease Management, Data Warehousing, Data Mining, Medical Services, Diabetes, Diseases, Medical Diagnosis, Hospitals, Evidence Based Medicine

## **I. INTRODUCTION**

The major challenge faced by the various healthcare organizations like high technology hospitals and many sanatoriums is the delivery of standard services at cheaper costs which can be afforded by every individual [10]. The decisions made in medical environment that give positive results are totally dependent on doctor's perception and knowledgeable information present in clinical databases. Some sorts of decisions are necessarily taken which may be in fact not a good unsympathetic decision and can lead to catastrophic results which are intolerable [11]. Most of sanatoriums use their clinical data to diagnose patients and producing positive results on patients. The knowledge hidden in the clinical databases procures us to making a poor decision in patient diagnosis. When Clinical decision support is integrated with the computer generated patient records then it could enhance patient safety, diminish medical errors, improve victim outcome and reduce unfavorable practice variation [13]. This hidden information may utilize to diagnose a patient who is suffering from particular diseases e.g. Heart Diseases and also predict the disease on the basis of symptoms.

Any doctor can use hidden information to treat his patient. Finding of evidence from knowledge extract from clinical databases is a great job in front of medical persons. EBM uses a data mining technology that makes it possible to automatically analyze huge clinical Databases and to discover patterns behind them [1]. The unification of evidence-based medicine rules into clinical decision-support systems will both improve quality and reduce costs of care, by recommending guidelines for only the most efficient treatments and medications. Internal clinical experience in integration with the external clinical expertise must be accessible to the healthcare specialists at the appropriate time and in the appropriate manner.

Data warehousing and Data Mining offers a comprehensive support for gathering, analyzing and presenting medical data. Clinical decisions are often made using the doctor's prescription and experience in his or her field rather than the knowledge base which is rich in data hidden in the database [11]. Usually such practices result in errors, wrong advice to the patients in case if the doctors are in fatigue and stress, unwanted biases and also leads to the extravagant medical price which directly affects the quality of services provided to patients [11].

Some sanatoriums deploy decision support systems, but they are generally limited in number. They can only provide the answer to the simple queries like "What is the age of person who is having a particular disease", "How many surgeries have taken place in the hospital in the last few days" or "How many patients are admitted in the hospital in the last 10 days" [1]. But, the existing decision support system cannot answer complex queries like "Whether the patient is suffering from the cancer or not", "If the patient is suffering from cancer then which medicine should be prescribed for the particular patient's cancer stage and case" or "Whether the person suffering from the cancer should be given with the treatment of chemotherapy" [10].

## **II. MOTIVATION**

The abundant amount of knowledge hidden in clinical databases that can be efficiently utilize to diagnosis of patient's diseases. This hidden information present in clinical databases can be used as guidelines after long process of lab research, specialist verification and positive result with patient diseases. A major challenge faced by the various healthcare organizations is providing the quality services at affordable costs producing the correct results. The term quality services means detecting patients correctly and conducting treatments that lead to effective results. Misguided clinical decisions can lead to disastrous consequences which are therefore unacceptable as it can harm the patient's health. Healthcare organizations must also minimize the cost of clinical tests and can help to achieve these results just by employing appropriate computer-based information and/or decision system based on expertise. Most of sanatoriums use their clinical data to diagnosis patients and producing positive results on patients. These systems usually generate huge amount of data which takes the input in the form of numbers, characters, graphs, charts and images. But unfortunately, this data is rarely used to support clinical decision making process. There is an abundance of hidden information in these data that is largely untapped which raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions on the basis of that data?"

## **III. PROBLEM DEFINITION**

Diabetes Mellitus is a disease which is attacking every single individual nowadays and its impact on the people leads to serious problems like cancer, high blood pressure etc. People around the world are using glucometer devices for testing their diabetes unknown of the fact that their results vary if another test is conducted after the first test. This leads to the error in detection of glucose with  $\pm 10$  mg/dl. So, there is a need to extract the hidden knowledge and patterns from the diabetes dataset and preprocessing of those data sets takes place. K-means Clustering algorithm to obtain the accuracy of the prediction is applied and also the new algorithm will be designed which will be the improved version of K-means to increase the accuracy and see the effect on the prediction of diabetes. Then the analysis of both the algorithms will be conducted.

## **IV. RESEARCH METHODOLOGY**

The new proposed algorithm will overcome the two disadvantages of K-means:

- 1) During each of the iteration of k- means algorithm it will calculate the distance of each data object to all the centers of cluster. This will take a lot of execution time for large datasets.
- 2) K- means will take whole dataset at once and make clusters.

Approach is used to solve the disadvantage is to take two data structures one for cluster and other for distance of data object from every cluster during each

iteration. By doing so, when the distance between the current data object and the new cluster center is calculated, if the distance that is computed is smaller than or equal to the distance to the old center, the data object will remain in its previous cluster that was assigned to it in previous iteration. Therefore there is no need to calculate the distance from this data object to other cluster centers will save lot of time. This will reduce the no. of iterations.

Approach to solve the second disadvantage is to divide the dataset into subsample then apply the k- means algorithm for each subsample this will minimize the probability of dividing one big cluster into two or more ones owing to the adoption of cluster error criteria.

### Proposed Algorithm

We have combined two approaches i.e. first dividing the dataset into sub samples and then applying the reduced iteration approach to each subsample to find the clusters. This will save lot of time and improve the performance for the large datasets.

- 1) Split the whole data into various subsamples.
- 2) Then apply the algorithm that will reduce the number of iterations.

The algorithm for the above is as:

1. Extract subsamples from the dataset

For each extracted subsample

2. Randomly select k objects from the subsamples as initial clusters centers.
3. Calculate distance between each data object then assign data object to the nearest cluster.
4. For each data object, find the closest center and assign data object to that cluster center.
5. Store the label of cluster center in which the data object is and the distance of data object to the nearest cluster and store them in array Cluster[ ] and the Dist[ ] separately.  
Set Cluster[i] =j, j is the label of nearest cluster.  
Set Dist[i] =d (di, cj), d (di, cj) is the nearest Euclidean distance to the closest center.
- 6) For each cluster j, recalculate the cluster center;  
Repeat
- 7) For each data object calculate its distance to the center of the present nearest cluster;
  - a) If this distance is less than Dist[i], the data object stays in the initial cluster;
  - b) Else

For every cluster center  $c_j (1 \leq j \leq k)$ , compute the distance  $d(d_i, c_j)$  of each data object to all the center, assign the data object  $d_i$  to the nearest center  $c_j$ .

Set Cluster[i]=j;

Set  $\text{Dist}[i] = d(d_i, c_j)$ ;

8. For each cluster center  $j(1 \leq j \leq k)$ , recalculate the centers;
9. Until the convergence criteria is met.
10. Output the clustering results;
11. Combine two nearest clusters into one cluster and recalculate the new cluster center for the combined cluster until the number of clusters reduces into  $k$ .

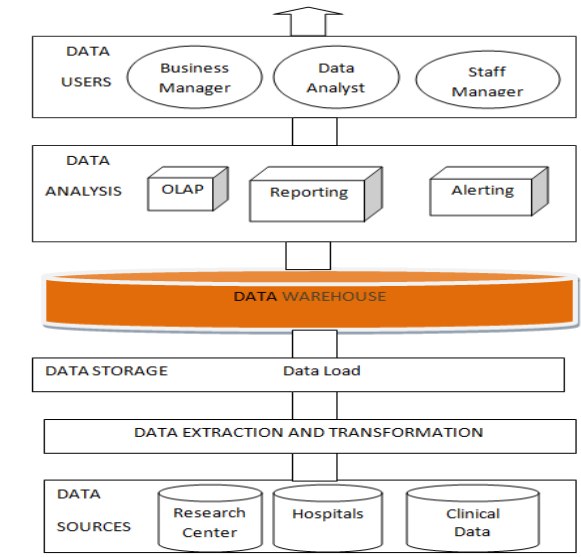
## V. DATA MINING AND DATA WAREHOUSING REVIEW

The term data warehouse was introduced by W. H. Immon as: "A Data Warehouse is a subject-oriented, time-variant, integrated, and non-volatile collection of data in support of management's decision making process" [3].

In contrast to traditional On-Line Transaction Processing (OLTP) applications, decision support usually places some different essentiality on database technology. Data warehouses provide storage, a sense of functionality and responsiveness to queries that are beyond the capacity of OLTP databases [5]. They contain historical, summarized, and consolidated data over probably extensive duration of time. Their size can be from hundreds of gigabytes to terabytes. There is a great need to provide decision-makers at all levels of management with information at the desired level of detail, to support their decision-making. Apart from performing regular, predefined reporting activities, a number of parallel users are submitting ad hoc and complicated queries. These queries require access to huge amount of records and cause numerous scan, join, and aggregate operations across the warehouse and results in query throughput and response times which are main issues in multi-user decision-support systems [4].

Figure 1 given below describes the architecture of Data Warehouse, ranging from source data collection to data delivery and then to the decision-makers. The source data is usually reserved in different source system having different formats.

- During the *extraction phase*, source data is collected from operational systems or external sources. The external data is often stored into spreadsheets, personal databases, web logs etc. It can be accessed directly or indirectly (in case of recovery systems).
- In the *transformation phase*, data which is collected is cleaned and converted into the specific format and made structure compatible with the DWH. Several Syntactic and semantic distinctions between operational sources of data are adjusted and then local logical models are portrayed and integrated into the global DWH data model. The mapping characteristics are captured and stored in the DWH as metadata.
- In the *data storage phase*, new data which is extracted and transformed is loaded into the data warehouse and integrated with the existing stored data. During this phase, data is usually restructured for optimized querying. Data loads need to be run on a regular basis in order to keep warehouse data accurate.



**FIGURE 1 DATA WAREHOUSE ARCHITECTURE**

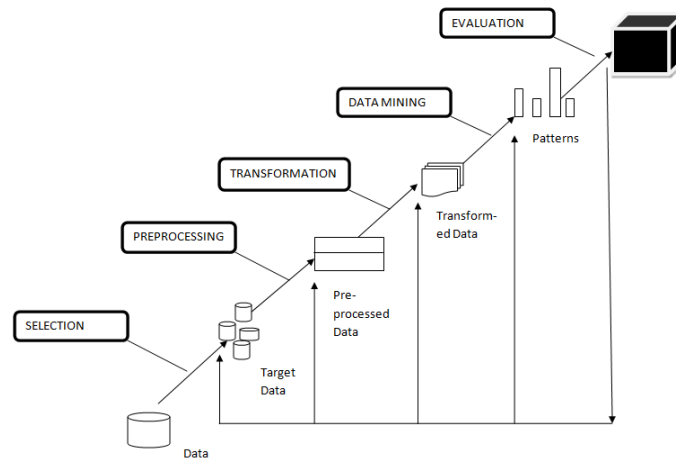
The term data mining generally refers to the process of automatically examining large databases to extract useful and hidden patterns. The same way knowledge discovery in broad area artificial intelligence which is also known as statistical analysis or machine learning, the concept of data mining evolves to discover hidden rules and patterns from the huge amount of data. The term Data mining is different from machine learning and statistics in the way that it manages large volumes of data, which is primarily stored on disk. That is, data mining deals with “knowledge discovery in databases.”

The extraction of non trivial, implied, hidden and actionable knowledge from large volume of datasets is performed by data mining. Discovery of the useful knowledge that can improve proficiency of processes cannot be handled manually. The aim of Data warehouses is to collect and archive important operational data and pure data. Data Warehouses are used to predict similar trends instances in the data and is used for decision support and performing the analysis on historical data and legacy data. One of the major tasks in data warehousing is to perform data cleansing on the data collected from various input sources. Schemas in data warehousing tends to be in multidimensional form, which involves one or a few some large fact tables and some amount of smaller dimension tables.

An online analytical processing (OLAP) tool provides the help to analysts so that they can view data in a summarized form in different ways and formats and with that view they can peep and view the proper functioning of an organization.

- OLAP tools work on data present in multidimensional form which is characterized by several attributes such as dimension attributes and measure attributes.
- The data cube comprises of data in multidimensional form which provides summarized information in different ways. Pre-computing of that data cube

- helps to speed up queries performed on the summaries of data.
- Cross-tab displays allow the users to view at a time only the two dimensions out of multidimensional data, along with the summaries of that data.
- There are the several basic operations that users perform with OLAP tools like Slicing and Dicing, Drill down and Roll up operations.



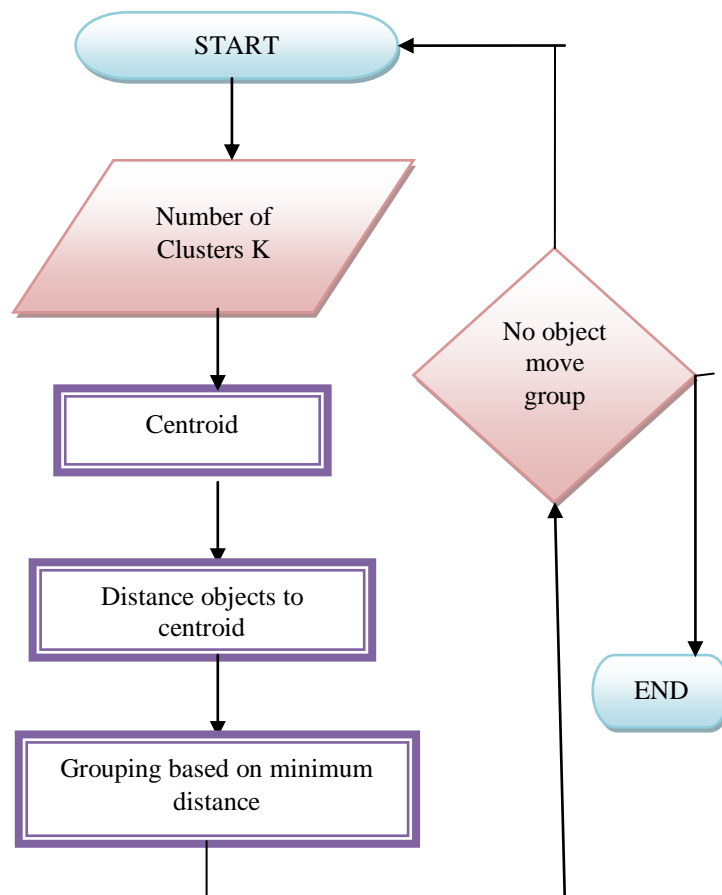
**FIGURE 2 KDD PROCESS OF DATA MINING**

## VI. K-MEAN CLUSTERING ALGORITHM

K-means clustering algorithm is an unsupervised method of learning. According to this, in the beginning we will determine “K” that is, the number of clusters and we will assume the centroid or center of these clusters using the distance between two objects and taking their mean value.

At the initial phase we can take any random objects and then apply the K means clustering algorithm which will do the three steps given below until convergence occurs. The convergence will occur until STABLE = no object in the group:

- Determine the centroid point.
- Compute the distance of each object to the centroid to form the groups.
- Group the object based on minimum distance and form the cluster.



**FIGURE 3 K-MEANS CLUSTERING ALGORITHM**

## VII. EXPERIMENTAL RESULTS

The tool which is used to implement this data mining approach is Weka 3.6.9, which is used to perform pre-processing of data, classification and clustering of data. The algorithms are being designed in Java Net beans IDE 8.0.

**TABLE 1 K-MEANS CLUSTERING ALGORITHM**

Clusterer output										
<b>mass</b>										
mean	18.561	37.3254	34.9624	31.7416	28.9775	31.0445	35.368	29.4171	36.6186	
std. dev.	13.8774	8.7732	5.5031	5.3704	5.9071	7.0182	6.2361	10.677	7.3488	
<b>pedi</b>										
mean	0.3203	0.7262	0.613	0.3705	0.4119	0.3692	0.487	0.6999	0.4946	
std. dev.	0.2175	0.5354	0.3298	0.2068	0.2436	0.2321	0.2919	0.4798	0.2621	
<b>age</b>										
mean	30.1742	28.3448	43.9476	43.7547	23.6227	36.6204	27.7527	56.2464	31.5495	
std. dev.	11.2186	5.3519	8.1344	11.0816	2.5065	11.944	4.8958	11.1521	6.6209	
<b>class</b>										
tested_negative	18.6122	11.1036	11.1079	68.9642	184.4546	114.3313	88.074	10.1124	2.2398	
tested_positive	4.4756	72.5373	88.0818	6.8737	10.0472	65.0047	6.0009	10.5016	13.4773	
[total]	23.0878	83.6409	99.1896	75.8379	194.5018	179.336	94.0749	20.614	15.7171	
Time taken to build model (full training data) : 163.99 seconds										
=== Model and evaluation on training set ===										
Clustered Instances										
0	17	( 2%)								
1	76	( 10%)								
2	104	( 14%)								
3	73	( 10%)								
4	208	( 27%)								
5	147	( 19%)								
6	83	( 11%)								
7	42	( 5%)								
8	18	( 2%)								

Table 1 show the results when we are performing the simple K-means Clustering algorithm. The time taken to build the model is 163.99 seconds. The results compared when are performing K-means with the reduced iterations and the K-means with merged reduced iterations and subsamples are as depicted below in table 2 and 3:

**TABLE 2 K-MEANS WITH REDUCED ITERATIONS**

Cluster centroids:							
Attribute	Full Data (768)	Cluster# 0 (54)	1 (311)	2 (252)	3 (19)	4 (116)	5 (16)
preg	3.8451	5.7222	1.6206	4.9246	3.3158	6.6638	3.9375
plas	120.8945	128.7778	107.3569	141.4087	98.4737	110.1466	138.875
pres	69.1055	77.2222	68.1093	75.3214	1.2632	75.1379	0
skin	20.5365	9.2778	22.2122	23.4524	2.2105	20.5259	1.875
linsu	79.7995	31.3148	86.2926	106.7063	1.3158	50.3707	0
mass	31.9926	29.7333	30.6871	35.2107	19.0947	31.3793	34.0687
pedi	0.4719	0.4687	0.4443	0.5581	0.3385	0.3875	0.4306
age	33.2409	57.8889	24.7074	37.4206	27.3684	36.7672	31.5
class		tested_negative	tested_negative	tested_positive	tested_negative	tested_negative	tested_positive

Time taken to build model (full training data) : 0.09 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	54 ( 7%)
1	311 ( 40%)
2	252 ( 33%)
3	19 ( 2%)
4	116 ( 15%)
5	16 ( 2%)

**TABLE 3 K-MEANS WITH REDUCED ITERATIONS AND SUBSAMPLES**

Cluster centroids:							
Attribute	Full Data (768)	Cluster# 0 (54)	1 (311)	2 (252)	3 (19)	4 (116)	5 (16)
preg	3.8451	5.7222	1.6206	4.9246	3.3158	6.6638	3.9375
plas	120.8945	128.7778	107.3569	141.4087	98.4737	110.1466	138.875
pres	69.1055	77.2222	68.1093	75.3214	1.2632	75.1379	0
skin	20.5365	9.2778	22.2122	23.4524	2.2105	20.5259	1.875
linsu	79.7995	31.3148	86.2926	106.7063	1.3158	50.3707	0
mass	31.9926	29.7333	30.6871	35.2107	19.0947	31.3793	34.0687
pedi	0.4719	0.4687	0.4443	0.5581	0.3385	0.3875	0.4306
age	33.2409	57.8889	24.7074	37.4206	27.3684	36.7672	31.5
class		tested_negative	tested_negative	tested_positive	tested_negative	tested_negative	tested_positive

Time taken to build model (full training data) : 0.09 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	54 ( 7%)
1	311 ( 40%)
2	252 ( 33%)
3	19 ( 2%)
4	116 ( 15%)
5	16 ( 2%)

The above results when compared shows that the time taken by K-means with reduced iterations is 0.09 seconds and that of K-means with merged reduced iterations and subsamples is 0.05 seconds. Now, if we increase the no. of clusters to 10 and then apply the K-means with reduced iterations and 10 clusters, the time taken is 0.12 seconds as shown in table 4:

**TABLE 4 K-MEANS WITH REDUCED ITERATIONS AND 10 CLUSTERS**

Attribute	Full Data (768)	0 (29)	1 (179)	2 (124)	3 (19)	4 (54)	5 (16)	6 (74)
preg	3.8451	4.7931	1.5531	2	3.3158	9.2963	3.9375	1.445
plas	120.8945	129.3793	98.2905	143.2016	98.4737	113.1296	138.875	129.52
pres	69.1055	74.3448	65.4302	72.6532	1.2632	77.3704	0	73.288
skin	20.5365	8.2759	22.9888	27.8468	2.2105	12.537	1.875	34.918
insu	79.7995	19.8621	59.7207	132.4758	1.3158	21.463	0	206.
mass	31.9926	26.9069	28.6335	36.5347	19.0947	31.45	34.0687	37.402
pedi	0.4719	0.4516	0.4163	0.5971	0.3385	0.4032	0.4306	0.568
age	33.2409	62.931	23.7095	30.0565	27.3684	44.963	31.5	26.135
class		tested_negative	tested_negative	tested_positive	tested_negative	tested_negative	tested_positive	tested_negative

Time taken to build model (full training data) : 0.12 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	29 ( 4%)
1	179 ( 23%)
2	124 ( 16%)
3	19 ( 2%)
4	54 ( 7%)
5	16 ( 2%)
6	74 ( 10%)
7	128 ( 17%)
8	77 ( 10%)
9	68 ( 9%)

**TABLE 5 K-MEANS WITH REDUCED ITERATIONS AND SUBSAMPLES WITH 10 CLUSTERS**

Attribute	Full Data (100)	0 (7)	1 (2)	2 (13)	3 (14)	4 (6)	5 (10)	6 (23)	7 (1)	8 (12)	9 (12)
preg	3.87	2.2857	3.5	5.2308	3	4.6667	6.6	1.6087	1	2.75	7.4167
plas	118.64	126.8571	82.5	128.4615	104.2143	129.3333	106.6	102.8696	0	136.5	153
pres	70.3	72	24	69.6923	74.3571	77.6667	71.9	65.2609	74	75.0833	72.5
skin	20.66	35.8571	11.5	1	3.1429	0	33.9	24.087	20	37.8333	30.6667
insu	75.77	262.4286	0	3.6923	9.2143	0	52.9	50.6087	23	138.0833	182.5
mass	32.656	37.9	31.3	34.3923	27.6357	30.9	33.64	28.6609	27.7	41.8083	32.775
pedi	0.454	0.5001	0.7135	0.5245	0.3309	0.248	0.2718	0.4691	0.299	0.4106	0.7335
age	32.35	27	25	38.9231	27.2857	53	39.7	24.087	21	28.3333	39.8333
class		tested_negative	tested_negative	tested_negative	tested_positive	tested_negative	tested_negative	tested_negative	tested_negative	tested_negative	tested_negative

Time taken to build model (full training data) : 0.08 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	65 ( 8%)
1	32 ( 4%)
2	95 ( 12%)
3	92 ( 12%)
4	53 ( 7%)
5	68 ( 9%)
6	187 ( 24%)
7	3 ( 0%)
8	80 ( 10%)
9	93 ( 12%)

**VIII. ACKNOWLEDGEMENT**

The author would like to thank Asst. Prof. Kundan Munjal of Lovely Professional University, Phagwara for the important collaboration and ideas which I received on this research. He always provides guidance about the current innovations which can be implemented in this research.

**REFERENCES**

- [1] Bei Andrea, Luca Stefano De , Ruscitti Giancarlo, Salamon Diego (2005), “Health Mining : a Disease Management Support Service based on Data Mining and Rule Extraction”, Proceeding of the 2005 IEEE, Symposium on Computer – Based Medical Systems, 27<sup>th</sup> Annual Conference September 1-4,2005, Shangai, China.
- [2] Durairaj M. and Vijtha C. (2014), “Educational Data Mining for Prediction of Student Performance Using Clustering Algorithms”, International Journal of Computer Science and Information Technologies, Vol. 5(4), 2014, 5987-5991, Tiruchirappali, India.
- [3] Fayyad, U, “Data Mining and Knowledge Discovery in Databases: Implications from scientific databases”, Proc. of the 9<sup>th</sup> Int. on the Scientific and Statistical Database Management, Olympia, Washington, USA, 2-11, 1997.
- [4] Giudici P, Wiley John (2003), “Applied Data Mining: Statistical Methods for Business and Industry”, New York.
- [5] Han J., Kamber M. (2006). “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers.
- [6] Mac Dougall Candice, Percival Jennifer and Mc Gregor Carolyu (2009), “Integrating Health Information Technology into Clinical Guidelines”, Annual International Conference of the IEEE, EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.
- [7] M Nirmala Devi , Balamurugan.S Appavu alias, U.V Swathi (2013), “An amalgam KNN to predict Diabetes Mellitus”, 2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology, Madurai, Tamil Nadu, India.
- [8] Ms. Wankhade Nishigandha V. and Mrs. Potey Madhuri A. (2013), “Transfer Learning Approach for Learning of Unstructured Data from Structured Data in Medical Domain”, 2013 IEEE 978-81-920249-7-4, Pune, India.
- [9] Obenshain, M.K. (2004), “Applications of Data Mining Techniques to Heath care Data”, Infection Control and Hospital Epidemiology, 25(8), 690-695, 2004.
- [10] Palaniappan Sellappan and Awang Rafiah (2008), “Intelligent Heart Disease Prediction System Using Data Mining Techniques “, International Journal of Computer Science and Network Security, VOL. 8 No. 8, August 2008, Selangor, Malaysia.
- [11] Parthiban Latha and Subramanian R. (2008) , “Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm” International Journal of Biological and Life Sciences 3:3 2008, Pondicherry.
- [12] Shouman Mai, Tumer Tim, Stocker Rob (2012), “Using Data Mining Techniques in Heart Disease Diagnosis and Treatment”, International Conference on Electronics, Communications and Computers 2012, IEEE, Northcott Drive, Canberra.

- [13] Srinivas K, Kavihta Rani B. and Dr. Govrdhan A. (2010), “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”, International Journal on Computer Science and engineering Vol. 02, No. 02, 2010 ,250-255, Jagtial, Karimnagar.
- [14] Sundar V Bata and Tevi T, Saravanan N (2012), “Development of a Data Clustering Algorithm for Predicting Heart”, International Journal of Computer Applications(0975-888) Volume 48- No. 7, June 2012, Coimbatore, India.
- [15] Thuraisingham, B. (2000), “A Primer for understanding and Applying Data mining”, IT Professional, 28-31, 2000.
- [16] Thangarasu Gunasekar and Assoc. Prof. Dr. Dominic P.D.D. (2014), “Prediction of Hidden Knowledge from Clinical Database using Data mining Techniques”, 2014 IEEE 978-1-4799-0059-6, Tronoh Perak, Malaysia.

