

A Comprehensive Analysis of Diabetes Mellitus using Data Mining Techniques

Deepa .S¹, Dr. B. Booba²

¹Research Scholar (UP19P9611053), Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram, Chennai-600117, India
Email: s.deepaonline[at]gmail.com

²Professor, Department of Information Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram, Chennai-600117, India

Abstract: Diabetes mellitus is a group of metabolic diseases in which the person has high blood glucose or blood sugar level either due to inadequate insulin production or because the body's cells do not respond properly to insulin or both. The most common sites of Diabetes have varied distribution among different geographical locations. The present study shows the comprehensive report of Diabetes Mellitus Patients of A & N Islands using data mining techniques. The study uses various data mining techniques such as classification, clustering, etc. to identify potential diabetes patients. This study also provides the details of diabetes patients based on several variables or data sets like Districts-South Andaman, North & Middle Andaman and Nicobar District, Area-Urban and Rural, Gender-Male and Female, Age, Eating Habits, etc. This study uses Descriptive Analysis i.e., Mean, Median, Mode and Standard Deviation, Correlation Analysis i.e., correlation of coefficient (r) and Differential Analysis i.e., 't' test, 'F' test as a Statistical Technique to find the relationship exists between several variables or data sets. The k-means clustering algorithm is used for partitioning the data into diabetes and non-diabetes clusters, where the initial cluster center is represented by the mean value of the weightage of significant patterns.

Keywords: Andaman & Nicobar Islands, Classification, Clustering, Diabetic patients, Data Mining Techniques, k-means Algorithm, Research Design, Statistical Techniques, Types of Diabetes-Type-I & Type-II

1. Introduction

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood sugar. Hyperglycaemia, or raised blood sugar, is a common effect of uncontrolled diabetes and over time leads to serious damage to many of the body's systems, especially the nerves and blood vessels.

In 2014, 8.5% of adults aged 18 years and older had diabetes. In 2019, diabetes was the direct cause of 1.5 million deaths and 48% of all deaths due to diabetes occurred before the age of 70 years.

Between 2000 and 2016, there was a 5% increase in premature mortality rates (i.e. before the age of 70) from diabetes. In high-income countries the premature mortality rate due to diabetes decreased from 2000 to 2010 but then increased in 2010-2016. In lower-middle-income countries, the premature mortality rate due to diabetes increased across both periods.

By contrast, the probability of dying from any one of the four main non-communicable diseases (cardiovascular diseases, cancer, chronic respiratory diseases or diabetes) between the ages of 30 and 70 decreased by 18% globally between 2000 and 2016.

The term "Diabetes Mellitus" describes a metabolic disorder of multiple etiology characterized by chronic hyperglycemia with disturbances of carbohydrate, fat (dyslipidaemia) and protein metabolism resulting from

defects in insulin secretion, insulin action, or both. The main symptoms are:

- 1) Polyuria (frequent urination)
- 2) Polydipsia (increased thirst)
- 3) Polyphagia (increased hunger)



Figure 1: Symptoms of Diabetes

The main types of diabetes are:

Type 1 Diabetes: It is due to the body's malfunction to produce insulin in the body, and requires the person to inject insulin. This form was previously referred to as

"Insulin-Dependent Diabetes Mellitus" (IDDM) or "Juvenile Diabetes".

Type 2 Diabetes: It is due to insulin resistance, a condition in which cells fail to use insulin properly, sometimes combined with an absolute insulin deficiency. This form was previously referred to as non insulin-dependent diabetes mellitus (NIDDM) or "adult-onset diabetes".

Type 2 Diabetes can be prevented after following healthy life style such as healthy diet, proper exercise or maintaining healthy weight.

The third main form, **Gestational diabetes** occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level. It may lead to type 2 DM.

Other types of diabetes include those caused by:

- Genetic defects of the beta cells, (the part of the pancreas that makes insulin) such as maturity-onset diabetes of the young (MODY) or neonatal diabetes mellitus (NDM).
- Diseases of the pancreas or conditions that damage the pancreas, such as pancreatitis and cystic fibrosis.
- Excess amounts of certain hormones resulting from some medical conditions such as cortisol in Cushing's syndrome that work against the action of insulin.
- Medications that reduce insulin action, such as glucocorticoids, or chemicals that destroy beta cells.

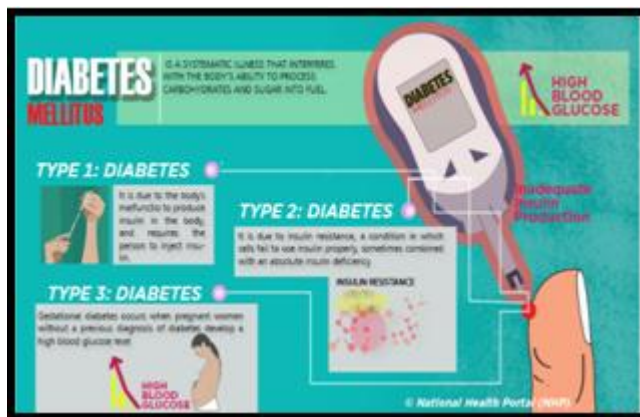


Figure 2: Types of Diabetes

As an alternative to the tedious physical storage of resources it is important to develop a data warehouse specific to diabetes disease and a data mining model to predict diabetes earlier. If a machine learning technique is developed to store a person's medical and general record and predict his predisposition towards diabetes, its type and exact diagnostic method, physicians can directly start treatment immediately without wasting the precious time in different methods of diagnosis. There have been multiple data mining techniques in health care and allied industries and specifically with respect to type-I & type-II of diabetes.

The present study is carried out for 500 Type 1 and Type 2 Diabetic patients of Andaman & Nicobar Islands in which the data has been collected from Diabetic Health Care

Centre.

All the patients diagnosed or suspected with diabetes, registered in Diabetic Health Care Centre (Shyamla Clinic). This research focuses on a comprehensive study of Diabetes Mellitus using various Data Mining Techniques.

2. Aims and Objectives of the Research

The aims and objectives of the study are stated in terms of - Primary and Secondary objectives. The primary objective is to identify the Type 1 and Type 2 diabetes patients of Andaman & Nicobar Islands using Data Mining Techniques. And the secondary objectives are to study the influence of the background variables like Age, Gender, Locality, Habits, Symptoms, Blood Sugar Level, Medications, and Type of Insulin, Diagnose and Treatment.

3. Delimitations of the Study

The study is only confined to the following:

- The study is only limited to the Type 1 and Type 2 Diabetes patients of Andaman and Nicobar Islands.
- It has the sample size of 250 males and 250 females Diabetes mellitus patients.

4. Review of Related Literature

The review of literature consists of two words namely 'review' and 'literature'. The term literature refers to the knowledge of a particular area of investigation of any discipline which includes theoretical, practical and its research studies. The term review refers to organize the knowledge of a specific area of research to evolve an edifice of knowledge to show that this study would be adding newer dimension to this field.

From the literature review it is learned that use of evolving IT systems in medical sciences to eradicate, diagnose, prevent diseases like diabetes and ameliorate the standard of living of patients with such life-threatening diseases has garnered the attention of IT researchers worldwide. Review of the literature on diabetes related databases helped to realize the fact that suitable data warehouse architecture should be implemented for the efficient functioning of the objective data analysis.

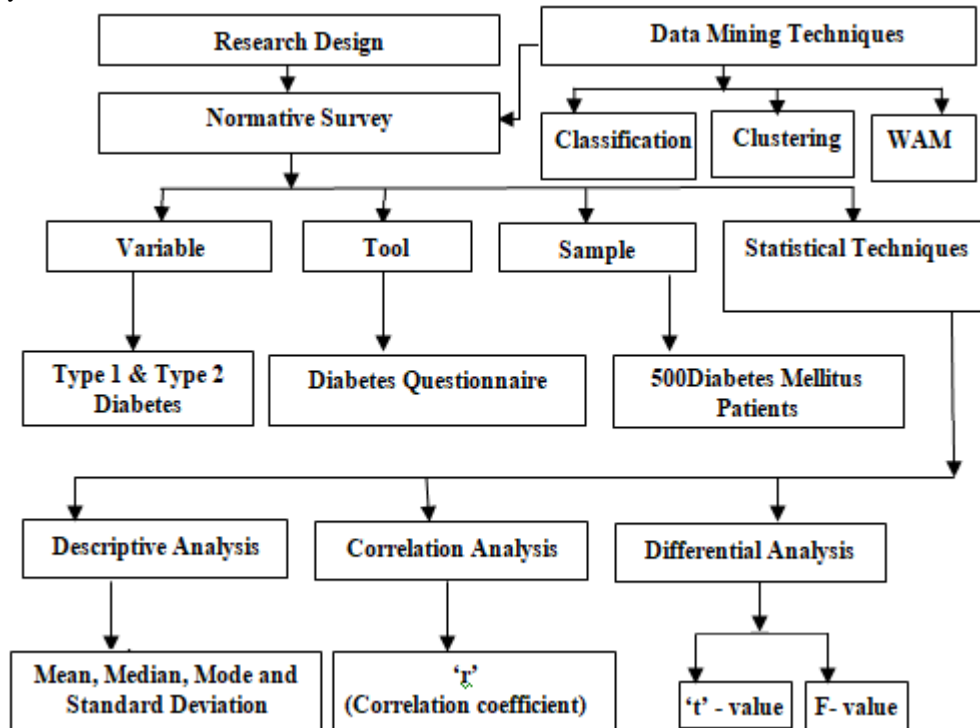
Review of the literature on diabetes related databases helped to realize the fact that suitable data warehouse architecture should be implemented for the efficient functioning of the objective data analysis.

5. Methodology of Research

To achieve the objectives, this study has used normative survey method to find out the influence of several background variables.

The Research Design provides planning on selection of subjects, data gathering devices, data analysis technique in relation to objectives of research. The research design is the

conceptual structure of the research procedure. The following tree diagram shows the complete research diagram of a study.



Here, the data will become from operational systems and external sources. To create the data warehouse, diabetes data will be extracted from source systems like questionnaire, diabetes institute database, etc. which will be cleaned (e.g., to detect and correct errors), transformed (e.g., put into subject groups or summarized), and loaded into a data store (i.e., placed into a data warehouse).

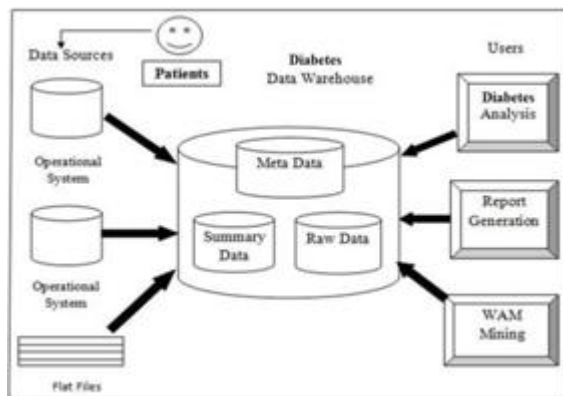


Figure 3: Diabetes Data Warehouse Architecture

k-means Clustering Algorithm for Classified Significant Pattern:

$$W(T, C) = \sum_{k=1}^K \sum_{i \in T_k} T_k d(i, C_k) \dots$$

Input: k: the number of clusters. D: training dataset containing n objects.

Output: A set of group of clusters

Step 1: Chooses two mean values from weightage of significant patterns as initial cluster centers;

Step 2: Assign each object to the cluster to which it is most

similar based on the mean value of the weight age.

Step 3: Update the cluster means by calculating the mean value of all the objects in the cluster.

Given K: number of clusters of a dataset containing a set of N entities; I, and M measurements; Here T = T diabetes types; C-Class;

k-means Clustering based Diabetes Detection

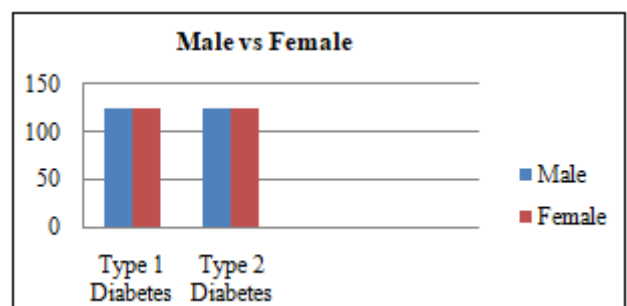
This method represents the patients total population distribution at computer system to reflect the impact of genetic factors of family history of diabetes. The weighted average value calculated using Weight of k-means of each parameter of user defined, addition to that 'I' user selected parameter as Blood group and Weight Loss.

6. Results

Following table shows the distribution of samples based on several background variables.

Table 1: Distribution of sample based on types of Diabetes

S. No.	Type of Diabetes	Male	Female	Total
1.	Type-I	125	125	250
2.	Type-II	125	125	250



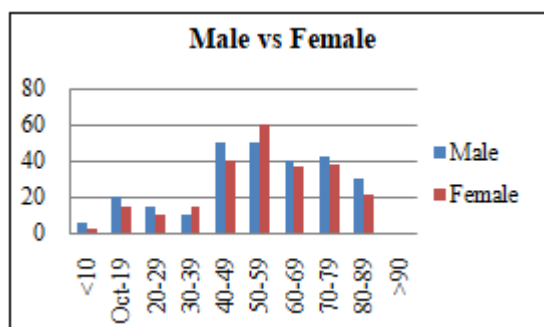
Graph 1: Sample distribution based on types of Diabetes

Correlation of Age and Sex of Diabetic patients

Through this study we can easily correlate the various age groups of diabetic patients with respect to their sex.

Table 2: Correlation of Age and Sex of Diabetes patient

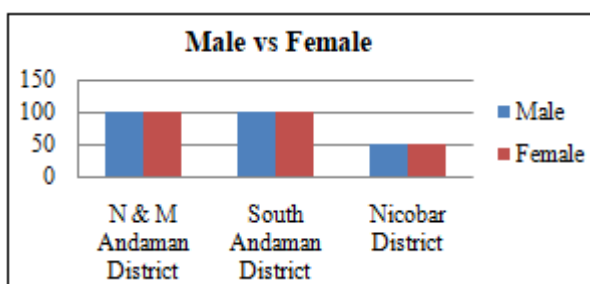
Age	Male	Female	Total
<10	5	2	7
10-19	20	15	35
20-29	15	10	25
30-39	10	15	25
40-49	50	40	90
50-59	50	60	110
60-69	40	37	77
70-79	42	38	80
80-89	30	21	51
>90	0	0	0



Graph 2: Sample distribution based on Age Group

Table 3: District-wise sample distribution

District	Male	Female	Total
North and Middle Andaman District	100	100	200
South Andaman District	100	100	200
Nicobar District	50	50	100



Graph 3: Sample distribution based on types of Localities/Districts

7. Summary of Findings

It is found that the males are more affected by diabetes as compared to females and also it is observed that South Andaman and North and Middle Andaman Districts have more number of diabetes patients as compared to Nicobar District. It is also observed that a multi dimensional diabetes data warehouse has also been built and used several data mining techniques to process and fetch the data as and when required.

8. Conclusions

There's no cure for type 1 diabetes. It requires lifelong disease management. But with consistent monitoring and adherence to treatment, you may be able to avoid more serious complications of the disease.

If you work closely with your doctor and make good lifestyle choices, type 2 diabetes can often be successfully managed.

An implementation of our new system to expose the diabetic risk factors and to ensure that people are provided with the information and support they need to adopt in a healthy lifestyles.

Four basic components of diabetes control-prevention, early detection, diagnosis, treatment and painkilling care-thus avoid and cure many diabetes, as well as palliative the suffered patients. Diabetes control aims to reduce the incidence or instance, morbidity and mortality of diabetes and to improve the quality of life of diabetes patients in a defined population, through the systematic implementation of evidence. An implementation of our new system to expose the diabetes risk factors and to ensure that people are provided with the information and support they need to adopt in a healthy lifestyles.

Diabetes detection and prevention is still a challenging for the upgraded and modern medical technology. After researching a lot of statistical analyses which is based on those people who are affected in various diabetes types are based on some general risk factors and symptoms have been discovered.

References

- [1] Amira Hassan Abed and Mona Nasr (2019), "Diabetes Disease Detection through Data Mining Techniques", Int.J. Advanced Networking and Applications, Volume:11 Issue: 01Pages:4142-4149 (2019)ISSN:0975-0290.
- [2] Seyed Ataaldin Mahmoudinejad Dezfuli et. al. (2019), "Early Diagnosis of Diabetes Mellitus Using Data Mining and Classification Techniques", Jundishapur Journal of Chronic Disease Care, Volume 8, Issue3.
- [3] Sneha N. and Tarun Gangil (2019), "Analysis of diabetes mellitus for early prediction using optimal features selection", Article No. 13.
- [4] Deepti Sisodia and Dilip Singh Sisodia (2018), "Prediction of Diabetes using Classification Algorithms", Procedia Computer Science, Volume 132, Pages1578-1585.
- [5] Deeraj Shettyet. al. (2017), "Diabetes disease prediction using data mining", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), INSPEC Accession Number: 17558894, DOI: 10.1109/ICIIECS.2017.8276012 (IEEE).
- [6] Nimna Jeewandar et. al. (2017), "Data Mining Techniques in Prevention And Diagnosis of Non Communicable Diseases", international journal of Research in Computer Applications and Robotics,

Volume 5, Issue, 11, Pp. 11-17, ISSN2320-7345.

- [7] Loannis Kavakiotis et. al. (2017), "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal (Elsevier), Volume15, Pages104-116.
- [8] Harleen and Bhambri Pankaj (2016), "A Prediction Technique in Data Mining for Diabetes Mellitus", Apeejay-Journal of Management Sciences and Technology, Volume 4, Issue 1, ISSN-2347-5005.
- [9] Shivakumar B and Alby S (2014), "A survey on data-mining technologies for prediction and diagnosis of diabetes", International Conference on Intelligent Computing Applications, ICICA2014 (2014) 167-173
- [10] Ravneet Jyot Singh and Williamjeet Singh (2012), "Data Mining in Health care for Diabetes Mellitus", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064
- [11] Miroslav Marinov et. al. (2011), "Data- Mining Technologies for Diabetes: A Systematic Review", Journal of Diabetes Science and Technology, Volume 5, Issue 6.
- [12] <https://www.indiatoday.in/education-today/gk-current-affairs/story/98-million-indians-diabetes-2030-prevention-1394158-2018-11-22>