Prediction of Diabetes Influcend Kidney Disease Using Data mining

S.Irfan banu[1], Mr. G. Ananthnath[2]

[1]Student, Dept. of Computer Applications, KMM Inistitute of PG Studies, Tirupathi, India

[2]Assistant Professor, Dept. of Computer Applications, KMM Inistitute of PG Studies, Tirupathi, India

Abstract- **Data mining can be viewed as a results of the normal development of data technology. Data is the process of discovering interesting patterns and information from large amounts of knowledge. The info sources will embrace databases, information warehouses, the Web, alternative data repositories or information that are streamed into the system dynamically. Data mining brings a collection of tools and techniques that may be applied to processed medical information to get hidden patterns that offer health care professionals a further supply of data for creating decisions. This survey analyzed totally different papers during which one or a lot of machine learning algorithms of knowledge mining used for the prediction of diabetes disease. Applying data mining techniques in medical field particularly in diabetes disease treatment will offer reliable performance.**

Index Terms- **Data mining, machine learning algorithm, diabetes disease.**

I. INTRODUCTION

Data mining, *the extraction of hidden predictive information from large databases,* is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations

The most commonly used techniques in data mining are:

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) .
- Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k $\geq$ 1). Sometimes called the k-nearest neighbor technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. These capabilities are now evolving to integrate directly with industry-standard data warehouse and OLAP platforms. The appendix to this white paper provides a glossary of data mining terms.

Diabetes mellitus is characterized by abnormally high levels of sugar (glucose) in the blood.

When the amount of glucose in the blood increases, e.g., after a meal, it triggers the release of the hormone insulin from the pancreas. Insulin stimulates muscle and fat cells to remove glucose from the blood and stimulates the liver to metabolize glucose, causing the blood sugar level to decrease to normal levels.

In people with diabetes, blood sugar levels remain high. This may be because insulin is not being produced at all, is not made at sufficient levels, or is not as effective as it should be. The most common forms of diabetes are type 1 diabetes (5%), which is an autoimmune disorder, and type 2 diabetes (95%), which is associated with obesity. Gestational diabetes is a form of diabetes that occurs in pregnancy, and other forms of diabetes are very rare and are caused by a single gene mutation.

For many years, scientists have been searching for clues in our genetic makeup that may explain why some people are more likely to get diabetes than others are. "The Genetic Landscape of Diabetes" introduces some of the genes that have been suggested to play a role in the development of diabetes.

Diabetes is classified by underlying cause. The categories are: type 1 diabetes—an autoimmune disease in which the body's own immune system attacks the pancreas, rendering it unable to produce insulin; type 2 diabetes—in which a resistance to the effects of insulin or a defect in insulin secretion may be seen; gestational diabetes; and "other types"

Type 2 diabetes commonly occurs in adults who are obese. There are many underlying factors that contribute to the high blood glucose levels in these individuals. An important factor is the body's resistance to insulin in the body, essentially ignoring its insulin secretions. A second factor is the falling production of insulin by the beta cells of the pancreas. Therefore, an individual with type 2 diabetes may have a combination of deficient secretion and deficient action of insulin.

In contrast to type 2, type 1 diabetes most commonly occurs in children and is a result of the body's immune system attacking and destroying the beta cells. The trigger for this autoimmune attack is not clear, but the result is the end of insulin production.

In this paper, we can also predict one more disease i.e., Chronic kidney disease (CKD) is normal and related to the expanded danger of cardiovascular ailment and end-arrange renal infection, which are possibly preventable through early recognizable proof and treatment of people in danger. In spite of the fact that hazard factors for event and movement of CKD have been distinguished, their utility for CKD chance stratification through forecast models stays indistinct. We basically surveyed chance models to anticipate CKD and its movement and assessed their appropriateness for clinical utilize.

CKD is extremely normal and is for the most part connected with maturing. The more seasoned you get, the more probable you are to have some level of a kidney ailment. It is assessed that one out of five men and one out of four ladies between the ages of 65 and 74 have some level of CKD.

The most widely recognized reason for CKD is harm caused by other constant conditions, for example, (hypertension) and diabetes.

CKD is more typical in individuals of the south Asian starting point (those from India, Bangladesh, Sri Lanka and Pakistan) and dark individuals than the all-inclusive community. The purposes behind this incorporate higher rates of diabetes in south Asian individuals and higher rates of hypertension in African or Caribbean individuals

ALGORITHM:

MACHINE LEARNING:
This is the algorithm part of the data mining process. It provides computers with the ability to learn without being explicitly programmed. This taxonomy or way of organizing machine learning algorithms is useful because it forces us to think about the the roles of the input data and the model preparation process and select one that is the most appropriate for our problem in order to get the best result.

Supervised Learning:
Input data is called training data and has a known label or result. A model is prepared through a training process where it is required to make predictions and is corrected when those predictions are wrong. The training process continues until the model achieves a desired level of accuracy on the training data.

Unsupervised Learning:
Input data is not labeled and does not have a known result. A model is prepared by deducing structures present in the input data. This may be to extract general rules. It may through a mathematical process to systematically reduce redundancy, or it may be to organize data by similarity.

Semi-Supervised Learning:

Input data is a mixture of labelled and unlabelled examples. There is a desired prediction problem but the model must learn the structures to organize the data as well as make predictions.

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbour method etc., are used for knowledge discovery from databases.

Classification:

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large.

Clustering:

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.

Predication:

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables.

Association rule:

Association and correlation is usually to find frequent item set findings among large

Neural networks :

Neural network is a set of connected input/output units and each connection has a weight present with it. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.

CONCLUSION

In this project, we are used different data mining techniques that can be engaged in computerized diabetes disease calculation systems and also chronic kidney diseases. Here different systems and different data mining classifications are difine in this system. Machine learning algorithm is used in that supervised, unsupervised, semi-supervised learning algorithms are there. By this algorithm to provide efficient and effective diabetes disease calculation and chronic kidney disease calculation. Here different techniques and classifications are used compared with other systems. So, this system will be provide effeciant accuracy.

REFERENCES

[1] American Diabetes Association: Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. Diabetes Care 26 (Suppl. 1):S51-S61, 2003

[2] American Diabetes Association: Nutrition recommendations and interventions for diabetes: a position statement of the American Diabetes Association. Diabetes Care 31 (Suppl 1):S61- 78, 2008

[3] American Diabetes Association: Diabetes Care – MNT. Diabetes Care 35 (Suppl. 1):S21-24, 2012.

[4] American Dietetic Association. Medical Nutrition Therapy Evidence-Based Guides to Practice. Nutrition Practice Guidelines for Type 1 and 2 Diabetes Mellitus, CD-ROM. 2001

[5] Delahanty LM, et al.: The role of diet behaviors in achieving improved glycemic control in intensively treated patients in the Diabetes Control and Complications Trial. Diabetes Care 16:1453-1458, 1993

[6] Franz MJ, et al.: Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. Diabetes Care 25:148-198, 2002

[7] Franz MJ, et al.: Cost-effectiveness of medical nutrition therapy provided by dietitians for persons with non-insulin-dependent diabetes mellitus. J Am Diet Assoc 95:1018-1024, 1995

[8] Franz MJ, et al.: Effectiveness of medical nutrition therapy provided by dietitians in the management of non-insulin-dependent diabetes mellitus: A randomized, controlled clinical trial. J Am Diet Assoc 95:1009-1017, 1995

[9] Padgett D, et al.: Meta-analysis of the effects of educational and psychosocial interventions on

the management of diabetes mellitus. J Clin Epidemiology 41:1007-1030, 1988

[10] American Dietetic Association, Position of the American Dietetic Association: Weight Management. J Am Diet Assoc, Vol. 109, Number 2: 330-346, 2011

[11] Gumbiner B: The treatment of obesity in type 2 diabetes mellitus. Prim Care 26(4):869-883, 1999

[12] National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health: Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults, 1999

[13] Thompson David, et al.: Estimated economic costs of obesity to U.S. business. American Journal of Health Promotion 13(2):120-127, 1998

[14] Wing RR, et al.: Caloric restriction per se is a significant factor in improvements in glycemic control and insulin sensitivity during weight loss in obese NIDDM patients. Diabetes Care 17(1): 30-36, 1994

[15] American Diabetes Association (Position Statement). Standards of medical care in diabetes. Diabetes Care 35 (Supplement 1): S4-S49, 2012

[16] Diabetes Prevention Program Research Group. The Diabetes Prevention Program: Description of lifestyle intervention. Diabetes Care 25: 2165-2171, 2002

[17] Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA & Nathan DM. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. New England Journal of Medicine,346: 393-403, 2002

[18] Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R,

[19] Zinman B. Impaired Fasting Glucose and Impaired Glucose Tolerance: Implications for care. Diabetes Care 30: 753-759, 2007

[20] Looijmans-Van Den Akker I, et al.: Clinical effectiveness of first and repeat influenza vaccination in adult and elderly diabetic patients. Diabetes Care 29:1771-1776, 2006