ANALYSIS OF COMMON DISEASES AMONG THE PEOPLE USING DATAMINING APPROACH

Soundaryajothi S[1], Annalakshmi V[2] , Ramalakshmi T[3]
[12]PG Student, [2]Assistant Professor
[1 2 3] Dept. of Computer Science,
[1 2 3] Jayaraj Annapckiam College for Women (Autonomous), Periyakulam-625 601, Tamil Nadu, India

Abstract—The analysis on "Common Diseases among the people above 30" is a study to mention the most frequently occurred diseases above the age people of 30. In early days, people had faced few diseases only at the age of 50. But in this generation, people are affecting early at or above the age of 30. The important instance to be noticed is that, the age limit had been decreased from 50 to 30 due to the major changes those we had adopted in our day today life. The factors which are considered as the main root cause to occur diseases in human body are lifestyle, food habits, work style, etc. Diseases may also occur among people due to their family situations, their own family tradition, some heredity reasons, the individual attitude based, environment, surroundings, work place, etc. The main intention of this case study is to learn the most commonly occurred diseases in this computer world.

Keywords— Data Mining, Apriority Algorithm, Diabetes, Blood Pressure, Stress, and Ulcer.

INTRODUCTION

The term disease broadly refers to any condition that impairs the normal functioning of the body. For this reason, diseases are associated with disfunctioning of the body's normal homeostatic process. Commonly, the term disease is used to refer specifically to infectious diseases, which are clinically evident diseases that result from the presence of pathogenic microbial agents, including viruses, bacteria, fungi, protozoa, multi cellular organisms, and aberrant proteins known as prions. An infection that does not and will not produce clinically evident impairment of normal functioning, such as the presence of the normal bacteria and yeasts in the gut, or of a passenger virus, is not considered a disease. By contrast, an infection that is asymptomatic during its incubation period, but expected to produce symptoms later, is usually considered a disease. Non-infectious diseases are all other diseases, including most forms of cancer, heart disease, and genetic disease. Chronic disease - disease that is a long term issue Incurable disease - disease that cannot be cured, also known as pathological disease Terminal disease - disease with death as an inevitable result of it In this paper we have done a detailed study on the Analysis of Common Diseases among people above 30 using Data mining approaches.

HISTORICAL BACKGROUND

A disease is considered rare when it affects one person out of 2000 or less. They are between 5000 and 8000 rare diseases, most of them genetic. A very rough estimate would be that in the world, one person out of 15 could be affected by a rare disease, this represents 400 million people worldwide of which 30 million Europeans and 25 million Americans. Rare diseases are serious chronic diseases, and are often life-threatening. While most genetic diseases are rare diseases, around 20% of rare diseases are not caused by genetic defects. There are very rare infectious diseases for instance, as well as auto-immune diseases and very rare poisonings. To date, the cause remains unknown for most rare diseases. This makes rare diseases truly a global health issue. In recent decades, considerable attention has been paid worldwide to stimulate the research, development and marketing of medicinal products for rare diseases. In the United States over 400 products have been approved as therapy in more than 200 rare diseases indications and in EU over 70 products for around 45 indications. Many orphan medicinal products are innovative, biotechnological products. Apart from treatments coming available, the introduction of various (research) programs and networks has advanced understanding and diagnosis of rare diseases as well.

Despite this positive development overall the rare disease burden continues to persist. This persistence is due to lack of knowledge/training, lack of or delayed diagnosis, limited disease understanding, lack of treatment, and lack or limited access to therapy or medical care. Being a complex and heterogeneous mosaic of an estimated 5 000-8 000 conditions, it has become clear that the research need can differ considerably between groups of rare diseases.

PRESENT CONDITION

At present the most commonly occurred diseases among few common diseases are Diabetes, Blood Pressure, Stress, and Ulcer.

Diabetes Mellitus: Type 1 and Type 2

In type 1 diabetes, the body does not produce insulin, and daily insulin injections are required. Type 1 diabetes is usually diagnosed during childhood or early adolescence and it affects about 1 in every 600 children. Type 2 diabetes is the result of failure to produce sufficient insulin and insulin resistance. Elevated blood glucose levels are managed with reduced food intake, increased physical activity, and eventually oral medications or insulin.

CAUSES:

Type 1 Diabetes

 • Caused by the immune destruction of the beta cells of the pancreas.

 • Antibodies to islet cells and insulin are present at diagnosis.

 • Insulin secretion gradually diminishes.

 • May present at any age, but most common in childhood and adolescence.

 • Insulin by injection is necessary for survival.

 • Contributing factors:

 – Genetic predisposition

 – Environmental triggers (infection or other stress)

Type 2 Diabetes

• Caused by insulin resistance in the liver and skeletal muscle, increased glucose production in the liver, over production of free fatty acids by fat cells and relative insulin deficiency.

• Insulin secretion decreases with gradual beta cell failure.

• Reductions in blood glucose levels often can be achieved with changes in food intake and physical activity patterns.

Oral medication and/or insulin injections are eventually required.

• Contributing factors:

– Obesity

– Age (onset of puberty is associated with increased insulin resistance)

– Lack of physical activity

– Genetic predisposition

– Racial/ethnic background

– Conditions associated with insulin resistance, (e.g., polycystic ovary syndrome)

BLOOD PRESSURE

In the majority of cases—over 90 percent—no specific cause for the elevated blood pressure can be identified. In this case, the elevated blood pressure is referred to as primary or essential hypertension. Some researchers believe that this type of high blood pressure may be due to hormonal factors relating to the handling of salt by the kidneys and/or to the elaboration of certain substances that cause constriction of blood vessels. These are probably genetically determined, but certain environmental factors, such as a high-salt, low-potassium diet and chronic stress, may play some role. In up to 10 percent of patients, high blood pressure may be a consequence of another disorder, or a side effect of medication. This type of hypertension is referred to as secondary hypertension. Lt is important to remember that these cases are relatively uncommon. However, some of the more common causes of secondary hypertension include the following.

Causes:

These are probably genetically determined, but certain environmental factors, such as a high-salt, low-potassium diet and chronic stress, may play some role. In up to 10 percent of patients, high blood pressure may be a consequence of another disorder, or a side effect of medication. This type of hypertension is referred to as secondary hypertension. It is important to remember that these cases are relatively uncommon.

STRESS:

Stress is actually a normal part of life. At times, it serves a useful purpose. Stress can motivate you to get that promotion

at work, or run the last mile of a marathon. But if you don't get a handle on your stress and it becomes long-term, it can seriously interfere with your job, family life, and health. More than half of Americans say they fight with friends and loved ones because of stress, and more than 70% say they experience real physical and emotional symptoms from it.

Read on to learn why you get stressed out, and how that stress might be affecting your health.

Causes:

Everyone has different stress triggers. Work stress tops the list, according to surveys. Forty percent of U.S. workers admit to experiencing office stress, and one-quarter say work is the biggest source of stress in their lives.

Causes of work stress include:

- Being unhappy in your job
- Having a heavy workload or too much responsibility
- Working long hours
- Having poor management, unclear expectations of your work, or no say in the decision-making process
- Working under dangerous conditions
- Being insecure about your chance for advancement or risk of termination
- Having to give speeches in front of colleagues
- Facing discrimination or harassment at work, especially if your company isn't supportive
- Life stresses can also have a big impact. Examples of life stresses are:
- The death of a loved one
- Divorce
- Loss of a job
- Increase in financial obligations
- Getting married
- Moving to a new home
- Chronic illness or injury
- Emotional problems (depression, anxiety, anger, grief, guilt, low self-esteem)
- Taking care of an elderly or sick family member

ULCER:

Stomach ulcers are painful sores that can be found in the stomach lining or small intestine. Stomach ulcers are the most visible sign of peptic ulcer disease. They occur when the thick layer of mucus that protects your stomach from digestive juices is reduced, thus enabling the digestive acids to eat away at the lining tissues of the stomach. Stomach ulcers are easily cured, but they can become severe without proper treatment.

Causes:

Stomach ulcers aren't necessarily caused by one single factor. The decrease in the stomach's mucus lining that leads to an ulcer is usually caused by one of the following:

- an infection with the bacterium Helicobacter pylori (H. pylori)
- long-term use of non-steroidal anti-inflammatory drugs (NSAIDs), such as aspirin and ibuprofen
- excess acid (hyperacidity) in the stomach, which may be related to genetics, lifestyle (stress, smoking), and certain foods
- Zollinger-Ellison syndrome, a rare disease that makes the body produce excess stomach acid
- Certain factors and behaviors can put you at higher risk for developing stomach ulcers:
- smoking
- frequent use of steroids (such as those for treating asthma)
- hypocalcaemia (overproduction of calcium)
- family history of stomach ulcers
- excessive consumption of alcohol

SIGNIFICANCE OF THE STUDY

1) Understanding how a patient's social history (travel, HIV risk factors, and exposures) can have a significant impact on the differential diagnosis and management of infections.

2) Understanding antibiotic selection and therapy including familiarity with major classes, choosing appropriate antibiotics and monitoring for antibiotic toxicities

3) Exposure to a broad range of major syndromes including community and hospital-acquired pneumonia, infective endocarditis, cellulitis, urinary tract infections and the evaluation of fever.

4) Appropriate use of diagnostic services including gram stain and culture, anti microbial sensitivity testing and other standard microbiology lab techniques

5) Understanding basic principles of infection control such as contact or respiratory isolation and contact tracing.

6) Exposure to critically ill patients and immune suppressed patients, as well as an understanding of their specific spectrum of diseases

OBJECTIVES OF THE STUDY

The objective of the study is to investigate the common diseases among the people of above 30. This is a case study of magnitude and listing common diseases in their earlier age.

- To analyze the common diseases among people above 30.
- To mention the frequently occurred diseases among the people.
- To provide the causes and symptoms of those commonly occurred disease.

Null hypothesis

In this case study, we assume some hypothesis. These hypotheses are based only the factors of common diseases occur to the people above the age 30. There is no significant difference in the diseases of the people based on gender with regarded to their age

- There is no significant difference in the causes of the diseases among the people
- There is no significant reason except food habits, lack of physical activity, etc.

RESEARCH METHODOLOGY

Methodology

The present study is an exploratory research conducted among the people in Then. In order to pursue the aims and objectives outlined in the introduction, a content analysis of information gained from a multimedia research process was conducted to establish the underlying trends in location to find common diseases. The first stage involved gathering of secondary information from people. The second stage involved identifying the age group among them and structuring a comparative analysis of the five identified parameters under each category. A summary of interpretations was also given. In the third stage, analysis was carried out by making specific assumptions in a hypothetical situation. In the last and the fourth stage, on the basis of the results and interpretations, specific postulates were framed, and on each postulate

hypotheses were framed that can be tested through quantitative research in the future. The above-mentioned stages have been described as objectives in the preceding paragraph.

Algorithm used

Data mining is the core process of knowledge discovery in database. It is the process of extraction of useful patterns from the large database. To analyze the large amount of collected information, the area of Knowledge Discovery in Database (KDD) provides techniques which extract interesting patterns in a reasonable amount of time .data mining is the application of efficient algorithms to detect the desired patterns contained within the given data. Data mining is the extraction of hidden descriptive or predictive information from large databases.

Association Rule Mining

Association rules mining are one of the major techniques of data mining. The purpose of association analysis is to figure out the hidden association and some useful rules of data base, and uses these rules to speculate and judge the unknown matter from the already known information. Association rule mining has many important applications in our life.

Association Rule

An association rule is one of the forms x=>y. and each rule has two basic needs: support and confidence. Things that occur often together can be associated to each other. These together occurring things form a frequent itemset. Conclusions based on the frequent itemsets make association rules.

APRIORI ALGORITHM

Apriori algorithm is a fundamental algorithm mining association rule. It contains two processes: • Detect all frequent itemsets by scanning db. • Form strong association rules in the frequent itemsets. Process one needs to scan DB several times, which consumes a lot of time and space. As a result, what needs to be improved is the mining competency of frequent group of things in DB. Apriori algorithm is a significant algorithm for mining frequent itemsets for Boolean association rules. Apriori algorithm is formed by Agrawal and Srikantin 1994. It is the most fundamental and important algorithm for mining frequent itemsets. Apriori is used to detect all frequent itemsets in a provided database

db.Thekeynote of Apriori algorithm is to form multiple passes over the database. It employs an repetitive approach called as a breadth-first search (level-wise search).

Key Concepts

• Frequent Item sets: The itemsets which has minimum help (denoted by li for ith-itemsets), Apriori property: any subgroup of frequent things must be frequent.

• Join Operation: To detect lk, a group of candidate k- group of things is developed by adding lk-1with itself.

How Apriori Works? Find All Frequent Itemsets.

Get Frequent Things: Things whose occurrence in database is more than or equal to the minimum help threshold.

Frequent Itemsets: Develop candidates from frequent things. Prune the results to detect the frequent itemsets. Develop strong association rules from frequent itemsets.Rules which satisfy the minimum support and minimum confidence threshold.

Association Rule:

Association rule of data mining involves picking out the unknown inter-dependence of the data and finding out the rules between those items [3]. Agrawal introduced association rules for point of sale (POS) systems in supermarkets. A rule is defined as an implication of the form A=>B, where A∩ B≠Ǿ. The left-hand side of the rule is called as antecedent. The right-hand side of the rule is called as consequent.

Support: I = { i1,i2,i3, … , im} is a collection of items. T be a collection of transactions associated with the items. Every transaction has an identifier TID [6]. Association rule A=>B is such that A Є I, B Є I.

A is called as Premise and B is called as Conclusion. The support ,S, is defined as the proportion of transactions in the data set which contains the item set.

Support(X=>Y) = Support (XUY) = P(XUY).

Confidence: The confidence is defined as a conditional probability

Confidence (X=>Y) = Support (XUY) / Support(X) = P(Y/X). Lift: is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e. lift = Pr(L,R) / Pr(L).Pr(R).
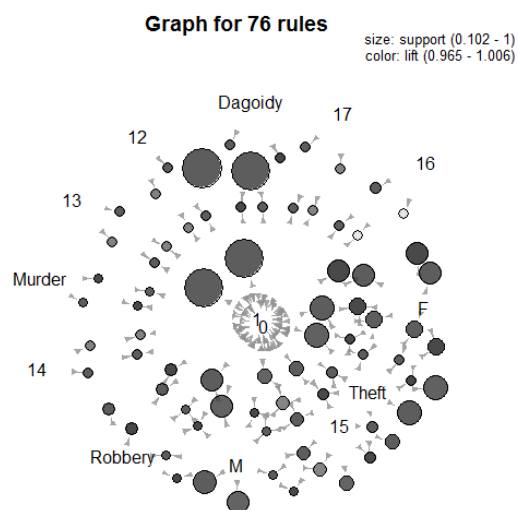
ANALYSIS OF DATA

Data Mining Definition:

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information-information that can be used to increase revenue, cuts costs, or both. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.
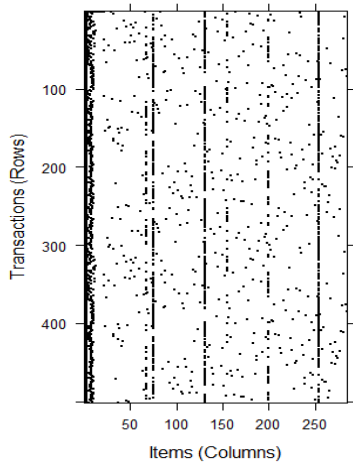
Analysis of data

The researcher collects more than 500 responses from samples all over Theni. She spent nearly 2 months to collect the fresh dada from end users. After collecting the information, all the details are fed into the software and checked for outlier. The cleaned data was analyzed using single attribute and multiple attributes. Name, Age, Gender. To process the data, I have installed the following libraries such as rules, a rules Viz, from cloud storage. Then the dataset was inserted into the R- tool for processing.
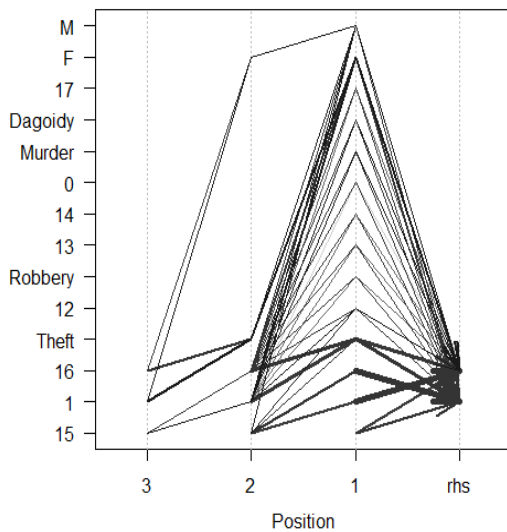
RESULTS AND DISCUSSION

This chapter discusses and analyzes the implementation results of the proposed work. The snapshot of the implementation details of the methodology are tested and evaluated. The experimental results are discussed in detail. The methodology consist of 3 phases, the results of each phase are discussed in the following sections.

Graph for 76 rules

Parallel coordinates plot for 76 rules

their work, interacting with their family members in their leisure time especially with the kids will deviate us from the mental pressure they have.

RECOMMENDATIONS

The case study on "Common Diseases Among People Above 30 Using Data Mining Approaches" results that, now a days the people above age 30 is commonly being affected by ulcer and stress caused by tension in their daily life. The major reasons for the occurrence of ulcer among people are their change in their recent life style. This means that the machine life had changed their life style which leads to such diseases among people. The food they take daily is also a reason for ulcer. Somebody will skip their food or improper intake of food leads to ulcer which causes stress in their machine life. The work tension and the mental pressure in each and every situation of their lives are also the major reasons for the disease. And finally to avoid these problems, people should take care of their health by proper intake of healthy food, reducing tension in their work, refreshing themselves from

BIBLIOGRAPHY

1. "Disease" at Dorland's Medical Dictionary "What is the deadliest disease in the world?". WHO. May 16, 2012. Retrieved December 7, 2014.

2. "Mental Illness—Glossary". US National Institute of Mental Health. Retrieved 2010-04-18.

3. "Regents Prep: Living Environment: Homeostasis". Oswego City School District Regents Exam Prep Center. Retrieved 2012-11-12.

4. "Illness" at Dorland's Medical Dictionary

5. Emson HE (April 1987). "Health, disease and illness: matters for definition". CMAJ 136 (8): 811–3. PMC 1492114. PMID 3567788.

6. McWhinney IR (April 1987). "Health and disease: problems of definition". CMAJ 136 (8): 815. PMC 1492121. PMID 3567791.

7. Hart BL (1988). "Biological basis of the behavior of sick animals". Neurosci Biobehav Rev 12 (2): 123–137. doi:10.1016/S0149-7634(88)80004-6. PMID 3050629.

8. Johnson R (2002). "The concept of sickness behavior: a brief chronological account of four key discoveries". Veterinary Immunology and Immunopathology 87 (3–4): 443–450.doi:10.1016/S0165-2427(02)00069-7. PMID 12072271.

9. Kelley KW, Bluthe RM, Dantzer R, Zhou JH, Shen WH, Johnson RW, Broussard SR (2003). "Cytokine-induced sickness behavior". Brain Behav Immun 17 (Suppl 1): S112–118. doi:10.1016/S0889-1591(02)00077-6. PMID 12615196.