Health Data Analytics with Cloud Based Solutions

Factors Affecting to Infect Cirrhosis

Submitted by:

142063 - A.M.U.Ishanika

142072 - M.L.S.K.Mahawela

142154 - K.M.M Madusarani

Supervised by: Mr. J.R.K.C. Jayakody

Department of Computing and Information Systems

Wayamba University of Sri Lanka

Kuliyapitiya

Table of Contents

1. ABSTRACT

Cloud computing has increased hospital flexibility in order to respond for dynamic changes and latest medical updates, in addition to demonstrate a great healthcare value by reducing costs, increasing productivity and security and improve data analysis with minimal management effort or service provider interaction. Cirrhosis is scarring of the liver as a result of long-term, continuous damage to the liver. There are many different causes of cirrhosis, but excessive alcohol consumption and tobacco consumption are the most common causes.

The objective of this paper is to review the factors' behavior which is affecting to the Cirrhosis cases which is showing symptoms. For this project dependent variable was number of cases which have the symptoms of Cirrhosis and age, year, Alcohol consumption, Tobacco consumption were the independent variables. This project will help to find whether user infect cirrhosis or not, when the user enters his/her age, gender, amount of alcohol per day and amount of tobacco per day and give the accurate result of probability of infect to cirrhosis and focusing the factors to reduce the use of alcohol and tobacco. Throughout this paper it's clear that, Alcohol liters per resident and Tobacco equivalents per resident are the main factors which is affected to the number of cases which shows the symptoms of cirrhosis. The best multiple regression model for cirrhosis cases which shows symptoms was fitted by using the analysis. The model satisfy the assumptions of the multiple linear regression. The model says that, if increase the alcohol consumption and tobacco consumption, the cirrhosis cases will be also increased.

2. INTRODUCTION

The healthcare sector is widely considered as one of the most important industries in information technology (Wager 2005). More and more, information technology has been considered as a practice that facilitates healthcare performance through using data and information efficiently within the healthcare sectors.

Cloud computing has increased hospital flexibility in order to respond for dynamic changes and latest medical updates, in addition to demonstrate a great healthcare value by reducing costs, increasing productivity and security and improve data analysis with minimal management effort or service provider interaction. Cloud computing reduce strain which caused by huge amount of clinical data. One of the cloud innovations is Phillips Health Suite platform that manages healthcare data and support doctors and patients. Phillips Health suite platform stores a huge amount of clinical and patient data which can be used directly in the future as an actionable data, a source of diagnosis analysis and disease prediction and prevention to increase patient care.

Cirrhosis is scarring of the liver as a result of long-term, continuous damage to the liver. It is a serious condition where healthy tissue in the liver is destroyed and replaced by scar tissue, which starts to block the flow of blood through your liver. There are many different causes of cirrhosis, but excessive alcohol consumption and tobacco consumption are the most common causes. Less common causes include hepatitis B infection, an inherited liver disease (such as hemochromatosis), and a condition called non-alcoholic steatohepatitis (NASH). The number of people with cirrhosis appears to be increasing. If person is a heavy drinker, his chances of developing cirrhosis are increased. However, cirrhosis of the liver is not just a condition that affects people who are dependent on alcohol (alcoholics). Heavy alcohol and tobacco consumption are the most common causes of cirrhosis of the liver. Therefore, the best way to prevent alcohol-related cirrhosis developing is to stick to the recommended limits. From this project person can know whether he has a risk of getting exposed to cirrhosis or not and hence can stick into tobacco and alcohol limits.

3. LITERATURE REVIEW

1. Healthcare Data Analytics

Hoyt, RE, Yoshihashi, A, Eds. (2014). Health Informatics: Practical Guide for Healthcare and Information Technology Professionals, Sixth Edition. Pensacola, FL, lulu.com

This research base around applying analytics to improve healthcare delivery is still in its early stages. There is an emerging base of research that demonstrates how data from operational clinical systems can be used to identify critical situations or patients whose costs are outliers. There is less research, however, demonstrating how this data can be put to use to actually improve clinical outcomes or reduce costs.

Studies using Electronic Health Record (HER) data for clinical prediction have been proliferating. One common area of focus has been the use of data analytics to identify patients at risk for hospital readmission within 30 days of discharge. A number of other critical clinical situations have been amenable to detection by analytics applied to EHR and other clinical data:

• Predicting 30-day risk of readmission and death among HIV-infected inpatients

• Identification of children with asthma

• Risk-adjusting hospital mortality rates

• Detecting postoperative complications

• Measuring processes of care

• Determining five-year life expectancy

• Detecting potential delays in cancer diagnosis

• Identifying patients with cirrhosis at high risk for readmission

• Predicting out of intensive care unit cardiopulmonary arrest or death

Clearly there is great promise ahead for healthcare driven by data analytics. The growing quantity of clinical and research data, along with methods to analyze and put it to use, can lead to improve personal health, healthcare delivery, and biomedical research. However, there is also a continued need to improve the completeness and quality of data as well as conduct research to demonstrate

how to best apply it to solve real-world problems. In addition, human expertise, including in informatics, will be required to optimally carry out such work.

2. Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications

Lo'ai A. Tawalbeh1,2, (Senior Member, Ieee), Rashid Mehmood3, (Senior Member, Ieee), Elhadj Benkhlifa4, And Houbing Song5, (Member, Ieee), Received August 16, 2016, accepted September 11, 2016, date of publication September 26, 2016, date of current version October 15, 2016.

In this paper, they discussed networked healthcare systems and the role that mobile cloud computing and big data analytics play in its enablement. The motivation and development of networked health care applications and systems was presented along with the adoption of cloud computing in healthcare. A Cloudlet based Mobile Cloud Computing infrastructure to be used for healthcare big data applications was described. The techniques, tools, and applications of big data analytics were reviewed. Healthcare applications require large amounts of computational and communication resources, and involve dynamic access to large amounts of data within and outside the health organization. This was discussed to be the main motivation for networked healthcare system where big data such as patient records need to be analyzed in real time, and this can implement efficiently via cloud and mobile cloud systems.

An important trend to enable next generation networked healthcare systems would be the networking and integration of healthcare and other smart city systems. Networked health care systems will eventually be designed as sustainable enterprise systems which will be part of networked smart city information and operations systems. Modeling methods that leverage high performance computing and big data technologies will be required in designing such complex networked health care systems. Further studies are needed on the integration of mobile cloud computing and healthcare applications to design realistic networked health care systems that are able to provide personalized medicine, reduce health care costs and facilitate better clinical and operational processes.

3. Population Health Data Science

Tomás J. Aragón, http://www.phds.io, San Francisco, California, USA

In here introduces R a programming language and environment for statistical computing and graphics to public health epidemiologists and health care analysts conducting population health analyses. Recent graduates come prepared with a solid foundation in epidemiological and statistical concepts and skills. However, what is sometimes lacking is the ability to implement new methods and approaches they did not learn in school. This is more apparent today with the emergence of data science and the new field of population health data science (PHDS) the art and science of transforming data into actionable knowledge to improve health.

PHDS is a transdisciplinary, rapidly emerging field that integrates the expertise from public health and medicine, mathematics, statistics, computer science, decision sciences, health economics, behavioral economics and human-centered design. PHDS is the future of public health data analysis and synthesis, and knowledge integration. Knowledge integration is the management, synthesis, and translation of knowledge into decision support systems to improve policy, practice, and ultimately population health.

In contrast to custom-made tools or software packages, R is a suite of basic tools for statistical programming, analysis, and graphics. One will not find a "command" for a large number of analytic procedures one may want to execute. Instead, R is more like a set of high quality carpentry tools (hammer, saw, nails, and measuring tape) for tackling an infinite number of analytic problems, including those for which custom-made tools are not readily available or affordable. I like to think of R as a set of extensible tools to implement one's analysis plan, regardless of simplicity or complexity. With practice, one not only learns to apply new methods, but one also develops a depth of understanding that sharpens one's intuition and insight. With understanding comes clarity, focused problem-solving, creativity, innovation, and confidence.

The goal is not to be comprehensive in each topic but to demonstrate how R can be used to implement a diversity of methods relevant to PHDS. My hope is that more and more epidemiologists will embrace R to become epidemiologic data scientists, or at least, include R in their epidemiologic toolbox.

4. Analysis of Research in Healthcare Data Analytics

Al Khatib et al, Healthcare Data Analytics, Australasian Conference on Information Systems, 2015, Sydney.

They conclude and relate that to the reason of why papers were decreased in the last 2 years, because quantitative studies are very expensive and costly, as well as it is focusing on involving professionals in healthcare analytics systems rather than patients, which will effect on having many inaccurate studies results, and create many struggles with healthcare analytics and kills it's future, therefore researchers will stop publishing papers as they lost their spirit in healthcare analytics, so we recommend to involve individuals, as it would help society to adopt and improve healthcare data analytics systems, as well as running these systems efficiently and smoothly. Also, from this study we have concluded that if Europe and Asia can collaborate and work together, then they can both get benefited from each other, in order to serve healthcare analytics researches in specific, and scientific academics studies in general, because English language in Asia is not popular, so English language in Asia and qualitative techniques can be improved more and more.

Finally, this paper is proposing a technique that will promise to leverage large amount of healthcare data properly, since doctors and nurses will be able to determine diseases and risks easily like some certain types of cancer, diabetes and blood pressure, as well as provide needed treatment in the right time. Also, enhance doctor's decision-making process by defining better care, developing drugs and vaccines along with a better treatment plan in order to reach patient satisfaction. Moreover, proposed technique will add a benefit of identifying risks early and mitigate it as much as possible. However, this study will need to push both doctors and patients to adopt new technique and collaborate together to reach high level of connection between both medical staff and patients in order to keep the system up to date and gather high quality of data. Also, this study will need from the individual to develop themselves and keep tracking their health conditions, but the problem is how to handle this with older people who are less attached and hard to convince to adopt new healthcare technologies and tools, as they consider this as a medical care issue involving medical staff and excluding their role in the medical care process. Generally, in healthcare sectors data analytics is very important and essential topic, since all the previous benefits we mentioned could lead for better choice of medical care practice and prevent illnesses.

5.Improving Healthcare Using Big Data Analytics

Revanth Sonnati, International Journal of Scientific & Technology Research Volume 6, Issue 03, March 2017

In daily terms we call the current era, which can also be named as the era of Big Data in the field of Information Technology. Our daily lives in today‘s world is rapidly advancing never quenching one’s thirst. The fields of science, engineering and technology are producing data at an exponential rate leading to Exabyte(s) of data every day. Big data helps us to explore and re-invent many areas not limited to education, health and law. The primary purpose of this paper is to provide an in-depth analysis in the area of Healthcare using the big data and analytics. The main purpose is to emphasize on the usage of the big data which is being stored all the time helping to look back in the history, but this is the time to emphasize on the analyzation to improve the medication and services. Although, many big data implementations happen to be in-house development, this proposed implementation aims to propose a broader extent using Hadoop, which just happen to be the tip of the iceberg. The focus of this paper is not limited to the improvement and analysis of the data; it also focusses on the strengths and drawbacks compared to the conventional techniques available.

The very first stage of the process is collecting the data from various repositories, sources and storing it in Hadoop Distributed File System (HDFS). Data can be both structured and unstructured like clinical analysis, pharmaceutical purchases, patient history, reports, medical emergencies, health index, and social media content, wearable devices and so on. The impact on the quality can be measured with the collection process.

One of the essential budget-oriented talk is healthcare and its cost, mainly in the countries having very less economic status, areas with bad health hygiene, increasing births and aged people. In this study, we propose to analyze the healthcare using the big data analytics specifically to any given geographic location and the data available. In addition, moving towards the big data storage and solutions would provide an efficient solution in contrast to the traditional storage solutions. Any further research can easily extend the system to improve the facilities and services.

4. METHODOLOGY

Variables:

- **Dependent variable**

  Number of people have the symptoms of cirrhosis

- **Independent variables**
  - Alcohol liters per resident
  - Tobacco equivalents per resident
  - Year
  - Age
  - Project Function

Number of people have the symptoms of cirrhosis

$$=\beta 0+\beta 1*(\text{Alcohol liters per resident}) +\beta 2*(\text{Tobacco equivalents per resident}) + \beta 3*(\text{Year}) +\beta 4*(\text{Age})$$

Data Collection

Secondary data obtained for the Project.

Used Software

- R
- Shiny Board

5. RESULTS

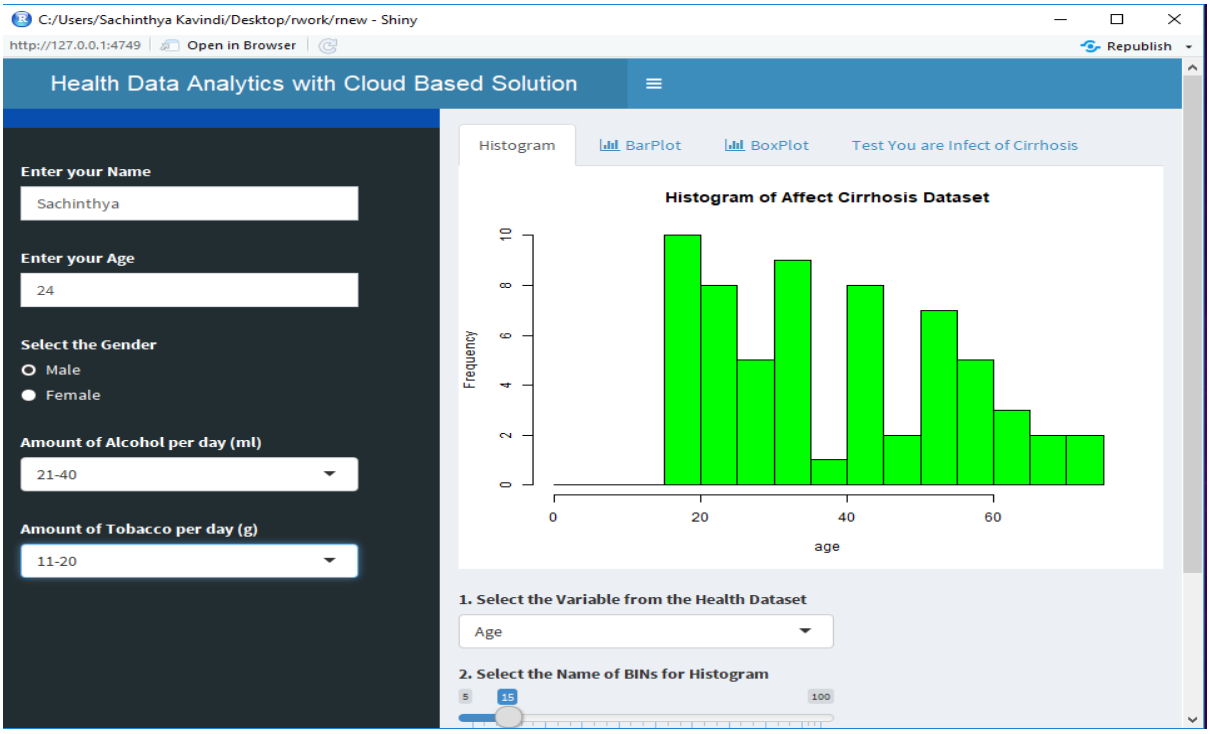

Figure 1: Histogram of Affect Cirrhosis

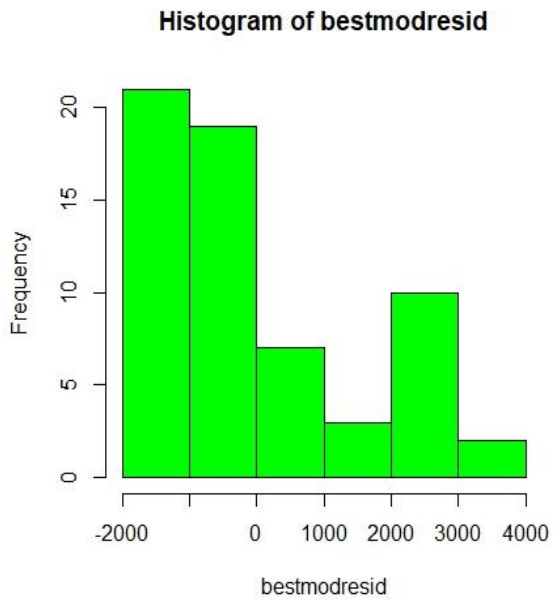

Figure 2: Histogram of the best model

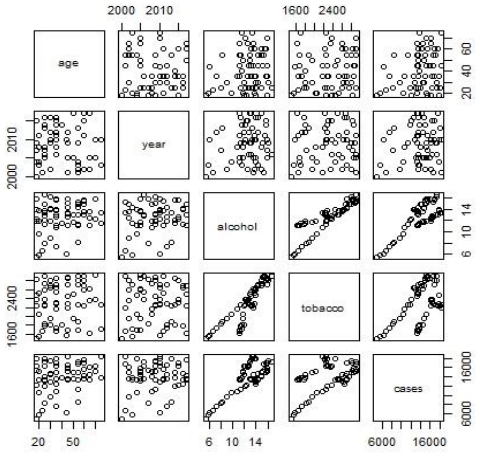

Figure 3: Linear Model

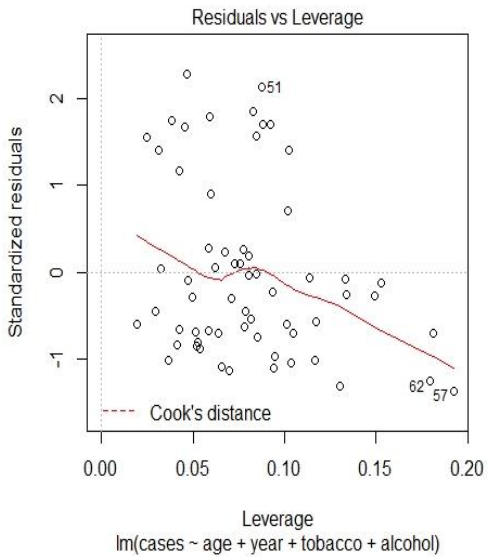

Figure 3: Effect on Alcohol

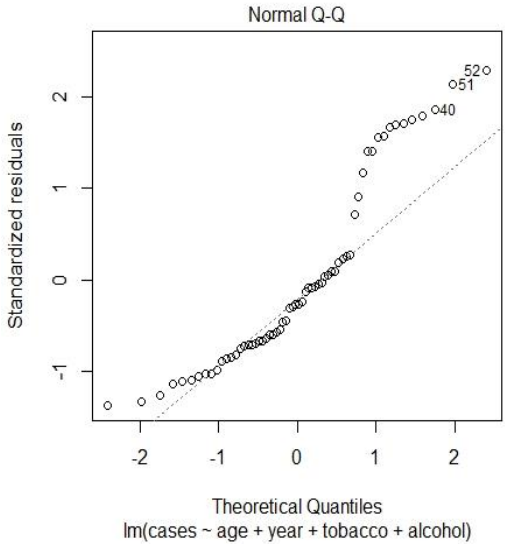

Figure 5: Effect on Tobacco

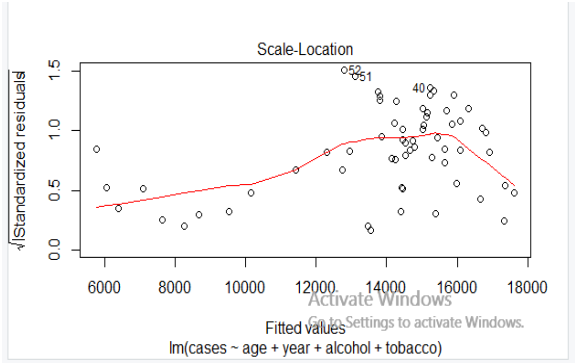

Figure 6: Effect on Age

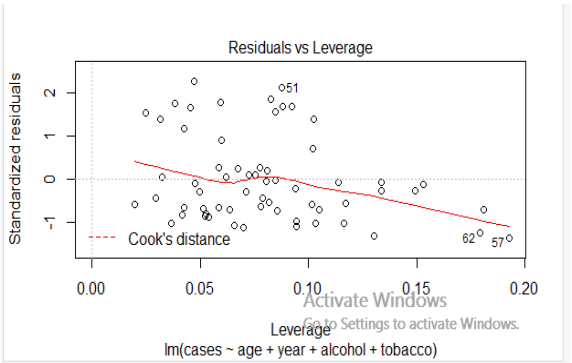

Figure 7: Effect on Year

6. DISCUSSION

This project was carried out to find the factors affecting to the number of cases which shows the symptoms of cirrhosis. The objective of this paper is to review the factors' behavior which is affecting to the Cirrhosis cases which is showing symptoms. For this project dependent variable was number of cases which have the symptoms of Cirrhosis and age, year, Alcohol consumption, Tobacco consumption were the independent variables. From analysis we can assume that most of the persons addicted to alcohol is age between 18-25. From less AIC value can fit the best model which is the combination of alcohol and tobacco are independent variables. Throughout this paper it's clear that, Alcohol liters per resident and Tobacco equivalents per resident are the main factors which is affected to the number of cases which shows the symptoms of cirrhosis.

7. CONCLUSION

The healthcare sector is widely considered as one of the most important industries in information technology. In this paper, we reviewed the factors mostly affect to the Cirrhosis death rate using analysis of different variables. From using user interface of this project, a person can know whether he will be getting exposed to Cirrhosis or not depending on the consumption of tobacco and alcohol as well as age and gender. Based on this review, we will be able to determine the factors affecting to the Cirrhosis and hence can reduce the usage of tobacco and alcohol.

In this project, consider four factors; age, year, alcohol consumption and tobacco consumption which is affecting to the cirrhosis cases which shows symptoms. Preliminary analysis shows that there is a relationship between cirrhosis cases and all the factors. But advanced analysis shows that cirrhosis death rate depends only on alcohol consumption and tobacco consumption.

The best multiple regression model for cirrhosis cases which shows symptoms was fitted by using the analysis. The model satisfy the assumptions of the multiple linear regression.

The model says that, if increase the alcohol consumption and tobacco consumption, the cirrhosis cases will be also increased.

8. REFERENCES

[1]. *Williom H.Samprit Chatterjee, Ali S.Hadi, Regression Analysis by Example, fifth edition, 2013, John Wiley and sons*

[2]. *Greene, Econometric Analysis, fifth edition, 2003, Pearson Education India.*

[3]. *Hoyt, RE, Yoshihashi, A, Eds. (2014). Health Informatics: Practical Guide for Healthcare and Information Technology Professionals, Sixth Edition. Pensacola, FL, lulu.com*

[4]. *Lo'ai A. Tawalbeh1,2, (Senior Member, Ieee), Rashid Mehmood3, (Senior Member, Ieee), Elhadj Benkhlifa4, And Houbing Song5, (Member, Ieee), Received August 16, 2016, accepted September 11, 2016, date of publication September 26, 2016, date of current version October 15, 2016.*

[5]. *Tomás J. Aragón, http://www.phds.io, San Francisco, California, USA*

[6]. *Al Khatib et al, Healthcare Data Analytics, Australasian Conference on Information Systems, 2015, Sydney.*

[7]. *Revanth Sonnati, International Journal of Scientific & Technology Research Volume 6, Issue 03, March 2017*

[8].https://www.youtube.com/results?search_query=rstudio

[9].https://rstudio.github.io/shinydashboard/index.html

[10].https://www.youtube.com/results?search_query=rstudio+connect+with+shinydshboard