Data Mining Awareness and Readiness in Healthcare Sector: a case of Tanzania

Salim Amour Diwani[1], Anael Sam[2]

[1] Information Communication Science and Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, P.O.BOX 447, Tanzania
diwania@nm-aist.ac.tz

[2] Information Communication Science and Engineering, Arusha, P.O.BOX 447, Tanzania
Anael.sam@nm-aist.ac.tz

Abstract

Globally the application of data mining in healthcare is great, because the healthcare sector is rich with tremendous amount of data and data mining becoming very essential. Healthcare sector collect huge amount of data in daily basis. Transferring data into secure electronic systems of medical health will saves lives and reduce the cost of healthcare services as well as early discovery of contiguous diseases with advanced collection of data. This study explore the awareness and readiness to implement data mining technology within healthcare in Tanzania public sector . This study is triangulated adopted online survey using Google doc and offline survey using presentation techniques through different hospital and distributed the questionnaires to the healthcare professionals. The issues explored in the questionnaire included the awareness of data mining technology, the level of understanding of, perception of and readiness to implement data mining technology within healthcare public sector. In this study we will analyze the data using SPSS statistical tool.

Keywords: *Data Mining , KDD, SPSS, healthcare, ICT,WHO, Healthcare, Dynamic Morphology, NM-AIST, NHIF, Paradigms, Positivists, HIS*

1. Introduction

In recent years, there has been a tremendous change in the adaptation and implementation of new technologies in both private and public sectors. Data mining is an emerging technology used in different types of organizations to improve the efficiency and effectiveness of business processes. Data mining techniques analyze large data sets to discover new relationships between the stored data values. Healthcare is an information rich industry, warehousing large amount of medical data. Health care industry finds it difficult to manage and properly utilize the huge medical data collected through different medical processes. Stored medical data collection is an asset for healthcare organizations if properly utilized. Healthcare industry uses data mining techniques to fully utilize the benefits of the stored medical datasets. This study investigates the readiness to adopt, explore and implement healthcare data mining techniques in Tanzania. The healthcare industry is growing rapidly. The healthcare expenditures are also increasing very high. Various healthcare organizations worldwide such as world health organization (WHO) are trying to provide quality healthcare treatments at cheaper costs. The government sectors are able to benefits the use of healthcare data mining technology in many ways. Examples providing self healthcare treatments, using scientific medical knowledge to provide healthcare services to everyone, minimize time to wait for the medical treatments, minimizing the delay time in providing medical treatment, providing various healthcare treatments based on the patients needs, symptoms and preferences.

2. Problem Description and Objectives

In Tanzania there has been no literature that has discussed the implementation of data mining technology within healthcare in the public sector . Various studies have shown that new technology have become popular and have been accepted by many people worldwide. This study seeks to address this problem by investigating the knowledge about and readiness with regard to the implementation of the data mining technology in healthcare in the public sector. Prior studies of data mining readiness and implementation have been undertaken in the private sector. Evidence suggests that personnel within private sector firms are aware and ready

to implement this technology. Studies in the area of telecommunication, banking, and insurance companies indicate that there is a level of optimism and innovativeness among employees indicating the potential to adopt data mining techniques. Readiness can be seen in terms of the adoption of or intent to adopt data mining technologies [1],[2],[3],[4].

3. Data Mining

Data Mining is the process of extracting useful information from large and mostly unorganized data banks. It is the process of performing automated extraction and generating predictive information from large data banks. It enables you to understand the current market trends and enables you to take a proactive measures to gain maximum benefit from the same. Refer to figure 1 below, data mining is a step in the knowledge discovery in databases (KDD) process and refers to algorithms that are applied to extract patterns from the data. The extracted information can then be used to form a prediction or classification model, identify trends and associations, refine an existing model, or provide a summary of the database being mined[5].
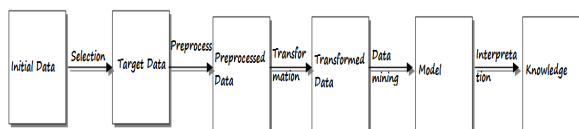

Fig. 1 KDD Process

4. Data Mining in Healthcare

The healthcare industry is growing abundantly. The healthcare expenditures are also increasing tremendously. Healthcare organization worldwide such as National Health Insurance Fund (NHIF) in Tanzania are trying to provide quality healthcare at cheaper cost. The healthcare organizations are adopting new technologies, which will help in early detection of life threatening diseases and lowering the medical cost. Data mining in healthcare is useful in segmenting patients into groups, identifying patients with recurring health problems, relation between

disease and symptoms, curbing the treatment cost, predicting medical diagnosis and can help physicians to discover which medications are most effectively and efficient for patients. Healthcare Data mining uses dynamic morphology for mining healthcare data. Dynamic morphology is the process of extracting healthcare data in order to discover new knowledge and relationships. Refer to figure 2 below, let say we want to mine the tuberculosis data using dynamic morphology we need to follow the following procedures:

- analyze stored medical data for TB

- identify causes or symptoms of TB

- Transform discovered knowledge using any one of the healthcare data transformation techniques to explore extra information

- store the results of knowledge transformation for further analysis
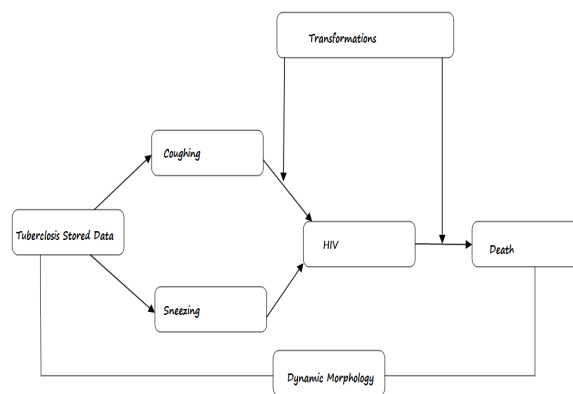

Fig. 2 Example of Dynamic Morphology for Tuberculosis

5. Data Mining Readiness

In order to accept and ready to apply or adopt data mining techniques, the major issues is the willingness and capability to accept the technology. Human resources primarily on their readiness toward accepting data mining technology can be argued as a major obstacle in adopting a new technology within any organization[3]. In this study issues whether people use such a technology or not and their ready to accept a technology or not are major concern. Hence, in order to discover the readiness towards data mining technology issue like clarity for business

strategy, user skills and experiences, organization and culture, technology and data quality needs to be considered.
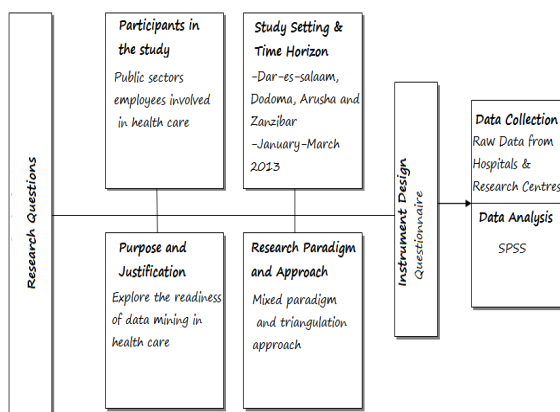
6. Research Design and Methodology

Figure 3: Modeling the research design [6]

The above figure explore the study design which will be adopted in this study. The design consists of research questions which need to be answered and analyzed, identify participants, location of the study and time horizon, purpose and justification of the study before discussing the research paradigm and approach considered. Each of this part of the research design will be discussed in detailed.

6.1 Participants

Participants in this study are professionals who working in medical fields in Tanzania public sectors. The respondent selected are physician, medical doctors, dentists, nurses, pharmacists, and hospitals body of directors. In this study individuals who are working in medical sectors are surveyed to address the basic research problem in the level of awareness and readiness to adopt data mining techniques. The study also seek if people are ready or not to adopt data mining techniques in healthcare. The questionnaires were prepared and available in English through Google docs and posted through emails of individuals who are working in medical field. Also the questionnaires were distributed manually in the hospital and research centers after conducting presentation which explained more about the study design.

6.2 Study Settings and Time Horizon

The setting selected for the study were hospitals in Dar-es-salaam, Tanga, Zanzibar, Arusha and Dodoma. This study was undertaken via presentation through the use of survey in the local hospital during their breakfast meeting. The questionnaire was distributed among respondent and later the respondent were contacted. The process of distributing and collecting the questionnaire begins in January 2013 and end in March 2013.

7. Purpose and Approach

The study aims to gain an understanding of data mining technology utilization in Tanzania healthcare public sector. The impact factors which data mining technology can be sought to be utilized such as level of awareness and readiness of data mining technology in healthcare public sector. Since there is little or no any paper which has been published in data mining technology in Tanzania. Also there are very few companies who implement and use data mining technology in their day to day activities such as in decision making process. This study aim to discover the level of awareness and readiness in data mining technology in healthcare sector and the willingness to accept or reject the implementation. This study will assist in filling the gap in the literature as well.

7.1 Paradigm

All researchers have different beliefs and ways of looking and interacting within their environment. As a result, the way in which research is conducted vary. However, there are certain standards rules that govern researchers actions and beliefs. A paradigm is "a broad view or perspective of something"[7]. Hence, definition of paradigm reveals how research could be affected and guided by a certain paradigm by stating, "paradigms are patterns of beliefs and practices that regulate inquiry within a discipline by providing lenses, frames and processes through which investigation is accomplished"[8]. This study utilized a positivist approach which explore social reality, such as attitudes of readiness toward data mining and their satisfaction toward healthcare system. The concept is based on the philosophical ideas of emphasizing observation and reason behind participant actions in adopting data mining technology for example, their knowledge and understanding about data mining, how they think about technology advancement in their work place.

7.2 Approach

This research is to be conducted in African developing countries. Consequently I expect issues such as national culture, form of government, and the politics of protest and shape the nature of this study in a different way as compared to a study conducted in the context of developed countries. Hence, this study utilized a triangulation approach to explore the survey data which will provide basic understanding about what happening in the industry in regard to data mining technology such as data mining awareness, attitudes towards data mining technology and software used related to data mining if any within healthcare sector. Also, if healthcare professionals are willing to adopt data mining technology and able to upgrade their skills.

7.3 Instrument Design

Questionnaires was designed as a quantitative part of this study using Google doc. There is no wrong or right answers in the questionnaire what we try to test is the level of awareness regard to data mining technology in healthcare and if the healthcare practitioners are willing to adopt a new technology or not. The design of the questionnaires of this study adopted several sources of data, including previous instrument developed by other researchers and literature related to data mining adoption. Types of questions were closed ended questions with ordered and unordered responses using Likert type scale like level of measurement for example 1=strongly disagree and 5 = strongly agree. Itemized rating scale was also developed for few questions. For example, scale from 1=poor to 5=excellent was used for question as to rate the actual performance on important factors for quality healthcare system. There is also similar scale 1=seldom to 5=very often was used for question requesting to indicate frequency of healthcare data were used in particular areas. The questionnaires were pre-tested by sending through the emails of NM-AIST colleagues to review the questions critically and getting the feedback from them about the content and form, and how to make it more appropriate and were to send it, in order to get valuable feedback.

7.4 Data Collection

A survey was conducted for 400 respondents from different parts of the country including Dar es salaam, Arusha, Zanzibar, Tanga and Dodoma. In Tanzania we have 5 medical University in operation with an approximate intake of 500 students, also we have more than 6000 medical doctors scattered around the country[8] . Structured questionnaires were designed and online survey was conducted to reach the targeted sample size of 400 respondents out of the entire population of the at least 6000 medical specialists. Different medical specialists (surgeon, gynecologists, physicians), doctors, Pharmacists, Nurse and information system officers were also involved in the survey.

8. Analysis and Findings

Refer to figure 4, 5 and 6 below, awareness and readiness of the respondents were assessed if they know or do not know any other health information systems and other questions were specific to data mining tool. 60% (241) of the respondents knew at least one health information system excluding data mining tool, 40% (158) did not know any health information system. Among them, 53% (211) respondents have once used it meanwhile 47% (189) have never used any health information system.

In the organization's perspective, 48% (190) of the organizations where respondents once worked at have an experience in a health information systems excluding data mining tool while 50% (202) of the organizations had no experience in health information systems and 2% (8) used other systems.

In order to gain more understanding of the respondents' readiness in adopting technology particularly in data mining, the following statements were provided to the respondents, these include whether technology gives the respondents greater control over his/her daily work activities, to what extent do they agree that technologies are much more convenient in the products and services, preference in using the most advanced technology available, efficient of technology in the respond's occupation, to keep up with the latest technological developments in his/her areas of interest, to have fewer problems than other people in making technology working for them, flexibility to learn about new and different technology, usefulness of the technology for accomplishment of any tasks and willingness to use data mining technology for analyzing medical data in addition to current methods. All these statements had to be responded in the rate of scale which includes strongly disagree, agree, neutral, Disagree and strongly disagree. Strongly disagree and disagree were put by a researcher to be negative response category and agree and strongly agree were put in the category of positive category. All the results are presented in the Chart below:
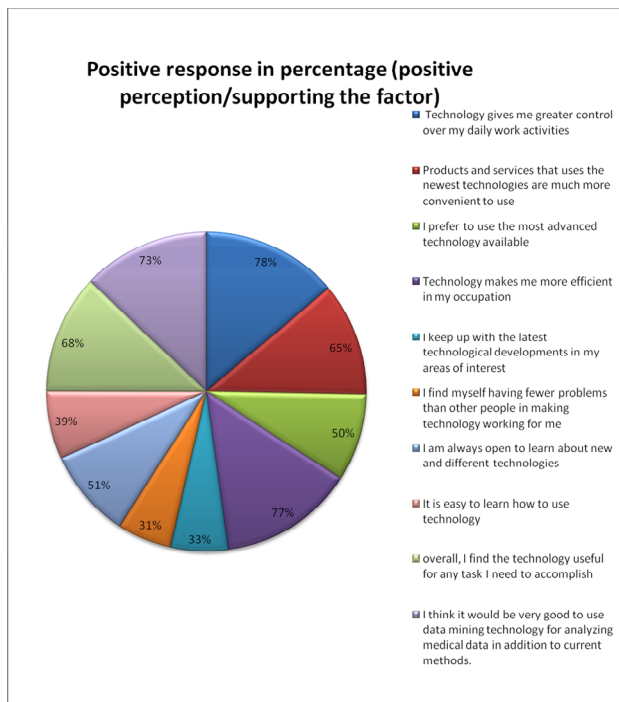
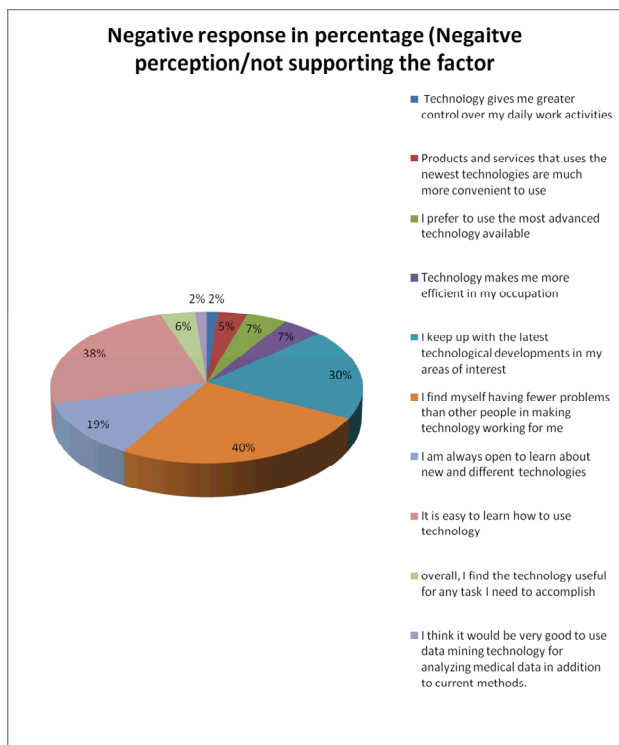Figure 4: Positive response in percentage supporting the factor

Figure 5: Negative response in percentage not supporting the factor

It is observed that new technology like data mining tools is essential to facilitate personal performance though

users will catch up with it slowly because it will not be easy for them to adapt to it quickly. Also, the majority of the respondents stayed neutral about new technology like data mining. This is due to limited knowledge of data mining tool.
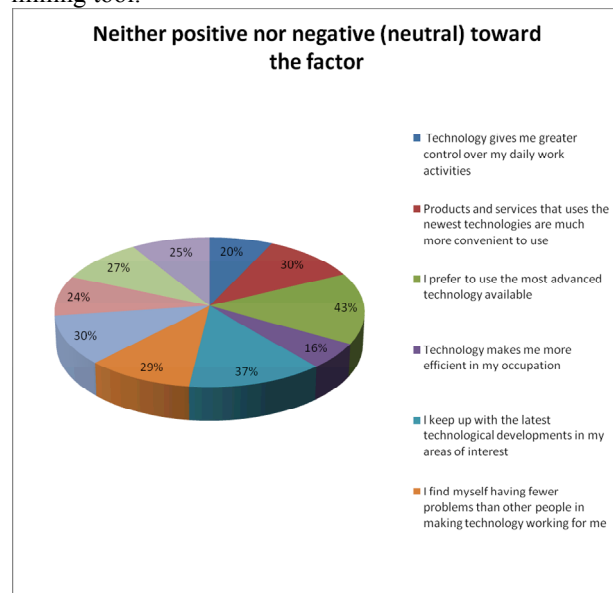

Figure 6: Neither positive nor negative toward the factor

9. Reliability

In this section the reliability of data in the study is discussed. Refer to table 1, 2 and 3 below, a reliability test was conducted using Cronbach's Alpha to measure the internal consistence or average correlation of the items. . It ranges from 0-1. The most reliable value is 0.7. In this study alpha obtained after analysis in SPSS was 0.974 which means the correlation of the research questions were most appropriate. The key variables used in the statistical analysis are: data mining readiness and awareness.

Table 1: Health Information System

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item Total Correlation	Cronbach's Alpha if Item Deleted
Do you know any HIS in Tanzania?	1495.5556	2614954.365	.304	.974
Have you ever used any HIS?	1495.5000	2614827.802	.307	.974

ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 1, No.7 , January 2014
ISSN : 2322-5157
www.ACSIJ.org

Table 2: Data Accuracy

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
The data which recorded is conforms to the actual value	1489.5000	2594194.972	.473	.974
data which recorded in your system is up to date	1489.4074	2594545.454	.469	.974
Complete : all relevance value for a certain variable are recorded	1487.9815	2597785.792	.366	.974
Consistent: representation of the data value is the same in all classes	1489.8519	2593776.166	.477	.974
Please rate the actual performance on each of those factors by your organization	1438.9630	2568839.206	.374	.974

Table 3: Medical Data

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Please indicate how frequently you use the medical data from HIS in each of the following areas- Decision Making and data analysis	1465.6667	2548146.038	.542	.974
Please indicate how frequently you use the medical data from HIS in each of the following areas- Performance measurement and planning	1462.0370	2536410.263	.608	.973

10. Discussion

Data mining in healthcare helps to improve healthcare standards. Data mining technology helps healthcare organizations to achieve various business goals such as low costs, increase revenue generation while maintaining high quality healthcare. Data mining in healthcare identified as the process of extracting and analyzing healthcare data for presentation in a format allowing the generation of information and the creation of knowledge through the analysis of information in order to enhance the decision making process within the public sector organization.

The major research problem addressed in this study was in Tanzania there is a lack of the status of implementation

ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 1, No.7 , January 2014
ISSN : 2322-5157
www.ACSIJ.org

of data mining techniques within healthcare in the public sector and benefits to be derived by implementing such technologies. The study sought to increase an understanding of data mining technology in Tanzania within public sector. Areas of interests within this study were data mining awareness and readiness to accept or reject technology in healthcare public sector, the reasons to accept or reject the technology and impacts of data mining technologies.

11. Conclusion

This paper discussed the awareness and readiness to adopt data mining techniques in healthcare sector in Tanzania. The statistical results of the questionnaire were collected and analyzed, and the issues discussed in the questionnaire were healthcare information system, data mining readiness, data mining technologies implementation and perception of data mining. The descriptive statistics has shown that many healthcare professionals in Tanzania they are not aware of the data mining technology and they never used it before. Hence they are willing to accept and adopting data mining technology if they get an opportunity to do so. Respondents identified the influencing factors for utilizing data mining and reason for not utilizing it. Each of the major research questions were analyzed and the results were given. Cronbach's Alpha was used to test the internal consistence of the questions; the alpha test was at least 0.7 which means the questions were most reliable. All the questions used likert scales, those scales which were coded as strongly agreed and agreed were considered to measure positive response while disagreed and strongly disagreed were considered to measure negative response. The findings showed the majority are aware of health information system but limited knowledgeable on data mining technology. About performance of health information system the majority of the respondents strongly agreed that the system should be easy to use, able to validate data, adequate and sufficient documentation for employees to follow and easy to modify and upgrade information. Also they suggested that sufficient financial resources, effective and adequate for the staffs is required in order to implement data mining in developing countries like Tanzania.

Acknowledgement

References

[1] Berger, C. (1999). 'Data mining to reduce churn'. Target Marketing, 22(8), 26.

[2] Chye, K. H., & Gerry, C. K. L. (2002). 'Data Mining and Customer Relationship Marketing in the Banking Industry'. Singapore Management Review, 24(2), 1-27.

[3] Dahlan, N., Ramayah, T., & Hoe, K. A. (2002). *Data Mining in the Banking Industry:* An Exploratory Study. The proceedings of the International Conference 2002. Paper presented at the 'Internet Economy and Business', Kuala Lumpur, 17-18th September 2002.

[4] Chun, S.-H., & Kim, S. H. (2004). 'Data mining for financial prediction and trading: application to single and multiple markets'. Expert System with Applications, 26, 131-139.

[5] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). 'The KDD process for extracting useful knowledge from volumes of data'. Association for Computing Machinery. Communications of the ACM, 39(11), 27-34.

[6] Cavana, R. Y., Delahaye, B. L., & Sekaran, U. (2001). Applied Business Reseach:Qualitative and Quantitative Methods. Sydney and Melbourne: John Wiley & Sons Australia, Ltd.

[7] Taylor, B., Kermode, S., & Roberts, K.(2007). Research in nursing and health care: Evidence for practice. (3rd ed.). South Melbourne, VIC: Thomson.

[8] Weaver, K., & Olson, J.K.(2006). Understanding paradigms used for nursing research. Journal of Advanced Nursing, 53(4), 459-469.

Salim Amour Diwani received his BS degree in computer science at Jamia Hamdard University, New Delhi, India in 2006 and Ms degree in computer science at Jamia Hamdard University in New Delhi, India in 2008. He is currently a PhD scholar in Information communication Science and Engineering at Nelson Mandela African Institution of Science and Technology in Arusha, Tanzania. His primary research interests are in Data Mining, Machine Learning and Database Management Systems.