Overview Applications of Data Mining In Health Care: The Case Study of Arusha Region

[1,]Salim Diwani, [2,]Suzan Mishol , [3,]Daniel S.Kayange ,[4,]Dina Machuve And [5,]Anael Sam

1,2,3,4,5,School of Computational and Communications Science and Engineering, Nelson Mandela-African Institution of Science and Technology (NM-AIST), Arusha, Tanzania.

ABSTRACT

Data mining as one of many constituents of health care has been used intensively and extensively in many organizations around the globe as an efficient technique of finding correlations or patterns among dozens of fields in large relational databases to results into more useful health information. In healthcare, data mining is becoming increasingly popular and essential. Data mining applications can greatly benefits all parties involved in health care industry. The huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform huge amount of data into useful information for decision making. This paper explores data mining applications in healthcare in Arusha region of Tanzania more particularly; it discusses data mining and its applications in major areas such as evaluation of treatment effectiveness, management of healthcare itself and lowering medical costs.

KEYWORDS*: Data mining, Healthcare application, Knowledge discovery, Data warehousing.*

I. INTRODUCTION

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using massive data stores to uncover previously unknown patterns. Data mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Data mining is not new idea, it has been used intensively and extensively by financial institutions for activities such as credit scoring and fraud detection[1, 2]; marketers for direct marketing and cross-selling; retailers for market segmentation and store layout; manufacturers for quality control and maintenance scheduling and it has been used in hospital care as well[2].Data mining has been becoming increasingly popular, it has been noted that several factors have motivated the use of data mining applications. The existence of medical insurance fraud and abuse, for example has led many healthcare insurers to attempt to reduce their losses by using data mining tools, the application has helped to help them find and track offenders. However fraud detection using data mining applications is prevailing in the commercial world for detection of fraudulent credit card transactions and fraudulent banking activities [3]. Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods hence calls for technological interventions so as to simplify management of those data. Data mining can improve decision making by discovering patterns and trends in large amounts of complex data. Such analysis has become increasingly essential as financial pressures have amplified the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue and operating efficiency while maintaining a high level of care.Healthcare organizations that perform data mining are better positioned to meet their long term needs; data can be a great asset to healthcare organizations, but they have to be first transformed into informationYet another factor motivating the use of data mining applications in healthcare is the realization that data mining can generate information that is very useful to all parties involved in the healthcare industry. For example, data mining applications can help healthcare insurers detect fraud and abuse, and healthcare providers can gain assistance in making decisions.. Data mining applications also can benefit healthcare providers such as hospitals, clinics, physicians, and patients by identifying effective treatments and best practices [4].

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using massive data stores to uncover previously unknown patterns. Data mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data.Data mining is an automated approach for discovering or inferring hidden patterns or knowledge buried in data. 'Hidden' means patterns that are not made apparent through casual observation[5].Data Mining is an interdisciplinary field that combines artificial intelligence, computer science, machine learning, database management, data visualization, mathematic algorithms, and statistics. Data Mining is a technology for knowledge discovery in databases (KDD). This technology provides different methodologies for decision making, problem solving, analysis, planning, diagnosis, detection, integration, prevention, learning and innovation[6].Data mining is a variety of techniques such as neural networks, decision trees or standard statistical techniques to identify nuggets of information or decision making knowledge in bodies of data, and extracting these in such a way that they can be put to use in areas such as decision support, prediction, forecasting, and estimation.

Data mining has been becoming increasingly popular, it has been noted that several factors have motivated the use of data mining applications. The existence of medical insurance fraud and abuse, for example has led many healthcare insurers to attempt to reduce their losses by using data mining tools, the application has helped to help them find and track offenders. However fraud detection using data mining applications is prevailing in the commercial world for detection of fraudulent credit card transactions and fraudulent banking activities [7]. Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods hence calls for technological interventions so as to simplify management of those data. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data. Such analysis has become increasingly essential as financial pressures have amplified the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue and operating efficiency while maintaining a high level of care.Yet another factor motivating the use of data mining applications in healthcare is the realization that data mining can generate information that is very useful to all parties involved in the healthcare industry. For example, data mining applications can help healthcare insurers detect fraud and abuse, and healthcare providers can gain assistance in making decision.

Data mining applications can benefit healthcare providers in Arusha regions such as hospitals, clinics, physicians, and patients by adopting new technologies, which will help in early detection of life threatening diseases and lowering the medical costs. The aims of quality healthcare services are:

- providing safe healthcare treatments
- using scientific medical knowledge to provide healthcare services to everyone
- providing various healthcare treatments based on the patient's needs, symptoms and preferences
- minimizing the time to wait for the medical treatment
- minimizing the delay time in providing medical treatment.
- Health determinants
- Inputs to the health system and related processes (e.g., health infrastructure, human and financial resources, equipment, policy, and organization)
- Health outcomes (e.g., mortality, morbidity, disability, well-being, and health status)

II. DATA MINING

Nowadays there is huge amount of data stored in real-world databases and this amount continues to grow fast as it creates both an opportunity and a need for semi-automatic methods that discover the hidden knowledge in such database. If such knowledge discovery activity is successful, discovered knowledge can be used to improve the decision making process of an organization. For instance data about a hospital's patient might contain interesting knowledge about which kind of patient is more likely to develop a given disease. This knowledge can lead to better diagnosis and treatment for future patients[8].Data mining and knowledge discovery is the name often used to refer to a very interdisciplinary field, which consists of using methods of several research areas to extract knowledge from real world data sets.There is a distinction between the terms data mining and knowledge discovery; the term data mining refers to the core steps of a broader process called knowledge discovery in database. In addition to the data mining step which actually extracts knowledge from data, the knowledge discovery process includes several preprocessing and post processing steps. The goal of data preparation methods is to transform the data to facilitate the application of a given data mining algorithms,

where the goal of knowledge refinement methods is to validate and refine discovered knowledge.The knowledge discovery is both iterative and interactive. It is iterative because the output of each step is often feedback to previous steps and typically many iterations of this process are necessary to extract high quality knowledge from data. It is interactive because the user or more precisely an expert in the application domain should be involved in this loop to help in data preparation, discovered-knowledge validation and refinement.

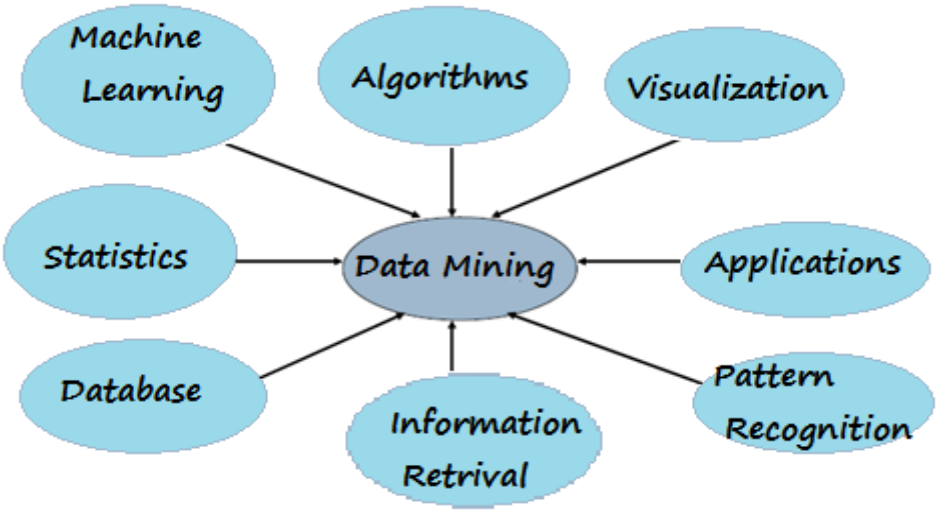

Fig 1: Data Mining Architecture

III. HEALTHCARE DATA MINING APPLICATIONS

There is vast potential for data mining applications in healthcare particularly in Arusha health centers. This is due to the fact that the use of technology can stand to provide accurate and more meaningful statistics of different activities going on within health centers. Generally, these activities can be grouped as the evaluation of treatment effectiveness, management of healthcare itself and customer relationship management[9].Treatment effectiveness: Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective,[10] for example the outcomes of patient groups treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective.Other data mining applications related to treatments include associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs and determining proactive steps that can reduce the risk of affliction.future needs of individuals to improve their level of satisfaction[11]. These applications also can be used to predict other products that a healthcare customer is likely to purchase, whether a patient is likely to comply with prescribed treatment or whether preventive care is likely to produce a significant reduction in future utilization.

IV. KNOWLEDGE DISCOVERY

Knowledge discovery is a non-tedious procedure for identifying effective and potential benefits amid data. It is known from Fig. 2 below that data mining is one of the important processes of knowledge discovery. From the definitions by the scholars, it is clear that the usage of data mining is an analysis process within a series of knowledge discovery[12]. As time changes the term ‘‘data mining’’ gradually replaces ‘‘knowledge discovery’’. From above, purpose of data mining is to uncover the rules that are helpful to decision process from massive data.

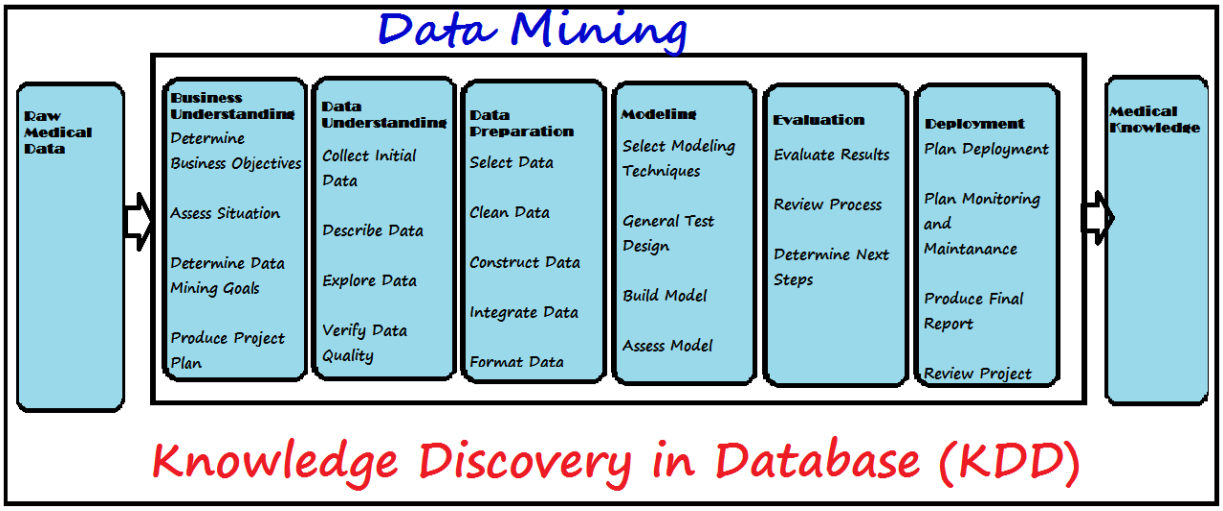

Fig 2: Knowledge Discovery

Knowledge Discovery As A Process

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
5. Data mining (an essential process where intelligent methods are applied in order to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user) Steps 1 to 4 are different forms of data preprocessing, where the data are prepared for mining. The data mining step may interact with the user or a knowledge base. The interesting patterns are presented to the user and may be stored as new knowledge in the knowledge base. Note that according to this view, data mining is only one step in the entire process, albeit an essential one because it uncovers hidden patterns for evaluation.

We agree that data mining is a step in the knowledge discovery process. However, in industry, in media, and in the database research milieu, the term data mining is becoming more popular than the longer term of knowledge discovery from data.

Data cleaning and data integration techniques may be performed on the data. The database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request.

V. LIMITATIONS OF DATA MINING

Data mining applications can greatly benefit the healthcare industry. However, they are not without limitations[13]. Healthcare data mining can be limited by the accessibility of data, because the raw inputs for data mining often exist in different settings and systems [14], such as administration, clinics, laboratories and more. Hence, the data have to be collected and integrated before data mining can be done. While several authors and researchers have suggested that a data warehouse be built before data mining is attempted, that can be a costly and time-consuming.Secondly, other data problems may arise whereby this include missing, corrupted, inconsistent or non-standardized data such as pieces of information recorded in different formats in different data sources. In particular, the lack of a standard clinical vocabulary is a serious hindrance to data mining applications. Thirdly, there may be ethical, legal and social issues, such as data ownership and privacy issues, related to healthcare data. Fourthly, the successful application of data mining requires knowledge of the domain areas as well as in data mining methodology and tools. Without a sufficient knowledge of data mining, the user may not be aware or be able to avoid the pitfalls of data mining.

VI. SUGGESTIONS

As it obvious that data mining is very important for healthcare and it can improve the situation from health centers to customers, we recommend that the government has to look at the technology, to invest on it, as we believe that it can prove great positive feedback to the lives of residents and organizations too. However, we recommend during implementation, government has to ensure the integration of data, text and digital diagnostic images for images such as X-rays, MRI's whereby by doing so will provide significant help for most residents who cannot afford to travel far for experts to read their images.

VII. CONCLUSION

Currently, Arusha region has few medical personnel and hospitals where data mining can be used as a substitute. We have suggested the implementation and use of data mining technology at Arusha region as this will improve the healthcare industry and well- being of the residents. In the paper we have proposed the integration of data and text mining, using digital diagnostic images which can be brought into healthcare data mining applications. It is our belief that the paper will be a contribution to the data mining and healthcare literature and practice. It also is hoped that this paper can help all parties involved in healthcare reap the benefits of healthcare data mining.

REFERENCES

[1] T.-H. Chen and C.-W. Chen, "Application of data mining to the spatial heterogeneity of foreclosed mortgages," Expert Systems with Applications, vol. 37, pp. 993-997, 2010.

[2] N. Padhy, "THE SURVEY OF DATA MINING APPLICATIONS AND FEATURE SCOPE Pragnyaban Mishra, Neelamadhab Padhy," 2012.

[3] R. Barker, "Online crisis communication response: A case study of fraudulent banking transactions in South Africa," Communicatio: South African Journal for Communication Theory and Research, vol. 37, pp. 118-136, 2011.

[4] H. C. Koh and G. Tan, "Data mining applications in healthcare," Journal of Healthcare Information Management—Vol, vol. 19, p. 65, 2011.

[5] McVey, S. 'Data Mining: The Brains Behind eCRM'. (2000). from http://www.technologyevaluation.com/Research/ResearchHighlights/BusinessApplications/2000/11/research_notes/prn_TU_BA_SRM_11_06_00_1.asp (17 May, 2004)

[6] Liao, S.-h. (2003). 'Knowledge Management Technologies and applications Literature review from 1995 to 2002'. *Expert System with Application*, 25, 155-164.

[7] A. A. Freitas, "Understanding the crucial role of attribute interaction in data mining," Artificial Intelligence Review, vol. 16, pp. 177-199, 2001.

[8] J. Wang, B. Zhou, and R. Yan, "Benefits and Barriers in Mining the Healthcare Industry Data," International Journal of Strategic Decision Sciences (IJSDS), vol. 3, pp. 51-67, 2012.

[9] F. Buttle, Customer relationship management: Routledge, 2012.

[10] K. J. Cios, W. Pedrycz, and R. Swiniarsk, "Data mining methods for knowledge discovery," Neural Networks, IEEE Transactions on, vol. 9, pp. 1533-1534, 1998.

[11] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques: Morgan kaufmann, 2006.

[12] H. Mannila, "Methods and problems in data mining," in Database Theory—ICDT'97, ed: Springer, 1997, pp. 41-55.

[13] Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. & O'Shea, M.J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. Journal of Healthcare Information Management, 15(2), 155-164.

[14] Benko, A. & Wilson, B. (2003). Online decision support gives plans an edge. Managed Healthcare Executive, 13(5), 20.

[15] Gillespie, G. (2000). There's gold in them thar' databases. Health Data Management, 8(11), 40-52.

[16] Kolar, H.R. (2001). Caring for healthcare. Health Management Technology, 22(4), 46-47.

[17] Relles, D. Ridgeway, G. & Carter, G. (2002). Data mining and the implementation of a prospective payment system for inpatient rehabilitation. Health Services & Outcomes Research Methodology, 3(3-4), 247-266.