

Title: Healthcare Dataset

Adrian Lockwood

Clafin University

HNTH 391: Honors Thesis Seminar

Dr. Shrikant Pawar

06 October 2023

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Karina Liles. She has played a helpful role in providing the smartwatches that were crucial in this research project. I would also like to thank her for giving the people who would be beneficial in this study. This research aims to explore the link between healthcare and data acquisition and their impact on people who want to improve their health and well-being. Dr. Liles' support has been indispensable in the completion of this study, and for that, I am truly grateful.

THESIS STATEMENT

This research project aims to prove that healthcare tactics can improve the general welfare of African Americans. This will be performed by a study of random individuals in the State of South Carolina. Using a collection of forms from African Americans, I will use visualization techniques to analyze the data collected. This will be conducted by using visualization techniques so the results of the output are displayed properly. The type of forms used will consist of health aspects including history, knowledge, and attitude. The tools used will help provide solutions in Healthcare for African Americans.

ABSTRACT

The integration of technology into daily life has generated great interest across various domains, including research, healthcare, and personal well-being. This thesis will provide an overview of categorizing the findings of Healthcare. The purpose of this research is to find visualization methods so we can collect various health data. This is ideal for tracking health in this project because we have a large amount of input from participants. The study will investigate if health literacy and data will help improve individuals' health efforts.

KEYWORDS AND ABBREVIATIONS

ARF- Align, Rank, Filter

EHR- Electronic Health Records

MEPS- Medical Expenditure Panel Survey

VISITORS- Visualization of Time-Oriented Records

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
THESIS STATEMENT	ii
ABSTRACT	iii
KEYWORDS AND ABBREVIATIONS	iv
INTRODUCTION	5
METHOD	
6	
RESULTS	7
CONCLUSION	8
DISCUSSION	9
REFERENCES	10

INTRODUCTION

These visualization methods will provide an overview of how healthcare tactics are constantly improving. These tools have changed a lot over the years in their functionalities and found applications in healthcare, education, and research. The research highlighted in this essay shows the potential of how these methods can be helpful in today's world. I aim to gather valuable health data by using these tactics to showcase and promote health literacy among people who lack the resources needed.

METHODS

There have been a few ways of using visualization when it comes to health. One of the most common has been using Healthcare models. Healthcare models have been very helpful as far as predicting a binary outcome. This means that it could detect responses such as "yes or no", and "True or False". A method that uses a strategy like this is called logistic regression. Logistic regression is a statistical method that uses parameters to predict the probability of an event. The method is based on independent variables (like patient data) to predict things like, risk assessment, early diagnosis, treatment planning, and clinical mortality. Logistic Regression is a great visualization method that has been improving greatly in the Healthcare Industry.

West (2015) also shows that there have been methods implemented for a different approach to information visualization. Innovation Visualization is "visualizations other than standard graphs

traditionally used for displaying healthcare information” (West 333). A study was conducted to evaluate innovative approaches to Electronic Health Records (EHR) in America. using visualization techniques reported between 1996 and 2013. This was done by using an electronic literature search called MEDLINE. It is considered “the most frequently used reference database in healthcare, and Web of Knowledge” (West 333). The study concentrated on records from a single patient and the representation of complex data connected to a single patient. By using this algorithm, it was able to find potential solutions on how to find better solutions for health.

The complexity of algorithms has been a great advancement in human life. One way this is beneficial is by finding the population cost of Healthcare Datasets. Healthcare costs continue to be a challenge in the United States. To tackle this issue, machine algorithms have been used to predict Healthcare costs. This has been done by using 3 methods Regression Tree, M5 Model Tree, and random forest. These models would usually estimate the costs by gathering the mean cost. Another study used was the Medical Expenditure Panel Survey (MEPS) dataset. This study lasted from 2004-2012.

Abudiyab showed Visualization in Healthcare. It involves “Visualization in healthcare involves the effective presentation of information through graphics, images, and videos” (Abudiyab 1)

Confidentiality is a major concern when collecting personal information. A common solution researchers use is Differential Privacy. This is where a company can get data from a user with a dataset without knowing about the individual personally. Many industries use this method of protecting the patients' confidentiality including tech and healthcare. After receiving

data, someone called a data custodian usually collects the data and gives what is called “private answers” to the person requesting. Therefore the data can be analyzed without revealing any sensitive information.

RESULTS

Some of the findings were astounding as these led to new solutions. In West (2015), it was found in the searches, of different types of applications. One of them being LifeLines 2. This is an application that visualizes categorial data from EHRs. It does this by using an operation called Align, Rank, and Filter (ARF) to collect patient data. This is different from standard visualization tactics as it is categorized by each feature so nothing gets misconstrued and can be compared easily.

Another method found by West is VISITORS (Visualization of Time-Oriented Records). This is another application similar to ARF with one exception. It uses previous health records applications to develop new ones when dealing with complex data that is temporal. This means it collects data that refers to events, dates, and times in the past, present, or future. When operated, it was found to be effective when examining patient EHRs. The results “found the system feasible for exploring longitudinal data for quality or clinical results” (West 334). The results were found not by using traditional graphs but by recognizing patterns and finding different values during a specified period.

Cost of Healthcare

It was revealed in Sushmita (2015) that the average healthcare cost of just one person is \$50,000. The demographics also showed that 9.08% of that comes from females while 23.09%

are male. The remaining 67.83% are unspecified. While MEPS had a broader view of medical events, the data collected was categorized by beneficiaries. The main method of this was to use a Regression Tree by establishing an independent variable between the events and the cost of each one. It also predicts “a response y (cost in our case) from inputs x_1, x_2, \dots, x_p ” (Newman 91). This collects a statistical analysis of the average cost of a Healthcare Dataset.

Using Artificial Intelligence

Dave et al. (2020) discussed how Explainable AI can use techniques to help out a Heart Disease Dataset. Artificial Intelligence has made society much easier and simpler for people to use and understand. It has been especially important to the Industry of Healthcare. While Humans have tried for centuries to figure out things like (1) medicine (2) practices, (3) laws, and (4) procedures, AI has come in to figure out potential problems hard to solve. With that, it has provided a sense of accuracy, security, and trust to these medical practitioners. After, the results will be able to determine how effective Explainable AI is to human society in healthcare.

They do this by using the UCI ML repository. This a collection of databases and datasets used to learn machine algorithms. These datasets are very helpful in detecting patients who suffer from Heart Disease. Explainable A.I. not only interprets certain results but also determines how they can get those results. Findings show that using the repository is so good and accurate, that physicians validate the information found. This further reinforces that Explainable A.I. does an outstanding job when looking at a Heart Disease Dataset.

Technology and the Future of Public Health

The article presented is about how Technology has impacted the lives of humans and their health. These technologies have been innovated for the welfare and simplicity of humans. These Healthcare innovations include Robotic surgeries which minimize risks and pain while increasing the accuracy and practicality of the procedure. Telemedicine has increased and improved in the last 3 years due to the COVID-19 pandemic. Devices like smartwatches make it easier to monitor and track health like heart rate hours slept and calorie intake and SmartInhalers track how much dosage is needed on a person suffering from asthma. Research shows that SmartInhalers are known to use less medicine than traditional inhalers.

REFERENCES

- Abudiyab, N. A., Alanazi, A. T., & Abudiyab, N. (2022). Visualization Techniques in Healthcare Applications: A Narrative Review. *Cureus*, 14(11).
- Banerjee, N., Rahmati, A., Corner, M. D., Rollins, S., & Zhong, L. (2007, September). Users and Batteries: Interactions and adaptive energy management in mobile systems. In *International conference on ubiquitous computing* (pp. 217-234). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Dave, D., Naik, H., Singhal, S., & Patel, P. (2020). Explainable AI meets Healthcare: A Study on Heart Disease Dataset. *ArXiv*.
- Purushotham, S., Meng, C., Che, Z., & Liu, Y. (2018). Benchmarking deep learning models on large healthcare datasets. *Journal of Biomedical Informatics*, 83, 112-134.
- Sushmita, S., Newman, S., Marquardt, J., Ram, P., Prasad, V., De Cock, M., & Teredesai, A. (2015). Population Cost Prediction on Public Healthcare Datasets. In *Proceedings of the 5th International Conference on Digital Health 2015 (DH '15)* (pp. 87-94). Association for Computing Machinery. <https://doi.org/10.1145/2750511.2750521>
- Technology and the future of public health. (2021). In K. Hayes (Ed.), *Global Public Health and Disease Control Sourcebook*. Omnigraphics. Retrieved November 8, 2023, from <https://search.credoreference.com/articles/Qm9va0FydGljbGU6NDk4ODM=?aid=101266>.

Vivian L West, David Borland, W Ed Hammond, Innovative information visualization of electronic health record data: a systematic review, *Journal of the American Medical Informatics Association*, Volume 22, Issue 2, March 2015, Pages 330–339, <https://doi.org/10.1136/amiajnl-2014-002955>