LOW COST DESIGN TO SUPPORT HEALTH CARE NEEDS OF UNDERSERVED POPULATIONS: USING SMS AND AI TO ENGAGE AT-RISK POPULATIONS

A Thesis Presented to

The Faculty of the Computer Science Program

California State University Channel Islands

In (Partial) Fulfillment

of the Requirements for the Degree

Masters of Science in Computer Science

By

Vijay Singh

December 14th 2017

APPROVED FOR THE COMPUTER SCIENCE PROGRAM

12/14/17

Advisor: Dr. Brian Thoms — Date

12/14/2017

Dr. Pawel Pilarczyk — Date

12/14/2017

Dr. Jason Isaacs — Date

APPROVED FOR THE UNIVERSITY

12/14/17

Dr. Joseph Shapiro — Date
Interim AVP and Dean
Extended University and International Programs

Non-Exclusive Distribution License

In order for California State University Channel Islands (CSUCI) to reproduce, translate and distribute your submission worldwide through the CSUCI Institutional Repository, your agreement to the following terms is necessary. The author(s) retain any copyright currently on the item as well as the ability to submit the item to publishers or other repositories.

By signing and submitting this license, you (the author(s) or copyright owner) grants to CSUCI the nonexclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that CSUCI may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that CSUCI may keep more than one copy of this submission for purpose s of security, backup and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. You also represent and warrant that the submission contains no libelous or other unlawful matter and makes no improper invasion of the privacy of any other person.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant CSUCI the rights required by this license, and that such third party owned material is clearly identified and acknowledged within the text or content of the submission. You take full responsibility to obtain permission to use any material that is not your own. This permission must be granted to you before you sign this form.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN CSUCI. YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT O F REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

The CSUCI Institutional Repository will clearly identify your name(s) a s the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

Low cost design to support health care needs of underserved populations: using SMS and AI to engage at-risk populations

Title of Item

Low cost system design, underserved populations, machine learning, healthy behavior

3 to 5 keywords or phrases to describe the item

Vijay Singh

Author(s) Name (Print)

12/14/2017

Author(s) Signature Date

ABSTRACT

This thesis focuses on building a system for underserved populations to manage their health by interacting through the short messaging system (SMS) of mobile devices. The proposed system is novel in that it does not rely on telecommunication data plans, which can be expensive and not affordable by underserved populations. Additionally, the system will afford novel mechanisms to help better track and measure healthy behaviors within this user population. In this study, I present the design, implementation and simulation of such a system that helps manage health information for these types of users (underserved population). The system utilizes SMS technology to aid in personal information storage and retrieval. SMS was chosen because it is an effective and efficient method to connect with these populations, many of whom currently have cellular devices. SMS also provides a low barrier to entry for those who do not have a cellular device. This management of personal health information is of concern for underserved populations facing chronic illnesses. The proposed system focuses on three steps; User Input, Data Processing and Information Output. First, data is collected from the user through questionnaire via SMS, which is curated to target patterns of healthy behaviors and irregularities in health. Next, a decision tree classification technique is used to create a machine-learning model by training on sample dataset. Finally, the system predicts if the user behavior is healthy or not by using the trained model and provide suggestions to manage their health properly.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor Dr. Brian Thoms for his patience, continuous support, motivation and immense knowledge. His guidance helped me to accomplish this research work and has played a pivotal role in enhancing my skills. The entire process has been very fascinating and inspiring.

I would also like to thank the committee members, Dr. Pawel Pilarczyk and Dr. Jason Isaacs for providing their valuable comments, suggestions and encouragement.

Table of Contents

Table of Contents of Figures

CHAPTER I: INTRODUCTION

1.1 Introduction

Reports by International Telecommunication Union (ITU) indicate that over 7 billion people or 95 percent of the global population now have access to mobile communication (ITU, 2016). Because of the globally available mobile phone network and the potential for real-time communication, development of mobile health (mHealth) interventions has happened rapidly over the past decade, helping to address discrepancies in healthcare service access and helping to increase healthcare outcomes (Fiordelli et al., 2013). In many situations, text messaging or Short Message Service (SMS) is preferred as the suitable tool for mHealth interventions (Finitsis et al. 2017). mHealth interventions are widely accepted because they can be sent, stored, answered and retrieved at the user's convenience, which is inexpensive and supports any mobile device, irrespective of their configurations (Thakkar et al., 2017).

As of 2015, an average of 169.3 billion text messages is sent each month, up from 110.4 billion in 2008 (CTIA, 2017). Hence, SMS will continue to play an essential role in mobile communication (Statista, 2017). Consequently, mHealth applications will continue to utilize SMS as a communication platform to reach remote and underserved populations.

In the present study, I build an easy to use system to manage personal health, focusing on underserved populations by interacting through their mobile devices using SMS technology. The system will help users to track and measure healthy behaviors. The system works in the following way. The system sends SMS with questions to each registered user every day and the response to each of these questions is stored on the server. Once a week the system applies a machine learning model created using decision tree algorithm, to predict if the user is healthy or not. In the coming section, I will describe the system design and its implementation. I will also talk about the study and tests conducted from a small population to show working of the system and results.

1.2 Healthcare Background

mHealth interventions can be classified into multiple categories based on the targeted health behavior. In Thakkar et al. (2016), mHealth technologies enhanced anti-retroviral

medication adherence, attendance of medical appointments in patients suffering from chronic disease including diabetes self-management, while Hall et al. (2015) found that mHealth technologies can play a formidable role in smoking cessation, weight loss and increased physical activity. Although the researchers suggested that the requirement for further research is to identify long-term intervention effects in both studies, each recognized features of mHealth that give success and estimate outcome measures other than self-reported adherence. mHealth is the promising domain of research. Hence, researchers and healthcare professionals are dedicated to advancing mHealth to enhance the quality and levels of patient health globally.

1.3 Underserved Populations

According to Rabinowitz et al. (2000), the term 'underserved population' is defined as the population which does not have access to primary healthcare services and are medically impoverished. The description for this group was not given only based on the attributes that are inherent of the individuals in the population, but also the environmental factors that might cause the demographic group to undergo challenges with respect to their health. Underserved populations were described as '*vulnerable populations*' by many well-respected sources. The racial and ethnic minorities, economically disadvantaged, low-income children, the uninsured people, the homeless, the elderly, those with chronic health conditions including severe mental illness, and those with Human Immunodeficiency Virus (HIV) can all be classified as vulnerable populations (AJMC, 2006). Furthermore, populations in rural areas who exhibit stumbling blocks on the availability of healthcare are also included. The factors such as ethnicity, race, gender, age, insurance coverage, income, and unavailability of usual source of care increased the susceptibility of these people. The societal parameters such as poverty, housing, and inadequate education added to their woes in addition to their existing health and healthcare issues (TPCHD, 2013).

1.4 Health in Underserved Populations

Several communities were mentioned as underserved by the U.S. Department of Health and Human Services (HHS), which categorized underserved populations based on Medically Underserved Area (MUA), Health Professional Shortage Area (HPSA), the communities and their people who are lacking health care needs, or Medically Underserved Population (MUP) (Medically Underserved Areas and Populations, 2016). The HHS provided the descriptions which identified these

groups as lacking dental health care workers and / or sufficient numbers of primary care workers to meet the service demands of these populations (USD of Health, 2016).

Many challenges surface because of the understaffed healthcare system in the case of the patients in these communities. Extended travel to access health care, waiting a long time for appointments and neglected health care are some of the challenges. The healthcare providers in the underserved communities must care for more patients whereas those in the non-underserved communities do not have much work burden. The patients' acuity is also higher in the underserved communities (TPCHD, 2013).

1.5 Thesis Breakdown

The remaining portion of the thesis is disserted in the following manner.

Chapter 2 summarizes relevant literature and introduces researcher related to underserved populations, mobile health studies and data mining techniques.

Chapter 3 provides a brief description of the architecture and the components used in the research is elucidated in this section.

Chapter 4 provides complete details into the construction, implementation procedure and working model for a proposed system for facilitating healthy outcomes through SMS design and data mining components.

Chapter 5 describes the implementation phase including simulations and tests with a sample population. Some analysis based on the use case test run.

Chapter 6 discusses the outcomes of the research work are explained as a conclusion and the future work is described as improvements of the work.

CHAPTER II: LITERATURE

2.1 Introduction

The present section reviews previous studies on the challenges faced by underserved populations in developed and developing nations followed by a review of mHealth projects and the factors that should be considered for this thesis. This section then discusses data mining and classification techniques involved in previous mHealth projects. The article will also address cloud computation in health care and concludes with the research gap this project aims to address.

2.2 Challenges to mHealth in Underserved Populations:

A developed nation is one which has a highly developed economy and tends to utilize advanced technological infrastructure when compared to other nations that are less industrialized. In this section, I examined previous papers pertaining to the challenges faced by underserved people dwelling in developed countries. It is crucial to develop a system that takes care of the major concern and previous obstacles faced by underserved populations while formulating any mHealth solution. In this way, we can understand the system requirements to build an efficient healthcare management system for this specific population of end-users.

When working with health data, it is important to follow a proper code of ethics. Alkabba et al. (2012) predicted the major 10 ethical issues as perceived by the participants facing the public and healthcare providers in Saudi Arabia: Patient's Rights, Equity of resource distribution, Confidentiality of patients, Patient Safety, Conflict of Interests, Ethics of privatization, Informed Consent, dealing with the opposite sex, Beginning and end of life, and Healthcare Team Ethics.

Hooker (2013) states that many citizens of the United States are medically isolated with its convoluted and complicated healthcare network. Community Health Clinics (CHCs) sprung up around the country to meet the needs of medically underserved populations in the 1960s. An understanding of why those working with the medically underserved and economically excluded do so is required because the exigency for services often outspreads beyond what is typically distributed in a traditional health care.

Strasser (2003) investigated around the global population of underserved populations to

find that individuals in rural areas are poorer than those in urban areas. Additionally, infant mortality rates in rural regions are 1.6 times that of urban regions in South Africa. 77 percent of rural children were more likely to be underweight or under height for their age; 56 percent lived over 5 kilometers away from the proper health facility; 75 percent of people living in rural areas were likely to be poor. Thus, the failure of proper health management can be attributed to the challenges as mentioned above. Reaching them directly to provide healthy information will help them to improve their overall health behaviors.

In reference to communication patterns among underserved populations, Best et al. (2017) investigated the communication gaps and opportunities to treat and screen cancer among underserved populations, who were medically and economically too poor to afford the treatment. They have designed the 'Sorenson's integrated model' that includes four coordinated dimensions such as access, understand, appraise and apply. The patient-level interaction should be improved to promote access and screening of cancer more than the system-level interactions. Addressing the entire society using a community-engaged approach was more advantageous; it was used to predict the people with the risk of cancer/ or medically underserved. They have come up with the solution that, better communication approaches were needed to increase their visibility and utilization.

Within mHealth, Hall et al. (2014) reviewed the impacts of mobile phones and its tremendous growth in the field of healthcare. They have considered the people from the low and middle-income countries (LMICs). The review was carried with the evidence of 76 peer-reviewed papers published related to the mHealth. The impressions of mHealth interventions reported in those papers were classified into common mHealth applications. The findings showed that there is strong support for growing demand for mHealth interventions in LMICs, especially in enhancing treatment compliance, appointment adherence, data collection and support networks, while there were some limitations related to the quantity and quality in these same mHealth services. Finally, they have concluded that by conducting a pilot study in the larger population and designing the mHealth application for the specific needs of the people could overcome these limitations.

Data mining techniques to facilitate better health outcomes within underserved communities is also prevalent in the literature. Aljumah (2011) investigated the datasets of Non-

Communicable Diseases (NCD) report of Saudi Arabia with the approval of World Health Organization (WHO). They employed regression techniques for mining health data, which helped to identify those methods of treatment best suited to treat hypertension. The Oracle Data Miner (ODM) was used as the platform for analyzing the treatment method adopted by the Blood Pressure (BP) patients in the age group of 15-64 years. The study concluded that prevalence of hypertension is getting increased in Saudi Arabia and 1/4th of its population was affected. They also suggested that controlling and managing BP at an early stage was critical to prevent adverse effects of the symptoms later on. These interventions could be in the form of suggestions to patients to change their dietary pattern, or through recommendations on medications, weight reduction, the cession of smoking and exercise.

2.3 Underserved U.S. Populations

When it comes to spending on Health Care, the U.S. spends more money than any other country in the world (Etehad & Kim, 2017). Unfortunately, many underserved population still do not receive critical health services, receive services too late, or receive lower quality services. Consequently, the government needs to offer better medical amenities to tackle the difficulties in the availability of medical services and the trend of increasing ailments by its Mobile Health Units (MHU) (International Committee of the Red Cross, 2006).

Vander Wielen et al. (2015) discovered that the medicinal services framework in the United States is loaded with aberrations related with a panoply of components including complex associations between race/ethnicity and financial status, geographic access to mind, and medical coverage status. For example, protection status, more commonly referred to as health insurance, is a critical requirement in cancer screening and a relationship between race/ethnicity and newborn child death rates is a common theme across scholarly writing. Additionally, people living in well-being proficient deficiency regions are more averse to receiving solutions for cardiovascular malady counteractive action, including statins and warfarin, particularly when uninsured.

More than 57 million people live in 5,864 assigned essential care deficiency territories in the United States (Nunez-Smith et al., 2011). People in these urban and rustic groups confront a deficiency of essential care suppliers in four essential care fortes: general or family home, general inside medication, pediatrics, and obstetrics and gynecology. Albeit essential care doctors make

up just 37% out of the doctor workforce, they give 56% of all doctor office visits. Specialists contend that the United States will confront a genuine lack of essential care doctors sooner rather than later, likely lessening further the entrance to essential watch over restoratively underserved people.

Vega et al. (2009) reviewed that the present epidemiologic investigations are deficiently intended to fight with the chain of multi-causality and time-requested relations that result in the multidimensional connection of small-scale and full-scale level causal variables. As Kaufman and Cooper (1999) have observed, social determinants are comprehensively connected to 3 spaces of race, sex, and class, "mirroring the complex pathways through which they are thought to influence or decide exposures, practices, physical constitution, and other immediate or contributory reasons for illness." Therefore, we make no affectation that we can show causality from existing epidemiologic confirmation. Instead, we apply deterministic thinking to the condition of learning concerning impacts of social imbalance and insignificance on the wellbeing and mortality of underserved populations. In particular, there is a great need to address the underserved Latino population in California, where, as reported in 2006, 22% currently live below the poverty line. This can be contrasted against the 10% of white, non-Latinos. Additionally, Latinos have reduced rates of medical coverage compared to other ethnic gatherings; 40% of Mexicans and Mexican Americans, 26% of Cubans, and 21% of Puerto Ricans were uninsured in 2006 as contrasted with 16% of white non-Latinos. This is a particularly articulated issue among outsiders since they are more averse to be qualified for open protection (National Academy of Sciences, 2006).

It is important to acknowledge the real medical issues facing underserved populations, including cardiovascular sickness, which is an essential driver of premature death, diabetes mellitus, which constituted the top weight of mortality differences (as characterized in this survey). Additional issues incorporated into the audit were business-related injuries, human immunodeficiency infection (HIV), liver ailment (counting liver cirrhosis), and particular sorts of stomach related framework diseases, including cervical, stomach, and liver cancer.

2.4 mHealth Projects

Ben-Zeev *et al.* (2015) studied that the healthcare initiatives for an innovative interdisciplinary field called Mobile Health (mHealth) are facilitated by the widespread use of

handheld devices and mobile phones. Evidence-based mHealth interventions for a variety of mental and physical health problems are being developed by many researchers worldwide (Heron & Smyth, 2010; Kaplan & Stone, 2013).

Bobrow *et al.* (2014) developed a new method, called the SMS text-message Adherence Support trial (StAR) that was aimed to examine the efficiency of the system in enhancing the treatment and controlling the BP than the traditional methods. The system is implemented in the practical environment for the treatment of hypertension in Cape Town and South Africa. The research was intended to create the program for the healthcare center appointment, a reminder for taking medications and information related to hypertension that is sent to the patients through an automated system with informational or interactive SMS depending on the queries. In contrary to this, this study focuses on developing a system that can gather day-to-day healthcare practices of patients via SMS and analyze their behavior by applying machine learning algorithms. The StAR trial system was economically affordable for the patients to communicate with the doctor and to quantify the results of the clinical tests. The research forms the platform for interventions of the mobile phones for the delivery of the healthcare information and retrieval of information from the patients those who are in a low-resource setting.

Bobrow *et al.* (2016) examined the influences of StAR for the automated supportive treatment for BP through the text messages. The new realistic tool was tested with the South African patients who are taking treatment for hypertension. The group was formed in the ratio of 1:1:1 for informative SMS, interactive SMS and usual care. The patients were monitored for 12 months from the date of the test. Out of 1,372 participants, 1,256 participants actively participated in the research. The results show that the 95% of participants in interactive SMS method showed the changes in their systolic BP. The results showed that the interactive method holds the proof for the successive recovery.

Sosa *et al.* (2017) developed the SMS/ MMS application concerned with the Sense Health for the diagnosis and treatment of neck and head malignancies. The application tends to collect and give the relevant information to the patients through messages. The research recommends that for the head and neck cancer patients, the digital intrusion methods may expand the chances of treatment in the postoperative care. The patients told that they find it

very useful, effective and easy to access. The research further aimed to focus on the people with cancer and provide them with suggestions regarding treatment and therapy according to their need. The review showed that the text messaging could reduce the money being paid to the third party and reach the patients in the rural area through text messaging.

Quintana and McWhirter (2017) designed an online and mobile system to assist and address the maternal and newborn health outcomes, antenatal care attendance, raising the rate of vaccination to the infants and to stop the unhealthy behaviors, such as smoking and consuming alcohol. The system will be successful only when it tries to fill the gaps in the cultural behavior of the people and how it addresses the issues given by the people. The level of acceptance of this new method depends on the trust level of the patients and healthcare providers. However, there were many deficiencies in the design and estimation of global maternal mobile health systems, the previous studies related to this research gave the trusted and promising chances for its successive adaptation in the future and the sustainable system to attain the long-term behavioral results.

Kazi *et al.* (2017) studied the viability of text messaging-based mHealth among patients from the rural and remote areas of Kenya. The mHealth intervention focused on the betterment of antenatal care and periodic immunization among children in that area. The survey was conducted for about a year and 92% of the participants likely to receive the SMS on a weekly basis from their consultants. The results indicated that pregnant women and their caregivers visiting the clinics were keen on getting these critical messages from their doctors. They also claimed that, if the proposed model was designed properly; it would be a new and innovative way to fetch the attention of the women in maternal care to improve their health and to attain the justifiable health goals.

Rehman *et al.* (2017) studied the existing works related to the proficiency of the mHealth inferences in controlling and improving the care for hypertension and hyperlipidemia. They have found that more research must be done related to address hyperlipidemia's risk factors and its treatment. It also noted that the resources that are available and the study population are crucial for effectively measuring a mHealth intervention. Potential solutions can be provided by SMS modalities in the low-income settings. In more resource-abundant settings, the use of wireless devices and mobile phone applications will only grow in the future. However, even though SMS

technology is relatively mature, considerable gaps in mHealth knowledge persist, including an absence of comparison between the lack of long-term outcomes data and different interventions, in spite of ample data. Evidence of cost-effectiveness is also essential. Expert guidelines using mHealth have been published lately for reporting of health interventions. These guidelines stress the significance of considerations like cost, local infrastructure, and contextual adaptability. For providing more meaningful context to understand the findings and improving the quality of mHealth reporting, future researchers should integrate these guidelines.

Rubinstein *et al.* (2016) proposed a mHealth intervention to assist in weight loss and improve the physical activities through the limited resources. They enrolled adult men and women with pre-hypertension from the urban areas of Buenos Aires, Argentina and, Guatemala. Qualified patients at the age of 30–60 years had the Systolic Blood Pressure (SBP) in the range of 120-140 mmHg and the Diastolic Blood Pressure (DBP) in the range of 80-89 mmHg. Participation required that patients not to take drugs for hypertension and had to own a mobile phone. The researchers examined healthy and unhealthy behaviors, usage of mobile phones and habit of text messaging among the people aged between 30 to 60 years. The study failed to show any improvement in the blood pressure level, but showed the significant changes in the case of body weight and dietary habits of the patients. They have observed that the dose-response effect has gained the attention of the people in low-resource (SMS) environment. They also suggested that more research has to be done in the low resource settings to provide healthcare for the patients with cardiovascular diseases.

Person *et al.* (2011) involved people infected with or having the risk of HIV or tuberculosis (TB) and analyzed their attitude, knowledge and behaviors toward mobile phone usage and text messaging regarding the health care among them. 241 participants had their mobiles out of 315 participants. The results surveyed that the participants prefer mHealth technology for overcoming difficulties in transportation, remembrance of appointments with the doctor, text messages to reminisce about the medications. An interesting finding was that positive attitudes towards healthcare-related text message reminders could be found in populations who did not currently possess mobile devices.

Ni *et al.* (2014) studied the Wireless Heart Health project launched by China for the

prevention and care of Cardio-Vascular Diseases (CVDs) in underserved communities and addressed the potential drawbacks in rural communities. The mHealth program was modeled in such a way to assist the people from rural areas who could not afford the care in the urban area. The population would be provided with 3G smartphones with web-based Electronic Medical Records (EMR) software, built-in ECG sensors, and Internet-ready workstations to community health clinics in the rural areas. However, there are many drawbacks, adults couldn't comprehend the system and giving prescriptions and treating patients through mHealth service was not authorized. Furthermore, there was a lack of regulations and uncertainties in the standard which make it challenging for companies to design the mobile application for mHealth care. But they hope that in the near future, Chinese rural people may find mHealth particularly useful. These people have limited access to quality medical care. The mobile technologies can reach these people across geographic and socio-economic boundaries. Such technology can potentially maximize the access to care and improve healthcare outcomes.

2.5 mHealth Studies related to Data mining

Data mining is defined as the process of extraction and mining of data from vast amounts of existing data (Han & Kamber, 2000). According to the report by Yuan et al. (2013), it is deemed that only in the United States, the use of data mining has aided to save more than 450 billion U.S. dollars every year through Health Informatics (Herland et al., 2014). In this regard, it is imperative to examine the different data mining techniques used in the healthcare sector.

Altaf *et al.* (2017) studied the applications of '*Association Rule Mining (ARM)*' in the Health Information System (HIS). The literature review from 2005 to 2014 explored that ARM technique is still widely preferred than the other mining techniques. The conclusion drawn from the literature surveyed explains that ARM is used as the decision support system for HIS because the HIS does not rely on ARM for making the decisions. The research recommends that HIS should use 'Big Data' or cloud computing technology for storing, retrieving and managing the data to improve its accuracy in data mining and data classification. Though ARM lacks in accuracy, it was applied to the relatively small number of well-known diseases with high mortality rate. Additionally, the drawback of applications associated with ARM must be predicted and suggestions must be given to eradicate the limitations found.

In research by Khamis *et al.* (2014), the k-NN (k-nearest neighbor) algorithm was applied to identify the chances of the presence of existing ailments and to increase the automated diagnosis of diseases with similar symptoms. The research considered several variables such as accidents history, the age of the patients, allergies, BP, total cholesterol, smoking habit, family history of heart disease, hypertension and diabetes, lack of physical activity, and obesity. At first, they have examined the accuracy and efficiency of the k-NN algorithm to improve its performance. The algorithm was applied to the patients who needed the emergency treatment and produced the results with low error rates and taken only a few minutes to diagnose the diseases correctly and give suggestions for therapy to the predicted disease. Secondly, they evaluated the factors that lead to the decline of the k-NN algorithm for data mining in the rural areas and found administrative expenses and classification efficiency played the significant role in implementing it.

Boukenze *et al.* (2016) conducted their research on the development of Big Data in a healthcare system and applied the '*decision tree learning algorithm*' to the data for classification and data mining. The research focused on classifying Chronic Kidney Disease (CKD) using C4.5 Decision Tree algorithm. Out of 396 instances, only four instances were misclassified with the accuracy of 99.63% and the error rate was 0.37%. They also claimed that C4.5 algorithm is excellent because it obtains the performance and accuracy value, KS (Kolmogorov-Smirnov value) as 0.97. The proposed C4.5 algorithm showed its excellence regarding precision, minimum execution time and accuracy which makes it a good classifier to be used in the medical field for classification and prediction. Accuracy and execution time of the C4.5 classifier makes it as a powerful machine learning tool to achieve accurate information from data logically and quote it as the eminent classifier of the knowledge.

Chaurasia and Pal (2013) proposed the data mining technique known as '*bagging algorithm*' for the accurate identification of the risk factors for heart disease with the fewer attributes. The tests were implemented using the Waikato Environment for Knowledge Analysis (WEKA) tool and the results were compared with the bagging and unbagging algorithm with 10-fold cross-validation. The proposed algorithm provided the result which can be easily read and interpreted by a human. The results obtained for the bagging algorithm indicated the accuracy around 85% and took 0.05sec to build the model and diagnose the patients with heart disease. This research finds its application predicting the people who need particular attention.

Omboni *et al.* (2016) described a novel data mining method for the classification of chronic disease using a U.S. risk behavior factor data set and demonstrated the application of the proposed method using a case study of depressive disorder. The proposed method involves three steps such as build the data mining model, innovative method to analyze the high dimension data and visualization method to communicate with the clinicians. This study focused on data mining techniques to classify depressed patients who were living with the childhood experience of sex abuse and mental torture. To visualize the classified results in the better manner, they have employed the techniques of the heat map, initial exploratory visualization and illustrative decision tree. This idea was more useful for the researchers to combine with the clinical domain expert for the classification and identification of disease. Correlation and causation are the significant problems associated with the data mining process. The classification methods used in the study mentioned above described the plausibility of variables obtained to be clear and the causation between the variable and the depression to be unclear. Hence, the model was considered as bidirectional and should be considered only as a correlational model and not as the prediction tool.

2.5.1 Data classification techniques

Adebayo (2017) developed a model that predicts and classifies hypertension with the level of risk associated with it. Patients with a history of receiving treatment for hypertension were selected and clustered using decision tree algorithms. The predictive model used the decision tree algorithm, C4.5 and ID3. The proposed method was simulated and tested using the WEKA tool. Outcomes revealed that the ID3 outperformed the C4.5 method with 100 percent accuracy of predicting the risk factors associated with hypertension. The variables associated found from the algorithm using the path rules assists the cardiologist to focus on the small risk factors for recognizing hypertension at the early stage.

Zriqat *et al.* (2017) developed a real-time smart medical decision support system based on data mining methods. They have selected the five data mining algorithms with massive data sets to access and examined the risk factors numerically associated with CVDs to compare the performance of the proposed classifiers. Classifiers such as Naïve Bayes, SVM, Decision Tree, Discriminant and Random Forest were implemented using the MATLAB tool with two data sets to prove their effectiveness. The results showed that all algorithms were predictive and gave an

equivalent answer as the traditional methods. The decision tree algorithm provided 99% of accuracy next to random forest algorithm and outperformed the remaining algorithms. However, the research did not apply their classification system for enhancing the medical care at the lower costs and to address the needs of the underserved people.

Prasad *et al.* (2016), used machine learning to predict positive disease growth in the thyroid gland and designed an expert advisory system through a hybrid architecture. Data was collected from a questionnaire, given to patients with the predisposition to thyroid ailments. The researchers proposed a learning method to diagnose thyroid conditions and results obtained could be easily interpreted by the doctor. The empirical studies show that the proposed algorithm can reduce the time and attributes required and increasing the accuracy level and can reduce the size of the data to be stored. So, the proposed method was combined with the RST and ML algorithm. This method utilizes the entropy to select the bin boundaries of the patients and the missing attributes values can also be restored and their similarities were measured. Thus, the hybrid architecture (combination of Rough dataset theory and Machine Learning algorithm) introduced in this research provides the accuracy in the range of 99-100% for diagnosing the thyroid and predicts severity level of disease.

2.6 Cloud computing and healthcare

Cloud computing considers any computing that is performed on the web / Internet. The Internet is not only about data storage, and Internet technologies can be used to implement high-resource operations, run application/software the same way our computer at home can do. Today, many widely used applications operate over the Internet, such as social networking software and banking applications, which use mobile phone software to interface with cloud computing services. The proposed system uses cloud computing to perform numerous operations (i.e., data mining) on the cloud server. The proposed system leverage computing programs constructed in PHP and Python, which run on the Internet.

In Botts *et al.* (2010), the impact of cyberinfrastructure on healthcare services was evaluated. The research gave attention to the design the architecture for Personal Health Record (PHR) called HealthATM. The proposed system combines services from Google's cloud computing environment. The HealthATM architecture was simple and the information from and

to patients was processed by the external services such as Google Health and HealthATM. The service was hosted on the web and used XHTML and CSS design to present information to patients in a standard web browser. The model discussed here explicitly focused on underserved populations seeking universal access to their healthcare. Moreover, the cloud computing environment provided flexibility for designing the affordable and scalable PHR systems. This architecture resembled a banking ATM and provided an easy interface for patients, consumers and healthcare providers to manage a patient's health. This system forms a great leap towards the better health management of the underserved application.

Parameswari and Prabakaran (2012) implemented an application for assisting and managing the health care system for the hospital applications on mobile internet in the cloud environment. The application enables the user to download the healthcare providers' application to get the appointment from specialists and can analyze the symptoms to get first aid. They also tried to implement the online interaction of patients with the doctors and able to manage the expenses not to affect the affordability for the underserved people. The study aimed to develop the automatic computing system that adjusts itself concerning the requirements of the administrator and integrate the new components easily. The proposed architecture employed the several modules in its design. The PHR application collects the records from the patients and stores it in the cloud. The cloud storage supervises and maintains the elements which are responsible for the operation of the server. The Cloud Service module, which handles and queues user requests, is connected to the Cloud Platform interface. The interface module that is in the cloud processes the user queues in the cloud service module. Registration is mandatory for the user to subscribe and avail the eHealth services and user needs to provide the required documents for registration to the administrator. The low cost and quick access to information in relation to the healthcare are provided by the healthcare community with the advent of this application. This platform offers the easy interaction between the hospital administrations and the person who is seeking for health-related information.

Nishantha *et al*. (2009) proposed the system that was a functionally oriented, dedicated system for healthcare organizations. They designed the Integrated Medical Information System (IMIS) with the joint effort of software engineers, healthcare providers and the Sri Lankan government. The effective medical collaboration forms the primary platform for the eHealth

services in Sri Lanka. IMIS was created by combining the International Medical Education and Collaboration System (IMECS) with the newly proposed hospital information system (HIMS). IMIS was the custom-made application designed for the institution and it does not need internet, high-speed transmission and costly equipment. A Sharable Content Model (was employed to retain data to enable flexibility in sharing and reusing. The contents may be of any format and are stored in the content store as ASCII data. IMECS also incorporates a web-based asynchronous system that manages the handles patient information system. The real-time application used the cloud environment and provided the sustainable eHealth environment.

Jiang *et al.* (2010) designed a BP monitoring system to diagnose hypertension at an early stage. The proposed architecture incorporates three parts such as server domain, user domain, and the channel domain. The experiments were conducted to evaluate the durability and reliability of the wearable BP measurement device. They developed a system that selected 4 standardized BP values at the outset, which were obtained from the device consists of 3-V digital power device (mercury manometer, digital oscilloscope, the PCB board, and gas balloon pump) Secondly, they presented the standardized BP value which was acquired from a gas balloon pump that was managed by a mercury manometer. Then, with a digital oscilloscope, they measured the digital value of the measurement device. At last, they gathered all data and evaluated the results. The examinations were meant to fill the hypertension control gaps, such as lack of hypertension records management and lack of continuous monitoring of the patient, etc. In the recent time, the proposed method was modeled and re-examined. Moreover, the result of the wrist BP measurement was sufficient to make a product from the prototype after progressive enhancement. They have concluded that the proposed work was feasible, and a redesigned system aims to be more affordable and reliable.

Pappachan *et al.* (2014) proposed a system called Rafiki which uses wearable and mobile computing devices which assists Community Health Workers (CHW) in decision making and enables alliance among them. Rafiki's semantic representation of data makes it simple to share information related to a symptom, disease, demographics of the patient, and diagnosis guidelines. CHWs use simple forms for gathering qualitative information about the patient by filling as the required fields to diagnose the disease. The forms represent the doctors and other health-care providers in health centers and hospitals while collecting information from the candidates. The

Rafiki system assists the Community Health Workers, guides by way of the diagnostic processing and facilitates the collaboration among personnel involved in community health-care. Besides, Rafiki offers different communication mechanisms to share useful information among personnel. The following advantages make Rafiki a successful one including proving access to health materials and providing of decision support for CHWs, usage of patient context to reduce the number of questions asked, detecting emergency situations automatically and providing general healthcare assistance.

For the management of hypertension, Omboni *et al.* (2016) concentrated on a specific application of telemedicine and surveyed the research related to hypertension. The remote data transmission of BP is enabled by Blood Pressure Tele-monitoring (BPT) and is also possible to transfer additional information on patients' health status from their remote to the hospital or the doctor's office. As per the current statistics, 19% of the measured population use a health app whereas half of the smartphone owners get the health information from their phones. The most popular apps among the hypertensive patients are those applications that can track function such as BPT. Therefore, in the management of hypertensive patients who seek the improvement in the quality of delivered care and to more efficiently address the cardiovascular consequences of high BP, e-health, more specifically mHealth and BPT, are becoming increasingly important. With This evidence, they have concluded that in hypertensive patients necessitating a tighter BP control, such as those at high-risk, BPT and telemedicine, may be useful. E-health solutions will assist in designing the networks between healthcare professionals for the management and improving the screening of hypertension and other related comorbidities. Thus, it is possible that an effective prevention of cardiovascular diseases can be brought into practice.

CHAPTER III: DESIGN, ARCHITECTURE AND COMPONENTS

3.1 Introduction

With the advent of cost-effective mobile phones, there has been an upsurge in mobile phone usage in low and middle-income countries. mHealth service is a powerful tool to combine the health care services with the individuals who are the registered users seeking for the improvement of health care. The real challenge for mHealth is to address the needs of the underserved people, especially in rural and remote areas. The work in this thesis focuses on understanding the demographics of the community of the user, usage of the mobile and feasibility among patients in rural and remote areas through SMS based mHealth intervention for improvements in diagnosis and treatment for major health concerns like Hypertension. The methods for achieving our goal are described below.

This chapter describes the testing of an implemented web application for healthcare domain and simulating the research parameter for the analysis of data. Testing of the web pages is related to the research questions that guided the study. The designed web server is simulated and tested with the local population to find out the difficulties faced by them to take the health advisory from the healthcare provider. The research process is required to determine the need for ongoing terminal care teaching about the health care among the selected population.

3.2 Research Gap

The research discussed in prior sections focused on the pressing need for additional mHealth solutions for underserved populations, one that incorporates data mining and classification techniques. This thesis aims to introduce an easy to use system that supports healthy behavioral outcomes by leveraging low-cost components while increasing access. While existing studies have helped underserved populations to a large extent, there are still challenges faced across these communities. I am to address some of the difficulties facing mHealth listed below and:

- Inspire patients to use mHealth care technology through easy to use, low-cost SMS interactions.

- Integrate complex data mining and data classification techniques that utilized patient-driven SMS data.
- Leverage cloud computing technologies to power both SMS communication and data mining classification.
- Overall, design a robust, yet affordable mHealth system, which can be leveraged by underserved populations.

3.3 Design

The present research aims to build a system for managing the health of underserved people by making the information easily accessible to them. The system utilizes short messaging services (SMS) for a low-cost and efficient way to connect with patients.

SMS via mobile phones, commonly referred to as text messaging, is an influential tool for behavioral change in patients. This includes underserved populations as well. Because of the ubiquity of mobile phones, SMS services are relatively less expensive than Internet data packages and are less complicated to use. The proposed work provides healthcare services to populations through SMS messages and monitors patients through simple predefined request-response data queries. This work should be of interest to underserved populations because of the low barriers to entry and low cost. Moreover, healthcare specialists can utilize patient responses to form a more comprehensive history of a patient and store them in the cloud for advanced data mining later on. Data mining tools such as classification and clustering algorithms can provide more significant analysis of a patient's data, which can be used to influence user behavior, provide health-related suggestions and/or expert medical advice and recommendations. Information within a patient's profile can be compared against standard health guidelines offered by healthcare experts. SMS messages are received daily from the individuals and notifications are sent on a weekly basis either on weekdays or weekends at a scheduled time for each user.

3.4 Design Goals

Specific design considerations for this system include the following:

1) Design the architecture for mHealth systems using the Twilio Application Programming Interface (API) which runs over SMS to interact with patients.
2) Implement a mHealth system that can monitor, evaluate and provide recommendations on their health status in real time.
3) Apply classification data mining techniques to analyze the trends and behavior of an individual.
4) Present the information back to users in an easy to use interface.

3.5 Architecture of the mHealth System

The general architecture of the mHealth system is illustrated in Figure 1. The proposed design of mHealth system constitutes different components such a cloud web server, central database management, data management module (the web portal for the health providers), Twilio server and the user. The server has a database management module for storing and handling patient data. The data management module is comprised of three sections, which is responsible for collecting the information from the user, storing and processing the information and mining the data for classification. The Twilio web service acts as a communication tool between the user and the system and sends suggestions to the user via text messages. It is used to programmatically send and receive SMS using its application programming interface (API).

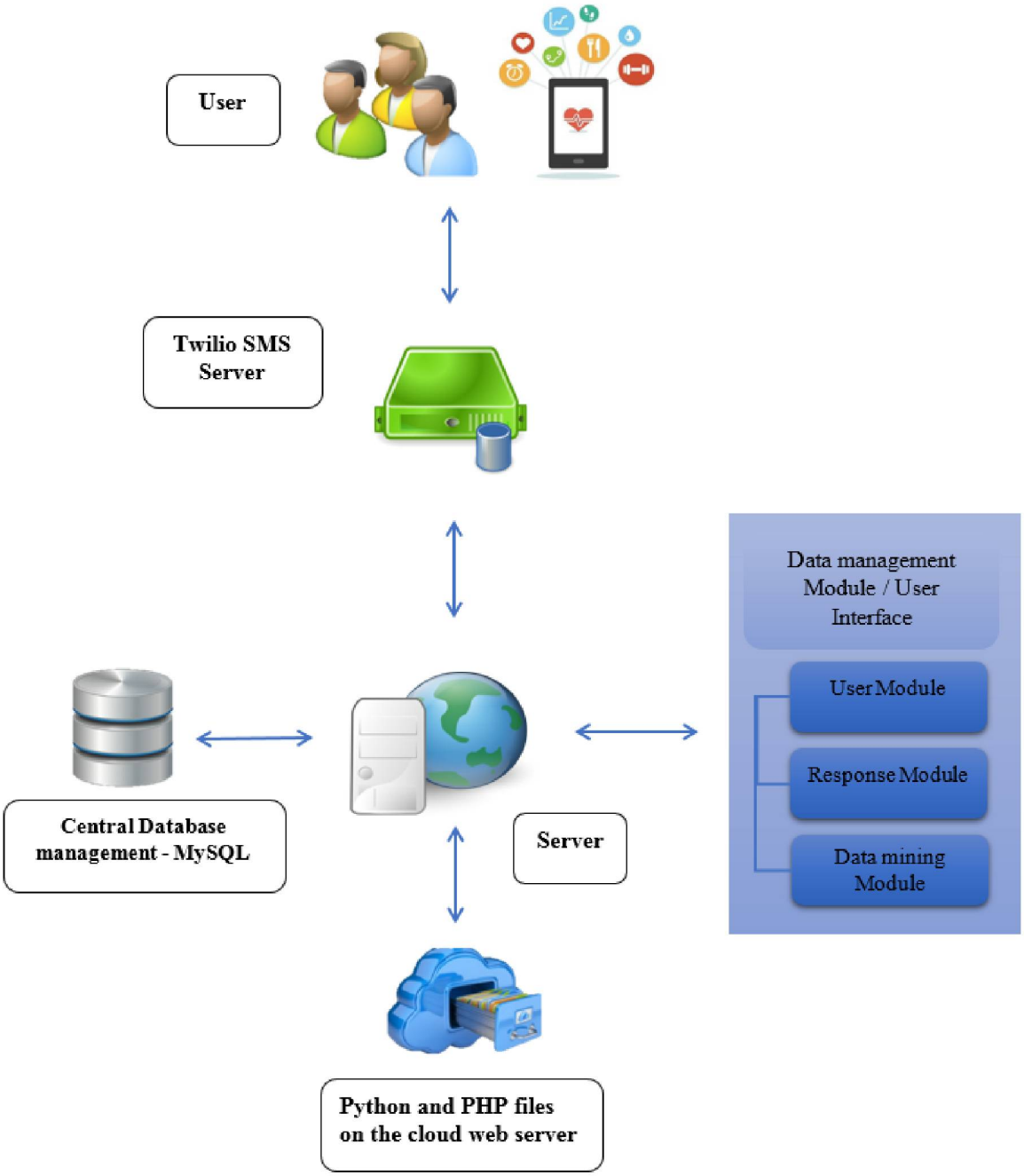


Figure 1: Architecture of the mHealth System

3.6 Framework

The process flow diagram includes the procedures of interaction between the user and the data center via SMS. The framework shown in Figure 2 was built in such a way that it collects information from the user on a daily basis, classifies and mines the user's data and gives valuable suggestions to improve their health. To implement the above-said process, the introduced framework comprised of the three modules. These modules consist of the input and output processing module (i.e., user module), a response module and data mining module.

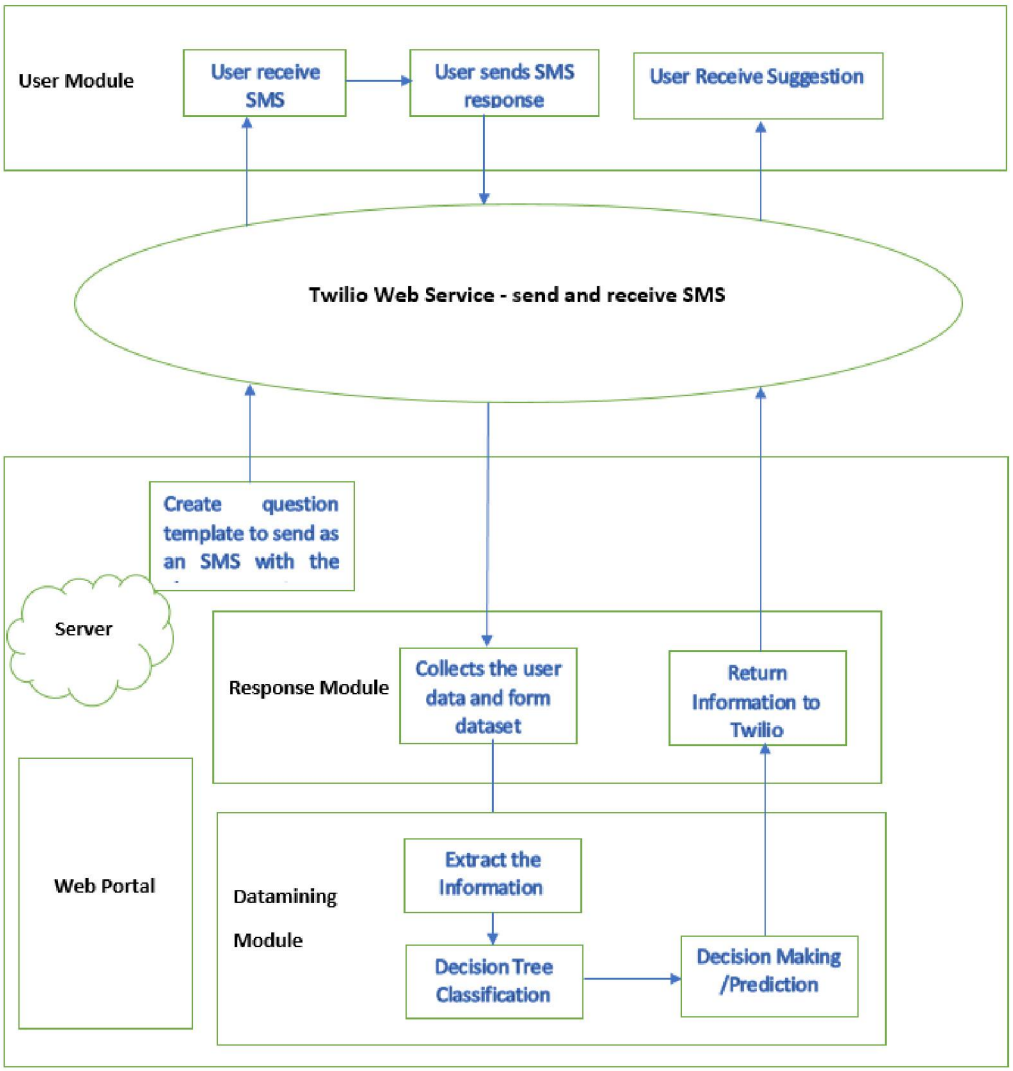


Figure 2: Framework

In the user module, data is collected from the user through the questionnaires and the responses from each user stored in the response module as the dataset. The system mainly concentrates and prepares the questionnaire on three specific behaviors, including sleeping, eating, exercise (i.e., the number of steps walked). Questions will be sent to the user and patients were asked to send the reply to the questions. Since there are three questions, the system will send one question at a time and wait for a response to that question before sending the next question. In case the user does not respond to the first question within an hour, the system will resend the same question again and wait for the user response. If the user still doesn't respond, the system will stop sending SMS to that user for that day. The message responses from an individual user are gathered classified and maintained as the dataset.

In the Data mining module, we use a pre-trained model based on the decision tree classification algorithm (model having an accuracy level of 98.3%) to predict the patient health status based on the dataset collected over a period of a week from the patients and provide patients with suggestions to better manage their health. The Twilio API acts as a communication module for sending and receiving responses from the user module. The web portal is designed where health providers can check on the health status of each patient. Health providers can access different sections of the web portal via database table views, which provide access to the SMS answers, questions and customizable reports.

3.7 SMS Technology

3.7.1 Twilio

The framework utilizes the Twilio application as the messaging platform for collecting information. Twilio is a messaging platform based on Platform as a Service (PaaS) model of cloud computing and it charges its user according to the usage which was developed by the company in San Francisco. Twilio authorizes and allows the programmer to design custom patterns using the programming languages such as PHP, Java, Python and Dot Net for sending and receiving text and phone calls.

Twilio is hosted using Amazon Web Services, which offers connectivity between the Public Switched Telephone Network (PSTN) and Hypertext Transfer Protocol (HTTP) through its

application programming interfaces (APIs). It adheres to the architectural design for enabling the protection against unexpected attacks, rewarded by Amazon Web Services in the year of 2011. Twilio also allows the open source community to produce much Open-Source Software (OSS) with security and integrity of data. 'Stashboard' is OSS programmed in Python was developed by Twilio that can run on any API to check and display the functionality of the API. Twilio also offers a local tunnel, which allows the programmer to display their local development to the public internet (Twilio, 2017). The open-source projects are available on the Twilio website are Flask Restful, shadow, and Banker's box.

3.8 Dataset Configuration

To make predictions on a given set of user input, the system was trained on a sample dataset. The dataset was prepared based on Table 1 as shown below and consists of four columns Age, Number of Hours Slept, Walking Steps per Day, Number of Calories consumed and was formed using information found on the internet (health.gov, 2017). The data requirements were based on user age, number of calories consumed, number of walking steps, number of hours of sleep in a day.

Age	**Number of Hours Slept**	**Walking Steps per day (500 steps = 5 mins)**	**Number of Calories Consumed per day**
15-17	8-10	11000-12000	2200-2400
18-20	7-9	11000-12000	2400-2600
20-50	7-9	7000-13000	2200-2600
50+	7-9	6000-8500	2000-2200

Table 1: Dataset for data mining

A dataset of 1,285,200 records were randomly generated using the above table. If the data (i.e., age, sleeping hours, walking steps and calories) is within the tolerance level, then the health status is "healthy" otherwise "unhealthy". The records are saved in a text file which is imported while training our model. This is an initial training set to show how the system might be used for underserved populations. This dataset is also used to train our decision tree classification model

which then predicts the health status of a new user input. The system performs various operations. In case the user is unhealthy, the system will send a text message stating the reasons for being unhealthy. For example, if the reason for being unhealthy is because the user has not slept for the required number of hours, then the text message will suggest them to have sleep for appropriate amount of time. The system will also compare the data with previous week to analyze if the health of a user has improved.

The figure 3 below shows a snapshot of the dataset consisting of 1,285,200 rows and five columns:

```
In [63]: runfile('C:/Users/MLPC2/Desktop/machine_learning.py', wdir='C:/Users/MLPC2/Desktop')
         Age   Sleep   Steps   Calories      Status
706467    34      10    7250       2800   Unhealthy
803014    37       5   12000        500   Unhealthy
1116206   46       1    1750       2700   Unhealthy
1771460   64       4    2250       2100   Unhealthy
1036082   43      12   14250        300   Unhealthy
97724     17      11   14500       1500   Unhealthy
381541    25       9   19750        200   Unhealthy
1526707   57       7    2750        800   Unhealthy
1189215   48       1   10250       1600   Unhealthy
1612829   59      13     250       3000   Unhealthy
661013    33       6    8500       2400   Unhealthy
110511    18       2    1000       2200   Unhealthy
1721558   62      13    6500        900   Unhealthy
1784896   64       9   14250       1700   Unhealthy
1195050   48       3   19000        100   Unhealthy
95961     17      10   19750       2200     Healthy
240047    21      11     500       1800   Unhealthy
1297691   51       1   14250       1200   Unhealthy
878196    39       6   18500        700   Unhealthy
1617247   59      14   17250        800   Unhealthy
1653729   60      15    1250       1000   Unhealthy
593702    31       8    7750        300   Unhealthy
1676398   61       9   10000       2900   Unhealthy
489579    28       9   20000       1000   Unhealthy
391763    25      14    4750       2400   Unhealthy
1745865   63       8    9000       1600   Unhealthy
1389864   53      10    2250       2500   Unhealthy
1331197   51      15   13500        800   Unhealthy
692301    34       4    9250       2200   Unhealthy
1436468   54      14   10750        900   Unhealthy
...      ...     ...     ...        ...         ...

69418     16      14   18500       2900   Unhealthy
1470382   55      13   13250       2300   Unhealthy
993442    42       9   18750       2300     Healthy
830535    38       2    1250       1600   Unhealthy
708809    34      11    6750       3000   Unhealthy
1233287   49       4   17500       1800   Unhealthy
987843    42       7   12250        400   Unhealthy
831378    38       2    8250       1900   Unhealthy
1243473   49       9    2500        400   Unhealthy
845890    38       8    9250       1100   Unhealthy
419603    26      10   16750       2400   Unhealthy
759461    36       2    9000       1200   Unhealthy
445223    27       6   10250       2400   Unhealthy
770495    36       7    1000        600   Unhealthy
1170088   47       8   10750       2900   Unhealthy
929596    40      13    6750       1700   Unhealthy
328568    24       2   18250        900   Unhealthy
1608746   59      11    6250       2700   Unhealthy
1484308   56       4    9250       2900   Unhealthy
359134    24      15   13000        500   Unhealthy
1721601   62      13    6750       2200   Unhealthy

[1285200 rows x 5 columns]

In [65]:
IPython console   History log
```

Figure 3: View of sample dataset

The above features were deemed as good since they are simple and independent. Selection

of features is important for machine learning solution. If the selection of features is bad, then no matter which algorithm we use, the results will not be accurate, a phenomenon known as garbage in, garbage out (GIGO). Since the dataset being inputted is labeled, this technique is referred to as supervised learning.

3.9 Data mining process

Data mining according to Ni *et al.* (2014) is defined as '*the process of the programmed analytical method to reveal the hidden relationships among data sets.*' Three commonly used data mining techniques are regression, classification and clustering. Clustering and classification are of prime importance when the datasets being processed are very large (Narwal & Mintwal, 2013). Clustering follows an unsupervised classification technique and forms the primary and basic step in the analysis to cluster the data. In explorative data mining, clustering is the usual process for statistical data analysis and has its application in the fields of machine learning, recognizing patterns, analysis of images, retrieval of information and bio-informatics.

3.9.1 Scikit- Learn – Machine Learning Library in Python

Scikit-learn is a machine learning library built for the Python programming language. Scikit-learn was first started as scikits.learn, a project by David Cournapeau, during Google's Summer of Code. The initial release of Scikit-learn was in June 2007 (http://scikit-learn.org/stable/about.html). Scikit learn can be used to solve a learning problem. A learning problem considers a set of n samples of data and then tries to predict properties of unknown data.

It is an efficient tool for data mining. Scikit-learn is built on the python libraries Scipy, Numpy and Matplotlib libraries. It has supervised and unsupervised learning algorithms and can be used to solve classification, regression, clustering and many more machine-learning problems. Machine learning is a subfield of computer science by which computers can learn themselves without being explicitly programmed. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data. (Machine learning, 2017). Machine learning algorithms construct models from inputted datasets and make data-driven predictions, rather than relying on an explicitly programmed set of instructions. Machine learning

algorithms are divided into supervised learning and unsupervised learning. Within supervised learning algorithms, every input is linked to the desired output, which, in this case, are the health attributes of a patient and whether or not these attributes represent healthy or unhealthy behaviors.

This section is designed to analyze the effectiveness of the web application developed and the answers provided by the users regarding the health care questions. The research has mainly focused on people with hypertension and provides advice to them for improving the current health status. The questionnaire was developed using the English language, and it was provided to the user daily. Instead of bulk messages, on a one-by-one basis, five questions are sent to patients on a fixed timeframe. Users must provide numerical input because the servers can easily process numeric values more than free-text formats. Every question is designed in such a way that included an example how the patient must answer the question. The data are collected on a regular basis to know the probability of developing the disease that is associated with hypertension. The reports section contains the details of the data that are mined and analyzed.

To predict if the person is healthy or not based on the inputs given by the user, a Decision Tree Classification Algorithm is used. The Decision Tree Classification Algorithm is a supervised learning algorithm and can be used for solving classification and regression problems. The decision tree creates a training model which can be used to make predictions by learning rules inferred upon by the training data. The decision tree solves the problem by creating a tree-like representation of the training data. Each internal and leaf node corresponds to attribute and class label respectively. The decision tree is easy to use data-structure and follows the same rules or approach for making decisions.

When formulating rules, the dataset is first divided into a training and testing set using the train_test_split function. In this case, 33% of the total data is considered as the testing set and the rest was training set. Once the data is divided into training and testing set, a fit method was applied on the decision tree classifier to train the model, which makes a tree representation of the dataset and is then used to make the prediction.

3.10 Summary

Text-messaging and SMS via mobile phone is an influential tool for results on health improvement and the behavioral change of underserved people. This work connects healthcare services to users through messages and monitors them through the dataset which they are providing. Data is passed back and forth between user and system through the Twilio API. Messages are received daily from the individuals and the notification from the system is sent on the weekly basis either on weekdays or weekends at the scheduled time for each user. Data is processed through a decision tree algorithm and information is extracted to provide users with additional services, including health trend monitoring.

CHAPTER IV: IMPLEMENTATION

4.1 Introduction

The goal of this thesis is to design a simple to use, a cost-effective system using SMS and artificial intelligence to help underserved populations in the rural and remote areas to better manage their health behaviors. As previously discussed, these populations are familiar with the text messaging services because of the technological advancements in the communication domain, which makes the mHealth developers an easy linkage with rural people.

The present chapter elaborates on the implementation of the mHealth system.

4.2 Implementation procedure

The model is developed using the PHP (Hypertext Preprocessor) platform. PHP is an open source general-purpose scripting language, which is used widely in interactive web development. Moreover, PHP can be embedded directly into HTML. Back-end processing, such as connecting to a database or displaying necessary patient-driven web content, is created by PHP and used for the basic functionalities such as Create, Read, Update, and Delete need to be enabled in the system.

Text editors, such as Sublime Text and Notepad++, were used for programming the backend system. XAMPP was employed as the web server for saving the website files created and to save the data obtained from the database apart from the usage of traditional web browsers which requires more RAM. These three components are required for developing the web page.

Once the PHP software was selected for creating the new projects, the administrator is provided with the default files for creating, editing and testing the files with .php extension. The steps involved in the process are:

- Setting up the server
- Creating the public HTML Pages
- Creating the database and its tables
- Adding users to the database
- User log-in: Authentication
- Setting up the home page for Logged-in users and Logging-out

- Testing Page Security
- Adding data to the list - User Access Only
- Displaying data on the homepage
- Editing Data
- Deleting data.

Figure 4 below shows Entity Relationship Diagram(ERD) of the database. Entity Relationship Diagram is a graphical representation of information which illustrates how different objects (or entities) in the system related to each other. For example, referring to the below figure we can define one to many relationships between customers and reports table, that means each customer can have multiple reports, as they are generated every week and so on.

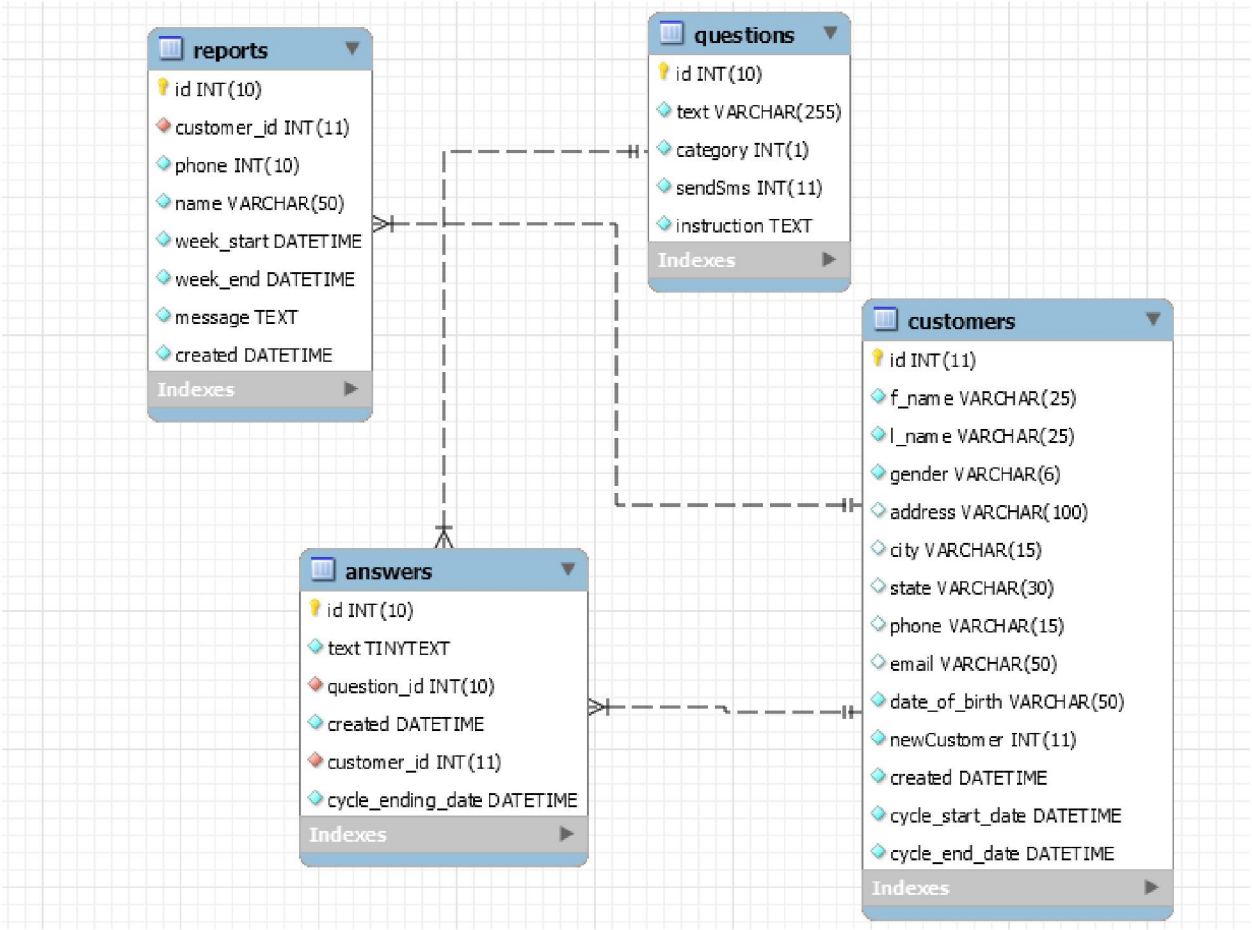


Figure 4: Entity Relationship Diagram

The data mining module is the only module that is developed using Python programming language. The reason for choosing Python is because python has the Scikit-learn library readily available, which includes a decision tree classification package for supervised machine learning. The Python file is stored in the same directory on the server where all the PHP files are stored. The Internet service provider is GoDaddy.com, which provides the option of creating a cron job, which kicks off the python script once every week to send out predictions and suggestions via SMS to the patients. GoDaddy, Inc. is an American company for Internet domain registrar and web hosting.

4.3 Modelling the System

4.3.1 SMS Module

The mHealth model is meant for doing the two automatic functionalities that are hosted on the server. First, it will send questions automatically to individual users at a particular date and time to provide suggestions either automatically on the basis of the answers gathered or manually by clicking on '*generate report*' button available inside reports section in the developed model. Clicking on '*generate report*' button inside the reports section will call the data mining Python script through PHP.

The script 'sms.php' was created to fetch stored patient's information from the database and, once the registration process is completed, it extracts one question at a time for a user and sends the question to the registered mobile number, instead of sending five questions at a time. It makes the user comfortable with the system for interacting with the application developed.

Once the user replies to question one, the system sends the next question and this process repeats until all the questions stored are finished. Twilio sends an SMS from the Twilio telephone number to the user. The Twilio number is added to the body of the SMS which is sent along with the questions to the customer on which the customer need to reply. It becomes complicated when the user provides a reply for a query that was not directed to the system; however, the reply gets forwarded to the Twilio which makes the API call to the system with the payload of the users. A separate file with the extension '.php' will force Twilio to perform an HTTP call. The incoming messages need to be configured in the Twilio account.

The set of questions framed for accessing and improving the health of underserved populations are described in the '*Questions*' tab in the menu bar of the admin module. The framed questions are displayed in Figure 5. Hence the framework contains the questions in relevance to hypertension and its control measures and diet followed by the patients. Some of the questions are:

1. How many calories did you consume today?
2. How many hours did you sleep?
3. How many steps did you walk today?

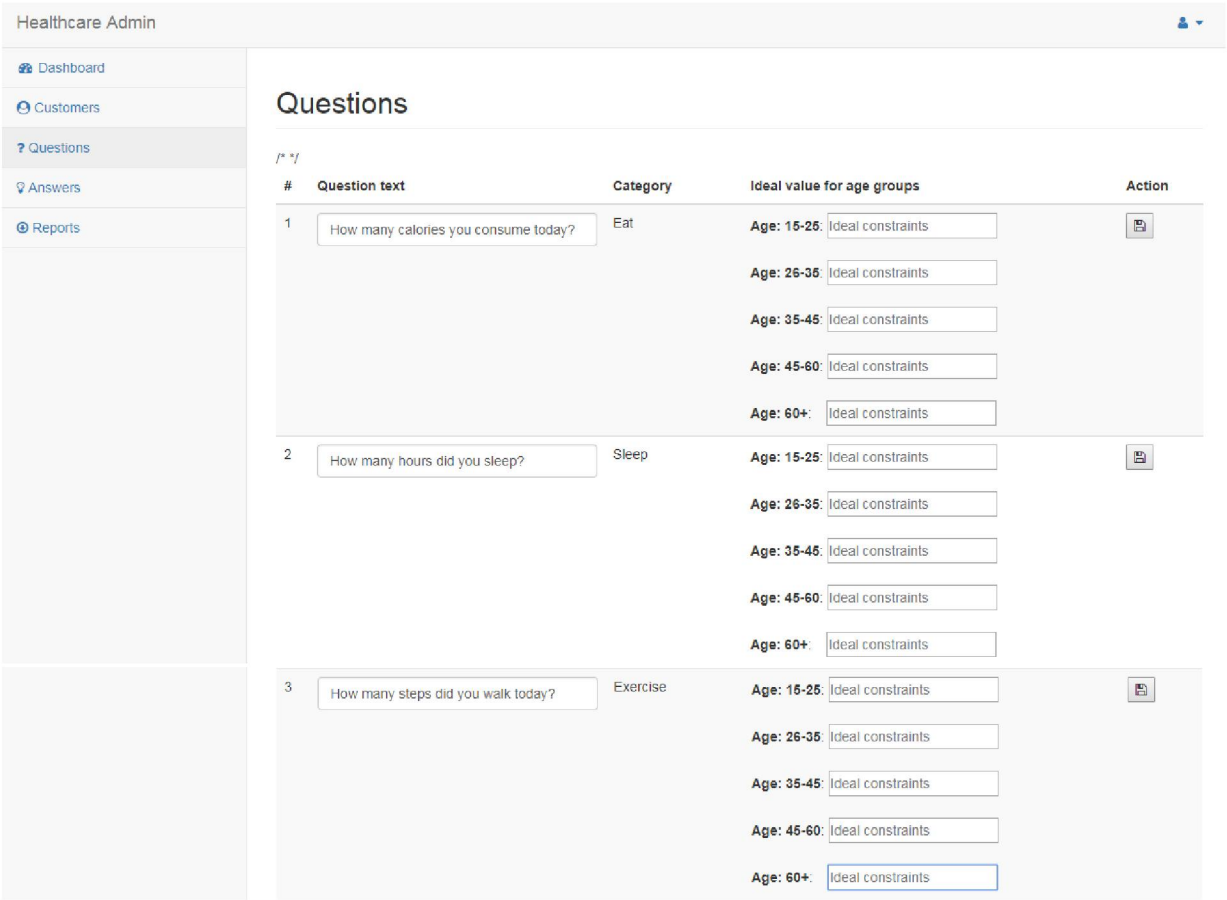


Figure 5: List of Questions

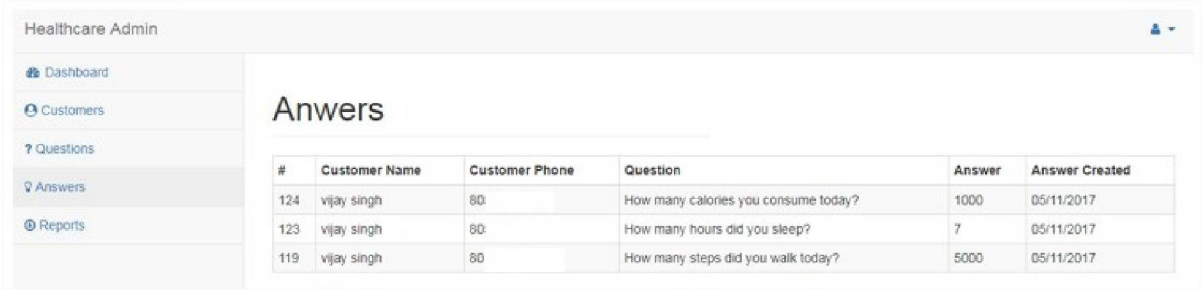


Figure 6: Sample responses from patients.

The database module shown in Figure 6 is displayed when the '*answers*' tab in the menu bar is selected. This section maintains the table for keeping the record of the answers provided by the users.

The messaging module uses 'Twilio', which is an external application created using the web hosting servers for sending and receiving the information from the user. The questions are sent to the patients and the answers are also obtained in the form of text messages through the unique number provided to the user in the text messages delivered to them. In case a user responds with invalid input, such as alphanumeric characters instead of numerals, the system uses regular expressions, a sequence of characters that define a search pattern (Regular Expression, Wikipedia, 2017) to parse and extract only proper numeric values.

Answers are stored in MySQL database tables, where the primary keys are set to the questions and the customers. As the table contains the answers that are provided by the customer for all the weeks, suggestions are created for each customer separately. Suggestions are provided every week based on the pre-trained data model generated on sample dataset using decision tree algorithm and the model can be updated with new dataset.

A python script connects to the database using MySQL client connector. Once the connection is established, the script gets the list of all patients where the cycle end date is the current system date. For every patient, the system checks for responses for each question in the past seven days and calculates the average of responses for each question, which is then fed as input to train model. The model then makes prediction on healthy behavior and send SMS to the patient with the prediction. Once the process completes, the cycle end data is updated to the current system date plus seven days.

4.3.2 Data mining

The data mining module is implemented using Python programming language. In the process of data mining, we want to train our model to make predictions based on the inputs. We train our model using Scikit-learn library integrated with python. More information about decision tree and classifier in Scikit-learn can be found on http://scikit-learn.org/stable/modules/tree.html and http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier respectively.

The Figure 7 below shows a sample code for using scikit-learn for creating train model:

```
#--------Importing the libraries ------------------------------------------------------------------

from sklearn import tree
import pandas as pd
from sklearn.cross_validation import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_squared_error
from sklearn.naive_bayes import GaussianNB
from sklearn.externals.six import StringIO
from sklearn.externals import joblib

#---------------Splitting the dataset into training and testing data ----------------------------

train_data, label_train_data, test_data, label_test_data = train_test_split (X, Y, test_size=0.33,
random_state=42)

#-------------Creating the model ------------------------------------------------------------------
clf = tree.DecisionTreeClassifier()
clf.fit(train_data,label_train_data)
clf.predict(test_data)
```

Figure 7: Sample decision tree python code

To create our decision tree model, we train our model using decision tree classifier on the

dataset, the text file we discussed in dataset section, consisting data of 1,285,200 rows and 5 columns. Columns within a dataset are commonly referred to as features of our dataset. The first 4 features in the dataset are Age (age of the patient), Sleep (number of hours the patient sleeps every day), Steps (number of steps the patient walks every day), Calories (number of calories the patient eat every day) which are the input features of the dataset and the 5th column is the status (Healthy or Unhealthy) which is the label for the dataset. In machine learning, a process can be trained to predict a label when provided a set of features as input and expected output. In this process, the dataset is read through the Panda library in python using read_csv function and then divided into features and labels.

In Figure 7, variable X are the features and Y are the labels. The features and labels are then passed through train_test_split function, a built-in function in the scikit-learn library to divide the dataset into training and testing data. In our case, we use 33% of the total dataset as the testing data and rest of the 67% of the data as training data.

Scikit-learn have a common method to process the data. Firstly, model.fit() is used to fit the training data. For applications involving supervised learning techniques, the fit function accepts two arguments: data (features) and labels. Once we trained our model, we save the model in pickle format using the joblib function. "Pickling" is the process whereby a Python object hierarchy is converted into a byte stream, and "unpickling" is the inverse operation, whereby a byte stream is converted back into an object hierarchy. It uses a powerful algorithm for serializing and de-serializing a Python object structure (https://docs.python.org/2/library/ pickle.html). The saved pickle file now has our model; we can use this pickle file as an input to predict any new set of input data using model.predict() function. This saves a lot of processing since the process will not need to train the model every time the system makes a prediction. However, the data is saved for re-configuring the classifier later on.

In Scikit-learn, Decision Tree algorithm is one of the many machine learning algorithm used to solve classification and regression problems. The algorithm is cost effective and feasible and can perform better than manual processing and programming. The library is also very easy to use. First, the dataset is loaded into the program and is divided into training and testing set. The decision tree classifier is fit on the data and the predictions are made on the training data. The predictions are made based on the prediction rules develop while creating the model. Figure 8

shows the decision tree model (with maximum depth=5), which will be used by our system to predict health status of the patient:

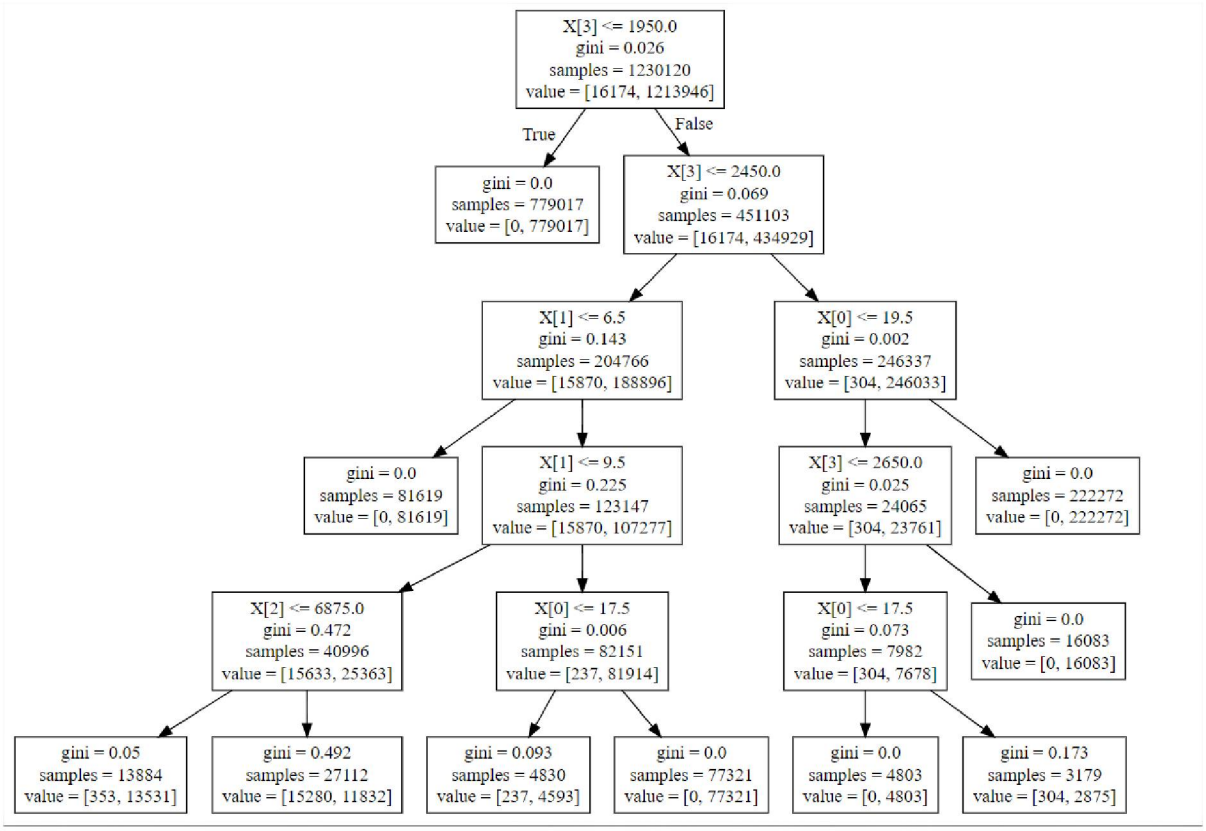


Figure 8: Decision Tree

The controlling factor of "how the Decision Tree Classifier should behave" depends on several factors, for example, Information Gain (IG). In the case above, the system analyzes each patient based on the responses received from them every week and predicts new results based on the created decision tree model.

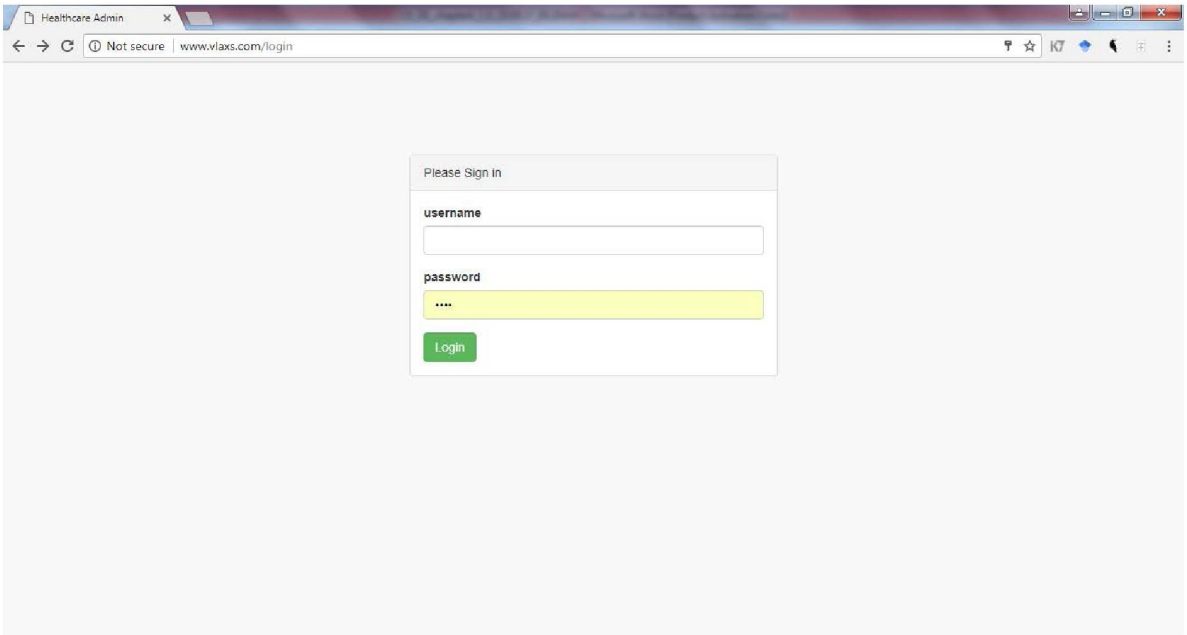
4.4 Implementation Results / Outcome:

This subsection provides the results obtained from the mHealth software designed. Figure 9 shows the screenshot of the webpage created as the administrator module. As discussed previously, the displayed webpage was created using PHP and HTML and MySQL are used as a front-end tool to create the web contents. Each block is created using the inbuilt commands available in the PHP platform that allows the programmer to edit the contents using the software languages with which they are familiar. The system does not store any private or

confidential patient information. All questions posed to patients are related to day-to-day health practices. However, this software could be integrated into a more comprehensive personal health information system.

To access patient details, a user logs into the mHealth website. For the security of the web page, 'authentication.php' is programmed and edited using the HTML. Once the login credential fields are updated with the correct username and password, by clicking on the '*login*' button the web page is opened into the administrator module. From there, a menu bar is provided for accessing site-wide users and questions and answers are provided on the left side of the page. On the right side of the top edge of the page, '*Setup*' allows for user-based profile information. This link also allows users to log in and log out of the system.

Figure 9: Login page of Administrator Module



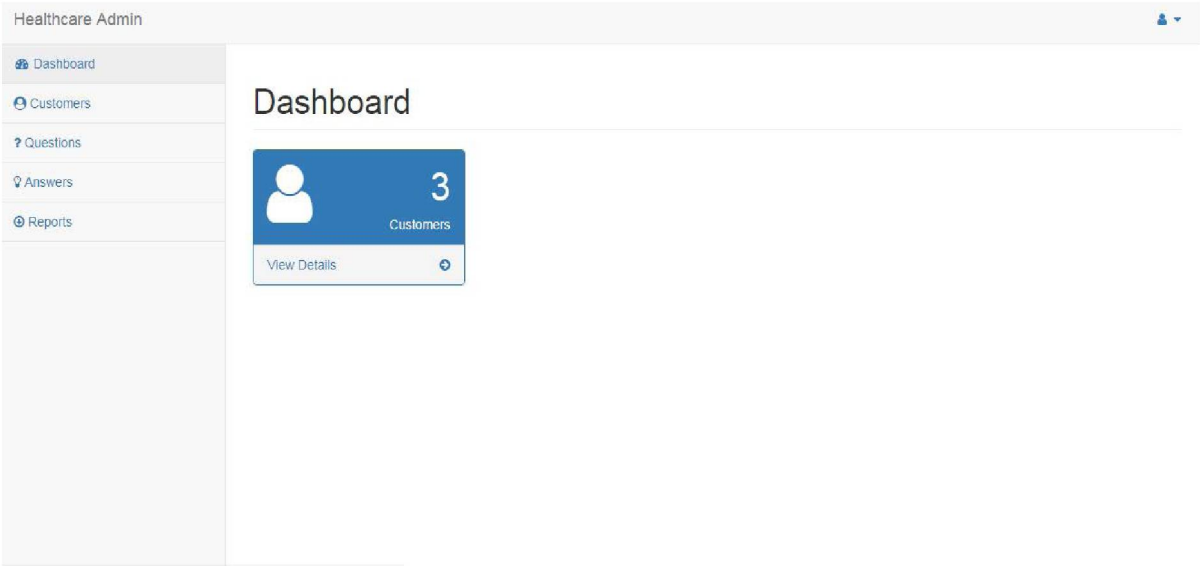


Figure 10: Screenshot of Admin Dashboard mHealth model

The Admin Dashboard is shown in Figure 10, contains the sections for dashboard, customers, questions, answers and reports. This module allows the administrator to set up the health care system. It provides the option for adding, removing the editing the user information who were registered with the mHealth system in the customers' section. The questions provided for monitoring the health care behavior of the individual is stored in the question section. Reports section allows you to see the reports that are the answers provided by the participants and the system generated suggestions which are provided to them based on their health status.

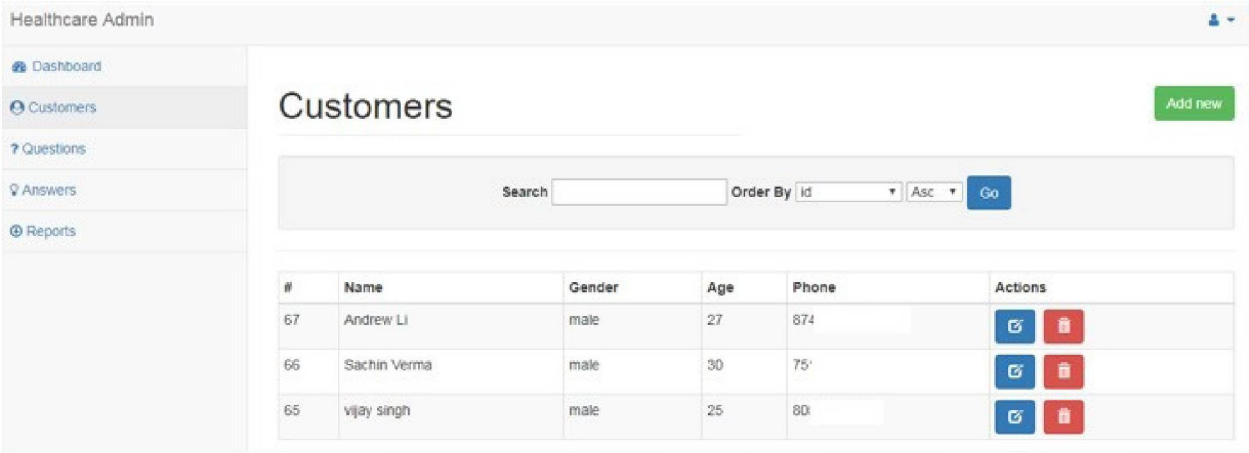


Figure 11: list of registered users

Figure 11 shows the list of the patients who are registered on the website. The menu bar is

used to check the list of patients who are enrolled. The administrator (i.e., care provider) can sort the details of the patients at his or her convenience. Additionally, administrators may select the 'go' button to edit the details such as first name, second name, gender, patient id and their mobile number, and demographic info are used to sort or arrange the list of persons who are receiving the healthcare through mobile.

The 'Add new' button enables administrators to add new users to a module. Selecting 'Add new' will initiate a pop-up window where administrators can input the demographic information of the user and add them to the healthcare seeker list.

In Figures 12 through 18, different views of database design are provided, and Figure 19 depicts how the patient may respond to questions via their mobile device.

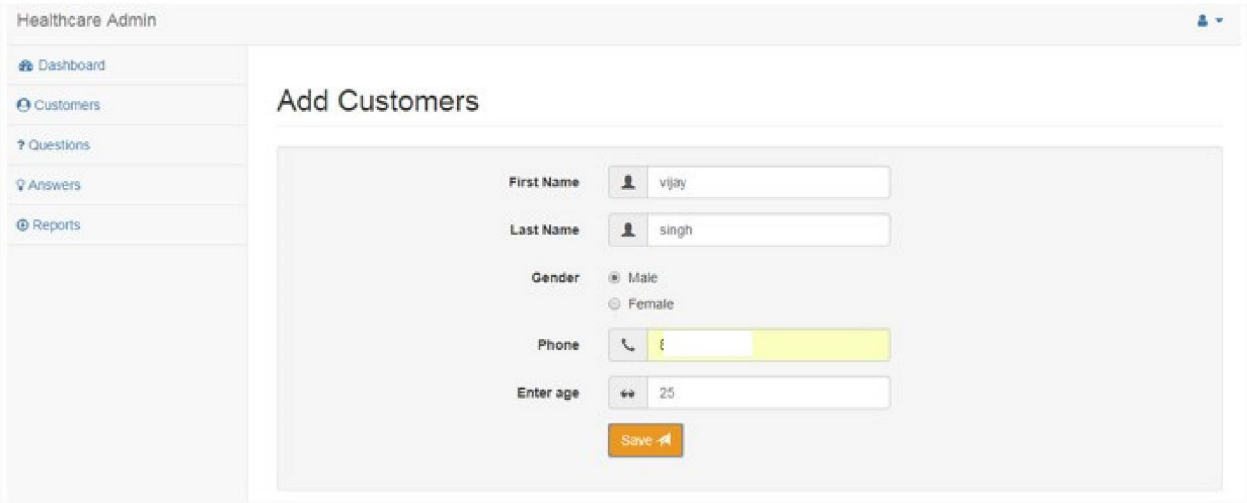


Figure 12: Add new patient/customer

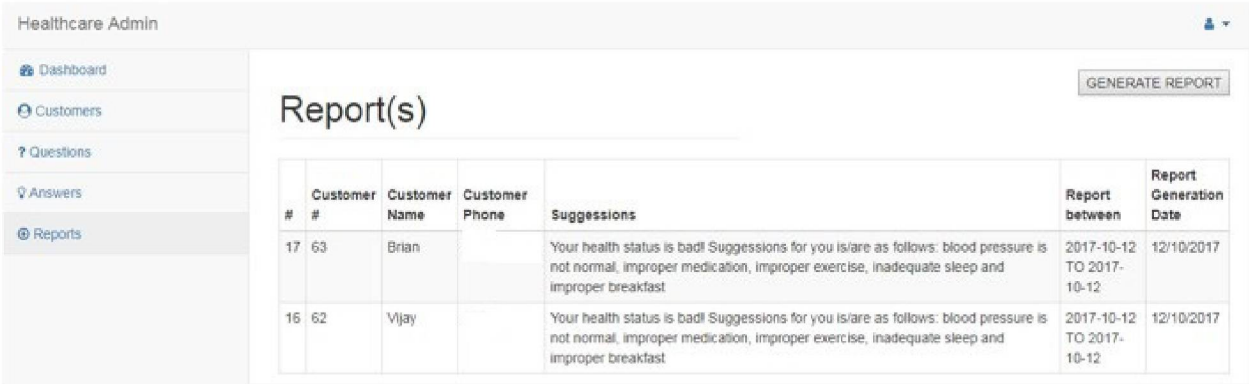


Figure 13: Generated health report

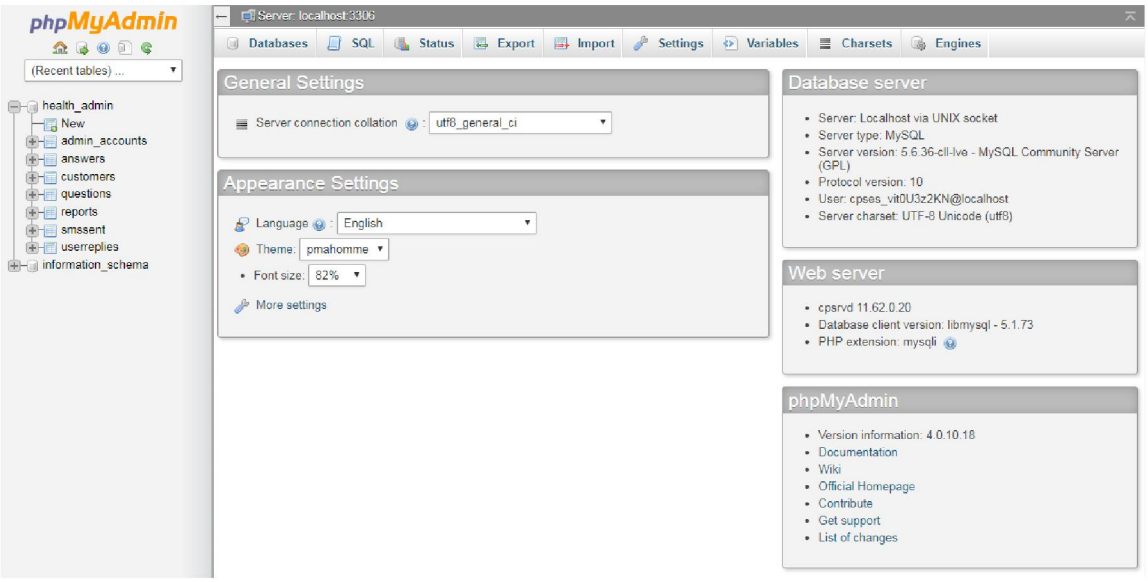


Figure 14: Database view - list of tables

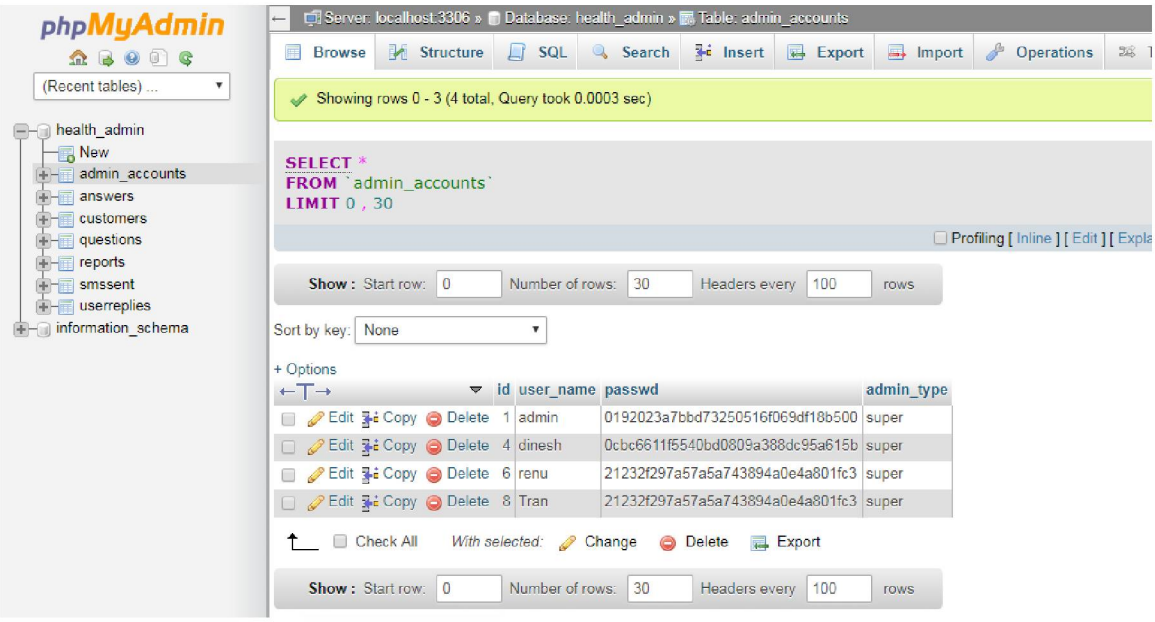


Figure 15: Database view - list of admin accounts

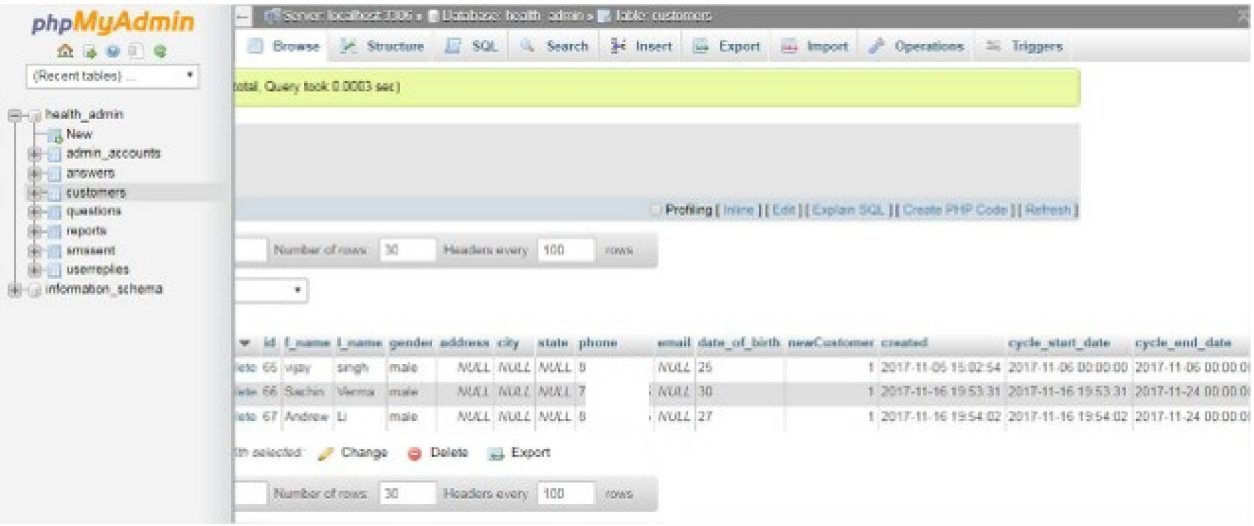


Figure 16: Database view - list of Users / Patients

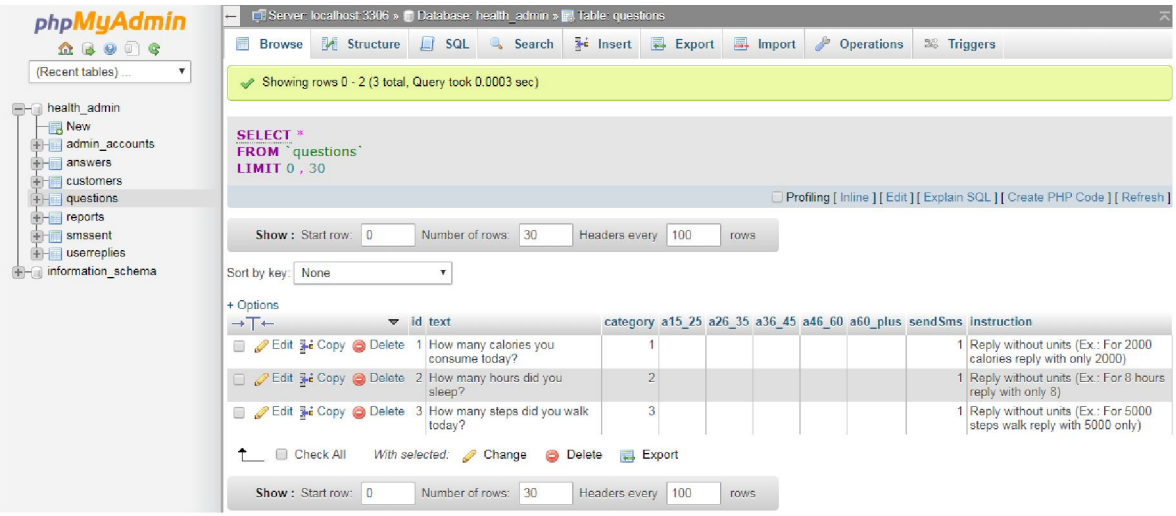


Figure 17: Database view- list of Questions

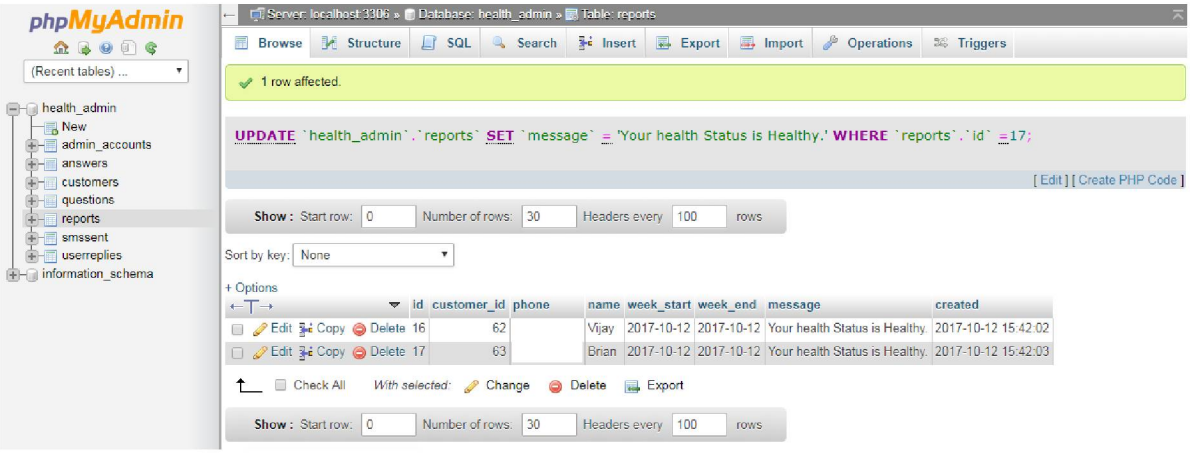


Figure 18: Database view - list of records showing health status

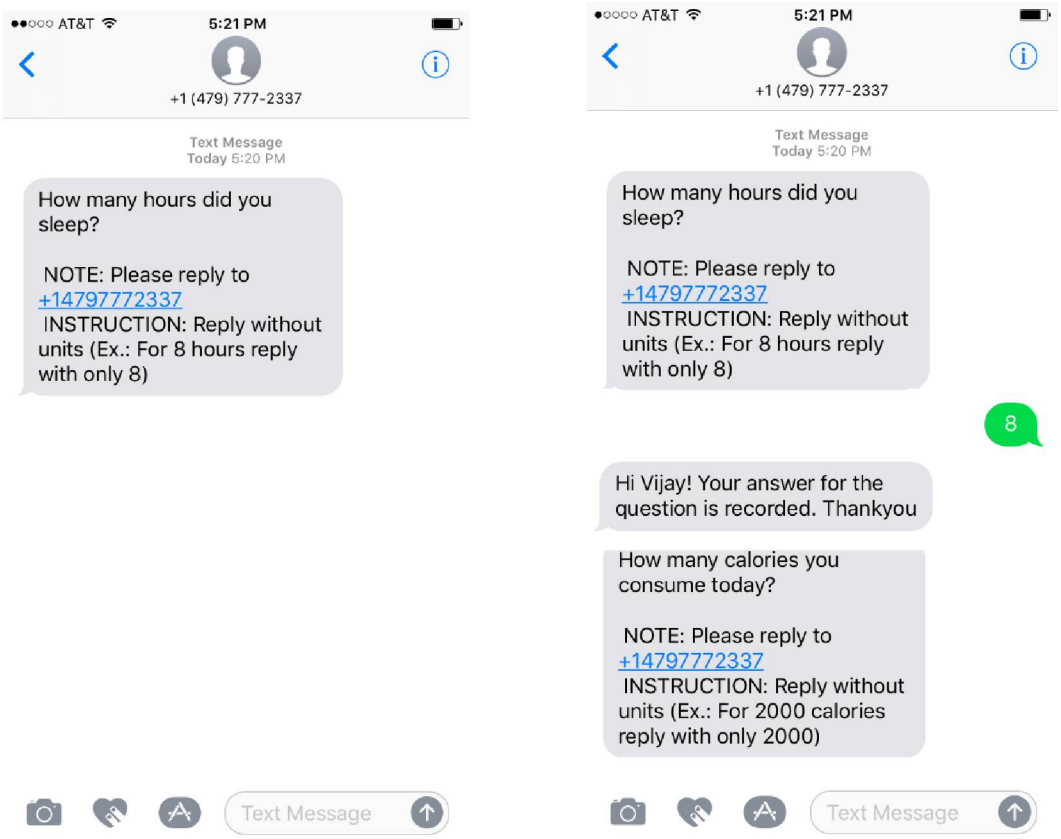


Figure 19: Phone view with question and user response

4.5 Summary

The primary goal of this project was to architect a system for engaging underserved populations in managing and monitoring their health behaviors using the simple messaging system of mobile devices. This chapter we deal with the implementation and creating the web application for the mHealth system. The displayed webpage is created using PHP and MySQL as backend services, while HTML, JavaScript and CSS are used as front-end tools for represented patient-driven data. Python was used to construct a data mining module using scikit-learn to make the prediction on patient's data collected over a period of past seven days.

CHAPTER V: SIMULATION TESTING

5.1 Introduction

Simulation/Use case testing is described as a small study that is performed before the execution of the preplanned project. This can be done to examine the proposed research criteria and its outcomes. Simulation testing allows for identifying adjustments that are necessary before a final commitment to the design must be made. A simulation was preferred over pilot testing since the goal is to show a proof-of-concept design. A pilot test would be the next logical step in this research to investigate how real-life users interact with the proposed design.

As a simulation, this study has its disadvantages. For one, it is not a hypothesis testing study in which safety, efficacy and effectiveness of the system are evaluated. Secondly, feasibility results cannot generalize beyond the inclusion and exclusion criteria of the simulated data. The simulation will, however, provide a proof-of-concept and provide guidelines for real-world user testing in the future.

5.2 Population

The goal of this system is to focus on underserved populations and those individuals without access to proper health services or cannot afford to partake in medical treatments. In our simulation test, we register participants by generating randomized data based on real types of underserved users. It is important to stress that the data supplied is an example of non-real, sample data. As such, we provide a proof-of-concept use case example, which can be applied later for underserved populations to manage their health.

5.2.1 Demographic Relationships and Study Variables

The generated participants were first registered on the healthcare application. The demographic data consisted of, age, gender and their mobile number of the participants. The study is conducted among the selected population with a questionnaire specifically related to hypertension.

The data to be collected aims to combat and/or minimize adverse effects of:

(1) Symptoms related to a chronic illness;

(2) Day-to-day pain and discomfort related to chronic illness;

(3) Complications related to the treatment of a chronic illness; and

(4) High healthcare costs are related to a chronic illness.

For each category, a set of three questions represents the functional activities of eating, sleeping and exercising performed by the respondents. Respondents should answer as many questions as possible, for the decision tree to make a better more informed decision.

User data will be collected and stored in the MySQL consisting of three sections for recovering and utilizing the data later. The first section stores demographic data such as age, sex, gender, and mobile number. A second section stores data that describes the answers provided by the respondents in correlation to the questionnaire. The third section stores result in the form of health status predictions/suggestions, which are generated by the system and includes answers provided by respondents. An analysis of the demographic data and generated suggestions were examined, and the association between the two terms was discussed with the aim of improving the health of the individuals.

5.3 Analysis and Results

Hypertension is a chronic health condition, which is a common problem all over the world. Studies have shown a correlation between patients having hypertension with their age, sleeping habits, physical activity and eating habits. With the increase in age comes an increase in the chances of developing hypertension. Additionally, people having daily sleep duration of ≤ 5 hours, 6 hours, 8 hours and ≥ 9 hours have a higher risk of hypertension, compared to people with 7 hours of daily sleep. The study also reveals that physical inactivity increases the risk of hypertension (Relationship between Duration of Sleep and Hypertension in Adults: A Meta-Analysis, 2015). Different studies have shown that the increase in body weight can cause an increase in blood pressure. Obesity is also a risk factor for hypertension. Therefore, it is necessary to promote a healthy dietary plan and develop a good lifestyle to prevent the development of hypertension to reduce the adverse effects related to its complications. More so, individuals who consume an appropriate number of daily calories and maintain an appropriate body weight can significantly

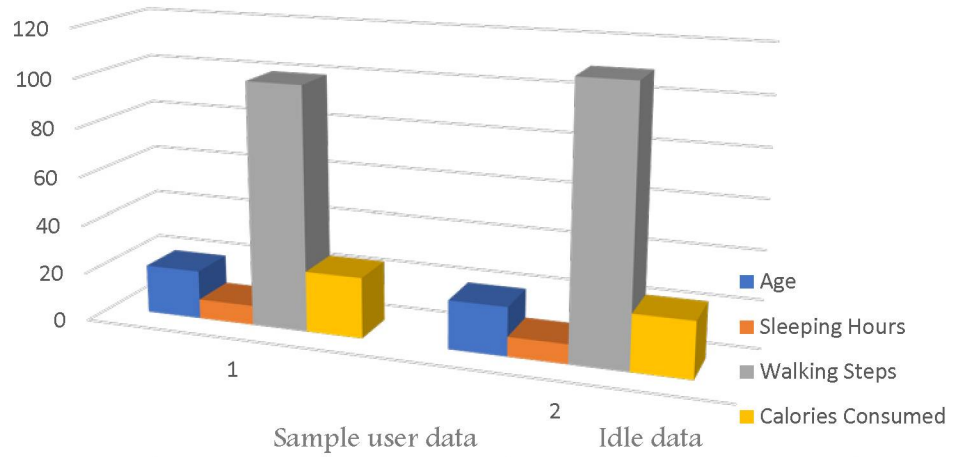
control their risk of developing hypertension (Analysis for hypertension and related risk factors of physical examination population, 2013).

To test and perform analysis, we use six different types of users. Of these users, three users have hypertension, while the other three users do not have hypertension. An analysis of the user input was based on age, the number of calories consumed daily, sleeping hours and walking minutes. The health status of a user is predicted by training the classification model using the decision tree classifier from the scikit-learn library. This process of predicting the user health is fast, easy and cost-effective. More so, it is not necessary to explicitly program the application for each input because manually programming an application would require huge lines of code and might still not able to properly code for every user input. For these reasons, a machine learning approach was applied. Machine learning requires only a few lines of code to predict user health outcomes based on the pre-trained classification model. By passing a subset of test data, the accuracy of the trained model, which predicted whether or not a user was healthy or not, was 98.3%. Table 2 shows that sample user data used for analysis.

S.No.	Age	Sleeping Hours (per day)	Walking Steps (per day)	Calories Consumed (per day)	Hypertension
User-1	20	8	10000	2500	No
User-2	22	9	1000	5500	No
User-3	35	5.5	1500	4500	Yes
User-4	40	7	9000	2500	No
User-5	65	7.5	10000	4000	Yes
User-6	70	6	2000	1500	Yes

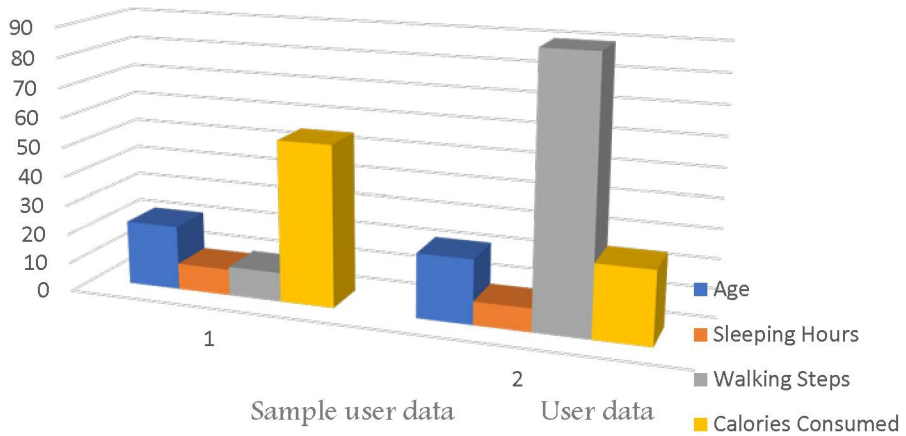
Table 2: User data for system analysis

Referring to the data in Table 2, it can be concluded that older aged users have a significantly higher risk of developing hypertension. Figures 20 through 24 below show the results of sample user data from Table 2 versus idle data from the dataset in Table 1, which is representative of a healthy person.



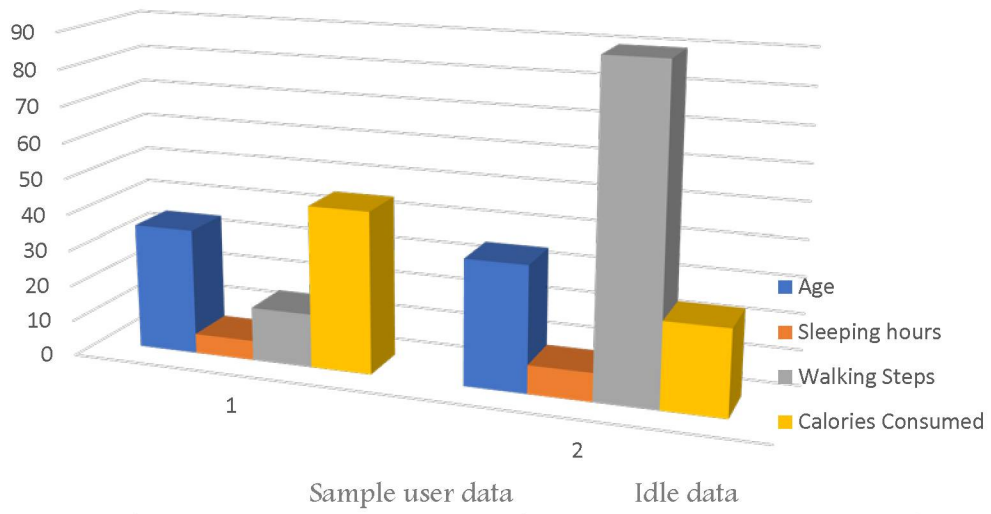
	1	2
Age	20	20
Sleeping Hours	8	8
Walking Steps	100	110
Calories Consumed	25	23

Figure 20: User-1 without hypertension



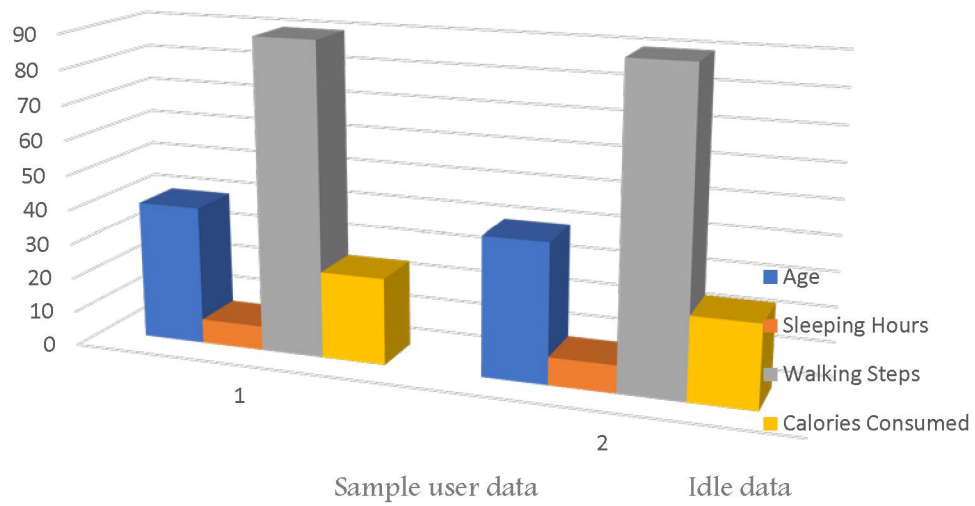
	1	2
Age	22	22
Sleeping Hours	9	8
Walking Steps	10	90
Calories Consumed	55	25

Figure 21: User-2 without hypertension



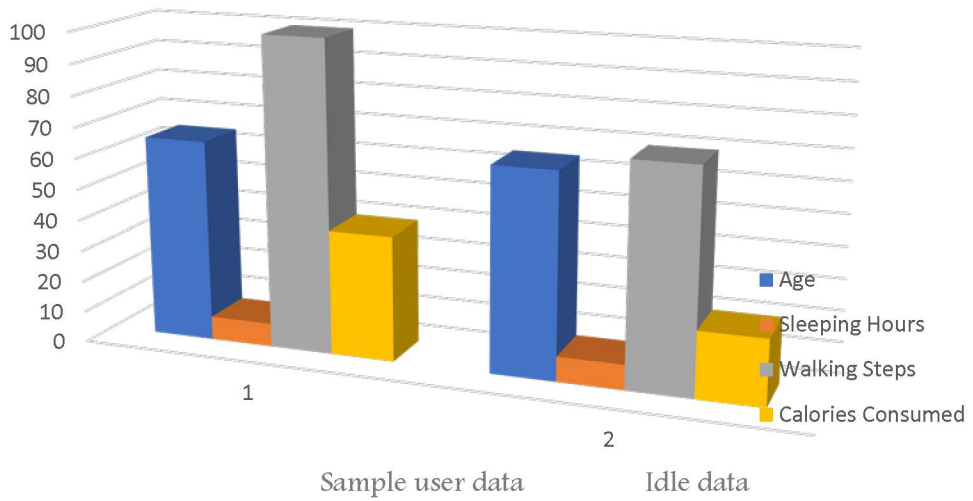
	1	2
Age	35	35
Sleeping hours	5.5	8
Walking Steps	15	90
Calories Consumed	45	24

Figure 22: User-3 with hypertension



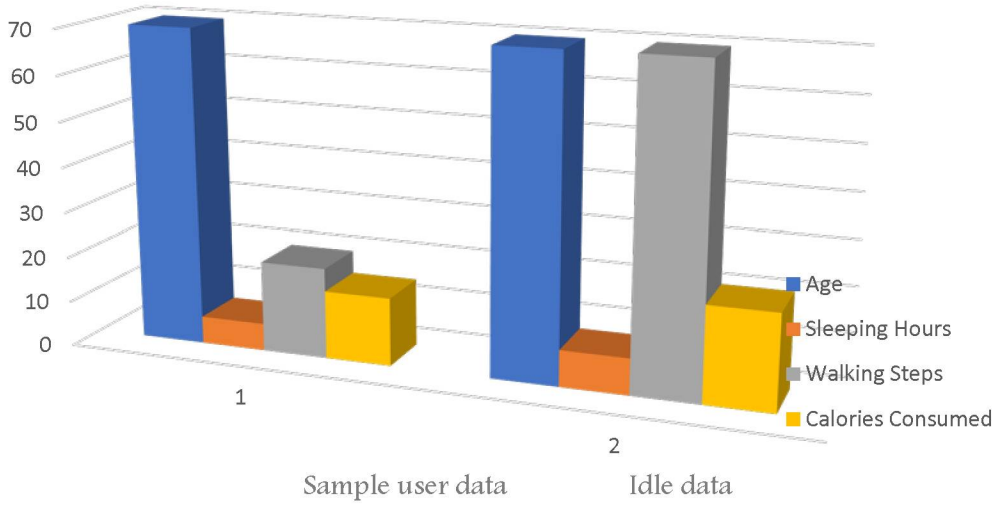
	1	2
Age	40	40
Sleeping Hours	7	8
Walking Steps	90	90
Calories Consumed	25	24

Figure 23: User-4 without hypertension



	1	2
Age	65	65
Sleeping Hours	7.5	8
Walking Steps	100	70
Calories Consumed	40	21

Figure 24: User-5 with hypertension



	1	2
Age	70	70
Sleeping Hours	6	8
Walking Steps	20	70
Calories Consumed	15	21

Figure 25: User-6 with hypertension

Figure 26 shows the system output for user-predictions based on the user input from Table 2. Here, it can be concluded that User-1 and User-4 have status "Healthy" and exhibit healthy behaviors compared to Users 2, 3, 5 and 6. These healthy behaviors are determined using the decision tree classifier based on the underlying data for diet, physical activity and sleep. For these types of users, the system will send a text message to keep them focused on healthy behaviors with the goal of reducing complications of hypertension. For User-2, the health status is unhealthy, but the user does not have hypertension (age=22). User-2 would typically be too young to have the disease, but continuing this kind of health behavior every day will result in higher risk of having hypertension at an older age. For these types of users, the system will send text messages of their health status as being "unhealthy" and suggest that the user perform a certain amount of physical activity and follow a healthy diet with the proper caloric intake. User-3, User-5 and User-6 already have the disease and the system classifies their health status as being "unhealthy". For these types of users, the system will suggest that users maintain proper sleep patterns, enough physical activity and follow a healthy diet with an appropriate number of calories intake. These users can use this system to keep track of their daily health habits, reduce the risk of hypertension disease and live a more healthy life. Health providers can use this mHealth system for their patients who are already suffering from hypertension or have systems that predispose them to the disease. Hypertension was chosen to show how the system can be used by users/patients, but the system can also be extended to other diseases.

```
--------------------------Health status prediction------------------------
Age =20  Hours of Sleep =8  Number of Walking Steps =10000  Number of Calories Consumed =2500  Health Status = :['Healthy']

--------------------------Health status prediction------------------------
Age =22  Hours of Sleep =9  Number of Walking Steps =1000  Number of Calories Consumed =5500  Health Status = :['Unhealthy']

--------------------------Health status prediction------------------------
Age =35  Hours of Sleep =5.5  Number of Walking Steps =1500  Number of Calories Consumed =4500  Health Status = :['Unhealthy']

--------------------------Health status prediction------------------------
Age =40  Hours of Sleep =7  Number of Walking Steps =9000  Number of Calories Consumed =2500  Health Status = :['Healthy']

--------------------------Health status prediction------------------------
Age =65  Hours of Sleep =7.5  Number of Walking Steps =10000  Number of Calories Consumed =4000  Health Status = :['Unhealthy'

--------------------------Health status prediction------------------------
Age =70  Hours of Sleep =6  Number of Walking Steps =2000  Number of Calories Consumed =1500  Health Status = :['Unhealthy']
>>>
```

Figure 26: User health prediction by proposed mHealth system

5.4 Summary

This section provided a detailed overview of simulated user test cases. The proposed system supports underserved populations living in rural and urban areas and aims to provide a low-cost means to manage the healthcare of this population via a mobile application. Moreover, the system specifically targets patients suffering from the chronic disease hypertension and advises patients towards improving their health status in better managing this disease. The primary interface between patient and system is a PHP script, which fetches stored customer details from a MySQL database and matches those users against a set of predefined questions. These questions are sent to patients via SMS on a daily basis, who will respond using the text-messaging feature of their mobile device. The system utilizes this data to track their health behaviors and determines their health status through a decision tree classification. This decision tree classifier makes suggestions on how patients might improve their health.

CHAPTER VI: CONCLUSIONS AND FUTURE WORK

6.1 Introduction

In this section, the main findings and general conclusions of the study are summarized and described. Moreover, the strengths and limitations of this thesis are taken into account. The suggestions for future research into healthcare domain are also presented. At the end of the chapter, recommendations for patients and system administrators and healthcare providers are also provided.

6.2 mHealth Summary

According to GSM Association (GSMA, 2015), there are more mobile phones than people on earth. In recent years, mHealth has become a platform for accessing and storing patient-driven data and processing that data into useful information. With a growth rate of 39% in 2016, it is no surprise that by 2018, the mHealth market is expected to increase in value to $11.8 billion. This is due to the realization that mHealth can provide value-added services to both health care providers as well as consumers. mHealth has proven to be an effective tool in enhancing the health services delivery in both decision support and information collection and management (Mosa et al.2012).

The mHealth system proposed in this thesis provides a system that is simple to use and is easily accessible by the targeted set of users; mainly underserved populations living in rural and urban areas. The underlying architecture developed in this thesis provides a '*server interaction module*', which synchronizes communication of underserved populations via the Twilio SMS API. This component was deemed critical for two reasons. First, SMS is a low-cost technology, which can be subsidized by health care providers and, second, the technology is easy to use and familiar to this user population who may lack technological literacy of more complex mobile applications. As such, this research assumes that users within underserved communities have mobile phones that support sending and receiving text messages.

The targeted population in this study was underserved populations, but a specific underserved population that suffers from hypertension. Hypertension is a chronic illness related to high blood pressure and is a long-term medical condition in which the blood pressure in the arteries is persistently elevated. The long-term effects of hypertension can place a patient at risk for

coronary artery disease, stroke, heart failure, vision loss, and kidney disease.

Recent research by Bobrow *et al.* (2016) focused on providing the suggestions to the people over the heart diseases through the Android mobile application. The study developed the application that is controlled by the administrator or the health care provider which provides suggestions to the users from the routine life habits that are closely associated with hypertension. These factors are considered to be effective in the existing models (Sosa et al. (2017), Ni et al. (2014)) for making a decision in the perfect manner.

Work by Bobrow *et al.* and others failed to address the difficulties faced by the underserved people. Targeting underserved populations was important because this group is more likely to not receive treatment due to remote access financial hardships. The proposed system was developed that targets specific healthy behaviors for patients with hypertension living in underserved communities. More specifically, a set of questions was constructed for managing and monitoring daily behaviors of these patients. These questions were sent via SMS daily and patients, in turn, could respond to these questions using their mobile device. These responses are aggregated and fed into a decision tree classifier, which can determine the health status of a patient. These results are then used to provide comments and suggestions from the healthcare provider and sent back to the patient via the mHealth application.

6.3 Conclusions and Future Work

Patient empowerment is a critical concept in healthcare and attempts to bring the patient into the forefront of their health services. mHealth is a critical new platform in the era of ubiquitous healthcare. mHealth has the scope of widening the potential market of healthcare services to many consumers, from motivating healthy consumers to helping consumers battling chronic illnesses. This research analyzed simulated user experiences of a proposed SMS-based mHealth application. The novel application is designed to improve the health conditions of patients suffering from hypertension by assessing their daily health behaviors. The system should be particularly interesting for populations living in remote areas where access to broadband Internet is not available. Instead, the system uses Twilio for sending SMS and is flexible with any mobile software and interactive with the patients.

This application offers a proof-of-concept mHealth application that utilizes SMS as the

primary mechanism for interacting with underserved populations. However, while the application offers many advantages, recommendations and future work are largely needed. Firstly, the system was not tested on a real population of users but rather was only run through simulated users. The next logical step would be to implement this system in some controlled experimental setting and test the impact of the system on a real set of users living with hypertension. Additionally, the system will need the input of experts such as medical doctors and healthcare providers to identify specific questions that can better support the healthy behaviors of populations living with hypertension. Finally, as a proof-of-concept, the system must also be stress-tested for its ability to scale to large amounts of data and system users. The simulations conducted only focused on a small subset of simulated users and did not test against large numbers of users accessing the system concurrently.

CHAPTER VII: REFERENCES

1. Adebayo, I.P. (2017). Predictive Model for the Classification of Hypertension Risk Using Decision Trees Algorithm. *American Journal of Mathematical and Computer Modelling*. [Online]. 2 (2). pp. 48–59. Available from: https://www.researchgate.net/profile/Peter_Idowu2/publication/314284545_Predictive_Model_for_the_Classification_of_Hypertension_Risk_Using_Decision_Trees_Algorithm/links/5988fcba6fdcc496162cc02/Predictive-Model-for-the-Classification-of-Hypertension-Risk.
2. Aljumah, A.A. (2011). Predictive Analysis on Hypertension Treatment using Data Mining Approach in Saudi Arabia. *Intelligent Information Management*. [Online]. 03 (06). pp. 252–261. Available from: http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/iim.2011.36031.
3. Alkabba, A., Hussein, G.A., Albar, A., Bahnassy, A. & Qadi, M. (2012). The major medical ethical challenges facing the public and healthcare providers in Saudi Arabia. *Journal of Family and Community Medicine*. [Online]. 19 (1). pp. 1. Available from: http://www.jfcmonline.com/text.asp?2012/19/1/1/94003.
4. Altaf, W., Shahbaz, M. & Guergachi, A. (2017). Applications of association rule mining in health informatics: a survey. *Artificial Intelligence Review*. [Online]. 47 (3). pp. 313–340. Available from: http://link.springer.com/10.1007/s10462-016-9483-9.
5. Ben-Zeev, D., Schueller, S.M., Begale, M., Duffecy, J., Kane, J.M. & Mohr, D.C. (2015). Strategies for mHealth Research: Lessons from 3 Mobile Intervention Studies. *Administration and Policy in Mental Health and Mental Health Services Research*. [Online]. 42 (2). pp. 157–167. Available from: http://link.springer.com/10.1007/s10488-014-0556-2.
6. Best, A.L., Vamos, C., Choi, S.K., Thompson, E.L., Daley, E. & Friedman, D.B. (2017). Increasing Routine Cancer Screening Among Underserved Populations Through Effective Communication Strategies: Application of a Health Literacy Framework. *Journal of Cancer Education*. [Online]. 32 (2). pp. 213–217. Available from: http://link.springer.com/10.1007/s13187-017-1194-7.
7. Bobrow, K., Brennan, T., Springer, D., Levitt, N.S., Rayner, B., Namane, M., Yu, L.-M., Tarassenko, L. & Farmer, A. (2014). Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt

trial) randomised controlled trial. *BMC Public Health*. [Online]. 14 (1). pp. 28. Available from: http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-28.

8. Bobrow, K., Farmer, A.J., Springer, D., Shanyinde, M., Yu, L.-M., Brennan, T., Rayner, B., Namane, M., Steyn, K., Tarassenko, L. & Levitt, N. (2016). Mobile Phone Text Messages to Support Treatment Adherence in Adults With High Blood Pressure (StAR): A Single-Blind, Randomized Trial. *Circulation*. [Online]. pp. 1–11. Available from: http://circ.ahajournals.org/lookup/doi/10.1161/CIRCULATIONAHA.115.017530.
9. Botts, N., Thoms, B., Noamani, A. & Horan, T.A. (2010). Cloud Computing Architectures for the Underserved: Public Health Cyberinfrastructures through a Network of HealthATMs. In: *2010 43rd Hawaii International Conference on System Sciences*. [Online]. 2010, IEEE, pp. 1–10. Available from: http://ieeexplore.ieee.org/document/5428374/.
10. Boukenze, B., Mousannif, H. & Haqiq, A. (2016). Predictive Analytics in Healthcare System Using Data Mining Techniques. In: *Computer Science & Information Technology (CS & IT)*. [Online]. 24 April 2016, Academy & Industry Research Collaboration Center (AIRCC), pp. 01–09. Available from: http://www.airccj.org/CSCP/vol6/csit65201.pdf.
11. Burns, M.N., Begale, M., Duffecy, J., Gergle, D., Karr, C.J., Giangrande, E. & Mohr, D.C. (2011). Harnessing Context Sensing to Develop a Mobile Intervention for Depression. *Journal of Medical Internet Research*. [Online]. 13 (3). pp. e55. Available from: http://www.jmir.org/2011/3/e55/.
12. Chaurasia, V. & Pal, S. (2013). Data Mining Approach to Detect Heart Dieses. *International Journal of Advanced Computer Science and Information Technology*. [Online]. 2 (4). pp. 56–66. Available from: https://poseidon01.ssrn.com/delivery.php?ID=724017066103123090127025027099088022026080077013030029112113031026093025106083091073011056033056027005107124092123030069071011046072056061077090125123081105121097089028058087083109115094090024082083094009101021 0.
13. CTIA (2017). *Background on CTIA's Wireless Industry Survey*. [Online]. 2017. CTIA. Available from: https://www.ctia.org/docs/default-source/default-document-library/ctia_survey_ye_2014_graphics.pdf?sfvrsn=2. [Accessed: 17 July 2017].
14. Finitsis, D.J., Pellowski, J.A. & Johnson, B.T. (2017). *Text Message Intervention Designs to Promote Adherence to Antiretroviral Therapy (ART): A Meta-Analysis of Randomized*

Controlled Trials. [Online]. 2017. PLoS. Available from: http://journals.plos.org.sci-hub.cc/plosone/article?id=10.1371/journal.pone.0088166#. [Accessed: 17 July 2017].

15. Fiordelli, M., Diviani, N. & Schul, P.J. (2013). *Mapping mHealth Research: A Decade of Evolution*. [Online]. Lugano. Available from: http://www.jmir.org/2013/5/e95/.
16. Hall, A.K., Cole-Lewis, H. & Bernhardt, J.M. (2015). Mobile Text Messaging for Health: A Systematic Review of Reviews. *Annual Review of Public Health*. [Online]. 36 (1). pp. 393–415. Available from: http://www.annualreviews.org/doi/10.1146/annurev-publhealth-031914-122855.
17. Hall, C.S., Fottrell, E., Wilkinson, S. & Byass, P. (2014). Assessing the impact of mHealth interventions in low- and middle-income countries: what has been shown to work. *Global Health Action*. [Online]. 7 (1). pp. 25606. Available from: https://www.tandfonline.com/doi/full/10.3402/gha.v7.25606.
18. Heron, K.E. & Smyth, J.M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology*. [Online]. 15 (Pt 1). pp. 1–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19646331.
19. Hooker, R.S. (2013). Working with the medically underserved. *Canadian Family Physician*. [Online]. 59 (4). pp. 339–340. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3625069/.
20. ITU (2016). *ICT Facts and Figures 2016*. [Online]. 2016. ITU. Available from: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx. [Accessed: 17 July 2017].
21. Jiang, J., Zhuangzhi Yan, Jun Shi, Kandachar, P. & Freudenthal, A. (2010). A Mobile Monitoring System of Blood Pressure for Underserved in China by Information and Communication Technology Service. *IEEE Transactions on Information Technology in Biomedicine*. [Online]. 14 (3). pp. 748–757. Available from: http://ieeexplore.ieee.org/document/5443609/.
22. Kaplan, R.M. & Stone, A.A. (2013). Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention. *Annual review of psychology*. [Online]. 64. pp. 471–98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22994919.

23. Kaufman, J.S. & Cooper, R.S. (1999). Seeking causal explanations in social epidemiology. *American journal of epidemiology*. [Online]. 150 (2). pp. 113–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10412955.
24. Kazi, A.M., Carmichael, J.-L., Hapanna, G.W., Wangoo, P.G., Karanja, S., Wanyama, D., Muhula, S.O., Kyomuhangi, L.B., Loolpapit, M., Wangalwa, G.B., Kinagwi, K. & Lester, R.T. (2017). Assessing Mobile Phone Access and Perceptions for Texting-Based mHealth Interventions Among Expectant Mothers and Child Caregivers in Remote Regions of Northern Kenya: A Survey-Based Descriptive Study. *JMIR Public Health and Surveillance*. [Online]. 3 (1). pp. e5. Available from: http://publichealth.jmir.org/2017/1/e5/.
25. Khamis, H.S., Cheruiyot, K.W. & Kimani, S. (2014). Application of k- Nearest Neighbour Classification in Medical Data Mining. *International Journal of Information and Communication Technology Research*. [Online]. 4 (4). pp. 121–128. Available from: https://www.researchgate.net/profile/Hassan_Shee/publication/270163293_Application_of_k-Nearest_Neighbour_Classification_in_Medical_Data_Mining/links/54a249fd0cf256bf8baf7fff.pdf.
26. Kitsiou, S., Thomas, M., Marai, G.E., Maglaveras, N., Kondos, G., Arena, R. & Gerber, B. (2017). Development of an innovative mHealth platform for remote physical activity monitoring and health coaching of cardiac rehabilitation patients. In: *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. [Online]. 2017, IEEE, pp. 133–136. Available from: http://ieeexplore.ieee.org/document/7897223/.
27. Narwal, S. & Mintwal, K. (2013). Comparison the Various Clustering and Classification Algorithms of WEKA Tools. *International Journal of Advanced Research in Computer Science and Software Engineering*. [Online]. 3 (12). pp. 866–878. Available from: https://www.ijarcsse.com/docs/papers/Volume_3/12_December2013/V3I12-0140.pdf.
28. Ni, Z., Wu, B., Samples, C. & Shaw, R.J. (2014). Mobile technology for health care in rural China. *International Journal of Nursing Sciences*. [Online]. 1 (3). pp. 323–324. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2352013214000726.
29. Nishantha, G.G.D., Uchida, Y., Cassim, M., Weerasinghe, J., Gunawardana, G., Hayashida, Y. & Goto, M. (2009). Towards a sustainable e-health deployment - An integrated medical information system for Sri Lankan case. In: *Proceedings of the 3d International ICST*

Conference on Pervasive Computing Technologies for Healthcare. [Online]. 2009, ICST. Available from: http://eudl.eu/doi/10.4108/ICST.PERVASIVEHEALTH2009.6011.

30. Omboni, S., Caserini, M. & Coronetti, C. (2016). Telemedicine and M-Health in Hypertension Management: Technologies, Applications and Clinical Evidence. *High Blood Pressure & Cardiovascular Prevention*. [Online]. 23 (3). pp. 187–196. Available from: http://link.springer.com/10.1007/s40292-016-0143-6.
31. Pappachan, P., Yus, R., Joshi, A. & Finin, T. (2014). Rafiki: A Semantic and Collaborative Approach to Community Health-Care in Underserved Areas. In: *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*. [Online]. 2014, pp. 1–10. Available from: http://ebiquity.umbc.edu/_file_directory_/papers/724.pdf.
32. Parameswari, R. & Prabakaran, N. (2012). An Enhanced Mobile HealthCare Monitoring System in Mobile Cloud Computing. *International Journal of Advanced Research in Computer and Communication Engineering*. [Online]. 1 (10). pp. 804 –807. Available from: http://ijarcce.com/upload/december/12-An Enhanced Mobile.pdf.
33. Person, A.K., Blain, M.L.M., Jiang, H., Rasmussen, P.W. & Stout, J.E. (2011). Text Messaging for Enhancement of Testing and Treatment for Tuberculosis, Human Immunodeficiency Virus, and Syphilis: A Survey of Attitudes Toward Cellular Phones and Healthcare. *Telemedicine and e-Health*. [Online]. 17 (3). pp. 189–195. Available from: http://www.liebertonline.com/doi/abs/10.1089/tmj.2010.0164.
34. Prasad, V., Rao, T.S. & Babu, M.S.P. (2016). Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms. *Soft Computing*. [Online]. 20 (3). pp. 1179–1189. Available from: http://link.springer.com/10.1007/s00500-014-1581-5.
35. Quintana, Y. & McWhirter, J. (2017). Culturally Appropriate Behavioral Change in Maternal Health: Role of Mobile and Online Technologies Over Time. In: V.L. Patel et al (ed.). *Cognitive Informatics in Health and Biomedicine*. [Online]. New Dehi: Springer International Publishing AG, pp. 217–233. Available from: http://link.springer.com/10.1007/978-3-319-51732-2_10.
36. Rehman, H., Kamal, A.K., Morris, P.B., Sayani, S., Merchant, A.T. & Virani, S.S. (2017). Mobile Health (mHealth) Technology for the Management of Hypertension and

Hyperlipidemia: Slow Start but Loads of Potential. *Current Atherosclerosis Reports*. [Online]. 19 (3). pp. 12. Available from: http://link.springer.com/10.1007/s11883-017-0649-y.

37. Rubinstein, A., Miranda, J.J., Beratarrechea, A., Diez-Canseco, F., Kanter, R., Gutierrez, L., Bernab?-Ortiz, A., Irazola, V., Fernandez, A., Letona, P., Mart?nez, H. & Ramirez-Zea, M. (2016). Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. *The Lancet Diabetes & Endocrinology*. [Online]. 4 (1). pp. 52–63. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2213858715003812.
38. Sosa, A., Heineman, N., Thomas, K., Tang, K., Feinstein, M., Martin, M.Y., Sumer, B. & Schwartz, D.L. (2017). Improving patient health engagement with mobile texting: A pilot study in the head and neck postoperative setting. *Head & Neck*. [Online]. 39 (5). pp. 988–995. Available from: http://doi.wiley.com/10.1002/hed.24718.
39. Statista (2017). *Smartphone user penetration as percentage of total global population from 2014 to 2020*. [Online]. 2017. Statista. Available from: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/. [Accessed: 17 July 2017].
40. Strasser, R. (2003). Rural health around the world: challenges and solutions*. *Family Practice*. [Online]. 20 (4). pp. 457–463. Available from: https://academic.oup.com/fampra/article-lookup/doi/10.1093/fampra/cmg422.
41. Thakkar, J., Kurup, R. & Laba, T.-L. (2016). for Medication Adherence in Chronic Disease A Meta-analysis. *JAMA Internal Medicine*. [Online]. 176 (3). pp. 340–349. Available from: http://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2484905.
42. Thakkar, J., Kurup, R., Laba, T.-L., Santo, K., Thiagalingam, A., Rodgers, A., Woodward, M., Redfern, J. & Chow, C.K. (2017). *Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease A Meta-analysis*. [Online]. 2017. HUB. Available from: http://sci-hub.cc/10.1001/jamainternmed.2015.7667. [Accessed: 17 July 2017].
43. TPCHD (2013). *Planning Process Tools: Information for Local Planners*. [Online]. 2013. TPCHD. Available from: https://www.tpchd.org/files/library/d30555beb1557e78.pdf. [Accessed: 17 July 2017].
44. Twilio (2017). *Twilio Announces Fourth Quarter And Full Year 2016 Results*. [Online]. 2017. Twilio. Available from: `. [Accessed: 20 June 2017].

45. VanderWielen, L.M., Vanderbilt, A.A., Crossman, S.H., Mayer, S.D., Enurah, A.S., Gordon, S.S. & Bradner, M.K. (2015). Health disparities and underserved populations: a potential solution, medical school partnerships with free clinics to improve curriculum. *Medical Education Online*. [Online]. 20 (1). pp. 27535–27539. Available from: https://www.tandfonline.com/doi/full/10.3402/meo.v20.27535.
46. Vega, W.A., Rodriguez, M.A. & Gruskin, E. (2009). Health Disparities in the Latino Population. *Epidemiologic Reviews*. [Online]. 31 (1). pp. 99–112. Available from: https://academic.oup.com/epirev/article-lookup/doi/10.1093/epirev/mxp008.
47. Zriqat, I.A., Zriqat, I.A. & Azzeh, M. (2017). *A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods*. [Online]. New York. Available from: https://arxiv.org/abs/1704.02799.
48. Medically Underserved Areas and Populations (MUA/Ps).(2016, October). Retrieved from *https://bhw.hrsa.gov/shortage-designation/muap.*
49. Estimated Calorie Needs per Day, by Age, Sex, and Physical Activity Level. (n.d.). Retrieved from *https://health.gov/dietaryguidelines/2015/guidelines/appendix-2/.*
50. National Sleep Foundation's sleep time duration recommendations: methodology and results summary (2015, March). Retrieved from *http://www.sleephealthjournal.org/article/ S2352-7218(15)00015-7/fulltext*
51. How many steps/day are enough? For older adults and special populations (July, 2011). Retreieved from *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169444/*
52. Machine learning (2017, November). Retreived from *https://en.wikipedia.org/wiki/Machine_learning*
53. Regular Expression (2017, December). Retrieved from https://en.wikipedia.org/wiki/Regular_expression
54. Wang, Y., Mei, H., Jiang,Y., Sun, W., Song, Y., Liu, S. & Jiang,F. (n.d., 2015).Relationship between Duration of Sleep and Hypertension in Adults: A Meta-Analysis. *Journal of Clinical Sleep Medicine*.
55. HHS (December, 2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *National Institutes of Health National Heart, Lung, and Blood Institute.* Retrieved from *https://www.nhlbi.nih.gov/files/docs/guidelines/express.pdf*

56. Qiao, S., Ye, Q., Dou, Y., Li, M., Kou, Y., Qian, D., Li, M., and Wang, G. (September 2013). Analysis for hypertension and related risk factors of physical examination population. PubMed Central. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3798214/.