A DATA DRIVEN, HOSPITAL QUALITY OF CARE PORTAL FOR THE PATIENT

COMMUNITY

by

SREEHARI BALAKRISHNA HEGDEN

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2016

Acknowledgements

Abstract

A DATA DRIVEN, HOSPITAL QUALITY OF CARE PORTAL FOR THE PATIENT

COMMUNITY

Sreehari B. Hegden, MS

The University of Texas at Arlington, 2016

Supervising Professor: Dimitrios Zikos

With the recent changes in health services provision, patients are members of a
consumer driven health care system. However, the health care consumers are not
presented with adequate opportunities to enhance their position in choosing high quality
hospital services. As a result, the demand for active patient participation in the choice of
quality and safe hospital services remained unaddressed. In this research work, we
developed MediQoC (Medicare Quality of Care), a data driven web portal for Medicare
patients, their caregivers and the health care insurance policy designers to grant access
to data-driven information about hospitals, and quality of care indicators. The portal which
utilizes the Medicare claims dataset enables the patients, caregivers and other
stakeholders the ability to locate high-quality hospital services for specific diseases and
medical procedures. MediQoC provides the users a list of eligible hospitals, and output
statistics on hospital stay attributes and quality of care indicators, including the
prevalence of hospital acquired conditions. It gives options for the users to rank hospitals
on the basis of the aforementioned in-hospital attributes and quality indicators. The
statistical module of the portal models the correlation between length of stay (LOS) and
discharge status attributes in each hospital for the given disease. Finally, the ranking
results are visualized as bar charts via MediQoC-viz, the visualization module of the

portal. The visualization module also makes use of Google Geocoding API to locate in map the nearest hospital to user's location. It also displays the location, distance and driving duration to the hospitals selected by the user from the ranked result list.

Table of Contents

List of Illustrations

List of Tables

Chapter 1

Introduction

Project Overview

With the recent changes in health services provision, patients are members of a consumer driven health care system. This thesis titled the 'A Data Driven, Hospital Quality of Care Portal for The Patient Community' is developed with the objective of creating a data driven web portal for Medicare patients and their caregivers as well as the health care insurance policy designers to grant access to data-driven information about hospitals, and quality of care indicators. The MediQoC (Medical Quality of care) portal we developed as part of this thesis utilizes the Medicare claims dataset of 2013 and enables the patients, caregivers and other stakeholders the ability to locate high-quality hospital services for specific diseases and medical procedures. MediQoC provides the users a list of eligible hospitals, and output statistics on hospital stay attributes and quality of care indicators, including the prevalence of hospital acquired conditions. It gives options for the users to rank hospitals based on the carefully selected performance indicators and the results are visualized as bar charts and also on Google maps considering the user's location as well.

The MediQoC web portal enables users to rank hospitals based on performance indicators like number of admissions in each hospital for each disease, the average length of stay (LOS) for each of these diseases in each hospital, the average daily and total charges in each hospital for each disease and the appropriateness indicator which is the ratio of number of cases under investigation to the total number of admissions in each hospital in percentage. Users can select any of the above performance indicators to have a grouping of hospitals under 5 classes as 5 stars, 4 stars, 3 stars, 2 stars and 1 star. The 5 stars hospitals will be the most suitable for the patient for the performance indicator

10

selected, the 4 stars the next and so on. The portal also enables users to filter these ranking results and get a list of hospitals with only 5 stars, 4 stars or more, only 4 stars and so on. One of the attribute in the result list table is the correlation between the length of stay and discharge status for the disease in each hospitals. Users can filter the results to see only the hospitals where higher the length of stay higher the chance of discharging alive or the hospitals where higher the length of stay higher the chance of discharging dead or the hospitals where a correlation could not be established based on the data available. Users can get above results either for a given Primary Diagnosis Code or a given Admission Code, the significance of which will be explained in further sections In the result list, the hospital mortality ratio for the given admission or primary diagnosis code is also listed for each hospital. Mortality ratio is the percentage of patients discharged as dead. Users can filter the results for mortality ratio below 26%, mortality ratio between 26% and 51%, between 51% and 76% and above 76%. Another information added in the list is the 'Subpart or Not' information. Each of the hospital is either a stand-alone institution or a subsidiary under such a parent or stand-alone institution. Users may prefer to go to a parent hospital which will have facilities for treatment of different diseases than to a subpart institution where there are only fewer facilities and infrastructure. MediQoC gives option to filter out parent or subpart institutions from the result list. Finally, for those looking for a particular hospital statistics for a given disease, MediQoC gives option to filter the hospital from result list by providing the National Provider Identifier (NPI). A National Provider Identifier (NPI) is a unique ten-digit identifier for the health care providers in the United States. It is assigned by the Centers for Medicare and Medicaid Services (CMS) [23]. In addition, the web page gives some information such as code type and description of the admission or primary

11

diagnosis code given by the user as input. This information is fetched from HIPAASpace ICD-9 Code Lookup website [28] on real-time.

Also, by clicking each hospital in this result list, users will be navigated to a different page where the selected hospital's statistics will be displayed. This includes ranking the different primary diagnosis codes or admission codes treated in this hospital based on number of cases reported, average length of stay, average daily and total charges. In addition, it displays the mortality ratio for each admission or primary diagnosis codes in this hospital as well as the appropriateness of these admission or primary diagnosis codes in this hospital. Star rating in the result table helps to identify the best treated diseases in the selected hospital. Also the selected hospitals location by zipcode, distance and driving duration to this hospital from user's current location are displayed in Google map with the help of Google Geocoding API [31]. Like in the diagnosis code statistics web page, here we show the NPI information such as the name of the hospital, mailing address, phone, practice location address and practice in real-time from HIPAASpace NPI Number Lookup website [29].

The MediQoC portal in addition to ranking based on the performance indicators also gives important statistical analysis based information on the percentage of hospital acquired cases of Pneumonia, Septicemia and Urinary Tract Infection. The hospitals can be ranked based on the sorted order of percentage of acquiring these diseases in the hospitals. This information in addition to helping patients to select a more reliable hospital for their diagnosis code also provides an opportunity for the hospital themselves to have a self-analysis on why their performance is bad or why the chances of getting a disease from these hospitals is more. This will indirectly improve the quality of available services extended by hospitals to patients which itself is an objective of developing this data-driven web portal.

Finally, the MediQoC portal also provides option to get ranked list of hospitals based on the Procedure Code which is the code for the various tests a patient has to take in course of his/ her treatment in the hospital. The results will easily identify some of the tests that results in higher expenditure for patients during the treatment in hospitals.

The web portal will also access users' location and shows list of hospitals nearest to them based on the zip code and clicking the hospital from the list will give detailed statistics of the hospital including the hospital specialty which will guide users to hospitals most suitable for them. As the nearest hospitals are shown in Google maps with distance and driving duration information, users will be able to find a suitable hospital in their locality itself instead of going out of the state or even farther.

While the results of MediQoC are historical data driven, it cannot assure 100% reliability. Things might have changed as time flies and constant update of data used is required to ensure continuous reliable services from the portal. MediQoC web portal at this stage does not provide personalized support. At this point MediQoC is a system which provides the same set of results to all of its users. In such a system, the hospital suitable for a particular patient need not be the best one for another patient with similar symptoms. MediQoC as it is now, considers only some of the selected performance indicators from the historical data and this does not mean the result will be the same if we considered other indicators. It is very difficult to identify and include all factors that affect the quality of care and this may lead to reporting false positive and false negative performance reports.

Problem Definition and Analysis

With the recent trends in the way that health care is delivered and seamless access to knowledge resources and online information, patients has enormous

13

opportunity to manage their own care. New technologies like high speed internet makes it possible and legislative changes facilitate it. Today patients can freely browse the internet for medical journals and use libraries to access health care information which were previously available only to purchase for professionals. Patients are also able to buy diagnostic kits and over-the-counter equivalent drugs online and get the results of tests online [18]. Patients being the end-users in this customer driven model drive the health care industry and can receive and transmit health-related data in real-time. It is the Affordable Care Act (ACA) and health insurance coverage changes that brought this shift. There is greater access to, and demand for health information via smartphone and patient portals as a result of these changes [19].

Patient groups and medical specialty societies pursue clinical registries as a clinical improvement strategy. There is an increased national focus on quality and cost and thus there is a high opportunity to leverage clinical registries to improve outcomes and appropriate utilization [17]. An increasing awareness on the quality of health care delivery is contributing towards patients, caregivers and patient advocates being more actively involved to the choice of quality health services. It is evident that medical errors threaten the quality of health care, increase health care costs, and cause thousands of deaths in U.S. hospitals each year [1]. The latter, being the eighth leading cause of death in the U.S. It is striking that almost half Americans believe that they had personally experienced a medical mistake [2].

Electronic medical records (EMRs) though are the only health information technology (HIT) application to have a clear and statistically significant effect on patient safety as compared to nurse charts, and picture archiving and communications systems (PACS), no evidence has been found to state existing HIT improved quality. EMRs are associated with reduced infection rates, but the effect is small [8]. Digital transformation

14

of health care through HIT can reduce costs and improve quality (Institute of Medicine 2001). One of the expected goals of HIT is improvements in health care quality. Though studies of existing HIT, most of which are specific, custom developed information systems (IS) at leading institutions, identified their positive impact on quality, including lower mortality rates and patient safety, broader assessments of these systems revealed not much quality gains or even negative effects associated with HIT adoption and implementation. The existence of abundant datasets in health care for researchers has created a unique opportunity for utilizing statistical approaches such as data mining for discovering more innovative ways to measure performance impacts of HIT than are currently available. With greater digitization of clinical data, more accurate measurements of quality than the norm today can be achieved [9].

In a consumer driven health care system which received a boost recently in the U.S [3], health care consumers have not yet foreseen an important opportunity to enhance their position in choosing high quality hospital services, minimizing the chances of malpractices and medical errors. In general, elements that health care consumers consider essential include their involvement in decisions, effective treatment, fast access to reliable health advice and provision of clear comprehensible information [20]. That said, the current high levels of low health literacy must be addressed for such decision making to be truly informed and informative. The ability to access and understand health information and the lack of ability to use information portals and difficulty understand mathematical and epidemiologic concepts such as probability and risk [21] are barriers that need to be addressed.

Some of the most important hospital qualities of care indicators include the mortality ratio due to various complications, medical conditions and clinical procedures, prevalence and mortality ratio of Hospital Acquired Conditions (HAC) [4], and the patient

15

readmission ratio. A key factor to improve quality of care is patient safety, which is considered the cornerstone of high-quality health care [5]. The Institute of Medicine defines quality of health care as the "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes". [6] Therefore a series of quality indicators including, but not limited to, death, disease, disability, discomfort, and dissatisfaction, achievement of self-care, and perception of being well cared for, are also used as positive components for quality of care [7].

Considering this existing scenario where there is more opportunity for statistical measurement of quality of care and there is an unfulfilled demand for active patient participation in the choice of quality and safe hospital services, we developed MediQoC (Medicare Quality of Care), a data driven web portal for Medicare patients, their caregivers and the health care insurance policy designers that provides information about hospital attributes and quality indicators as well as access to a hospital ranking system for the aforementioned parameters. The output may either be based on health related symptoms, or a known condition and medical procedure that a person may need to undergo.

Medicare Claims Data

Medicare is a United States health insurance program for Americans of age 65 or above, as well as for Americans under age of 65 but with certain disabilities and Americans of all ages having permanent kidney failure. There are more than 49 million patients who are beneficiaries of the Medicare in the United States for the year 2015 and Medicare is the primary payer for the 47.2 percent of total in-patient hospital costs in the United Sates.

The MediQoC portal uses mainly two datasets. The main dataset is the Medicare in-hospital claims file which contains records for the 100% Medicare beneficiaries who used hospital in-patient services during the year 2013. At present, the MediQoC uses only the records in this dataset for those who used hospital in-patient services only in Texas, United States. The second dataset used is the National Provider Identifier (NPI) dataset with comprehensive information on each of the registered NPI to date. This dataset includes information such as NPI, its registered address, its specialty information as Taxonomy codes, if it is a subpart or not, zip code etc.

ICD-9

ICD stands for the International Classification of Diseases, which is the widely used short-form name for the International Statistical Classification of Diseases and Related Health Problems. ICD is the international "standard diagnostic tool for epidemiology, health management and clinical purposes". [30] It is maintained by the World Health Organization (WHO). ICD basically provides a system of diagnostic codes for classifying diseases as well as various symptoms and other findings, complaints or causes. ICD-9 is the Ninth Revision of the ICD. The ICD-9-CM which stands for the "International Classification of Diseases, 9th Revision, Clinical Modification" is an adaption created by the National Center for Health Statistics (NCHS) and maintained jointly by the NCHS and the Centers for Medicare & Medicaid Services (CMS) in the United States. This is updated every year on October 1. Usually this contains two or three volumes, in which the first two are the diagnosis codes whereas the volume three contains the procedure codes for the various diagnostics tests. For example, in ICD-9 code '486' corresponds to Pneumonia and 3893 stands for the procedure code: Venous catheterization, not elsewhere classified.

Targeted users

Hospital rankings are not new and death rates, patient safety and hospital reputation are few of the many factors considered by most of such systems. In this study we developed dedicated, evidence based, data driven online portal where the users can choose selected performance indicators and rank the hospitals based on them. The web application makes the interpretation of results easier with the help of visualization based on graphs or column charts. The targeted users of this application would be mainly below three categories:

1. General public (the patients)
2. Hospitals or doctors or caregivers
3. Administrators or sales persons in health insurance sector

General public (the patients)

The general public or the patients specifically, can use the system to select the hospital best suitable for their treatment. The system will help them in decision making regarding which hospital to be selected based on their own priorities of quality of care. Thus MediQoC ensures active participation of patient community in making decisions on selecting the high-quality hospital services which are appropriate for specific symptoms and medical procedures. Patients are given the choice of selecting the quality parameter to rank the result list. Finally, the location services showing the distance and driving duration to the hospitals leaves the option to patients to confirm their decision of selection.

Hospitals or doctors or caregivers

Hospitals or doctors or caregivers can also use this system to monitor the performance of the hospitals. Mortality ratio information gives opportunity for hospitals to do a self-check on their quality of care and provide them an indication of improving the infrastructure or treatment facilities. Hospital acquired percentage of diseases or symptoms also help hospitals to identify sources of in hospital infections and eliminate them. The statistical correlation between length of stay and discharge status equips hospitals to identify the right amount of treatment required in diseases having a positive correlation (higher length of stay implies higher chances of discharging alive), whereas helps to identify sources of increasing complications for diseases with negative correlation (higher length of stay implies higher chances of discharging dead).

Administrators or sales persons in health insurance sector

Administrators in health insurance sectors can use this system to classify hospitals based on the selected performance indicators. The results will equip them with enough statistics to provide an insight on the performance of hospitals and thus the quality of care provided by the hospitals. They can also get to see the factors like mortality ratio and hospital acquired percentages of each disease or symptoms and these results will help them in formulating insurance policies and decisions. Hospitals unnecessarily doing overtreatment can be identified with ranking based on length of stay attribute. Ranking by cost parameter will list the most expensive hospitals and proper investigation of these hospitals can help detect fraud or waste of money and resources if any.

Chapter 2

Background and Related Work

This chapter gives details of some of the prior works and systems in the health care information technology sector aiming at improving the quality of care, hospital rankings and use of Medicare claims data. There is lots of past and ongoing research in health care information technology sector, and hospital rankings which makes use of the clinical registries or Medicare claims datasets.

Existing Systems

According to the Medicare Payment Advisory Commission's 'Report to the Congress New Approaches in Medicare' of June 2004 [10], health information technology (HIT) is often associated with following technologies and terms. 'Electronic health record' (EHR) is an automated order-entry and patient-tracking system providing real-time access to patient data. 'Computerized provider order entry' (CPOE) is a medication ordering and fulfillment system. 'Clinical decision support system' (CDSS) provides physicians and nurses with real-time diagnostic and treatment recommendations. 'Picture archiving and communications system' (PACS) captures and integrates diagnostic and radiological images to store them to a medical record. Bar coding is used to matching drugs to patients by using bar codes on both the medications and patients' arm. 'Radio frequency identification' (RFID) tracks patients throughout the hospital, and links lab and medication tracking through a wireless communications system. 'Automated dispensing machines' (ADMs) distribute medication doses. 'Electronic materials management' (EMM) is used to track and manage inventory of medical supplies, pharmaceuticals, and other materials. 'Interoperability' refers to electronic communication among organizations so that the data in one IT system can be incorporated into another. Though the adoption

of HIT applications is aimed at improving quality of patient care, further research is needed to establish the link between HIT and quality. Quality health care relies on ability of caregivers, patients and those with the right information at the right time to make the right decisions.

There is hospital rankings conducted by various organizations year to year. However, for example in case of renal transplant, hospital rankings by US News and World Report may not correlate with actual outcomes after renal transplant. There were more than half of the hospitals not ranked in the US New and World Report Top 50 but had higher than expected patient and graft survival rate [12]. Clinical registries' role in health care will have an increasing importance with a number already established in cardiac surgery [11]. One of the ranking of hospitals by the quality of care for medical conditions [13] work used the 1991 and 1992 MedisGroups National Comparative Data Bases and the ranking methodology was based on the expected number of deaths in a given hospital. This was computed by estimating the probability of death for each patient at a given hospital using their logistic regression model for death and then summing the probabilities of death for all patients in the hospital.

For research or quality improvement efforts in relation to high rates of abnormal screening mammography, Medicare claims data might be feasible data source [14]. An algorithm was developed to identify women with incident breast cancer using claims data from 1995 breast cancer patients from the SEER-Medicare database and claims data from 1995 Medicare control subjects and the algorithm was validated on 1994 claims from breast cancer subjects and controls [15]. An attempt to rank hospitals on surgical mortality used the national Medicare data from 2003 to 2006 for three surgical procedures: coronary artery bypass grafting (CABG), abdominal aortic aneurysm (AAA)

repair, and pancreatic resection by examining the implications of reliability adjustment on hospital mortality with surgery [16].

Large number of research publications on HIT, hospital ranking and use of Medicare claims data exists. However, the review of all those publications is beyond the scope of this thesis. In addition to the above mentioned works, there are online portals publicly available to all to perform hospital comparisons. One such commonly used system is the 'Hospital Compare' website by "Medicare.gov" [26].

The Hospital Compare website by Medicare.gov compares up to 3 hospitals from a list of hospitals available at a given location. The site accepts a zip code and list hospitals within 25 miles of the given zip code and allows user to select up to 3 from the list to compare. The selected hospitals are compared based on general information (hospital type, availability of emergency services etc.), survey of patients' experiences, timely and effective care information (how fast the outpatients were taken to tests), rate of complications, rate of readmissions and deaths, use of medical imaging and payment and value of care. But the system is not tailored for different use cases. In other words, the parameters to compare cannot be selected by users and whenever the users select the hospitals to compare they are going to get the same results every time. The static information about hospitals for the parameters is put side by side and users can figure out which one has better values for the factors.

However, users cannot rank hospitals for a given disease or symptoms by a quality of care factor like length of stay or total cost or mortality ratio available in a list. Also, there is no option for relative ranking or a 5-star ranking system for the above factors in which users can select any of the above performance indicators to have a grouping of hospitals under 5 classes as 5 stars, 4 stars, 3 stars, 2 stars and 1 star. This brings the significance of having a data driven quality of care portal for patients where

they can get a ranked list of hospitals based on a selected quality of care parameter. MediQoC is a step towards this direction.

Proposed System

This study developed an online portal available to all where the users can select the carefully selected performance indicators from a drop down menu and get a ranked list of all hospitals based on that for a given disease or symptoms or procedure code. Users can also select any of the performance indicators to have a grouping of hospitals under 5 classes as 5 stars, 4 stars, 3 stars, 2 stars and 1 star. The 5 stars hospitals will be the most suitable for the patient for the performance indicator selected, the 4 stars the next and so on. The performance indicators include:

1. Average length of stay

2. Average total charges

3. Appropriateness in percentage

4. Mortality ratio in percentage

5. Correlation between length of stay and discharge status

6. Percentage of hospital acquired conditions

7. Count of admissions

8. Average daily charges

Average length of stay

Average length of stay is the average of length of stay in all records for the given hospital and disease.

Average total charges

Average total charges is the average of total charges in all records for the given hospital and disease.

Appropriateness in percentage

Appropriateness indicator is the ratio of number of cases under investigation to the total number of admissions in each hospital for the given disease represented in percentage.

Mortality ratio in percentage

Mortality ratio is the ratio of patients discharged as dead to the total number of patients for the given disease in each hospital represented in percentage.

Correlation between length of stay and discharge status

Relationship between LOS and discharge status for a given disease in each hospital. Often this has no effect to the patient safety, but this is not always the case. For this reason we explored possible correlation between LOS and discharge status per NPI for each disease. A positive correlation implies higher the LOS, the higher the discharged alive rate. A negative correlation implies higher the LOS, the lower the discharged alive rate.

Percentage of hospital acquired conditions

Using the diagnosis code 1 to 25 and present on admission diagnosis indicator 1 to 25, we find whether the given disease is hospital acquired or not. If a disease is diagnosed the corresponding ICD-9 code is listed under any of the diagnosis code 1 to

25 in the Medicare claims data and if we look the corresponding present on admission diagnosis indicator field and the value is 1, it means the symptom or disease was present on admission and is not hospital acquired. If the value is 0, it means the disease was not present on admission and hence it is in-hospital acquired. For example, if 'DIAGNOSIS_CODE_15' is 486 (ICD-9 code for Pneumonia) and if the 'PRESENT_ON_ ADMISSION_DIAGNOSIS_INDICATOR_15' is 0, then Pneumonia is in-hospital acquired for that particular case. We take the count of in-hospital acquired cases for the given diagnosis code in each NPI and find its ratio to the total number of cases in which the diagnosis code is reported in any of the 1 to 25 diagnosis code fields for the particular NPI. Now this value is expressed in percentage and used for ranking. In this thesis, we found the percentage of hospital acquired conditions only for three diseases as an experimental model:

1. Pneumonia: ICD-9 486

2. Septicemia: ICD-9 0389

3. Urinary Tract Infection: ICD-9 5990

These diseases were chosen because of their high frequency in hospital-acquired diseases list.

Count of admissions

Count of admissions is calculated as the total of number of admissions for the disease or symptom in each hospital.

Average daily charges

Average daily charges is the average of daily charges in all records for the given hospital and disease. Daily charges are not available in Medicare claims data and are calculated as the ratio of total charges to the length of stay.

The portal will also enable users to filter the ranking results and get a list of hospitals with only 5 stars, 4 stars or more, only 4 stars and so on. Users can falso ilter the results to see only the hospitals where higher the length of stay higher the chance of discharging alive or the hospitals where higher the length of stay higher the chance of discharging dead or the hospitals where a correlation could not be established based on the data available. Users can filter the results for mortality ratio below 26%, mortality ratio between 26% and 51%, between 51% and 76% and above 76%. Another information added in the list is the 'Subpart or Not' information. The portal gives option to filter out parent or subpart institutions from the result list.

In addition, the web page will give some information such as code type and description of the admission or primary diagnosis code given by the user as input. This information is fetched from HIPAASpace ICD-9 Code Lookup website on real-time. The portal will also show the typical diseases that are treated in each NPI. Also the hospital's location by zip code, distance and driving duration to this hospital from user's current location will also be displayed in Google map with the help of Google Geocoding API. The web portal will also list hospitals nearest to the user based on the zip code. The NPI information such as the name of the hospital, mailing address, phone, practice location address and practice in real-time from HIPAASpace NPI Number Lookup website will also be shown.

Chapter 3

System Design

The proposed data driven, hospital quality of care portal (MediQoC) for the patient community has the four below components:

1. The client

2. The web server

3. The application server

4. The comma separated values (csv) data file
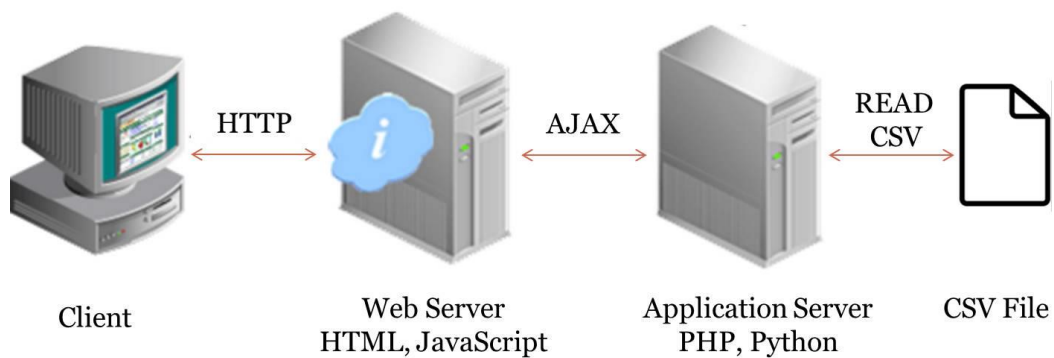

Figure 3-1 The MediQoC Architecture

The Client

A client accessing the portal from a web browser represents the end user. End user such as patients access MediQoC using a web browser like Google Chrome and navigates through the portal. The patient selects the ranking parameters and clicks the 'Relative Ranking' button in the portal to see the ranked list of hospitals.

The Web Server

A web server in the internet with HTML and JavaScript libraries enabled receives the requests from client and calls the appropriate application server functions and returns the response from application server back to the end users via the client. The web server also loads the result table and column charts in '.html' format written by the Python script in the application server into the client browser.

The Application Server

An application server with PHP and Python is the core component of the system. All the data manipulation is performed by this server based on the request received from user via the web server. The application server is in fact not separated from the web server, but is a separate component working on the same server machine. The web server calls the PHP script corresponding to the user request. The PHP script calls the corresponding Python script as a command in the exec() function. The Python script reads the dataset and performs the needed query corresponding to user request and returns the response back to PHP script. The PHP script returns this response back to the web server. The Python script also writes the result table and column charts as '.html' files into the web server.

The application server uses Python language to perform the data processing. The server uses 'Anaconda' [32] distribution of 'Python' [33] and specifically the 'Pandas' [34], 'NumPy' [35] and 'SciPy' [36] packages. Pandas is used to read the csv file (dataset) into a DataFrame. Numpy is used to apply mathematical functions on the DataFrame values and Scipy is used for scientific computing on the DataFrame values. The statistical correlation between LOS and discharge status is calculated using the 'pointbiserialr()' function in Scipy. The application server also uses the 'Bokeh' [37]

package in Python for visualization to plot the column charts. More details on these packages are provided in the Appendix A.

The Comma Separated Values (CSV) Data File

The csv file is the Medicare dataset in the '.csv' format. The NPI data file is also another csv file used by the application server to get the results. The application server reads the csv files into a Pandas DataFrame and applies the required queries and manipulations to get the result list.

Data Design

As mentioned in the system design, this system reads data from csv files and hence no relational database is used in the system. The reasons for not converting the csv data to a relational database are mentioned below:

1. There is no unique key in the Medicare claims dataset and hence storing it to relational database is not advisable

2. No update or insert is required to the data, hence relational database is not necessary

3. As lot of mathematical functions is to be applied to the data, Python along with its scientific computing packages are ideal to use and pandas package in Python is fast in terms of reading the csv data to DataFrame which is a tabular representation of the data. The use of Scientific Python reduces the lines of code required to access and manipulate the data.

4. With more than a million rows in the dataset, performance of accessing the data from csv is much better than reading and manipulating the data from a relational database

There are primarily two datasets used in this system:

1. Medicare Claims dataset

2. National Provider Identifier data

Medicare Claims dataset

The main dataset used in this system is the Medicare in-hospital claims file which contains records for the 100% Medicare beneficiaries who used hospital in-patient services during the year 2013. Medicare claims dataset of 2013 is made available by the Centers for Medicare and Medicaid Services (CMS). At present, the MediQoC uses only the records in this dataset for those who used hospital in-patient services only in Texas, United States. This dataset for Texas in its csv format has a size of about 501.3 MegaBytes, has 1035685 tuples and about 214 columns. We again removed records with primary diagnosis codes which are not having a total of at least 50 cases. This is because, the disease is not frequent and we cannot get reliable information from those records. The 214 attributes can be broadly classified into below six categories:

1. Admission information and demographics

2. Discharge information

3. Clinical outcomes

4. Hospital procedures

5. Diagnoses

6. Cost of care and diagnoses related groups

We primarily used only the below 56 attributes from this dataset:

1. National Provider Id (NPI)

2. State

3. Length of Stay

4. Total Charges

5. Discharge Status

6. Admission Code

7. Diagnosis Codes 1 to 25

8. Procedure Codes 1 to 25

9. Present On Admission Diagnosis Indicator 1 to 25

Below are the details of these attributes.

'NATIONAL_PROVIDER_ID'

The National Provider Identifier (NPI) is a unique ten-digit identifier for the health care providers in the United States. It is assigned by the Centers for Medicare and Medicaid Services (CMS). 'NATIONAL_PROVIDER_ID' column shows the NPI of the institution which handled the case. This is not a primary key. The dataset has no key as it has just records of various claims and each record has an NPI. There can be many records with same NPI and same or different admission and primary diagnosis code.

E.g.: 1003833013

'STATE'

'STATE' represents the state in the United States in which the case was recorded. MediQoC uses only those records with state value as 'TX' (Texas).

'LENGTH_OF_STAY'

The number of days the patient is admitted in the hospital for the particular case or record.

'TOTAL_CHARGES'

The total amount (rounded to whole dollars) of all charges (covered and non-covered) for all services provided to the beneficiary for the stay.

'DISCHARGE_STATUS'

The condition of the patient when discharged after the specific case or record.

- A: patient discharged alive from hospital

- B: patient discharged dead from hospital

- C: still a patient

'ADMITTING_DIAGNOSIS_CODE'

'ADMITTING_DIAGNOSIS_CODE' is the ICD-9 code for the diseases a patient was recorded with during the time of admission. There is only one admission code assigned in each case or record and it may or may not be the same as the primary diagnosis code. There can be many cases in which a patient is admitted for one particulat ICD-9 code but diagnosed to have a different ICD-9 code as the primary diagnosis code.

'DIAGNOSIS_CODE_x'

'DIAGNOSIS_CODE_x' where x is from 1 to 25 fields indicate the ICD-9 code for the diseases a patient is diagnosed to have. A patient can be assigned maximum up to 25 diseases during a single admission. 'DIAGNOSIS_CODE_1' is referred to as the primary diagnosis code in this thesis. The 'DIAGNOSIS_CODE_1' to 25 is used along with the 'POA_DIAGNOSIS_INDICATOR_1' to 25 to calculate the percentage of hospital acquired diseases.

The 'ADMITTING_DIAGNOSIS_CODE' and 'DIAGNOSIS_CODE_x' is the ICD-9-CM code which is replaced by an equivalent ICD-10-CM code (or codes) when the United States transitioned from ICD-9-CM to ICD-10-CM on October 1, 2015. The dataset is still having the ICD-9 codes.

E.g.: V5789: Care involving other specified rehabilitation procedure

0389: Unspecified septicemia

'PROCEDURE_CODE_x'

'PROCEDURE_CODE_x' where x is from 1 to 25 fields indicate the ICD-9 code for the various tests a patient had to take during the course of the treatment. A patient can be assigned maximum up to 21 procedure codes in a single admission.

E.g.: 9670: Continuous invasive mechanical ventilation of unspecified duration

'POA_DIAGNOSIS_INDICATOR_x'

'POA_DIAGNOSIS_INDICATOR_x' where x is from 1 to 25 fields indicate whether the particular diagnosis code was present on admission (POA). A patient can be diagnosed with maximum of 25 diagnosis codes (diseases) in a single admission. The POA_DIAGNOSIS_INDICATOR indicates if each of these 25 diseases were present while admitting to hospital. If a disease was not POA, but the patient record has it, it means the disease was acquired during the course of treatment in hospital. The 'POA_DIAGNOSIS_INDICATOR _1' to 25 is used along with the 'DIAGNOSIS_CODE_1' to 25 to calculate the percentage of hospital acquired diseases. Below are POA values and description from CMS.gov Hospital-Acquired Conditions (Present on Admission Indicator) page [25]:

• Y: Diagnosis was present at time of inpatient admission

- N: Diagnosis was not present at time of inpatient admission

- U: Documentation insufficient to determine if the condition was present at the time of inpatient admission

- W: Clinically undetermined. Provider unable to clinically determine whether the condition was present at the time of inpatient admission

- 1: Unreported/Not used. Exempt from POA reporting. This code is equivalent to a blank.

A complete list of attributes in the Medicare claims dataset is listed in the Appendix B.

National Provider Identifier data

The second dataset used is the National Provider Identifier (NPI) dataset: 'npidata_20050523-20150607.csv' with comprehensive information on each of the registered NPI to date. The NPI data also is available to download from Centers for Medicare and Medicaid Services (CMS) [24]. This dataset in its csv format has a size of about 5.6 GigaBytes, has 1048575 tuples and about 329 columns. This dataset includes information such as NPI, its registered address, its specialty information as Taxonomy codes, if it is a subpart or not, zip code etc. We used the National Provider Identifier (NPI) of the two datasets as a common key to link the data and included the subpart information in the ranking results table and also enabled Geolocation based services by using the zip code information.

Below are the details of the attributes used:

'NPI'

National Provider Identifier same as the Medicare claims dataset. 'NPI' is the unique key in this dataset. 'NPI' is used the key to get subpart information in this dataset and added to the result.

'Is Organization Subpart'

A heath care organization with multiple functional components or physical locations can get them their own NPIs and refer to as 'Subpart NPI's [22]. The functional component can be a laboratory or surgical department which may be part of a larger organization but can be separately licensed and may bill separately for health care services. The functional component or location is a subpart of a larger NPI.

Below are the values for this field and its description:

- X: Not Answered

- Y: Yes, Entity Type 2 Provider (Organization) is a Subpart

- N: No, Entity Type 2 Provider (Organization) is not a Subpart

'ploczip'

Zip code for each of the NPI.

Technical Specification

Software Requirements

1. Client

- Operating System   : Microsoft Windows NT/2000 (or higher)

- Client side   : JavaScript, HTML

- Browser   : Google Chrome 44.0.2403.155 m or higher

2. Server

- Operating System   : Microsoft Windows NT/2000 (or higher)

- Server side   : PHP 5.5.12

Python 2.7.8 (Anaconda 2.1.0 64 bit) with Pandas
(0.16.1), Scipy (0.15.1), NumPy (1.9.2)

Bokeh (0.8.2)

Hardware Requirements

1. Client

- Processor   : Minimum- Intel Core 2 Duo
- Memory   : Minimum- 4 GB

2. Server

- Processor   : Minimum- Intel Core i7 or a processor with 2.6 GHz speed or
                         more
- Memory   : Minimum: 8 GB, Recommended: 12 GB or more

Performance Constraints

Memory Constraints

The system as works on dataset with millions of records require reasonable amount of memory. A system with at-least 8 GB RAM is required to load the dataset and query on it.

Speed Constraints

PC with processor of 2.6 GHz speed or above is required for the application server.

Chapter 4

System Implementation and Testing

System Implementation

The design of system is converted to code in a programming language during coding

phase. The web interface is designed to have mainly 4 ranking options:

1. Admission/ Primary Diagnosis Code Statistics Per National Provider ID

2. National Provider ID Statistics Per Admission/ Primary Diagnosis Code

3. Hospital Acquired Primary Diagnosis Code Statistics Per National Provider ID

4. Procedure Code Statistics Per National Provider ID

The results for each of the above use cases are shown below.

Admission/ Primary Diagnosis Code Statistics per National Provider ID

The user wants to get a ranked list of top 5 hospitals with lowest average total

charge for cases with primary diagnosis code Pneumonia (ICD-9 486) and selects total

charges as the relative ranking attribute and submits the form. See Figure 4-1 on page 39

for the form submitted.

The web portal displays the ranked list in a table, see Table 4-1 on page 40 and

displays the column chart plotting average total charge for each NPI in the result; see

Figure 4-2 on page 41. Finally, the information about the submitted primary diagnosis

code is also obtained in real-time from HIPAASpace ICD-9 Code Lookup website; see

Figure 4-3 on page 41.

☐ Admission/ Primary Diagnosis Code Statistics Per National Provider ID

Sort by Performance Indicators!

Enter number of NPI to display in results (Required):	5
Select Code Type (Required):	○ Admission Code ⊙ Primary Diagnosis Code
Enter Diagnosis Code 1 (Required):	486
Minimum no. of cases required to consider the Diagnosis Code 1:	50
Select Parameter to Sort (Required):	Average Total Charge ▼
Select Sort Order (Required):	Ascending ▼
Select Relative Ranking Attribute (Required):	Average Total Charge ▼
Enter the only NPI to see in result (Optional):	
Select Star Rating Filter (Optional):	--- Select Star Rating Filter --- ▼
Select Mortality Ratio Filter (Optional):	--- Select Mortality Ratio Filter --- ▼
Select NPI Subpart Filter (Optional):	--- Select NPI Subpart Filter --- ▼
Select Correlation Filter (Optional):	--- Select LOS vs. Discharge Status Correlation Filter --- ▼
	Get Relative Ranking

Figure 4-1 User Selects the Performance Indicators and Submits Form.

Table 4-1 Results for Top 5 Hospitals with Lowest Average Total Charge for Primary Diagnosis Code Pneumonia (ICD-9 486) with

Star Rating for Average Total Charge.

| NPI	Avg. Total Charge (in US$)	Avg. Days of Stay	Mortality Ratio	Appropriateness (in %)	Correlation	Is Organization Subpart	Star Rating By Avg. Total Charge
1295739258	1095	1	0	14.29	Could Not Establish a Correlation	N	5 Star
1467434225	1226	1	100	50	Could Not Establish a Correlation	N	5 Star
1306844519	1431	1	0	100	Could Not Establish a Correlation	Info Unavailable	5 Star
1134116049	1500	4	0	33.33	Could Not Establish a Correlation	N	5 Star
1366519027	1840	2	0	25	Could Not Establish a Correlation	Info Unavailable	5 Star
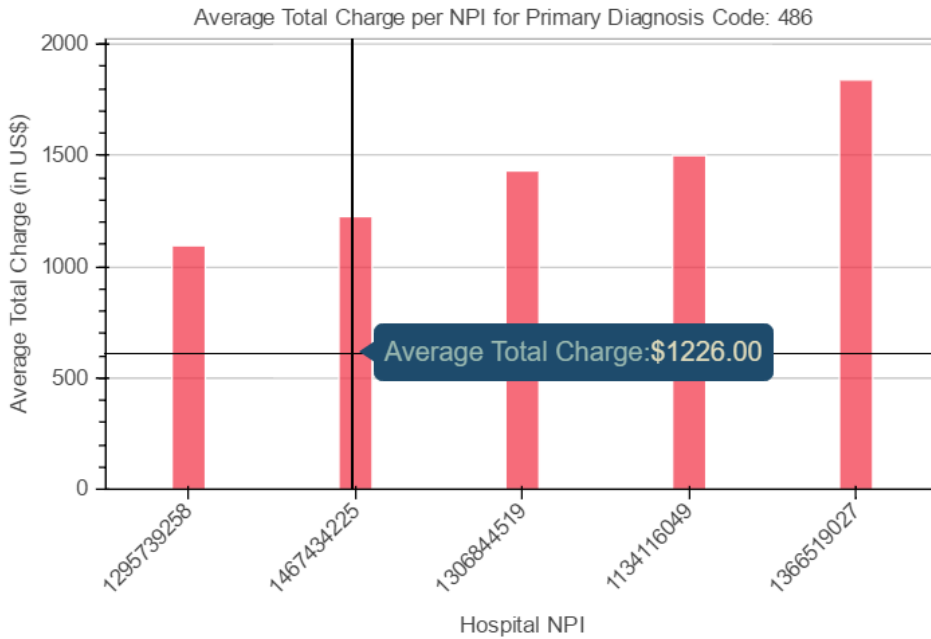
Figure 4-2 Column Chart Plotting the Average Total Charge for the NPIs with Lowest

Average Total Charge for Pneumonia.

Figure 4-3 Diagnosis Code Information Obtained from HIPAASpace ICD-9 Code Lookup

Website in Real-time.

National Provider ID Statistics Per Admission/ Primary Diagnosis Code

The user wants to get a ranked list of typical diseases treated in a national provider. The user needs top 5 primary diagnosis codes with lowest average total charge in NPI: 1386868388 and selects total charges as the relative ranking attribute and submits the form.

The web portal displays the ranked list in a table, see Table 4-2 on page 42 and displays the column chart plotting average total charge for each primary diagnosis codes in the result; see Figure 4-4 on page 43. Finally, the information about the submitted NPI is also obtained in real-time from HIPAASpace NPI Number Lookup website; see Figure 4-5 on page 43. The distance to this national provider from user's current location is displayed in map; see Figure 4.6 on page 44.

Table 4-2 Results for Top 5 Primary Diagnosis Codes with Lowest Average Total Charge in NPI 1386868388.

Primary Diagnosis Code	Avg. Total Charge (in US$)	Avg. Days of Stay	Mortality Ratio	Appropriateness (in %)
29284	9116	1	0	0.11
29652	11137	3	0	0.11
311	13493.5	3.6	0	1.1
29663	14818	5	0	0.11
29623	15252.25	5	0	0.44

Figure 4-4 Column Chart Plotting the Average Total Charge for the Primary Diagnosis

Codes with Lowest Average Total Charge in NPI 1386868388.

Figure 4-5 NPI Information Obtained from HIPAASpace NPI Number Lookup Website in

Real-time.
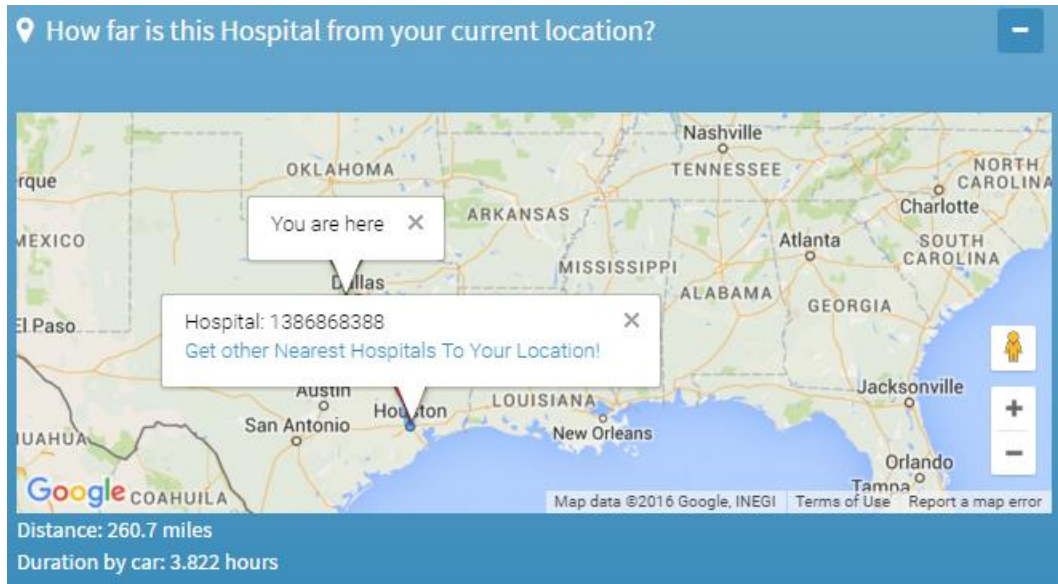
Figure 4-6 Distance to NPI 1386868388 from User's Location.

Hospital Acquired Primary Diagnosis Code Statistics Per National Provider ID

The user wants to get a ranked list of top 5 hospitals with highest hospital acquired percentage for cases with primary diagnosis code Septicemia (ICD-9 0389) and selects hospital acquired percentage as the relative ranking attribute and submits the form.

The web portal displays the ranked list in a table, see Table 4-3 on page 45 and displays the column chart plotting hospital acquired percentage for each NPI in the result; see Figure 4-7 on page 46. Finally, the information about the submitted primary diagnosis code is also obtained in real-time from HIPAASpace ICD-9 Code Lookup website; see Figure 4-8 on page 46.

Table 4-3 Results for Top 5 Hospitals with Highest Hospital Acquired Percentage for Primary Diagnosis Code Septicemia (ICD-9 0389) with Star Rating for Hospital Acquired Percentage.

| NPI	Avg. Total Charge (in US$)	Avg. Days of Stay	Mortality Ratio	Appropriateness (in %)	Correlation	Is Organization Subpart	Star Rating By Percentage of Hospital Acquired Septicemia	Hospital Acquired Septicemia (in %)
1396765681	38707	8	0	11.11	Could Not Establish a Correlation	N	1 Star	100
1417010653	32521.429	2.1428571	14.29	0.48	Could Not Establish a Correlation	Info Unavailable	3 Star	57.692308
1962504340	113059.67	11	33.33	0.17	Could Not Establish a Correlation	Info Unavailable	3 Star	53.333333
1093719932	35343	9.3333333	0	8.82	Could Not Establish a Correlation	N	3 Star	50
1093740128	7200	2	0	3.85	Could Not Establish a Correlation	N	3 Star	50
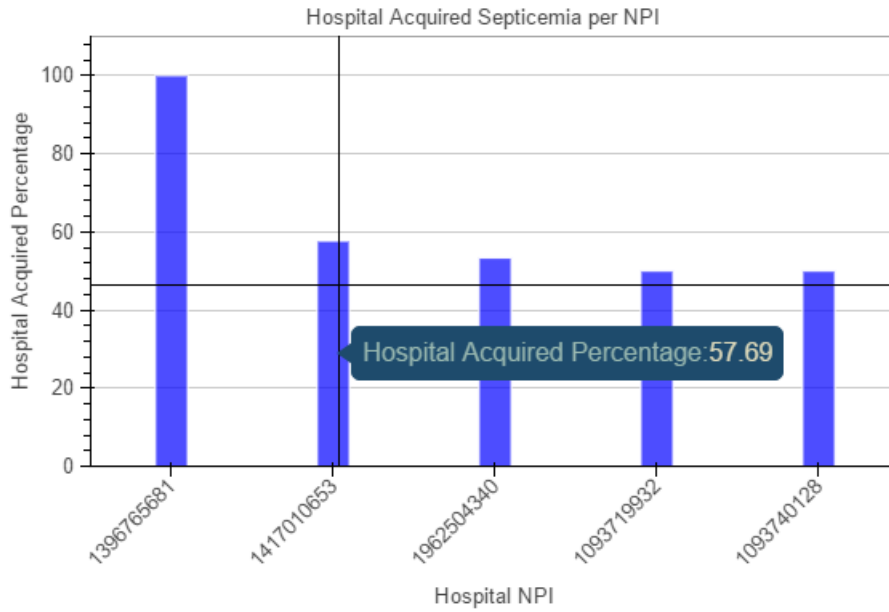
Figure 4-7 Column Chart Plotting the Hospital Acquired Percentage for the NPIs with

Highest Hospital Acquired Percentage for Septicemia.

Figure 4-8 Diagnosis Code Information Obtained from HIPAASpace ICD-9 Code Lookup

Website in Real-time.

Procedure Code Statistics Per National Provider ID

The user wants to get a ranked list of top 5 hospitals with lowest average length of stay for cases with procedure code 1 with ICD-9 9670 and selects average length of stay as the relative ranking attribute and submits the form.

The web portal displays the ranked list in a table, see Table 4-4 on page 48 and displays the column chart plotting average length of stay for each NPI in the result; see Figure 4-9 on page 48.

Table 4-4 Results for Top 5 Hospitals with Lowest Average Length of Stay for Procedure Code 1 with ICD-9 9670 with Star Rating for Average Length of Stay.

NPI	Avg. Total Charge (in US$)	Avg. Days of Stay	Appropriateness (in %)	Is Organization Subpart	Star Rating By Avg. Stay
1003813742	2065	1	1.56	N	5 Star
1063411767	16044	1	0.04	N	5 Star
1184622847	5805	1	0.01	N	5 Star
1548232044	38202.5	1	0.06	Info Unavailable	5 Star
1730132234	25829	1	0.01	N	5 Star
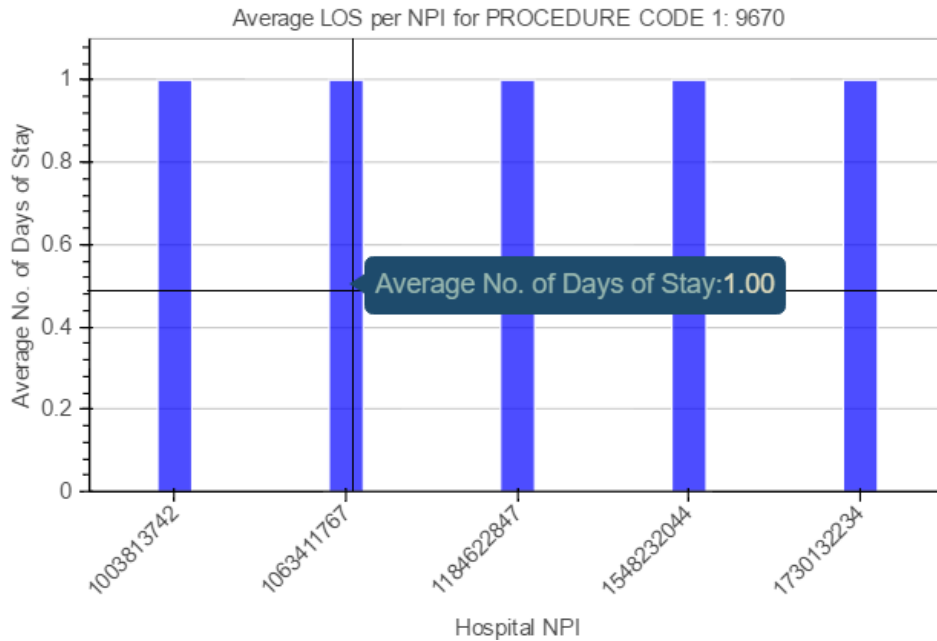
Figure 4-9 Column Chart Plotting the Average Length of Stay for the NPIs with Lowest

Average Length of Stay for Procedure Code 1 with ICD-9 9670.

Test Criteria and Methods

Testing is essential to find any errors in the system. The results of query and 'groupby()' functions were verified by obtaining the results for single NPI or primary diagnosis code and manually verifying the data. Results were also verified with the help of data analysis software like Microsoft Excel.

This system's online portal was tested for input form validation and output verification was also performed. The website is fully functional and best viewed in Google Chrome browser of version 44.0.2403.155 m or higher. The website is fully functional also in Microsoft Edge 25.10586.0.0. The website is known to have issues with running the JavaScript functions in the web page when loaded in Mozilla Firefox browser.

Chapter 5

Conclusion and Future Enhancements

Conclusion

Active patient participation in choice of quality and safe hospital services is not yet completely fulfilled. We wanted this thesis as a successful step towards this direction. This system is data driven based on historical data. The system provides a user friendly web interface for users to select from given list of performance indicators such as mortality ratio, percentage of hospital acquired diseases, length of stay and rank hospitals based on the selected factors. The system empowers patients, their caregivers and family doctors who want to direct their patients to appropriate services. The system also acts as a pressure to the health care systems to provide more quality services. It can potentially be used by insurance organizations which are going to be evaluating hospitals for more precise reimbursement coverage.

Future Enhancements

We identified some of the areas of improvement for this system and will be considered in future enhancements of the system. The dataset we used for this system did not include patient satisfaction survey data. The rankings are therefore purely based on historical quantitative data and not based on any feedback. We want to figure out a way to incorporate patient feedback data into some sort of quantitative value for ranking. The system is limited to Medicare patient records of Texas. Though we consider this as a representative model and expect similar behavior or results for Medicare records of other states, we want to expand this system with ability to rank hospitals of all states by taking into account the claims records for all states. The system though accurate, can be made more efficient in terms of handling large data file and returning the results faster. The

system will have to implement parallel processing to achieve this. Finally, we plan to integrate machine learning functionalities for predictive modeling. This will make the system provide personalized results for individual patients by taking into account the patient demographics, medical history and preferences.

Appendix A

Related Terms

Medicaid

Medicaid is another health coverage plan that covers millions of Americans. It covers people of different age groups like eligible low-income adults, children, pregnant woman, elderly adults and people with disabilities. It is funded jointly by states and the Federal Government in the United States [27].

ICD-10

Beginning October 1 2015, in the United States ICD-10 codes will be used instead of the ICD-9 codes.

Python

Python is a powerful and fast interpreted programming language which is developed under an OSI-approved open source license.

Anaconda

Anaconda is a free Python distribution containing the most popular Python packages for science, math, engineering, and data analysis.

Pandas

'pandas' is an open source, BSD-licensed Python library. 'pandas' have tools for reading and writing data between in-memory data structures and different formats like CSV and HTML. The library has fast and efficient 'DataFrame' object for data manipulation. DataFrame when used with IPython notebook can provide a tabular representation of data read from csv files making it easier in data analysis tasks.

NumPy

NumPy is the fundamental package for scientific computing with Python. It contains a powerful N-dimensional array object for faster processing

SciPy

The SciPy library is a library for Python mainly for providing scientific computing operations.

Bokeh

Bokeh is an interactive visualization library in Python. It enables to include interactive visualization plots to html.

Appendix B

Medicare Claims Dataset

Medicare Claims Dataset

The Medicare Claims Dataset has 214 columns. Once the file is loaded, out of the 214 attributes, we retained only below attributes (not in the order as in dataset) on which our study is focused on:

1. 'NATIONAL_PROVIDER_ID'
2. 'DISCHARGE_STATUS'
3. 'LENGTH_OF_STAY''
4. 'TOTAL_CHARGES'
5. 'POA_DIAGNOSIS_INDICATOR_1'
6. 'POA_DIAGNOSIS_INDICATOR_2'
7. 'POA_DIAGNOSIS_INDICATOR_3'
8. 'POA_DIAGNOSIS_INDICATOR_4'
9. 'POA_DIAGNOSIS_INDICATOR_5'
10. 'POA_DIAGNOSIS_INDICATOR_6'
11. 'POA_DIAGNOSIS_INDICATOR_7'
12. 'POA_DIAGNOSIS_INDICATOR_8'
13. 'POA_DIAGNOSIS_INDICATOR_9'
14. 'POA_DIAGNOSIS_INDICATOR_10'
15. 'POA_DIAGNOSIS_INDICATOR_11'
16. 'POA_DIAGNOSIS_INDICATOR_12'
17. 'POA_DIAGNOSIS_INDICATOR_13'
18. 'POA_DIAGNOSIS_INDICATOR_14'
19. 'POA_DIAGNOSIS_INDICATOR_15'
20. 'POA_DIAGNOSIS_INDICATOR_16'
21. 'POA_DIAGNOSIS_INDICATOR_17'

22. 'POA_DIAGNOSIS_INDICATOR_18'

23. 'POA_DIAGNOSIS_INDICATOR_19'

24. 'POA_DIAGNOSIS_INDICATOR_20'

25. 'POA_DIAGNOSIS_INDICATOR_21'

26. 'POA_DIAGNOSIS_INDICATOR_22'

27. 'POA_DIAGNOSIS_INDICATOR_23'

28. 'POA_DIAGNOSIS_INDICATOR_24'

29. 'POA_DIAGNOSIS_INDICATOR_25'

30. 'DIAGNOSIS_CODE_1'

31. 'DIAGNOSIS_CODE_2,

32. 'DIAGNOSIS_CODE_3'

33. 'DIAGNOSIS_CODE_4'

34. 'DIAGNOSIS_CODE_5'

35. 'DIAGNOSIS_CODE_6'

36. 'DIAGNOSIS_CODE_7'

37. 'DIAGNOSIS_CODE_8'

38. 'DIAGNOSIS_CODE_9'

39. 'DIAGNOSIS_CODE_10'

40. 'DIAGNOSIS_CODE_11'

41. 'DIAGNOSIS_CODE_12'

42. 'DIAGNOSIS_CODE_13'

43. 'DIAGNOSIS_CODE_14'

44. 'DIAGNOSIS_CODE_15'

45. 'DIAGNOSIS_CODE_16'

46. 'DIAGNOSIS_CODE_17'

47. 'DIAGNOSIS_CODE_18'

48. 'DIAGNOSIS_CODE_19'

49. 'DIAGNOSIS_CODE_20'

50. 'DIAGNOSIS_CODE_21'

51. 'DIAGNOSIS_CODE_22'

52. 'DIAGNOSIS_CODE_23'

53. 'DIAGNOSIS_CODE_24'

54. 'DIAGNOSIS_CODE_25'

55. 'PROCEDURE_CODE_1'

56. 'PROCEDURE_CODE_2'

57. 'PROCEDURE_CODE_3'

58. 'PROCEDURE_CODE_4'

59. 'PROCEDURE_CODE_5'

60. 'PROCEDURE_CODE_6'

61. 'PROCEDURE_CODE_7'

62. 'PROCEDURE_CODE_8'

63. 'PROCEDURE_CODE_9'

64. 'PROCEDURE_CODE_10'

65. 'PROCEDURE_CODE_11'

66. 'PROCEDURE_CODE_12'

67. 'PROCEDURE_CODE_13'

68. 'PROCEDURE_CODE_14'

69. 'PROCEDURE_CODE_15'

70. 'PROCEDURE_CODE_16'

71. 'PROCEDURE_CODE_17'

72. 'PROCEDURE_CODE_18'

73. 'PROCEDURE_CODE_19'

74. 'PROCEDURE_CODE_20'

75. 'PROCEDURE_CODE_21'

76. 'PROCEDURE_CODE_22'

77. 'PROCEDURE_CODE_23'

78. 'PROCEDURE_CODE_24'

79. 'PROCEDURE_CODE_25'

80. 'ADMITTING_DIAGNOSIS_CODE'

81. 'STATE'

Appendix C

System Screenshots

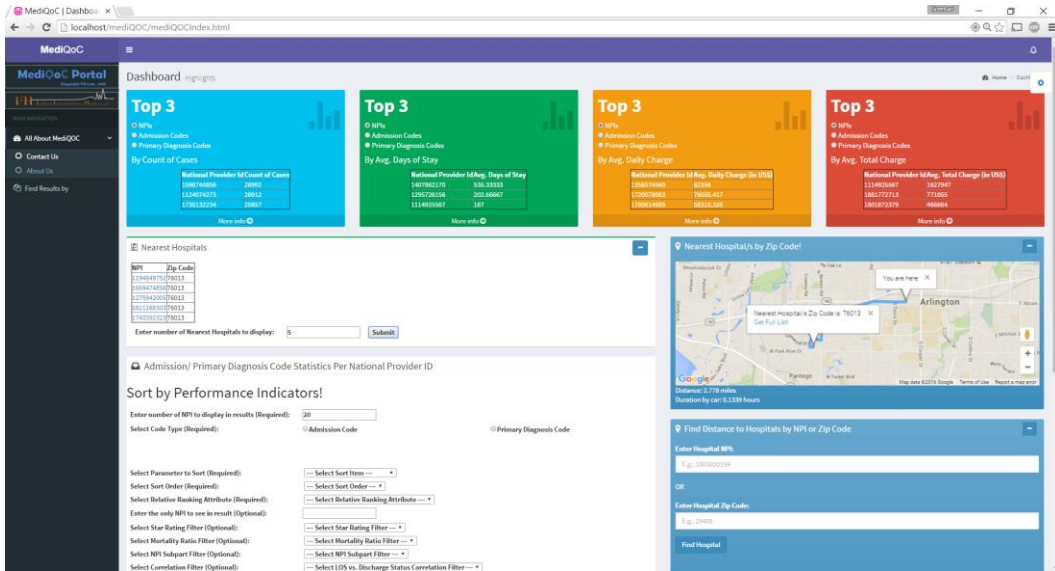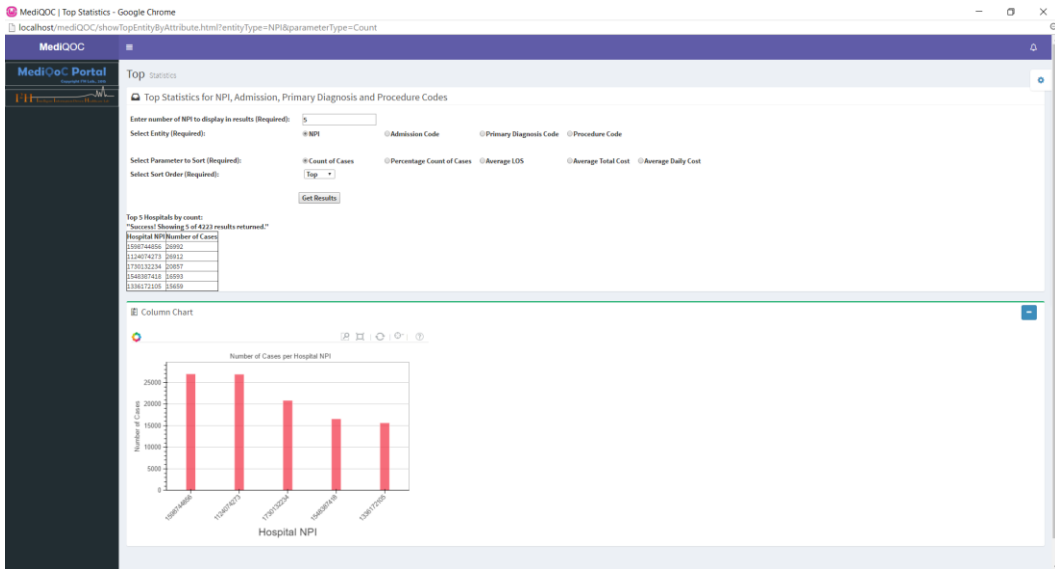Figure 0-1 MediQoC Home Page with Highlights and Nearest Hospitals Location and List.

Figure 0-2 Top 5 NPIs by Number of Admissions.

🖴 Admission/ Primary Diagnosis Code Statistics Per National Provider ID

Sort by Performance Indicators!

Enter number of NPI to display in results (Required): `20`

Select Code Type (Required): ⦿ Admission Code   ○ Primary Diagnosis Code

Enter Admission Code (Required): `4280 Y`

Select Parameter to Sort (Required): `Average Daily Charge ▼`

Select Sort Order (Required): `Descending ▼`

Select Relative Ranking Attribute (Required): `Count of Cases ▼`

Enter the only NPI to see in result (Optional): [   ]

Select Star Rating Filter (Optional): `--- Select Star Rating Filter --- ▼`

Select Mortality Ratio Filter (Optional): `--- Select Mortality Ratio Filter --- ▼`

Select NPI Subpart Filter (Optional): `--- Select NPI Subpart Filter --- ▼`

Select Correlation Filter (Optional): `--- Select LOS vs. Discharge Status Correlation Filter --- ▼`

[Get Relative Ranking]

Results for Admission/ Primary Diagnosis Code: 4280 Y
"type 'exceptions.ValueError'"
0 Results! Try ranking with a different code for Admission Code or change filters for NPI or star rating or mortality ratio or NPI subpart or correlation filter.
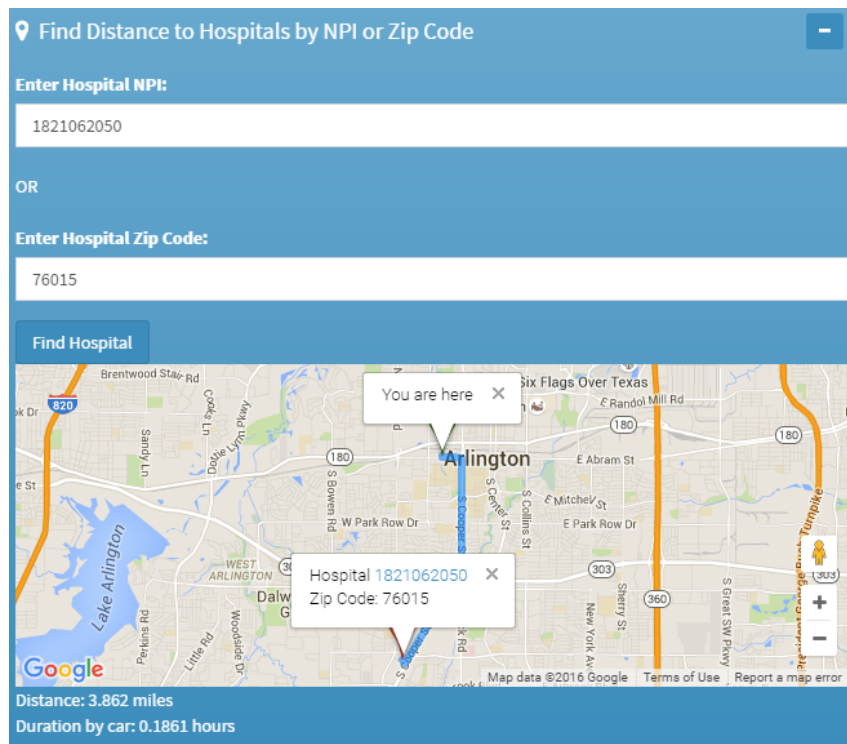
Figure 0-3 No Results Returned for Admission Code 4280 Y.

Figure 0-4 See Location, Distance and Duration by Giving NPI.

60

Admission/ Primary Diagnosis Code Statistics Per National Provider ID

Sort by Performance Indicators!

Enter number of NPI to display in results (Required):	20
Select Code Type (Required):	○ Admission Code ● Primary Diagnosis Code
Enter Diagnosis Code 1 (Required):	486
Minimum no. of cases required to consider the Diagnosis Code 1:	50
Select Parameter to Sort (Required):	Count of Cases ▼
Select Sort Order (Required):	Descending ▼
Select Relative Ranking Attribute (Required):	Count of Cases ▼
Enter the only NPI to see in result (Optional):	
Select Star Rating Filter (Optional):	--- Select Star Rating Filter --- ▼
Select Mortality Ratio Filter (Optional):	Below 26% ▼
Select NPI Subpart Filter (Optional):	Not a Subpart ▼
Select Correlation Filter (Optional):	+ve, Higher LOS => More Discharged Alive ▼

Get Relative Ranking

Results for Admission/ Primary Diagnosis Code: 486
"Success! Showing 6 of 6 results returned."

NPI	Case Count	Avg. Daily Charge (in US$)	Avg. Total Charge (in US$)	Avg. Days of Stay	Mortality Ratio	Appropriateness (in %)	% of Diagnosis Code in this NPI out of total count of this Diagnosis Code	Los vs Discharge Status Correlation: R Value	Los vs Discharge Status Correlation: P Value	Correlation	Is Organization Subpart	Star Rating By No. of Cases	Count of Pneumonia Diagnosed	Count of Pneumonia Present on Admission	Count of Hospital Acquired Pneumonia	Hospital Acquired Pneumonia (in %)
1023013448	320	7065.9901	28340.094	4.34375	3.75	4.87	0.85	0.049761275	0.10977761	+ve, Higher LOS means more chance of Discharging Alive	N	3 Star	872	824	47	5.3899083
1437148020	90	3077.5796	13075.178	4.5888889	6.67	9.05	0.24	0.010245757	0.26937334	+ve, Higher LOS means more chance of Discharging Alive	N	1 Star	149	138	10	6.7114094

Figure 0-5 Applying Filters to Ranked List.

References

[1] SoRelle, Ruth. "Reducing the rate of medical errors in the United States." Circulation 101.3 (2000): e39-e40.

[2] Louis and Harris Associates, 100 million Americans see medical mistakes directly touching them as patients, friends, relatives, National Patient Safety Foundation (2007).

[3] Goodman, John C., Consumer Driven Health Care. Networks Financial Institute Policy Brief, Indiana State University. (2006). doi:10.2139/ssrn.985572

[4] Dale S, Janice R, Luke H, Teresa D, 2014 CAHPS Health Plan Survey Database, 2014

[5] Mitchell, Pamela H., Defining Patient Safety & Quality Care, Patient Safety and Quality: An Evidence-Based Handbook for Nurses, Rockville (MD): Agency for Healthcare Research & Quality (US); 2008

[6] Lohr K. Committee to Design a Strategy for Quality Review and Assurance. In: Medicare, editor. Medicare: a strategy for quality assurance. Vol. 1. Washington, DC: National Academy Press; 1990.

[7] Mitchell, Pamela H, Ferketich S, Jennings BM., Quality health outcomes model. American Academy of Nursing Expert Panel on Quality Health Care 1998.

[8] Parente S. T., McCullough J. S. Health information technology and patient safety: Evidence from panel data. Health Affairs (2009) 28(2):357–360.

[9] Press, In. "The Digital Transformation of Healthcare: Current Status and the Road Ahead."

[10] Medicare Payment Advisory Commision. 2004. Report to the Congress: New Approaches in Medicare. Washington, D.C.: MedPAC, pp.160.

[11] Hickey GL, Grant SW, Cosgriff R, Dimarakis I, Pagano D, Kappetein AP, Bridgewater B. Clinical registries: governance, management, analysis and applications. Eur J Cardiothorac Surg. 2013;44:605–614.

[12] Durinka, Joel, et al. "MP9-19 DO HOSPITAL RANKINGS CORRELATE WITH ACTUAL OUTCOMES?." *The Journal of Urology* 193.4 (2015): e109.

[13] Williams, S., et al. "Ranking hospitals by the quality of care for medical conditions: the role of complications." *Transactions of the American Clinical and Climatological Association* 107 (1996): 263.

[14] Hubbard, Rebecca A., et al. "Identification of abnormal screening mammogram interpretation using Medicare claims data." *Health services research* 50.1 (2015): 290-304.

[15] Nattinger, Ann B., et al. "An algorithm for the use of Medicare claims data to identify women with incident breast cancer." *Health services research* 39.6p1 (2004): 1733-1750.

[16] Dimick, Justin B., Douglas O. Staiger, and John D. Birkmeyer. "Ranking hospitals on surgical mortality: the importance of reliability adjustment." *Health services research* 45.6p1 (2010): 1614-1629.

[17] Berwick D, Jain S, Porter M. Clinical registries: the opportunity for the nation. HealthAffairs Blog, May 2011. Available online at http://healthaffairs.org/blog/2011/05/ 11/clinical-registries-the-opportunity-for-the-nation/ (Last accessed July 30, 2014).

[18] Herrick, Devon M. "Consumer-Driven Health Care: The Changing Role of the Patient." Available online at http://www.ncpa.org/pub/st276/

[19] Khan, Fauzia The Shift to Consumer Driven Healthcare. The future of patient engagement. February 2014. Available online at http://healthcare-executive-

insight.advanceweb.com/Features/Articles/The-Shift-to-Consumer-Driven-Healthcare.aspx.

[20] Picker Europe. Our mission and values. http://www.pickereurope.org/our-mission-and-values.html

[21] Quick guide to health literacy http://health.gov/communication/literacy/quickguide/factsbasic.htm

[22] https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Practice%20Facility%20Profile/Organization_and_Support_NPI.pdf

[23] CMS website. Available from: https://www.cms.gov/

[24] CMS website. Available from: https://www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/NationalProvIdentStand/DataDissemination.html

[25] CMS website. Available from: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Coding.html

[26] Medicare website. Available from: https://www.medicare.gov/hospitalcompare/search.html

[27] Medicaid website. Available from: https://www.medicaid.gov/medicaid-chip-program-information/medicaid-and-chip-program-information.html

[28] HIPAASpace ICD-9 Code Lookup website. Available from: https://www.hipaaspace.com/Medical_Billing/Coding/International_Classification_Of_Diseases/ICD9_Codes_Lookup.aspx

[29] HIPAASpace NPI Number Lookup website. Available from: https://www.hipaaspace.com/Medical_Billing/Coding/National_Provider_Identifier/NPI_Number_Lookup.aspx

[30] "International Classification of Diseases (ICD)". World Health Organization. Archived from the original on 12 February 2014. Retrieved 14 March 2014.

[31] Google Geocoding website. Available from: https://developers.google.com/maps/documentation/geocoding/intro

[32] Anaconda website. Available from: https://www.continuum.io/downloads

[33] Python website. Available from: https://www.python.org/

[34] Pandas website. Available from: http://pandas.pydata.org/

[35] NumPy website. Available from: http://www.numpy.org/

[36] SciPy website. Available from: https://www.scipy.org/scipylib/index.html

[37] Bokeh website. Available from: http://bokeh.pydata.org/en/0.10.0/index.html

[38] SciPy statistical functions website. Available from: http://docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.chi2_contingency.html .

Biographical Information

Sreehari Balakrishna Hegden was born in Kerala, India in 1989. He received his Bachelor degree in Computer Science and Engineering from Mahatma Gandhi University, Kerala, India in 2011. He joined the University of Texas at Arlington, Texas in Fall 2014 to pursue his Master's degree in Computer Science. Sreehari has worked in Accenture Services Pvt. Ltd., Bangalore as a Software Engineering analyst on a Java based middleware integration platform called webMethods for three years. His areas of interests include Data Mining, Data Analysis and Modeling, Big Data and Cloud Computing.