UNIVERSITY OF CHICAGO

Group Members: ***Michael Diarra***
Marin Gow
Zain Iqbal
Zain Jafri
Venku Buragadda

MSCA 31012 -01 - Data Mining Principles

Project Description – Due on 05/02/2019 at 5:59pm

look at this start up!
healthatscale.com 16M Aetna
Lytic.tech/patient.html

I - PROJECT SUMMARY

According to the Alliance for Aging Research, every year 1.7 million Americans develop hospital acquired infections. 99,000 of those patients die. Yet most American do not have access to a readily available hospital ranking system that would help them make an informed decision when choosing a healthcare facility (in a non-emergency setting; e.g., planned/elective procedures, etc.).

This comes as a contrast with today's consumer expectations. Today consumers have come to rely on their peers' ratings to evaluate the quality of many products. Yelp, Google Reviews, Amazon product recommendations are good examples of such reliance on fellow consumers' reviews.

Our objective is to bring that peer review experience to the medical field and help patients make informed decisions based on the experience of their fellow patients. As such, we will first look to understand the relationship between patients' feedback on their use of specific healthcare facilities and the statistics of complications and death of those healthcare facilities. Based on our identification of the feedback areas most correlated with health outcomes, we will seek to build a recommendation model in Python that will suggest facilities that are highly rated in the most relevant areas.

The expected outcome is a solution that could potentially evolved into a mobile application that prospective and current patients will use to make more informed decisions about their hospital choices in non-emergency situations while also having the opportunity to contribute to the ranking by providing their feedback using the same mobile application.

II – DATASETS

- Hospital Patient Survey Data: A list of hospital ratings for the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). HCAHPS is a national, standardized survey of hospital patients about their experiences during a recent inpatient hospital stay.
- Hospital Complications and Deaths: Complications and deaths - provider data. This data set includes provider-level data for the hip/knee complication measure, the CMS Patient Safety Indicators, and 30-day death rates.

Data Overview

- National data set covering Medicare hospitals
- The dataset has a number of dimensions, therefore there should be a good number of dimensions and records to analyze
- There will be a fair amount of data cleaning/manipulation required. For example, in the quality data set we would need to pivot out the 'HCAHPS Question' field and in the complications data set pivot out the 'Measure Name' field.

Group Members: ***Michael Diarra***
Marin Gow
Zain Iqbal
Zain Jafri
Venku Buragadda

MSCA 31012 -01 - Data Mining Principles

Project Description – Due on 05/02/2019 at 5:59pm

Expected use of the datasets: Evaluate the relationship between hospital quality measures (based on patient survey data) and the prevalence of complications and deaths from inpatient stays. Many of the quality measures focus on the performance of hospital staff and how they take care of their patients (among other measures). The goal would be to identify quality measures that indicate poor performance in terms of complications and deaths. This could be used to recommend to patients and insurers which hospitals to use for non-emergency services.

TEAM MEMBERS ROLES

Each team member will be participating in every task listed below. The table below shows the name of the team member responsible for the coordination of the work related to each specific task. All team members will contribute to each major task.

Team Leads	Task
Michael Diarra	Scoping
Marin Gow	Data Preparation
Venku Buragadda	Exploratory Data Analysis
Zain Iqbal	Model Development
Zain Jafri	Class Presentation Document

everyone should be included in each step
That's ok to do most of the work based on task but everyone should be able to answer questions for entire project.

Please think of a graph based model to design a solution. It's not a must but I'd recommend to model the problem such as:

nodes: patient or hospital
attributes: patient and/or hospital quality measures, complications etc.
edges: ?