Define Problem / Problem Understanding

|  |  |
| --- | --- |
| Assignment Date | 12 may 2023 |
| Team ID | NM2023TMID03508 |
| Project Name | Estimetion and prediction of hospitalization and medical care costs |

**1.Define Problem / Problem Understanding**

A problem statement is a clear, concise, and specific description of an issue or challenge that needs to be addressed. It outlines the context and scope of the problem, the potential causes and consequences, and the goals or objectives that need to be achieved through addressing the problem.

### 1.1 Specify The Business Problem

**Estimation And Prediction Of Hospitalization And Medical Care Costs**

Medical costs are one of the most common recurring expenses in a person's life.

Based on different research studies, BMI,ageing, smoking, and other factors are

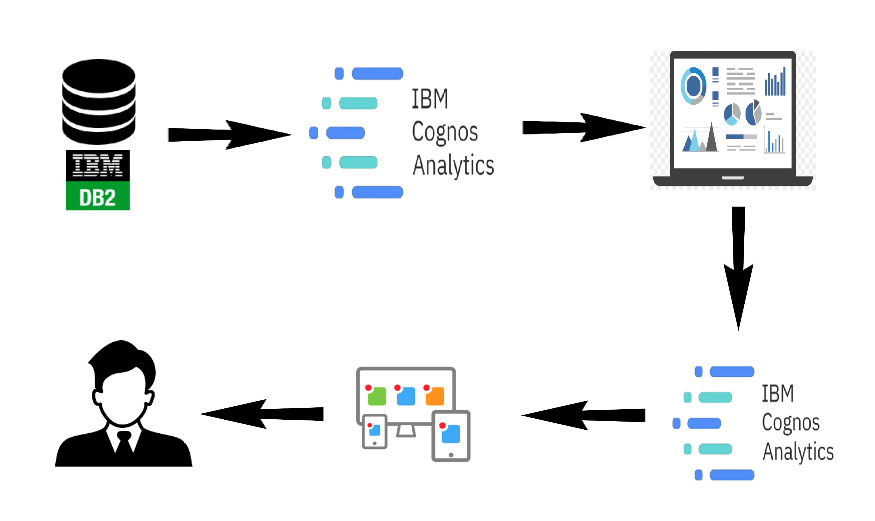
all related to greater personal medical care costs.The estimates of the

expenditures of health care related to obesity are needed to help create cost-

effective obesity prevention strategies.Obesity prevention at a

Young age is a top concern in global health,clinical practice, and public health.

**Technical Architecture:**



**1.2 Business requirements:**

This phase focuses on finding and understanding the questions of interest

from a business perspective, and transforming these demands into a data

mining problem definition. This first phase is related to the DSR’s Relevance

Cycle, due to the information exchange with the environment (organization

and stakeholders), and is decisive to make all the subsequent decisions during

the project. From this cooperation, the researcher needs to compose a set

of success criteria (metrics) to know what a "good" model will look like and

thus ensure that the answers to the problem would assist a decision-making

process. Within this phase, there are three tasks described next.

backcall==0.1.0

certifi==2019.11.28

Click==7.0

colorama==0.4.3

cycler==0.10.0

decorator==4.4.1

Flask==1.1.1

Flask-Cors==3.0.9

graphviz==0.13.2

ipython>=7.16.3

ipython-genutils==0.2.0

itsdangerous==1.1.0

jedi==0.15.2

Jinja2==2.11.3

joblib==0.14.1

kiwisolver==1.1.0

MarkupSafe==1.1.1

matplotlib==3.1.2

numpy==1.22.0

pandas==0.25.3

parso==0.5.2

pickleshare==0.7.5

prompt-toolkit==3.0.2

pydotplus==2.0.2

Pygments==2.7.4

pyparsing==2.4.6

python-dateutil==2.8.1

pytz==2019.3

scikit-learn==0.22

scipy==1.4.1

six==1.13.0

sklearn==0.0

traitlets==4.3.3

wcwidth==0.1.8

Werkzeug==0.16.0

wincertstore==0.2

gunicorn==20.0.4

### 1.3 Literature Survey

### Survey 1:

### Hindawi

### Journal of Healthcare Engineering Volume 2022, Article ID 7969220,

### <https://doi.org/10.1155/2022/7969220>

### The prevalence of obesity, which is defined as a [body mass index](https://www.sciencedirect.com/topics/medicine-and-dentistry/body-mass-index) (BMI) greater than 30, has increased dramatically in the United States since the late 1990s. So much so that recently obesity has been officially recognized as a disease by the American Medical Association, an action that could put more emphasis on the health condition by doctors and insurance companies to minimize its adverse effects. Currently, rates of obesity exceed 30% in most sex and adult age groups, whereas its prevalence among children and adolescents, defined as a BMI of more than the 95th percentile, has reached 17%.

**SURVEY 2:**

Walter A. Kukull, Thomas D. Koepsell, Douglas A. Conrad, Virginia Immanuel, Jan Prodzinski and Carol Franz

Medical Care

[Vol. 24, No. 10 (Oct., 1986)](https://www.jstor.org/stable/i290435), pp. 961-966 (6 pages)

Published By: Lippincott Williams & Wilkins

<https://www.jstor.org/stable/3765075>

In settings where an itemized hospital bill is not generated, estimation of hospitalization charges for research or administrative purposes can be a laborious task. This article examines the extent to which the number of hospital days spent outside an intensive care unit (ICU), number of days in an ICU, number of laboratory tests performed, number of x-rays, and number of surgeries can be used in a multiple regression equation to impute inpatient charges for a sample of 103 hospitalizations at a Veterans Administration hospital. These predictor variables, all of which are readily ascertained in a brief medical record review, accounted for about 97% of the variance in imputed hospital charges. The bootstrap method was applied for validation of the prediction equation. Application of the method described here may be of value to researchers concerned with hospital charge estimation in non-fee-for-service settings.

SURVEY 3:

# **Predicting Patient treatment costs using Machine Learning.**

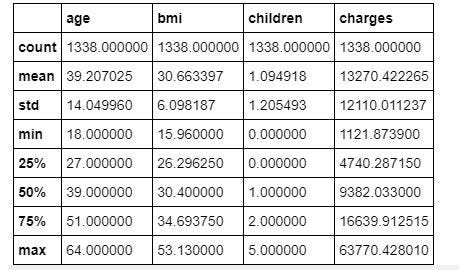
Data Analytics Engineering Graduate Student at Northeastern. Ex Senior Data

Engineer & IBM Certified Data Scientist.

[https://thomasgeorgethomas.com](https://thomasgeorgethomas.com/)

We try to identify numerical and categorical data. We proceed to collect basic descriptive stats using describe(). We try to understand what the data looks like and what it is trying to tell us.

data.describe()



**SURVEY 4:**

# **Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques**

Tripoliti - Cited by 203

Heart failure is a serious condition with high prevalence (about 2% in the adult

population in developed countries, and more than 8% in patients older than

75 years). About 3–5% of hospital admissions are linked with heart failure

incidents. Heart failure is the first cause of admission by healthcare

professionals in their clinical practice. The costs are very high, reaching up to

2% of the total health costs in the developed countries. Building an effective

disease management strategy requires analysis of large amount of data, early

detection of the disease, assessment of the severity and early prediction of

adverse events. This will inhibit the progression of the disease, will improve the

quality of life of the patients and will reduce the associated medical costs.

Toward this direction machine learning techniques have been employed. The

aim of this paper is to present the state-of-the-art of the machine learning

methodologies applied for the assessment of heart failure. More specifically,

models predicting the presence, estimating the subtype, assessing the severity of

heart failure and predicting the presence of adverse events, such as

destabilizations, re-hospitalizations, and mortality are presented.

**1.4 Social Or Business Impact.**

* **Social Impact:** Customers can make more informed decisions about their travel plans and compare prices and services more easily. Increased Accessibility and improved customer experience
* **Business Model/Impact**: Competitive Advantage, Innovation and Improved Business strategy can be achieved by analysing

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
|  | Problem Statement (Problem to be solved) | Improving the accuracy of the model |
|  | Idea / Solution description | Medical costs are one of the most common recurring expenses in a person’s life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. The estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. |
|  | Novelty / Uniqueness | New accuracy from by this model |
|  | Social Impact / Customer Satisfaction | All customer used to medicine with low cost.customer easy to facilities of medicine. |
|  | Business Model (Revenue Model) | Generate revenue by selling our application |
|  | Scalability of the Solution | 85% - 95% |

