ANALYTICS FOR HOSPITALS' HEALTH-CARE DATA

1. INTRODUCTION

1.1 PROJECT OVERVIEW:

Today’s healthcare industries are moving from volume-based business into value-based business, which requires an overwork from doctors and nurses to be more productive and efficient. This will improve healthcare practice, changing individual life style and driving them into longer life, prevent diseases, illnesses and infections.

This paper aims to proof that healthcare data analytics techniques are not efficient enough and suitable anymore these days in order to manage big data issue and improve healthcare data analytics due to the rapid growth and evolution of technology. Moreover, it’s also aims to promise professionals of a better quality of medical results, as well as reduce time needed to analyze healthcare data by keeping systems up to-date and sorting medical data in a logical structure along with accessing and retrieving patient’s historical data fast and smoothly. Stakeholder 2 Doctors and nurses.

Professionals (doctors and medical stuff) will be benefited for sure and they will use proposed technique, since it will reduce their time and efforts, therefore this technique focused also on adding a true assistance to their job to run smoothly as its really stressful and valuable, so they need a technique that facilitates their job and save their efforts such as: retrieving historical and old data quickly, sorting data in a logical structure way and keep it up to date, which will help them to discover hidden patterns and extract information effortlessly and efficiently. Moreover, professionals will be satisfied as they will touch that this technique will

provide and additional source of knowledge to make a better decision (support decision making process) which is really needed to improve individuals' healthcare and increase their reputation.

1.2 PURPOSE:

Data analytics in health care is vital. It helps health care organizations to evaluate and develop practitioners, detect anomalies in scans and predict outbreaks in illness, per the Harvard Business School. Data analytics can also lower costs for health care organizations and boost business intelligence.

Professionals (doctors and medical stuff) will be benefited for sure and they will use proposed technique, since it will reduce their time and efforts, therefore this technique focused also on adding a true assistance to their job to run smoothly as its really stressful and valuable, so they need a technique that facilitates their job and save their efforts such as: retrieving historical and old data quickly, sorting data in a logical structure way and keep it up to date, which will help them to discover hidden patterns and extract information effortlessly and efficiently. Moreover, professionals will be satisfied as they will touch that this technique will provide and additional source of knowledge to make a better decision (support decision making process) which is really needed to improve individuals' healthcare and increase their reputation.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM:

TITLE: Data-Driven Methods for Typical Treatment Pattern Mining

DESCRIPTION:

A huge volume of digitized clinical data is generated and accumulated rapidly since the widespread adoption of Electronic Medical Records (EMRs). This paper discusses the research background - big data analytics in healthcare, the research framework of big data analytics in healthcare, analysis of medical process, and treatment pattern mining. Then the challenges for data-driven typical treatment pattern mining are highlighted, including similarity measure between treatment records, typical treatment pattern extraction, evaluation and recommendation, when considering medical information in EMRs. Furthermore, three categories of typical treatment patterns are mined from doctor order content, duration, and sequence view respectively, which can provide a data-driven guideline to achieve the “5R” goal for rational drug use and clinical pathways.

TITLE: A Systematic Mapping Study

DESCRIPTION:

The current study performs a systematic literature review (SLR) to

synthesise prior research on the applicability of big data analytics (BDA) in healthcare. The SLR examines the outcomes of 41 studies, and presents them in a comprehensive framework. The findings from this study suggest that applications of BDA in healthcare can be observed from five perspectives, namely, health awareness among the general public, interactions among stakeholders in the healthcare ecosystem, hospital management practices, treatment of specific medical conditions, and technology in healthcare service delivery. This SLR recommends actionable future research agendas for scholars and valuable implications for theory and practice.

TITLE: Exploring big data analytics in health care.

DESCRIPTION:

Cost optimization is one of the major issues in health care as it has become very difficult in fetching patient's information across huge data bases. Here, various data mining techniques such as SVM, Decision Trees etc. have been discussed in order to address various healthcare issues. Later on Big Data Analytics tools were addressed on top of data mining techniques in health care sector, as the health care industry is one of the leading sectors where huge revenue will be generated across globe as the numbers of patients are increasing drastically with the population. In future Machine learning with Big Data has lot of scope in healthcare as so many new diseases are coming into lie light across the world.

TITLE: Big Data Analytics Framework for Opinion Mining of Patient Health

Care Experience

DESCRIPTION:

Preciously administration might be able to acknowledge the crucial decision-making process where the new investigations would be accounted for different research avenues.

TITLE: A systematic review of health care big data.

DESCRIPTION:

Analyzing different perspectives about the concept of big data in healthcare Exploring the origins of healthcare big data Identifying tools and techniques for healthcare big data Analytics.

2.2 REFERENCES:

1. Abbott, PA & Coenen, A 2008, 'Globalization and advances in information and communication technologies: The impact on nursing and health', Nursing Outlook, vol. 56, no. 5, pp 238- 246.
2. Al-Azzawi, H. 2014. “Caradigm healthcare analytics.” http://www.caradigm.com/media/68911/CaradigmWP-Healthcare-Analytics-Jan-2014-US-EN.PDF Retrieved 09 August, 2015.
3. Bakshi, K. 2014, ‘Considerations for big data: architecture and approaches’, Aerospace conference, IEEE, pp. 1-7.
4. Bertsimas, D; Bjarnadóttir, M; Kryder, J; Pandey, R; Vempala, S; Wang, G.

(2008), ‘Algorithmic prediction of health-care costs’, Operations Research, vol. 56, no. 6, pp. 1382-1557.

5. Beyer, K., Ercegovac, V., Gemulla, R., Balmin, A., Eltabakh, M., Ozcan, F. and Shekita, E. 2011. “Jaql: A Scripting Language for Large Scale Semi-Structured Data Analysis.” http://web.cs.wpi.edu/~meltabakh/Publications/Jaql-PVLDB2011.pdf Retrieved 09 August, 2015.

2.3 PROBLEM STATEMENT DEFINITION:

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS:

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to helps teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it.

The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

3.2 IDEATION & BRAINSTORMING:

3.3 PROPOSED SOLUTION:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To precisely determine the patient's length of Hospital stay
2.	Idea / Solution description	Gather patient data from medical facilities Examining the specifies of the patient's documents Create a Cognos Analytics dashboard to show Patient data

3.	Novelty / Uniqueness	Accurate understanding of the factors associating with the LOS and progressive improvements in processing and monitoring may allowmore efficient management of the LOS inpatients.
4.	Social Impact/ Customer Satisfaction	A shorter LOS reduces the risk of acquiring staph infections and other healthcare-relatedconditions, frees up vital bed spaces,and cuts overall medical expences.
5.	Business Model (Revenue Model)	The length of stay LOS is an important indicatorof the efficiency of hospital management.Reduction in the number of inpatients days results in decreased risk of infection and medication side effects,improvement in the quality of treatment, and increased hospital profit withmore efficient bed management.
6.	Scalability of the Solution	Remote patient monitoring systems enablingeffective distance treatment. Patient portals that allow people to better manage their health themselves.

3.4 PROBLEM SOLUTION FIT:

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration throughForm Registration throughGmail
FR-2	User Confirmation	Confirmation viaEmail Confirmation via Message
FR-3	User Login	Login throughGmail
FR-4	Uploading the data	Uploading intothe IBM Cognos analytics
FR-5	Preparing the data	Data module creation
FR-6	Data module creation	Performing predictive analysis
FR-7	Visualizing the data	Creating dashboards

4.2 Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

FR	Non-Functional	Description

No.	Requirement	
NFR-1	**Usability**	This Dashboards are designed to offer a Comprehensive overview of patient's LOS, and do so through the use of data visualization tools like charts and graphs.
NFR-2	**Security**	The Dashboard helps to indicate the current threatlevel to the Hospitals; an indication of events and incidents that have occurred; a record of authentication errors; unauthorized access
NFR-3	**Reliability**	This dashboard will be consistent and reliable to the users and helps the user to use in effective, efficientand reliable manner
NFR-4	**Performance**	This dashboard can scan the backend users and analyzing the frequency in which they visit the dashboard helpsunderstand how useful and helpful the datadisplayed is for tasks

NFR-5	**Availability**	The dashboard can available to meet user’s demand in timely mannerand it is also helps to provide necessary information to the user’sdataset
NFR-6	**Scalability**	The layersused in thedashboard are a hosted feature layer,feature layer view,or hosted tilelayer.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data centers and leaves the system, what changes the information, and where data is stored.

5.2 SOLUTION & TECHNICAL ARCHITECTURE:

SOLUTION ARCHITECTURE:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

1. Find the best tech solution to solve existing business problems.
2. Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
3. Define features, development phases, and solution requirements.
4. Provide specifications according to which the solution is defined, managed,

and delivered.

TECHNICAL ARCHITECTURE:

5.3 USER STORIES:

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION:

Use the below templateto create productbacklog and sprint schedule

Sprint	**Functional Requirement(Epic)**	**User Story Number**	**User Story/ Task**	**Story Points**	**Priority**	**Team Members**
Sprint-1	Registration	USN-1	As a user,I can register for the application byentering my email, password, and confirming my password.	10	High	DEEPAN GOPIN TH DEEPAN CHAKKARAVARITHY

Sprint-1	Data uploading	USN - 2	As a user, I will be uploading my data intothe cognos analytics	10	High	DEEPAN FARMAN

-

-

Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
np.set_printoptions(suppress=True)
import warnings
warnings.filterwarnings('ignore')
```

%importing dataset

```
train = pd.read_csv('train.csv')
test = pd.read_csv('test.csv')
train.head()
train.info()
train.Stay.unique()
test.isnull().sum().sort_values(ascending = False)
```

```
# Dimension of train dataset
train.shape
# Dimension of test dataset
test.shape
```

```
# Number of distinct observations in train dataset
for i in train.columns:
    print(i, ':', train[i].nunique())
```

-

-

-

Result analytics

Data preparation

```
#Replacing NA values in Bed Grade Column for both Train and Test datssets
train['Bed Grade'].fillna(train['Bed Grade'].mode()[0], inplace = True)
test['Bed Grade'].fillna(test['Bed Grade'].mode()[0], inplace = True)
```

%train the dataset

```
train['City_Code_Patient'].fillna(train['City_Code_Patient'].mode()[0],
inplace = True)
test['City_Code_Patient'].fillna(test['City_Code_Patient'].mode()[0],
inplace = True)
```

%model design

```
model = Sequential()
model.add(Dense(64, activation='relu', input_shape = (254750, 20)))
model.add(Dense(128, activation='relu'))
model.add(Dense(256, activation='relu'))
model.add(Dense(512, activation='relu'))
model.add(Dense(512, activation='relu'))
model.add(Dense(11, activation='softmax'))
```

```
model.compile(optimizer= 'SGD',
              loss='categorical_crossentropy',
              metrics=['accuracy'])
# Genrating tensorboard
!tensorboard --logdir logs_keras
```

%pretrain the dataset

```
# Genrating tensorboard
```

```
!tensorboard --logdir logs_keras
```

Case id

RESULT AND SCREN SHOT

`

Analytical result

Patient analytical results view

CONCLUSION

A COVID-19 Analytics Agenda HDMS recognized the opportunity to leverage our clinical and analytic experts to offer content to support companies that extend even beyond our own client base. We created an Analytics Agenda and shared supporting resources to best leverage health analytics to support your organization A massive amount of data in various forms needs to be handled for any healthcare application and data type, data size and other features are significant in data handling. With the growth of big data in biomedical and healthcare communities, accurate analysis of medical data has the benefits of early disease detection, improved patient care, and effective community services. Because of its significance, there is a need to develop efficient and better-performing algorithms, techniques, and tools to analyze multi-modal medical big data from the gene level to the clinical level. However, the traditional algorithms used are not capable of analyzing such complex data. Machine learning algorithms with a good fit for these kinds of data and analytics are required. Also, different regions exhibit unique characteristics of certain regional diseases, and this may weaken the

prediction of disease outbreaks.

ADVANTAGES AND DISADVANTAGE

1. Its Can Provide The Facility Of Covid Patient Details
2. Organization-Wide Assessment Of Current Conditions
3. We Can Provide The Analytical Results
4. Its Only Analytical Result
5. Its not user friendly

VIDEO LINK:

https://drive.google.com/file/d/1icz2jsPbed8d6IJrMv-pZyHk563FxzNv/view?usp=share_link

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-37880-1660361404