Predicting Post-Procedural Complications on MIMC-III Dataset

****

Master of Science in Data Science

Faisal Maqbool

MSDS17027

Session: 2017 – 2019

DEPARTMENT OF COMPUTER SCIENCE

INFORMATION TECHNOLOGY UNIVERSITY

LAHORE, PAKISTAN

****

Predicting Post Procedural Complications on MIMIC-III Data

A thesis submitted in partial fulfillment of the requirements for the

Degree of Master of Science in

Data Science

Faisal Maqbool

Dr. Saeed Ul Hassan

Committee Member Name

Committee Member Name

**Declaration**

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. I also declare that this work is the result of my own investigations, except where identified by references and free from plagiarism of the work of others.

Signature: ………….

Student Name

Date: ………………..

The undersigned hereby certify that they have read and recommend the thesis entitled “…………………………..” by ………………….. for the degree of Master of Science in Data Science.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor Name (ITU), Thesis Advisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Committee Member Name (ITU), Thesis Committee Member

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Committee Member Name, Thesis Committee Member

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Chairperson Name, Chairperson of the Department

**Acknowledgment**

First and foremost, I would like to express sincere gratitude to my advisor Dr. Saeed Ul Hassan for being the supportive advisor I could have asked for. His precious insights, guidance and support throughout the research journey has not only helped me become a better student but also examine data for research and contribute to previously done research. Though he is very busy, he always made me feel like a priority. He took time to review my findings, answer questions, and give thoughtful feedback for which I am extremely grateful. This thesis would not have been possible without them.

I would like to thank Dr. Mohsin Ali for his help, advice and support throughout my work and for agreeing to serve on my thesis committee.

Moreover, I would like to thank my family especially my parents and my friends for me letting my follow my dreams and for their encouragement which has been the vital part of the research.

**Table of Contents**

Acknowledgments

Contents

**No table of contents entries found.**

**List of Tables**

**List of Figures**

**List of Abbreviations**

MIMIC:

PHI:

ICU:

ETL:

ICD:

**Abstract**

The advancements in bioinformatics and health care sector has inspired researchers to develop systems that can mimic work of doctors because quantifying patient health and predicting outcomes is an important problem in critical care research. But, predicting outcomes for critically ill patients admitted in intensive care units requires specific characteristics of clinical data: quality, volume, access and dimensionality. In this Master Thesis, an analysis of the data from critical patients was carried out in order to study the influence of several factors for predicting the post-procedural complications and seeking if those complications can lead to mortality of patients. To derive insights for that, we used well-known clinical dataset named Medical Information Mart for Intensive Care III (MIMIC-III).

**.. Results Pending**

1. **Chapter: 1**

# Introduction

With the initiation of digital technology, advanced techniques are increasingly making it possible to utilize big data to more precisely risk evaluation and predict how an individual patient will behave based on a given diagnose or procedure. Intensive care unit (ICU) is a ward in hospital, where critically ill patients are admitted requires accurate predictors that can help doctors with the assessment of severity of illness.

In this thesis we investigate the different methodologies for extracting, transforming and loading (ETL) processes, which obtain data from original source to perform informative analysis and features extraction to aid model to predict post-procedural (diagnoses & procedure) complications of critically ill patients and investigating those complications if those can lead to mortality of patient or not. The methods use demographics, data from different hospital system, lab events, diagnoses, notes and other engineered information regarding each patient. The database used for the study is Medical Information Mart for Intensive Care **MIMIC-III** [1] which comes from health service with anonymized data for protecting health information (PHI).

Other researches about MIMIC-III data is also presented to motivate our problem, establish understanding of dataset, key findings and recommendations for future investigations. The question of predicting post-procedural complications from data science perspective and critical health perspective is not only important for doctors, administrators but also for the patient as well. For administrators this would help managing patients and required resources. Avoiding predicted complications can further be avoided if such information is known during the stay of patient at ICU.

# ICD and Complications

The International Classification of Diseases (ICD) is the foundation for the identification of health trends and statistics globally, and the international standard for reporting diseases and health conditions. It is the diagnostic classification standard for all clinical and research purposes. Under revision of ICD9 codes, the code 996 defines complications particular to certain specified procedures and diagnoses. Most complications are caused due to cardiac, vascular or other used devices and some of them relates to reaction caused due to a procedure performed. In our work, we are focusing on such complications and investigating if those can lead to the mortality of patient.

# Why COMPLICATIONS?

Extended diagnoses and patients stay at the hospital is associated with not only the health of patient, cost, increased number of deaths but also increased number of readmissions. Each of these parameters defines the hospital performance. So, our focus is to produce insights that can complement these parameters ranging from cost to patient health.

# APPROACH Followed



# Description of the Content

The thesis is organized as follows: In chapter **1** we introduce the problem, explains purpose of our work, methodologies followed and detail of contents. In Chapter **2**, we define the data source MIMIC-III in detail. In Chapter **3**, we explained the ETL process to obtain data and process it effectively for posterior use. In Chapter **4**, we explain the preprocessing and analysis of data which is used of feature engineering and model selection process. In Chapter **5**, results and etc………

References

[1] Johnson, Alistair EW, Tom J. Pollard, Lu Shen, H. Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. "MIMIC-III, a freely accessible critical care database." *Scientific data* 3 (2016): 160035.

**THESIS GUIDELINES**

**Thesis Cover:**

For thesis, the following color scheme will be used:

* Bachelor Thesis: Blue
* Master Thesis: Dark Maroon
* Ph.D. Thesis: Black

**Format for Spine:**

* For spine of the thesis, the student will put ***Degree title i.e. MSCS***, ***Thesis title*** and mention ***Session***.

**Font Size and Style:**

* The ***Headings*** in the thesis will be in ***Times New Roman*** with font size ***16.***
* The **Body** of the thesis will be in ***Times New Roman*** with font size ***12.***
* The line spacing will be *1.5.*

**Referencing:**

* The student should follow ***IEEE*** referencing style.

**Appendices:**

* Provide appendices at the end of the thesis labeling as appendix A, appendix B and so on.