



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**SCHOOL OF COMPUTING**  
Faculty of Engineering

Project Proposal Form MCST1043  
Sem: 2 Session: 2024/25

## SECTION A: Project Information.

---

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCST1043)**

Student Name: Guo Yachao

Metric Number: MCS241039

Student Email & Phone: guoyachao@ggraduate.utm.my

Project Title: Medical data analysis and prediction of sepsis patients

Supervisor 1: \_\_\_\_\_

Supervisor 2 / Industry

Advisor(if any): \_\_\_\_\_

## SECTION B: Project Proposal

---

### Introduction:

Sepsis is a systemic inflammatory syndrome triggered by the body's abnormal immune response to infection, leading to organ dysfunction and even failure. It is a life-threatening emergency. Sepsis is not a single disease but a complex pathological process triggered by infections (such as bacterial, viral, fungal, etc.), representing a major challenge in critical care medicine.

### Problem Background:

Sepsis is a life-threatening clinical emergency caused by an abnormal overreaction of the patient's body to infection, requiring urgent treatment to be initiated within one hour ("golden hour" principle) to reduce the risk of multi-organ failure and death. At its core is systemic inflammatory response syndrome (SIRS), which results from immune system imbalance and can rapidly progress to septic shock (Septic Shock) or irreversible organ damage.

### Problem Statement:

The urgency and time sensitivity of sepsis are indeed central challenges in clinical decision-making. Given that the condition progresses rapidly (potentially worsening to septic shock or multi-organ failure within hours), doctors often cannot afford to wait for the results of traditional laboratory tests (such as blood cultures, cytokine analysis) that take several hours or even days. This contradiction has fueled an urgent need for rapid diagnostic techniques and real-time

predictive tools.

**Aim of the Project:**

The purpose of the project is to analyze patients' electronic health records, machine learning and artificial intelligence to provide effective tools to speed up the prediction of an upcoming septic shock, the prediction of the sequential organ failure, and the prediction of survival or mortality of the patient.

**Objectives of the Project:**

- 1.To analyze a dataset of electronic health records of 364 patients collected between 2014 and 2016.
- 2.To employ several machine learning methods to predict it.
3. To use a data mining approach to identify the most important dataset features in relation to targets, and compared these results with the results achieved through a standard biostatistics approach.

**Scopes of the Project:**

In the present study, an electronic health record dataset of patients with cardiovascular heart disease was analyzed: each patient had 29 clinical features, including binary for survival, binary for septic shock and values for a sequential organ failure assessment (SOFA) scores.

**Expected Contribution of the Project:**

The outcomes of this project will contribute to effective methodologies to predict septic shock, SOFA score, and survival of patients diagnoses with sepsis, from their electronic health records data.And regarding clinical feature ranking, these results showed that Random Forests feature selection identified several unexpected symptoms and clinical components as relevant for septic shock,SOFA score, and survival. These discoveries can help doctors and physicians in understanding and predicting septic shock.

**Project Requirements:**

Software: free R programming language,construction of the dataset,common machine learning packages(randomForest, caret,e1071, keras, ROSE, DMwR,mltools, DescTools)

Hardware: Laptop or desktop computer with sufficient processing power and storage capacity

Technology/Technique/Methodology/Algorithm: Matthews correlation coefficient (MCC),MLP( Multilayer perceptron),MS( Model selection),PCC(Pearson correlation coefficient), $R^2$  ( Coefficient of determination),

**Type of Project (Focusing on Data Science):**

☐ Data Preparation and Modeling

☒ Data Analysis and Visualization

☐ Business Intelligence and Analytics

☒ Machine Learning and Prediction

☐ Data Science Application in Business Domain

**Status of Project:**

☒ New

☐ Continued

If continued, what is the previous title?

**SECTION C: Declaration**

**I declare that this project is proposed by:**

☒ Myself

☐ Supervisor/Industry Advisor ( )

Student Name: Guo Yahao

Signature

Date

**SECTION D: Supervisor Acknowledgement**

The Supervisor(s) shall complete this section.

**I/We agree to become the supervisor(s) for this student under aforesaid proposed title.**

Name of Supervisor 1:

Signature

Date

Name of Supervisor 2 (if any):

Signature

Date

**SECTION E: Evaluation Panel Approval**

The Evaluator(s) shall complete this section.

**Result:**

☐ FULL APPROVAL ☐ CONDITIONAL APPROVAL (Major)\*

☐ CONDITIONAL APPROVAL (Minor) ☐ FAIL\*

\* Student has to submit new proposal form considering the evaluators' comments.

**Comments:**

Name of Evaluator 1:

Signature

.....  
Date

Name of Evaluator 2:

Signature

.....  
Date