10/24/23

C964: Computer Science Capstone

Madison Maguire

011029252

## Table of contents

[Table of contents 1](#_Toc149052284)

[Part A: Letter of Transmittal 2](#_Toc149052285)

[Part B: Project Proposal Plan 4](#_Toc149052286)

[Project Summary 4](#_Toc149052287)

[Organizational Need 4](#_Toc149052288)

[Context and Background 4](#_Toc149052289)

[Deliverables 5](#_Toc149052290)

[Client Benefit 5](#_Toc149052291)

[Data Summary 5](#_Toc149052292)

[Implementation 6](#_Toc149052293)

[Timeline 7](#_Toc149052294)

[Evaluation Plan 8](#_Toc149052295)

[Resources and Costs 8](#_Toc149052296)

[Part C: Application 9](#_Toc149052297)

[Part D: Post-implementation Report 10](#_Toc149052298)

[Solution Summary 10](#_Toc149052299)

[Data Summary 10](#_Toc149052300)

[Machine Learning 11](#_Toc149052301)

[Validation 12](#_Toc149052302)

[Visualizations 12](#_Toc149052303)

[User Guide 14](#_Toc149052304)

[Reference Page 16](#_Toc149052305)

# Part A: Letter of Transmittal

October 24th 2023

Dr. Robert Kelso

Sacred Heart Hospital

12629 Riverside Drive

North Hollywood, CA 91607

Dear Dr. Kelso,

Thank you for speaking with me on the phone today. It was a pleasure to make your acquaintance. I am following up on our discussion regarding sepsis patient interventions. As you know, sepsis is a severe and potentially fatal disease, so it is crucial to identify patients at risk of developing it. As your hospital is initiating a pilot program to combat sepsis and enhance patient outcomes, my company can help you achieve your goals.

We propose creating an application that utilizes machine learning to determine patients' sepsis risk. Machine learning is a cutting-edge tool in data science that enables us to use algorithms to identify patterns in large amounts of data. We can then leverage these patterns to predict outcomes of current or future patients based on their biographical data, such as their sex or age.

This tool could be incredibly beneficial to your staff. When a patient is admitted, a healthcare worker can enter their information into our tool, and it will predict the likelihood of the patient surviving. Hospital staff can then target patients with a low survival probability for specialized interventions to improve their outcomes.

If we proceed with the project, it will cost approximately $67,600 and take roughly twelve weeks to complete. The data we will use to train the model is freely available, and there are no ethical concerns since the data used is entirely anonymous.

We have been developing software for hospitals for 20 years and are passionate about improving patient outcomes. We have recently completed similar projects for Seattle Grace Hospital, County General Hospital, and Princeton-Plainsboro Teaching Hospital. We would love to work with your hospital to achieve its patient-care objectives.

I look forward to hearing from you.

Sincerely,

Madison Maguire

Madison Maguire

# Part B: Project Proposal Plan

# Project Summary

This document describes a proposed machine learning solution for identifying at-risk sepsis patients. Hospital staff will then use this information to implement targeted interventions that improve patient outcomes.

## Organizational Need

Sepsis is a severe infection that can harm vital organs and even result in death. To combat this illness, Sacred Heart Hospital is initiating a trial program that employs targeted interventions for patients who are at risk. As a part of this program, the hospital staff must identify such patients to receive specialized care. The hospital requires a quick way to identify these patients at check-in.

## Context and Background

Sacred Heart Hospital was established in North Hollywood in 1952, and it is currently a 160-bed facility with approximately 1,000 employees. The hospital is renowned for its unconventional approach to medical issues, and it is particularly famous for introducing the Cox-Dorian program, which was designed to measure the impact of humor on patients during their post-surgery recovery.

## Deliverables

The follow items are proposed deliverables for this project:

* Dataset used to train the model
* Model source code
* Working implementation of the model on the network
* Application documentation

## Client Benefit

A significant benefit that Sacred Heart Hospital will receive is the ability of any staff member to calculate the chances of a patient surviving a sepsis episode. The application is user-friendly and requires minimal training, saving time and money. Once the staff member enters the patient's details into the system, the application calculates and displays the survivability percentage. This will enable the medical staff to make informed decisions about patient care, ultimately leading to improved patient outcomes.

## Data Summary

The model will be trained using a dataset obtained from Kaggle, an online data science platform. The dataset consists of 110,204 records of Norwegian sepsis admissions, which include information about each patient's age, sex, previous sepsis episodes, and the patient's outcome.

This data will be made available on the hospital's internal network, giving access to authorized personnel. The data is stored in CSV format, easily accessible, and editable with various applications. The hospital's sepsis patient data can be incorporated into the CSV file to integrate the information into the model.

The project requires collecting data from a large variety of sepsis patients. The data must include biographical patient information but must be anonymized to protect patient privacy. The selected dataset meets these requirements.

As the data is highly anonymized, there is no breach of patient privacy and, therefore, no ethical concerns regarding its use.

## Implementation

The project will follow the CRISP-DM methodology:

1. Business Understanding – Sacred Heart’s patient and business needs will be examined to ensure the solution meets them. We will look at their current situation and any long-term objectives.
2. Data Understanding – The dataset will be examined so we understand that data we are using with the model.
3. Data Preparation – Any preparation work that needs to be done on the data will be implemented. Any obvious outliers can be discarded.
4. Modeling - The modeling algorithm will be implemented on the training dataset. Different algorithms will be tried to compare against each other.
5. Evaluation - The modeling results will be evaluated to determine which tested algorithm produced the best results. Depending on the results, we may iterate further or proceed to the next stage.
6. Deployment - When satisfied with the evaluation results, the model and will be deployed by making the application accessible to staff on the client’s network. We will then review the project to ensure it meets the client’s objectives. Documentation will be delivered, and Sacred Heart will take over the maintenance of the project.

## Timeline

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone or deliverable | Duration  (hours or days) | Projected start date | Anticipated end date |
| Client examines proposal and signs off. | 7 days | 11/04/23 | 11/11/23 |
| Data for dataset collected and documented. | 7 days | 11/12/23 | 11/19/23 |
| Modeling algorithms implemented against dataset. Results documented. | 14 days | 11/20/23 | 12/04/23 |
| Results from modeling algorithm test interpreted, and the best implemented. | 7 days | 12/05/23 | 12/12/23 |
| Implementation presented to client for final sign off. | 7 days | 12/13/23 | 12/20/23 |
| Application developed and integrated with model. | 28 days | 12/21/23 | 01/18/24 |
| Project wind down. Any issues are resolved. Documentation handed over to client. | 7 days | 01/19/24 | 01/26/24 |
| Final sign off. | 7 days | 01/27/24 | 02/03/24 |

## Evaluation Plan

We will create three datasets based on the original data. The training dataset will be utilized to train the model, while the evaluation dataset will be used to evaluate the model's performance during the training phase. Finally, we will use the testing dataset to assess the accuracy of the model.

## Resources and Costs

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| Hosting equipment | Server equipment utilized for hosting the dataset and performing analysis. | $10,000 |
| Machine Learning engineers | Creates the dataset and implements the algorithms. | 800 hours @ $60 per hour  = $48,000 |
| App developers | Creates the application and integrates the model | 160 hours @ $40 per hour  = $6,400 |
| Technical writers | Creates the documentation | 80 hours @ $40 per hour  = $3,200 |
|  | **Total** | $67,600 |

# Part C: Application

The code is hosted on GitHub:

<https://github.com/MsBadison/WGU-Capstone/blob/main/Sepsis.ipynb>

The application is accessed through Google Collab:

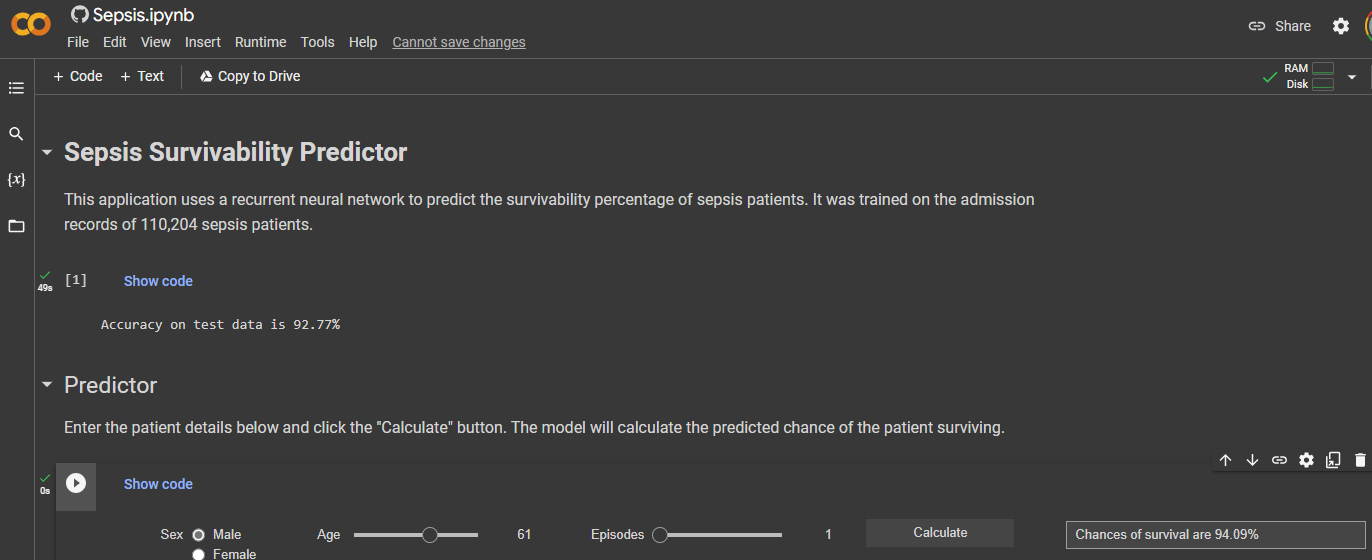
<https://colab.research.google.com/github/MsBadison/WGU-Capstone/blob/main/Sepsis.ipynb>

The dataset is hosted online and can be downloaded from:

<https://drive.google.com/uc?export=download&id=10iVApqE-lAabAsqspOqiEgHx6I18LVIC>

The following application files will also be submitted:

1. Sepsis.ipynb
2. s41598-020-73558-3\_sepsis\_survival\_primary\_cohort.csv



# Part D: Post-implementation Report

# Solution Summary

Sacred Heart Hospital has set a goal to reduce the number of deaths caused by sepsis. The hospital plans to implement targeted interventions to lower the mortality rate of high-risk patients. To achieve this, the hospital needs an efficient way to identify at-risk patients.

To address this issue, we have analyzed the outcomes of past patients with sepsis to identify any patterns in the data that may help predict mortality rates. Our application uses machine learning to analyze the outcomes of 110,204 patients with sepsis. Training a model on this data and creating a user interface allows the hospital to make predictions based on past outcomes and develop effective interventions to benefit their **sepsis** patients.

## Data Summary

The application uses the *Sepsis Survival Minimal Clinical Records* dataset from Kaggle to train its model. This dataset contains the records of 110,204 Norwegian patients admitted to the hospital with sepsis. It provides information on each patient's age, sex, number of prior sepsis episodes, and eventual patient outcome.

The CSV file was hosted on the network and read using Pandas' *read\_csv* method to import the dataset into the application. The dataset's simple format makes it easy to work with, and it is centrally accessible, allowing the hospital to continue adding patient records to train the model further.

## Machine Learning

Our application utilizes a sequential recurrent neural network to predict the survival rate of sepsis patients. The prediction is based on the analysis of the outcomes of thousands of prior patients, which enables us to create a highly accurate model to predict the survivability of current and future patients. To develop this model, we utilized TensorFlow, a widely used machine learning platform that provides powerful data tools. We also employed the Keras library to simplify TensorFlow usage and reduce the development and maintenance time. The application has been written in Python, hosted on GitHub, and can be accessed through Google Collab.

Our model development began with finding a suitable dataset. We chose a dataset from Kaggle that detailed the outcomes of 110,204 sepsis patients. The dataset also contained their age, sex, and the number of previous sepsis episodes. This provided us with enough data to train the model, which allowed us to enter the same patient details to predict the survivability of current and future patients.

We split the dataset into training, evaluation, and testing data. We used the training and evaluation data to train the model and the testing data to check its accuracy after training. This approach allowed us to make changes to the model and study how those changes impacted its accuracy.

We used a sequential model, which was the best choice for our data as it provided good accuracy. ReLu activation was chosen for the input and hidden layers due to its computational efficiency, allowing us to optimize the available resources while maintaining good performance. Also, we selected a Sigmoid function for the output to output a number between 0 and 1, which can then be multiplied by 100 to obtain a survivability percentage.

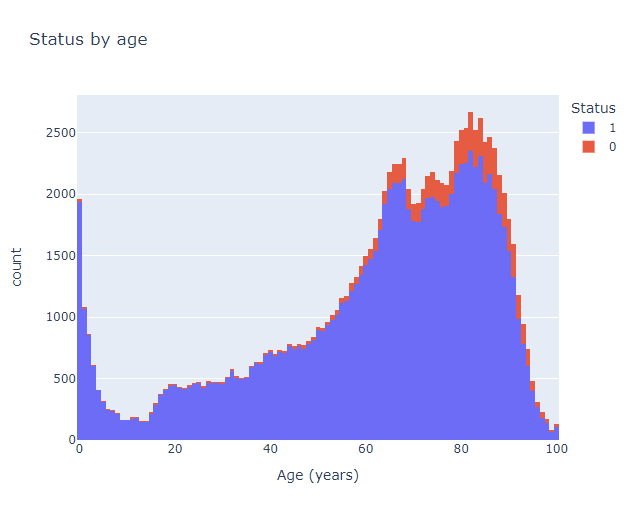
## Validation

We used Keras's evaluation method and our test data to evaluate our model. This method returns a value between 0 and 1, representing the model's accuracy. We assigned the value to a variable, which was then multiplied by 100 to obtain a percentage. We printed the percentage value with an accuracy of 2 decimal places. After evaluating our model using the test data, we achieved an accuracy of 92.77%.

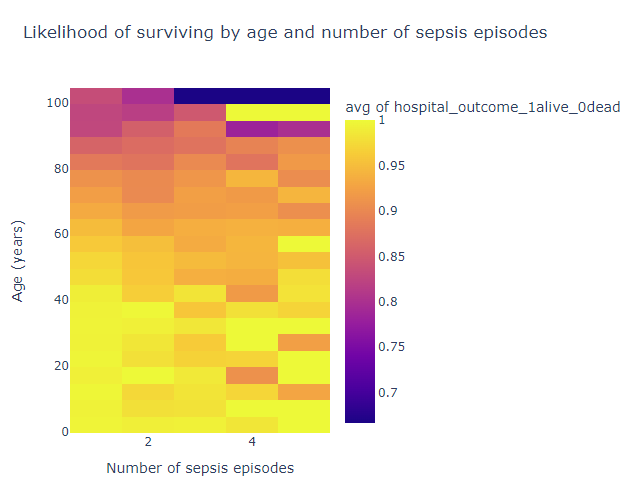
## Visualizations

As part of the project, three visualizations were selected and included. The first two are related to the data used to train the model, while the third is related to the model itself.

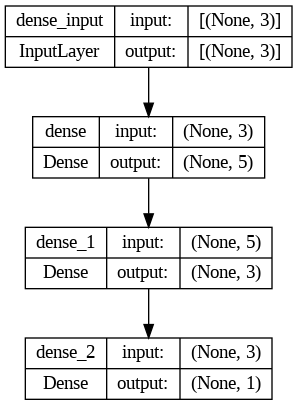
The first visualization is a stacked histogram that displays the number of patients categorized by age and survival status. Patients who survived are represented in blue, while those who died are represented in red. The graph clearly shows that the risk of sepsis infection increases with age, and so does the likelihood of it being fatal. There is an interesting spike at the beginning of the graph, which we attribute to newborns.



The second visualization is a heatmap that displays the average survival rate according to age and the number of sepsis episodes. Brightly colored groups indicate higher survival rates, while darkly colored groups indicate lower survival rates. Based on the heatmap, young individuals with few sepsis episodes have a higher survival rate than older individuals with multiple sepsis episodes.



The final visualization displays the model utilized in the project, illustrating the layers used and their input and output counts. The bottom layer contains a Sigmoid activation function with one output.



## User Guide

The application is hosted on GitHub and accessed through Google Collab. To start it, please follow these steps:

1. Open a web browser. (tested on Microsoft Edge and Google Chrome)
2. Go to <https://colab.research.google.com/github/MsBadison/WGU-Capstone/blob/main/Sepsis.ipynb>
3. Click the *Sign in* button and log into your Google account.
4. Select *Run all* from the *Runtime* menu at the top of the screen.
5. If you receive a warning, click *Run anyway*.
6. Wait for the application to finish loading.
7. Once it has loaded, the application will display the model's accuracy.
8. To use the predictor, enter the patient information using the radio buttons and sliders and click *Calculate*. The application will display the calculated survivability percentage in the box on the right.

# Reference Page

No references were used when writing this report.