# Final Project Proposal

**Problem Statement:** Sepsis is a life-threatening medical condition that is caused by the spread of infection in one's body, resulting in multiple organ failure or even death in many cases. Since it is a serious issue which can cause death, prediction of survival in such cases is among the top priorities of medical community today. Though we have a lot of sophisticated medical laboratories that can provide useful information about the patient, yet these things take their time, and might not be available immediately, not allowing medical practitioners to detect an urgent life threat and treat accordingly. Machine learning can be applied in this setting to get faster results. It can help build a model which can take in variables that can be easily retrieved, and if dataset available is large enough, can give accurate predictions of survival based on a handful of related inputs only.

**Dataset:** The dataset comprises of $19{,}052$ hospitalized subjects between 2011 and 2012 in Norway who were diagnosed with infections, systemic inflammatory response syndrome (SIRS) or septic shock. For every admission, the dataset contains information about patient's age, sex, septic episode number and hospitalization outcome (dead or alive).

| Feature | Explanation | Measurement | Range |
|---|---|---|---|
| Age | Age of patient | Integer | [0,100] |
| Episode Number | Number of septic episodes experienced | Integer | [1,5] |
| Sex | 0: male, 1: female | Boolean | 0,1 |
| Survival | 0: dead, 1: alive | Boolean | 0,1 |

Some rows from the dataset:

| age_years | sex_0male_1female | episode_number | hospital_outcome_1alive_0dead |
|---|---|---|---|
| 7 | 1 | 1 | 1 |
| 17 | 0 | 2 | 1 |
| 70 | 0 | 1 | 1 |
| 76 | 0 | 1 | 1 |
| 8 | 0 | 1 | 1 |

The dataset can be found at the following link:
https://archive.ics.uci.edu/ml/machine-learning-databases/00628/

**Variables:** The main aim of this project is to predict the hospital outcome (alive or dead) using relevant machine learning models. The prediction is based on three easily available features of hospitalized patients, i.e., sex, age and septic episode number (which can be easily found from the patient's medical history).