# HW1 Report
# Daniel Shemesh 211388251 & Shir Turgeman 315047324

## Executive summary

The goal of this project was to predict whether a patient in intensive care suffers from sepsis approximately 6 hours before being identified as suffering from sepsis, based on clinical data about their medical condition over time. We used four different machine learning algorithms to train and test prediction models on a dataset of intensive care patients from two US hospitals. We optimized the hyperparameters of each algorithm and evaluated the performance of each model using the F1 score. The best model was XGBClassifier with an F1 score of 0.75 on the test set. We also analyzed the importance of different features and the limitations of our approach.

## Exploratory Data Analysis

The dataset consists of patients' PSV files, each containing demographic and medical data about one patient in intensive care. Each file has several columns, representing different features. The last column is SepsisLabel, which indicates whether the patient suffered from sepsis within 6 hours of the current row. The dataset is imbalanced, as only a minority of the rows have SepsisLabel equal to 1.

## Feature Engineering

We decided to use all 40 features except for SepsisLabel as input for our prediction model, as we assumed that they all have some predictive power for sepsis risk.

## Prediction

We used four different machine learning algorithms to build prediction models: Logistic Regression (LR), Random Forest Classifier (RFC), Support Vector Classifier (SVC), and XGBoost Classifier (XGB). We used Optuna to optimize the hyperparameters of each algorithm by maximizing the F1 score.

## Summary & Discussion

In this project, we aimed to predict sepsis risk in intensive care patients using machine learning models. We used a dataset of clinical data and performed exploratory data analysis, feature engineering, and prediction tasks. We used four different machine learning algorithms and optimized their hyperparameters. We evaluated their performance using F1 score on a test set. We found that XGB was the best model with an F1 score of 0.75 on the test set.