

## Heinz 95-845: Project Proposal

**Kevin McGrady**

KEVINMCG/KEVINMCG@ANDREW.CMU.EDU

*Heinz College*

*Carnegie Mellon University*

*Pittsburgh, PA, United States*

### 1. Proposal Details (10 points)

Please provide information for the following fields. Your proposal write-up should be less than 2 pages.

#### 1.1 What is your proposed analysis? What are the likely outcomes?

For the final project in Heinz 95-845, I propose to investigate severe sepsis. This will take one of three forms: predicting the onset of severe sepsis, examining the efficacy of the treatment of severe sepsis, or predicting readmissions following severe sepsis. The determination of the exact analysis will be made following additional sepsis research, data examination, and machine learning skill development.

#### 1.2 Why is your proposed analysis important?

Severe sepsis is a leading cause of death in hospitals and is a major component of readmissions costs.

#### 1.3 How will your analysis contribute to existing work? Provide references.

This analysis was inspired by the work published by Henry, Hager, Pronovost, and Saria. Their targeted real-time early warning score (TREWScore) was intriguing. If I were to choose to study signals predicting severe sepsis, I would examine their selected features to ensure that they are in fact routinely available in electronic health records. I would also investigate other methods of survival analysis. Finally, I would explore the variation amongst the types of intensive care units. The treatment of sepsis requires further research as does the readmission prediction to understand possible analytical contribution.

#### 1.4 Describe the data. Please also define $Y$ outcome(s), $U$ treatment, $V$ covariates, $W$ population as applicable.

The data will be drawn from the MIMIC II database. The population will be those individuals admitted to intensive care units at Beth Israel Deaconess Medical Center in Boston, MA from 2001 to 2008. If the first study is chosen, the outcome would be severe sepsis; there would be no treatment; the covariates will be basic demographics and other selected features. If the second study is chosen, the outcome would be severe sepsis, the treatment would be treatment for severe sepsis; the covariates will be basic demographics and other selected features. If the third study is chosen, the outcome would be readmission; there would be no treatment; the covariates will be basic demographics and other selected features.

**1.5 What evaluation measures are appropriate for the analysis? Which measures will you use?**

Standard machine learning metrics for classification problems will be employed including: area under the (ROC) curve, precision, and recall.

**1.6 What study design, pre-processing, and machine learning methods do you intend to use? Justify that the analysis is of appropriate size for a course project.**

In each case, feature selection algorithms will need to be applied to the dataset. If the first study is chosen, I will be using various methods of survival analysis. If the second or third study is chosen, I will be using a variety of classification algorithms. It is unclear at this point what I can or cannot complete in an analysis on severe sepsis within the confines of this course, which is why I am leaving some options open in terms of study design.

**1.7 What are possible limitations of the study?**

The limitations of this study, outside of timeframe (two months), current level of medical knowledge (near zero), and current level of machine learning skill (novice), is the dataset itself. The question will be how well can the results of this study be generalized given the idiosyncrasies of this medical center and the patients that it receives.

**1.8 References**

Henry, Katharine E., et al. "A targeted real-time early warning score (TREWScore) for septic shock." *Science Translational Medicine* 7.299 (2015): 299ra122-299ra122.