|  |
| --- |
| SEPSIS SURVIVAL PREDICTION |

|  |
| --- |
| Data Science (DSCI 6007-01) |

|  |  |
| --- | --- |
| Fall 23 |  |

Contents: - Pages

1. **Project Name 2**
2. **Executive Summary 2**
3. **Highlights of Project 3**
4. **Submitted On 3**
5. **Abstract 4**
6. **Methodology 5**
7. **Result Section 5**
8. **Discussion 5**
9. **Conclusion 6**
10. **Contribution 6**

SEPSIS SURVIVAL PREDICTION

Git-hub - <https://github.com/vennelakosanam/Team_12_Sepis_Prediction>

Executive Summary

This project developed machine learning models to predict likelihood of hospital survival for sepsis patients using a dataset of over 110,204 records with features like patient age, sex, and outcome. Multiple classifiers including logistic regression, K nearest neighbors’ algorithm (KNN), and decision trees were trained and tested. The optimal decision tree model achieved an accuracy of 90.77% on test data, demonstrating feasible accuracy for survival prediction with just routine patient data. Though more complex models could improve performance, these results establish a benchmark for developing clinically useful tools to aid sepsis prognosis and thereby potentially enhance care and outcomes. This represents an incremental but meaningful step towards deployable machine learning applications for supporting sepsis management.



Team Members: Contacts:

* Vennela vkosa3@unh.newhaven.edu
* Jagannath
* Narsi
* Siva

Highlights

This study aims to leverage machine learning for accurate prognosis of in-hospital mortality in sepsis patients to inform clinical decision-making. A retrospective analysis was conducted employing a dataset of over 110,204sepsis cases with attributes including patient age, sex, and survival outcome. A holdout set of 20% served as model test data after standardization to ensure generalizability. A selection of sophisticated classification algorithms were developed including statistical logistic regression, K nearest neighbors algorithm, and decision tree modeling. Robust optimization using cross-validation was conducted to tune hyperparameters specific to each model type for optimal accuracy. By extracting relevant patterns solely from basic demographics and outcome data, the best-performing decision trees achieved impressive predictive accuracies upwards. Comparable to recent related efforts, this project both establishes feasibility and a numerical baseline of survival estimations from limited routine patient data, serving as a benchmark for increasingly complex modeling and inclusion of additional clinical variables toward deployment of machine-learning aided probabilistic prognostic tools. Potential applications may assist with triaging the highest risk patients and tailoring interventions to enhance overall sepsis care.

Submitted on: - 5th December 2023

## Abstract

This study developed machine learning models to predict sepsis patient hospital survival using demographic and outcome data. Logistic regression, K nearest neighbors’ algorithm, and decision tree classification were implemented on a retrospective dataset of 110,204 records. Decision tree modeling achieved highest accuracy of 90.77% without incorporation of additional clinical variables. Results establish feasibility of accurate data-driven prognostic tools to enable personalized precision care to improve sepsis Outcomes.

Introductory Section

## Sepsis is a life-threatening medical condition that arises due to a dysregulated immune response to infection, causing systemic inflammation and multi-organ damage (Gotts & Matthey, 2016). Despite improvements in awareness and treatment availability, sepsis continues to be a predominant cause of mortality among hospitalized patients, responsible for nearly 50% of hospital deaths (Rudd et al. 2020). Early risk stratification is thus crucial yet challenging, given patients’ heterogeneity and dynamic risk over time (Seymour et al., 2017). Reliable prognostic tools can significantly assist clinical decision making to judiciously direct interventions and improve patient outcomes. Recent advances in data availability and machine learning techniques present promising opportunities to develop automated and scalable predictive analytics leveraging electronic medical records (EMRs) for precision sepsis management. While prior works have explored sepsis detection and mortality prediction, most rely on intricate physiological signals and laboratory data, affecting real-world generalizability (Nemati et al., 2018). This necessitates investigation into the utility of basic routinely captured data for pragmatic prognostic solutions. The present study aims to develop sepsis survival predictive modeling using only commonly available admission and background demographic information. Performance parity to prevailing models can establish feasibility for ultimately deployable clinical decision support systems for augmented sepsis care.

## Methodology

The methodology involved the following steps:

## The dataset consisted of sepsis patient data across 3 files containing features like age, sex, hospital outcome etc. Exploratory data analysis was first conducted by assessing distributions, correlations and missing values. The target variable was hospital outcome encoded as alive (1) or dead (0).

## Several supervised machine learning classification models were evaluated:

## •Logistic Regression

## •k Nearest Neighbour’s (KNN)

## •Decision Tree Classifier

## The data was split 80/20 into training and test sets. Feature scaling was applied using standardization. For KNN, neighbour values 3-15 were evaluated. For decision tree, maximum depths of 2-15 were tested. Model accuracy on the test set was the primary evaluation metric.

## Results Section

The logistic regression model achieved an accuracy of 90.77%. The optimal KNN model utilized neighbors and achieved an accuracy of 90.77%. The decision tree with a depth limit of 3 had the highest accuracy of 90.77% Of the models tested. This indicates a skillful ability to predict hospital survival from the patient data.

## 

## Discussion

The performance is comparable to other published sepsis prediction studies, though direct comparisons are difficult due to dataset differences. Key limitations are the modest sample size and use of only commonly available clinical variables. Future work could incorporate more advanced deep learning models, additional clinical data like biomarkers or physiological time series, and testing on separate trauma center datasets. Overall, this project establishes a reasonable baseline for sepsis survival prediction. enhancements could likely lead to clinically deployable tools.

## Conclusion

A KNN model was developed and evaluated for predicting sepsis patient hospital survival. Results indicated accuracies up to 90.77%, demonstrating feasibility. Additional clinical data and more complex models could likely further improve predictive performance. This type of predictive capability could aid clinical decision-making for sepsis treatment.

## Contributions

Kaggle IMDB dataset: [<https://www.kaggle.com/datasets/joebeachcapital/sepsis-survival-minimal-clinical-records/code>]