

Project Synopsis
On
Sepsis Prediction Using Machine Learning

Submitted as a part of course curriculum for

Bachelor of Technology
in
Computer Science



Submitted by

Yash Puri (2000290120196)
Yash Goel (2000290120195)
Vishal Verma (2000290120193)

Under the Supervision of

Ms. Neha Shukla
Assistant Professor,
Department of Computer Science

KIET Group of Institutions, Ghaziabad
Department of Computer Science
Dr. A.P.J. Abdul Kalam Technical University
2023-2024

TABLE OF CONTENTS

TITLE

DECLARATION

CERTIFICATE

ACKNOWLEDGEMENT

ABSTRACT

LIST OF FIGURES

LIST OF ABBREVIATIONS

CHAPTER 1 INTRODUCTION

1.1 Introduction

1.2 Problem Statement

1.3 Objective

1.4 Scope

CHAPTER 2 LITERATURE REVIEW

CHAPTER 3 PROPOSED METHODOLOGY

CHAPTER 4 TECHNOLOGY USED

CHAPTER 5 DIAGRAMS

CHAPTER 6 CONCLUSION

REFERENCES

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students:

Name: Yash Puri, Yash Goel, Vishal Verma

Roll No.: 2000290120196, 2000290120195, 2000290120193

Date: 14-January-2024

CERTIFICATE

This is to certify that Project Report entitled “**Sepsis Prediction Using Machine Learning**” which is submitted by **Yash Puri, Yash Goel, Vishal Verma** in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: 14-January-2024

Supervisor Signature

Ms. Neha Shukla
Assistant Professor,
Department of Computer Science

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project undertaken during B.Tech. Fourth Year. We owe a special debt of gratitude to Ms. Neha Shukla (Assistant Professor), Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for her constant support and guidance throughout the course of our work. Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Shrivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:

Date : 14-January-2024

Name : Yash Puri, Yash Goel, Vishal Verma.

Roll No: 2000290120196, 2000290120195, 2000290120193.

ABSTRACT

Machine Learning is a bunch of algorithms which allows enormous software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic fundamental of machine learning is to build models and employ algorithms that can take input data and use its algorithms and analysis to predict an output while updating outputs as new data becomes available.

In this project we are trying to calculate the probability of onset of sepsis as soon as the patient visits the hospital for treatment. For achieving this we are making a machine learning model in which we will use certain machine learning algorithms and methods of machine learning and will choose the best one depending on the accuracy levels and performance. To achieve this first we will train our dataset which will include dataset of patient including various vital parameters etc. Then we will train our model on these dataset after training of our dataset we find out its accuracies and error or if any fixes needed in the model. When we are done with training of our model on these dataset now we will check our model with new data or we can say with testing datasets. Then we will find our model which is best working with less error values and more accuracy. In this prediction model we will try python libraries such as numpy, seaborn, matplotlib, open cv and some other algorithms and classification techniques as per our requirements if necessary. These libraries and models are proposed and will choose the best one at the time of implementation.

LIST OF FIGURES

FIG-1.1: Workflow Model Diagram

FIG-1.2: Architecture

CHAPTER 1-INTRODUCTION

Sepsis is a hazardous condition that happens when the body's reaction to any kind of contamination causes tissue harm, organ failure, or even demise of the person. Generally, the body releases natural chemicals into the blood circulation system in order to fight the infection which is inside. Sepsis occurs when the body's response to these chemicals is out of balance, this can damage many organ systems. Sepsis is caused by infection and can happen to anyone. It is most common and dangerous for senior citizens, pregnant ladies, kids below one-year-old, persons suffering from chronic conditions, such as diabetes, kidney disease, lung disease, or even cancer, as they have weak immune systems. This disease is a major health concern for the public in terms of morbidity, health care expenses and mortality. Detecting at early stages, with antibiotic treatment the outcomes can be improved. Though many professional care societies have proposed new methods in recognizing sepsis, the central requirement for early identification and treatment remains neglected. It can be treated if it can be recognised at early stages. Several examinations have demonstrated that delays in finding and treatment of sepsis can prompt high death rates. Our main motto is to detect sepsis as soon as the patient visits the emergency department for the treatment.

We are going to use various python libraries such as numpy, pandas, matplotlib, seaborn etc and various machine learning algorithms to train our model on dataset. All these would be used to create a prediction model and then comes the stage of testing of our model on testing dataset which contains new data and we will compare it the difference between the actual values and predicted values if the difference or error value is less then it will be the best prediction model. This will be a machine learning prediction model. According to us this model will prove to be very helpful in predicting the onset of sepsis accurately.

PROBLEM STATEMENT

Our problem statement in simple words would be like:

Sepsis is a potentially life-threatening condition that occurs when the body's response to an infection damages its own tissues. By detecting sepsis at early stages, helps in saving the lives of people. For the patient the practicality of detecting sepsis disease occurrence in development is an important factor in the result. Late prediction of sepsis is potentially life-threatening, and also consumes heavy hospital resources. By using Machine Learning models predicting sepsis consumes limited resources and can assume the risk of prediction to be minimal time.

- We are going to analyse different algorithms on performance.
- This will be ml model for prediction of sepsis.
- This model will predict the onset of sepsis in minimal time possible.

OBJECTIVE

Our objective of this project is to provide a machine learning model which will be effective for the following:-

- To identify the best classifier among Logistic Regression, KNN, Naive Bayes and Random Forest for prediction of sepsis at early stages.
- To design and develop a website and integrate with the model.
- Takes patient data with vital parameters and predict sepsis with accuracy.
- And some similar related functions like these.

SCOPE

After discussion with the team and with our guide, we can consider the scope of this project as follows:

- We can create a website for calculating or predicting the onset of sepsis when the patient arrives the hospital for treatment.
- Healthcare professionals can easily use this model to quickly get information regarding presence of sepsis to improve the treatment outcomes.
- The project is only useful for doctors who has knowledge on the disease and its treatment.

CHAPTER 2 -LITERATURE REVIEW

2.1. Predicting Infections Using Computational Intelligence – A Systematic Review [1]

In this research paper the author, computational intelligence can be useful to predict the risk of infection in patients, raising early alarms that can aid medical teams to respond as quick as possible. In the study the most widely addressed infection is by far sepsis, followed by infection and surgical site infections. Most works use machine learning techniques, SVM, logistic regression ,random forest and naive bayes are used for early identification of infections which shows better accuracy but the implementation of computational intelligence in healthcare can be expensive, particularly if it involves the development of new technologies or the restructuring of existing healthcare system.

Increased attention to large scale machine learning is also due to the spread of very large datasets across many modern applications and also due to the accuracy of implementing these techniques.

2.2. A Deep Learning-Based Sepsis Estimation Scheme [2]

By this research paper the author is trying to empower the great success of machine learning and its various algorithms in helping many people with their particular needs. The objective of this research is to design and implement a machine learning (ML) based technique that can predict cases of septic shock and extreme sepsis and assess its effects on medical practice and the patients. Data from the laboratory tests serve as the primary early indicator of septic shock by confirming the presence of toxins. The patients should have had positive blood culture during their interaction with the hospital. Data from the laboratory tests serve as the primary early indicator of septic shock by confirming the presence of toxins.

The layers become more and accuracy of the model becomes less to predict cases of extreme sepsis and septic shock.

2.3-Machine learning for health care diagnostics [3]

By this research paper first of all the author sets up the base for health diagnostics and limitations in health care diagnostics also how can they be coped by using machine learning and its various methods and algorithms. In this research paper the author tries to explain the current limitations in health care and the benefits of introducing machine learning for diagnostics. Author conveys that systems based on machine learning can be used to find various complications before they occur at very early stage of diagnosis which will prove to be very useful in treatment if diagnosed at an early stage. This paper

basically contrasts on the importance of the systems based on machine learning and its algorithms in health care for diagnostics.

Very huge and enormous dataset are available in It department of various hospitals which includes the past history of patients, disease encountered, disease diagnosed, treatment given or cure such data sets from these hospitals can be used to train our machine learning algorithms on the basis of outcome. Author says that it is comparatively easy to find patterns in such a huge dataset and predict outcomes using machine learning.

Machine learning systems can be advanced and trained with more huge number of dataset to be capable of giving more accurate results as those given by a well trained clinician.

2.4. Learning representations for the early detection of sepsis with deep neural networks [4]

In this research paper the author aimed to construct early-stage sepsis detection models using deep learning algorithms and methodologies and to assess the viability and improvement of the novel deep learning approach against that of the regression method utilizing traditional “temporal feature extraction”. They achieved enhanced performance with input to output layer that is fed forward in only direction neural networks utilizing extended short-term memory and enhanced performance with deep neural networks by comparing them to reference features.

We can understand more about the neural networks by taking a look into an example which is one of the most simplest example which is a fully connected three layer model that consists of input layer, the hidden layer and at last the output layer.

Training of neural networks can be difficult and can be very time consuming also but once we have trained our neural network model than they can prove to be very fast and efficient.

The use of deep neural networks with input to output layer feeding in only one direction has made the detection process faster and more efficient but due to the complexity of deep learning algorithms, reproducing the same results on different datasets or with different parameters can be challenging. This makes it difficult to evaluate and compare the performance of different models.

2.5. Coronavirus disease (COVID 19): Analysis and Prediction using machine learning techniques [5]

As we all know from the history coronavirus is as deadly virus as it appears to be, from SARS COV in the seafood market near Wuhan in China during the end of year 2019. And it quickly spread across the world which made a dreadful situation worldwide across the globe. Due to sudden spread and lack of specific treatment people faced major challenges.

It was observed that symptoms such as fever, throat irritation, breathing issues, but it is observed that people having diabetes or patients with blood pressure problems had more severe symptoms as compared to other people.

In some research it is also shown that covid reduced the count of wbc, rbc which were one of the major causes of fatality rate. With the outbreak of covid 19 machine learning and data science has played a crucial role in prediction of upcoming trends of covid 19 for safety measures and implementation of various precautions that had been taken to combat or to fight the battle against covid on huge scale more effectively. With this paper the author came up to conclusion that the predicted outcomes of naive based system are very close enough to the actual confirmed cases. However other techniques of machine learning such as linear regression and support vector machine do not stand up on the marked line. And thus this can be concluded that in such cases naive bayes is more likely to be preferred in future for prediction purposes. Finally at last we can conclude that AI and ML can learn to detect an outbreak at an early stage.

2.6. Role of data science in managing Covid 19 pandemic [6]

In this research paper the author is trying to convey that data science along with machine learning and its different prediction algorithms has always played an important and vital role in fighting against deadly pandemic which was popularly known as SARS COV 2. From making categories of the covid zones based on the severity and density of infected patients to deciding the dosage and medicine for the patient on the severity of infection as a cure or treatment. Statistical analysis and mathematical model are always key towards data science field which can be further categorized into data management, data visualization and statistical machine learning with the help of its various algorithms which can further be used for organization, sorting, processing and many times for real time data analysis. The data science expert must build their model keeping in mind the limitations and shortcomings also that are associated with it. They have to consider their models with certain reservations. The catch here proposed here by the author is that these models are acting just as an helping hands towards successful fight against covid 19 or towards the successful treatment of covid 19 infection on such a large scale.

CHAPTER 3 –PROPOSED METHODOLOGY

This section will include flow chart that how the system will be built step by step for predicting sepsis along with its explanation.

3.1-Flow chart

Import Libraries
(Pandas, NumPy, Matplotlib)



Read Dataset
(Split dataset into training and testing parts)



Data Preprocessing



Remove unnecessary columns and units.
Impute missing values using forward fill and backward fill.
Drop remaining missing values.
Standardize the data using StandardScaler



Data Analysis



Data Balancing
(Undersample majority class to balance dataset)



Model Training
- Splitting data into training & testing sets
- Train on different Classifiers
- Apply XGBoost Classifier



Model Evaluation
(Accuracy, Precision,
Mean Absolute error,
Confusion Matrix)



Save the best performing model

3.2- Algorithm proposed

- Collect our dataset
- Refine it and handle if any errors
- Data preprocessing
- Find correlation in dataset if any
- Define the model
- Train the model using different algorithms
- Check model on testing data
- Model evaluation
- Choose the best performing model
- If desirable then use this model to predict sepsis on real values/ scenario

CHAPTER 4-TECHNOLOGY USED

After taking some consideration, finding some results from the web, discussing with guide and other experts we decided to use the technology which is mainly based on machine learning with the following machine learning algorithms and techniques:-

1) Python libraries:-

- Numpy
- Panda
- Matplotlib
- Seaborn
- Scikit

2) Algorithms

- KNN classifier
- Naïve bayes
- Random forest
- XG boost

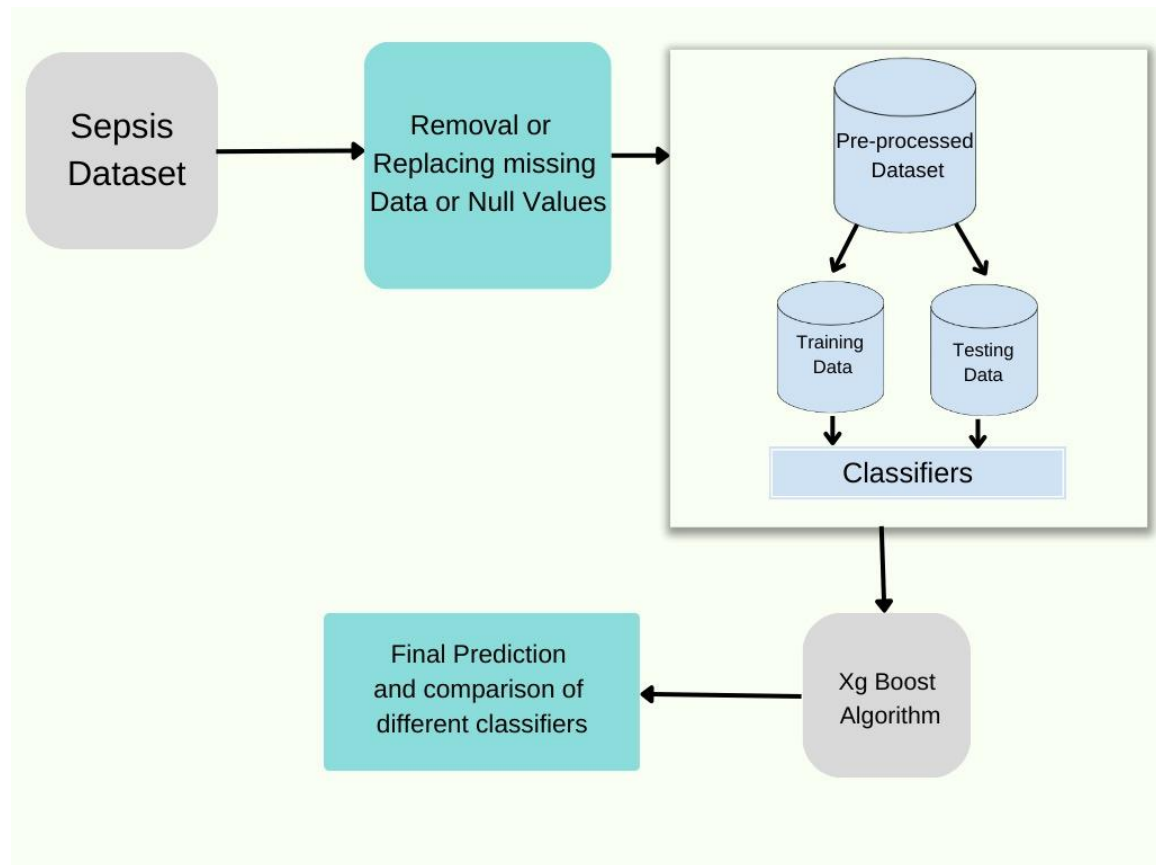
3) HTML and CSS

Features :

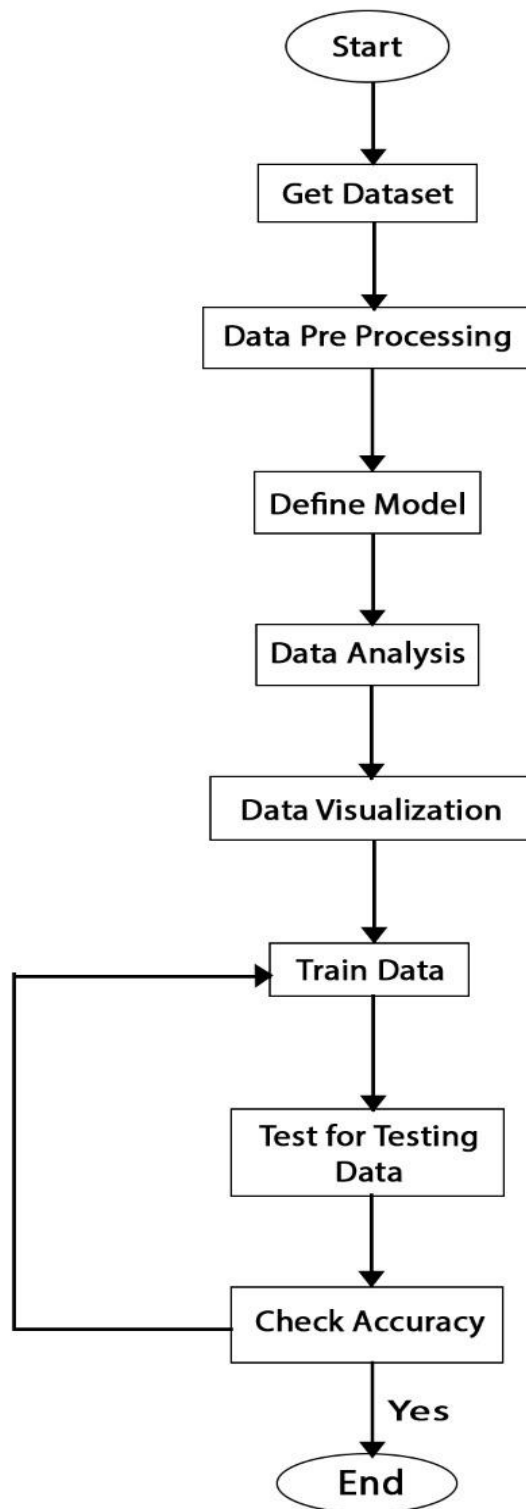
- Huge amount of data can be analysed in minimal time.
- No human intervention required for detection.
- Can predict sepsis with more accuracy.
- Lesser possibility of false results.
- Save resources and time.

DIAGRAMS

ARCHITECTURE



WORKFLOW



CHAPTER 6-CONCLUSION

In conclusion of this project, we would conclude that this project will provide a great help to future as already discussed in the future scope section. This project is mainly based on machine learning algorithms and will work as an assist for healthcare professionals to easily detect sepsis at an early stage.

Our main motive of this was to calculate the onset of sepsis without any delay to improve the patient outcomes such as lower mortality and morbidity rates. This project will be of great help for both patients as well as doctors to take necessary actions as early as possible. This project can also be used to provide educational content about the technologies used to concerned people for making more advancements in such machine learning models to increase their accuracy and reliability with time.

REFERENCES

- [1] - Baldominos, A., Puello, A., O'gul, H., A,suro'glu, T., Colomo- Palacios, R. (2020).Predicting infections using computational intelligence—a systematic review. IEEE Access, 8, 31083-31102.
- [2] - Al-Mualemi, B. Y., Lu, L. (2020).A deep learning-based sepsis estimation scheme. IEEE Access, 9, 5442-5452.
- [3]. Shashua, A. (2009). Introduction to machine learning: Class notes 67577. arXiv preprint arXiv:0904.3664.
- [4] - Kam, H. J., Kim, H. Y. (2017).Learning representations for the early detection of sepsis with deep neural networks.Computers in biology and medicine, 89, 248-255. [5]. Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM, 42(11), 30-36.
- [5] - Zhao, X., Shen, W., Wang, G. (2021). Early prediction of sepsis based on machine learning algorithm. Computational Intelligence and Neuroscience, 2021.
- [6] - Wang, Z., Yao, B. (2021). Multi-branching temporal convolutional network for sepsis prediction. IEEE Journal of Biomedical and Health Informatics, 26(2), 876-887.
- [7] - Scherpf, M., Gr'aber, F., Malberg, H., Zaunseder, S. (2019).Predicting sepsis with a recurrent neural network using the MIMIC III database. Computers in biology and medicine, 113, 103395.
- [8] - Fu, M., Yuan, J., Lu, M., Hong, P., Zeng, M. (2019, September).An ensemble machine learning model for the early detection of sepsis from clinical data. In 2019 Computing in Cardiology IEEE.