CHAPTER 1

INTRODUCTION

1.1 About

There are times when we need a doctor all of a sudden but sometimes they are not available due to some reason and we are left in trouble. The system we have proposed is user friendly to get help and advice on health issues immediately through the online healthcare system. Now a days, with the help of the statistics and posterior distribution the problems are swiftly and easily. As the Bayesian statistics has a great success rate in the field of economic, social science and a few other fields just like that, in medical fields, people have solved various medical problems that are tiresome to be settled in classic statistics by classification and can be solved easily. Naive Bayes is among the basic common classification techniques introduced by Reverend Thomas Bayes. The classification rules which help in solving the prediction of disease are generated by the samples trained by themselves and help in solving the problem easily. It is approximated that greater than 70% of people in India are prone to various body dis- eases like viral, flu, cough, cold etc. in intervals of 2 months. As many people don’t understand that the general body diseases could be symptoms of something more harmful, 25% of this population dies or gets some serious medical problem because of ignoring the early general body symptoms and this is a very serious condition that we are facing and the problem can be proven to be a very dangerous situation for the population and can be alarming if the people will continue ignoring these diseases. Hence identifying or predicting the disease at the very basic stage is very important to avoid any unwanted problems and deaths. The systems which are available now a days are the systems that are either dedicated to a particular disease or are in development or the research for solving the algorithms related to the problem when it comes to generalized disease. The main motive of the proposed system is the prediction of the commonly occurring diseases in the early phase as when they are not checked or examined they can turn into a disease more dangerous disease and can even cause death. The system applies data mining techniques, decision tree algorithms, Naive Bayes algorithm and Random Forest algorithm. This system will predict the most possible disease based on the given symptoms by the user and precautionary measures required to avoid the aggression of disease, it will also help doctors to analyze the patterns of diseases in the society. This project is dedicated to the Disease prediction System that will have data mining techniques for the basic stages of the dataset and the main model will be trained using the Machine Learning (ML) algorithms and will help in the prediction of general diseases.

1.2 Data Mining and Machine Learning Algorithm

The Data Mining and the Machine Learning Algorithms are used for the prediction of Disease in the Project. There are different Data Mining and Machine Learning used for the purpose of correcting and evaluating the dataset and then testing the dataset on the basis of train score and the test score of the ML model.1.2.1 Data Analysis and Data Mining The Data Mining is a process in which raw data is prepared and structured from the unstructured data as to take meaningful information from the data which can be used in the project. Task of making data organized and reflective about data is to way to get what this information does the data contains in it and what it does not have in it. There are so many different types of methods in which the people can make use of data analysis. It is simply very easy to use data during the analysis phase and get to some certain conclusions or some agendas. The analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the objective of highlighting useful information, suggesting conclusions, and supporting decision making which are helpful to the user. Data analysis has multiple facets and approaches, encompassing diverse techniques under an array of names, in different business, science, and social science domains. Data Mining is the discovery of unknown information found in databases, data mining functions has some different methods for clustering, classification, prediction, and associations. In the data mining important application is that of mining association rules, association rules was first introduced in 1993 and are used to identify relationships among a set of items in databases these different properties are not based on the properties of the data, but rather based on co-occurrence of the data items. The Data mining helps in giving new and different perspectives for data analysis the main role of data mining is to extract and discover new knowledge from data. In the past few years, different methods have been coined and developed about the capabilities of data collection and data generation, data collection tools have provided us with a huge amount of data, data mining processes have integrated techniques from multiple disciplines such as, statistics, machine learning, database technology, pattern recognition, neural networks, information retrieval and spatial data analysis. The data mining techniques have been used in many different fields such as, business management, science, engineering, banking, data management, administration, and many other applications.

1. Two-component is used to introduce data mining techniques first one is the database, and the second one is machine learning. The database provides data management techniques, while machine learning provides methods for data analysis. But to introduce machine learning methods, it used algorithms.

2. Data Mining utilizes more data to obtain helpful information, and that specific data will help to predict some future results. For example, In a marketing company that utilizes last year's data to predict the sale, but machine learning does not depend much on data. It uses algorithms. Many transportation companies such as OLA, UBER machine learning techniques to calculate ETA (Estimated Time of Arrival) for rides is based on this technique.

3. Data mining is not capable of self-learning. It follows the guidelines that are predefined. It will provide the answer to a specific problem, but machine learning algorithms are self-defined and can alter their rules according to the situation, and find out the solution for a specific problem and resolves it in its way.

4. The main and most important difference between data mining and machine learning is that without the involvement of humans, data mining can't work, but in the case of machine learning human effort only involves at the time when the algorithm is defined after that it will conclude everything on its own. Once it implemented, we can use it forever, but this is not possible in the case of data mining.

5. As machine learning is an automated process, the result produces by machine learning will be more precise as compared to data mining.

6. Data mining utilizes the database, data warehouse server, data mining engine, and pattern assessment techniques to obtain useful information, whereas machine learning utilizes neural networks, predictive models, and automated algorithms to make the decisions.

1.2.1 Artifical Intelligence

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of artificial intelligence is [machine learning](https://www.investopedia.com/terms/m/machine-learning.asp) (ML), which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Deep learning techniques enable this automatic learning through the absorption of huge amounts of unstructured data such as text, images, or video.

When most people hear the term artificial intelligence, the first thing they usually think of is [robots](https://www.investopedia.com/sectors-and-industries-analysis-4689756). That's because big-budget films and novels weave stories about human-like machines that wreak havoc on Earth. But nothing could be further from the truth.

Artificial intelligence is based on the principle that human intelligence can be defined in a way that a machine can easily [mimic it and execute tasks](https://www.investopedia.com/terms/k/knowledge-engineering.asp), from the most simple to those that are even more complex. The goals of artificial intelligence include mimicking human cognitive activity. Researchers and developers in the field are making surprisingly rapid strides in mimicking activities such as learning, reasoning, and perception, to the extent that these can be concretely defined. Some believe that innovators may soon be able to develop systems that exceed the capacity of humans to learn or reason out any subject. But others remain skeptical because all cognitive activity is laced with value judgments that are subject to human experience.

As technology advances, previous benchmarks that defined artificial intelligence become outdated. For example, machines that calculate basic functions or recognize text through optical character recognition are no longer considered to embody artificial intelligence, since this function is now taken for granted as an inherent computer function.

AI is continuously evolving to benefit many different industries. Machines are wired using a cross-disciplinary approach based on mathematics, computer science, linguistics, psychology, and more.

The applications for artificial intelligence are endless. The technology can be applied to many different sectors and industries. AI is being tested and used in the healthcare industry for dosing drugs and doling out different treatments tailored to specific patients, and for aiding in surgical procedures in the operating room.

Other examples of machines with artificial intelligence include computers that play chess and [self-driving cars](https://www.investopedia.com/articles/investing/052014/how-googles-selfdriving-car-will-change-everything.asp). Each of these machines must weigh the consequences of any action they take, as each action will impact the end result. In chess, the end result is winning the game. For self-driving cars, the computer system must account for all external data and compute it to act in a way that prevents a collision.

Artificial intelligence also has [applications in the financial industry](https://www.investopedia.com/financial-advisor/how-ai-shaping-advisory-landscape/), where it is used to detect and flag activity in banking and finance such as unusual debit card usage and large account deposits—all of which help a bank's fraud department. Applications for AI are also being used to help streamline and make trading easier. This is done by making supply, demand, and pricing of securities easier to estimate.

Artificial intelligence can be divided into two different categories: weak and strong. [Weak artificial intelligence](https://www.investopedia.com/terms/w/weak-ai.asp) embodies a system designed to carry out one particular job. Weak AI systems include video games such as the chess example from above and personal assistants such as Amazon's Alexa and Apple's Siri. You ask the assistant a question, and it answers it for you.

[Strong artificial intelligence](https://www.investopedia.com/financial-technology-and-automated-investing-4689759) systems are systems that carry on the tasks considered to be human-like. These tend to be more complex and complicated systems. They are programmed to handle situations in which they may be required to problem solve without having a person intervene. These kinds of systems can be found in applications like self-driving cars or in hospital operating rooms.

ince its beginning, artificial intelligence has come under scrutiny from scientists and the public alike. One common theme is the idea that machines will become so highly developed that humans will not be able to keep up and they will take off on their own, redesigning themselves at an exponential rate.

Another is that machines can hack into people's privacy and even be weaponized. Other arguments debate the ethics of artificial intelligence and whether intelligent systems such as [robots](https://www.investopedia.com/articles/active-trading/081315/how-code-your-own-algo-trading-robot.asp) should be treated with the same rights as humans.

Self-driving cars have been fairly controversial as their machines tend to be designed for the lowest possible risk and the least casualties. If presented with a scenario of colliding with one person or another at the same time, these cars would calculate the option that would cause the least amount of damage.

Another contentious issue many people have with artificial intelligence is how it may affect human employment. With many industries looking to [automate certain jobs](https://www.investopedia.com/articles/markets-economy/091316/3-ways-robots-affect-economy.asp) through the use of intelligent machinery, there is a concern that people would be pushed out of the workforce. Self-driving cars may remove the need for taxis and car-share programs, while manufacturers may easily replace human labor with machines, making people's skills obsolete.

AI is used extensively across a range of applications today, with varying levels of sophistication. Recommendation algorithms that suggest what you might like next are popular AI implementations, as are chatbots that appear on websites or in the form of smart speakers (e.g., Alexa or Siri). AI is used to make predictions in terms of weather and financial forecasting, to streamline production processes, and to cut down on various forms of redundant cognitive labor (e.g., tax accounting or editing). AI is also used to play games, operate autonomous vehicles, process language, and much, much, more.

In healthcare settings, AI is used to assist in diagnostics. AI is very good at identifying small anomalies in scans and can better triangulate diagnoses from a patient's symptoms and vitals. AI is also used to classify patients, maintain and track medical records, and deal with health insurance claims. Future innovations are thought to include AI-assisted robotic surgery, virtual nurses or doctors, and collaborative clinical judgment.

1.2.2 Machine Learning

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/topics/artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

IBM has a rich [history](https://www.ibm.com/ibm/history/ibm100/us/en/icons/ibm700series/impacts/) with machine learning. One of its own, Arthur Samuel, is credited for coining the term, “machine learning” with his [research](https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/636026949/report_frank_gabel.pdf) (PDF, 481 KB) (link resides outside IBM) around the game of checkers. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat seems trivial, but it’s considered a major milestone in the field of artificial intelligence.

Over the last couple of decades, the technological advances in storage and processing power have enabled some innovative products based on machine learning, such as Netflix’s recommendation engine and self-driving cars.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

Machine learning algorithms are typically created using frameworks that accelerate solution development, such as TensorFlow and PyTorch.

since deep learning and machine learning tend to be used interchangeably, it’s worth noting the nuances between the two. Machine learning, deep learning, and neural networks are all sub-fields of artificial intelligence. However, neural networks is actually a sub-field of machine learning, and deep learning is a sub-field of neural networks.

Neural networks, or artificial neural networks (ANNs), are comprised of node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network by that node. The “deep” in deep learning is just referring to the number of layers in a neural network. A neural network that consists of more than three layers—which would be inclusive of the input and the output—can be considered a deep learning algorithm or a deep neural network. A neural network that only has three layers is just a basic neural network.

Deep learning and neural networks are credited with accelerating progress in areas such as computer vision, natural language processing, and speech recognition.

Machine learning is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect. We probably use a learning algorithm dozens of time without even knowing it. Applications of Machine Learning include:

* Web Search Engine: One of the reasons why search engines like google, bing etc work so well is because the system has learnt how to rank pages through a complex learning algorithm.
* Photo tagging Applications: Be it facebook or any other photo tagging application, the ability to tag friends makes it even more happening. It is all possible because of a face recognition algorithm that runs behind the application.
* Spam Detector: Our mail agent like Gmail or Hotmail does a lot of hard work for us in classifying the mails and moving the spam mails to spam folder. This is again achieved by a spam classifier running in the back end of mail application.
* Augmentation:Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias.
* Automation:Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential process steps in manufacturing plants.
* Finance Industry:Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.
* Government organization:The government makes use of ML to manage public safety and utilities. Take the example of China with the massive face recognition. The government uses Artificial intelligence to prevent jaywalker.
* Healthcare industry: Healthcare was one of the first industry to use machine learning with image detection.
* Marketing:Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, marketing department relies on AI to optimize the customer relationship and marketing campaign.

Today, companies are using Machine Learning to improve business decisions,increase productivity, detect disease, forecast weather, and do many more things. With the exponential growth of technology, we not only need better tools to understand the data we currently have, but we also need to prepare ourselves for the data we will have. To achieve this goal we need to build intelligent machines. We can write a program to do simple things. But for most of times Hardwiring Intelligence in it is difficult. Best way to do it is to have some way for machines to learn things themselves. A mechanism for learning – if a machine can learn from input then it does the hard work for us. This is where Machine Learning comes in action. Some examples of machine learning are:

* Database Mining for growth of automation: Typical applications include Web-click data for better UX( User eXperience), Medical records for better automation in healthcare, biological data and many more.
* Applications that cannot be programmed: There are some tasks that cannot be programmed as the computers we use are not modelled that way. Examples include Autonomous Driving, Recognition tasks from unordered data (Face Recognition/ Handwriting Recognition), Natural language Processing, computer Vision etc.
* Understanding Human Learning: This is the closest we have understood and mimicked the human brain. It is the start of a new revolution, The real AI. Now, After a brief insight lets come to a more formal definition of Machine Learning
* Arthur Samuel(1959): “Machine Learning is a field of study that gives computers, the ability to learn without explicitly being programmed.”Samuel wrote a Checker playing program which could learn over time. At first it could be easily won. But over time, it learnt all the board position that would eventually lead him to victory or loss and thus became a better chess player than Samuel itself. This was one of the most early attempts of defining Machine Learning and is somewhat less formal.
* Tom Michel(1999): “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” This is a more formal and mathematical definition

1.2.3 Machine Learning Algorithms

The ML is a small part of Artificial Intelligence (AI) which is used in the computation work and the analysis work in the AI. The ML algorithms are used to find different patterns and different structures in the dataset which is provided to the dataset, the ML algorithms are used to give a large computation capabilities to the system by which a large amount of data is given to the model for the purpose of training and testing the data, the ML algorithms are used in decision making process the model which is prepared by using the ML has a large amount of data in it which makes it a very good for the process of decision making. ML algorithms have very high computational power and are proven to be very helpful in today’s world. Different types of ML algorithms are organized into different ways, based on the desired outcome of the algorithm. Common algorithm types include:

* Supervised learning — The supervised learning algorithm can apply what has been learned in the past to new data using labelled examples to predict the future events. Starting from analysis of a known training dataset. This algorithm is used to provide targets for any new values after sufficient amount of training of the model.
* Unsupervised learning — Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. This algorithm shows how the system can infer a function to describe a hidden structure from unlabeled data.
* Semi-supervised learning — This category of the ML algorithms falls somewhere be-tween the supervised learning and the unsupervised learning algorithm which combines both labeled and unlabeled examples to generate an appropriate function or classifier which is used to make a model for the purpose of prediction or classification.
* Reinforcement learning — This is the algorithm where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.
* Transduction — This algorithm is similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs.
* Learning to learn — This method is where the algorithm learns its own inductive bias based on previous experience.

The performance and computational analysis of ML algorithms is a branch of statistics known as computational learning theory.

Machine learning is about designing algorithms that allow a computer to learn. Learning is not necessarily involves consciousness but learning is a matter of finding statistical regularities or other patterns in the data. Thus, many machine learning algorithms will barely resemble how human might approach a learning task. However, learning algorithms can give insight into the relative difficulty of learning in different environments.

Machine learning is made up of three parts:

* The computational algorithm at the core of making determinations.
* Variables and features that make up the decision.
* Base knowledge for which the answer is known that enables (trains) the system to learn.

Initially, the model is fed parameter data for which the answer is known. The algorithm is then run, and adjustments are made until the algorithm’s output (learning) agrees with the known answer. At this point, increasing amounts of data are input to help the system learn and process higher computational decisions.

CHAPTER 2

LITERATURE SURVEY

* In model [1] Machine learning algorithms for early sepsis detection in the emergency. Department: a retrospective study. Here the proposed model is to find The primary outcome of the model performance that predicted the final diagnosis of sepsis determined by the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, and predictive performance compared with those of the reference models (quick sequential organ failure assessment [qSOFA],modified early warning score [MEWS], and systemic inflammatory response syndrome [SIRS]) using the testing dataset.

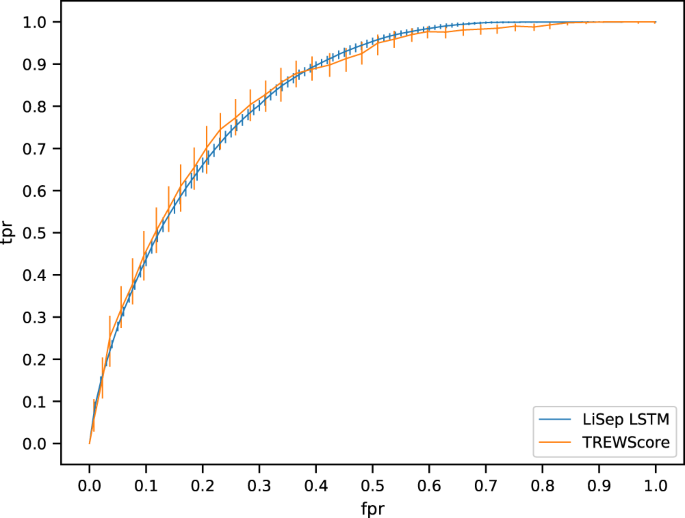


Figure 2.1: Primary prediction of sepsis disease

* The model [2] Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. Here the proposed model is to use the clinical text models include narrative notes written by nurses, physicians, and specialists in varying situations. This is often combined with common structured data such as demographics, vital signs, laboratory data, and medications. Area under the receiver operating characteristic curve (AUC) comparison of ML methods showed that utilizing both text and structured data predicts sepsis earlier and more accurately than structured data alone. No meta-analysis was performed because of incomparable measurements among the 9 included studies.

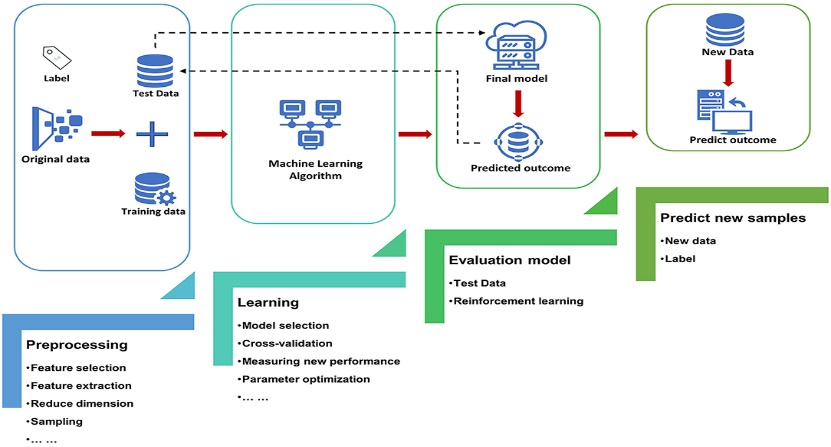


Figure 2.2: The clinical data for detection of sepsis.

* The model [3] A Machine Learning Model for Early Prediction and Detection of Sepsis in Intensive Care Unit Patients. In this work, machine learning models for the prediction of different stages of sepsis are proposed. Here, a frame of the problem of early prediction and detection of different stages of sepsis as a classification problem is developed. The objective is to continuously update the predicted probability that the encounter will result in sepsis using all accessible patients’ information up till that time.
* In the model [4] Interpretable Machine Learning for Early Prediction of Prognosis in Sepsis: A Discovery and Validation Study. The current research was a single-center retrospective modeling study using the MIMIC-IV database, which restricted us from identifying the causal relationship between features and outcomes. Thus, further prospective randomized clinical trials are required to validate the effectiveness for our model. Secondly, many unmeasured confounders may contribute to the impact on the prognosis of death for septic patients, such as racial and treatment strategies. We may therefore have neglected some important features used for the XGBoost model due to the limitation of MIMIC-IV.
* In the model proposed by [5] Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. The primary outcome was all-cause in-hospital mortality which was measured as the patient’s status at discharge. Secondary outcomes were the difference between a patient’s SOFA score at the time of the alert and 72 h after the alert and hospital length of stay after the alert until discharge among survivors. SOFA scores were calculated using the worst measurements taken between 48 h and 72 h after the alert. For patients who died or were discharged before 72 h after the alert, the SOFA score was based on the worst measurements from the 24 h preceding death or discharge.



Figure 2.3:TREWS machine learning-based early warning system for sepsis.

CHAPTER 3

SYSTEM ANALYSIS

* 1. Existing Systems
* The existing system for detection of sepsis is done based on the diseased user input data, where the results are approximate and depend only on the genuinity of their input.
* At present only web apps are available with inconsistent dashboard and lack of storing huge patient history.
* Also the current system can only detect sepsis 6 hours before the clinical time.
* The existing proposed system depends on the biomarkers read through sensors and as well as user input. Here a comparative data analysis is done between the above data with sepsis disease symptoms dataset to extract accurate information for predicting disease.
* At present only web apps are available with inconsistent dashboard and lack of storing huge patient history. In the proposed system a UX is designed to get the demographic details and the premedical history of the user through the Mobile APP and pushed to cloud database..\

3.2 Proposed Systems

* Our proposed system uses a combination of biomarkers to diagnose sepsis. which can say the sepsis detection in effective manner.
* In our proposed system we are using the 34 parameters to detect the sepsis by analysis of dataset.
* It uses two types of input data:

1.the data that are collected from the patients.

2.And we are also using the data from sensor.

* Then by performing comparative smart data analysis between the acquired sensor data, user input with the gained parameters of the disease symptoms dataset.
* Automatic comparative data analysis is done between the biomarkers read through sensors, user input data with sepsis disease symptoms dataset to extract accurate information for predicting disease.
* The hidden information in the health care data set can be used for affective decision making to calculate the probability od sepsis.
* The accuracy of disease prediction is 84.5% and the system is able to give the risk associated with disease which is lower risk of sepsis or higher.

CHAPTER 4

SYSTEM DESIGN

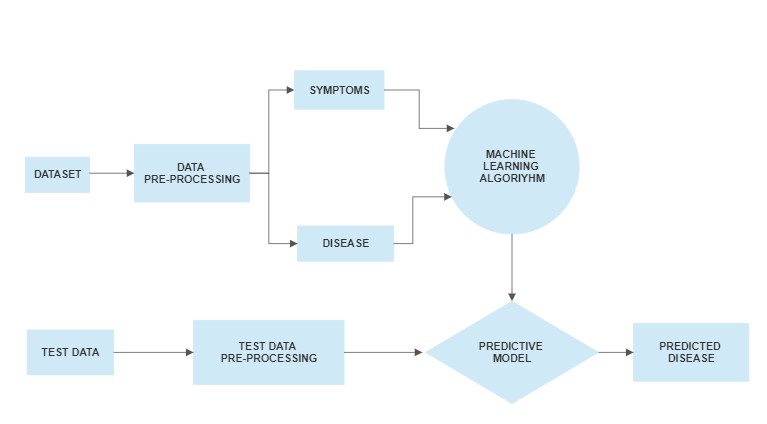
4.1 Machine Learning Model

Figure 4.1: Detailed Design Of Model

Data mining techniques are used in the project to see whether the dataset is good for prediction or not. Various data mining libraries used in the project are:

1. Scipy : This is used for implementing scientific computing in Python programming language. It is a collection of mathematical algorithms and convenience functions built on Numpy: Following are some of the functionalities it provides Special Functions (special), Integration (integrate), Optimization (optimize), Fourier Transforms (interpolate), Signal Processing (signal), Linear Algebra (linalg), Statistics (stats), File IO (io) etc. In this project stats (Statistics) library of this package is primarily used.

2. Sklearn: This stands for Scikit learn and is built on the Scipy package. It is the primary package being used in this project. It is used for providing an interface for supervised and unsupervised learning algorithms. Following groups of models are provided by sklearn Clustering, Cross Validation, Datasets, Dimensionality Reduction, Ensemble methods, Feature extraction, Feature selection, Parameter Tuning, Manifold Learning, Supervised Models.

3. Numpy : It is a library for the Python programming language, adding support for multidimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. It provides functions for Array Objects, Routines, Constants, Universal Functions, Packaging etc. In this project, it is used for performing multi-dimensional array operations

4. Pandas : This library is used to provide high-performance, easy-to-use data structures and data analysis tools for the Python programming language. It provides functionalities like table manipulations, creating plots, calculate summary statistics, reshape tables, combine data from tables, handle times series data, manipulate textual data etc. In this project it is used for reading csv files, comparing null and alternate hypothesis etc.

5. Matplotlib : It is a library for creating static, animated, and interactive visualizations in Python programming language. In this project it is used for creating simple plots, sub- plots and its object is used alongside with the seaborn object to employ certain functions such as show, grid etc. A %matplotlib inline function is also used for providing more concise plots right below the cells that create that plot.

6. Seaborn : It provides an interface for making graphs that are more attractive and interactive in nature. It is based on the matplotlib module. These graphs can be dynamic and are much more informative and easier to interpret. It provides different presentation formats for data such as Relational, Categorical, Distribution, Regression, Multiples and style and color of all these types. In this project, they are used for creating complex plots that use various attributes.

7. Warning : It is used for handling any warnings that may arise when the program is running. It is a subclass of Exception.

8. Stats : This library is used to incorporate statistics functionality in Python programming language. This library is included in the scipy package. This library is not directly used rather the required functions are directly imported as and when required i.e. for Measures of Central Tendency, Measures of Variability. The functions used can be for simple concepts like mean, median, mode, variance, percentiles, skew, kurtosis, range, Cumulative Distribution Function (CDF), Probability Distribution Function (PDF), stats (used for returning mean, variance, skew, kurtosis) etc, to complex hypothesis tests like chi2 contingency (used for chi-square test), testing(used for performing t-test), ks-2samp (used for performing Kolmogorov-Smirnov test) etc.

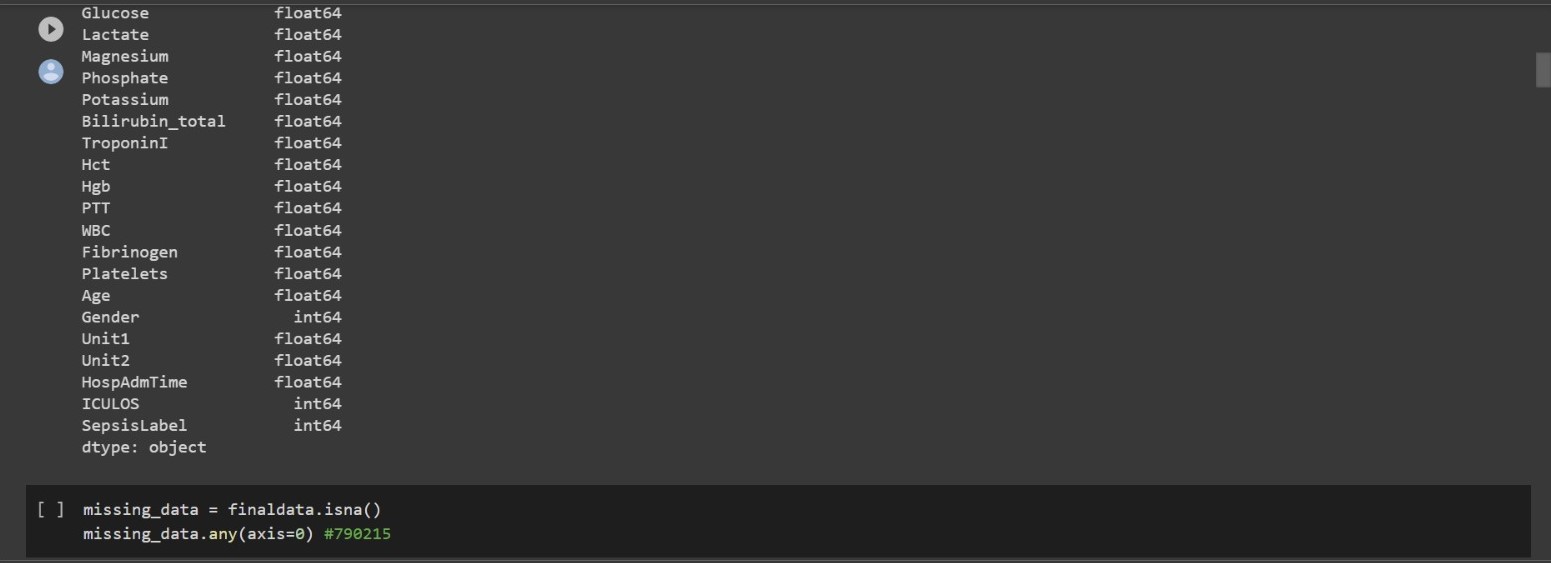
**CHAPTER 5**

**EXPERIMENTAL ANALYSIS**

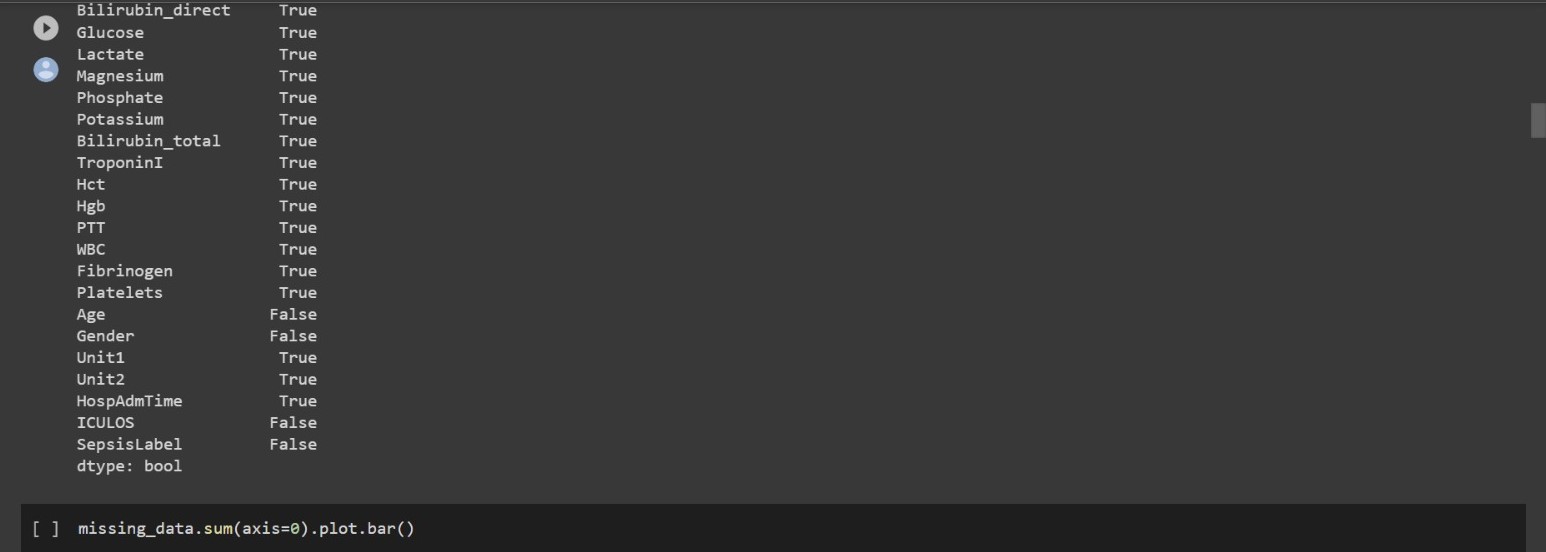
**5.1 Screen shots**

****

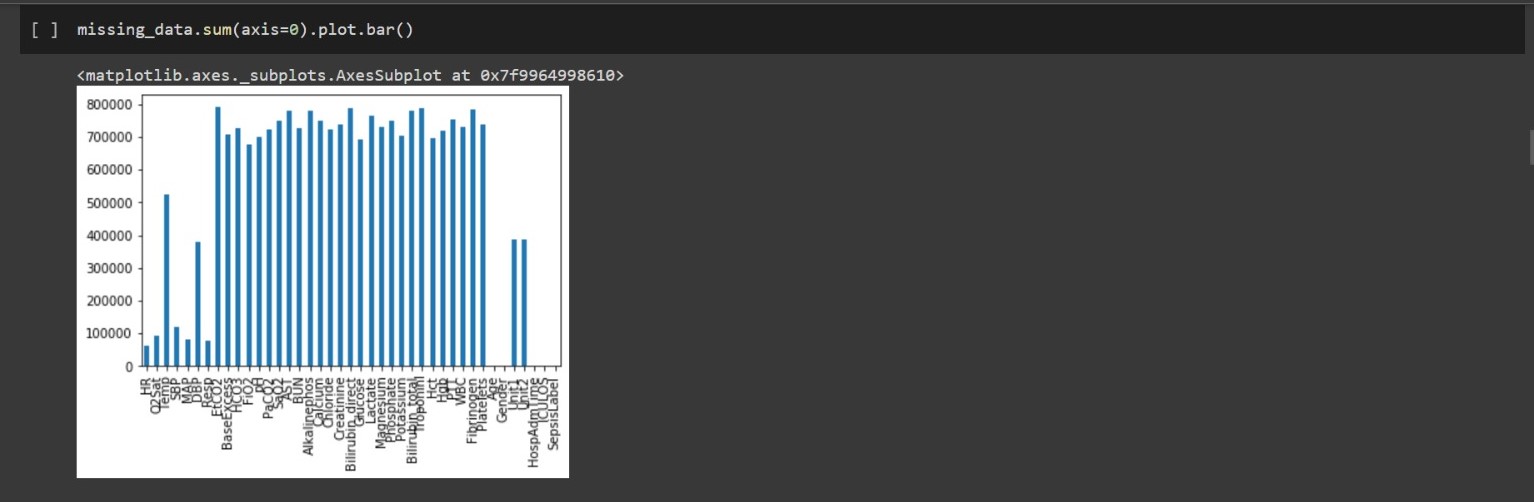
**Figure 5.1:**

****

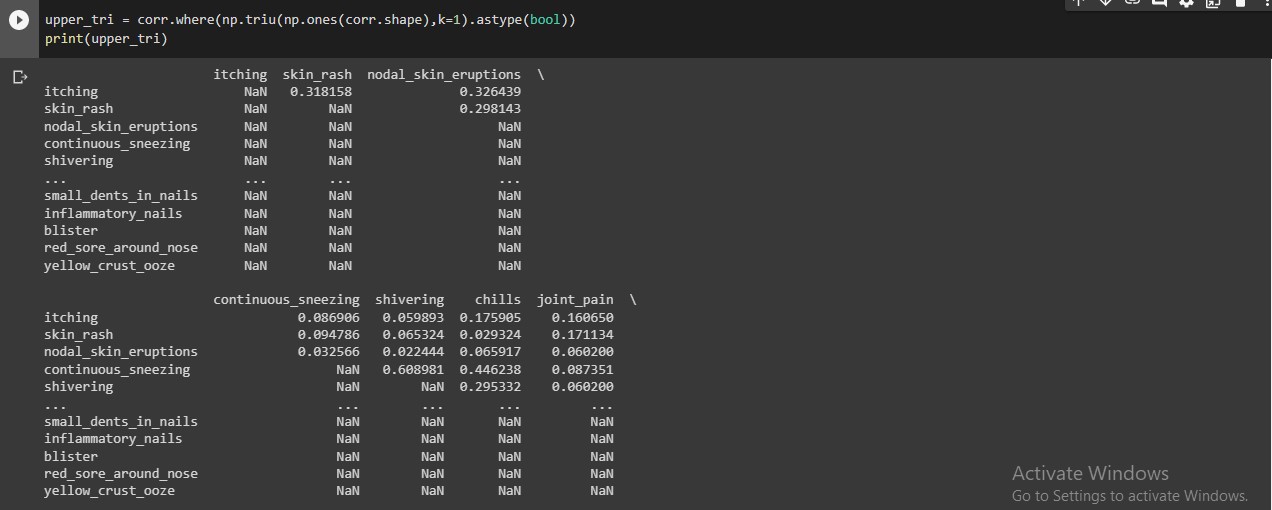
**Figure 5.2:**

****

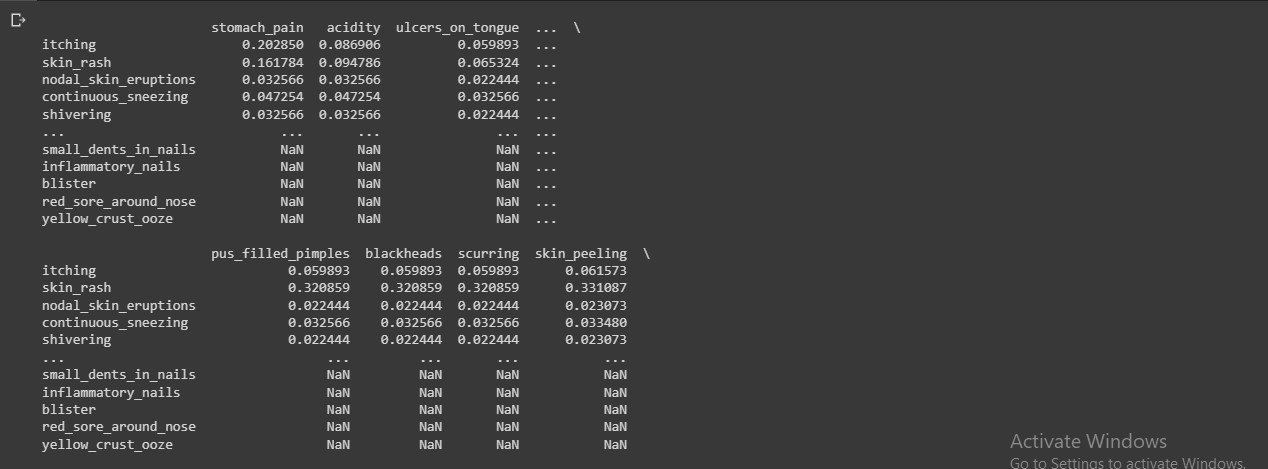
**Figure 5.3:**

****

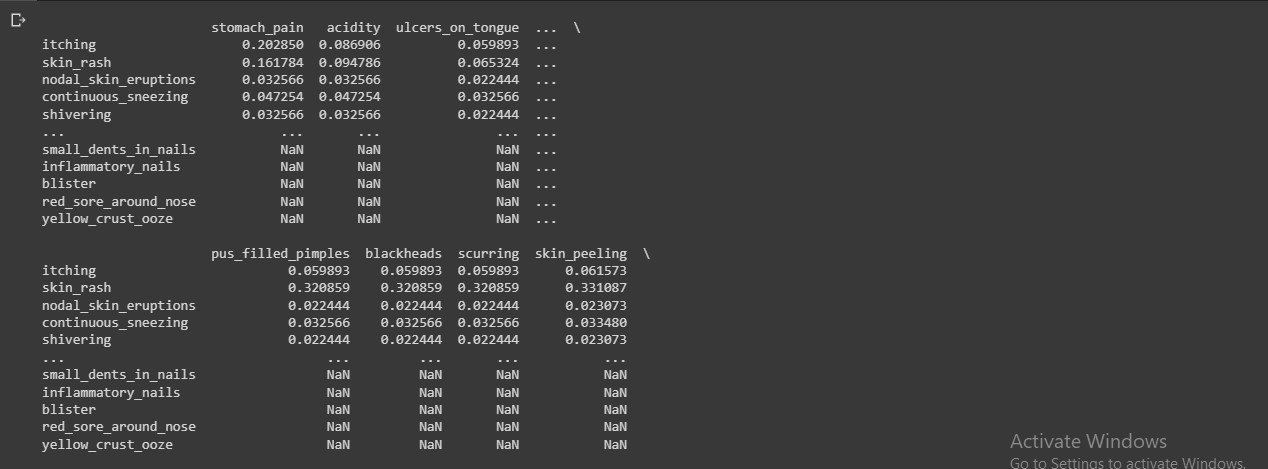
**Figure 5.4:**

****

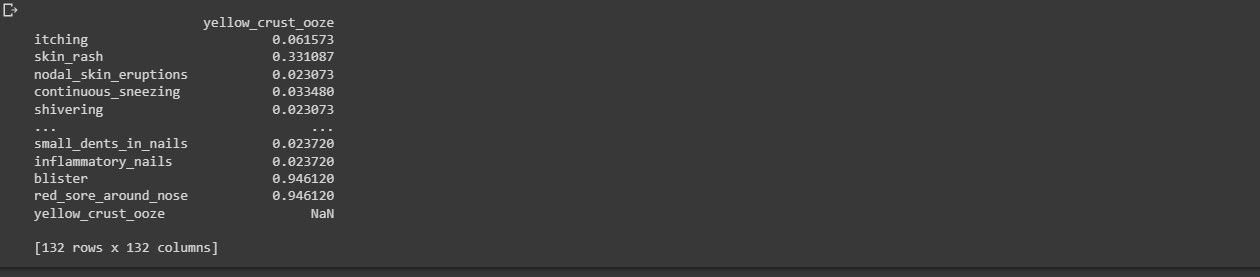
**Figure 5.5: Shaping the Training data**

****

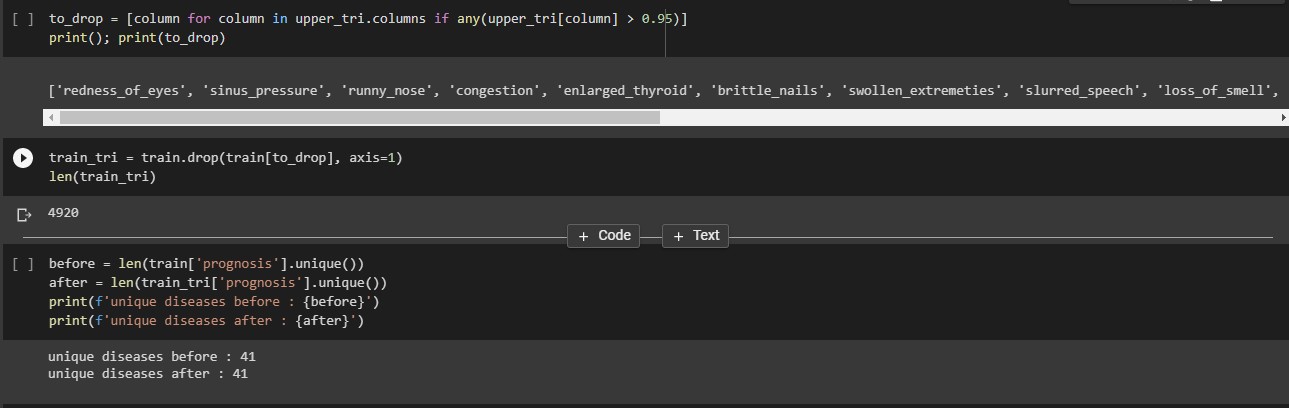
**Figure 5.6: Shaping the Training data**

****

**Figure 5.7: Shaping the Training data**

****

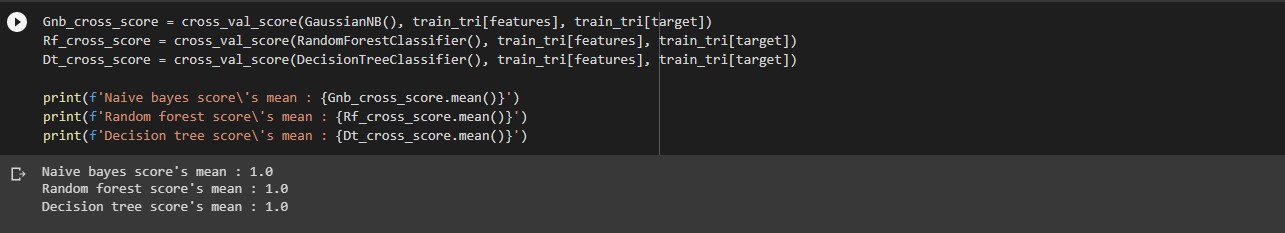
**Figure 5.8: Shaping the Training data**

****

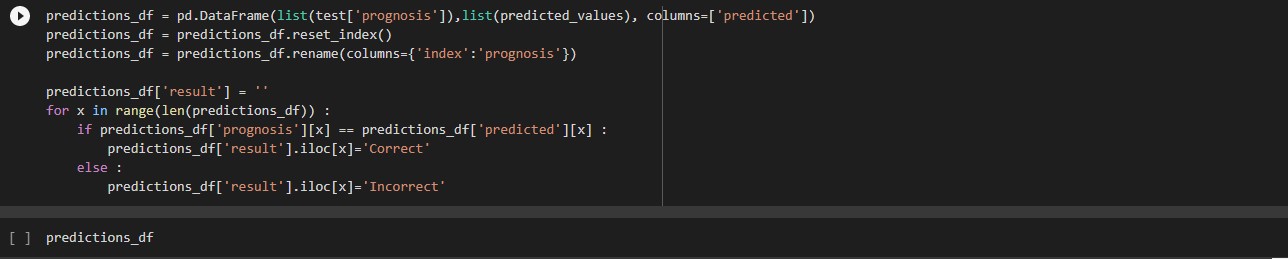
**Figure 5.9: Prognosis**

****

**Figure 5.10: Datatypes**

****

**Figure 5.11: Machine Learning Algorithms**

****

**Figures 5.12: Predictions**

**CHAPTER 6**

**CONCLUSION**

* In this study, we have showed that by using an LSTM network for early detection of septic shock, we can detect patients up to 20 hours earlier than a Cox proportional hazards model, at similar sensitivity and specificity, when the models are trained using the same features and target definitions. This finding is relevant since early detection and treatment of septic shock is essential for maximising the patient’s chance of survival.
* This paper proposes a machine learning model for early prediction and detection of sepsis in intensive care unit patients.
* First of all, the missing data are collected using the imputation process and applying matrix factorization to improve the model’s performance. Secondly, different models like SVM, RF, NB, LR, and XGBoost are developed using various machine learning packages.
* Then, the proposed ensemble method is proposed, which combines SVM, RF, NB, LR, and XGBoost.
* The proposed method delivered a good classification that improves the proposed performance.
* This model is beneficial to the patients admitted to the intensive care unit. This work can be extended by collecting geographical patient data to visualize more signs and symptoms of patients to feed more data to the machine learning model.
* The medical dataset contains a lot of missing data that could degrade the model performance; it requires more advanced data imputation techniques to handle this problem.
* Approaches were heterogeneous, but studies showed that utilizing both unstructured text and structured data in ML can improve identification and early detection of sepsis.
* The machine learning models demonstrated superior performance in prediction of sepsis diagnosis among emergency patients compared with that using the traditional screening tools. Further studies are needed to determine whether the models will enhance physicians’ judgments and improve patient outcomes.
* We have demonstrated the potential of machine-learning approaches for predicting outcome early in patients with sepsis. The SHAP method could improve the interpretability of machine-learning models and help clinicians better understand the reasoning behind the outcome.

**CHAPTER 7**

**FUTURE ENHANCEMENT**

* Today’s

**APPENDIX 1**

**INTEGRATED CODE**

import pandas as pd

import seaborn as sns

import numpy as np

import os

import glob as gb

import matplotlib.pyplot as plt

file\_path = "/home/milan/Desktop/Ys\_PCE/data/training\_setA/training/\*.psv" finaldata = pd.concat([pd.read\_csv(file,sep="|") for file in gb.glob(file\_path)])

finaldata.shape

(790215, 41)

finaldata.dtypes

HR float64

O2Sat float64

Temp float64

SBP float64

MAP float64

DBP float64

Resp float64

EtCO2 float64

BaseExcess float64

HCO3 float64

FiO2 float64

pH float64

PaCO2 float64

SaO2 float64

AST float64

BUN float64

Alkalinephos float64

Calcium float64

Chloride float64

Creatinine float64

Bilirubin\_direct float64

Glucose float64

Lactate float64

Magnesium float64

Phosphate float64

Potassium float64

Bilirubin\_total float64

TroponinI float64

Hct float64

Hgb float64

PTT float64

WBC float64

Fibrinogen float64

Platelets float64

Age float64

Gender int64

Unit1 float64

Unit2 float6 HospAdmTime float64

ICULOS int64

SepsisLabel int64

dtype: object

missing\_data = finaldata.isna()

missing\_data.any(axis=0) #790215

HR True

O2Sat True

Temp True

SBP True

MAP True

DBP True

Resp True

EtCO2 True

BaseExcess True

HCO3 True

FiO2 True

PaCO2 True

SaO2 True

AST True

BUN True

Alkalinephos True

Calcium True

Chloride True

Creatinine True

Bilirubin\_direct True

Glucose True

Lactate True

Magnesium True

Phosphate True

Potassium True

Bilirubin\_total True

TroponinI True

Hct True

Hgb True

PTT True

WBC True

Fibrinogen True

Platelets True

Age False

Gender False

Unit1 True

Unit2 True

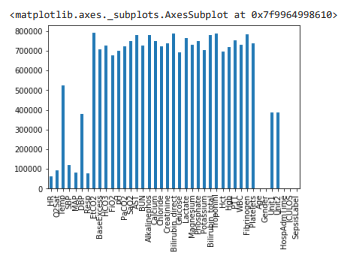
HospAdmTime True

ICULOS False

SepsisLabel False

dtype: bool

missing\_data.sum(axis=0).plot.bar()



missingpercentage = missing\_data.sum(axis=0)/finaldata.shape[0] missingpercentage

HR 0.077433

O2Sat 0.120320

Temp 0.662243

SBP 0.152112

MAP 0.102324

DBP 0.481258

Resp 0.097768

EtCO2 1.000000

BaseExcess 0.895749

HCO3 0.919494

FiO2 0.858070

pH 0.885329

PaCO2 0.912318

SaO2 0.950444

AST 0.985042

BUN 0.918407

Alkalinephos 0.985407

Calcium 0.950244

Chloride 0.916761

Creatinine 0.933579

Bilirubin\_direct 0.998504

Glucose 0.877684

Lactate 0.965651

Magnesium 0.922197

Phosphate 0.949512

Potassium 0.891376

Bilirubin\_total 0.987734

TroponinI 0.998779

Hct 0.882237

Hgb 0.911643

PTT 0.951525

WBC 0.924896

Fibrinogen 0.992369

Platelets 0.934829

Age 0.000000

Gender 0.000000

Unit1 0.488683

Unit2 0.488683

HospAdmTime 0.000010

ICULOS 0.000000

SepsisLabel 0.000000

dtype: float64

#rowwise na analysis

missing\_data.any(axis=1).sum()

790215

finaldata.shape

(790215, 41)

**APPENDIX 2**

**REFERENCES**

[1] Dahiwade, D., Patle, G., and Meshram, E. (2019). “Designing disease prediction

model using machine learning approach.” 2019 3rd International Conference on

Computing Methodologies and Communication (ICCMC), IEEE. 1211–1215.

[2] Fitriyani, N. L., Syafrudin, M., Alfian, G., and Rhee, J. (2019). “Development of

disease prediction model based on ensemble learning approach for diabetes and

hypertension.” IEEE Access, 7, 144777–144789.

[3] Hong, W., Xiong, Z., Zheng, N., and Weng, Y. (2019). “A medical-history-based

potential disease prediction algorithm.” IEEE Access, 7, 131094–131101.

[4] Kohli, P. S. and Arora, S. (2018). “Application of machine learning in disease

prediction.” 2018 4th International Conference on Computing Communication and

Automation (ICCCA), IEEE. 1–4.

[5] Krishnan.J, M. “Prediction of heart disease using machine learning algorithms.

[6] Patil, M., Lobo, V. B., Puranik, P., Pawaskar, A., Pai, A., and Mishra, R. (2018). “A Proposed model for lifestyle disease prediction using support vector machine.” 2018 9th International Conference on Computing, Communication and Networking Technologies(ICCCNT), IEEE. 1–6.

[7] Shobana, V. and Kumar, N. (2017). “A personalized recommendation engine for

prediction of disorders using big data analytics.” 2017 International Conference on

Innovations in Green Energy and Healthcare Technologies (IGEHT), IEEE. 1–4

[8] Priyanka Sonar, Prof. K. JayaMalini, “Diabetes Prediction Using Different Machine Learning Approaches”, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4 [9] Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, “Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm”, 2018 21st International Conference of Computer and Information Technology (ICCIT)

[10] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, “Diabetes Disease Prediction Using Data Mining”, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

[11] Zhilbert Tafa, Nerxhivane Pervetica, Bertran Karahoda, “An Intelligent System for Diabetes Prediction”, 4thMediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro