**PROJECT REPORT TEMPLATE**

1. Title of the project
   1. Introduction
2. Review of Literature
3. Data Collection
4. Methodology
   1. Exploratory Data Analysis
   2. Figures and Tables
   3. Statistical Techniques and Visualization
   4. Data Modelling and Visualization
5. Findings and Suggestions
6. Conclusion

**1. HEALTH MONITORING**

**1.1 INTRODUCTION:**

Monitoring patients in intensive care unit (ICU) is a critical task done by nurses and doctors. ICU is a special department in hospitals for patients who are suffering from dangerous diseases and having critical states. These patients need an intensive care from medical staff and they have to be always monitored using monitoring systems. Monitoring system is considered as a fundamental tool.

Monitoring system measures several medical parameters that indicate the state of patients. Each measured parameter has a threshold set by doctors. In case of having values exceeds their threshold an alarm is triggered. This alarm is an indication for nurses and doctors that state of patient is no more stable but very critical. As a result, this patient needs a particular treatment or his state will be more critical and he can even die. Unfortunately, sometimes there may be false alarms triggered by the current monitoring system. In fact, the monitoring system can trig alarms that do not indicate a real critical state but, in some cases they are due to a wrong setting of parameters, or a bad setting of monitoring devices. Besides, the monitoring systems do not take into account of the relation between the measured parameters. It separately measures each parameter which can lead to false alarms. Hence, false alarms present a real danger for the patient life. They do not report the real state of patients which can make the monitoring task more complicate. Furthermore, the working condition of the medical staff become more difficult .since, such alarms disturb them and make patients under more pressure.

As a result, avoiding these false alarms become more and more necessary. Several researches have focused in this problem and many works have been proposed to avoid decrease the high number of false alarms in intensive care unit. In this model we use machine learning algorithms to reduce high level of false alarms in intensive care unit (ICU).

**Machine learning** is an application of Artificial Intelligence that provides systems ability to learn automatically and improve from experience without being explicitly programmed. Now days we train and teach the computers to program themselves and to make decisions using data we provide. Let the data do work instead.

Machine learning works on statistical data. In machine learning we predict the future and works only with numeric data

**Features of Machine learning:**

* It uses data to detect patterns in a dataset and adjust programs accordingly.
* It focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.
* It enables computers to find hidden insights using iterative algorithms without being explicitly programmed.
* Machine Learning is a method of data analysis that automates analytical model building.

**Python libraries for data science:**

Many popular python tool boxes or libraries are available. Some of them are

* Numpy
* Pandas
* Scikit-learn
* Matplotlib
* Seaborn

Here Matplotlib and Seaborn are Visualization libraries

**Numpy:**

Numpy stands for numerical python. Numpy is used for all numerical computations in python. It is also used for metrics computations and vectorized computations. Here datasets works on arrays.

**Pandas**:

Perfect tool for data wrangling designed for quick and easy data manipulation, aggregation.

* Pandas also works for numeric Boolean and string. Pandas is an open source library.
* There are two data structures in pandas.

1. Series: Series is one dimensional array like object containing array of data and an array of labels called index.

2. Dataframes: Python dataframe is a data structure containing an ordered collection of columns. Dataframe has both row and column i.e., rows and columns.Each column may hold numeric string or Boolean values.

**Scikit learn Library:**

Scikit learn is simple and efficient. It focuses more on data mining and data analysis.It is built on **Numpy, Matplotlib, and open source.**

**Matplotlib**:

Matplotlib is the father library of data visualization with python. It was created by John Hunter. It is an excellent 2D and 3D graphics library for generating scientific figures. It enables you to make bar charts, scatter plots, line charts, histograms, pie-charts, contour plots, and quiver plots.

**Seaborn Library:**

It is focussed on visual of statistical models which include heat maps and depict overall distributions.

**Process of performing machine algorithm is**

Step-1: Collection of data

Step-2: Data Wrangling

Step-3: Analyse model

Step-4: Train algorithm

Step-5: Test Algorithm

Step-6: Deployment

**2. Literature Review:**

**Health monitoring** means monitoring different medical parameters such as systolic blood pressure, diastolic blood pressure, temperature of the body, pulse, etc….

This project describes the steps of building efficient monitoring system in intensive care unit (ICU) . It deals with large data sets and solves the main problems of current monitoring system. In fact the current monitoring system in ICU has many issues to detect real states of patient’s namely critical and normal states. It frequently generates a high number of false alarms having bad effects on the working conditions. These alarms can threat the patient life by misleading medical staff.

Our aim is to avoid false alarms and keep a high level of sensitivity by improving the current monitoring system. Now we have used Generated model which is based on **Logistic Regression**. It is a classification type algorithm. Logistic Regression has produced the best results compared to the other algorithms and based on different evaluation criteria. Logistic Regression deals with large data sets to generate the actual output. By using the general model there may be chance to get false output. This false output can be rectified by using this Logistic Regression algorithm. All experimental results using real medical databases have been analysed and have proved the performance of Logistic Regression.

**3. DATA COLLECTION:**

**Dataset:**

A data set (or dataset) is collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows.

Dataset used in this project is collecting from different sensors connected to different patients having similar disease.

**The Attributes that are used in our dataset are:-**

* **Systolic Blood Pressure:**

It refers to the amount of **pressure** in your arteries during the contraction of your heart muscle.

* **Diastolic Blood Pressure:**

Diastolic blood pressure (the bottom number or lowest blood pressure between heart beats) gives the best idea of your risk of having a stroke or heart attack.

* **Pulse**:

 A pulse represents the tactile arterial palpation of the heartbeat by trained fingertips.

* **Temperature:**

Normal human body temperature, also known as normothermia or euthermia, is the typical temperature range found in humans.

**4. METHODOLOGY**

Regression and classification are categorized under the same umbrella of supervised machine learning. Both share the same concept of utilizing known datasets (referred to as training datasets) to make predictions.

**Prediction in Logistic Regression algorithm:**

Logistic regression is another technique borrowed by machine learning from the field of statistics. Logistic Regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome. Outcome is a binary class type.

y = b0+b1\*x

P = 1/1+e^-y (Sigmoid function)

Ln (p/1-p) = b1+b1\*x

**4.1 Exploratory Data Analysis:**

Exploratory data analysis (EDA) is an approach to analysing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA), which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA. The objectives of EDA are to:

* Suggest hypotheses about the causes of observed phenomena
* Assess assumptions on which statistical inference will be based
* Support the selection of appropriate statistical tools and techniques
* Provide a basis for further data collection through surveys or experiments.

**4.1.1 Figures and Tables:**

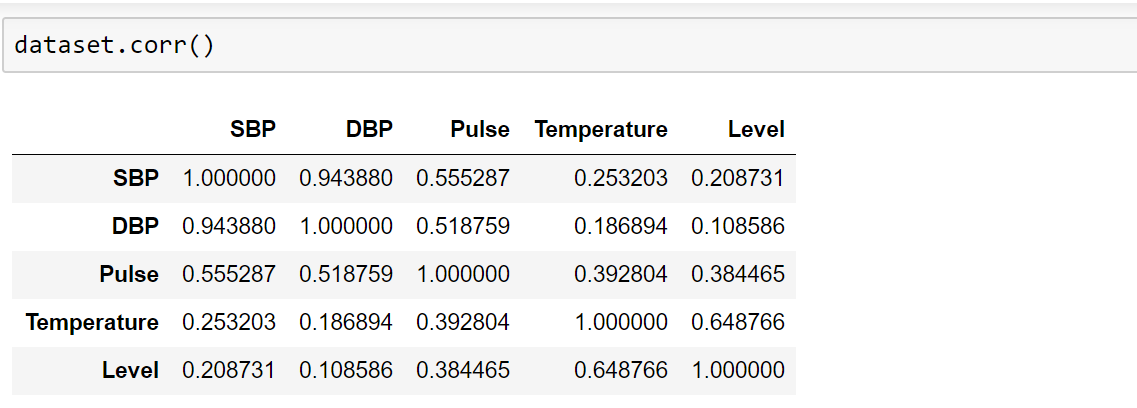
****

Fig 4.1.1(a): Correlation table

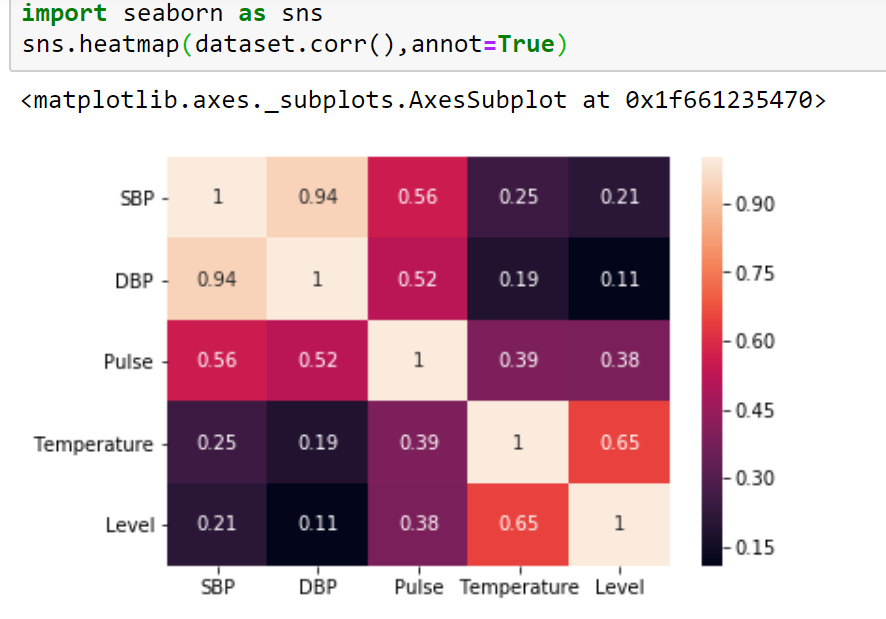
****

Fig 4.1.1(b): Seaborn

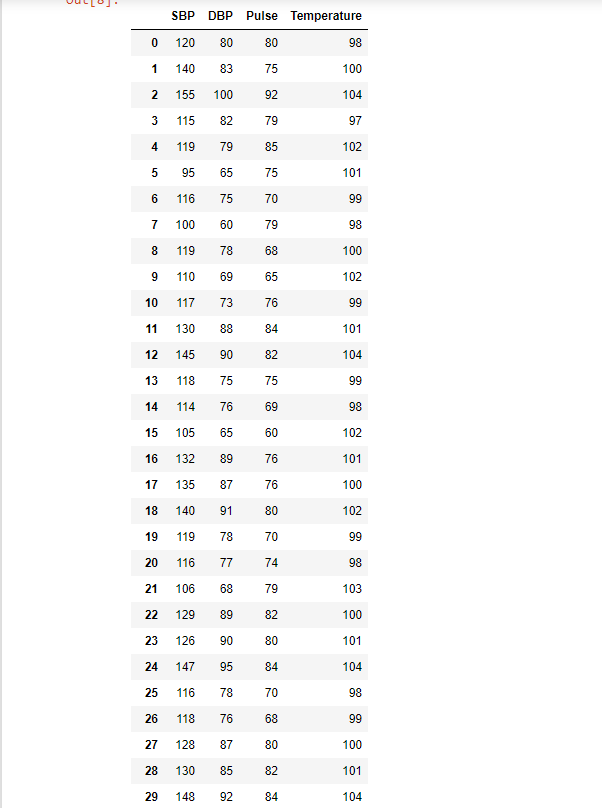
****

Fig 4.1.1(c) Sample Dataset

**4.2 Statistical Techniques and Visualization:**

Statistics is a collection of tools that you can use to get answers to important questions about data. You can use descriptive statistical methods to transform raw observations into information that you can understand and share. You can use inferential statistical methods to reason from small samples of data to whole domains. Statistics is a pillar of machine learning. You cannot develop a deep understanding and application of machine learning without it.

* Problem Framing: Requires the use of exploratory data analysis and data mining.
* Data Understanding: Requires the use of summary statistics and data visualization.
* Data Cleaning. Requires the use of outlier detection, imputation and more.
* Data Selection. Requires the use of data sampling and feature selection methods.
* Data Preparation. Requires the use of data transforms, scaling, encoding and much more. Model Evaluation. Requires experimental design and resampling methods.
* Model Configuration. Requires the use of statistical hypothesis tests and estimation statistics. Model Selection. Requires the use of statistical hypothesis tests and estimation statistics. Model Presentation. Requires the use of estimation statistics such as confidence intervals. Model Predictions. Requires the use of estimation statistics such as prediction intervals.

**NumPy** is a commonly used Python data analysis package. By using NumPy, you can speed up your workflow, and interface with other packages in the Python ecosystem, like scikitlearn, that use NumPy under the hood. NumPy was originally developed in the mid-2000s, and arose from an even older package called Numeric. This longevity means that almost every data analysis or machine learning package for Python leverages NumPy in some way.

* Import the numpy package.
* Pass the list of lists wines into the array function, which converts it into a NumPy array.
* Exclude the header row with list slicing.
* Specify the keyword argument dtype to make sure each element is converted to a float.

**Pandas** - is an open source python library that is built on top of NumPy. It allows you do fast analysis as well as data cleaning and preparation. Pandas is hands down one of the best libraries of python. It supports reading and writing excelspreadsheets, CVS's and a whole lot of manipulation. It is more like a mandatory library you need to know if you‘re dealing with datasets from excel files and CSV files. i.e., for Machine learning and data science. This is part one of Pandas tutorial. I‘m not going to cover everything possible with pandas, however, I want to give you a taste of what it is and how you can get started with it. This tutorial is going to be super short just introducing you to Series object of pandas. As other libraries, you‘d import pandas and reference it as **Pd**

**Matplotlib** - is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib. Matplotlib was originally written by John D. Hunter, has an active development community, and is distributed under a BSD-style license. Michael Droettboom was nominated as matplotlib's lead developer shortly before John Hunter's death in August 2012, and further joined by Thomas Caswell. As of 23 June 2017, matplotlib 2.0.x supports Python versions 2.7 through 3.6. Matplotlib 1.2 is the first version of matplotlib to support Python 3.x. Matplotlib 1.4 is the last version of Matplotlib to support Python 2.6. Matplotlib has pledged to not support Python 2 past 2020 by signing the Python 3 Statement. Pyplot is a Matplotlib module which provides a MATLAB-like interface.

Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source. Several toolkits are available which extend Matplotlib functionality. Some are separate downloads, others ship with the Matplotlib source code but have external dependencies.

**Plotting from a script** - If you are using Matplotlib from within a script, the function plt.show() is your friend. plt.show() starts an event loop, looks for all currently active figure objects, and opens one or more interactive windows that display your figure or figures. The plt.show () command does a lot under the hood, as it must interact with your system's interactive graphical backend. The details of this operation can vary greatly from system to system and even installation to installation, but matplotlib does its best to hide all these details from you.

**4.3 Data Modelling and visualization:**

Node-RED is a flow-based development tool for visual programming developed originally by IBM for wiring together hardware devices, APIs and online services as part of the Internet of Things. Node-RED provides a web browser-based flow editor, which can be used to create JavaScript functions. Elements of applications can be saved or shared for re-use. The runtime is built on Node.js. The flows created in Node-RED are stored using JSON. Since version 0.14 MQTT nodes can make properly configured TLS connections.

In 2016, IBM contributed Node-RED as an open source JS Foundation project.

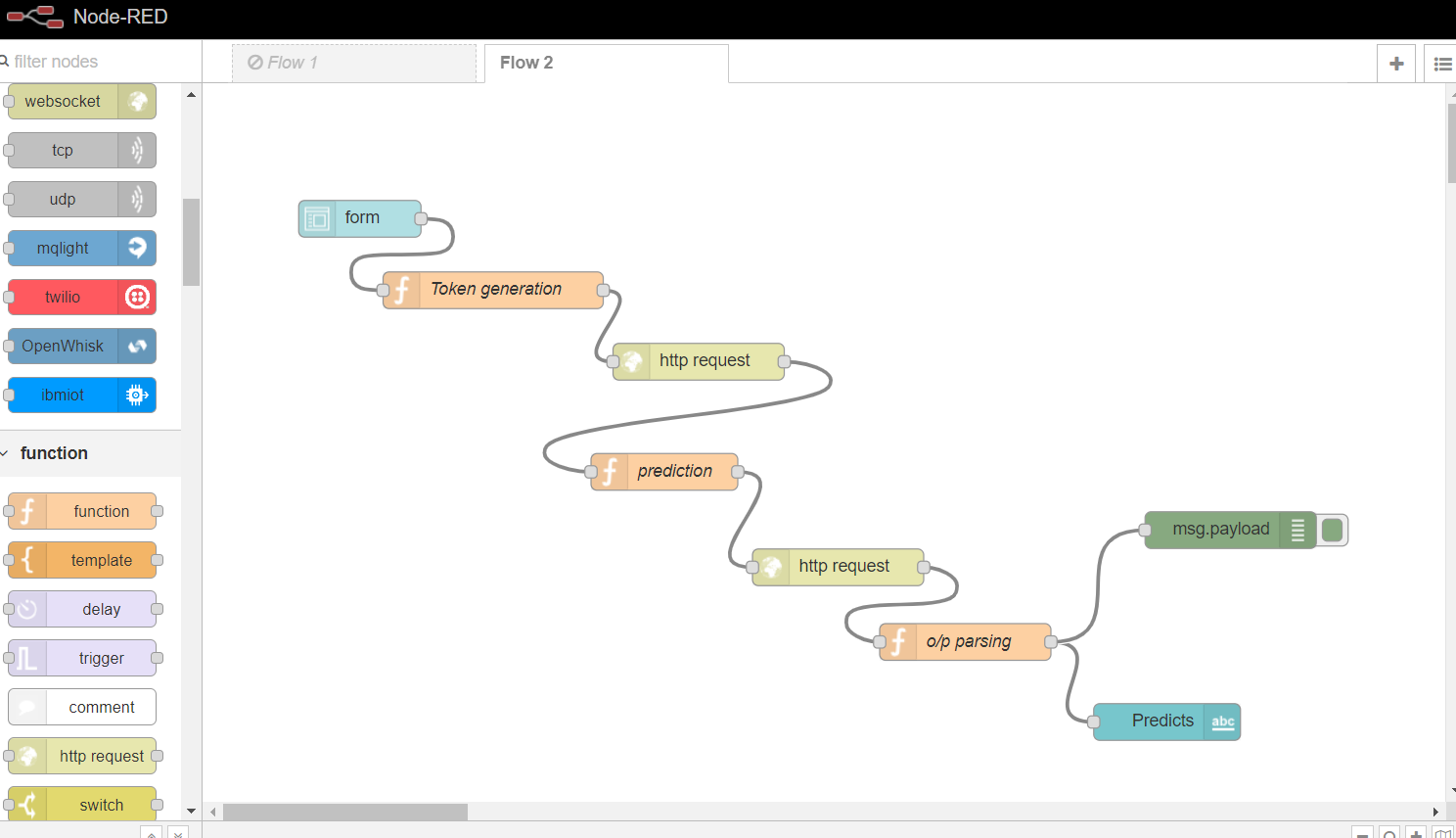


Fig 4.3.1: Node-Red Flow

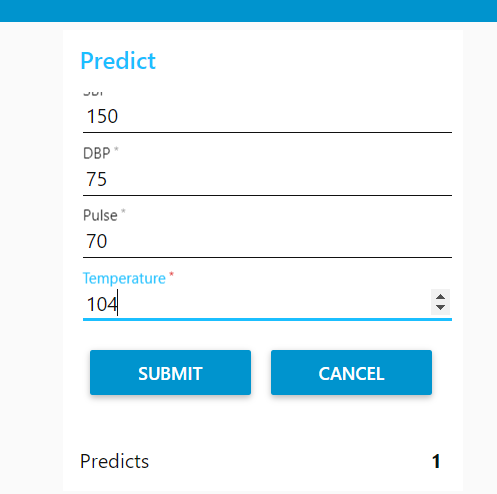


Fig 4.3.2: User Interface

**5. FINDINGS AND SUGGESTIONS**

**Findings**:

The Health Monitoring helps the doctors to know about the medical parameters of the patients. We predict the future health rate based on the data extracted from previous datasets. The health monitoring will help the doctors for predicting the health analysis of the patients.

**Suggestions:**

With predictive analytics, you can get the actual medical parameters. Instead of Logistic Regression, there are many number of possible algorithms not yet mentioned, including:

* K Nearest Neighbours
* Support Vector Machine
* Decision Tree
* Random Forest

**6. CONCLUSION**

With the help of machine learning technology, it has become easy to find out relation and patterns among various data‘s. The work in this project mainly revolves around predicting the health problem which may occur in future. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation using Logistic Regression Algorithm. The model predicts the health condition and Data visualization helps in analysis of data set and prediction of crimes. The graphs include bar, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding different health monitoring datasets that can help in capturing the factors that can help in keeping a person healthy.