PREDICTION OF REQUIREMENT OF INTUBATION

**Data Analytics**

**Semester 7**

**BE 2**

**BY:**

**Aakash Kale 4241**

**Anuj Kanetkar 4242**

**Shriniwas Nayak 4256**

Contents

[Acknowledgement 2](#_Toc21903363)

[Problem Statement 3](#_Toc21903364)

[Abstract 3](#_Toc21903365)

[Objective 4](#_Toc21903366)

[Introduction 4](#_Toc21903367)

[Project Development Outline 4](#_Toc21903368)

[Scope 5](#_Toc21903369)

[Functions 6](#_Toc21903370)

[Test Cases 6](#_Toc21903371)

[Results 7](#_Toc21903372)

[CONCLUSION 8](#_Toc21903373)

[**REFERENCES** 9](#_Toc21903374)

# Acknowledgement

We Akash Kale (4241), Anuj Kanetkar (4242), Shriniwas Nayak (4256) would like to extend our sincere gratitude towards our Data Analytics Lab Lecturer Prof HP Channe.

It is because of her guidance and support that we have been able to complete the project in time and as per the requirements.

We take this opportunity to thank our college Pune Institute of Computer Technology (PICT) for providing all the required infrastructure and facilities required for the completion of the project.

We would also like to thank our family and friends for their love and support without which the project could not have been completed.

# Problem Statement

Data related to patients admitted for op compound poisoning is available but no actionable insight can be obtained from it yet, condition of patients can be improved if it can be predicted in advance if the patient will require intubation in future. The data may not be complete and may require pre-processing.

# Abstract

Data mining is used widely today to provide solutions to many problems, increase profits, improve processes and at times for research and development. Data mining are particularly useful in this example as the presented data can be used for predicting the outcome for a patient using supervised learning techniques.

The project can be broadly classified into these three categories:

* Data Analysis
* Data Cleaning and Pre-processing
* Data Classification

The project uses three classification algorithms and provides with accuracy obtained from each of these. Libraries used for the same include numpy, scikitlearn, matplotlib amongst others.

The project is built in modular fashion to help in dynamic and plug and play programming, the project is capable of analysing any other similar file however pre-processing is data specific.

# Objective

The mini project aims to attain the following objectives:

1. Clean data by using various data cleaning techniques
2. Pre-process the data so that it can be used for analysis and statistical computation
3. Analyse provided data present data in readable format for any user
4. Visualize data in order to assess the spread and density of data
5. Classify the data as per the labels and predict the outcome for similar data

# Introduction

Today use of fertilizers and pesticides in the field of farming is rampant and also excessive at times. Farmers and other workers involved in the supply chain and utility chain of these fertilizers are being exposed to dangerous levels of potentially hazardous chemicals and as a reason of which prove to be at risk of a number of related diseases. Namely op compound poisoning can be identified as a main reason for many ailments. Data related to these patients can help to analyse the effect of the same on different class of patients thereby providing an insight into the nature of the problems caused and help to reduce their effect if not eliminate them completely.

Classification of the data using different techniques can help to identify if a patient will require intubation in future. Intubation is a standard procedure that involving a passing a tube into tracheal track. Ability to predict this in future will help to know the chances of survival of the patient.

# Project Development Outline

For the successful implementation of the project we undertook the following activities:

1. Brainstorming for project field selection
2. Requirement gathering for the selected topic
3. Requirement analysis of the collected requirements
4. Finalizing requirements and problem statement.
5. Brainstorming for data analysis techniques
6. Incorporation of changes
7. Implementation of individual modules
8. Testing of individual modules
9. Exception handling in individual modules
10. Integration of modules
11. Integration testing
12. Integrated Exception handling
13. Final testing
14. Project Presentation
15. Project Report Submission

# Scope

The project deals with patient data but as per the instructions of the source of the data names and registration ids of the patients have not been used in the project and no group member had access to them during any time of the project.

The output of the project consists of these:

* Analysis Reports
* Pre-processing Reports
* Classification Reports
* Confusion Matrix
* Histograms

The project can be further expanded into a full fledge system by adding following functionality:

1. Automated report mailing to concerned doctor
2. Creating database and combing data from different centres
3. Analyse demographic effects using current reports

# Functions

The project uses following functions to achieve the required objectives:

* isFloat()
* writeData()
* cleanData()
* readFile()
* findMean()
* findMode()
* findMedian()
* findQuartile()
* findStandardDeviation()
* analyze()
* preprocess()
* naivebayes()
* SVM()
* DecisionTree()

# Test Cases

|  |  |  |
| --- | --- | --- |
| INPUT | OUTPUT | RESULT |
|  |  |  |
| |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | -1.6102 | 2.475113 | 1.878729 | -0.88491 | 0 | -2.82761 | 2 | 1.119648 | 2.030995 | 1.898731 | | **Yes** | **Pass** |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 0.571204 | -0.2451 | -0.78142 | 2.317695 | 1 | 0.740692 | 0 | -1.37861 | -1.01259 | 0.750883 | 0 | | **No** | **Pass** |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 0.328826 | -0.92516 | -0.78142 | 0.525588 | 1 | 1.065083 | 0 | 0.388449 | -1.38527 | 0.750883 | 0 | | **No** | **Pass** |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | -1.6102 | -0.2451 | 0.548655 | -0.71729 | 1 | -0.55687 | 1 | 1.363381 | 0.850831 | 0.062175 | 1 | | **Yes** | **Pass** |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 1.783094 | 1.455032 | 0.548655 | -0.92602 | 1 | -0.23248 | 1 | 0.02285 | -0.79519 | -1.08567 | 1 | | **Yes** | **Pass** |

# Results

**========================================================**

*Naive Bayes :*

*precision recall f1-score support*

*0 0.82 1.00 0.90 9*

*1 1.00 0.87 0.93 15*

*accuracy 0.92 24*

*macro avg 0.91 0.93 0.91 24*

*weighted avg 0.93 0.92 0.92 24*

*SVM :*

*precision recall f1-score support*

*0 0.90 1.00 0.95 9*

*1 1.00 0.93 0.97 15*

*accuracy 0.96 24*

*macro avg 0.95 0.97 0.96 24*

*weighted avg 0.96 0.96 0.96 24*

*Decision Tree :*

*precision recall f1-score support*

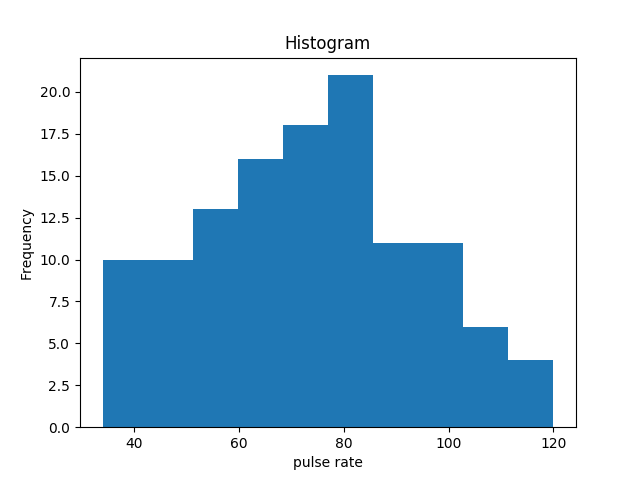
*0 0.69 1.00 0.82 9*

*1 1.00 0.73 0.85 15*

*accuracy 0.83 24*

*macro avg 0.85 0.87 0.83 24*

*weighted avg 0.88 0.83 0.84 24*



Pulse rate Histogram

# Conclusion

We have successfully completed the project with the required functionalities. We have also studied the future prospects of the project. Created required visualizations for any user and generated analysis reports.

# 

# **REFERENCES**

[1] <https://scikit-learn.org/> (Date visited : 13/9/19 time : 15:30)

[2] <https://matplotlib.org/> (Date visited : 15/9/19 time : 15:30)

[3] <https://en.wikipedia.org/wiki/Data_analysis> (Date visited : 11/9/19 time : 17:30)