# 1.0 Problem Definition

Triage is the primarily risk assessment and a form of prioritization delegating task of patient care based on the illness/ injury, severity, prognosis, and resource availability. The purpose of triage is to identify patients that need an immediate attention and medical care.

The primarily risk assessment focuses on the assessment of the patients’ risk based on several metrics such as age, gender, chest pain, blood pressure, cholesterol, max heart rate, exercise angina, plasma glucose, skin thickness, insulin, BMI, diabetes pedigree, hypertension, heart disease, residence type, smoking status and categorizes them into level of triage whereby blue is denoting as least severity and does not require an immediate sought of medical attention whereas red is denoting as the most severe that required immediate attention from physician. The ordinal level goes from least to most severe by tagging with blue, green, yellow, orange and red.

Logically says, patient considered to be at low risk for the evaluation if in young age, no chest pain, normal range of blood pressure, cholesterol level, heart rate, insulin level, BMI etc. The correlation of these metrics is in ambiguity and therefore, there is a need to further exploit and understand the pre-set dataset to well define the triage categorization.

## 1.1 Research Goal

* The objective is to assess the relationship between the attributes (metrics) and the triage categorization.

## 1.2 Objective

* Identify the significant metric that has the pivotal role in triage
* Investigate the relationship between the attributes and the triage categorization by using statistical hypothesis testing.

# 2.0 Dataset Characteristics

The Triage dataset is a multivariate, and the associated task is a multiclass classification. It consists of 5109 of instances and 17 attributes. Table 2.0.1 depicts a total of 9 qualitative attributes which subdivided into ordinal and nominal and 8 quantitative attributes which subdivided into discrete and continuous.

Table 2.0.1 Description of the dataset

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attribute | Description | Datatype |
| 1 | Age | Age of patient | Quantitative – discrete |
| 2 | Gender | Sex of patient [“1” (male),”0” (female)] | Qualitative – binary |
| 3 | Chest pain type | [“0” (no chest pain), “1” (typical angina), “2” (atypical angina), “3” (non-anginal pain), “4” (asymptomatic)] | Qualitative – nominal |
| 4 | Blood pressure | mmHg | Quantitative – discrete |
| 5 | Cholesterol | mg/ DL | Quantitative – discrete |
| 6 | Max heart rate | Beats per minute (bpm) | Quantitative – discrete |
| 7 | Exercise Angina | [“0” (no angina), “1” (present angina)] | Qualitative – binary |
| 8 | Plasma glucose | mg/ DL | Quantitative – discrete |
| 9 | Skin\_thickness | mm | Quantitative – discrete |
| 10 | Insulin | pmol/L | Quantitative – discrete |
| 11 | BMI | Body mass index | Quantitative – discrete |
| 12 | Diabetes\_pedigree | Scores likelihood of diabetes based on family history | Quantitative – continuous |
| 13 | Hypertension | [“0” (no hypertension), “1” (present hypertension)] | Qualitative – binary |
| 14 | Heart\_disease | [“0” (no heart disease), “1” (present heart disease)] | Qualitative – binary |
| 15 | Residence\_type | [“urban”, “rural”] | Qualitative – binary |
| 16 | Smoking\_status | [“never smoked’, “formerly smoked”, “smoked”, “unknown”] | Qualitative – nominal |
| 17 | Triage | Classification of emergency patients according to their urgencies [“Red’ (immediate evaluation by physician), “Orange” (emergent), “Yellow” (potentially unstable), “Green” (non-urgent), “Blue” (minor injuries or complaints)] | Qualitative – ordinal |