|  |
| --- |
|  |

|  |
| --- |
| **KPI Design Document** |
| Heart Failure |
|  |
| A chronic condition in which the heart doesn't pump blood as well as it should. This Document will help understanding the heart condition of the person who is being examined or diagnosed. With our understanding on the given data, we can aware the diagnosed person or their families about their health condition so that they can take proper precautions or treatments. |
|  |
| **Tanu Roy(DBA20013)** |
| **7/18/2021** |
|  |

Table of Contents

[Overview 2](#_Toc76344314)

[Introduction 2](#_Toc76344315)

[High-Level Scope 2](#_Toc76344316)

[KPI 2](#_Toc76344317)

[Death Events 2](#_Toc76344318)

[Business Requirement 2](#_Toc76344319)

[Data Sources 2](#_Toc76344320)

[Business Logic in ML 2](#_Toc76344321)

[Conclusion 2](#_Toc76344323)

# Overview

## Introduction:

Heart failure can occur if the heart cannot pump (systolic) or fill (diastolic) adequately.

Symptoms include shortness of breath, fatigue, swollen legs and rapid heartbeat.

Treatments can include eating less salt, limiting fluid intake and taking prescription medication. In some cases a defibrillator or pacemaker may be implanted. Treatment depends on the severity and cause of the disease. In people with chronic stable mild heart failure, treatment commonly consists of lifestyle modifications such as [stopping smoking](https://en.wikipedia.org/wiki/Smoking_cessation), [physical exercise](https://en.wikipedia.org/wiki/Physical_exercise), and dietary changes, as well as medications.

Older are more likely than younger people to suffer from cardiovascular disease, which is problems with the heart, blood vessels, or both. **Aging** can cause changes in the heart and blood vessels that may increase a person's risk of developing cardiovascular disease.

The **Anaemia** itself can worsen cardiac function, both because it causes cardiac stress through tachycardia and increased stroke volume, and because it can cause a reduced renal blood flow and fluid retention, adding further stress to the heart.

**Creatine Phosphokinase** or **CPK** is also a factor which can results in Heart Failure. When the total CPK level is very high, it most often means there has been injury or stress to muscle tissue, the heart, or the brain. Muscle tissue injury is most likely. When a muscle is damaged, CPK leaks into the bloodstream. Finding which specific form of CPK is high helps determine which tissue has been damaged.

**High blood pressure** increases the force of blood through your arteries and can damage artery walls. Having both high blood pressure and **diabetes** can greatly increase your risk for heart disease.

A normal **Ejection Fraction** doesn't always mean your heart is healthy. You could have heart failure with preserved ejection fraction (HFpEF). It happens when your heart muscle thickens to the point that the left ventricle holds less than the usual amount of blood.

**Platelets** can detect a disruption in the lining of a blood vessel and react to build a wall to stop bleeding. **Platelets** form a **platelet** plug to stop bleeding from an injured blood vessel. In **cardiovascular disease**, abnormal clotting occurs that can result in heart attacks or stroke.

A significant subset of patients with **heart failure** experience small to moderate rise in **serum creatinine** (RSC) in the setting of otherwise beneficial therapies such as aggressive diuresis or renin-angiotensin-aldosterone system (RAAS) inhibition.

**Congestive heart failure** (CHF) causes a decrease in cardiac output and circulating blood volume, which in turn triggers a compensatory response aimed at preserving blood pressure. This stimulates the body to retain both water and **sodium.**

Although men tend to develop coronary artery disease earlier in life, after age 65 the risk of heart disease in women is almost the same as in men. Women have many of the same risk factors for **heart disease** as men.

The buildup of plaque also makes it more likely that blood clots will form in your arteries. Blood clots can partially or completely block blood flow. Over time, **smoking** contributes to atherosclerosis and increases your risk of having and dying from **heart disease**, **heart failure**, or a **heart attack**.

Patients with **heart failure** are at high risk of readmission to hospital—recent data suggest that almost a third of the **patients** discharged alive after an episode of decompensation are readmitted because of the same problem within 6 months.

## High Level Scope:

We will be looking forward to do an exploratory data analysis in PYTHON using Regression to get the factor which affects the most and Machine Learning models such as Confusion matrix, Random Forest, Bagging, Boosting and many more to see the accurate model and predict the results on new data. At last we have to fill some KPI reports.

# KPI

## Death Events:

### Business Requirements:

Our goal is to reduce the number of Death Events by analysing the factors such as smoking, creatinine phosphokinase, anaemia and take precautions. Some factors are non-changable such as Age, gender and in such scenario the patient will be advised accordingly if the factors affect the death event. With the help of Machine Learning algorithm and Regression and we will come to a conclusion which patient is more likely to get a Heart Attack next time if the precautions were not taken care of.

### Data Sources:

We have an authentic data-table from <http://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records> on Heart Failure Clinical Records dataset with 300 records and 12 attributes.

### Business Logic(ML):

import numpy as np

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import confusion\_matrix

import statsmodels.api as sm

import seaborn as sns

from sklearn.metrics import confusion\_matrix

from sklearn.model\_selection import KFold

from sklearn.model\_selection import cross\_val\_score

from sklearn.ensemble import AdaBoostClassifier

from sklearn.linear\_model import SGDClassifier

from sklearn.ensemble import BaggingClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.linear\_model import LogisticRegression

from sklearn.feature\_selection import SelectKBest

from sklearn.feature\_selection import chi2

import matplotlib.pyplot as plt

from sklearn.metrics import confusion\_matrix, accuracy\_score

df=pd.read\_csv('/content/heart\_failure\_clinical\_records\_dataset.csv')

train = df.iloc[:,:-1]

target = df.iloc[:,-1]

print(train,target)

df.describe()

df.isna().sum()

X = df.iloc[:,:-1]

Y = df.iloc[:,-1]

print(X,Y)

X.columns

bestfeatures = SelectKBest(score\_func=chi2)

fit = bestfeatures.fit(X,Y)

df1scores = pd.DataFrame(fit.scores\_)

df1columns = pd.DataFrame(X.columns)

featureScores = pd.concat([df1columns,df1scores],axis=1)

featureScores.columns = ['Specs','Score']  #naming the dataframe columns

print(featureScores.nlargest(25,'Score'))  #print best and top best features

corrmat = df.corr()

top\_corr\_features = corrmat.index

#plt.figure(figsize=(50,50))

#plot heat map

#g=sns.heatmap(df1[top\_corr\_features].corr(),annot=True,cmap="RdYlGn")

corr = df1.corr()

sns.heatmap(corr)

from sklearn.model\_selection import train\_test\_split

from sklearn import datasets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y)

import numpy as np

from sklearn.linear\_model import LinearRegression

regr = LinearRegression(fit\_intercept = True, normalize = True, copy\_X = True, n\_jobs = 2).fit(X\_train,y\_train)

regr.predict(X\_test)

regr.score(X\_train,y\_train)

regr.coef\_

import numpy as np

from sklearn.linear\_model import LinearRegression

regr = LinearRegression(fit\_intercept = True, normalize = True, copy\_X = True, n\_jobs = 2).fit(X\_train,y\_train)

regr.predict(X\_test)

regr.score(X\_train,y\_train)

regr.coef\_

!pip install scikit-plot

from imblearn import over\_sampling

from imblearn.over\_sampling import RandomOverSampler

from imblearn import over\_sampling

#Sampling process----------------------------------------------------------------

ada = over\_sampling.ADASYN(sampling\_strategy='minority', random\_state=0)

ros = RandomOverSampler(sampling\_strategy='minority', random\_state=0)

sm = over\_sampling.SMOTE(sampling\_strategy='minority', random\_state=0)

clf1 = DecisionTreeClassifier(max\_depth=5,splitter="best",min\_weight\_fraction\_leaf=0.025, criterion='gini' ,min\_impurity\_decrease=0.0001,max\_features='sqrt',random\_state=0,min\_samples\_split=2,min\_samples\_leaf=10)

#for nestimators in [5,10,20,80, 90, 100, 120,130]:

X\_sm,y\_sm = ros.fit\_resample(X\_train,y\_train)

#clf1  =RandomForestClassifier(n\_estimators=nestimators, )

model\_1 = clf1.fit(X\_sm,y\_sm)

Y\_pred = model\_1.predict\_proba(X\_test)

ruc\_score = roc\_auc\_score(y\_true = y\_test, y\_score =Y\_pred[:,1]  )

print(ruc\_score)

scikitplot.metrics.plot\_roc(y\_test,Y\_pred )

best\_threashold =  best\_th( y\_test, Y\_pred[:,1] )

y\_pred = binaryclaasification(best\_threashold,  Y\_pred[:,1])

print(confusion\_matrix( y\_test,y\_pred ))

print(classification\_report( y\_test,y\_pred ))

### Conclusion:

Our ML model gives 62% accuracy so that It is helpful to predict the seriousness of a patient’s getting a Heart Failure.