PROJECT REPORT FORMAT

1. INDRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms.The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

The general term used to cover malfunctions of the heart is Heart Disease, or sometimes Cardiac Disease ("Cardiac" is a Latin term for the heart). Though there are multiple forms of heart disease, our discussion focuses on the two most common: Heart Attack and Heart Failure. This document is designed to teach you about heart attacks and heart failure: what causes these diseases, what forms these diseases take, and what can be done to treat these diseases when they occur. As both of these diseases are to some extent avoidable, we have also provided a discussion of preventative steps you can take to decrease your chances of having to deal with heart disease, or to minimize the negative effects of existing heart disease.

1.1 PROJECT OVERVIEW

- Blood vessel disease, such as coronary artery disease
- Irregular heartbeats (arrhythmias)
- Heart problems you're born with (congenital heart defects)
- Disease of the heart muscle
- Heart valve disease
- Many forms of heart disease can be prevented or treated with healthy lifestyle choices.
- A congenital heart defect is a problem with the structure of the heart that a child is born with.
- Some congenital heart defects in children are simple and don't need treatment. Other congenital heart defects in children are more complex and may require several surgeries performed over a period of several years.
- Learning about your child's congenital heart defect can help you understand the condition and know what you can expect in the coming months and years.

2 . Purpose

- The objective of this project is to check whether the patient is likely tNowadays, humans face various diseases due to the current environmental condition and their living habits. The identification and prediction of such diseases at their earlier stages are much important, so as to prevent the extremity of it.

- Predictive research is chiefly concerned with forecasting (predicting) outcomes, consequences, costs, or effects. This type of research tries to extrapolate from the analysis of existing phenomena, policies, or other entities in order to predict something that has not been tried, tested, or proposed before.In general, prediction is the process of determining the magnitude of statistical variates at some future point of time.

- The be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc. A dataset is selected from the UCI repository with patient's medical history and attributes.

2. LITERATURE SURVEY

1. Bellaachia and Guven (2005) proposed predicting breast cancer lastingness using data mining method.The authors have examined three data mining methods such as Naïve Bayes, propagated neural networks, and c4.5 decision tree algorithms. Naïve Bayes method is the first method that uses the Bayesian method, because of its simple, clear, and fast predictive nature. The second method is artificial neural networks (ANNs) that uses multilayer network with transmission utilisation. Finally, they used c4.5 decisiontree algorithms. On the whole, the authors' work shows that the preliminary results are challenging prediction problem in medical data sets. Data mining is the apt technology to predict patterns in the health sector data set. Though it is tedious to make the prediction of few diseases such as heart attack, due to its complexity, such tasks need more skill. Masethe and Masethe (2014) discussed to determine heart disease using classification algorithms. Few data mining algorithms such as j48, Naïve Bayes, REPTREE, and classification and regression trees (CART) are applied to predict heart attacks. The author's research work result shows that prediction accuracy is 99%, and j48, REPTREE, and CART gave a prediction model of 89 cases with a risk factor positive for heart attacks. From these techniques, it was identified that prediction of diagnoses can be done by data-mining algorithms.

2. A medical data of large size need powerful data analysis tools for processing Data mining techniques can also be used for the diagnosis and predictive analysis. Ramaraj and Thanamani (2013) proposed predictive analytics methods to identify heart diseases. The authors' aim was to design a predictive method for heart disease detect Classification accuracy report among various data mining techniques with the difference inrates is provided in analysis part. The authors' final result shows that CN2Rule performs classification more accurately than the other methods. Nasridinov

et al. (2014) discussed a study on crime pattern prediction using data mining techniques. The authors analysed many data mining techniques with generated test data to determine the best method to perform crime pattern prediction task.Specifically, the authors did an extensive performance analysis of various data mining prediction algorithms such as support vector machine (SVM), decision tree, neural network, k-nearest neighbour, and Naïve Bayes. The authors assumed that wearable sensor devices are attached to the clothes of the user of the proposed method. It captures the inner temperature and heartbeat of a user and sends these data to the server to perform emotion mining.

3. Chandra Shekar et al. (2012) make up a better algorithm for prediction of heart disease using case-based machine learning-based methods technique on non-binary data sets. Mining frequent item-sets in non-binary search space presented fascinating challenges over conventional mining in binary search space. Initially, the non-binary search space needs innovative tactics to calculate support and must be active. As there is a chance of removal of candidate item-set from the non-binary data set due to pruning, applying it at a higher level may become frequent. Support calculation and candidate generation at each level are carried out using separate mechanism. The author’s final result was a prototype for generating frequent item sets for non-binary data set that was developedMining frequent item-sets in non-binary search space presented fascinating challenges over conventional mining in binary search space. Initially, the non-binary search space needs innovative tactics to calculate support and must be active. As there is a chance of removal of candidate item-set from the non-binary data set due to pruning, applying it at a higher level may become frequent. Support calculation and candidate generation at each level are carried out using separate mechanism. The author’s final result was a prototype for generating frequent item sets for non-binary data set that was developed.

4. Kone and Karwan (2011) predicted the expense incurred in delivering bulk (liquefied) gas to new customers making use of a multifactor linear regression model. Development of a single model, i.e. evaluating all the observations one time, leads to poor prediction outcomes. Hence, before regression analysis, a novel supervised learning method is utilised for grouping the customers who have similarity in some or the other perception. Hyperboxes are used to denote classes on customers, and subsequently, a linear regression model is developed within every class. To increase with the combination of data classification and regression, the accuracy of the prediction is indicated. Bhat et al. (2011) presented a new preprocessing phase along with imputation of missing value for numerical and also categorical data. A hybrid combination consisting of classification and regression trees (CART), genetic algorithms for imputing the missing sequential values and self-organising feature maps (SOFM) for imputing the categorical values are used in the work.A linear regression model is developed within every class. To increase with the combination of data classification and regression, the accuracy of the prediction is indicated.

5. Various data mining methodologies were used for predicting the heart disease by Soni et al. (2011). The accuracy of those algorithms are verified, in which the accuracies of Naïve Bayes, ANN and decision tree are said to have accomplished a respective 86.53, 85.53, and 89%. The data mining algorithms such as ANN, decision trees, and C4.5 apply ECG signals to analyse the heart disease. Decision tree algorithm is found to be the best and obtains 97.5% accuracy. The C4.5 algorithm yields an accuracy of 99.20%, whereas Naïve Bayes algorithm produces 89.60% of accuracy (Aneeshkumar and Venkateswaran, 2012). Therefore, these algorithms are employed for estimating the supervision over liver disorder.C5.0 is a classification algorithm that is applied on huge data sets. It overcomes C4.5 in terms of the memory and speed along with the performance. This technique divides the sample depending on the field which provides the high information gain. Later, the obtained sample subset received earlier will be divided. The action will persist till the sample

subset cannot be further divided. At last, the lowest level split in the sample subsets that have less than acceptable level contribution for the model will be eliminated. C5.0 methodology easily deals with the missing attribute and the multivalued attribute from data set (Patil et al., 2012).

2.1 Existing Problem

1. Data insight**:** As mentioned here we will be working with the heart disease detection dataset and we will be putting out interesting inferences from the data to derive some meaningful results.
2. EDA**:** Exploratory data analysis is the key step for getting meaningful results.
3. Feature engineering**:** After getting the insights from the data we have to alter the features so that they can move forward for the model building phase.

2.2 REFERENCE

1. Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics—2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2008; 117(4):e25–146. [Pub Med]

2. National Heart Lung and Blood Institute Fact Book, Fiscal Year 2006. Bethesda, Md: National Heart Lung and Blood Institute, National Institutes of Health; 2006. [12 April 2011]. Last accessed at http://www.nhlbi.nih.gov/about/factbook-06/toc.htm on.

3. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA. 2001; 285(19):2486–97. [Pub Med]

4. Hayden M, Pignone M, Phillips C, et al. Aspirin for the primary prevention of cardiovascular events: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2002; 136(2):161–72. [Pub Med]

.

5. The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Arch Intern Med. 1997;157(21):2413–46. [Pub Med]

6. Grundy SM, Bazzarre T, Cleeman J, et al. Writing Group I. Prevention Conference V. Beyond secondary prevention: identifying the high-risk

patient for primary prevention. Medical Office Assessment. Circulation. 2000; 101(1):e3–11. [Pub Med]

7. Pearson TA, Blair SN, Daniels SR, et al. AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update. Circulation. 2002; 106(3):388–91. [Pub Med]

8. Montgomery AA, Fahey T, MacKintosh C, et al. Estimation of cardiovascular risk in hypertensive patients in primary care. Br J Gen Pract. 2000; 50(451):127–8. [PMC free article] [Pub Med]

9. Chatellier G, Blinowska A, Menard J, et al. Do physicians estimate reliably the cardiovascular risk of hypertensive patients? Medinfo. 1995;8(Pt 2):876–9. [PubMed]

10. Grover SA, Lowensteyn I, Esrey KL, et al. Do doctors accurately assess coronary risk in their patients? Preliminary results of the Coronary Health Assessment Study. BMJ. 1995; 310(6985):975–8. [PMC free article] [Pub Med]

11. Friedmann PD, Brett AS, Mayo-Smith MF. Differences in generalists' and cardiologists' perceptions of cardiovascular risk and the outcomes of preventive therapy in cardiovascular disease. Ann Intern Med. 1996; 124(4):414–21. [Pub Med]

12. Meland E, Laerum E, Lehmann EH. Assessment of coronary heart disease risk, II: a clinical multicentre study of general practitioners' risk assessment. Fam Pract. 1994;11(2):122–6. [Pub Med]

13. Meland E, Laerum E, Stensvold I. Assessment of coronary heart disease risk, I: a postal inquiry among primary care

physicians. Fam Pract. 1994;11(2):117–21. [Pub Med]

14. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Münster (PROCAM) study. Circulation. 2002; 105(3):310–5. [Pub Med]

15. Pocock SJ, McCormack V, Gueyffier F, et al. A score for predicting risk of death from cardiovascular disease in adults with raised blood pressure, based on individual patient data from randomised controlled trials. BMJ. 2001; 323(7304):75–81. [PMC free article] [Pub Med]

16. Tunstall-Pedoe H. The Dundee coronary risk-disk for management of change in risk factors. BMJ. 1991; 303(6805):744–7. [PMC free article] [Pub Med]

17. Menotti A, Lanti M, Puddu PE, et al. Coronary heart disease incidence in northern and southern European populations: a reanalysis of the Seven Countries Study for a European coronary risk chart. Heart. 2000; 84(3):238–44. [PMC free article] [Pub Med]

18. Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. Am Heart J. 1991; 121(1 Pt 2):293–8. [Pub Med]

19. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation. 1998; 97(18):1837–47. [Pub Med]

20. Menotti A, Puddu PE, Lanti M. Comparison of the Framingham risk function-based coronary chart with risk function from an Italian population study. Eur Heart J. 2000; 21(5):365–70. [Pub Med]

21. Game FL, Bartlett WA, Bayly GR, et al. Comparative accuracy of cardiovascular risk prediction methods in patients with diabetes mellitus. Diabetes Obes Metab. 2001; 3(4):279–86. [Pub Med]

22. Sheridan S, Pignone M, Mulrow C. Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians. J Gen Intern Med. 2003; 18(12):1039–52. [PMC free article] [Pub Med]

23. D'Agostino RB Sr, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA. 2001; 286(2):180–7. [Pub Med]

24. Liao Y, McGee DL, Cooper RS. Prediction of coronary heart disease mortality in blacks and whites: pooled data from two national cohorts. Am J Cardiol. 1999; 84(1):31–6. [Pub Med]

25. Liao Y, McGee DL, Cooper RS, et al. How generalizable are coronary risk prediction models? Comparison of Framingham and two national cohorts. Am Heart J. 1999; 137(5):837–45. [Pub Med]

26. Grundy SM, D'Agostino RB Sr, Mosca L, et al. Cardiovascular risk assessment based on US cohort studies: findings from a National Heart Lung and Blood Institute workshop. Circulation. 2001; 104(4):491–6. [Pub Med]

27. U.K. Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. BMJ. 1998; 317(7160):703–13. [PMC free article] [Pub Med]

28. Haffner SM, Lehto S, Ronnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med. 1998; 339(4):229–34. [Pub Med]

29. Buyken AE, von Eckardstein A, Schulte H, et al. Type 2 diabetes mellitus and risk of coronary heart disease: results of the 10-year follow-up of the PROCAM study. Eur J Cardiovasc Prev Rehabil. 2007; 14(2):230–6. [Pub Med]

30. Diverse Populations Collaborative Group. Prediction of mortality from coronary heart disease among diverse populations: is there a common predictive function? Heart. 2002; 88(3):222–8. [PMC free article] [Pub Med]

31. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III): final report. Circulation. 2002; 106(25):3143–421. [PubMed]

32. Brindle P, Beswick A, Fahey T, et al. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. Heart. 2006; 92(12):1752–9. [PMC free article] [Pub Med]

33. Eichler K, Puhan MA, Steurer J, et al. Prediction of first coronary events with the Framingham score: a systematic review. Am Heart J. 2007; 153(5):722–31. [Pub Med]

34. Wilson PW, Bozeman SR, Burton TM, et al. Prediction of first events of coronary heart disease and stroke with consideration of adiposity. Circulation. 2008; 118(2):124–30. [Pub Med]

35. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ. 2008; 336(7659):1475–82. [PMC free article] [Pub Med]

36. Simmons R K, Sharp S, Boekholdt SM, et al. Evaluation of the Framingham risk score in the European Prospective Investigation of Cancer-Norfolk cohort:doseaddingglyhemoglobiimprove the prediction of coronary heart disease events? Arch Intern Med. 2008; 168(11):1209–16. [Pub Med]

37. Gaziano TA, Young CR, Fitzmaurice G, et al. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-Up Study cohort. Lancet. 2008; 371(9616):923–31. [PMC free article] [Pub Med]

38. Yang X, So WY, Kong AP, et al. Development and validation of a total coronary heart disease risk score in type 2 diabetes mellitus. Am J Cardiol. 2008; 101(5):596–601. [Pub Med]

39. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008; 117(6):743–53. [Pub Med]

40. Becker A, Leber A, Becker C, et al. Predictive value of coronary calcifications for future cardiac events in asymptomatic individuals. Am Heart J. 2008; 155(1):154–60. [Pub Med]

2.3 Problem Statement Definition

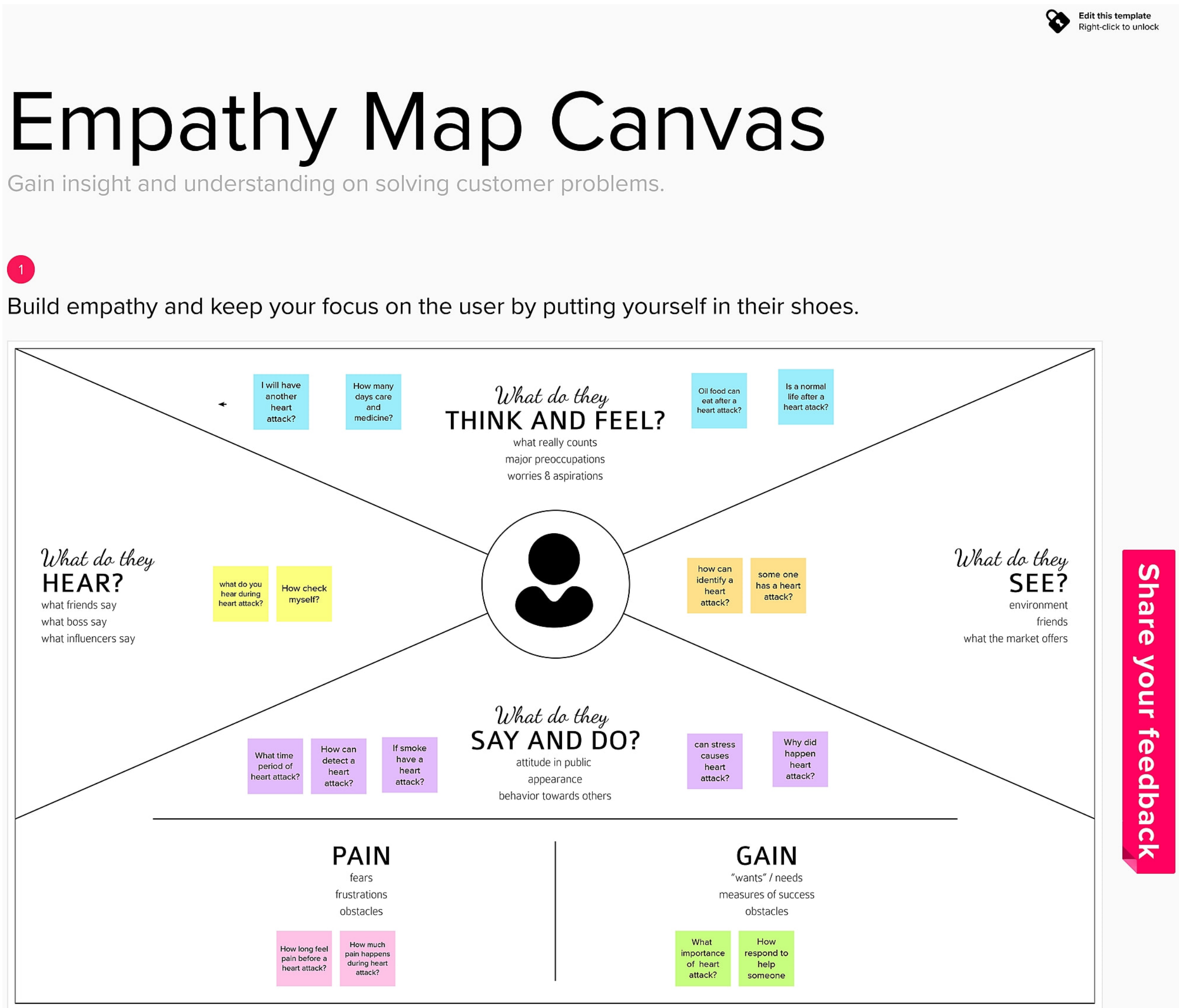
Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will beto predict accurately with few tests and attributes the presence of heart disease. Attributesconsidered form the primary basis for tests and give accurate results more or less.Manymore input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors'intuition and experience rather than on the knowledge rich data hidden n the data set and databases.

This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systemsystematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. According to (Wurz & Takala, 2006) the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending.

The successful application of data mining in highly visible fields like e-business, marketing andretail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP



3.2 IDEATION

A cross sectional study of 103 patients with coronary artery disease aged between 35 and 87 years, in a public health clinic, all diagnosed with coronary artery disease who were under medical supervision of a cardiologist, was conducted. The criteria for inclusion in this study were: have medical follow-up; have watched the initial interview and have it performed in its entirety; have sufficient cognitive capacity to understand the instructions given; and are 18 or over years old. Patients signed a consent form and were aware of the experimental protocol (approved by the Ethics Committee of the Universidade Federal do Rio de Janeiro) before the start of the participation. Patients were evaluated with Mini International Neuropsychiatric Interview (MINI 5.0) [12]; this is considered to be an instrument that has the default template a short structured interview (approximately 25 minutes) for the assessment of the existence of Axis I psychiatric disorders according to DSM-IV and the 10th revision of International Classification Of Diseases (ICD-10), and in accordance with the criteria of cut-off point of the current risk of suicide, the scores are classified as follows: 1–6 = mild; from 6 to 9 = moderate and ≥10 = high.

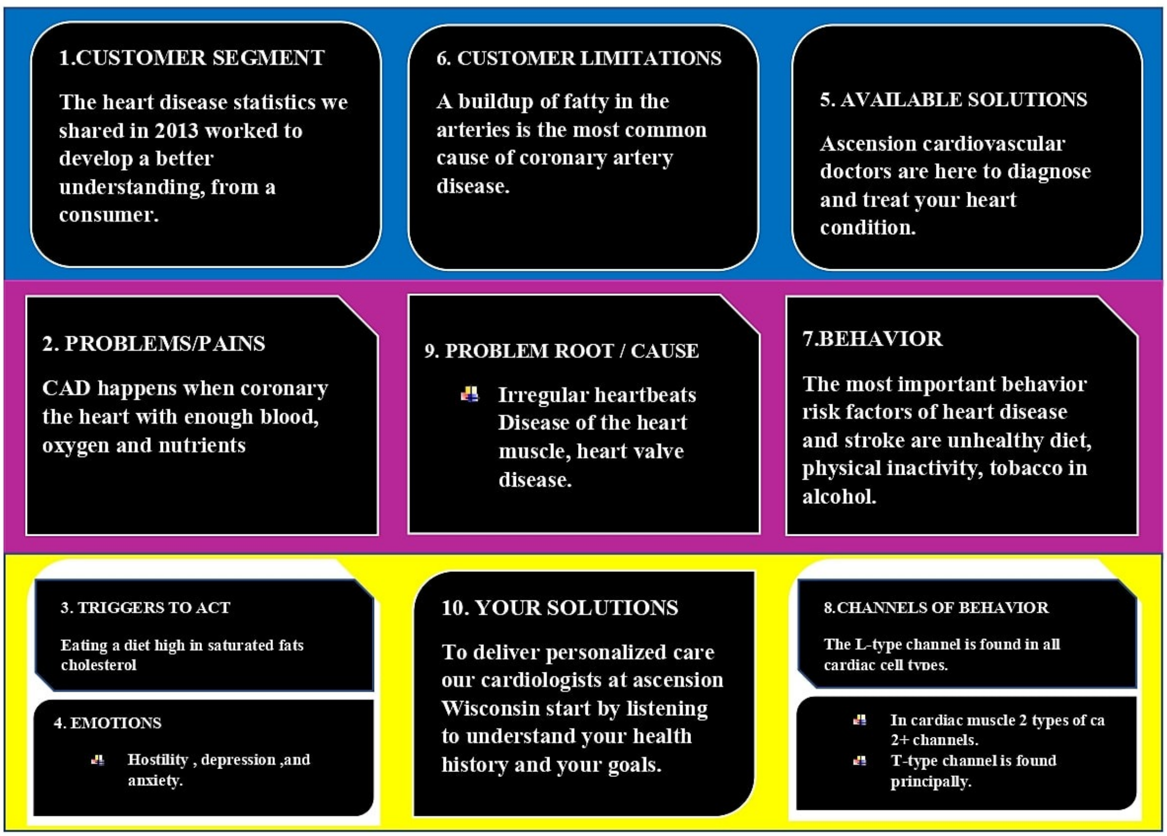
The Beck Depression Inventory (BDI) is an instrument applied to identify and quantify symptoms of depression. This consists of 21 items that assess cognitive components, affective, somatic and behavioral depression. The BDI is an investigation of sadness, pessimism, sense of failure, lack of satisfaction, a feeling of guilt, feeling of punishment, auto depreciation charges, suicidal ideas, bouts of crying, irritability, social downturn, indecision, distortion of body image, inhibition for work, sleep disturbance, fatigue, loss of appetite, weight loss, and somatic concern. For samples of patients with affective disorder recommended cut-off points are as follows: 10, no depression or symptoms of depression minimal; 10–18, mild depression to moderate; 19–29, moderate to severe depression; and 30–63, severe depression . For the evaluation of anxiety and depression, Hospital anxiety and Depression Scale (HADS) was applied. This instrument consists of 14 questions – seven for anxiety and seven for depression – with a response scale ranging from zero to three, and maximum score for both mental symptoms. Scores of cut-off points for both subscales were: HAD-anxiety (HAD it)—without anxiety 0 to 8 and with anxiety, ≥9, and scale HAD-depression (HAD-D)—no depression, 0–8 and depressed ≥9. Other instrument used was Beck Suicidal Ideation Scale (BSI).

The BSI is an instrument for measuring the presence of suicidal ideation, wishes, attitude and suicide plans. This scale was developed based on psychiatric patients, adults admitted and outpatients. The scale consists of 21 items, each with response alternatives 0 to 2 points; it assesses three dimensions of suicidal ideation: active, passive and prior suicide attempt. With a cut-off point of ≥8, suicidal ideation was considered clinically significant. The social and demographic descriptive data, including gender, age, education, occupation, religion, children, psychiatric or psychological treatment past or current and the use of psychotropic substances were also checked by means of a registration form. For statistical analysis descriptive statistics was used for social and demographic data, considering the raw data and percentage or mean values and standard deviation.

3.3 PROBLEM SOLUTION FIT

VISUALINZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

PROBLEM SOLUTION FIT



4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Business rules	Eat oil fish Quit smoking
FR-4	Audit tracking	Maintaining clinical care of parents has been the top priority of the cardiovascular community.
FR-5	External interfaces	Atherosclerosis Deep vein thrombosis Pericardial disease
FR-6	Reporting requirements	Shortness of breath while at a rest,not related to exercise or exertion

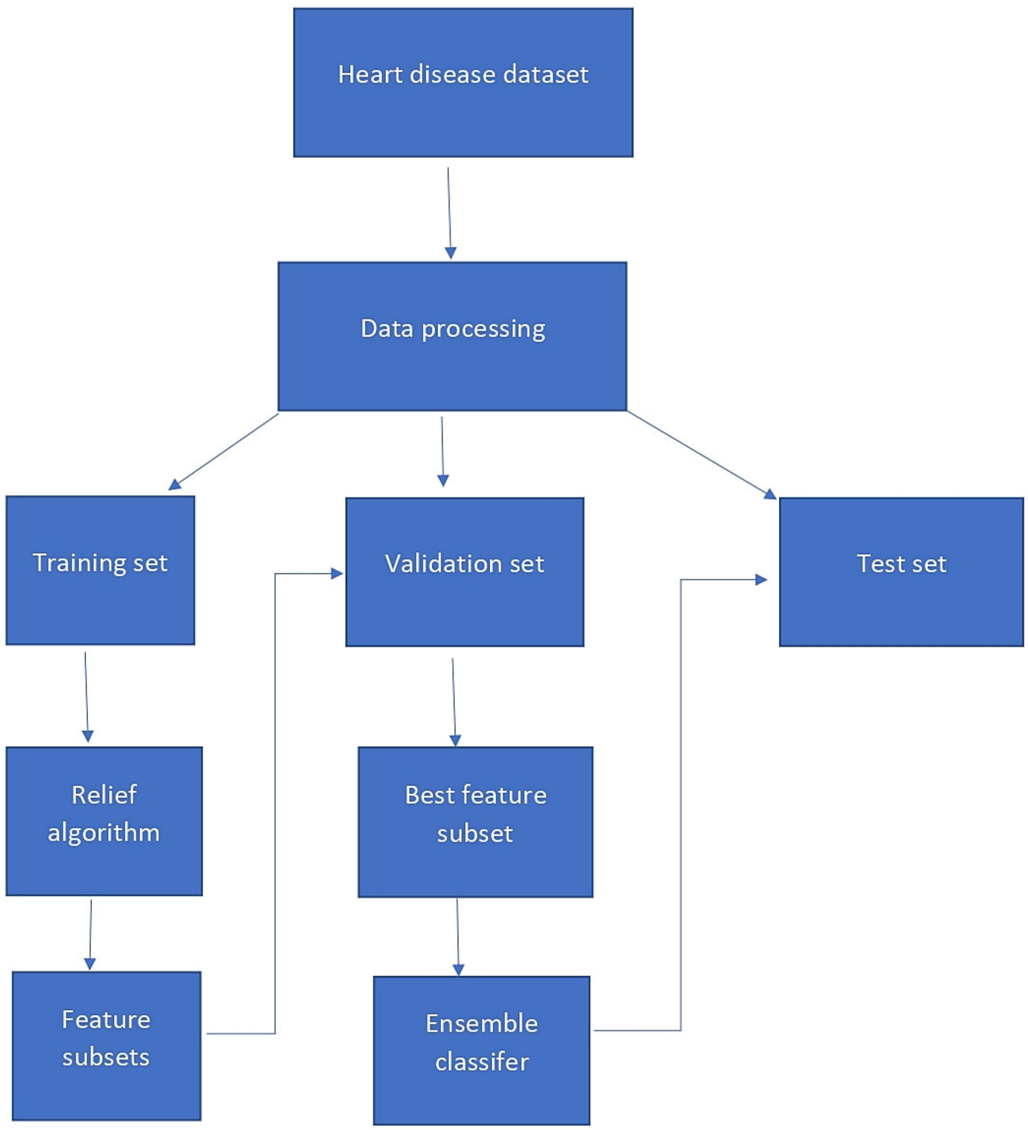
Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

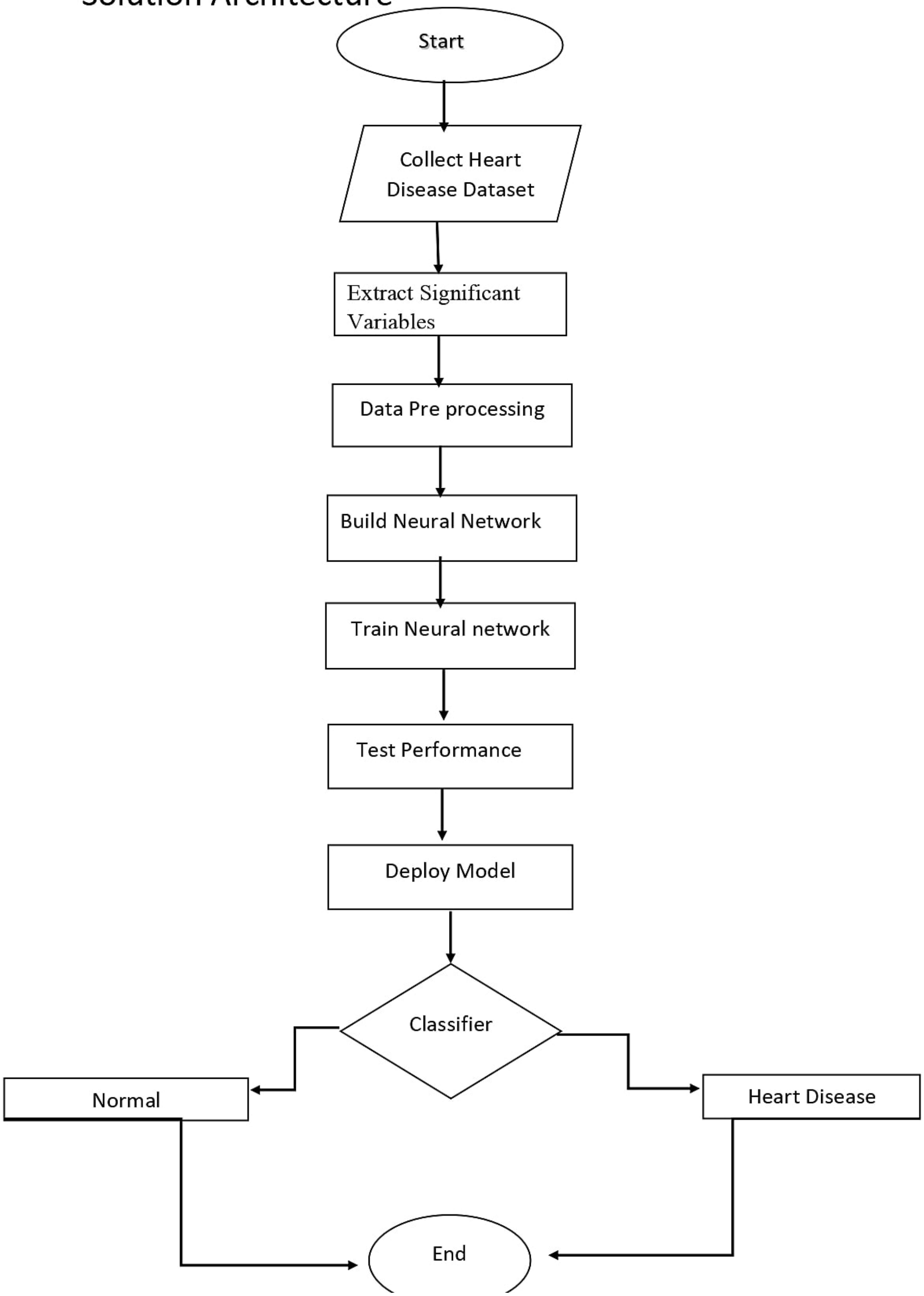
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The usability evaluations of selected heart disease monitoring mobile application are available on google play and google apple store.
NFR-2	**Security**	The social security administration operates two program including supplemental security insurance and security disability insurance.
NFR-3	**Reliability**	The reliability and validity of this questionnaire have been previously studied.The reliability was slightly better in patients whose test interval was<14days.
NFR-4	**Performance**	Physical limitations Symptoms Numerous hospitalization.

5. PROJECT DESIGN

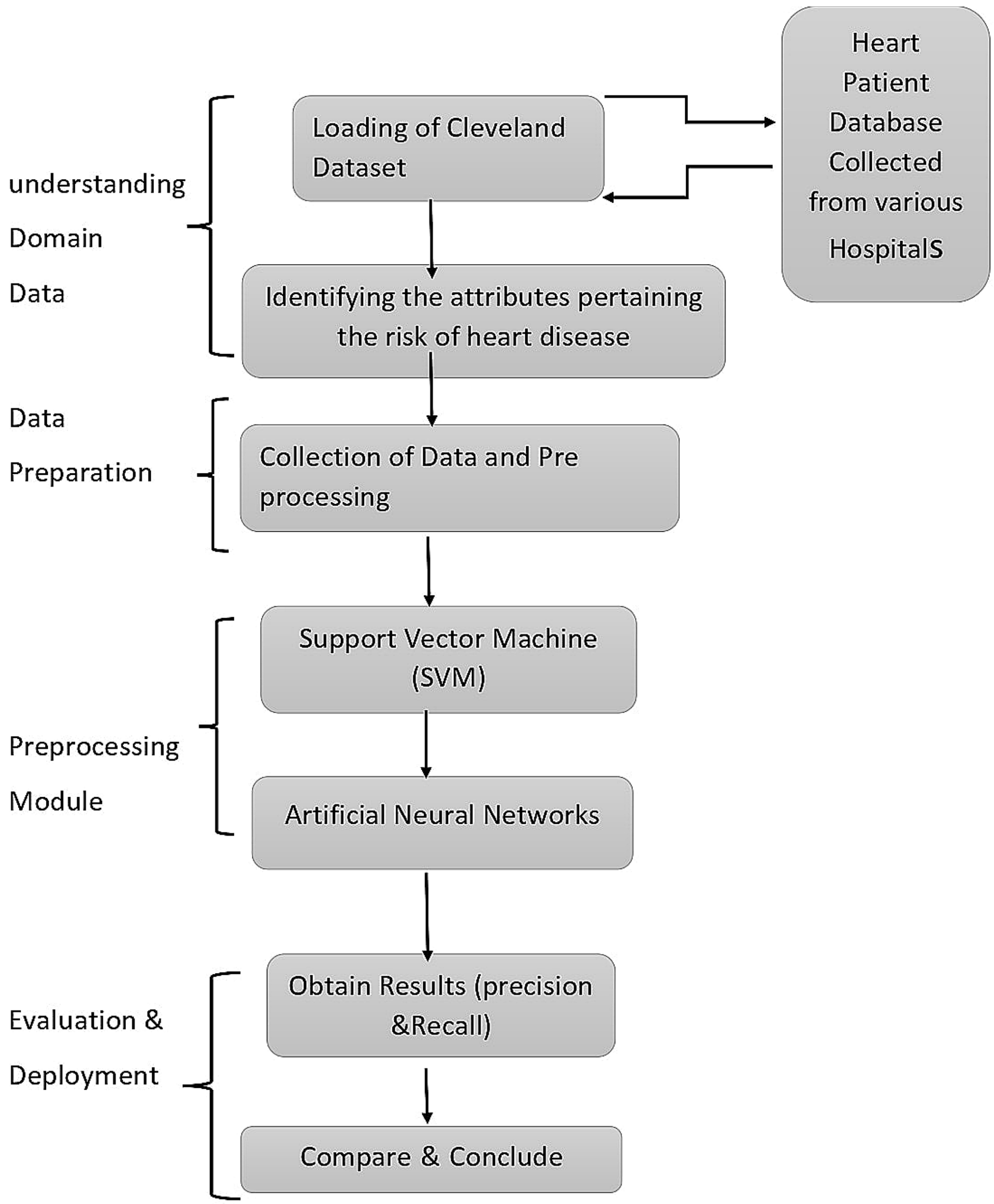
5.1 DATA FLOW DIAGRAM



Solution Architecture



5.2 TECHNICAL ARCHITECTURE



6. PROJECT PLANNING AND SCHEDULING

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations / Graphs – 4
2.	Data Responsiveness	A cardiologist measures vitals & hands you this data to **perform Data Analysis** and **predict** whether certain patients have Heart Disease.
3.	Amount Data to Rendered (DB2 Metrics)	Among the many applications of mass spectrometry, biomarker pattern discovery from protein mass spectra has aroused considerable interest in the past few years.
4.	Utilization of Data Filters	The BP algorithm has served as a useful methodology to train multilayer perceptron for a wide range of applications.
5.	Effective User Story	No of Scene Added - 2
6.	Descriptive Reports	No of Visulizations / Graphs - 3

7.CODING & SOLUTIONING

```
# 1 = heart import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from keras.models import Sequential
from keras.layers import Dense
from tensorflow.keras.optimizers import Adam
from keras.callbacks import ModelCheckpoint
from sklearn.metrics import classification_report, accuracy_score
# import the heart disease dataset
dataset = "http://archive.ics.uci.edu/ml/machine-learning-databases/heart-
disease/processed.cleveland.data"
column_names =
['age','sex','cp','trestbps','chol','fbs','restecg','thalach','exang','oldpeak','slope'
,'ca','thal','class']
# read the csv
dataset = pd.read_csv(dataset, names=column_names)

# remove missing data with "?"
df = dataset[~dataset.isin(['?'])]

#drop rows with NaN values from DataFrame
df = df.dropna(axis=0)

# checking data type of the dataframe
print (df.dtypes)
```

```
# transform data to numeric because ca and thal are object datatypes
data = df.apply(pd.to_numeric)
print(data.dtypes)
# plot histograms for each variable
data.hist(figsize = (15, 15))
plt.show()
# create X and Y datasets for training
X = data.iloc[:,0:13]
y = data.iloc[:,-1]
# changing class column to binary.
# 0 = no heart diseasedisease

data["class"] = np.where(data["class"] > 0, 1, data["class"])


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,shuffle
= True)

# create model
model = Sequential()
model.add(Dense(10, input_dim=13, kernel_initializer='normal',
activation='relu'))
model.add(Dense(8, kernel_initializer='normal', activation='relu'))
model.add(Dense(4, kernel_initializer='normal', activation='relu'))
model.add(Dense(1, activation='sigmoid'))

filepath="CNN_Model-{epoch:02d}-{val_accuracy:.2f}.h5"
checkpoint = ModelCheckpoint(filepath, monitor='val_accuracy', verbose=1,
save_best_only=True, mode='max')
callbacks_list = [checkpoint]
```

```
# compile model
model.compile(loss='binary_crossentropy',
optimizer=Adam(learning_rate=0.001), metrics=['accuracy'])

train_model = model.fit(X_train, y_train, epochs=60, batch_size=8, verbose
= 1, validation_data=(X_test,y_test),callbacks=[callbacks_list])

plt.plot(train_model.history['accuracy'], marker='.')
plt.plot(train_model.history['val_accuracy'], marker='.')
plt.title('model accuracy')
plt.xlabel('epoch')
plt.ylabel('accuracy')
plt.grid()
plt.legend(['accuracy', 'val_accuracy'], loc='lower right')
plt.savefig('model_accuracy_binary.png')
plt.close()

plt.plot(train_model.history['loss'], marker='.')
plt.plot(train_model.history['val_loss'], marker='.')
plt.title('model loss')
plt.xlabel('epoch')
plt.ylabel('loss')
plt.grid()
plt.legend(['loss', 'val_loss'], loc='upper right')
plt.savefig('model_loss_binary.png')
plt.close()

# Rounding the prediction if it is a binary problem
pred = np.round(model.predict(X_test))
print(pred)
```

```python
for i in pred:
    if(np.max(i) > 0.5):
        print("Have Heart Disease")
    else:
        print("Does not have heart disease")


print('Classification Accuracy: '+str(accuracy_score(y_test, pred) * 100)+'
%')
```

8. TESTING

8.1 TEST CASES

- Blood tests. ...
- Electrocardiogram (ECG) ...
- Exercise stress test. ...
- Echocardiogram (ultrasound) ...
- Nuclear cardiac stress test. ...
- Coronary angiogram. ...
- Magnetic resonance imaging (MRI) ...
- Coronary computed tomography angiogram (CCTA)
- Electrocardiogram (ECG or EKG). An ECG is a quick and painless test that records the electrical signals in the heart. ...
- Holter monitoring. ...
- Echocardiogram. ...
- Exercise tests or stress tests. ...
- Cardiac catheterization. ...
- Heart (cardiac) CT scan. ...
- Heart (cardiac) magnetic resonance imaging (MRI) scan.

8.2 USER ACCEPTANCE TESTING

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

3. Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3

9. RESULT

9.1 PERFORMANCE METRICS

The performance metrics, Accuracy, Precision and Recall, obtained by the four machine learning algorithms are compared. These mathematical models are being proposed to get the better analysis on the performance metrics. As per the comparisons shown in the above, the performance metrics of this paper are more accurate than the previous papers. Hence, the results of performance metrics meet the expectation of implementation. It is observed that the highest accuracy is obtained by decision tree, the highest precision is obtained by Logistic Regression and the highest recall is obtained by KNN. The random forest algorithm gives promising results across all the performance metrics.

10. ADVANTAGE & DISADVANTAGES

ADVANTAGES

- To maintain cholesterol and blood pressure. High cholesterol levels increase your risk for developing cardiovascular disease, but that's not all. ...
- To reduce feelings of depression. ...
- To lower your risk of developing dementia.
- Strengthen your heart and blood vessels.
- Improve the flow of oxygen throughout your body.
- Lower your blood pressure and cholesterol.
- Reduce your risk for heart disease, diabetes, Alzheimer's disease, stroke, and some kinds of cancer.

DISADVANTAGES

1. Both the heart and the coronary arteries that supply the heart with blood are in a vulnerable state after a coronary artery bypass graft, particularly during the first 30 days after surgery. Some people who have a coronary artery bypass graft have a heart attack during surgery, or shortly afterwards.
2. In a small percentage of patients with stents, blood cells can become sticky and clump together to form a small mass – or clot. When a blood clot forms, it can block the free flow of blood through an artery and may cause a heart attack or even death.

11. CONCLUSION

This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in building the given model are Logistic regression, Random Forest Classifier and KNN [22]. The accuracy of our model is 87.5%.

This predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure

12. FUTURE SCOPE

- The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc.

- Conditions that fall within the scope of heart disease include cardiac arrhythmias, high blood pressure, heart failure, coronary artery disease, valve disorders, and congenital heart defects, among others.

- Here are the major career opportunities one can choose from these different specialities after completing graduation in Cardiology: Cardiologist. Cardiac Surgeon (Cardiothoracic Surgeon) Clinical Nurse Specialist in Cardiology

- The scope of Cardiology and Cardiovascular Medicine journal includes coronary artery and valve diseases, interventional and pediatric cardiology, cardiovascular surgery, cardiomyopathy, and heart failure, arrhythmias and stimulation, cardiovascular imaging, vascular medicine and hypertension, epidemiology and risk ...

13. APPENDIX

Some possible complications of an appendectomy include: Bleeding. Wound infection. Infection and redness and swelling (inflammation) of the belly that can occur if the appendix bursts during surgery.

DEMO LINK :

https://www.mediafire.com/file/har4rubqhptmb8l/PNT12022TMID49136.mp4/file