

Sudden Cardiac Arrest Patient Identification

Heart-Well Hospitals

- Nikhita Veluri
 - Paresh Natarajan
 - Shreyas R. Nadig
- 

Table of contents

01

Introduction to the
problem
statement

02

Introduction to the
data

03

Exploratory data
analysis

04

Data pre-
processing and
wrangling

05

Models building
and evaluations



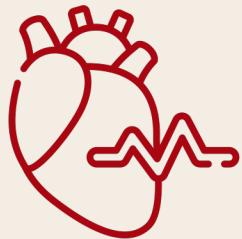
01

Introduction to the problem statement

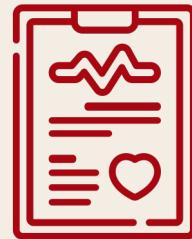




Heart-Well Hospitals



Identifying patients who potentially could be subject to **sudden cardiac arrest**



Data extracted from **patients' history**



We have been asked to will build several **prediction models** to identify patients who could develop a SCA

•
•
•

02

Introduction to the Data



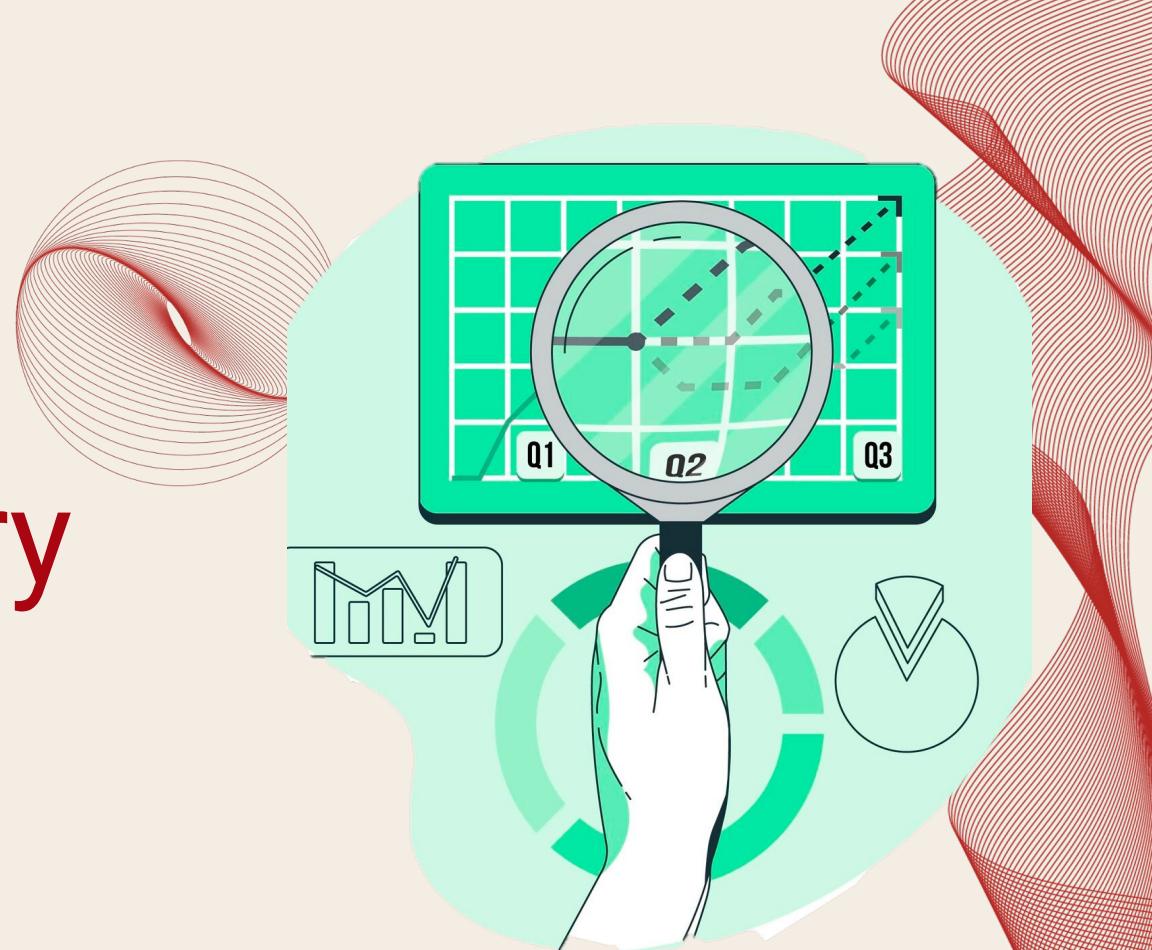
Patient History Data

Column Name	Description	Datatype
PatientName	The patient's name – this is an identifiable feature.	String
Age	The age of the patient is years	Integer
Sex	The Gender of the patient, where M=Male and F=Female.	Character (Flag)
ECG-Resting	The electrocardiogram results while patient is resting – Normal=Normal, ST=Patient has ST-T wave abnormality, and LVH	String (Flag)
ST-Slope	The ST segment/heart rate slope of the peak exercise as a predictor of coronary artery disease – Up=Sloping up, Flat=flat (no slope), and Down=Sloping down.	String
BloodPressure-Resting	The systolic blood pressure part of the reading, which is the pressure in the arteries as the heart pumps out blood during each beat – Normal is 120.	Integer
HeartRate-Max	The maximum measurement recorded for heart rate achieved - Numerical value that should be between 60 and 202.	Integer
ChestPainType	The type of chest pain, where ASY=Asymptomatic, ATA=Atypical Angina, NAP=Non-Anginal Pain, and TA=Typical Angina.	String (Flag)
Cholesterol	The cholesterol level in the blood - Numerical value where normal level for adults is below 200mg/dL.	Integer
BloodSugar-Fasting	The Glucose level while patient is fasting – Numerical Value, where normal level is 100mg/dL or less.	String (Flag)
ExerciseAngina	The patient was subject to exercise-induced angina - Y=Yes and N=No.	Character (Flag)
OldPeak	It is a term used in heart disease diagnosis that refers to the ST depression induced by exercise relative to rest – Numerical value ranging from 0 to 6.21.	Float
SCA	Sudden Cardiac Arrest is the output or target class – 1=Heart Arrest and 0=None.	Integer (Flag)

• • •

03

Exploratory Data Analysis



Understanding the data

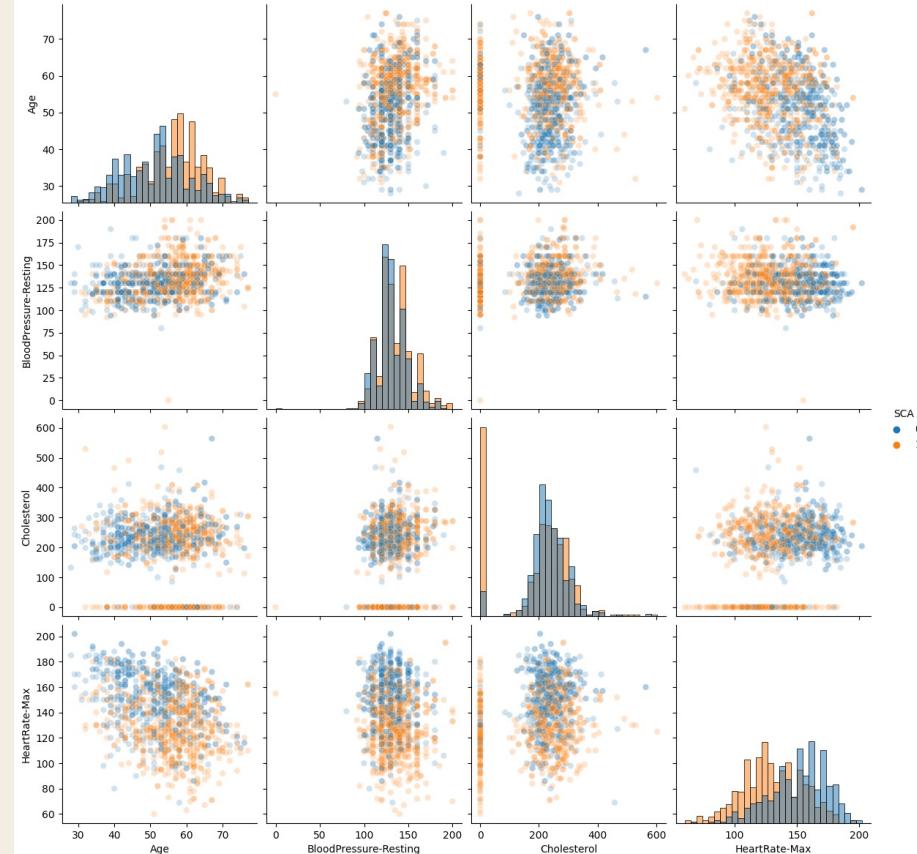
Summary Statistics

	Age	Resting BP	Max HR	Cholesterol	OldPeak	SCA
count	1221	1221	1221	1221	1221	1221
mean	53.7	132.22	139.99	210.68	0.92	0.53
std	9.34	18.29	25.44	100.42	1.09	0.49
min	28	0	60	0	-2.6	0
25%	47	120	122	188	0	0
50%	54	130	141	228	0.6	1
75%	60	140	160	269	1.6	1
max	77	200	202	603	6.2	1

Correlation Matrix

	Age	Resting BP	Max HR	Cholesterol	OldPeak	SCA
Age	1.00	0.26	-0.37	-0.04	0.25	0.26
Resting BP	0.26	1.00	-0.10	0.10	0.17	0.12
Max HR	-0.37	-0.10	1.00	0.24	-0.19	-0.41
Cholesterol	-0.04	0.10	0.24	1.00	0.06	-0.20
OldPeak	0.25	0.17	-0.19	0.06	1.00	0.40
SCA	0.26	0.12	-0.41	-0.20	0.40	1.00

Data vs SCA (target variable)



o
o
o

04

Data pre-processing and wrangling



Making the data model ready

Step 1

172

Zero Values from Resting BP
and Cholesterol removed

303

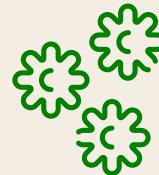
Duplicate values identified
and removed

11

Outliers identified and
removed from the data

Step 2

Feature Engineering



Introduced a new derived feature, “**Heart Risk**” which considered the Age, Resting BP, Max Heart Rate and Cholesterol. This was done to increase the correlation of the variables with the target data (SCA).

$$\text{Heart Risk} = \text{Age}/(\text{Resting BP} + \text{Cholesterol} + \text{Max Heart Rate})$$

Step 3

X → 1

Categorical encoding done
to convert all categorical
variables into numerical
features

1 → 1

Feature scaling was
performed to make sure that
the data is in the same range
for effective trend realization

○ ○ ○

05

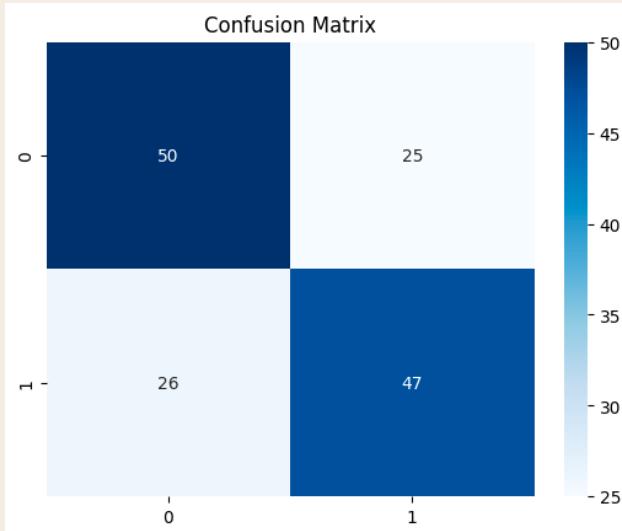
Model Building and Evaluations



Selecting the best suited model

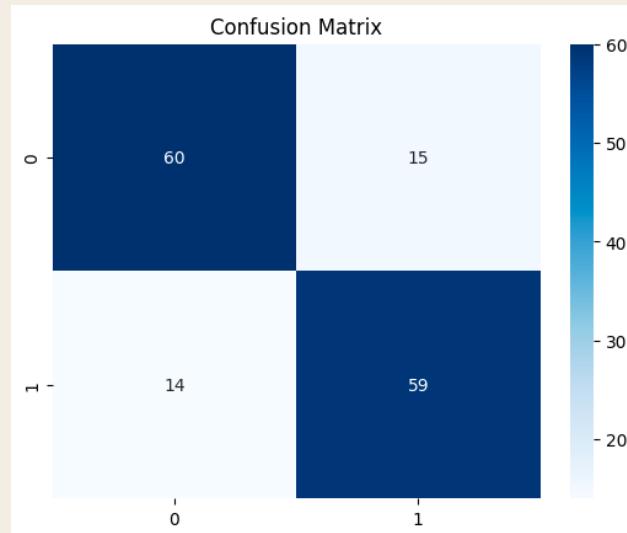
Support Vector Machine

Accuracy	0.66
Precision	0.66
Recall	0.66
F1-Score	0.66
Cross-Validation Accuracy	0.67



K-Nearest Neighbors

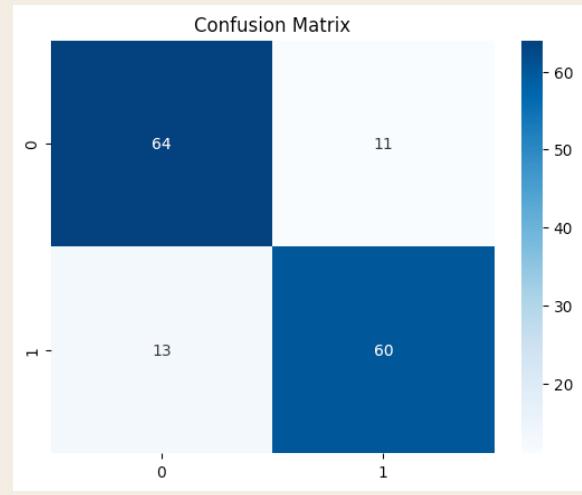
Accuracy	0.80
Precision	0.80
Recall	0.80
F1-Score	0.80
Cross-Validation Accuracy	0.79



Selecting the best suited model

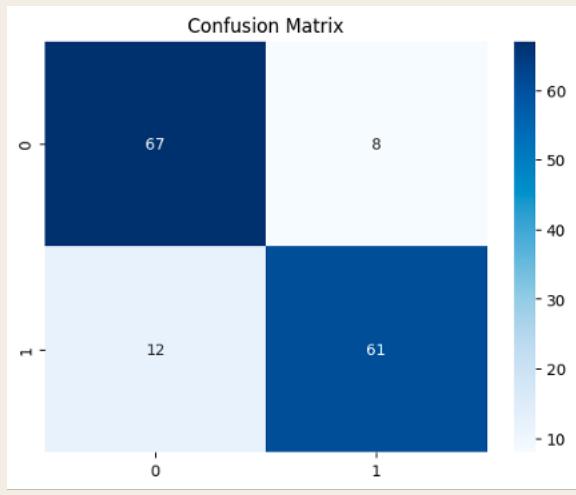
Decision Tree

Accuracy	0.84
Precision	0.84
Recall	0.84
F1-Score	0.84
Cross-Validation Accuracy	0.78



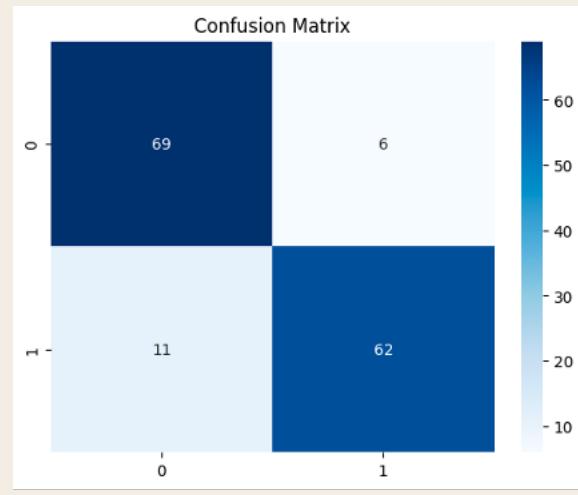
Gradient Boost

Accuracy	0.86
Precision	0.87
Recall	0.86
F1-Score	0.86
Cross-Validation Accuracy	0.85



Random Forest

Accuracy	0.89
Precision	0.89
Recall	0.89
F1-Score	0.88
Cross-Validation Accuracy	0.86

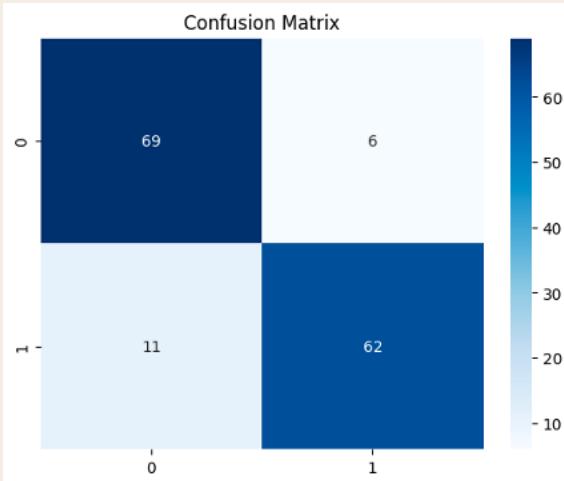


Model Winner



Random Forest

Accuracy	0.89
Precision	0.89
Recall	0.89
F1-Score	0.88
Cross-Validation Accuracy	0.86



Thank You