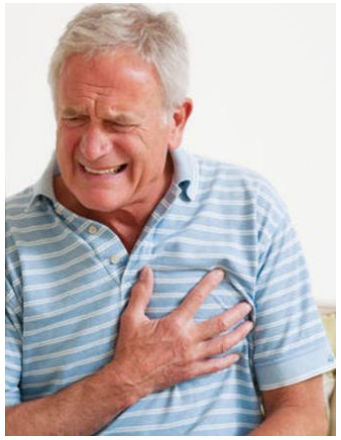
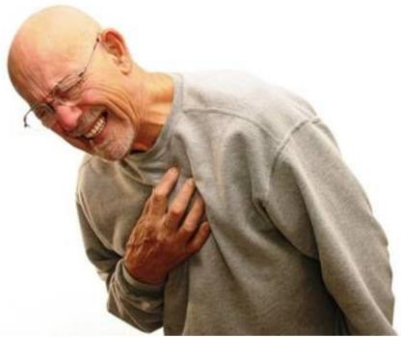


Heart Disease Detection



Team Members:

Dongxu Zheng (dz2353)

Shijun Zhu (sz2664)

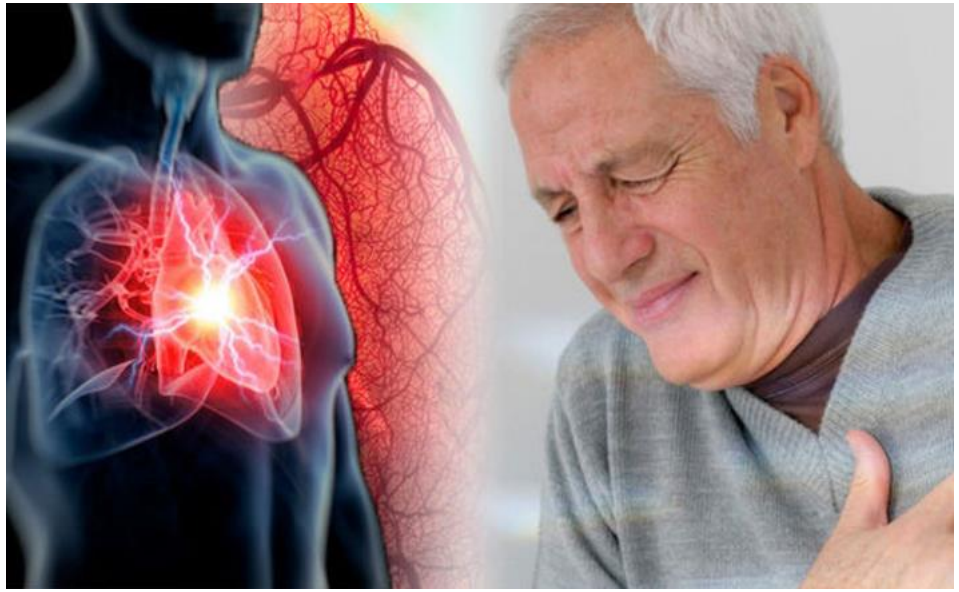
1

Introduction

Heart Disease Detection

Some facts about

Heart Disease



A heart attack occurs when the blood flow that brings oxygen to the heart muscle is blocked

610,000

Leading cause of death

About **610,000 people** die of heart disease in the United States every year–that's **1 in every 4 deaths.**

735,000

Yearly

Every year about **735,000 Americans** have a heart attack. Of these, **525,000** are a first heart attack.

Our Goal

Building a model to assist doctors to detect whether these patients are under a high risk of heart disease

Data Set Introduction

http://archive.ics.uci.edu/ml/datasets/Heart+Disease

Cleveland database

14 attributes 303 samples

UCI Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Heart Disease Data Set

Download: Data Folder, Data Set Description

Abstract: 4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

Data Set Characteristics:	Multivariate	**Number of Instances:**	303	**Area:**	Life
Attribute Characteristics:	Categorical, Integer, Real	**Number of Attributes:**	75	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	654686

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slop	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1

Project Overview

Data Pre-processing

Data exploration & cleaning

1

Pickle generator

Save the result of Classification

3

2

Build Model

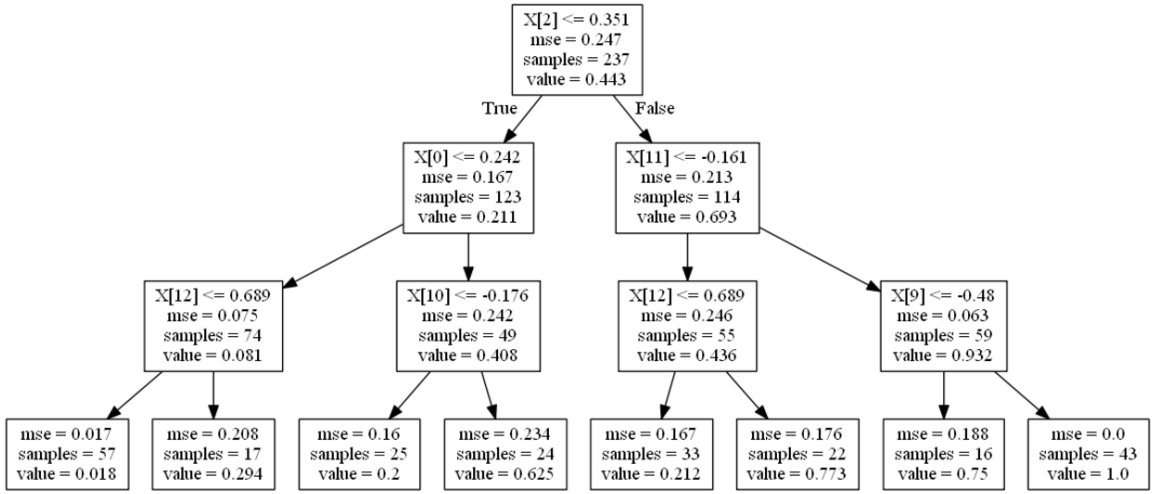
Compare the performance of KNN; Logistic Regression; Decision Tree; Random Forest

4

Build website

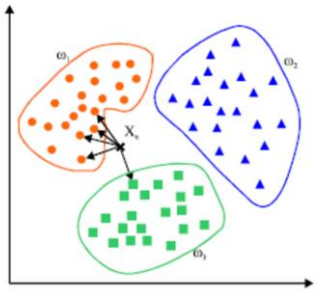
Use flask to build our heart disease prediction website

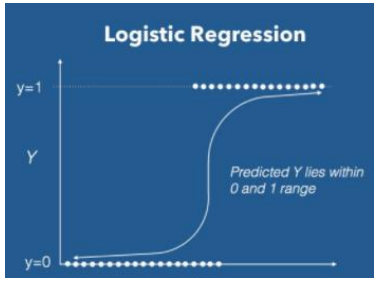
2

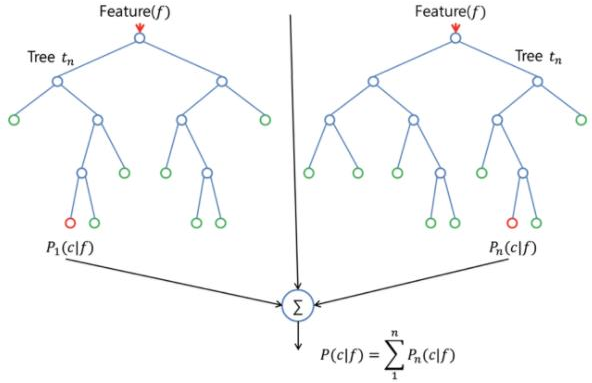


Building Model

Comparison of four models:
1) KNN 2) Logistic Regression
3)Decision Tree 4) Random Forest







Model Selection

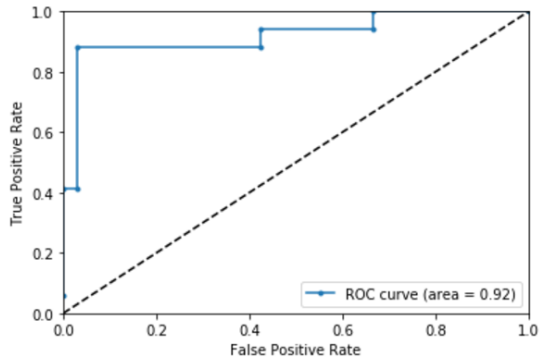
Based on Accuracy, ROC Curve

Choose

Logistic Regression

R^2 = 0.843

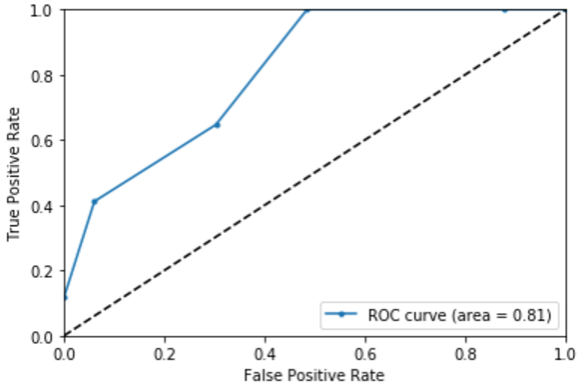
Area under the ROC curve : 0.921569



KNN Classification

84% Accuracy

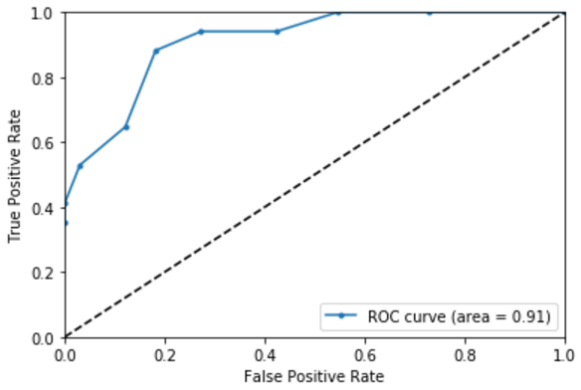
Area under the ROC curve : 0.809269



Random Forest Classifier

R^2 = 0.774

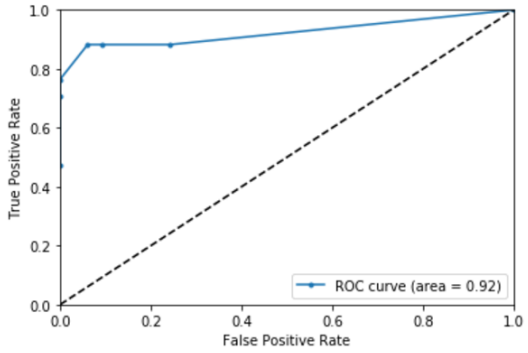
Area under the ROC curve : 0.911765



Decision Tress

R^2 = 0.604

Area under the ROC curve : 0.923351

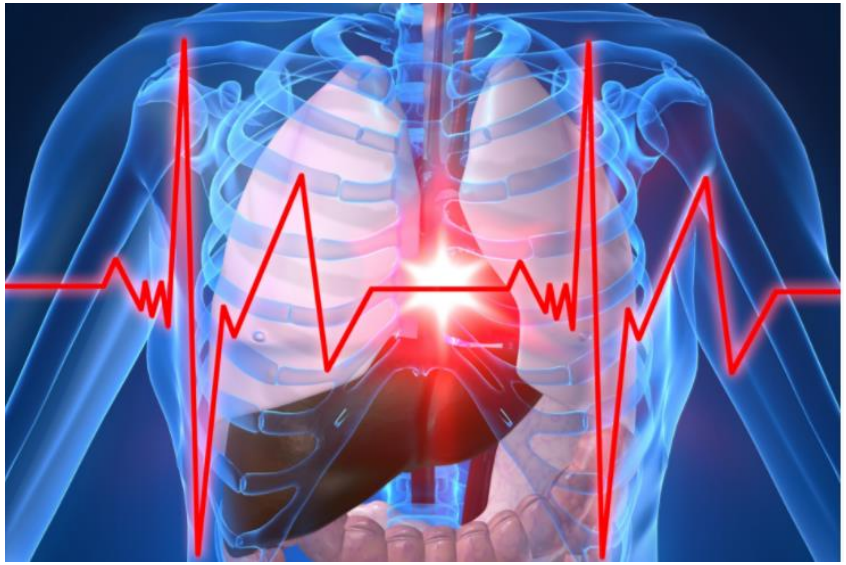


The optimal depth is 3. R^2 is not good enough. AUC is 0.923.

3

Website

Base on Flask

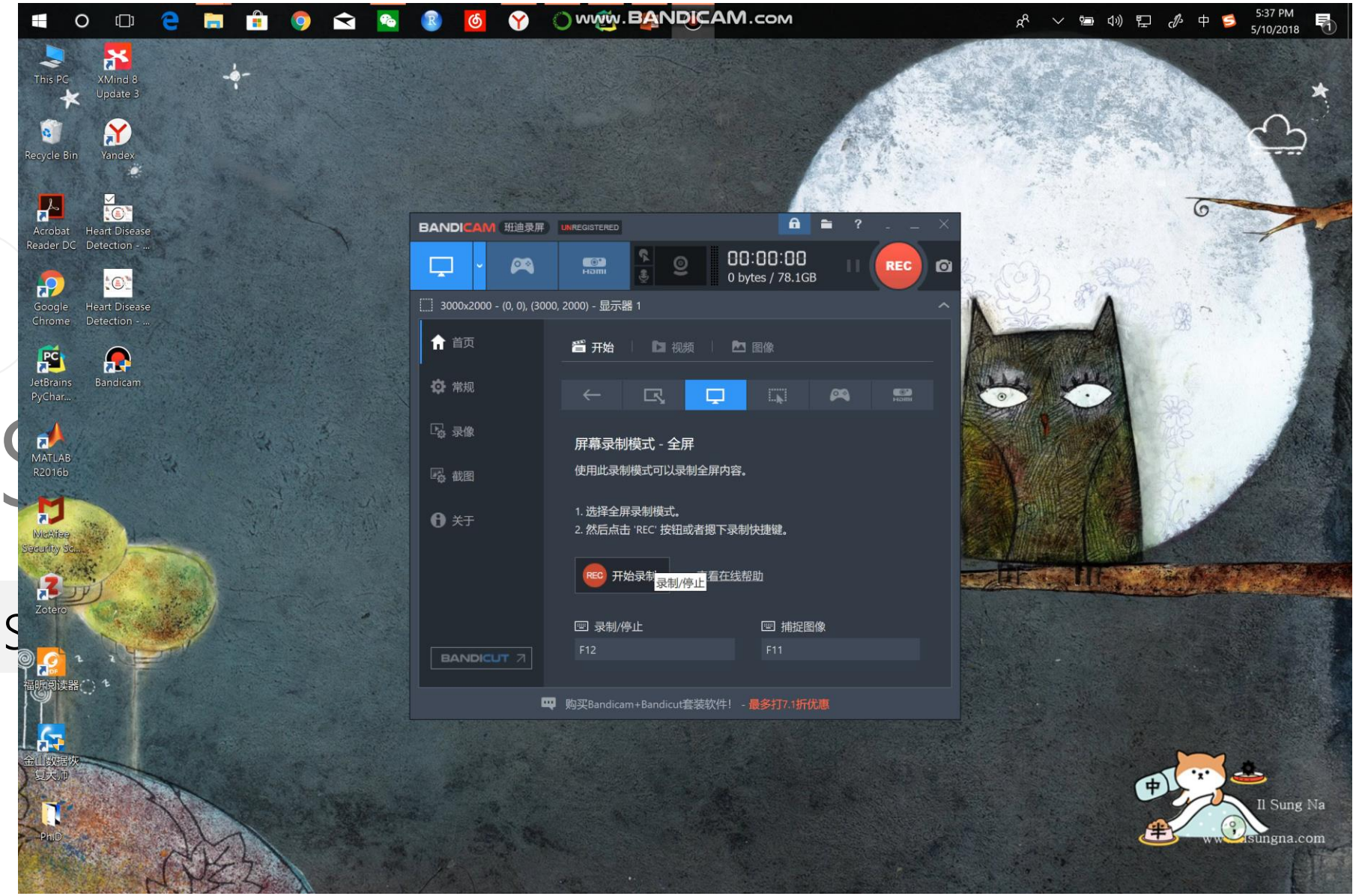


3

Webs

Base on Flas

“



Website
Interface

Heart Disease Detection

Age:

Sex:
Female

Chest pain type:
typical angina

Resting blood pressure:
mmHg

Serum cholestoral:
mg/dl

Fasting blood sugar over 120 mg/dl?
no

Resting electrocardiographic results:
0
Value 0: normal
Value 1: ST-T wave abnormality
Value 2: left ventricular hypertrophy

Maximum heart rate achieved:

Exercise induced angina?
yes

ST depression induced by exercise relative to rest:

The slop of the peak exercise ST segment:
upsloping

Number of major vessels colored by flourosopy:

thal:
normal

Submit

Healthy

Warning