

DAYANANDA SAGAR COLLEGE OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING

Minor Project- Report Aug-2021-2022

Course Faculty: Dr. Vinothini C

Course Name & Code: Cloud & Big Data Laboratory with Minor Project & 18CS7DLCBL

Semester: 7

Date: 19-1-2022

TITLE OF THE PROJECT	Heart Disease Analysis		
STUDENT NAME	Gautam Kumar	Rithvik K Bhat	Vignesh K
USN	1DS18CS710	1DS18CS732	1DS18CS744
INDIVIDUAL CONTRIBUTION	AWS Random Forest Algorithm	Decision Tree Algorithm Random Forest Algorithm	AWS Decision Tree Algorithm
GUIDE	Dr. Vinothini C		
PROJECT ABSTRACT	Heart Disease prediction is one of the most complicated tasks in medical field. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. This project uses the heart disease data set available on the AWS cloud S3 platform. The project predicts the chances of heart disease and classifies patient's risk level by implementing two data mining techniques they are Decision Tree and Random Forest. Thus, this project analyses the performance of the two machine learning algorithms. The trial results verify that Random Forest algorithm has achieved the highest accuracy compared to Decision Tree.		
PLATFORM USED (H/W & S/W TOOLS TO BE USED)	Jupyter Notebook & Amazon Web Services S3		

DAYANANDA SAGAR COLLEGE OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING

INTRODUCTION

Human heart is the main part of the human body which regulates blood flow throughout our body. Any irregularity to the heart can cause distress in other parts of the body and can be classified as a heart disease. Today, heart disease is one of the primary reasons for the occurrence of most deaths. Heart disease may occur due to unhealthy lifestyle, smoking, alcohol and high intake of fat which may cause hypertension.

The main challenge in today's healthcare is provision of best quality services and accurate diagnosis. The accuracy in management of a disease lies in the proper time of detection of that disease. The project makes an attempt to detect these heart diseases at an early stage to avoid disastrous consequences.

Records of large set of medical data are available for analyzing and extracting valuable knowledge from it. Data mining techniques are used for extracting valuable information from the large data available. The medical database consists of discrete information which makes decision making a complex and tough task. Machine Learning (ML) which is a subfield of data mining handles large scale well-formatted dataset efficiently. In the medical field, machine learning can be used for diagnosis, detection and prediction of various diseases. The main goal of our project is to provide a tool for doctors to detect heart disease at an early stage. This in turn will help to provide effective treatment to patients and avoid severe consequences. ML plays a very important role to detect the hidden discrete patterns and analyze the given data. This project shows the performance analysis of two ML techniques: Decision Tree and Random Forest for predicting heart disease at an early stage.

DESIGN

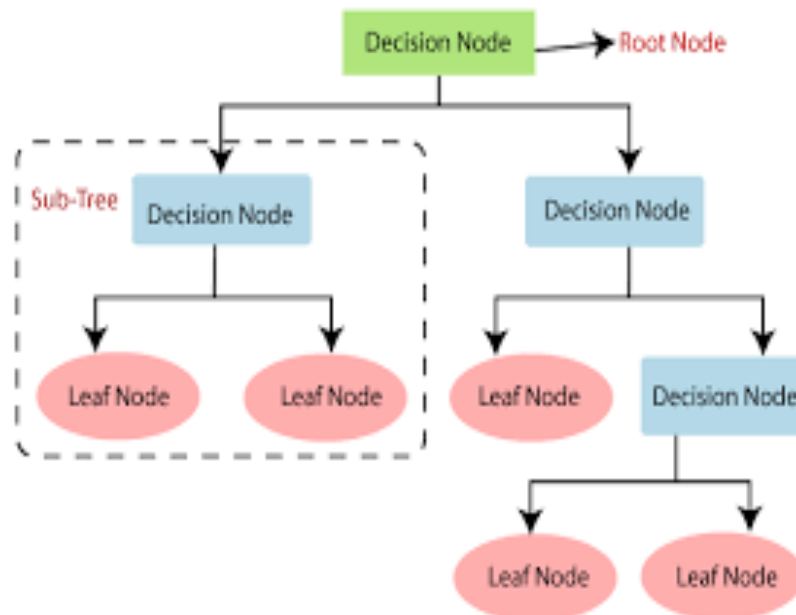
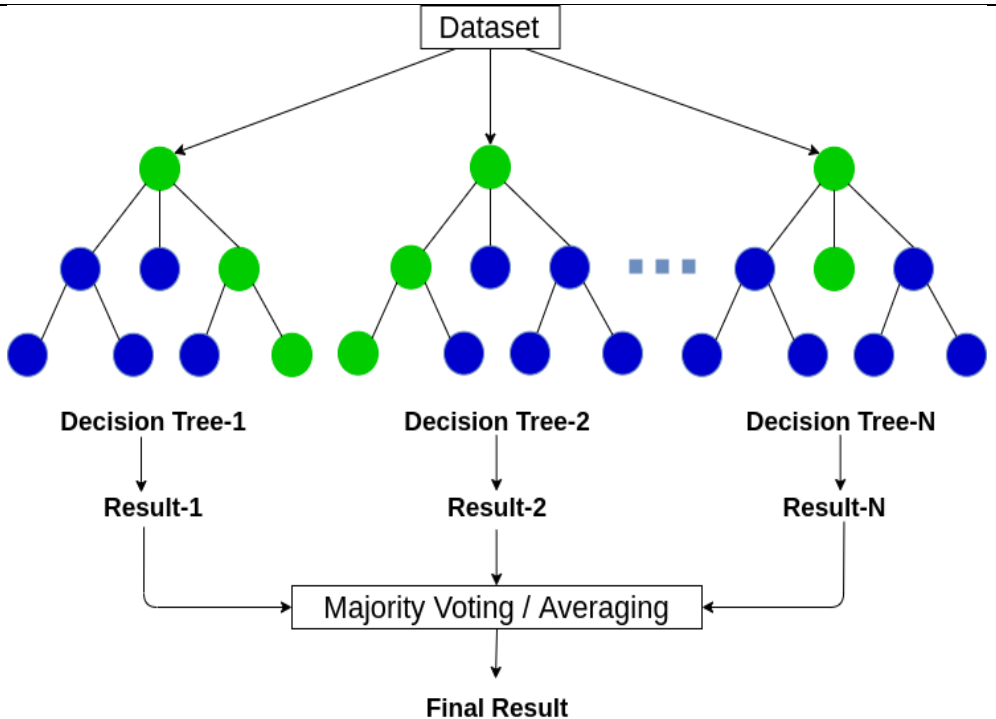


Fig. 1 – Decision Tree Algorithm

DAYANANDA SAGAR COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING



	 <p style="text-align: center;">Fig.2 – Random Forest Algorithm</p>
<p>PROJECT SOURCE CODE LINK (GITHUB/ GOOGLE DRIVE)</p>	<p>https://github.com/gautamK007/Heart-Disease-Analysis</p>
<p>CONCLUSION /FUTURE ENHANCEMENT</p>	<p>Predicting heart disease is one of the most difficult problems in medicine. In the realm of healthcare, data science is critical for examining massive amounts of data. Because predicting cardiac illness is a difficult undertaking, it is necessary to automate the process in order to avoid the risks connected with it and to inform the patient well in advance. This project makes use of the AWS cloud S3 platform's heart disease data set. The study uses two data mining approaches, Decision Tree and Random Forest, to predict the likelihood of heart disease and classify the risk level of patients. As a result, the performance of the two machine learning algorithms is examined in this study. The trial results show that the Random Forest algorithm gives the best results. For future enhancements, the project can be used to a variety of machine algorithms. Using visuals to detect cardiac disease could be a future improvement as well.</p>

DAYANANDA SAGAR COLLEGE OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING

UI SCREENSHOTS

The top screenshot shows the Amazon S3 Management Console interface. The left sidebar contains navigation links for Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, Access analyzer for S3, Storage Lens, Dashboards, AWS Organizations settings, and Feature spotlight. The main content area displays the 'Buckets (1)' page for the bucket 'bigdatatestver1' in the Asia Pacific (Mumbai) region. The bucket's access is set to 'Bucket and objects not public'. The bottom screenshot shows the 'Objects (1)' page for the same bucket. It displays a single object named 'heart.csv' with a size of 10.8 KB and a storage class of 'Standard'. The object was last modified on December 17, 2021, at 20:40:04 (UTC+05:30). The interface includes various action buttons like 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Create folder', and 'Upload'.

Amazon S3 Buckets (1) Info

Name	AWS Region	Access	Creation date
bigdatatestver1	Asia Pacific (Mumbai) ap-south-1	Bucket and objects not public	December 16, 2021, 16:45:36 (UTC+05:30)

Amazon S3 bigdatatestver1

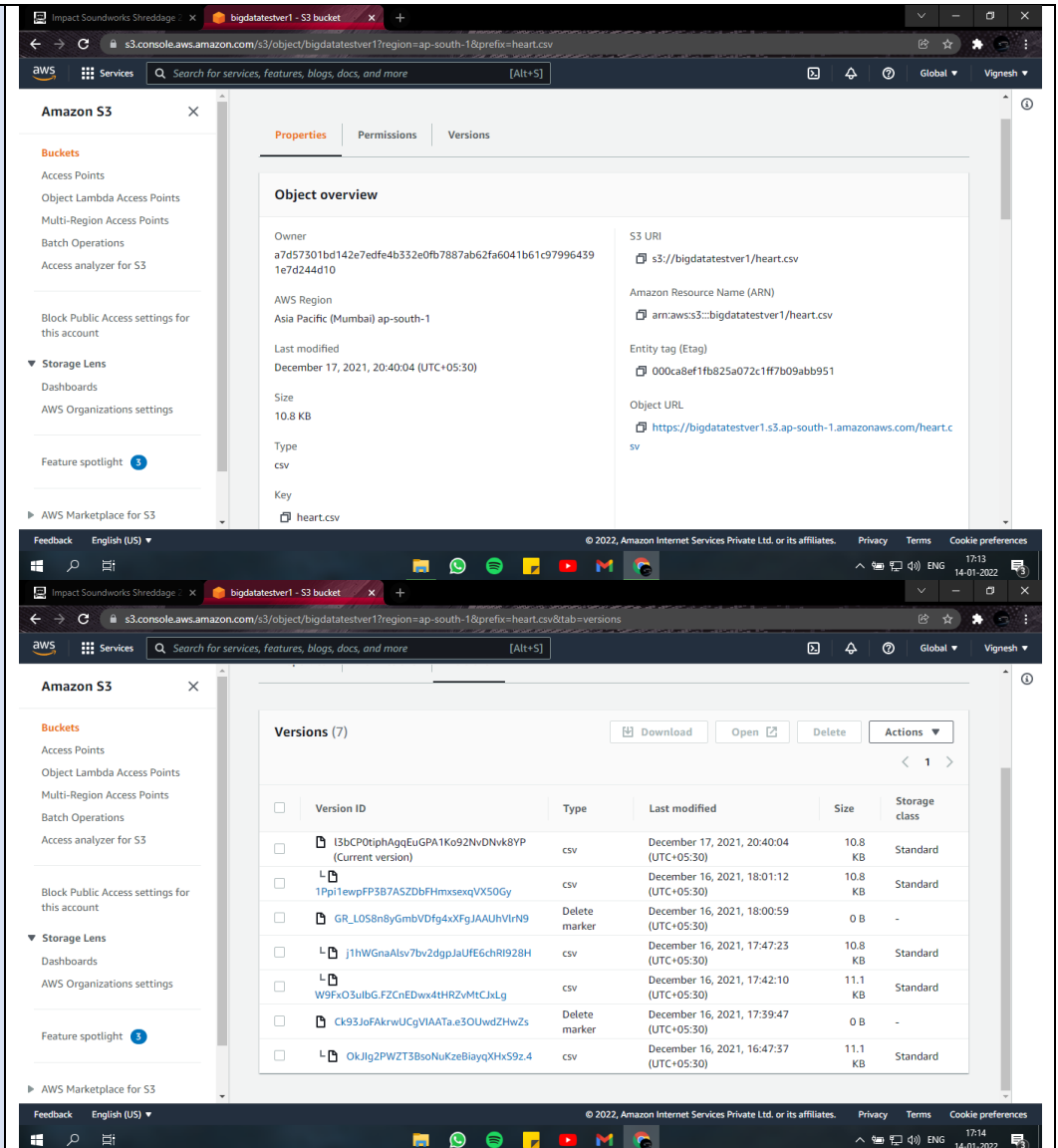
bigdatatestver1 Info

Objects (1)

Name	Type	Last modified	Size	Storage class
heart.csv	csv	December 17, 2021, 20:40:04 (UTC+05:30)	10.8 KB	Standard

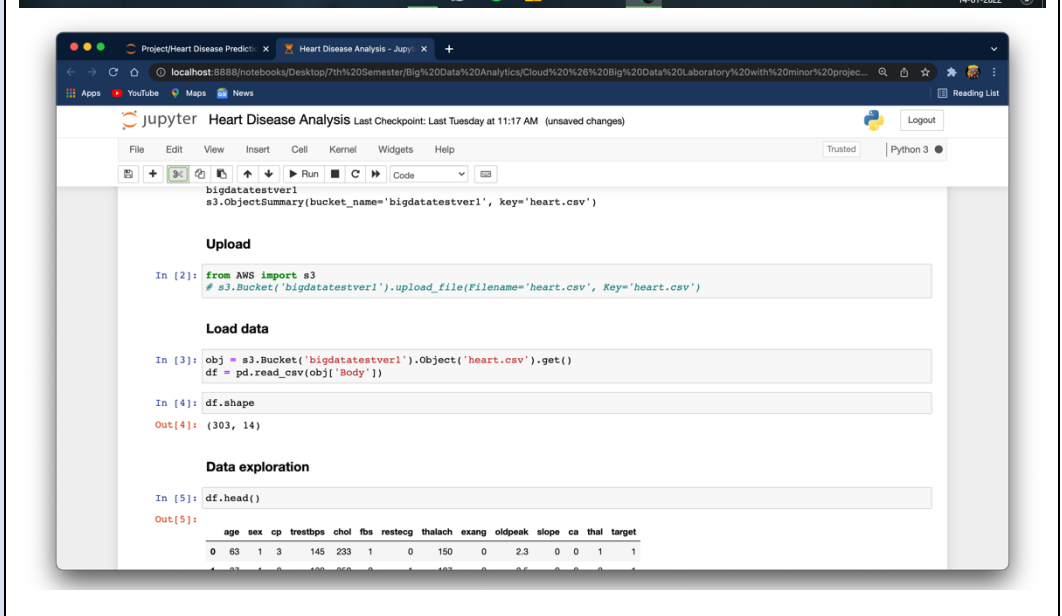
DAYANANDA SAGAR COLLEGE OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING



The screenshot displays the AWS S3 console interface. The left sidebar shows the navigation menu with options like Buckets, Access Points, and Storage Lens. The main content area is divided into two sections: 'Object overview' and 'Versions (7)'. The 'Object overview' section shows details for the object 'heart.csv', including its owner, AWS Region (Asia Pacific (Mumbai) ap-south-1), last modified date (December 17, 2021, 20:40:04 UTC+05:30), size (10.8 KB), type (csv), and key (heart.csv). The 'Versions (7)' section shows a list of object versions with columns for Version ID, Type, Last modified, Size, and Storage class. The current version is '13bCP0tiphAggEuGPA1Ko92NvDNvk8YP (Current version)'.

Version ID	Type	Last modified	Size	Storage class
13bCP0tiphAggEuGPA1Ko92NvDNvk8YP (Current version)	csv	December 17, 2021, 20:40:04 (UTC+05:30)	10.8 KB	Standard
1Pp11ewpFP3B7ASZDbFHmsenqVX50Gy	csv	December 16, 2021, 18:01:12 (UTC+05:30)	10.8 KB	Standard
GR_L058n8yGmbVDFg4xJfGJAAUHVrN9	Delete marker	December 16, 2021, 18:00:59 (UTC+05:30)	0 B	-
j1HWGnaAlsv7bv2dgpJaUFE6chRi928H	csv	December 16, 2021, 17:47:23 (UTC+05:30)	10.8 KB	Standard
W9Fx03ulbG.FZCnEDwx4tHRZvMtCJkLg	csv	December 16, 2021, 17:42:10 (UTC+05:30)	11.1 KB	Standard
Ck93JoFAkrwUCgVIAATa.e3OUwdZhwZs	Delete marker	December 16, 2021, 17:39:47 (UTC+05:30)	0 B	-
OkJlg2PWZT3BsoNuKzeBlayqXHxS9z4	csv	December 16, 2021, 16:47:37 (UTC+05:30)	11.1 KB	Standard



The screenshot displays a Jupyter Notebook interface with the title 'Heart Disease Analysis'. The notebook contains code for uploading a file to S3 and loading data. The code is as follows:

```
bigdatatestver1
s3.ObjectSummary(bucket_name='bigdatatestver1', key='heart.csv')

Upload

In [2]: from AWS import s3
# s3.Bucket('bigdatatestver1').upload_file(Filename='heart.csv', Key='heart.csv')

Load data

In [3]: obj = s3.Bucket('bigdatatestver1').Object('heart.csv').get()
df = pd.read_csv(obj['Body'])

In [4]: df.shape
Out[4]: (303, 14)

Data exploration

In [5]: df.head()
Out[5]:
```

age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1

DAYANANDA SAGAR COLLEGE OF ENGINEERING

COMPUTER SCIENCE & ENGINEERING

