VISVESVARAYA TECHNOLOGICAL UNIVERSITY



BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

"Machine learning algorithms predicting the Risk of chronic disease"

Submitted in partial fulfilment for the award of degree(18ECI85)

BACHELOR OF ENGINEERING IN Electronics and communication Engineering

Submitted by:

SOUNDARYA S 1AT18EC144



Conducted at **COMPANY NAME**



ATRIA INSTITUTE OF TECHNOLOGY BANGALORE Department of Electronics and Communication Engineering

ATRIA INSTITUTE OF TECHNOLOGY BANGALORE

Department of Electronics and Communication Engineering



CERTIFICATE

This is to certify that the Internship titled "Machine Learning algorithm for predicting the Risk of chronic diseases" carried out by Miss SOUNDARYA S, a bonafide student of Atria Institute of Technology, Bangalore in partial fulfillment for the award of Bachelor of Engineering, in Electronics and communication Engineering under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18ECI85)

Signature of Guide	Signature of HOD	Signature of Principal
	External Viva:	
Name of the Examiner		Signature with Date
1)		
2)		
· · · · · · · · · · · · · · · · · · ·		

Internship report 2022-2023 2

DECLARATION

I, **SOUNDARYA S**, final year student of Electronics and communication Engineering , Atria Institute of Technology, declare that the Internship has been successfully completed, in **ICSoln Technology**. This report is submitted in partial fulfillment of the requirements for award of Bachelors in Engineering in Electronics and communication Engineering, during the academic year 2022-2023.

Date: 05/10/2022 :

Place: Bangalore

USN: 1AT18EC144

NAME : Soundarya S

OFFER LETTER

INTERNSHIP OFFER LETTER



Date: 2nd September, 2022

Name: Soundarya S USN: 1AT18EC144

Dear Student,

We would like to congratulate you on being selected for the Machine Learning With-Python(Research Based) Internship position with ICsoln, effective Start Date 2nd September, 2022, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a parttime job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of Machine Learning With Python(Research Based) through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Ganesh G
Product Manager
ICSOLN
GMCH Hostel Rd, Kachari Basti Rd
Ganeshguri, Guwahati, Assam781005

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several important

persons. We take this opportunity to express our gratitude to all who have helped us to complete

the Internship.

We express our sincere thanks to our Principal, for providing us adequate facilities to undertake

this Internship.

We would like to thank our Head of Dept, for providing us an opportunity to carry out

Internship and for his valuable guidance and support.

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate

Prof, for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended

during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the

Internship.

Last but not the least, we would like to thank our parents and friends without whose constant

help, the completion of Internship would have not been possible.

NAME: SOUNDARYA S

USN: 1AT18EC144

ABSTRACT

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. Our project is implemented using machine learning technique which is based on a simple algorithm based on feature extraction and model creation. It involves the calculation of feature parameters from the raw data. The features are used to train a model that learns to produce the desired output labels.

Among various life-threatening diseases, heart disease has garnered a great deal of attention in medical research. The diagnosis of heart disease is a challenging task, which can offer automated prediction about the heart condition of patient so that further treatment can be made effective. The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. The available heart disease database consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the database.

Table of Contents

Sl no	Description	Page no
1	Company Profile	
2	About the Company	
3	Introduction	
4	System Analysis	
5	Requirement Analysis	
6	Design Analysis	
7	Implementation	
8	Snapshots	
9	Conclusion	
10	References	

Internship report 2022-2023 7

	CHAPTER 1 COMPANY PROFILE	
Internship report	2022-2023	8

1. COMPANY PROFILE

A Brief History of ICSoln Technologies

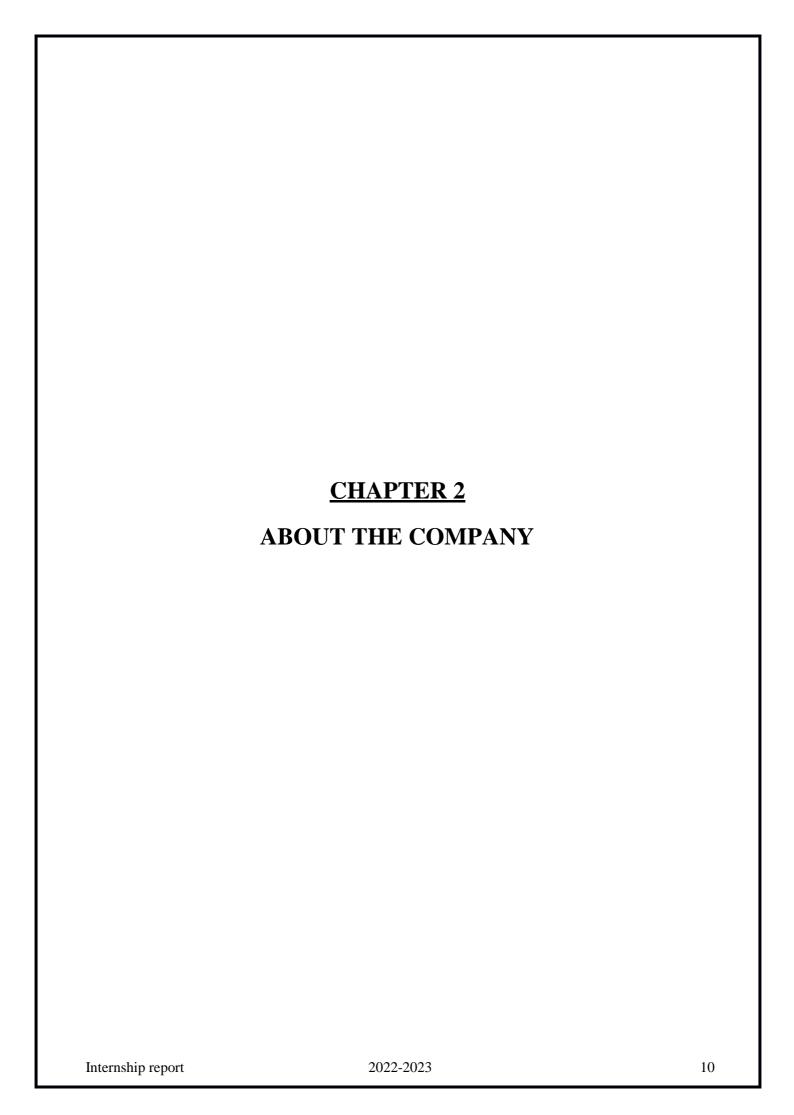
ICSoln Technologies, was incorporated with a goal "To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

ICSoln Technologies, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. ICSoln Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence "Technology helps you to Delight your Customers" and that is what we want to achieve.



2. ABOUT THE COMPANY



ICSoln Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, ICSoln Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well". Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

Products of ICSoln Technologies.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and zutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but software development is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client- side logic) runs in a web browser. Common web applications include web mail, online

retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can beconsidered as a specific variant of client—server software where the client software is downloaded to the client machine when visiting the relevant web page, using standardprocedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, andacts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security- related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a "smart" client that performs all the work and queries a "dumb" server, or a "dumb" client that relies on a "smart" server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn't allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It is encompasses many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

ICSoln Technologies plays an essential role as an institute, the level of education, development of student's skills are based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at ICSoln Technologies gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of ICSoln Technologies always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by ICSoln Technologies.

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Training

	CHAPTER 3	
	INTRODUCTION	
Internship report	2022-2023	14

3. INTRODUCTION

Introduction to ML

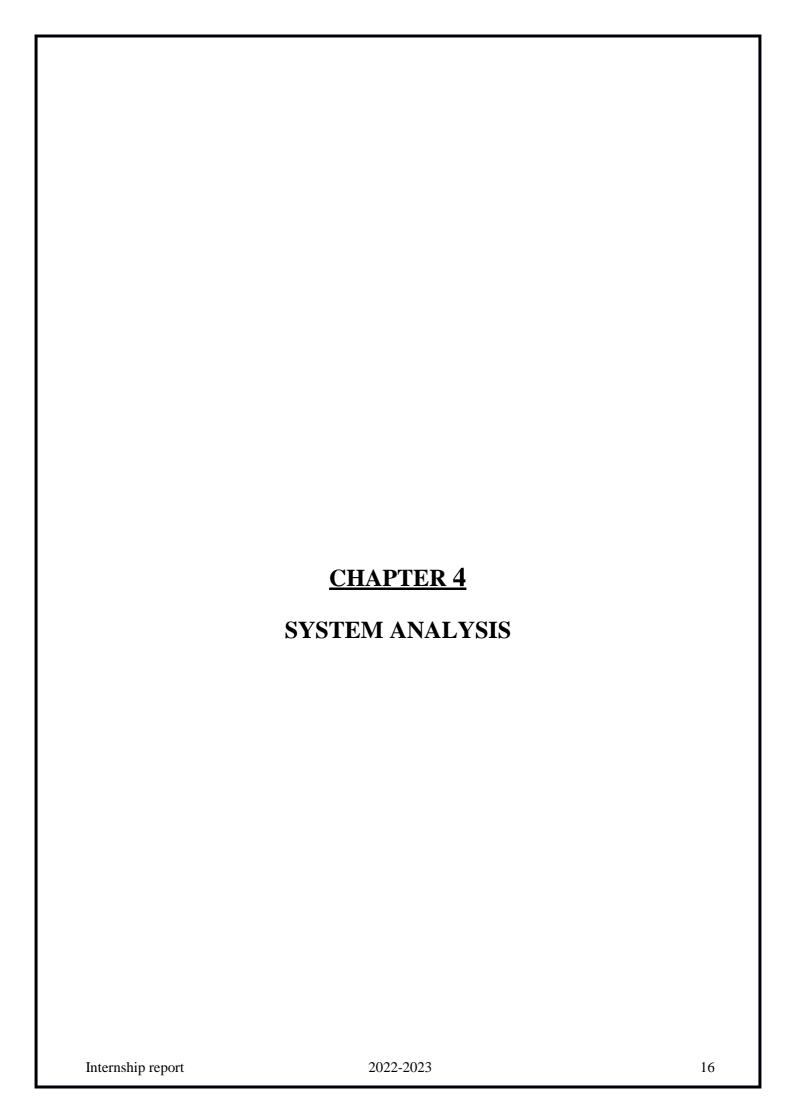
Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data. ML has a modules facility to support the incremental construction of large programs. An ML program is constructed as an interdependent collection of structures which are glued together using functions. Separate compilation is supported through the ability to export and import functions.

Machine Learning can be thought of as the study of a list of sub-problems, such as decision making, clustering, classification, forecasting, deep-learning, inductive logic programming, support vector machines, reinforcement learning, similarity and metric learning, genetic algorithms, sparse dictionary learning, etc. Supervised learning, or classification is the machine learning task of inferring a function from a labeled data. In Supervised learning, we have a training set, and a test set. The training and test set consists of a set of examples consisting of input and output vectors, and the goal of the supervised learning algorithm is to infer a function that maps the input vector to the output vector with minimal error. In an optimal scenario, a model trained on a set of examples will classify an unseen example in a correct fashion, which requires the model to generalize from the training set in a reasonable way. In layman's terms, supervised learning can be termed as the process of concept learning, where a brain is exposed to a set of inputs and result vectors and the brain learns the concept that relates said inputs to outputs. A wide array of supervised machine learning algorithms are available to the machine learning enthusiast, for example Neural Networks, Decision Trees, Linear Regression, Support Vector Machines, Random Forest, Naïve Bayes Classifier, Bayes Net, Majority Classifier etc., and they each have their own merits and demerits. There is no single algorithm that works for all cases, as merited by the No free lunch theorem. In our project, we try and find patterns in a classification of audio, and attempt to throw various intelligently-picked algorithms at the data, and see what sticks.

Problem Statement

Machine Learning algorithm for predicting the Risk of chronic diseases.

Implementing a Machine Learning algorithm to predict the risk of Heart disease in a person Using the existing data of the people.



4. SYSTEM ANALYSIS

1. Existing System

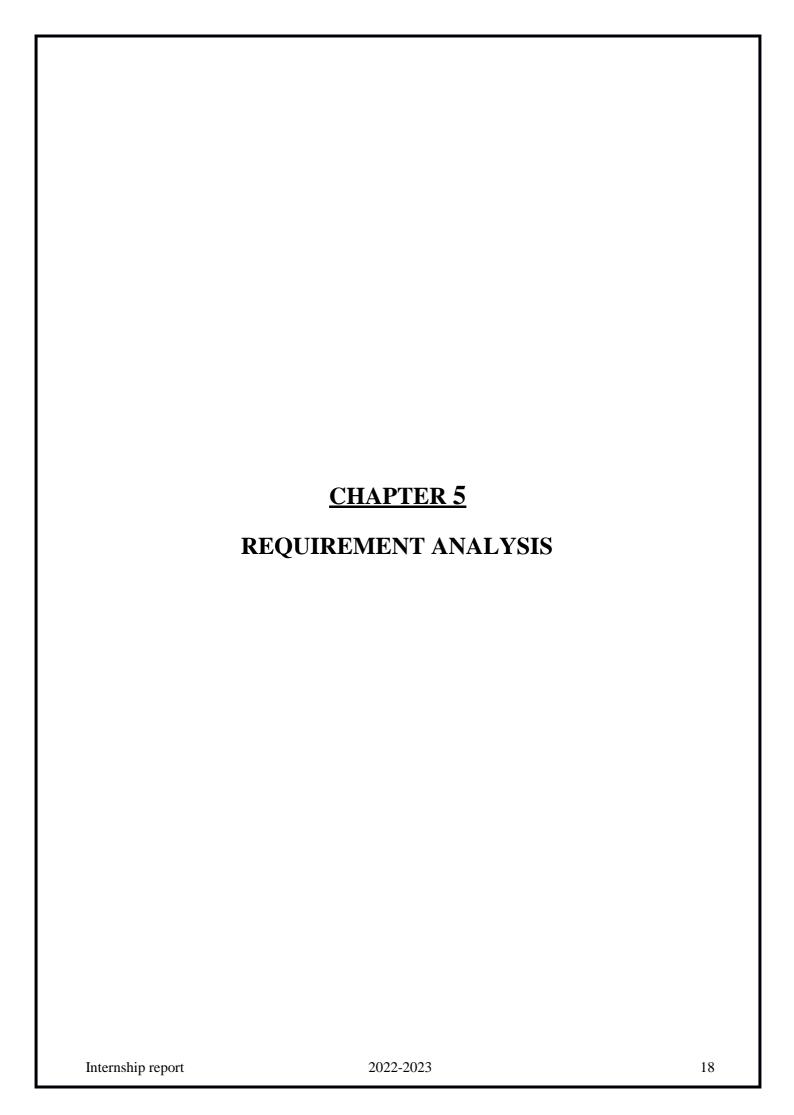
The first and the most basic objective is to use a system is to read the dataset and provide more accurate results.

2. Proposed System

The objective of the proposed system is to take the existing module and test it with various datasets, noting down the conclusions, and using different functions or modules, updated ones, along with different algorithms to make more accurate algorithm. At the end of the day, they are machines and they can only perform so much, we need to try to improve the model as much as we can to prove efficient.

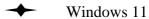
3. Objective of the System

The objective of the system is to provide the information if the person has the risk of heart disease or not. The more data we will be provided, the higher will be the performance.



5. <u>REQUIREMENT ANALYSIS</u>

Hardware Requirement Specification



→ 4 GB RAM recommended

Software Requirement Specification

Language: Python
Jupyter notebook

Libraries: numpy, sklearn, pandas

	CHAPTER 6 DESIGN ANALYSIS	
Internship report	2022-2023	20

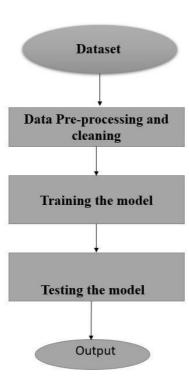
6. DESIGN & ANALYSIS

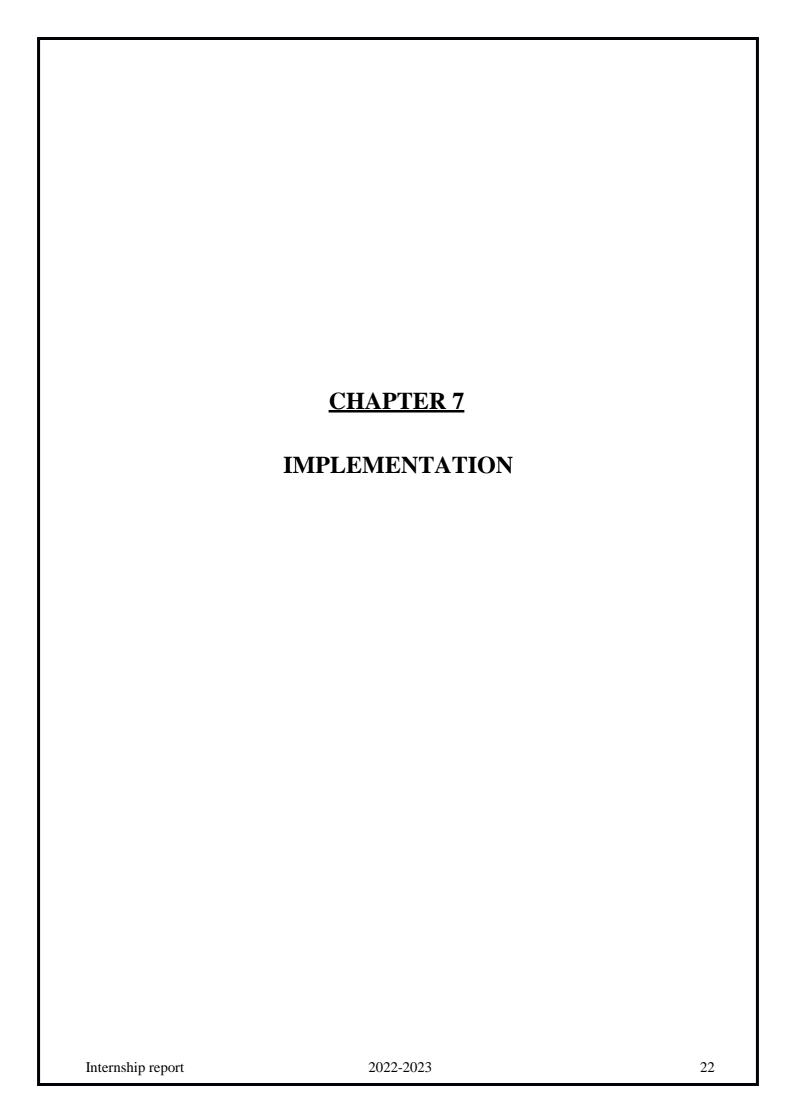
Gathering the Data: Data preparation is the primary step for any machine learning problem. We will be using a dataset from Kaggle. This dataset consists of one CSV file which will be used for both for training and for testing. There is a total of 14 columns and 1025 rows describing the data.

Cleaning the Data: Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine learning model. So it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are numerical.

Model Building: After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We will be using this cleaned data to train the algorithm.

Inference: After training the model we will be predicting the if the person suffers from heart disease or not from the input parameters by using the prediction model. The algorithm predicts in a more robust and accurate way.





7. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

TESTING

The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

- 1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
- 2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
- 3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

	CHAPTER 8	
	SNAPSHOTS	
Internship report	2022-2023	24

8. SNAPSHOTS

```
#loading .csv file data to a Pandas Dataframe
heart_data =pd.read_csv('..\ML-python\heart.csv')
heart_data
# heart_data.head()
# heart_data.tail()

v 0.8s
```

	age	sex	ср	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	output
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

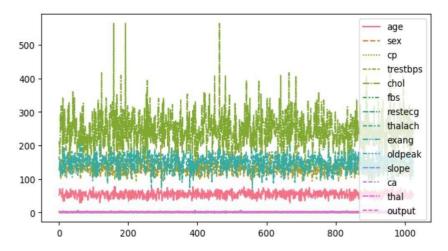
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows × 14 columns

```
plt.figure(figsize=(8,4))
sns.lineplot(data = heart_data)

4.9s
```

<AxesSubplot:>



heart_data.info() ✓ 0.6s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):

#	Column	Non-N	Wull Count	Dtype
0	age	1025	non-null	int64
1	sex	1025	non-null	int64
2	ср	1025	non-null	int64
3	trestbps	1025	non-null	int64
4	chol	1025	non-null	int64
5	fbs	1025	non-null	int64
6	restecg	1025	non-null	int64
7	thalach	1025	non-null	int64
8	exang	1025	non-null	int64
9	oldpeak	1025	non-null	float64
10	slope	1025	non-null	int64
11	ca	1025	non-null	int64
12	thal	1025	non-null	int64
13	output	1025	non-null	int64
	722 9 2			

dtypes: float64(1), int64(13)

memory usage: 112.2 KB



```
heart_data['output'].value_counts()

✓ 0.6s

Python
```

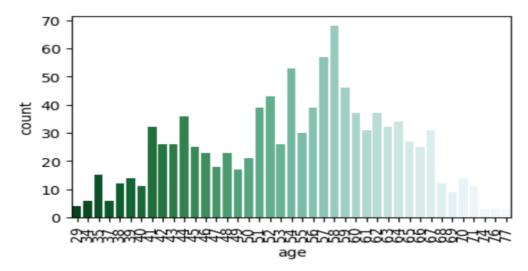
1 526 0 499

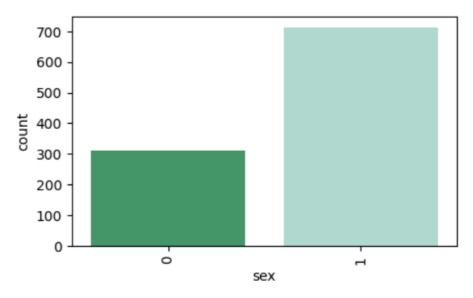
Name: output, dtype: int64

```
age sex cp trestbps chol fbs restecg thalach exang oldpeak \
0
     52
         1
            0
                  125
                       212
                            0
                                 1
                                         168
                                                0
                                                      1.0
     53
                   140
1
                                         155
                                               1
         1 0
                      203
                            1
                                   0
                                                      3.1
2
                                         125
        1 0
                   145 174
                           0
                                    1
                                               1
                                                      2.6
                      203
                                         161
     62
        0 0
                   138
                       294
                             1
                                    1
                                         106
                                                0
                                                      1.9
                   . . .
                        . . . . . . .
                                         . . .
                                               . . .
                                                      . . .
1020
    59
        1
            1
                   140
                       221
                            0
                                  1
                                         164
                                               1
                                                      0.0
1021
     60
                   125
                       258
                             0
                                         141
1022
     47
                       275
         1
            0
                   110
                            0
                                    0
                                         118
                                                1
                                                      1.0
1023
                                                      0.0
    50 0 0
                   110 254 0
                                  0
                                         159
                                                0
1024 54 1 0
                   120 188 0
                                  1
                                         113
                                                      1.4
    slope ca thal
       2
          2
0
       0
               3
1
2
       0
3
       1 3
1020
      2 0
1021
1022
      1 1
               2
1023
      2 0
               2
```

```
age sex cp trestbps chol fbs restecg thalach exang oldpeak \
             0
                                  1
                                                   0
0
     52
          1
                    125
                         212
                              0
                                            168
                                                         1.0
                    140
1
          1 0
                         203
                                      0
                                            155
                                                   1
                                                         3.1
     53
                               1
2
     70
          1 0
                    145
                        174
                             0
                                      1
                                           125
                                                   1
                                                         2.6
                    148
                         203
                                            161
                    138
                        294
                               1
                                            106
                                                         1.9
                    . . .
                                            . . .
                                                         . . .
1020
     59
          1
             1
                    140
                         221
                              0
                                     1
                                            164
                                                   1
                                                         0.0
1021
     60
          1
                    125
                         258
                                      0
                                            141
                                                         2.8
                                 0
1
1022
     47
                         275
                                            118
          1
             0
                    110
                              0
                                                   1
                                                         1.0
1023
     50 0 0
                    110
                        254 0
                                            159
                                                   0
                                                         0.0
1024 54
        1 0
                    120 188
                             0
                                            113
                                                         1.4
```

```
slope ca thal
      2 2
0
       0
          0
1
2
       0
          0
       1
1020
      2 0
1021
1022
       1 1
       2 0
1023
               2
```





Internship report 2022-2023 28

```
#Accuracy prediction on the Training data

X_train_prediction = model.predict(X_train)

training_data_accuracy = accuracy_score(X_train_prediction,Y_train)

print("Accuracy of the Model on Training Data: ",training_data_accuracy)

#Accuracy prediction on the Test data

X_test_prediction = model.predict(X_test)

test_data_accuracy = accuracy_score(X_test_prediction,Y_test)

print("Accuracy of the Model on the Test Data: ",test_data_accuracy)

✓ 0.7s
```

Accuracy of the Model on Training Data: 0.8524390243902439
Accuracy of the Model on the Test Data: 0.8048780487804879

```
#Predictive System
input_data = (58,0,0,100,248,0,0,122,0,1,1,0,2)
inputData_as_numpyArray = np.asarray(input_data)
inputData_reshaped = inputData_as_numpyArray.reshape(1,-1)

predict = model.predict(inputData_reshaped)
print(predict)

if(predict[0] == 0):
    print("the person suffers from heart disease")
else:
    print("the person does not suffer from heart disease")
```

[1] the person does not suffer from heart disease

```
#Predictive System
input_data = (554,1,0,124,266,0,0,109,1,2.2,1,1,3)
inputData_as_numpyArray = np.asarray(input_data)
inputData_reshaped = inputData_as_numpyArray.reshape(1,-1)
predict = model.predict(inputData_reshaped)
print(predict)

if(predict[0] == 0):
    print("the person suffers from heart disease")
else:
    print("the person does not suffer from heart disease")
```

[0]
the person suffers from heart disease

	CHAPTER 9 CONCLUSION	
Internship report	2022-2023	31

9. CONCLUSION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- ❖ Automation of the entire system improves the efficiency
- ❖ It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- **!** It gives appropriate access to the authorized users depending on their permissions.
- ❖ It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- ❖ The System has adequate scope for modification in future if it is necessary.

10. REFERENCE

Geeks for Geeks

https://www.geeksforgeeks.com

Blog: Stack overflow

https://www.stackoverflow.com

W3 Schools

https://www.w3schools.com