VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Machine Learning Algorithms for Predicting the Risks of Chronic Diseases"

Submitted in partial fulfillment for the award of degree(18CSI85)

BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE & ENGINEERING

Submitted by:

SAAIMA NISHAT

1BI19CS133

Conducted at
ICsoln

BANGALORE INSTITUTE OF TECHNOLOGY
Department of CSE
K.R. Road, V.V. Pura, Bengaluru-560 004

BANGALORE INSTITUTE OF TECHNOLOGY
Department of CSE
K.R. Road, V.V. Pura, Bengaluru-560 004

CERTIFICATE

This is to certify that the Internship titled **"Machine Learning Algorithms for Predicting the Risks of Chronic Diseases"** carried out by **Ms. Saaima Nishat,** a bonafide student of Bangalore Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering**, in **CSE** 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 the Internship prescribed for the course Internship / Professional Practice (18CSI85)

Signature of Guide **Signature of HOD** **Signature of Principal**

External Viva:

Name of the Examiner Signature with Date

1)_____

2)_____

D E C L A R A T I O N

I, **Saaima Nishat**, final year student of CSE, Bangalore Institute of Technology - 560 004, declare that the Internship has been completed, in **ICsoln**. This report is submitted in partial fulfillment of the requirements for the award of Bachelor's Degree in CS&E, during the academic year 2022-2023.

Date: 05-10-2022 :

Place: Bangalore

USN: 1BI19CS133

NAME: Saaima Nishat

OFFER LETTER

Internship Offer Letter

Date: **2nd September, 2022**

Name: **Saaima Nishat**
USN: **1BI19CS133**

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 part-time 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, Assam-
781005

A C K N O W L E D G E M E N T

This Internship is a result of the 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 with adequate facilities to undertake this Internship.

We would like to thank our Head of Dept of CS&E, for providing us an opportunity to carry out our Internship and for his valuable guidance and support.

We would like to thank ICsoln Software Services for guiding us during the period of internship.

We express our 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 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 the Internship would have not been possible.

SAAIMA NISHAT
1BI19CS133

ABSTRACT

Technological development, including machine learning, has a huge impact on health through effective analysis of various chronic diseases for more accurate diagnosis and successful treatment. In the field of biomedical and healthcare communities, accurate prediction plays a major role to find out the risk of the disease in the patient. The only way to overcome the mortality due to chronic diseases is to predict it earlier so that disease prevention can be done. Such a model is a Patient's need in which Machine Learning is highly recommendable. But the precise prediction based on symptoms becomes too difficult for the doctor. The correct prediction of disease is the most stretching task. To overcome this problem data mining plays an important role to predict the disease. This study analyzes chronic diseases using machine learning techniques based on a chronic disease dataset from the UCI machine learning data warehouse. We use Heart disease, Kidney disease, Cancer disease, and Diabetes disease datasets, To build reliable prediction models for these chronic diseases using data mining techniques. The most relevant features are selected from the dataset for improved accuracy and reduced training time. The system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease Prediction is done by implementing the Logistic Regression. By using logistic regression, random forest, and decision tree we are predicting diseases like Diabetes, Heart, Cancer, and Kidney. For each chronic disease, diverse models, techniques, and algorithms are used for predicting and analyzing. The paper comprises a conceptual model that integrates the prediction of the most common chronic diseases.

Table of Contents

Sl no	Description	Page no
1	Company Profile	8
2	About the Company	10
3	Introduction	14
4	System Analysis	16
5	Requirement Analysis	18
6	Design Analysis	20
7	Implementation	25
8	Snapshots	27
9	Conclusion	34
10	References	36

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of ICsoln Technologies

ICsoln Technologies was incorporated with a goal of “To provide high quality and optimal Technological Solutions to business requirements of our clients”. Every business is 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 client's requirements and provide them with tailor-made technological solutions.They also understand that the 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 client's requirements.

ICsoln Technologies, strive to be the front runner in creativity and innovation in software development through its 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 client's demands better. ICsoln Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder if 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, and reliability; to put it in one sentence ”Technology helps you to Delight your Customers” and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

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 the right mix of professionals as a stakeholder to help us serve our clients to the 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 solutions. The 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 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 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 tutorials. 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 webmail, 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 isdownloaded 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 usecase. In applications that 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. Some 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 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 a 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 that follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches in these kinds of applications are a major concern because they can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and some key operational areas 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 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 markup. 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 marks up 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, and development of students' 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 give you the facility of skilled employees so that you do not feel unsecured aboutyour academics. Personality development and academic status are some of those things which lie in the mentor's hands. If you are trained well then you can do well in your future and knowing the importance of ICsolnTechnologies 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 mentor can develop their skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by ICsolnTechnologies.

• Core Java and Advanced Java

• Web services and development

• Dot Net Framework

• Python

• Selenium Testing

• Conference / Event Management Service

• Academic Project Guidance

• The Job Training

• Software Training

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

Introduction to ML

Machine learning is programming computers to optimize a performance using example data or past data. Machine learning is the study of computer systems that learn from data and experience. ML is categorized as supervised (i.e., consists of output variables that are predicted from input variables) or unsupervised (i.e., deals with clustering of different groups for a particular intercession). ML is used to determine complex models, and extract medical knowledge, exposing novel ideas to professionals, and specialists. In clinical practice, ML predictive models can highlight strengthen rules in the decision-making regarding individual patient care. These are also capable of autonomous diagnosis of different diseases under clinical rules. The incorporation of these models in drug prescription can save doctors and offer new medical opportunities for identification.

Problem Statement

Machine Learning Algorithms for Predicting the Risks of Chronic Diseases

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. Existing System

Chronic diseases are growing to be one of the prominent causes of death worldwide. There is an increasing percentage of the world population facing the adverse health effects of living. In general, the patient's reports have to be carefully scrutinized by doctors to make a diagnosis of the disease. Since the diagnosis is manual sometimes it is difficult for the doctors to treat patients efficiently. The number of people suffering from Chronic Diseases is rising day by day. Conventional Health Care is passive. Due to this type, patients can die due to a lack of proper treatment during emergencies such as cardiac arrest. The key to improving Health Care efficiency is to reduce the mortality rate due to lack of proper treatment and to transform the passive Health Care program into a continuous one at a reduced cost.

2. Proposed System

Due to the low-progress nature of Chronic Diseases, it is important to make an early prediction and provide effective medication. Therefore, it is essential to propose a decision model which can help to diagnose chronic diseases and predict future patient outcomes.

3. The objective of the System

While there are many ways to approach this in the field of AI, the present study focuses distinctly on ML predictive models used in the diagnosis of Chronic Diseases. In comparison to the conventional data analysis techniques, we will be able to find promising results that enhance the quality of patient data and inspect specific items that are related to ML algorithms in medical care. The main purpose of our project is to make hospital tasks easy and to develop efficient and feasible software that replaces the manual prediction system with an automated healthcare management system. Our project enables healthcare providers to improve operational effectiveness and reduce medical errors and time consumption. If the disease can be predicted, then early treatment can be given to the patients which can reduce the risk of life and save the life of patients. The cost to get treatment for diseases can also be reduced to an extent by early recognition.

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- Operating System — Ubuntu or Microsoft Windows 10. I recommend updating Windows 10 to the latest version before proceeding forward.
- Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher. An AMD equivalent processor will also be optimal.
- RAM — 8 GB minimum, 16 GB or higher is recommended.
- Graphics Processing Unit (GPU) — NVIDIA GeForce GTX 960 or higher. AMD GPUs are not able to perform deep learning regardless.

Software Requirement Specification

- Download Anaconda
- Install Anaconda & Python
- Start and Update Anaconda
- Install CUDA Toolkit & cuDNN
- Create an Anaconda Environment
- Install APIs (TensorFlow & Keras)

CHAPTER 6

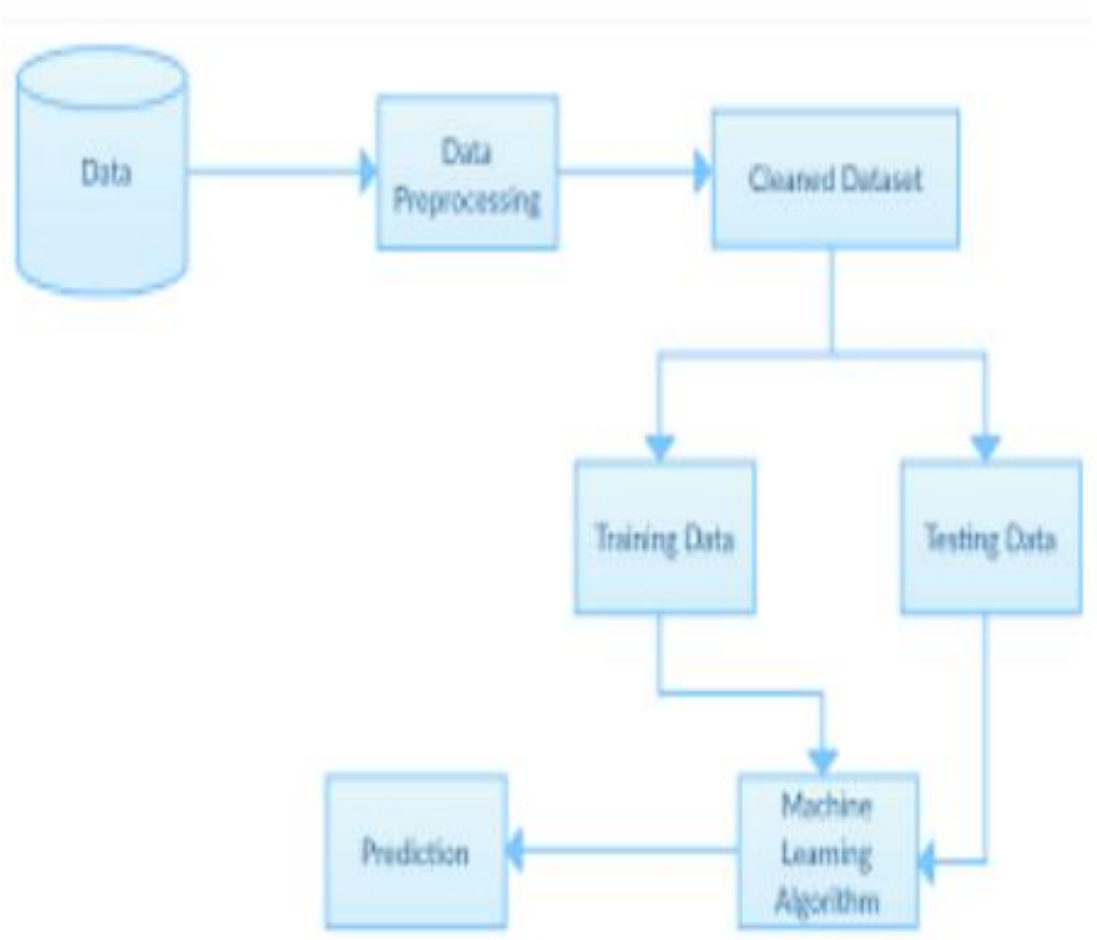
DESIGN ANALYSIS

6. DESIGN & ANALYSIS

SYSTEM DESIGN

The design goals consist of various designs which we have implemented in our system "Chronic Disease Prediction Using Machine Learning". This system is built with various designs such as a data flow diagram, sequence diagram, class diagram, use case diagram, and activity diagram. We have designed our system in such a way that the registration process is solely done by the administrator. After the registration process, the users i.e. doctors can log in into the system using their credentials. Based on the inputs/attributes given, doctors will be able to predict the chronic disease accordingly

SYSTEM ARCHITECTURE

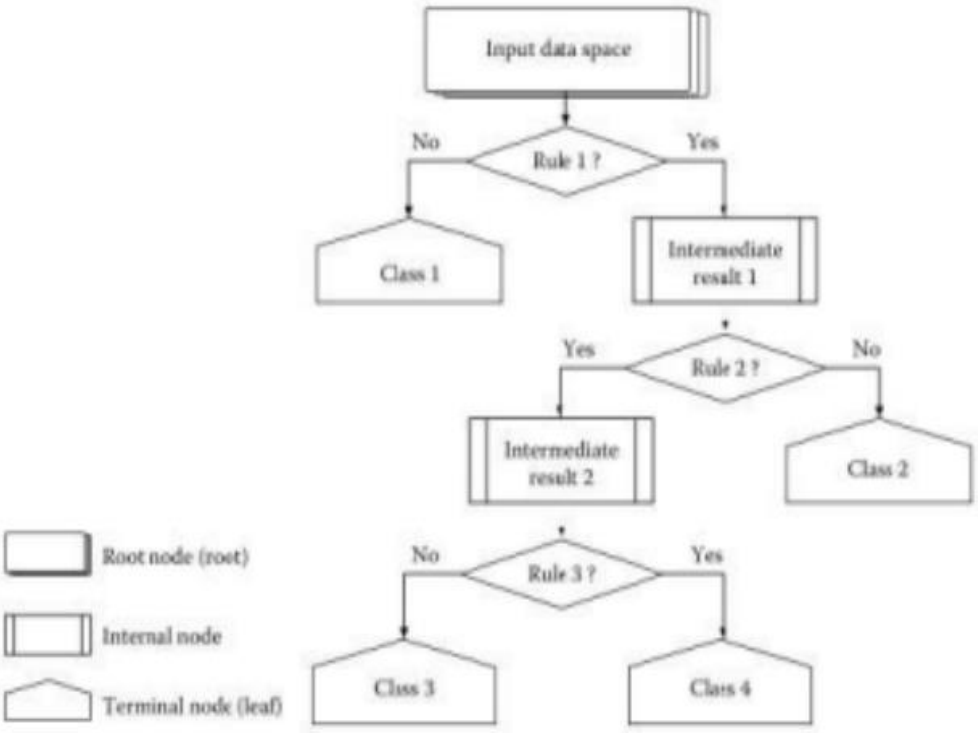


MODELS

There are four different kinds of models present in our project to predict the disease these are

- Decision tree
- Random forest tree
- Gaussian Naïve Bayes
- KNN

The decision tree is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification of the image. It is used for the classification of very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes, and leaf. Roots consist of the attribute which has the most effect on the outcome, leaf tests for the value of a certain attribute, and leaf give out the output of the tree.



Random Forest Algorithm is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps –

1. It chooses random data samples from the dataset.
2. It constructs decision trees for every sample dataset chosen.
3. At this step every predicted result will be compiled and voted on.
4. At last most voted predictions will be selected and presented as a result of classification.

In this project, we have used random forest classifier with 100 random samples, and the result given is ~95% accuracy.

K Nearest Neighbour is a supervised learning algorithm. It is a basic yet essential algorithm. It finds extensive use in pattern finding and data mining. It works by finding a pattern in data that links data to results and it improves upon the pattern recognition with every iteration. We have used K Nearest Neighbour to classify our dataset and achieved ~92% accuracy.

Naïve Bayes algorithm is a family of algorithms based on naïve bayes theorem. They share a common principle that every pair of predictions is independent of each other. It also makes an assumption that features make an independent and equal contribution to the prediction. In our project, we have used the naïve Bayes algorithm to gain a ~95% accurate prediction.

LIBRARIES

In this project standard libraries for database analysis and model creation are used. The following are the libraries used in this project:

1. Tkinter: It's a standard GUI library of python. Python when combined with Tkinter provides a fast and easy way to create GUI. It provides a powerful object-oriented tool for creating GUI. It provides various widgets to create GUI some of the prominent ones being:

- Button
- Canvas
- Label
- Entry
- Check Button
- List box
- Message
- Text
- Messagebox

Some of these were used in this project to create our GUI namely messagebox, button, label, Option Menu, text and title. Using tkinter we were able to create an interactive GUI for our model.

2. Numpy: Numpy is the core library of scientific computing in python. It provides powerful tools to

deal with various multi-dimensional arrays in python. It is a general-purpose array processing package. Numpy's main purpose is to deal with the multidimensional homogeneous array. It has tools ranging from array creation to its handling. It makes it easier to create a n dimensional array just by using np.zeros() or handle its contents using various other methods such as replace, arrange, random, save, and load it also helps I array processing using methods like sum, mean, std, max, min, all, etc Array created with NumPy also behave differently than arrays created normally when they are operated upon using operators such as +,-,*,/. All the above qualities and services offered by NumPy array make it highly suitable for our purpose of handling data. Data manipulation occurring in arrays while performing various operations needs to give the desired results while predicting outputs requires such high operational capabilities.

3. Pandas: it is the most popular python library used for data analysis. It provides highly optimized performance with back-end source code purely written in C or python.

Data in python can be analyzed in 2 ways

- Series
- Dataframes

Series is a one-dimensional array defined in pandas used to store any data type. Dataframes are two-dimensional data structures used in python to store data consisting of rows and columns. Pandas data frame is used extensively in this project to use datasets required for training and testing the algorithms. Dataframes make it easier to work with attributes and results. Several of its inbuilt functions such as replace were used in our project for data manipulation and preprocessing.

4. Sklearn: Sklearn is an open-source python library with implements a huge range of machine-learning, pre-processing, cross-validation, and visualization algorithms. It features various simple and efficient tools for data mining and data processing. It features various classification, regression, and clustering algorithms such as support vector machine, random forest classifier, decision tree, gaussian naïve-Bayes, and KNN to name a few. In this project, we have used sklearn to get the advantage of inbuilt classification algorithms like decision trees, random forest classifiers, KNN, and naïve Bayes. We have also used inbuilt cross-validation and visualization features such as classification reports, confusion matrix, and accuracy score.

CHAPTER 7

IMPLEMENTATION

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 in the new system to 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 its constraints on implementation, design of methods to achieve the changeover, and an evaluation of change over methods apart 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 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 that will help automate the process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirementsare satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, wherein each module is tested to provide its correctness, validity and also determine any missing operations, and verify whether the objectives have been met. Errors are noted down and corrected immediately.

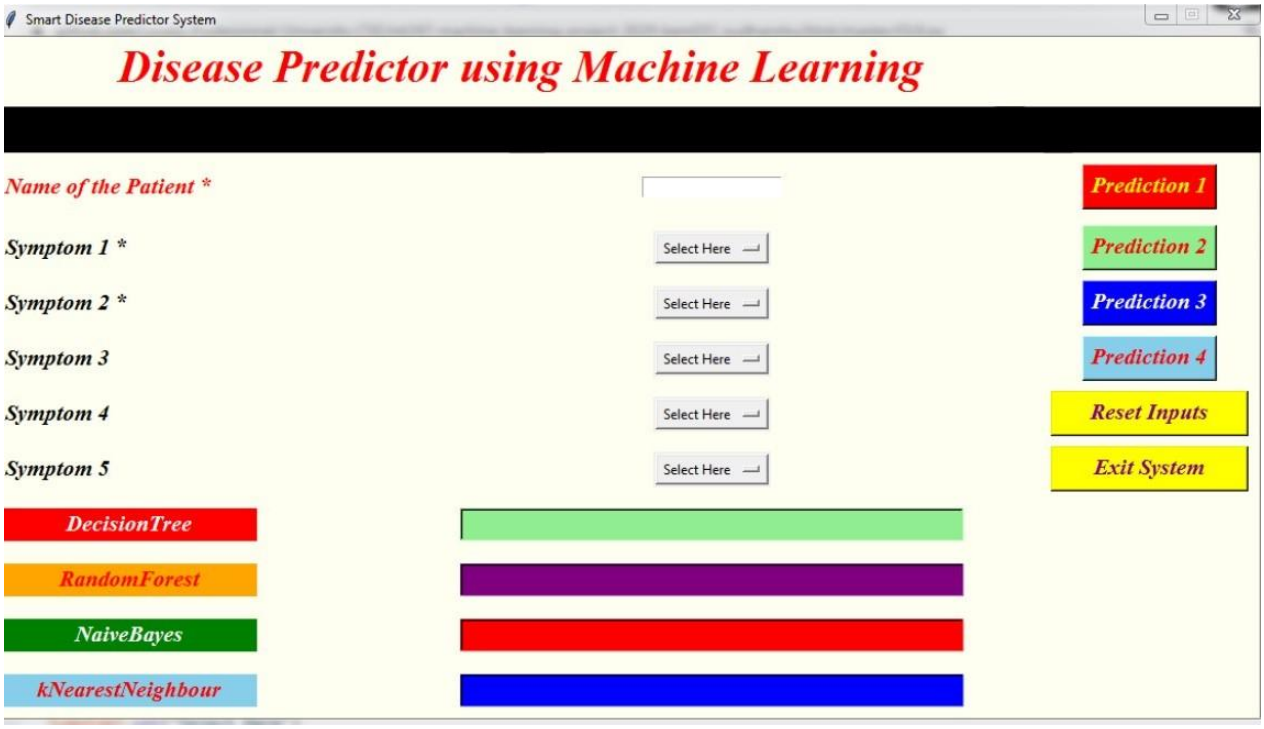
2. Unit testing is an important and major part of the project. So errors are rectified easily in particular modules and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted on individual modules.

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

8. SNAPSHOTS



GUI made for this project is a simple tkinter GUI consisting of labels, messagebox, button, text, title and option menu

```
#Importing Libraries
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from tkinter import *
import numpy as np
import pandas as pd
import os
```

Importing libraries

```
root = Tk()
pred1=StringVar()
def DecisionTree():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn import tree

        clf3 = tree.DecisionTreeClassifier()
        clf3 = clf3.fit(X,y)

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=clf3.predict(X_test)
        print("Decision Tree")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

        for k in range(0,len(l1)):
            for z in psymptoms:
                if(z==l1[k]):
                    l2[k]=1

        inputtest = [l2]
        predict = clf3.predict(inputtest)
        predicted=predict[0]

        h='no'
        for a in range(0,len(disease)):
            if(predicted == a):
                h='yes'
                break


if (h=='yes'):
    pred1.set(" ")
    pred1.set(disease[a])
else:
    pred1.set(" ")
    pred1.set("Not Found")
#Creating the database if not exists named as database.db and creating table if not exists named as DecisionTree using sqlite3
import sqlite3
conn = sqlite3.connect('database.db')
c = conn.cursor()
c.execute("CREATE TABLE IF NOT EXISTS DecisionTree(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringVar,Symtom4 TEXT,Symtom5 TEXT,Disease StringVar)")
c.execute("INSERT INTO DecisionTree(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(NameEn.get(),Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get(),pred1.get()))
conn.commit()
c.close()
conn.close()

#printing scatter plot of input symptoms
#printing scatter plot of disease predicted vs its symptoms
scatterinp(Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get())
scatterplt(pred1.get())
```

Decision Tree Algorithm

```
pred2=StringVar()
def randomforest():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.ensemble import RandomForestClassifier
        clf4 = RandomForestClassifier(n_estimators=100)
        clf4 = clf4.fit(X,np.ravel(y))

        # calculating accuracy
        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=clf4.predict(X_test)
        print("Random Forest")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

        for k in range(0,len(l1)):
            for z in psymptoms:
                if(z==l1[k]):
                    l2[k]=1

        inputtest = [l2]
        predict = clf4.predict(inputtest)
        predicted=predict[0]

        h='no'
        for a in range(0,len(disease)):
            if(predicted == a):
                h='yes'
                break
        if (h=='yes'):
            pred2.set(" ")
            pred2.set(disease[a])
        else:
            pred2.set(" ")
            pred2.set("Not Found")
#Creating the database if not exists named as database.db and creating table if not exists named as RandomForest using sqlite3
import sqlite3
conn = sqlite3.connect('database.db')
c = conn.cursor()
c.execute("CREATE TABLE IF NOT EXISTS RandomForest(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringVar,Symtom4 TEXT,Symtom5 TEXT,Disease StringVar)")
c.execute("INSERT INTO RandomForest(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(NameEn.get(),Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get(),pred2.get()))
conn.commit()
c.close()
conn.close()
#printing scatter plot of disease predicted vs its symptoms
scatterplt(pred2.get())
```

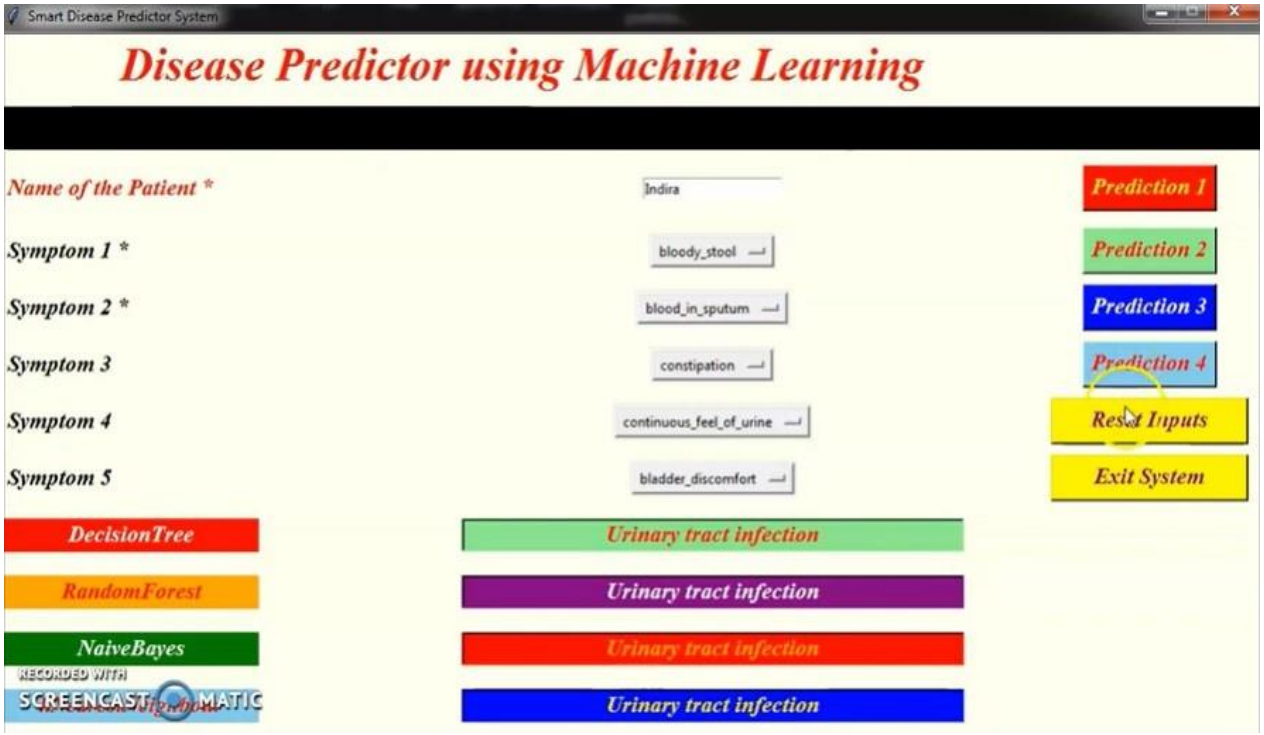
Random Forest Algorithm

```
pred4=StringVar()
def KNN():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.neighbors import KNeighborsClassifier
        knn=KNeighborsClassifier(n_neighbors=5,metric='minkowski',p=2)
        knn=knn.fit(X,np.ravel(y))

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=knn.predict(X_test)
        print("kNearest Neighbour")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

        for k in range(0,len(l1)):
            for z in psymptoms:
                if(z==l1[k]):
                    l2[k]=1

        inputtest = [l2]
        predict = knn.predict(inputtest)
        predicted=predict[0]

        h='no'
        for a in range(0,len(disease)):
            if(predicted == a):
                h='yes'
                break
if (h=='yes'):
    pred4.set(" ")
    pred4.set(disease[a])
else:
    pred4.set(" ")
    pred4.set("Not Found")
#Creating the database if not exists named as database.db and creating table if not exists named as KNearestNeighbour using sqlite3
import sqlite3
conn = sqlite3.connect('database.db')
c = conn.cursor()
c.execute("CREATE TABLE IF NOT EXISTS KNearestNeighbour(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringVar,Symtom4 TEXT,Symtom5 TEXT,Disease StringVar)")
c.execute("INSERT INTO KNearestNeighbour(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(NameEn.get(),Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get(),pred4.get()))
conn.commit()
c.close()
conn.close()
#printing scatter plot of disease predicted vs its symptoms

scatterplt(pred4.get())
```

K Nearest Neighbor Algorithm

```
pred3=StringVar()
def NaiveBayes():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.naive_bayes import GaussianNB
        gnb = GaussianNB()
        gnb=gnb.fit(X,np.ravel(y))

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=gnb.predict(X_test)
        print("Naive Bayes")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
        for k in range(0,len(l1)):
            for z in psymptoms:
                if(z==l1[k]):
                    l2[k]=1

        inputtest = [l2]
        predict = gnb.predict(inputtest)
        predicted=predict[0]

        h='no'
        for a in range(0,len(disease)):
            if(predicted == a):
                h='yes'
                break
        if (h=='yes'):
            pred3.set(" ")
            pred3.set(disease[a])
        else:
            pred3.set(" ")
            pred3.set("Not Found")
    #Creating the database if not exists named as database.db and creating table if not exists named as NaiveBayes using sqlite3
    import sqlite3
    conn = sqlite3.connect('database.db')
    c = conn.cursor()
    c.execute("CREATE TABLE IF NOT EXISTS NaiveBayes(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringVar,Symtom4 TEXT,Symtom5 TEXT,Disease StringVar)")
    c.execute("INSERT INTO NaiveBayes(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(NameEn.get(),Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get(),pred3.get()))
    conn.commit()
    c.close()
    conn.close()
    #printing scatter plot of disease predicted vs its symptoms
    scatterplt(pred3.get())
```

Naive Bayes Algorithm


Smart Disease Predictor System
Disease Predictor using Machine Learning
Name of the Patient *
Indira
Symptom 1 *
bloody_stool
Symptom 2 *
blood_in_sputum
Symptom 3
constipation
Symptom 4
continuous_feel_of_urine
Symptom 5
bladder_discomfort
Prediction 1
Prediction 2
Prediction 3
Prediction 4
Reset Inputs
Exit System
DecisionTree
Urinary tract infection
RandomForest
Urinary tract infection
NaiveBayes
Urinary tract infection
Urinary tract infection
RECORDED WITH
SCREENCAST-O-MATIC

Prediction

CHAPTER 9

CONCLUSION

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 authorized users depending on their permissions.
- It effectively overcomes the delay in communications.
- Updating information becomes so easier
- System security, data security, and reliability are the striking features.
- The System has adequate scope for modification in the future if it is necessary.

10. REFERENCE

[1] Hamet P., Tremblay J. Artificial intelligence in medicine. Metabolism. 2017; 69:S36–S40. doi: 10.1016/j.metabol.2017.01.011. [PubMed] [CrossRef] [Google Scholar].

[2] Johnson K.W., Soto J.T., Glicksberg B.S., Shameer K., Miotto R., Ali M., Dudley J.T. Artificial intelligence in cardiology. J. Am. Coll. Cardiol. 2018; 71:2668–2679. doi: 10.1016/j.jacc.2018.03.521. [PubMed] [CrossRef] [Google Scholar].

[3] Bini S. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? J. Arthroplast. 2018; 33:2358–2361. doi: 10.1016/j.arth.2018.02.067. [PubMed] [CrossRef] [Google Scholar].

[4] Kotsiantis S.B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques. Emerg. Artif. Intell. Appl. Comput. Eng. 2007; 160:3–24. [Google Scholar].

[5] Deo R.C. Machine Learning in Medicine. Circulation. 2015; 132:1920–1930. doi: 10.1161/CIRCULATIONAHA.115.001593. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[6] Battineni G., Sagaro G.G., Nalini C., Amenta F., Tayebati S.K. Comparative Machine-Learning Approach: A Follow-Up Study on Type 2 Diabetes Predictions by Cross-Validation Methods. Machines. 2019; 7:74. doi: 10.3390/machines7040074. [CrossRef] [Google Scholar].

[7] Polat H., Mehr H.D., Cetin A. Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. J. Med. Syst. 2017; 41:55. doi: 10.1007/s10916-017-0703-x. [PubMed] [CrossRef] [Google Scholar]