

Design and Development of Smart Health Prediction Using Data Mining

BY

MD ASIQUL ISLAM

ID: 152-15-6065

This Report Presented in Partial Fulfillment of the Requirements for the Degree
of Bachelor of Science in Computer Science and Engineering.

Supervised By

Dr. Sheak Rashed Haider Noori

Associate Professor and Associate Head

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

October 2020

APPROVAL

This Project/internship titled “**Design and Development of Smart Health Prediction Using Data Mining**”, submitted by **Md Asiqul Islam**, ID No: **152-15-6065** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **8 October, 2020**.

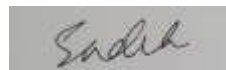
BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Md. Sadekur Rahman
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Subhenur Latif
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Saddam Hossain
Assistant Professor

Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori, Associate Professor and Associate Head**, Department of CSE, **Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr. Sheak Rashed Haider Noori

Associate Professor and Associate Head

Department of CSE

Daffodil International University

Submitted by:



Md Asiqul Islam

ID: 152-15-6065

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

At First I would like to express my heartiest thanks and gratefulness to almighty God for His divine blessing make me possible to complete the final year project successfully.

I really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori**, Associate Professor and Associate Head of CSE, Faculty of Science and Information Technology, Daffodil International University, Dhaka. Their Deep Knowledge & keen interest with supportive instructions helped us in the field of deep learning based research; finally I completed my work on “**Design and Development of Smart Health Prediction Using Data Mining**”. Their endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain**, Professor and Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish my project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again I want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This project is a result of hard work and all those inspirations and assistance.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

This project can be used for the data mining techniques such as medical field, research field, and educational field and various aspects. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. As per the modern technology huge improvement has been made in computer field and therefore there is no need to deal with such a large amount of data at a same time. A major objective of this paper is to evaluate data mining techniques in clinical and health care applications to develop an accurate decision. It is a famous and powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different databases to transform it into new researches and results. It makes use of Artificial Intelligence, machine learning and database management techniques to extract new patterns from large data sets and the knowledge associated with these patterns. By using this technique data can be extracted automatically or semi automatically. The different parameters included in data mining include clustering, forecasting, path analysis and predictive analysis.

CONTENTS	PAGE
Board of Examiners	i
Declaration	ii
Acknowledgment	iii
Abstract	iv

CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	1-5
1.1 Introduction	1
1.2 Motivation	2
1.3 Rational of the Study	2
1.4 Advantages and Restriction	3
1.4.1 Advantages	3
1.4.2 Restriction	4
1.5 Expected Output	4
1.6 Report Layout	4
CHAPTER 2: BACKGROUND	6-14
2.1 Introduction	6
2.2 Related Works	6
2.3 Research Summary	13
2.4 Scope of the problem	13
2.5 Challenges	14
CHAPTER 3: Research Methodology	15-21
3.1 Introduction	15
3.2 Research Subject and Instrumentation	15
3.3 Materials and Methods	15
3.3.1 Classification	15
3.3.2 Gaussian Naïve Byes	17
3.3.3 Decision Tree	18
3.3.4 Support Vector Machine	19

3.4 Data Collection Procedure	20
3.4.1 Data Collection	20
3.4.2 Data Transformation	20
3.4.3 Data Pre-Processing	20
3.4.4 Feature Extraction	20
3.4.4.1 Data Mining	20
3.5 Implementations Requirements	21
CHAPTER 4: Architectural Design	22-27
4.1 Planning	22
4.2 Flow of Working Procedure	24
4.3 Database	25
4.4 Implementation	27
CHAPTER 5: Experimental Result and Discussion	28-30
5.1 Introduction	28
5.2 Experimental Results	28
5.3 Descriptive Analysis	29
5.4 Testing	30
5.5 Summary	30
CHAPTER 6: Summary, Conclusion, Recommendation, Implication	31-32
6.1 Summery of the Study	31
6.2 Conclusion	31
6.3 Recommendation	31
6.4 Implication	32
Appendix	33-33
Appendix A: Project Reflection	33
Appendix B: Related Issues	33
REFERENCES	34-35

LIST OF FIGURES

FIGURES	PAGE
Figure 3.3.1: Classification	16
Figure 3.3.2: Gaussian Naïve Byes formula	17
Figure 3.3.3: Decision Tree	18
Figure 3.3.4: Support Vector Machine	19
Figure 4.1: Use case diagram	23
Figure 4.2: Data Flow Diagram 1	24
Figure 4.3: Entity Relationship Model	26
Figure 4.4: Class Diagram	27

LIST OF TABLE

TABLES	PAGE
Table 1- Predictive Accuracy of Bayes and other Technique	7
Table 2- Sample Data Sets	28
Table 3- Sample Output	29

CHAPTER 1

Introduction

1.1 Introduction

As the utilization of science and innovation, health prediction is a workmanship to anticipate malady changes and conduct in the subject of time for human. Health prediction is one of the most explored issues everywhere on the world in the past quite a few years deductively and innovatively. Prediction is one of the fundamental uses of data mining. Data Mining is utilized to investigate and break down enormous measure of data, produce designs, assemble significant data.

Health prediction achieves ailment how the circumstance will change or change the Symptoms of a human body. In this time of hardware health issues are watched and recorded by a medical clinic or a specialist. At present, when one encounters explicit contamination, by then the individual needs to visit to pro which is dreary and costly too. Also if the customer is a long way from expert and clinical centers it may be hard for the customer as the disease can't be perceived. Along these lines, if the above cycle can be finished utilizing a mechanized program which can spare time just as cash, it could be simpler to the patient which can make the cycle simpler. Anticipating health utilizing a PC is known as algorithmic health prediction.

Health prediction is an unpredictable and long cycle for general individuals as well as for researcher and specialists. The right symptoms of infections are fundamental for this sort of utilizations. Exact logical prediction is horrible employment due to the bizarre propensity for the malady. Researchers utilize different strategies for understanding the symptoms of an infection.

Smart Health Prediction is an electronic application that predicts the disease of the customer concerning the signs given by the customer. Smart Health Prediction framework has data sets gathered from various health related destinations. With the

assistance of Smart Health Prediction the client will have the option to know the likelihood of the malady with the given symptoms.

1.2 Motivation

On the off chance that we see cautiously over the most recent couple of a long time there are so numerous new maladies are appearance and human are tainted by them. Some of them are slaughtered endless individuals. In any case, each malady had some extraordinary sort of symptoms. It might have occurred so frequently that you or someone yours need authorities help rapidly, yet they are not available on account of some clarification. The Health Prediction framework is an end customer maintain and internet meeting adventure. Here we propose a framework that licenses customers to get second course on their health issues through a shrewd health care framework on the web. The framework is dealt with various signs and the contamination or malady related with those frameworks. The framework grants customer to share their results and issues. It by then estimates customer's results to check for various ailments that could be connected with it. Here we use some vigilant data mining techniques to calculate the most definite malady that could be connected with patient's appearances. Later on, we need to select ourselves with enormous data investigation. Along these lines, we need to make something that encourages us to prevail throughout everyday life.

Data mining is where a tremendous measure of data is gathered and being handled to compost some valuable data for example data. Everybody needs the synopsis and suitable data which is conceivable by it, as it's anything but a simple should however it turns out conceivable by a progression of techniques and science.

1.3 Rational of the study

The purpose of data burrowing is for finding structure through unstructured data. Data mining pulls out the significance of tumultuous data and its found models are clearly with discretionary data. It moreover uses from very all the more probable get designs,

models, associations lastly anticipates the customer direct and symptoms. Along these lines, we comprehend the figuring to improve the prediction.

We present the systems and uses of information mining in therapeutic and enlightening pieces of the Clinical Predictions. In restorative and human services handle, an immense amount of data is transforming into open because of accessibility of PCs. Such a larger than usual measure of data can't be prepared to make wellbeing expectations in the beginning period and make treatment calendars to analyze. Our point is to evaluate the systems of information preparing in the fields of clinical and social insurance to create right decisions. It likewise offers a nearby trade of therapeutic data dealing with methodologies which may improve different pieces of Clinical Predictions. It's a latest astonishing development that is of high eagerness for the PC world. It uses already existing data in a couple of databases to redo it into new researches and results. From huge educational assortments, to isolate new models and the data related to these models data mining uses AI and database the board. Particularly the task is to get data by the techniques for customized or self-loader. The various boundaries encased in data getting ready fuse batching, assessing, way examination and insightful assessment.

1.4 Advantages and Restrictions

1.4.1 Advantages

This undertaking intends to give a web stage to anticipate the events of ailment based on different symptoms. The client can choose different symptoms and can discover the ailments with their probabilistic figures.

The advantages we have found:

- Client can look for specialist's assistance anytime of time.
- Client can discuss their ailment and get moment finding.
- Doctors get more clients online.

1.4.2 Restrictions

The limitations of this project are given below:

- Disease Predictor doesn't recommend meds of the infection.
- Past history of the disease has not been thought of.
- The project isn't completely robotized. It needs specialists for full analysis.

1.5 Expected Output

- We will get information about the specialized doctor for the exact disease we predict.
- We can easily predict the disease by using this system with the help of symptoms.
- A patient can easily talk about their illness and get instant solution.
- Getting huge amount of data about disease and their symptoms.

1.6 Report Layout

Chapter 1: This section is about the Introduction of our investigation.

In this Chapter we will examine about Introduction, Motivation, Rationale of the Study, Advantages and Restrictions, Expected Output, Report Layout are portrayed.

Chapter 2: This section is about the Background of the assessment.

In this Chapter we will examine about Related Works, Research Summary, Scope of the Problem and Challenges are depicted.

Chapter 3: This section is about the Research Methodology of the assessment.

In this Chapter we will examine about Research Subject and Instrumentation, materials and strategy, Data Collection Procedure, Implementation Requirements are portrayed.

Chapter 4: This section is about Architectural Design of the assessment.

In this Chapter we will examine about Planning, Flow of Working Procedure, Database and Implementation are investigated.

Chapter 5: This section is about Experimental Results and Discussion of the assessment.

In this Chapter we will examine about Experimental Results, Descriptive Analysis, Testing and Summary are portrayed.

Chapter 6: In this part Summary, Conclusion, Recommendation, and Implication for Future Study is portrayed.

CHAPTER 2

Background

2.1 Introduction

There is no uncertainty that health prediction is significant and important in our life. There are endless explores and examination had done about health prediction. By doing heaps of study, we have realized that numerous metrological office are attempting to improve the exactness of health prediction however none of them are hundred percent precise. Thus, we have concluded ourselves to work in this field.

2.2 Related Works

K.M. Al-Aidaros, A.A. Bakar and Z. Othman “has coordinated the assessment for the best clinical discovering mining strategy. For this makers differentiated Naïve Baeyes and five distinct classifiers for instance Logistic Regression (LR), KStar (K*), Decision Tree (DT), Neural Network (NN) and a basic standard based calculation (ZeroR). For this, 15 certifiable clinical issues from the UCI AI store (Asuncion and Newman, 2007) were picked for surveying the presentation things being what they are. In the assessment it was found that NB beats various figuring’s in 8 out of 15 educational lists so it was assumed that the perceptive precision brings about Naïve Baeyes is better than various systems” [1].

Table 1- Predictive Accuracy of Bayes and other Technique

Medical Problems	NB	LR	K*	DT	NN	ZeroR
Breast Cancer wise	97.3	92.98	95.72	94.57	95.57	65.52
Breast Cancer	72.7	67.77	73.73	74.28	66.95	70.3
Dermatology	97.43	96.89	94.51	94.1	96.45	30.6
Echocardiogram	95.77	94.59	89.38	96.41	93.64	67.86
Liver Disorders	54.89	68.72	66.82	65.84	68.73	57.98
Pima Diabetes	75.75	77.41	70.19	74.49	74.75	65.11
Haeberman	75.36	74.47	73.73	72.16	70.32	73.53
Heart-c	83.34	83.7	75.18	77.13	80.99	54.45
Heart Statlog	84.85	84.04	73.89	75.59	81.78	55.56
Heart-b	83.95	84.23	77.83	80.22	80.07	63.95
Hepatitis	83.81	83.89	80.17	79.22	80.78	79.38
Lung Cancer	53.25	47.25	41.67	40.83	44.08	40
Lymphpgraphy	84.97	78.45	83.18	78.21	81.81	54.76
Postooerative Patient	68.11	61.11	61.67	69.78	58.54	71.11
Primary tumor	49.71	41.62	38.02	49.39	40.38	24.78
Wins	8\15	5\15	0\15	2\15	1\15	1\15

Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, Nicholas Christakis and Albert Laszlo Barabasi “have discovered that worldwide therapy of constant malady is neither time or cost productive. So the creators directed this exploration to foresee future illness hazard. That’s why the Consideration was utilized (which depends just on a patient's clinical history utilizing ICD-9-CM codes so as to foresee future ailments chances). CARE joins communitarian sifting techniques with grouping to foresee every patient's most prominent ailment hazards dependent on their own clinical history and that of comparative patients. Creators have likewise depicted an Iterative adaptation, ICARE, which fuses group ideas for improved execution. These epic frameworks require no specific data and give expectations to ailments of numerous sorts in a solitary run. The noteworthy future sickness inclusion of ICARE speaks to more precise early admonitions for a large number of maladies, some even a very long time ahead of itime [2]. Applied to maximum capacity, the Consideration system can be utilized investigate a more extensive ailment chronicles, propose already unconsidered concerns, and encouraging conversation about early testing and avoidance” [3].

JyotiSoni, Ujma Ansari, Dipesh Sharma and SunitaSoni “have done this research research paper into provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in the present clinical examination especially in Heart Disease Prediction. Number of analysis has been directed to look at the presentation of prescient data mining strategy on the equivalent dataset and the result uncovers that Decision Tree beats and at some point Bayesian grouping is having comparable precision as of choice tree yet other prescient techniques like KNN, Neural Networks, Classification dependent on bunching isn't performing great” [4].

Shadab Adam Pattekari and AsmaParveen, “has directed an exploration utilizing Naïve Bayes Calculation to anticipate the heart maladies where client furnishes the information which is contrasted and prepared arrangement of qualities. So from this exploration, patients had the option to give their fundamental data which is contrasted and the information and the coronary illness is anticipated” [5].

M.A.NisharaBanu, B Gomathy, “utilized clinical information mining procedures like affiliation rule mining, characterization, grouping I to separate the different sorts of heart based issues. Choice tree is made to show each possible aftereffect of a choice. Different guidelines are made to get the best outcome. In this investigation age, sex, smoking, overweight, liquor admission, glucose, hear rate, pulse are the limits used for making the decisions. Risk level for different limits are taken care of with their id's expanding (1-8). ID lesser than of 1 of weight contains the normal level of desire and higher ID other than 1 incorporate the higher peril levels .K-implies gathering technique is used to inspect the model in the dataset. The figuring bundles information's into k get-togethers. Each point in the dataset is given out to the shut gathering. Each gathering place is recomputed as the typical of the concentrations in that bundle” [6].

Min Chen, “proposes AI calculations for viable prediction of chronic disease flare-up in disease-successive networks. Experiment the adjusted prediction models over genuine medical clinic data gathered from focal China in 2013-2015. To defeat the difficulty of fragmented data uses a dormant factor model to reconstruct the missing data on a local ongoing disease of cerebral dead tissue. Also, various locales show unique characteristics of certain provincial diseases, which may weaken the prediction of disease outbreaks. Proposes another convolution neural network based multimodal disease hazard prediction (CNN-MDRP) calculation utilizing organized and unstructured data from emergency clinic. Here focused on the two sorts of organized and unstructured data in the territory of clinical large data analytics. Compared to a few regular prediction calculations, the prediction accuracy of our proposed calculation arrives at 94.8% with a convergence speed which is quicker than that of the CNN- based unmoral disease hazard prediction (CNN-UDRP) calculation” [7].

Feixiang Huang, “applies the data mining cycle to predict hypertension from tolerant clinical records with eight other diseases. An example with the size of 9862 cases has been studied. The test was extricated from a genuine Health care Information System database containing 309383 clinical records. Here, saw that the dispersion of patient diseases in the medical database is imbalanced. Under-testing procedure has been applied to create preparing data sets, and data mining Weka instrument has been

utilized to produce the Naive Bayesian and J-48 classifiers. Furthermore, a gathering of J-48 classifiers was made attempting to improve the prediction execution, and rough set instruments were utilized to lessen the troupe dependent on the thought of second-request estimation. Trial results indicated a little improvement of the gathering approach over unadulterated Naïve Bayesian and J-48 in exactness, affectability and F-measure” [8].

Sujatha R, “enhances mining in health sector assumes a significant function for revealing new patterns which thus accommodating for all the gatherings related with this field. It is another incredible innovation which is of high enthusiasm for PC world. It is a sub field of software engineering that utilizes previously existing data in various databases to change it into new investigates and results. It utilizes Artificial Intelligence, Machine Learning and database the executives to separate new examples from enormous data sets and the information related with these examples. The genuine assignment is to extricate data via programmed or self-loader implies. The various boundaries remembered for data mining incorporate bunching, gauging, way investigation and prescient examination” [9].

Pinky Saikia Dutta, “depicts about data mining in disease prediction is a sub field of software engineering that utilizes previously existing data in various databases to change it into new explores and results. In this data rich world, individuals are running out of data. This can involve hazard for the individual who needs quick cures with respect to their chronic frailty. To unfurl this obstacle, the idea of data mining is the most appropriate. Here, the customary methodologies have been supplanted by smart advancements . The principle reason for data mining application in healthcare framework is to build up a robotized apparatus for recognizing and scattering significant healthcare data. In this framework, we have introduced an online application for Predicting diseases dependent on client input symptoms. It predicts plausible diseases by mining data sets and gives medicinal answers for viable treatment” [10] [11] .

Ravi Aavula, “propose e-health applications in light of the fact that counseling a specialist is a very clear thing in our everyday life, except the accessibility of the specialist during the hour of our prerequisite is flighty. So as to defeat this issue a

proposition of android application is made, this smart health application empowers clients to get moment report on their health issues through a shrewd health care application on the web. This e-health application empowers client to communicate their symptoms and issues. It at that point measures user's issues and symptoms to check for different health gives that could be related with the symptoms given by the client. In the event that the application can't give a specific arrangement, at that point it encourages the client to under-go tests like blood test, CITI filter as needs be" [12].

Priyanka Vijay Pawar, "depicts about the significance of smart advancements over customary methodologies has been coordinated in each order of science. Here, propounds a philosophy to foresee diseases dependent on client input symptoms. They have fabricated a model to show the proficiency of these techniques which will advise clients about the disease they are experiencing. It predicts likely diseases by mining data sets and gives proposed specialists and therapeutic arrangements. It will likewise manage the clients by offering tips to carry on with a healthy life, some eating regimen tips and furthermore handiness of plants and food things. The creators recognized likelihood of diseases utilizing apriori method" [13].

Krishnaiah V, "shows a review of the utilizations of data mining strategies, clinical, research, and instructive parts of clinical predictions. In clinical and health care zones, because of guidelines and because of the accessibility of PCs, a lot of data is opening up. From one perspective, professionals are required to utilize this data in their work at the same time, simultaneously; such a lot of data can't be handled by people in a brief timeframe to make finding, visualization and therapy plans. A significant goal of this paper is to assess data mining methods in clinical and health care applications to create precise choices. The paper likewise gives an itemized conversation of clinical data mining strategies can improve different parts of Clinical Predictions" [14].

Rawan Ali, "makes an online discussion venture for health prediction framework. It permits clients to get moment direction on their health issues through a canny health care framework. Through online contains data of different symptoms and the

disease/sickness related with those symptoms. It additionally has a possibility for clients of sharing their symptoms and issues. Data mining strategy is utilized for prediction for quiet disease and patient can take the medication legitimately from the drug store if specialist can't make prediction. With the use of big data in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. To overcome the difficulty of incomplete data, uses a latent factor model to reconstruct the missing data. A convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm” [15].

Md. Tahmid Rahman Laskar, “proposes ADPS due to rapid proliferation of internet technology and handheld devices has opened up new avenues for online healthcare system. There are instances where online medical help or healthcare advice is easier or faster to grasp than real world help. People often feel reluctant to go to hospital or physician on minor symptoms. However, in many cases, these minor symptoms may trigger major health hazards. As online health advice is easily reachable, it can be a great head start for users. Moreover, existing online health care systems suffer from lack of reliability and accuracy. Herein, we propose an automated disease prediction system (ADPS) that relies on guided user input. The system takes input from the user and provides a list of probable diseases. The accuracy of ADPS has been evaluated extensively. It ensured an average of 14.35% higher accuracy in comparison with the existing solution” [16].

2.3 Research Summary

From research paper, books, different article we got lots of information about health prediction related works. There are completed a tiny work about this topic in Bangladesh. In almost research paper, they predict the behavior of disease. This is not suitable for large and complex data set and model. We collect a bulk of historical data from different websites, hospitals and doctors. Then we create data model for various disease from dataset and apply different algorithm to get desired output.

2.4 Scope of the Problem

In this paper a health prediction model is compiled for predicting the disease successfully. Our model of the health data with consist of time series is very careful and those data are transformed with information where the transformation are done using useful data mining technique. The data mining technique are used to uncover the invisible pattern of data and create a link among the different attributes with the health condition.

It is true that there are many automated weather prediction system with attractive and smart looking graphical user interface in market. Lots of websites, android and IOS apps are available. People are cheated by thinking that all of them are very accurate. But it is not true. We find many of them are fraud and give unrealistic results. In order to reduce this hassle our system will provide appropriate level of prediction. This system will help specially those people who stay irregular hour in their place of employment. Lastly, bad prediction is very harmful for human daily and makes the necessary decision. So, we take this opportunity to work in this platform.

2.5 Challenges

During working on this research we face enormous problem which we had deal very carefully. Some of those major challenges are mentioned bellow-

- Finding lots of data approximately 20 years historical data.
- Selecting a good API to get real time data and updating the daily data.
- Making the data set suitable for our system and correcting the huge amount of errors.
- Apply the relative algorithm.
- Choosing the best algorithm and improve the internal workflow of it.

CHAPTER 3

Research Methodology

3.1 Introduction

The primary work of our research is collecting data. We must know about the history of disease. Modern methods are not nearly being so accurate without numerical prediction, which uses mathematical formula, equations for health prediction. Such prediction requires most powerful super computer and real time observational data. It is not a single person job. So, we need a company which consists of much small station to collect data locally and store all of them in a prime data center.

3.2 Research Subject and Instrumentation

We are working on health prediction using data mining technique. First of all, we collect the historical data from different websites. The data was appeared in various form such as csv, excel, JSON. We collected all them together and merge in a structural database management system to work easily. We also done adding or deleting row, editing cell, correcting data types, avoiding unrealistic value or attribute to get a clean data set. After doing this, we choose a portion of dataset according to desired date for making model. After that we split the model into two parts i.e. Train model and target model and then applying different classification algorithm.

After getting output we compare the predicted value with actual situation or data. We calculated the accuracy for choosing the best algorithm suitable for our goal.

3.3 Materials and Methods

3.3.1 Classification:

Classification is related to data mining, the technique of discovering a pattern, model and similar object that details and discriminates data categories, classes, ranges, groups and concepts. Classification is a problem of determining the category of a set of observation.

Example: if I am feeling cold and headache then in this circumstances, a classification is needed to predict the class labels such as fever Yes or No. It is done in two steps:

- Learning Step: Building of classification model. There are many types of algorithm that are used to create a classifier using the model generated by a certain dataset. This model will be trained to perform accurate prediction.
- Classification Step: Now the trained model is used to classify a test data. We enter a test data of certain category and get a class as our expected output.

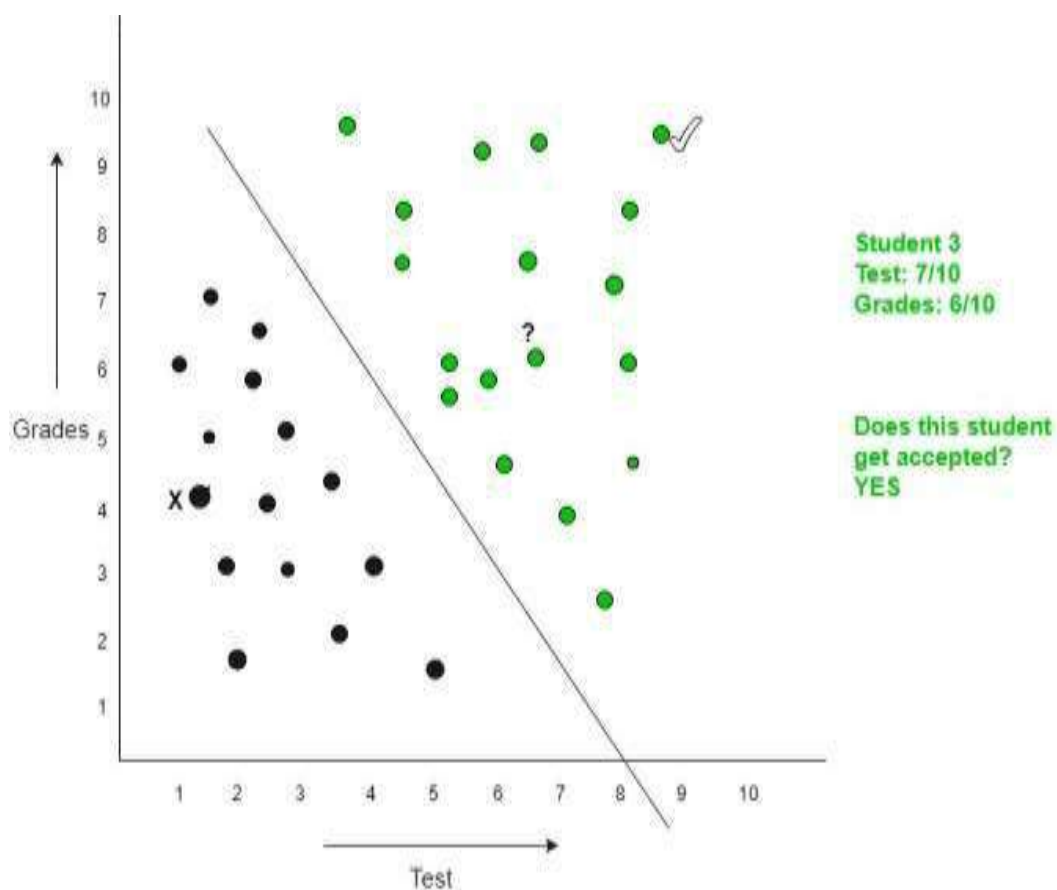


Figure 3.3.1: Classification

3.3.2 Gaussian Naïve Byes:

Bayes Theorem: Bayes' Theorem is created by Thomas Bayes portrays the likelihood of an occasion dependent on gained information on conditions.

$$P(A|B) = \frac{P(B|A)(P(A))}{P(B)}$$

Figure 3.3.2: Gaussian Naïve Byes formula

Here we clarify the condition below-

- A and B are two occasions.
- P (A) and P (B) are the conceivable outcomes of occasion which is autonomous from one another.
- P (B|A) is the chance of B under the state of A and the other way around.

Naïve Bayes has two classifier for example Bernoulli innocent Bayes and Gaussian credulous Bayes. We utilize Gaussian credulous Bayes since it is generally well known and straightforward.

3.3.3 Decision Tree:

Decision Tree, one of the most mainstream and incredible asset is utilized for arrangement and prediction. It is a chart like tree which assists with settling on decision. In decision tree each inner hub demonstrates a test on trait, each branch means a result of test and each leaf hub holds a class name.

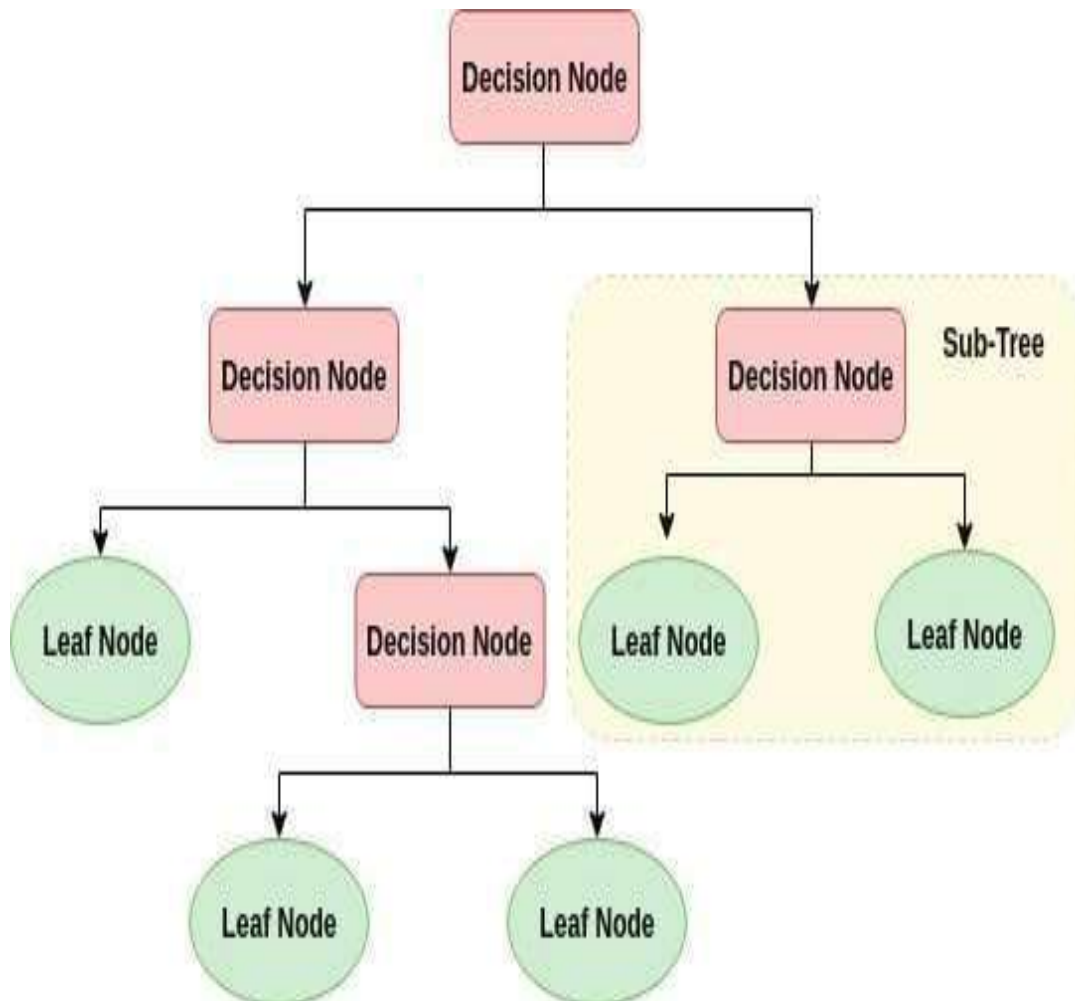


Figure 3.3.3: Decision Tree

3.3.4 Support Vector Machine:

Support Vector Machine is set up on the idea of decision sheet that characterize the limits. A decision sheet or line are utilized which isolates a lot of item as indicated by their diverse class participation. A graphical portrayal is given beneath-

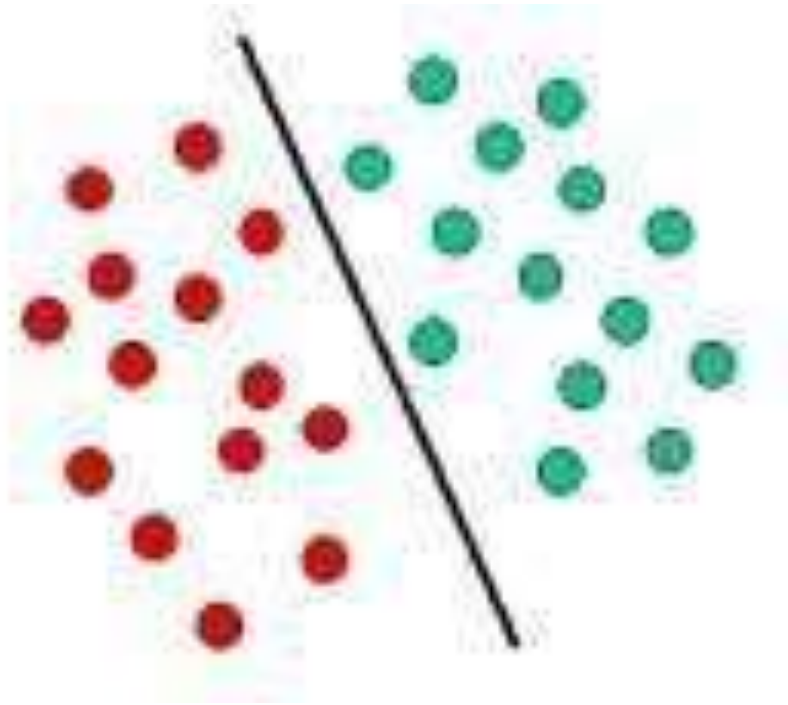


Figure 3.3.4: Support Vector Machine

In the above picture there are a lot of red and green and balls. We put a sheet in the center. We sort the balls individually by keeping the red balls in left half of the sheet and the green balls in right half of sheet. In this way, we said there are two class Green and Red. On the off chance that another ball like blue shaded is put in left half of the sheet then it is named Red or in the event that it is set in right half of the sheet, at that point it is named Green.

3.4 Data Collection Procedure

We had followed a few stages to gather data which is utilized in data mining strategies.

3.4.1 Data Collection:

Most importantly it is exceptionally hard to locate an appropriate dataset that fulfill us to arrive at our objective. We found a little asset of authentic data structure so numerous site, clinics and specialists.

3.4.2 Data Transformation:

The data was in endless arrangements. We read the each phone of data line savvy automatically and afterward changed over it to Sqlite3 design. We characterize a capacity without anyone else for carrying out this responsibility.

3.4.3 Data Pre-processing:

Data pre-preparing is engaged with various undertaking like eliminating invalid cell, including new cell, changing data type, rectifying invalid or ridiculous worth.

3.4.4 Feature Extraction:

Rest of the imperative was utilized for extra result in our application. They were fundamentally unrelated and no excess was found between them.

3.4.4.1 Data Mining:

Here we are examining the returns dataset by utilizing GNB, SVM and Decision Tree. At that point we locate the better anticipated outcome contrasting and the genuine data. We compute the exactness for knowing which calculation works better over the time. We characterize a few credits from the dataset that are anticipated by calculation.

3.5 Implementations Requirements

- HTML is used to display content in the browser.
- CSS is used to properly align the HTML content.
- Bootstrap framework is used for design the application.
- PHP is used for developing the application.
- Javascript Also use for developing the application.
- Creately is used for constructing figures.

CHAPTER 4

Architectural Design

4.1 Planning

Use Case Diagram: In Unified Modeling Language (UML) use case diagram primer structure of framework properties, modules and prerequisites for another framework plan which is immature. Use case center around the normal conduct of a framework and nothing about the genuine technique for manufacture. A striking highlights of use case diagram is that it encourages us to plan a framework from end user's point of view. Use case diagram is extremely basic. It shows a framework initially without enumerating anything.

It is useful for visiting the framework conduct subject to user term. It just sums up the inward engineering of a framework. In prior, it is said that a use case diagram is the blue-print of a framework. Here is our blue print:

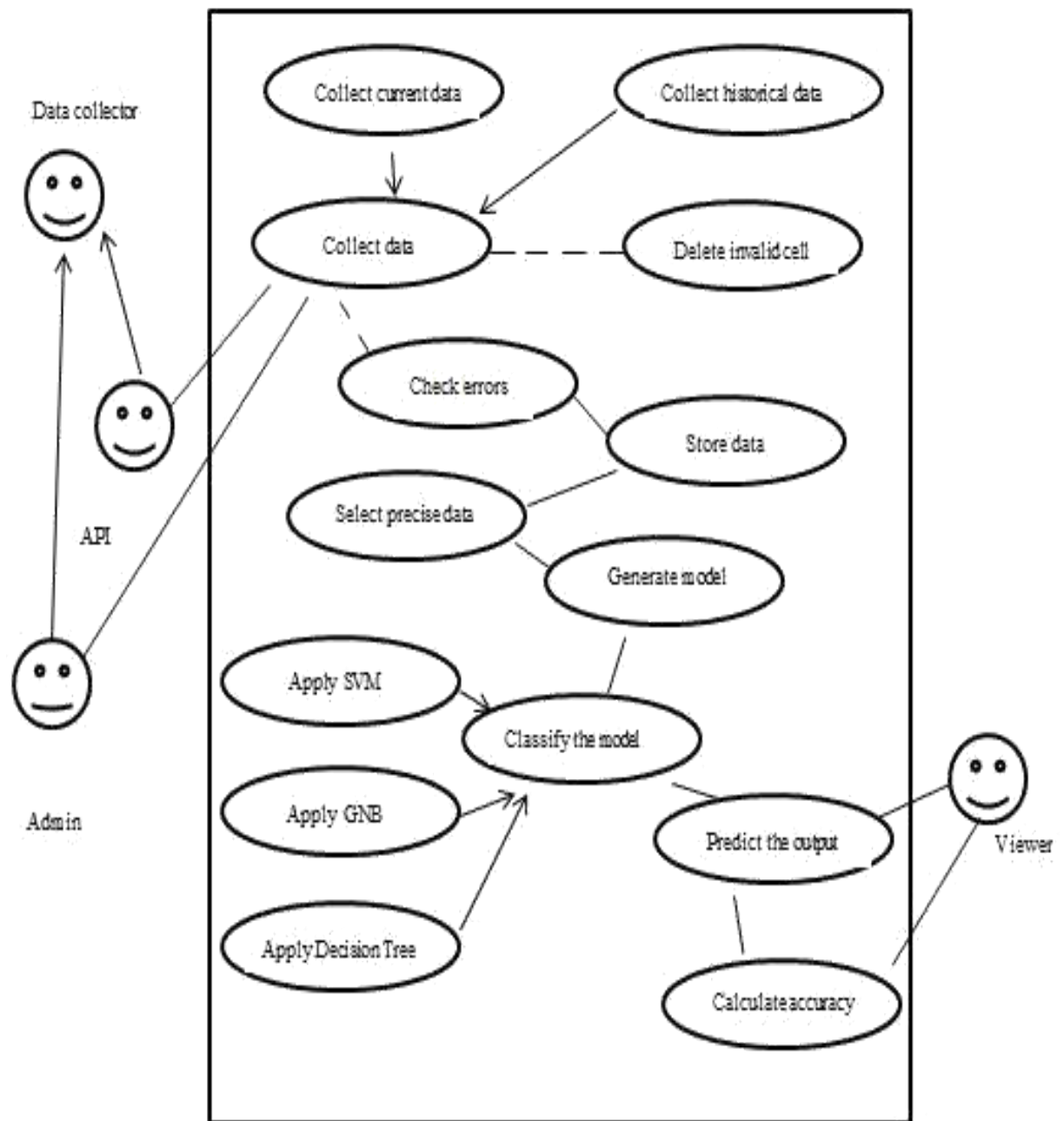


Figure 4.1: Use case diagram

4.2 Flow of Working Procedure

Data Flow Diagram: A Data Flow Diagram (DFD) shows the flow of data in any cycle or framework. It uses some clear images like rectangular, circle and bolt. It shows input-yield and the two elements of front-end and back-end. It depicts how info or crude data is made from user and each step of inner preparing before coming to administrator. Data Flow Diagram can be low level, mid-level and inside and out level. They can investigate the specific framework. DFD is the one of the simplest method to depict something that is hard to portray in word. It works for both specialized, non-specialized, from designers to CEO.

DFD speaks to the capacity, module, measure, which control, catch, store, recover, actualize, circulates data between a framework and its current circumstance and between its inner segments. It is regularly used for following explanation referenced beneath:

- Uses straightforward and less documentation.
- Flow of sensible data of the framework.
- Make a rundown or decide of physical necessities of framework.

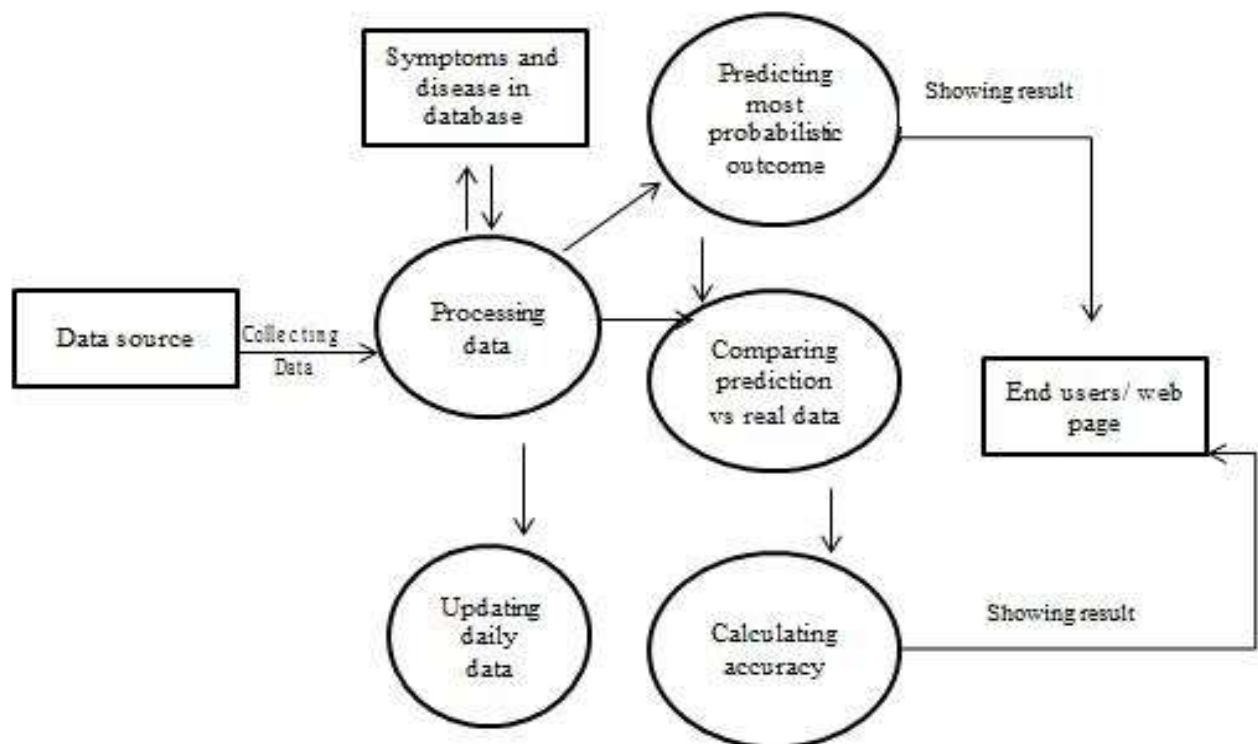


Figure 4.2: Data Flow Diagram 1

4.3 Database

Entity Relationship Model (E-R Diagram): Now daily, database is a basic aspect of a product framework. For using E-R Diagram in database designing ensure us to deliver great database plan. An E-R Diagram shows connection among the elements. Tables are known as entity. A table has essential key, unfamiliar key, and determined quality, and multivalued property, compound characteristic and so forth. A table has halfway or full reliance, 1 to 1, 1 to numerous relationships with other table. Database likewise has week entity set and seven days entity set must have seven days relationship.

E-R Diagram speaks to these relations graphically. It is useful to a database designer to plan a database and comprehend its framework from E-R Diagram. Our framework has a database. It contains all the information of Doctors, Patient and most important thing is the data of all the symptoms of disease. This data help to improve our prediction accuracy and also help to keep information of user.

This is our Entity Relationship model of our database. Each table has diverse quality which some of them are unpredictable and some of them are inferred and so on.

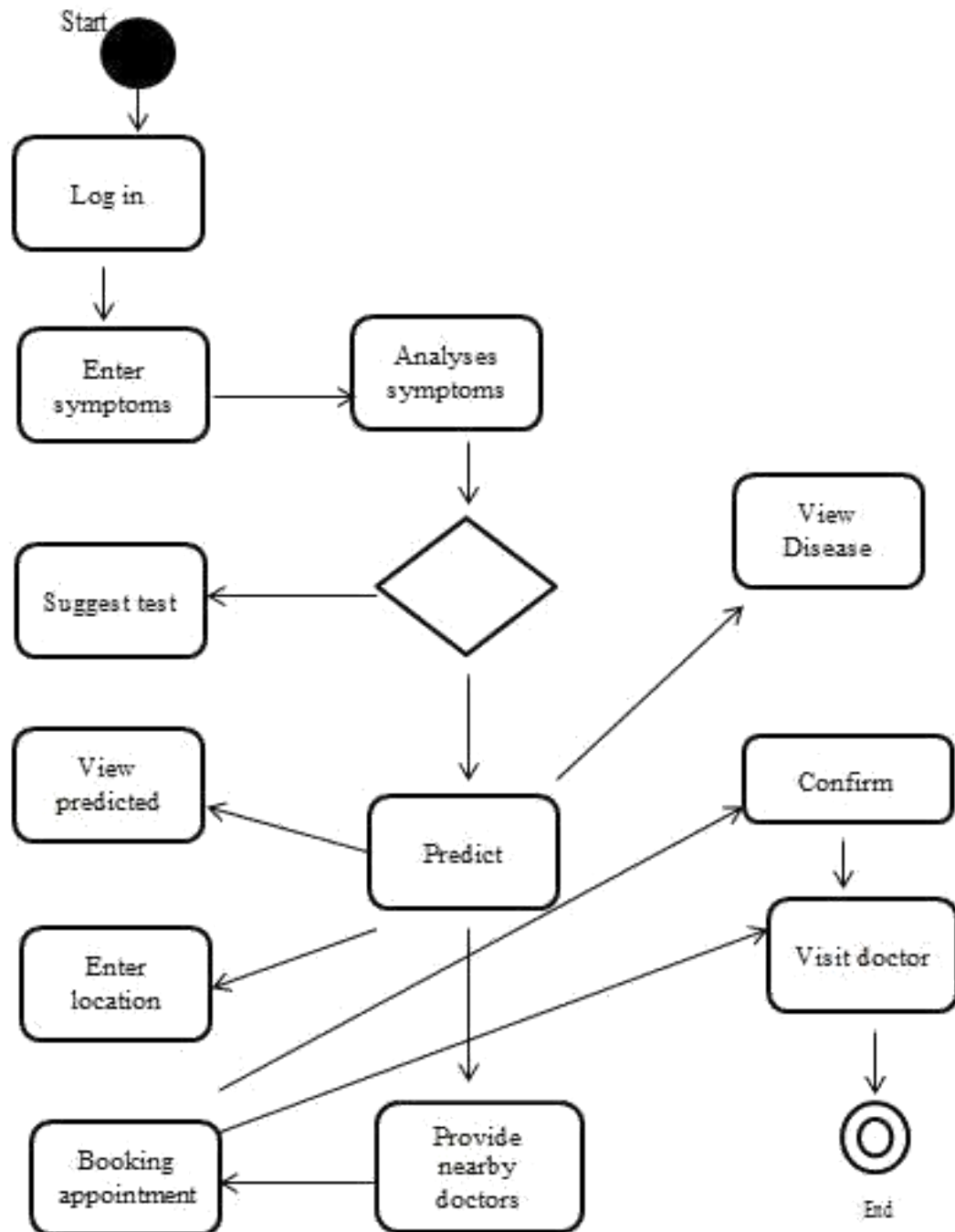


Figure 4.3: Entity Relationship Model

4.4 Implementation

Class Diagram: Class diagram shows us a static perspective on a framework. It clarifies all sort of qualities, techniques for various class and make a graphical portrayal about their connection among themselves. It is more close to execution of a framework. Practically all item arranged language uphold class diagram. Article situated writing computer programs is the assortment of class. We can undoubtedly compose vital code from a class diagram.

In class diagram each class spoke to by a rectangular shape. It rectangular has three sections. The upper bit demonstrates the class name. The center segment is rundown of all property with their data types. The last part hold all capacity or strategy with their boundary and return data type. A characteristic or strategy can be public, private and secured. It tends to be spoken to utilizing three documentation that is +, - , # individually.

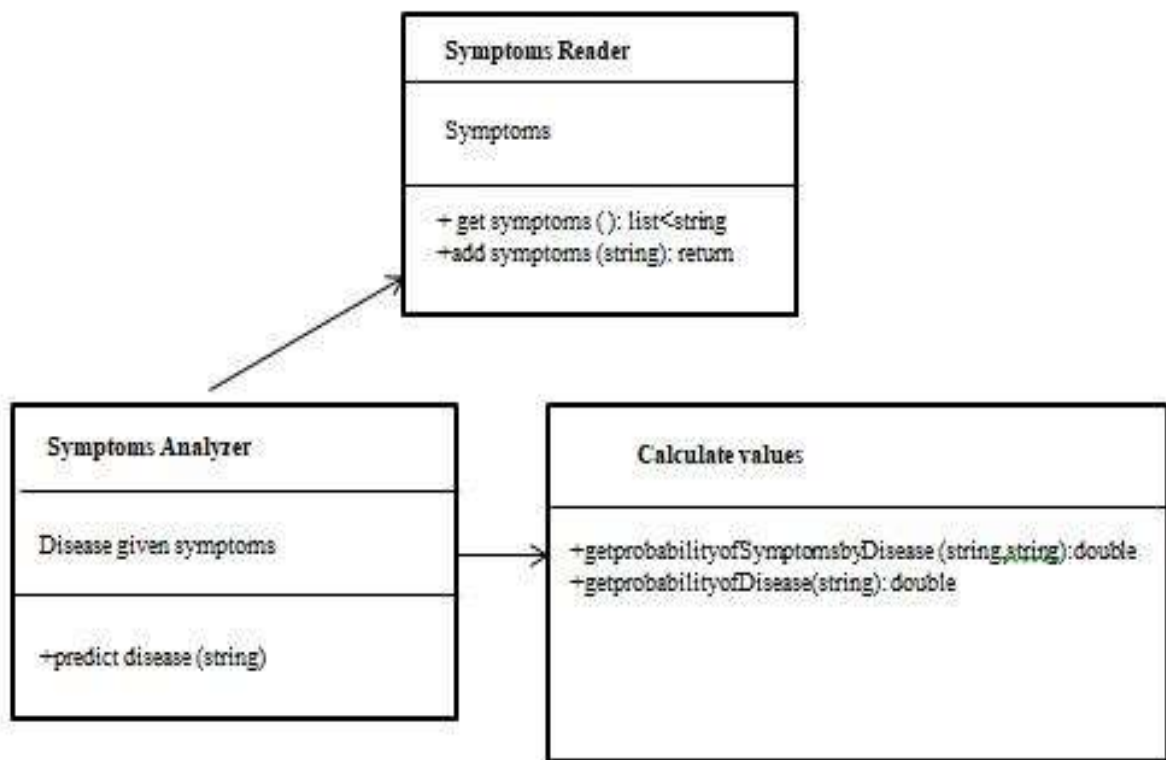


Figure 4.4: Class Diagram

CHAPTER 5

Experimental Results and Discussion

5.1 Introduction

This part will portray the consequences of our examination just as depicted about the objective of study which was decide to foresee the disease. Finally what we anticipated from our venture and really what we got. Here we will show our yield that we had got and quickly portray about it. To get more effective outcome about predictions we need to do some refresh in future.

5.2 Experimental Results

In this examination SVM, GNB, Decision Tree models are intended to anticipate the disease. Rundown of side effect which to be foresee disease. For indicating the outcome we take some example dataset underneath-

Table 2- Sample Data Sets

Symptom	Disease
Fever, Cough, Tiredness, Aches and pains, Headache, Loss of taste or smell. Diarrhea	Corona
Fever, Cough, Headache, Sore throat, Runny nose, Fatigue	Flu
Headache, Rashes, Fever	Dengue

Naïve Bayes classifier uses the accompanying principle to classify the datasets:

$$\prod_{i=1}^n P(\text{symptom} \mid \text{Disease})$$

User offers contribution to the framework. The information comprises of symptoms. The customer means the signs because of which the customer is feeling unwell.

- Fever
- Rashes
- Headache

The "Smart Health Prediction" system predicts the disease as demonstrated by the data instructive assortments and processes the probability of the disorder.

The example output is given as:

Table 3- Sample Output

Disease Name	Probability
Dengue	0.9%
Corona	0.4%
Flu	0.5%

5.3 Descriptive Analysis

We made a sqlite3 database containing three tables for example one for momentum symptoms of disease, another for anticipated disease yield and the last one for precision of prediction. In current table we put all chronicled data gathered from various sites. We figure out this table as day and hour.

The subsequent table is for anticipated data. We put away every day prediction of three kinds of calculation independently. This will make our prediction more precise step by step.

In conclusion, we contrasted our anticipated yield and real data and determined exactness. At that point we put away these precision in exactness table for contrasting the regular precision among them.

The huge classes in the application are:

Symptoms Reader:

This class is the run first when the client needs for illness expectation

Input: User chooses the manifestations from the rundown.

Output: The chose manifestations are placed in the rundown.

Symptoms Analyzer:

Input: Takes the client input for example side effects.

Output: Predicts the infection

Calculate Values:

Here the real numerical calculation happens.

5.4 Testing

The experiment intended for the task is given below:

Experiment I: Submit the side effects from the rundown
Precondition: The application is open.
Suspensions: The side effects for the infection are accessible
Test steps: 1. Select the checkbox from the rundown 2. Select submit
Expected Result: The side effects chose ought to be submitted and further dissected to figure the likelihood of the disease.

5.5 Summary

In spite of the fact that researchers are attempting to foresee disease from an earlier time however there is an a lot of issues. Health is anticipated through the issues. Besides, there are endless diseases which symptoms are exceptionally near one another. Likewise, they can change whenever. Disease isn't anticipated too for it tumultuous conduct. Thus, record-breaking prediction isn't acceptable as what we anticipated. In any case, yet we create it for the most part. In the point of view of Bangladesh, this kind of exploration isn't adequate and all around created. We are not really attempting from our level best to satisfy the prerequisites with the goal that it help us to fill our need. Shockingly there are a few constraints, for example, data isn't accessible for all the areas. We all realize that machine or PC works by relying upon guidance or calculation. It doesn't take any choice by its own. It just follows what we give as program. Machine can choose what's going on or right. It just takes a choice from prepared boundary or past data. We can no control the disease or any regular occasions. Yet, we are attempting our level best to make comparability by attempting an ever increasing number of times. In this manner, in the event that anybody feels any missing or improvement of our framework, don't hesitate to tell us.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication

6.1 Summery of the Study

There is no such prediction which gives us the total outcome and all of prediction has a few needs, constraints, blunders. Be that as it may, we were attempting our level best as per our capacity to improve the prediction result. There are immense measure of prediction application, division, research group and sites whose work constantly to improve the outcome. We have additionally remembered ourselves for this excursion. We learn about heaps of procedure, strategies, innovation, gadget identified with health prediction. From every one of them we had center just around the improvement of algorithmic application. We pick data mining for this reason. The execution of this way to deal with improves the prediction of various symptoms. We discovered three distinctive classification calculations, for example, Naïve Bayes, SVM, Decision Tree which are appropriate to for us.

6.2 Conclusion

Our examination and framework offers a superior and efficient health prediction model utilizing classification calculation. Classification is such theme that is used in data mining and AI generally. This can be helping us to get a dependable health prediction. Out of the specialized subject, our framework has great effect in our everyday life and our day by day works. We can take ideal choice in our existence with assistance of this framework. Step by step our model will give better outcomes. Because of refreshing our data sets consistently more unequivocally every day, the profundity of our model is expanding step by step. We trust after a specific time, our framework will give the yield more close to the real worth.

6.3 Recommendations

Here we used chronicled data to do the prediction. However, there is a genuine downside of the chronicled data extraordinarily the strategy of gathering and recording of data. There were numerous mistakes in authentic data that caused us to baffle to such an extent. Wrong dataset causes awful prediction. Along these lines, we suggest some extra necessities. In the event that we include equipment instruments such data assortment reason, at that point it will be so sensible and help us to get best prediction.

6.4 Implication for Further Study

Health prediction isn't as a basic errand as like other data mining and data prediction segment. From the result of this examination we will improve a framework. Health prediction isn't just programming or programming occupations. It additionally relies much upon equipment. Because of the count of enormous data we need a higher design PC or super PC and a major data community. Along these lines, this is the entirety of our likely arrangements.

APPENDIX

Appendix A: Project Reflection

We started our task work from most recent couple of months. Our essential target was to improve the precision. We search, download and read several exploration papers. We found a great deal of article about health prediction cycle and read them with cognizant psyche. We had most issues to gather data. There were a couple of locales which gave us a colossal sum data in free. Our next test was to change, altering, remedy of data. This cycle was incorporated to data preprocessing. Our next test was to change over this framework to an application stage for making this user cordial for our users. After quite a while dedicated and our hearted endeavors we complete our motivation.

Appendix B: Related Issues

Health prediction is unpredictable and it requests a difficult aptitudes and huge measure of data. We use classification to classify the diverse health credits as indicated by their class. We used sqlite3 database the executives framework to store data in database and web as GUI.

References

- [1] Z. Othman, K.M. Al-Aidaroos A.A. Bakar, "Medical Data Classification with Naive Bayes," *Information Technology Journal*, pp. 1166-1174, 2012.
- [2] K.Vasanthi, K.Bala Sita Ramireddy, Ravi Kumar Tenali, G.Pooja reddy M.Trinath basu, "Smart E-Health Prediction System Using Data," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 6, pp. 787-791, April 2019.
- [3] Nicholas Blumm, Nicholas Christakis, Albert-László Barabasi, Darcy A. Davis Nitesh V. Chawla. (2008, January) ACM Digital Library. [Online].
<https://dl.acm.org/doi/10.1145/1458082.1458185>
- [4] Dipesh Sharma, Jyoti Soni Ujma Ansari, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43-48, March 2011.
- [5] Shadab Adam Pattekari Asma Parveen, "PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES," *International Journal of Advanced Computer and Mathematical Sciences*, vol. 3, no. 3, pp. 290-294, 2012.
- [6] M.A.Nishara Banu B Gomathy, "DISEASE PREDICTING SYSTEM USING DATA," *International Journal of Technical Research and Applications*, vol. 1, no. 5, pp. 41-45, December 2013.
- [7] Kai Hwang, Lu Wang, Lin Wang, Min Chen Yixue Hao, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," *IEEE Access*, vol. 5, pp. 8869-8879, 2017.
- [8] Chien-Chung Chan, Feixiang Huang Shengyong Wang, "Predicting disease by using data mining based on healthcare information system," *IEEE International Conference on Granular Computing*, pp. 191-194, 2012.
- [9] Anitha Nithya R, Sujatha R Sumathy R, "A Survey of Health Care Prediction Using Data Mining," *International Journal of Innovative Research in Science*, vol. 5, no. 8, 2016.
- [10] Sunayana Dutta, Tridisha Das, Sweetey Buragohain, Susmita Sarma Pinky Saikia Dutta, "A Survey on Smart Health Care Using Data Mining," *ijcsetional Journal of Computer Sciences and Engineering neerin*, vol. 4, no. 7, pp. 74-79, December 2016.
- [11] Sunayana Dutta, Tridisha Das, Sweetey Buragohain, Susmita Sarma Pinky Saikia Dutta. (2016, December) *ijcsetional Journal of Computer Sciences and Engineering*. [Online].
https://www.ijcseonline.org/spl_pub_paper/NCRITET-IJCSE-2016016.pdf
- [12] N.Raviteja, K.Shashank, Ravi Aavula M.Kruthini, "Smart Health Consulting Android System," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 6, no. 3, October 2017.
- [13] Pallavi Chitte, Priyanka Vijay Pawar Megha Sakharam Walunj, "Estimation based on Data Mining Approach for Health Analysis," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 4, 2017.
- [14] N. Subhash Chandra, V. Krishnaiah G. Narsimha, "A Study On Clinical Prediction Using Data Mining Techniques," *International Journal of Computer Science Engineering and Information Technology Research (IJCEITR)*, vol. 3, no. 1, 2013.
- [15] Huda Saeed Salman, AmeerMohyAdden, Mona Mohmmad Rawan Ali AL jraib Wafamoqhas, "Biohouse Journal of Computer Science," *International Journal Series*, 2017.

- [16] Abu Raihan Mostofa Kamal, NafiulRashid, Md. Tahmid Rahman Laskar Md. Tahmid Hossain, "Automated Disease Prediction System (ADPS): A User Input-based Reliable Architecture for Disease Prediction," *International Journal of Computer Applications*, vol. 133, no. 15, 2016.
- [17] Siddhesh Murudkar, Prince Mishra, Prof. Krishna Kumar Tripathi Shubham Jawadwar, "A Smart Health Prediction Using Data Mining ," *International Research Journal of Engineering*, vol. 5, no. 4, pp. 2905-2908, April 2018.
- [18] Pallavi Chitte Priyanka, Vijay Pawar Megha Sakharam Walunj, "Estimation based on Data Mining Approach for Health Analysis," *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*, vol. 3, no. 1, 2013.

Final_Report_152-15-6065_Asiqul.pdf

ORIGINALITY REPORT

3%

SIMILARITY INDEX

0%

INTERNET SOURCES

0%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Palos Verdes High School

Student Paper

3%

2

Submitted to International School of
Management and Technology (ISMT), Nepal

Student Paper

<1%

Exclude quotes On

Exclude matches < 10 words

Exclude bibliography On