A PRELIMINARY PROJECT REPORT ON

Machine Learning for Prediction and Diagnosis of Cardiovascular Diseases

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

Babhulgaonkar Shashwat	Exam No:
Bendkule Pritam	Exam No:
Suryavanshi Jayesh	Exam No:
Sanap Vivek	Exam No:

DEPARTMENT OF COMPUTER ENGINEERING
K. K. Wagh Institute Of Engineering Education & Research

Hirabai Haridas Vidyanagari, Amrutdham, Panchavati, Nashik-422003

SAVITRIBAI PHULE PUNE UNIVERSITY

2018-19

K. K. Wagh Institute Of Engineering Education & Research

CERTIFICATE

This is to certify that the Project Entitled

Machine Learning for Prediction and Diagnosis of Cardiovascular Diseases

Submitted by

Babhulgaonkar Shashwat	Exam No:
Bendkule Pritam	Exam No:
Suryavanshi Jayesh	Exam No:
Sanap Vivek	Exam No:

is a bonafide work carried out by Students under the supervision of Prof. L. A. Patil and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University, for the award of Bachelor of Engineering (Computer Engineering).

Prof. L. A. Patil
Internal Guide
Dept. of Computer Engg.

Prof. Dr. S. S. Sane
H.O.D
Dept. of Computer Engg.

Prof. Dr. K. N. Nandurkar
Principal

Place: Nashik

Date:

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on ***Machine Learning for Prediction and Diagnosis of Cardiovascular Diseases****.*

We would like to take this opportunity to thank our internal guide ***Prof. L. A. Patil*** *for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful.*

We are also grateful to ***Prof. Dr. S. S. Sane****, Head of Computer Engineering Department, KKWIEER for his indispensable support, suggestions.*

In the end our special thanks to ***the Computer Engineering Department staff*** *for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for our Project.*

Babhulgaonkar Shashwat
Bendkule Pritam
Suryavanshi Jayesh
Sanap Vivek

(B.E. Computer Engg.)

ABSTRACT

Diagnosis and prediction of cardiovascular disease has often become a challenge faced by doctors and hospitals In India as well as abroad. Despite major transformations in lifestyles of people and advancements in medical domain; heart attacks still hold a major share in the global death rate. The ambiguity in diagnosis of most heart diseases lies in the intricate grouping of clinical and pathological data which may introduce misinterpretation of data among clinical experts, doctors and researchers. Ultimately, the problem lies within making decisions concerned with predicting and later diagnosing the heart diseases. These decisions can have a drastic effect on life of a person. The proposed approach to use machine learning for prediction as well as diagnostic purposes can play a very important role in this area. Various Machine Learning techniques can be used for classifying healthy people from the ones suffering from heart diseases. This work intends to presents a comprehensive review of prediction of Cardiac diseases by using Machine Learning based approach.

Key words: Cardiovascular diseases, Machine learning, clinical, pathological.

INDEX

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Millions die every year because of heart diseases across the globe. The major challenges lie in an early detection of these diseases and diagnosis. The proposed system aims to hasten the decision making process of patients as well as doctors, clinical researchers.

Various types of heart diseases are:

1)Coronary heart disease

2)Cardiomyopathy

3)Cardiovascular disease

4)Ischaemic heart disease

5)Heart failure

6)Hypertensive heart disease

7)Inflammatory heart disease

8)Valvular heart disease

Common risk factors of heart disease include:

1) High blood pressure

2) Abnormal blood lipids

3) Use of tobacco

4) Obesity

5) Physical inactivity

6) Diabetes

7) Age

8) Gender

9) Family generation

Data mining is the process of automatically extracting knowledgeable information from huge amounts of data. It has become increasingly important as real life data enormously increasing. Heart disease prediction system can assist medical professionals in predicting state of heart, based on the clinical data of patients fed into the system. There are many tools available which use prediction algorithms but they have some flaws. Most of the tools cannot handle big data. There are many hospitals and healthcare industries which collect huge amounts of patient data which becomes

difficult to handle with currently existing systems. Machine learning algorithm plays a vital role in analyzing and deriving hiddenKnowledge and information from these data sets. It improves accuracy and speed.

1.2 PROBLEM DEFINITION

Prediction of Heart Diseases Using Machine Learning and Data MiningAlgorithms and Tools-Design System H with Database D and applying algorithm A for detecting a disease C.

CHAPTER 2

LITERATURE SURVEY

- Intelligent Heart Disease Prediction System Using Data Mining Techniques
  Authors: Sellappan Palaniappan, Rafiah Awang The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not mined to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naive Bayes and Neural Network.

- Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design
  Authors: M. Raihan, Saikat Mondal, Arun More, Md. Omar Faruqe Sagor, Gopal Sikder, Mahbub Arab Majumder, Mohammad Abdullah Al Manjurand Kushal Ghosh An Android based prototype software has been developed by integrating clinical data obtained from patients admitted with IHD(Ischemic Heart Disease). The clinical data from 787patients has been analyzed and correlated with the risk factors like Hypertension, Diabetes, Dyslipidemia (Abnormal cholesterol), Smoking, Family History, Obesity, Stress and existing clinical symptom which may suggest underlying non detected IHD. The data was mined with data mining technology and a score is generated. Risks are classified into low, medium and high for IHD.

- Analysis of Data Mining Techniques for Heart Disease Prediction

  Authors: Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin Heart disease is considered as one of the major causes of death throughout the world. It cannot be easily predicted by the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. This paper addresses the issue of prediction of heart disease according to input attributes on the basis of data mining techniques. We have investigated the heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a standard data set as well as a collected data set. Based on performance factor SMO and Bayes Net techniques show optimum performances than the performances of Kstar, Multilayer Perceptron and J48 techniques.

- Machine Learning Application to Predict the Risk of Coronary Artery Atherosclerosis

  Authors:Soodeh Nikan, Femida Gwadry-Sridhar, and Michael Bauer Coronary artery disease is the leading cause of death in the world. In this research, we propose an algorithm based on the machine learning techniques to predict the risk of coronary artery atherosclerosis. A ridge expectation maximization imputation (REMI) technique is proposed to estimate the missing values in the atherosclerosis databases. A conditional likelihood maximization method is used to remove irrelevant attributes and reduce the size of feature space and thus improve the speed of the learning. The STULONG and UCI databases are used to evaluate the proposed algorithm. The performance of heart disease prediction for two classification models is analyzed and compared to previous work. Experimental results show the improved accuracy percentage of risk prediction of proposed method. The effect of missing value imputation on the prediction performance is also evaluated and the proposed REMI approach performs significantly better than conventional techniques.

- Prediction of Heart Disease Using Neural Network

  Authors : Tulay Karaylan, Ozkan Kilic

  Heart disease is a deadly disease that large population of people around the world suffers from. When considering death rates and large number of people who suffers from heart disease, it is revealed how important early diagnosis of heart disease. Traditional way of diagnosis is not sufficient for such an illness. Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way. In this paper, a heart disease prediction system which uses artificial neural network backpropagation algorithm is proposed. 13 clinical features were used as input for the neural network and then the neural network was trained with backpropagation algorithm to predict absence or presence of heart disease with accuracy of 95 percent.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

3.1.1 Project Scope

Comparative study of different Machine Learning algorithms will be done. The most efficient algorithm of them will be found and used. Create a Machine Learning program trained with a different data set. Develop a User-interactive application for Heart Disease Prediction.

3.1.2 Operational Details

[A] Support Vector Machine : Support vector machine (SVM) are supervised learning method that analyze data used for classification and regression analysis. It is given a set of training data, marked as belonging to either one of two categories, an SVM training algorithm then builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier . An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. The points are separated based on hyper planes that separate them. When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups.In the project, we have used this algorithm to classify the patients into groups according to the risk posed to them based on the parameters provided.

[B] Decision Trees : A decision tree is a flowchart-like structure in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules. In decision analysis, a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated. A decision tree

consists of three types of nodes: Decision nodes represented by squares Chance nodes represented by circles End nodes represented by triangles Decision trees are commonly used in operations research and operations management. If, in practice, decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities. Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods. In this project, we have used the decision tree to classify the patient according to the 14 parameters provided by the user.

[C] Naive-Bayes Algorithm : In machine learning we are often interested in selecting the best hypothesis (h) given data (d).In a classification problem, our hypothesis (h) may be the class to assign for a new data instance (d).One of the easiest ways of selecting the most probable hypothesis given the data that we have that we can use as our prior knowledge about the problem. Bayes Theorem provides a way that we can calculate the probability of a hypothesis given our prior knowledge.

Naive Bayes Classifier: In natural language processing, Naive Bayes Classifier is used to categorise the input data. It works on the basis of probability.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability — Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

1. P(c—x) is the posterior probability of class (target) given predictor (attribute).
2. P(c) is the prior probability of class.
3. P(x—c) is the likelihood which is the probability of predictor given class.
4. P(x) is the prior probability of predictor.

Success Conditions: Correct prediction of the Cardiovascular disease with an acceptable accuracy.

Failure Conditions: Incorrect classification of Cardiovascular disease.

3.1.3 Assumptions and Dependencies

1. The proposed system module of heart disease prediction using machine learning is only focus on heart diseases.

2. The workability of the proposed system modules such as those dealing with SVM with Database Schema is assumed.

3. The Database schema is depending on requirements of user.

4. Users are assumed to have fair estimate systems execution time, so that the detection of diseases and its results within time.

3.2 FUNCTIONAL REQUIREMENT

Format of data plays crucial part in this application. At the time of uploading the user data application will check its proper file format and if it not as per need then ERROR dialog box will be prompted. There will be the following three algorithms implemented: Support Vector Machine (SVM) Decision Tree Nave Bayes Algorithm The algorithms have been trained using the data set obtained from Various Resources. 75 percent of the entries in the data set have been used for training and the remaining 25 percent for testing the accuracy of the algorithm. Furthermore, some steps have been taken for optimizing the algorithms thereby improving the accuracy. These steps include cleaning the dataset and data preprocessing. The algorithms were judged based on their accuracy and it was observed that the SVM was the most accurate out of the three with 64.4 percent efficiency. Hence, it was selected for the main application. The main application is a web application which accepts the various parameters from the user as input and computes the result. The result is displayed along with the accuracy of prediction. Inputs: Data set, User Data Outputs: Result, Accuracy.

3.3 NON FUNCTIONAL REQUIREMENTS

3.3.1 Performance Requirements

Security vulnerabilities allow software to be abused in ways that the developers never intended. Imagine being able to design a hammer that can only hammer nails and nothing else. By building robust software security requirements, you can lock down what your software does so that it can only be used as intended. But this proposed project is user based project so we dont need any authentication for security of application.

3.3.2 Safety Requirements

This product uses the data of users for prediction purposes only, as is evident as there is no Name input field. Thus, the privacy of the user is preserved. But the program uses open source libraries and any developer having doubts about safety can directly check the source code.

3.3.3 Security Requirements

Security vulnerabilities allow software to be abused in ways that the developers never intended. Imagine being able to design a hammer that can only hammer nails and nothing else. By building robust software security requirements, you can lock down what your software does so that it can only be used as intended. But this proposed project is user based project so we dont need any authentication for security of application. Data will not be shared to other organizations. Data will be safe and secure.

3.3.4 Software Quality Attributes

Accuracy: The degree of accuracy with which the system will detect the disease. To determine the accuracy of the proposed system confusion matrix will be used. **Execution speed:** The system should be able to process the data fast, also classify it as and provide the result accordingly.

3.4 SYSTEM REQUIREMENTS

3.4.1 Database Requirements

Field Description Range and Values Age Age of the patient 0-100 in years Sex Gender of the patient 0-1 (1: Male 0: Female) Chest Pain Type of chest pain 1-4 (1: Typical Angina, 2: Atypical Angina, 3: Non-anginal, 4: Asymptotic) Resting Blood Pressure Blood pressure during rest mm Hg Cholesterol Serum Cholesterol mg / dl Fasting Blood Sugar Blood sugar content before food intake if ¿120 mg/dl 0-1 (0: False, 1: True) ECG Resting Electrocardiographic results 0-1 (0: Normal, 1: Having ST-T wave) Max Heart Rate Maximum heart beat rate. Beats/min Exercise Induced Angina Has pain been induced by exercise 0-1 (0: No, 1: Yes) Old Peak ST depression induced by exercise relative to rest 0-4 Slope of Peak Exercise Slope of the peak exercise ST segment 1-3 (1: Up sloping, 2: Flat, 3: Down sloping)

3.4.2 Software Requirements (Platform Choice)

Language : Python
Framework : Python
OS : Windows 10 or Ubuntu Linux 18.04
IDE : Microsoft Visual Studio

3.4.3 Hardware Requirements

Processor : Intel Core i5 or above
Graphics Card : Nvidia GTX 650 or AMD RX450 or above
Hard Disk : 80GB
RAM : 4GB or more

3.5 ANALYSIS MODELS : SDLC MODEL TO BE APPLIED

The analysis model used for the development of this system is Iterative model. An iterative life cycle model does not start with a full specification of requirements. In this model, the development begins by specifying and implementing just part of

the software, which is then reviewed in order to identify further requirements. Moreover, in iterative model, the iterative process starts with a simple implementation of a small set of the software requirements, which iteratively enhances the evolving versions until the complete system is implemented and ready to be deployed. Each release of Iterative Model is developed in a specific and fixed time period, which is called iteration.

Furthermore, this iteration focuses on a certain set of requirements. Each cycle ends with a usable system i.e., a particular iteration results in an executable release. Iterative Model allows accessing previous phases, in which the changes are made accordingly. The final output of the product is revived at the end of the Software Development Life Cycle (SDLC). Typically iterative development is used in conjunction with incremental development, in which a longer software development cycle is split into smaller segments that are built upon each other. Hence, iterative model is used in following scenarios:

- When the requirements of the complete system are clearly defined and understood.
- The major requirements are defined, while some functionalities and requested enhancements evolve with the process of the development process.
- A new technology is being used and is being learnt by the development team, while they are working on the project.

Since this suits well for this project, this model is being used.

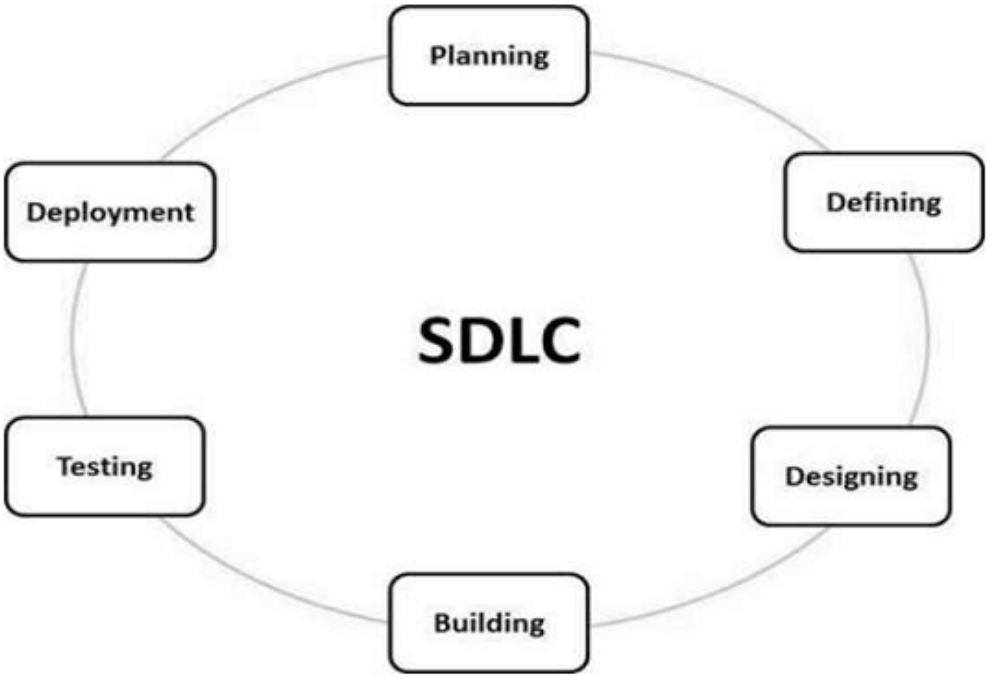


Figure 3.1: Software Development Life Cycle.

3.6 SYSTEM IMPLEMENTATION PLAN

Table 3.1: Project Implementation plan

Start Phase	End Phase	Description	Duration(days)
25/07/2018	31/07/2018	Preparatory Phase	7
02/08/2018	22/08/2018	Requirement Gathering	20
23/08/2018	26/08/2018	Approval Phase	3
27/08/2018	01/09/2018	Abstract and Synopsis	5
01/09/2018	10/10/2018	Presentation and Report Writing	30
20/10/2018	25/10/2018	Project Evaluation	5
26/10/2018	06/11/2018	Project Module 1 Designing	11
07/11/2018	25/11/2018	Project Module 2 Designing	18
27/11/2018	14/12/2018	Project Module 3 Designing	17
15/12/2018	25/12/2018	Project Evaluation	10
26/12/2018	10/01/2018	Testing	15
10/01/2018	25/01/2018	Delivery	15

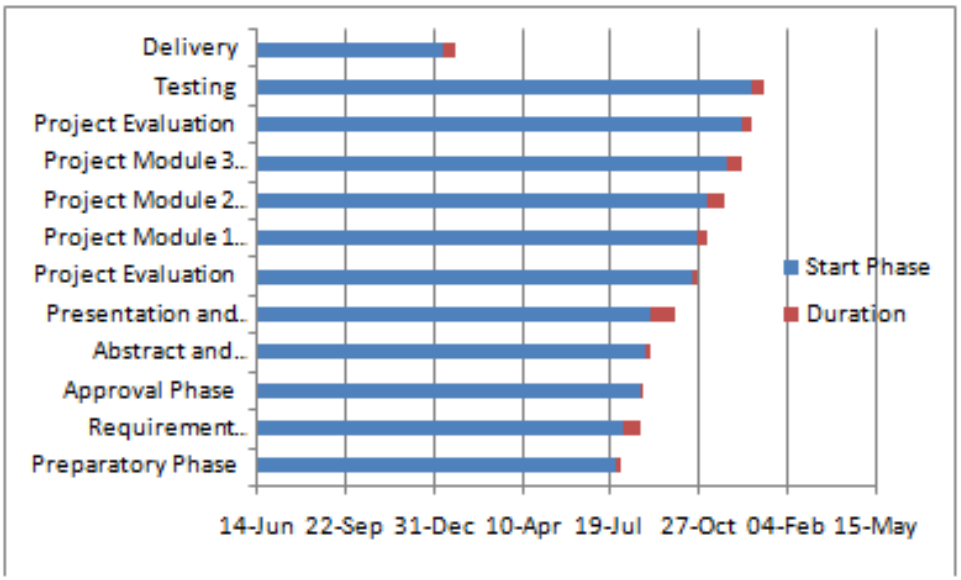


Figure 3.2: Gantt Chart.

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

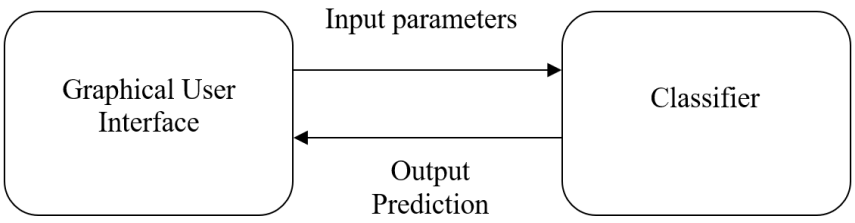


Figure 4.1: System Architecture.

4.2 DATA FLOW DIAGRAM

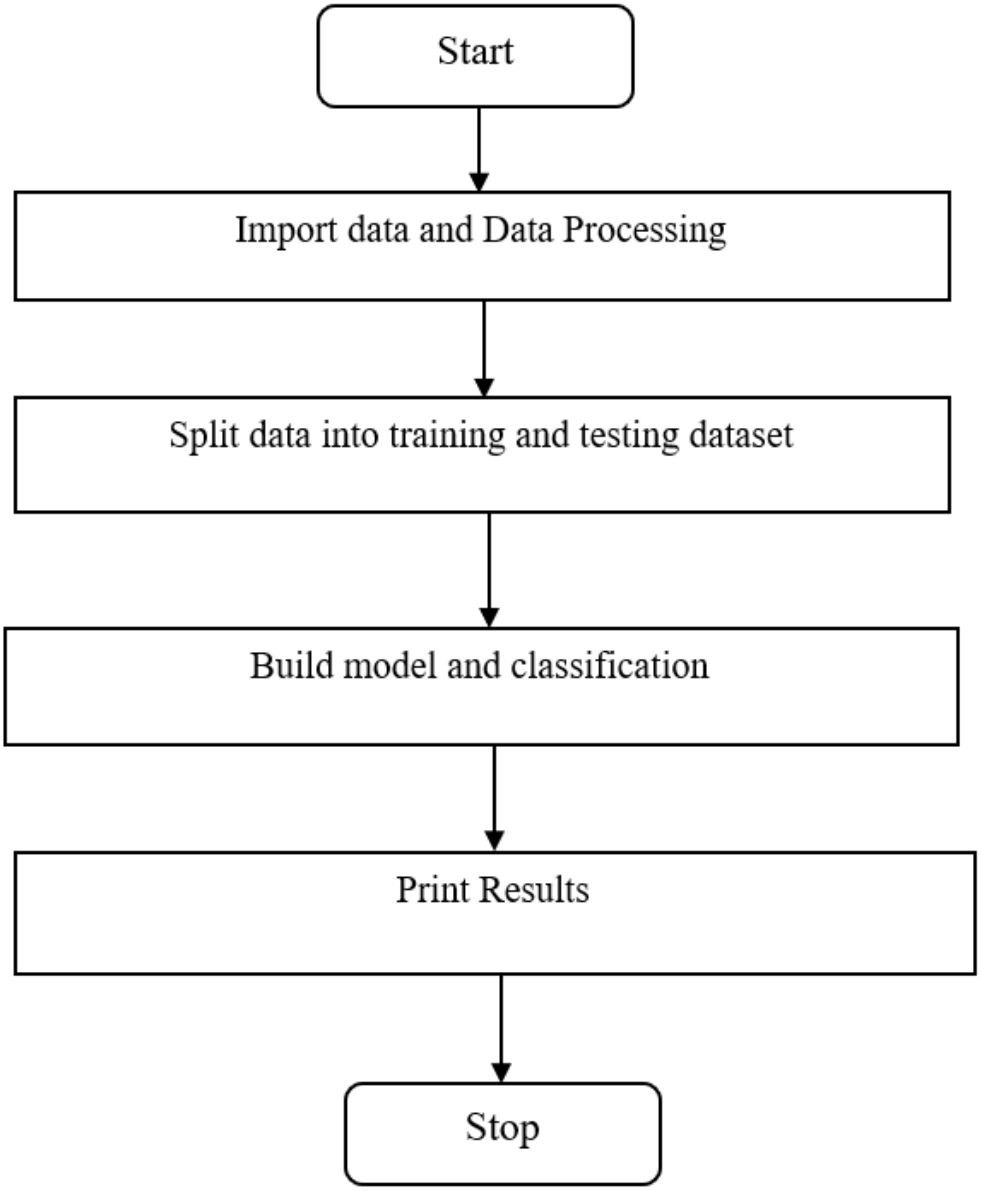


Figure 4.2: Data Flow Diagram.

4.3 UML DIAGRAM

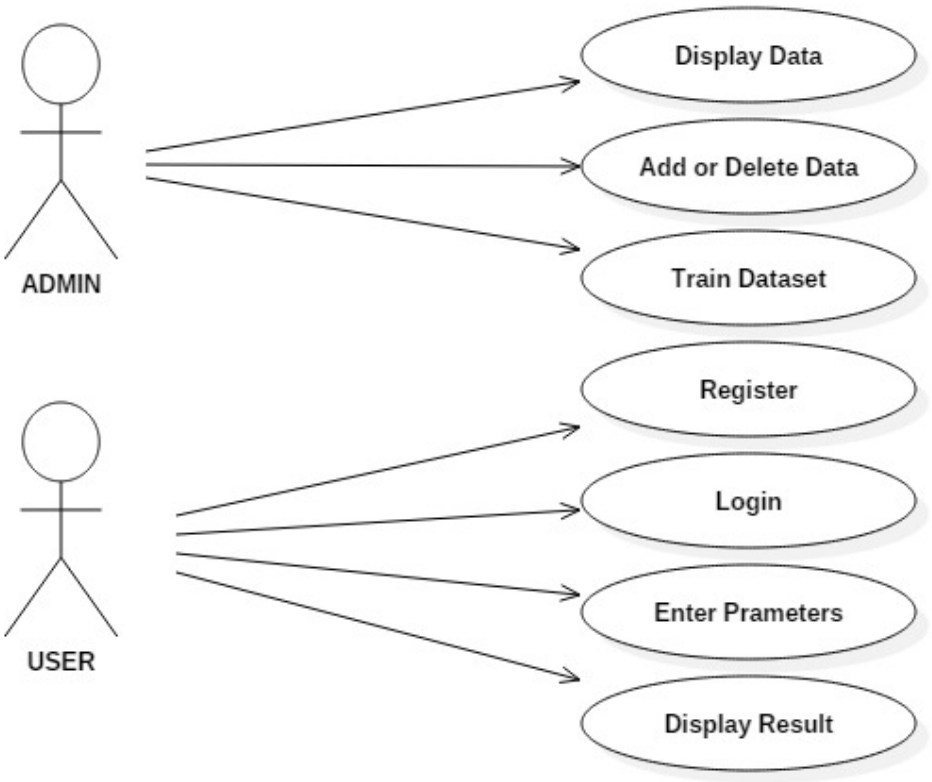


Figure 4.3: UML Diagram.

CHAPTER 5

OTHER SPECIFICATION

5.1 ADVANTAGES

- Requires a small amount of training data to estimate the parameters.
- Cost Effective
- Easy to implement.
- Good results obtained in most of the cases.

5.2 LIMITATIONS

- Assumption: class conditional independence, therefore loss of accuracy.
- Practically, dependencies exist among variables.
- Dependencies among these cannot be modeled by Nave Bayesian Classifier.

5.3 APPLICATIONS

Medical diagnosis plays vital role and yet complicated task that needs to be executed efficiently and accurately. To reduce cost for achieving clinical tests an appropriate computer-based information and decision support should be aided. Data mining is the use of software techniques for finding patterns and consistency in sets of data. Also, with the advent of data mining in the last two decades, there is a big opportunity to allow computers to directly construct and classify the different attributes or classes. Learning of the risk components connected with heart disease helps medicinal services experts to recognize patients at high risk of having Heart disease. Statistical analysis has identified risk factors associated with heart disease to be age, blood pressure, total cholesterol, diabetes, hyper tension, family history of heart disease, obesity and lack of physical exercise, fasting blood sugar etc.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Recognizing the disease is mainly the purpose of the proposed approach which can recognize the heart diseases with little computational effort. This approach can be used for the medical applications like detection and classification of diseases of heart with suitable classifier.

The proposed system will describes a possible,feasible approach for diagnosis. The proposed system addresses how the disease analysis is possible for the heart diseases detection, the analysis of the various diseases present in the heart can be effectively detected in the early stage before it will led to disastrous consequences.

ANNEXURE A

What are NP, P, NP-Complete and NP-Hard Problems?

P is set of plant leaves that can be solved by a feature extraction algorithm. NP is set of Decision making problems that can be solved ANN algorithm. P is subset of NP informally, NP is set of decision problems which can be solved by ANN algorithm, an algorithm that always makes a right guess from leaves dataset. NP-complete problems are the hardest problems in NP set.

A Decision problem L is NP complete if:

1. L is in NP.

2. Every problem in NP is reducible to L in polynomial time.

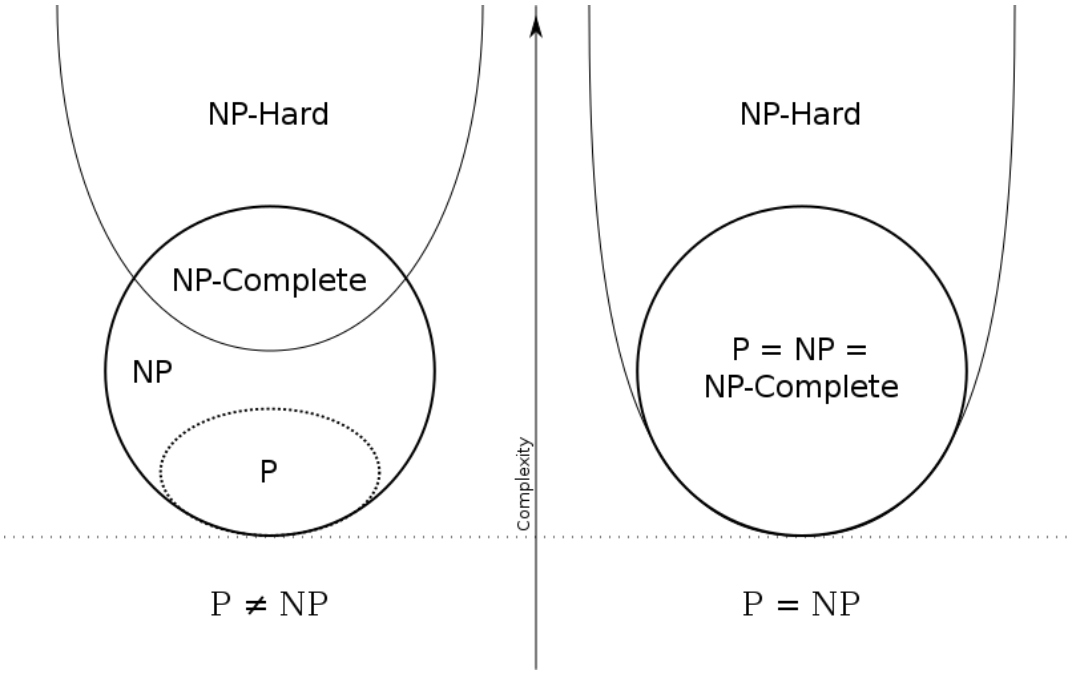


Figure A.1: NP, NP-hard and P Problems.

The problem in NP-Hard cannot be solved in polynomial time, until P = NP. If a problem is proved to be NPC, there is no need to waste time on trying to find an efficient algorithm for it. So proposed mathematical methods for cardiovascular disease detection system is NP-Complete. Because its reduces polynomial time in every next steps.

Our proposed project problem is NP-Complete problem because project contains decision making for detection of heart diseases.

ANNEXURE B

- PAPER: Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Asmita Mukherjee ”Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review”

  REVIEW: Comparative study by Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Asmita Mukherjee have given the Heart Disease Diagnosis and Prediction Using Machine Learning and Data. It is observed that hybrid models give very high accuracy if proper combinations of different algorithms are chosen. One of the major drawbacks of these works is that the main focus has been on the application of classification techniques for heart disease prediction, rather than studying various data cleaning and pruning techniques that prepare and make a dataset suitable for mining. It has been observed that a properly cleaned and pruned dataset provides much better accuracy than an unclean one with missing values. Selection of suitable techniques for data cleaning along with proper classification algorithms will lead to the development of prediction systems that give enhanced accuracy.

- PAPER:Prof. (Dr.) Kanak Saxena, Prof Head, Department of Computer Application, S.A.T.I., Richa Sharma,Efficient Heart Disease Prediction System using Decision Tree , In Proc. of UAI-99, pp.101108.

  REVIEW: Cardiovascular disease (CVD) is a big reason of morbidity and mortality in the current living style. Identification of Cardiovascular disease is an important but a complex task that needs to be performed very minutely, efficiently and the correct automation would be very desirable. Every human being can not be equally skillful and so as doctors. An automated system in medical diagnosis would enhance medical care and it can also reduce costs. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the users requirement. The performance of the system is evaluated in terms of classification accuracy and the results shows that the system has great potential in predicting the heart disease risk level more accurately.

- PAPER:SP Rajamhona, C. Akalya Devi, K Umamaheshwari, R Kirubha, K Karunya,Analysis of Neural Network Based Heart Prediction system,International Issues of Computer science, June 2017,pp, 150-154
  REVIEW: Heart Disease is one of the major reasons for increase in the death rate. Health care is one among the most important beneficial of huge knowledge and anlytics. Exacting medical data is progrsively becoming more and more necessary for prediction and treatment of high death rate. Terabytes of data produced everyday.Quality service are avoid poor clinical decisions taht lead to disastrous consequences. The hospitals can make sure of appropriate decision support systems thus minimizing the cost of clinical tests.

- PAPER:Prabakaran.N and Kannadasan.R,Prediction of Cardiac Disease Based on Patients Symptoms , Journal of Health care Engineering, 2017, VOL 2017
  REVIEW: In this paper a reliable multi process method combining decision tree techniques and clustering to build a cardiac arrest risk prediction system is proposed. Cardiac arrest has become the leading cause of death worldwide. The most effective way to reduce such deaths is to detect its symptoms earlier. Many people avoid such screening de effective preventive strategy. This system can also be used as a source of record with detailed patient history in hospitals as well as help doctors to concentrate on particular therapy for any patient.

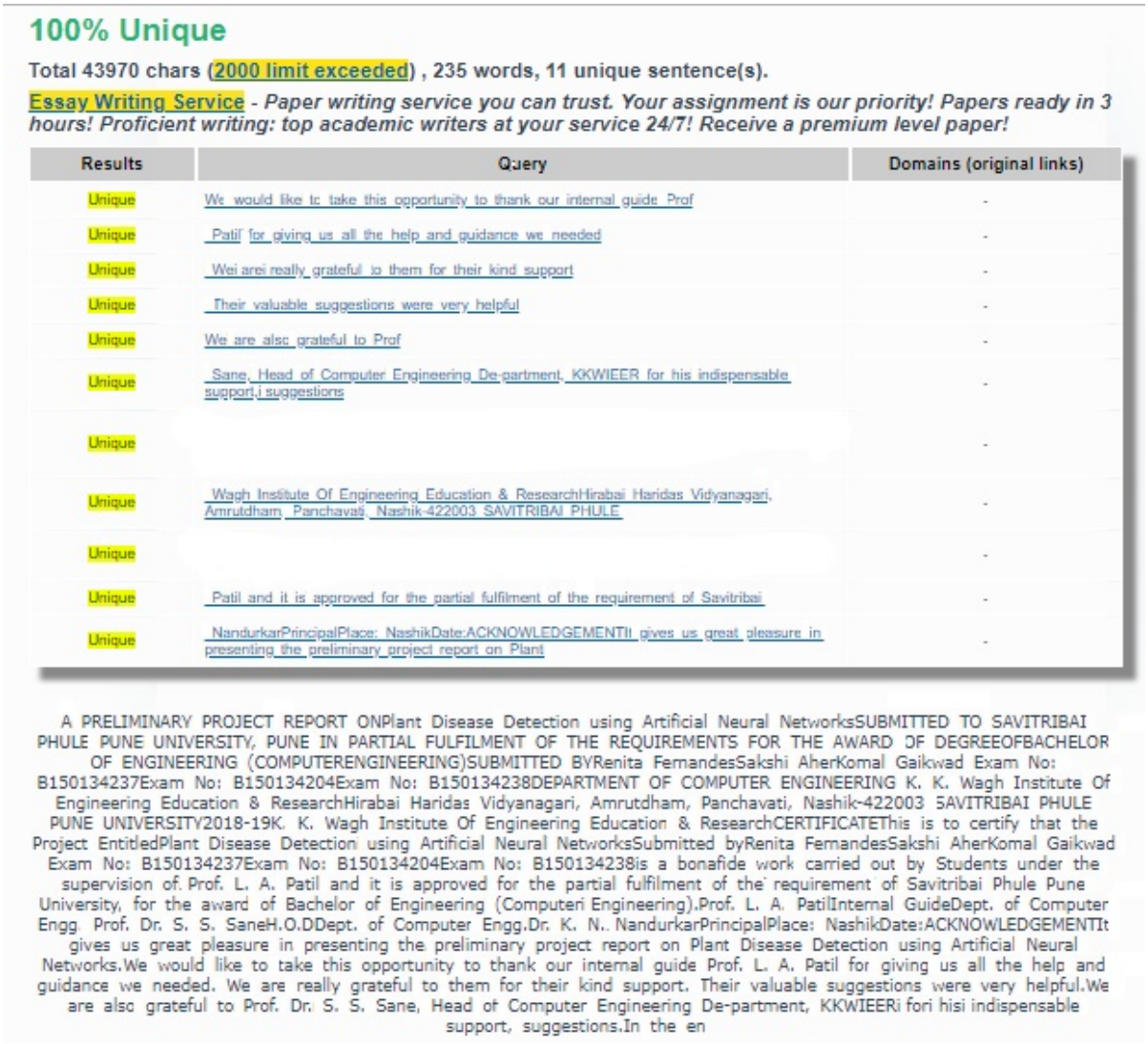
- PAPER: Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD),Prediction of Heart Disease Using Machine Learning ,Journal of Health care Engineering, 2018, VOL 2018
  REVIEW: With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. Thus we propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc. The machine learning algorithm neural networks

has proven to be the most accurate and reliable algorithm and hence used in the proposed system.

- PAPER:Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep,Heart Disease Prediction using Evolutionary Rule Learning
  REVIEW: TIn modern society, Heart disease is the noteworthy reason for short life. Large population of people depends on the healthcare system so that they can get accurate result in less time. Large amount of data is produced and collected by the healthcare organization on the daily basis. To get intriguing knowledge, data innovation permits to extract the data through automization of processes. Weighted Association Rule is a type of data mining technique used to eliminate the manual task which also helps in extracting the data directly from the electronic records. This will help in decreasing the cost of services and also helps in saving lives. In this paper, we will find the rule to predict patients risk of having coronary disease. Test results have shown that vast majority of the rules helps in the best prediction of coronary illness.

ANNEXURE C

PLAGIARISM REPORT



100% Unique

Total 43970 chars (2000 limit exceeded) , 235 words, 11 unique sentence(s).

Essay Writing Service - *Paper writing service you can trust. Your assignment is our priority! Papers ready in 3 hours! Proficient writing: top academic writers at your service 24/7! Receive a premium level paper!*

Results	Query	Domains (original links)
Unique	We would like to take this opportunity to thank our internal guide Prof	-
Unique	Patil for giving us all the help and guidance we needed	-
Unique	Wei arei really grateful to them for their kind support	-
Unique	Their valuable suggestions were very helpful	-
Unique	We are alsc grateful to Prof	-
Unique	Sane, Head of Computer Engineering De-partment, KKWIEER for his indispensable support,i suggestions	-
Unique		-
Unique	Wagh Institute Of Engineering Education & ResearchHirabai Haridas Vidyanagari, Amrutdham Panchavati Nashik-422003 SAVITRIBAI PHULE	-
Unique		-
Unique	Patil and it is approved for the partial fulfilment of the requirement of Savitribai	-
Unique	NandurkarPrincipalPlace: NashikDate:ACKNOWLEDGEMENTIt gives us great pleasure in presenting the preliminary project report on Plant	-

A PRELIMINARY PROJECT REPORT ONPlant Disease Detection using Artificial Neural NetworksSUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREEOFBACHELOR OF ENGINEERING (COMPUTERENGINEERING)SUBMITTED BYRenita FernandesSakshi AherKomal Gaikwad Exam No: B150134237Exam No: B150134204Exam No: B150134238DEPARTMENT OF COMPUTER ENGINEERING K. K. Wagh Institute Of Engineering Education & ResearchHirabai Haridas Vidyanagari, Amrutdham, Panchavati, Nashik-422003 SAVITRIBAI PHULE PUNE UNIVERSITY2018-19K. K. Wagh Institute Of Engineering Education & ResearchCERTIFICATEThis is to certify that the Project EntitledPlant Disease Detection using Artificial Neural NetworksSubmitted byRenita FernandesSakshi AherKomal Gaikwad Exam No: B150134237Exam No: B150134204Exam No: B150134238is a bonafide work carried out by Students under the supervision of. Prof. L. A. Patil and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University, for the award of Bachelor of Engineering (Computeri Engineering).Prof. L. A. PatilInternal GuideDept. of Computer Engg. Prof. Dr. S. S. SaneH.O.DDept. of Computer Engg.Dr. K. N.. NandurkarPrincipalPlace: NashikDate:ACKNOWLEDGEMENTIt gives us great pleasure in presenting the preliminary project report on Plant Disease Detection using Artificial Neural Networks.We would like to take this opportunity to thank our internal guide Prof. L. A. Patil for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful.We are alsc grateful to Prof. Dr. S. S. Sane, Head of Computer Engineering De-partment, KKWIEERi fori hisi indispensable support, suggestions.In the en

Figure C.1: Plagiarism report.

ANNEXURE D

REFERENCES

1. Prediction of heart disease using a hybrid technique in data mining classification Ankita Dewan ; Meghna Sharma

   2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)

2. Heart disease prediction system based on hidden nave bayes classifier M. A. Jabbar ; Shirina Samreen

   2016 International Conference on Circuits, Controls, Communications and Computing (I4C) Year: 2016

3. Heart Diseases Prediction with Data Mining and Neural Network Techniques Bandarage Shehani Sanketha Rathnayakc ; Gamage Upeksha Ganegoda

   2018 3rd International Conference for Convergence in Technology (I2CT) Year: 2018

4. Prediction of Cardiac Disease Based on Patient's Symptoms N. Prabakaran ; R. Kannadasan

   2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) Year: 2018

5. Comparative Analysis of Classification Approaches for Heart Disease Prediction S. M. M. Hasan ; M. A. Mamun ; M. P. Uddin ; M. A. Hossain

   2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) Year: 2018

6. Analysis of Neural Networks Based Heart Disease Prediction System SP Rajamhoana ; C. Akalya Devi ; K. Umamaheswari ; R. Kiruba ; K. Karunya ; R. Deepika

   2018 11th International Conference on Human System Interaction (HSI) Year: 2018