International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

Heart Disease Identification Method Using Machine Learning Classification in E Healthcare

Santhosh V [1], Dr.E. Pandiyan[2]
1.IIMCA, Sri Muthukumaran Institute of Technology, Mangadu
2.Professor,MCA, Sri Muthukumaran Institute of Technology, Mangadu

Submitted: 15-08-2022 Revised: 27-08-2022 Accepted: 30-08-2022

ABSTRACT
This Project work which initiated at an early detection of all the probable symptoms and signs which might further lead to detection of heart diseases using data collected from previous patients as well as data input received from the user at that particular time. Current scenario of health-care data used for surveillance are no longer simply a time building series of aggregate daily counts. Instead, a wealth of proposed spatial as well as temporal demographic, and symptom information is available at the data presented during the time of execution. Our proposed method incorporates all such information that is being used as a classification approach that compares recent healthcare data against data from that particular baseline distribution. In addition, the data sample data used is first train and test the system using machine learning approaches. the proposed system trained to be Logistic Regression, K-Nearest Neighbors, Random Forest, XGBoost algorithm are used. Then proposed test scores have been evaluated. Classifier is further chosen to make predictions.

I. INTRODUCTION

Cardiac arrest is one of the fatal attacks in the world that results in the supremacy of death. Heart attacks is caused by a sudden occurrence of coronary thrombosis, typically which results in the death of a particular heart muscle and sometime scan be fatal. A heart attack happens if the flow of oxygen-rich blood to a section of heart muscle suddenly becomes blocked and the heart can't get oxygen. When plaque builds up in the arteries, the condition is called atherosclerosis. The professional build-up of plaques presents in the arteries occur over many years. Eventually, an area of plaque can rupture (break open) inside of an artery. This can cause a blood clot that can be formed on the surface of the plaque. The flow of blood through the clot becomes large. If any blockage present isn't treated quickly, the portion of heart muscle that can be fed by the artery can lead to death of that particular artery.

Healthy heart tissue is replaced with scar tissue. This heart damage may not be obvious, which further caused long-lasting and severe problems. A majority of the heart attacks occur as a result proportion to coronary heart disease. Coronary heart disease is a condition in which a wax like substance that can be termed as plaque builds up inside of the coronary arteries. Only early prediction could help to better diagnose the cardiac problems at the benign stage to save a person's life. A less initiated common cause of heart attack is that of severe spasm or rather tightening of a coronary artery. The spasm can cutoff the flow of the blood through the artery. Atherosclerosis does not show effect on the spasms present in the coronary arteries. Heart attacks that can be associated with or can lead to severe problems that can diminish the health of an individual, such as heart failure and also can lead to life-threatening arrhythmias. Heart failure is a condition in which the heart can't pump enough blood to meet the body's needs. Irregular heartbeats are called Arrhythmias. Ventricular fibrillation present, is connected to a life-threatening arrhythmia that led to death if not treated the right away.

There are several factors that could affect a person's predisposition for Cardiac disease. Education is an important indicator of socioeconomic status that is associated with the occupation and also among the other factors affecting an individual's life style. A number of studies in developed countries have shown that Cardiac disease incidence varies between people with different levels of education.

1.ARTIFICIAL INTELLIGENCE:
2.MACHINE LEARNING
3.LOGISTIC REGRESSION ALGORITHM
4.RANDOM FOREST ALGORITHM
5.XGBOOST

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

6.KNN (K-NEAREST NEIGHBOUR)
LITERATURE REVIEW

[1] Title: Prediction of heart disease by using machine learning

Authors: Rohit Murty, Satish Patle, Saurabh Bute, Sneha Bhilkar, Durga Wanjari - 2020
Description:

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.

[2] Title: Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset
Authors: Tapas Ranjan Baitharua, Subhendu Kumar Panib - 2016
Description:

Accuracy in data classification depends on the dataset used for learning. Now-a-days the most important cause of death for both men and women is due to the Liver Problem. The healthcare industry collects a huge amount of data which is not properly mined and not put to the optimum use. Discovery of these hidden patterns and relationships often goes unexploited. Our research focuses on this aspect of Medical diagnosis by learning pattern through the collected data of Liver disorder to develop intelligent medical decision support systems to help the physicians. In this paper, we propose the use decision trees J48, Naive Bayes, ANN, ZeroR, 1BK and VFI algorithm to classify these diseases and compare the effectiveness, correction rate among them. Detection of Liver disease in its early stage is the key of its cure. It leads to better performance of the classification models in terms of their predictive or descriptive accuracy, diminishing of computing time needed to build models as they learn faster, and better understanding of the models. In this paper, a comparative analysis of data classification accuracy using Liver disorder data in different scenarios is presented. The predictive performances of popular classifiers are compared quantitatively.

[3] Title: Improving Disease Prediction by Machine Learning
Authors: Smriti Mukesh Singh1, Dr. Dinesh B. Hanchate2 – 2018
Description:

These days utilization of Big Data is expanding in biomedical and human services groups, exact investigation of medicinal information benefits early malady discovery, quiet care and group administrations. Fragmented therapeutic information lessens examination precision. The machine learning calculations are proposed for successful expectation of ceaseless infection. To beat the trouble of deficient information, Genetic algorithm will be utilized to remake the missing information. The dataset comprises of structured data and unstructured data. To extract features from unstructured data RNN algorithm will be utilized. Framework proposes SVM calculation and Naive Bayesian calculation for sickness expectation utilizing unstructured and structured information individually from hospital information. Community Question Answering (CQA) system is additionally proposed which will foresee the inquiry and answers and will give proper responses to the clients. For that, two calculations are proposed KNN and SVM. KNN algorithm will perform classification on answers and SVM calculation will perform classification on answers. It will help client to discover best inquiries and answers identified with infections.

[4] Title: Heart Disease Prediction System Using Data Mining Techniques
Authors: Abhishiek Taneja – 2015
Description:

In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person so instantly that it hardly gets any time to get treated with. So diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to develop a cost-effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately, most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

a complex activity. In this paper using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question.

[5] Title: Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques

Authors: Chaitrali S. Dangare, Sulabha S. Apte - 2016

Description:

The Healthcare industry is generally "information rich", but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. Advanced data mining techniques are used to discover knowledge in database and for medical research, particularly in Heart disease prediction. This paper has analyzed prediction systems for Heart disease using a greater number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. This research paper added two more attributes i.e., obesity and smoking. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database. The performance of these techniques is compared, based on accuracy. As per our results accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively. Our analysis shows that out of these three classification models Neural Networks predicts Heart disease with highest accuracy.

II. PROBLEM STATEMENT

A heart attack happens if the flow of oxygen-rich blood to a section of heart muscle suddenly becomes blocked and the heart can't get oxygen. The heart damages caused long-lasting and severe problems. A majority of the heart attacks occur as a result proportion to coronary heart disease. Coronary heart disease is a condition in which a wax like substance that can be termed as plaque builds up inside of the coronary arteries. Only early prediction could help to better diagnose the cardiac problems at the benign stage to save a person's life.

III. EXISTING SYSTEM

- Existing system developed Heart Disease classification system by using machine learning classification techniques and the performance of the system was 77% in terms of accuracy.
- Cleveland dataset was utilized with the method of global evolutionary and with features selection method.
- DBP algorithm along with FS algorithm and performance was not good

DISADVANTAGES

- Existing system performance is very low
- Computationally complex
- More Execution time required to generate results

DEVELOPMENT PROCESS :-

REQUIREMENT ANALYSIS :

Requirements are a feature of a system or description of something that the system is capable of doing in order to fulfill the system's purpose. It provides the appropriate mechanism for understanding what the customer wants, analyzing the needs assessing feasibility, negotiating a reasonable solution, specifying the solution unambiguously, validating the specification and managing the requirements as they are translated into an operational system.

- **PYTHON**
- **ANACONDA**
- **ANACONDA NAVIGATOR**
- **JUPYTER NOTEBOOK**

RESOURCE REQUIREMENTS:

SOFTWARE REQUIREMENTS:

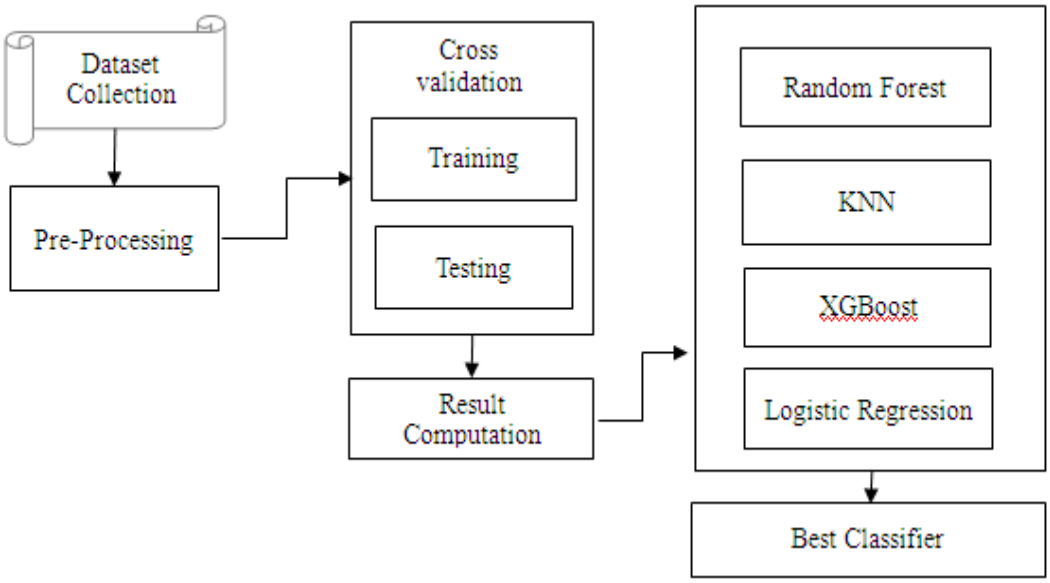
OPERATING SYSTEM	WINDOWS 7 OR LATER
SIMULATION TOOL	ANACONDA (JUPYTER BOOK)
DOCUMENTATION	MS-OFFICE

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

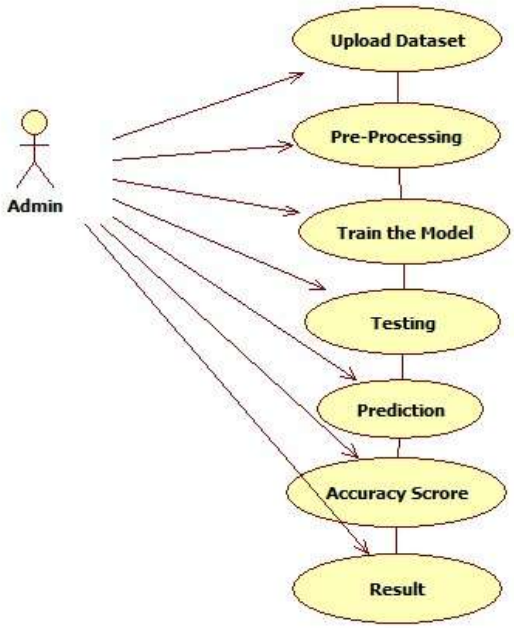
HARDWARE REQUIREMENTS:

CPU type	I5 and Above
Ram size	4GB
Hard disk capacity	80 GB
Keyboard type	Internet keyboard
Monitor type	15 Inch colour monitor
CD -drive type	52xmax

SYSTEM ARCHITECTURE



USE CASE DIAGRAM



International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

PROPOSED SYSTEM

- Our proposed system involves Logistic regression, XG Boost, Random Forest classifier, KNN Algorithm in Machine Learning concept used to train the dataset.
- Thus, preventing Heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives.
- This is where Machine Learning comes into play. Machine Learning helps in predicting the Heart diseases, and the predictions made are quite accurate.
- The project involved analysis of the heart disease patient dataset with proper data processing. Then, different models were trained and predictions are made with different algorithms KNN, Decision Tree, Random Forest, Logistic Regression etc.
- This is the jupyter notebook code and dataset I've used for my Kaggle kernel 'Binary Classification with Sklearn and Kera's'

ADVANTAGES

- Easy detection of the Cardiac disease with the concluded technique.
- Time consuming.
- Best accuracy Model helps in better treatment as early.
- Detection of best Model will quick the treatment which is life saving

SYSTEM MODULES:

- Module 1: Dataset Collection and pre-processing
- Module 2: EDA concept
- Module 3: Train the model
- Module 4: Evaluation
- Module 5: Comparison of existing model
- Module 6: Performance analysis

Module 1: Dataset Collection and Pre-processing

A dataset (or data set) is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the dataset in question. It lists values for each of the variables, such as height and weight of an object. Each value is known as a datum.

We have chosen to use a publicly-available Healthcare dataset which contains a relatively small number of inputs and cases. The data is arranged in such a way that will allow those trained in medical disciplines to easily draw parallels between familiar statistical and novel ML techniques. Additionally, the compact dataset enables short computational times on almost all modern computers.

The sklearn preprocessingpackage provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators.

In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate. The behaviors of the different scalers, transformers, and normalizers on a dataset containing marginal outliers are highlighted in Compare the effect of different scalers on data with outliers.

Standardization, or Mean removal and Variance Scaling

Standardization of datasets is a common requirement for many machine learning estimators implemented in scikit-learn; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with zero mean and unit variance**.**

Scaling features to a range

In practice we often ignore the shape of the distribution and just transform the data to center it by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation.

For instance, many elements used in the objective function of a learning algorithm (such as the RBF kernel of Support Vector Machines or the l1 and l2 regularizes of linear models) assume that all features are centered on zero and have variance in the same order. If a feature has a variance that is orders of magnitude larger than others, it might dominate the objective function and make the estimator unable to learn from other features correctly as expected.

An alternative standardization is scaling features to lie between a given minimum and maximum value, often between zero and one, or so that the maximum absolute value of each feature is scaled to unit size. This can be achieved using MinMaxScaler or MaxAbsScaler, respectively.

The motivation to use this scaling includes robustness to very small standard deviations of features and preserving zero entries in sparse data.

MaxAbsScaler works in a very similar fashion, but scales in a way that the training data lies within the range [-1,1] by dividing through the largest maximum value in each feature. It is meant for data that is already centered at zero or sparse data.

Normalization

Normalization is the process of scaling individual samples to have unit norm. This process can be useful

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

if you plan to use a quadratic form such as the dot-product or any other kernel to quantify the similarity of any pair of samples.
This assumption is the base of the Vector Space Model often used in text classification and clustering contexts.

Module 2:EDA concept

Exploratory Data Analysis or (EDA) understands the data sets by summarizing their main characteristics often plotting them visually. This step is very important especially when we arrive at modeling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plot, Scatter plot and many more. It often takes much time to explore the data. Through the process of EDA, we can ask to define the problem statement or definition on our data set which is very important. So in this tutorial, we will explore the data and make it ready for modeling.

1. Importing the required libraries for EDA

The libraries that are used in order to perform EDA (Exploratory data analysis)

2. Loading the data into the data frame.

Loading the data into the pandas data frame is certainly one of the most important steps in EDA, as we can see that the value from the data set is comma-separated. One thing to remember in this step is that uploaded files will get deleted when this runtime is recycled.

3. Checking the types of data

The datatypes because sometimes the MSRP or the price of the car would be stored as a string or object, if in that case, we have to convert that string to the integer data only then we can plot the data via a graph.

4. Dropping irrelevant columns

This step is certainly needed in every EDA because sometimes there would be many columns that we never use in such cases dropping is the only solution. In this case, the columns such as Engine Fuel Type, Market Category, Vehicle style, Popularity, Number of doors, Vehicle Size doesn't make any sense to me so I just dropped for this instance**.**

5. Renaming the column

In this instance, most of the column names are very confusing to read, so I just tweaked their column names. This is a good approach it improves the readability of the data set.

6. Dropping the duplicate rows

This is often a handy thing to do because a huge data set as in this case contains more than 10, 000 rows often have some duplicate data which might be disturbing, so here I remove all the duplicate value from the data-set. For example prior to removing I had 11914 rows of data but after removing the duplicates 10925 data meaning that I had 989 of duplicate data.

7. Dropping the missing or null values.

An outlier is a point or set of points that are different from other points. This is mostly similar to the previous step but in here all the missing values are detected and are dropped later. Now, this is not a good approach to do so, because many people just replace the missing values with the mean or the average of that column.

8. Detecting Outlier

Sometimes they can be very high or very low. It's often a good idea to detect and remove the outliers. Because outliers are one of the primary reasons for resulting in a less accurate model. Hence it's a good idea to remove them. The outlier detection and removing that I am going to perform is called IQR score technique. Often outliers can be seen with visualizations using a box plot.

Module 3: Train the Model

This stage is to form evaluation the models based on the input data. For our purpose of study, we are going to implement to train the model using four types of algorithm in machine learning to predict heart disease.

K-nearest neighbor

The K-nearest neighbors' algorithm is a supervised classification algorithm method. It classifies objects dependent on nearest neighbor. It is a type of instance-based learning. The calculation of distance of an attribute from its neighbors is measured using Euclidean distance. It uses a group of named points and uses them on how to mark another point. The data are clustered based on similarity among them, and is possible to fill the missing values of data using K-NN. Once the missing values are filled, various prediction techniques apply to the data set. It is possible to gain better accuracy by utilizing various combinations of these algorithms.

Random forest algorithm

Random forest algorithm is a supervised classification algorithmic technique. In this algorithm, several trees create a forest. Each individual tree in random forest lets out a class expectation and the class with most votes turns into a model's forecast. In the random forest classifier, the greater number of

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1480-1488 www.ijaem.net ISSN: 2395-5252

trees gives higher accuracy. The three common methodologies are:

- Forest RI (random input choice);
- Forest RC (random blend);
- Combination of forest RI and forest RC.

It is used for classification as well as regression task, but can do well with classification task, and can overcome missing values. Besides, being slow to obtain predictions as it requires large data sets and more trees, results are unaccountable.

Logistic regression algorithm

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

XGB Classifier

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, MPI) and can solve problems beyond billions of examples.

Module 4: Evaluation

This stage is to form evaluation the models based on the input data. For our purpose of study, we are going to implement the model using XGboost classifier algorithm in Machine Learning.

We will split the data set into test and train set. After splitting the data first have to train the data and test the data using XGboost classifier, Logistic regression, KNN, Random Forest in Machine learning techniques.

We will initialize the classifier model. We will set two hyperparameters namely max_depth and n_estimators. These are set on the lower side to reduce overfitting.

All right, we will now perform cross-validation on the train set to check the accuracy. The accuracy is slightly above the half mark. This can be further improved by hyperparameter tuning and grouping similar stocks together. The XGBoost python model tells us that the pct_change_40 is the most important feature of the others. Since we had mentioned that we need only 7 features, we received this list. Here's an interesting idea, why don't you increase the number and see how the other features stack up, when it comes to their f-score. You can also remove the unimportant features and then retest the model.

Module 5: Comparison of Existing model

This module includes comparison of existing system algorithm accuracy and our proposed model accuracy. Our aim is to improve the accuracy score. So, we have to change the model in our project.

In existing system have implemented on naïve bayes, Logistic Regression, Decision Tree, and KNN. These algorithms are used in existing system that given 85% accuracy. But Comparison of our project, we have done higher accuracy compared to existing system.

Module 6: Performance Analysis

The next stage is to predict the results using Classifier. The best method for the training and test data set is definitely given has the best results for Classification Accuracy and Recall for both validation cases. Now we forward this Random Forest classifier to next stage to predict the disease that may further lead to a cardiac arrest.

The results are compared using a confusion matrix. The consistency of a classification model can be well visualized with a tabular form also called Confusion Matrix (or "classifier") which shows its results over a set of known test data.

SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth N with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis are

i. Economical Feasibility
ii. Technical Feasibility
iii. Social Feasibility

Economic Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company

can pour into the research and development of the system is limited. The expenditures must be justified. Thus, the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel

TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub – assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS
1. UNIT TESTING
2. INTEGRATION TESTING
3. FUNCTIONAL TEST
4. SYSTEM TEST
5. WHITE BOX TESTING
6. BLACK BOX TESTING
7. UNIT TESTING:

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Integration test, System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

IV. CONCLUSION AND FUTURE WORK

In this project a reliable multi process method, Machine Learning concept to build a heart disease risk prediction system is proposed and Evaluate High accuracy had done comparatively Existing system. Heart disease has become the leading cause of death worldwide. The most effective way to reduce such deaths is to detect its symptoms earlier. The overall aim is to define various data mining techniques useful in effective heart disease prediction. Efficient and accurate prediction with a lesser number of attributes and tests is our goal. In this study, I consider only essential attributes. I applied four data mining classification techniques, K-nearest neighbour, XGboosting, Logistic regression and random forest. The data were pre-processed and then used in the model. random forest are the algorithms showing the best results in this model. I found the accuracy after implementing four algorithms to be lowest in K-nearest neighbors. We can further expand this research incorporating other data mining techniques such as time series, clustering and association rules. Considering the limitations of this study, there is a need to implement more complex and combination of models to get higher accuracy for early prediction of heart disease.

REFERENCES

[1]. Baitharu, Tapas Ranjan, and Subhendu Kumar Pani. "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset." Procedia Computer Science 85 (2016): 862-870.

[2]. Gavhane, Aditi, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. "Prediction of heart disease using machine learning." In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1275-1278. IEEE, 2018.

[3]. Singh, Smriti Mukesh, and Dinesh B. Hanchate. "Improving disease prediction by machine learning." Int J Res Eng Technol 5, no. 6 (2018): 1542-1548.

[4]. Taneja, Abhishek. "Heart disease prediction system using data mining techniques." Oriental Journal of Computer science and technology 6, no. 4 (2013): 457-466.

[5]. Dangare, Chaitrali S., and Sulabha S. Apte. "Improved study of heart disease prediction system using data mining classification

techniques." International Journal of Computer Applications 47, no. 10 (2012): 44-48.

[6]. Thomas, J., and R. Theresa Princy. "Human heart disease prediction system using data mining techniques." In 2016 international conference on circuit, power and computing technologies (ICCPCT), pp. 1-5. IEEE, 2016.

[7]. Kaur, Beant, and Williamjeet Singh. "Review on heart disease prediction system using data mining techniques." International journal on recent and innovation trends in computing and communication 2, no. 10 (2014): 3003-3008.

[8]. Meyer, Alexander, Dina Zverinski, Boris Pfahringer, Jörg Kempfert, Titus Kuehne, Simon H. Sündermann, Christof Stamm, Thomas Hofmann, Volkmar Falk, and Carsten Eickhoff. "Machine learning for real-time prediction of complications in critical care: a retrospective study." The Lancet Respiratory Medicine 6, no. 12 (2018): 905-914.

[9]. Rajkomar, Alvin, Michaela Hardt, Michael D. Howell, Greg Corrado, and Marshall H. Chin. "Ensuring fairness in machine learning to advance health equity." Annals of internal medicine 169, no. 12 (2018): 866-872.

[10]. Rajamhoana, S. P., C. Akalya Devi, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika. "Analysis of neural networks based heart disease prediction system." In 2018 11th International Conference on Human System Interaction (HSI), pp. 233-239. IEEE, 2018.

[11]. Ramalingam, V. V., Ayantan Dandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." International Journal of Engineering & Technology 7, no. 2.8 (2018): 684-687.

[12]. Kohli, Pahulpreet Singh, and Shriya Arora. "Application of machine learning in disease prediction." In 2018 4th International conference on computing communication and automation (ICCCA), pp. 1-4. IEEE, 2018.

[13]. Marimuthu, M., M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra. "A review on heart disease prediction using machine learning and data analytics approach." International Journal of Computer Applications 181, no. 18 (2018): 20-25.

[14]. Beyene, Chala, and Pooja Kamat. "Survey on prediction and analysis the occurrence of heart disease using data mining techniques." International Journal of Pure and Applied Mathematics 118, no. 8 (2018): 165-174.

[15]. Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." International Journal of Intelligent Engineering & Systems 12, no. 1 (2019): 242-252.