**CHAPTER-3**

**SYSTEM ANALYSIS**

**3. SYSTEM ANALYSIS**

**EXISTING SYSTEM**

In this system, the input details are obtained from the patient. Then from the user inputs using MI. techniques heart disease is Analysed Now, the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart donase patients collected from the UC laboratory is used to discover patterns with NN, DT, Support Vector machines SVM, and Naïve bayes The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 87% for measure, competing with the other existing methods

**DISADVANTAGES**

1. Predict of heart stroke disease results is not accurate.
2. Cannot handle enormous datasets for patient records.

Prediction of cardiovascular disease

results is not accurate

Algorithms: KNN, LOGISTIC REGRESSION

**KNN:**

K-Nearest Neighbor (KNN) It is a simple algorithm that stores all the available cases and classifies the new data or case based on similarity measure. K means the number of nearest neighbors which are voting class of new or testing data. To calculate the least distant k points, mathematical equations like Euclidean distance, Manhattan distance, etc are employed. It is also called Lazy Learner because it does not have a discriminative function from the training data. It memorizes the training data and there is no learning phase of the model.

K-nearest neighbors classifier implements learning based on the nearest neighbors of each query point, where 'k' is an integer value specified by the user. The value of 'k' depends on block of code.

Steps in K-Nearest Neighbour include:

• Pick a value for K.

• Calculate the distance of unknown case from all cases. . Find the K observation in the training data that are nearest to the measurement

of the unknown data point.

• Predict the response of the unknown data point using the most popular value from the KNN.

• Assign the new data points to that category for which the number of the neighbor is maximum.

**LOGISTIC REGRESSION:**

Logistic Regression- It is a method to predict a dependent variable given a set of independent variables such that dependent variable is categorical. The dependent vs independent variable is mapped to a sigmoid function. So, value of dependent variable is found either 0 or 1 for any value of independent variable. It gives the probability of occurrence of an event and several results like accuracy, ROC curve, F1 score, precision, recall, confusion matrix, etc can be obtained from logistic regression.

Logistic Regression is a classification algorithm for categorical values. In Logistic regression, we use one or more than one independent variables (X) to predict an outcome dependent variable (y), Logistic Regression tries to predict a categorical or discrete target field(label) instead of a numeric one, such as yes/no, true/false, successful/unsuccessful, etc.

In Logistic regression independent variable(X) should be continuous, if it is categorical, they are transformed into a numeric value. The logistic regression makes use of model called the logistic function i.e. sigmoid function to extract the output of a linear equation which is between 0 and 1. The logistic function is defined as:

Proposed System

The proposed technique focuses on predicting Heart stroke, It classifies the person risk level by implementing various machine learning algorithms like naive bayes and svm(support vector machine)

**Advantages:**

1. Increased accuracy for effective heart stroke disease diagnosis.
2. Handles roughest(enormous) amount of data using svm(support vector machine) and naïve bayes.
3. Reduce the time complexity of doctors.
4. Cost effective of patients.

Algorithms: NAIVE BAYES, SVM

**NAIVE BAYES:**

Naive Bayes This algorithm is based on the Bayes theorem. The assumption taken here is that all the input features or attributes are independent of each other and it provides the conditional probability as the output based on the input parameters. Bayes theorem calculates the posterior probability of an event (A) given some prior probability of event B represented by P(A/B) as shown in equation 1:

**SVM:**

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

• SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

• SVM algorithm can be used for Face detection, image classification, text categorization, etc.

Types of SVM

SVM can be of two types:

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Define number of features+1 as F and SVs+1 as SV

FOR each SV Read streamed data Store into array\_SVs [SV][F]

END FOR

Convert it to float

END FOR

Read streamed data

Convert it to float

Store into array ay [0] (b value)

FOR cach SV

Read streamed data

Convert it to float Store into array \_ay [SV]

END FOR

FOR each feature

Read streamed data

Convert it to float

Store into array test [F]

END FOR

FOR each feature

Clear array AC [F]

END FOR

FOR each SV

FOR each feature of the SV

array AC [F]+= array ay [SV] array\_SVs [SV][F] END FOR

END FOR

FOR each feature

Distance value+array\_AC [F] array\_test [F]

Distance value--b

END FOR

IF (Distance value th) THEN

ELSE

RETURN 1

RETURN-1

END IF

**Feasibility Study:**

An important outcome of preliminary investigation is the determination that the system request is feasible. This is possible only if it is feasible within limited resource and time. The different feasibilities that have to be analyzed are

* Operational Feasibility
* Economic Feasibility
* Technical Feasibility

**Operational Feasibility**

Operational Feasibility deals with the study of prospects of the system to be developed. This system operationally eliminates all the tensions of the Admin and helps him in effectively tracking the project progress. This kind of automation will surely reduce the time and energy, which previously consumed in manual work. Based on the study, the system is proved to be operationally feasible.

##### Economic Feasibility

Economic Feasibility or Cost-benefit is an assessment of the economic justification for a computer based project. As hardware was installed from the beginning & for lots of purposes thus the cost on project of hardware is low. Since the system is a network based, any number of employees connected to the LAN within that organization can use this tool from at anytime. The Virtual Private Network is to be developed using the existing resources of the organization. So the project is economical feasible.

##### Technical Feasibility

According to Roger S. Pressman, Technical Feasibility is the assessment of the technical resources of the organization. The organization needs IBM compatible machines with a graphical web browser connected to the Internet and Intranet. The system is developed for platform independent environment. Python, Django, HTML, SQL server and Web Logic Server are used to develop the system. The technical feasibility has been carried out. The system is technically feasible for development and can be developed with the existing facility.

**FUNCTIONAL REQUIREMENTS**

* Take the patient medical history and predict based on that data perform some operations like testing and training
* The result will be accurate by using the svm(support vector machine ) algorithm
* Performing analysis on accuracy of test data set
* This project performs preprocessing on the given

**NON FUNCTIONAL REQUIREMENTS**

* **Usability:** On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. This project can be used to cardiologists to analyze the heart diseases.
* **Efficiency:** Thus the experimental results show that the proposed features selection algorithm select features that are more effective and obtains high classification accuracy than the standard feature selection algorithms.
* **Performance:** To test the performances of the classifiers, various performance evaluation metrics are computed. All the experiments have been performed in a python environment using different machine learning libraries.