



Combining machine learning and deep learning for the prediction of heart disease

Dr.Nithya.R

Department of MCA
Dhanalakshmi Srinivasan College
of Engineering and Technology

Mr.Deva.P

Department of MCA
Dhanalakshmi Srinivasan College
of Engineering and Technology

Abstract: Heart is the main organ that pumps blood to the body and for proper functioning of body. Heart disease is a fatal human disease increasing globally in both developed and undeveloped countries and consequently, causes death. Normally, in this disease, the heart fails to supply a sufficient amount of blood to other parts of the body in order to accomplish their normal functionalities. It associates many risk factors in heart disease and a need of the time to get accurate, reliable, and sensible approaches to make an early diagnosis to achieve prompt management of the disease. Data mining is a commonly used technique for processing enormous data in the healthcare domain. Researchers apply several data mining and machine learning techniques to analyse huge complex medical data, helping healthcare professionals to predict heart disease. The proposed method is to build a machine learning model capable of classifying the person has heart disease or not. Different algorithms are compared and the best model is used for predicting the outcome.

I. INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.

Data Scientist:

Data scientists examine which questions need answering and where to find the related data. They have business acumen and analytical skills as well as the ability to mine, clean, and present data. Businesses use data scientists to source, manage, and analyze large amounts of unstructured data.

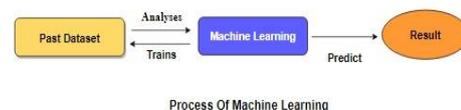
ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

MACHINE LEARNING

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a

simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms.



Preparing the Dataset:

This dataset contains 271 records of features extracted from Heart Patients, which were then classified into 2 classes:

- Happen
- Not Happen

Existing System:

The Acoustic analysis has the potential to be a relatively low cost and non-invasive diagnostic tool for point-of-care assessment and remote monitoring of LVAD recipients. Prior work on acoustic analysis in LVAD recipients has focused on analysis of LVAD sounds to detect pump thrombosis. The analysis of heart sounds in LVAD recipients has been limited by the dominant LVAD sounds present within their precordial sounds. We have developed a novel signal processing pipeline to mitigate LVAD-generated sounds within precordial sounds recorded from LVAD recipients, potentially rendering automated heart sound analysis in this population feasible.

Disadvantages:



1. Analysing the sound graphs and getting the sound it's a complex process.
2. Better classification can be done by proper comparison of algorithms.

Proposed System:

The proposed method is to build a machine learning model for classification of heart disease. The process carries from data collection where the past data related to heart disease are collected. Data mining is a commonly used technique for processing enormous data in the healthcare domain.

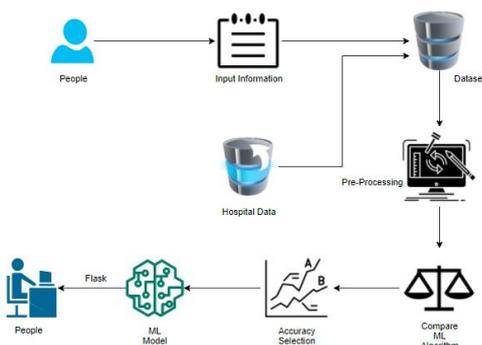
The heart disease if found before proper treatment can save lives. Machine learning is now applied and mostly used in health care where it reduces the manual effort and better model makes error less which leads in saving the life.

The data analysis is done on the dataset proper variable identification done that is both the dependent variables and independent variables are found. Then proper machine learning algorithm are applied on the dataset where the pattern of data is learnt. After applying different algorithms a better algorithm is used for the prediction of outcome.

Advantages:

1. The data used here will not as complex as sound graphs and analysis of data will be easy.
2. Performance metrics of different algorithm are compared and the better prediction is done.

System Architecture



MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data
 output : getting accuracy

Random Forest Classifier

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision *trees*, resulting in a *forest of trees*, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

The following are the basic steps involved in performing the random forest algorithm:

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.

In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output). The final value can be calculated by taking the average of all the values predicted by all the trees in forest. Or, in case of a classification problem, each tree in the forest predicts the category to which the new record belongs. Finally, the new record is assigned to the category that wins the majority vote.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data
 output : getting accuracy

Decision Tree Classifier

It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Assumptions of Decision tree:

- At the beginning, we consider the whole training set as the root.

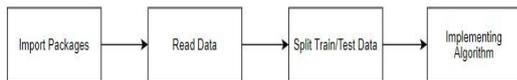


- Attributes are assumed to be categorical information gain, attributes are assumed to be continuous.
- On the basis of attribute values records are distributed recursively.
- We use statistical methods for ordering attributes as root or internal node.

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed.

This process is continued on the training set until meeting a termination condition. It is constructed in a top-down recursive divide-and-conquer manner. All the attributes should be categorical. Otherwise, they should be discretized in advance. Attributes in the top of the tree have more impact towards in the classification and they are identified using the information gain concept. A decision tree can be easily over-fitted generating too many branches and may reflect anomalies due to noise or outliers.

MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

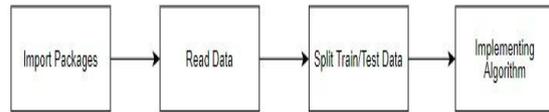
input : data
 output : getting accuracy

Naive Bayes algorithm:

- The Naive Bayes algorithm is an intuitive method that uses the probabilities of each attribute belonging to each class to make a prediction. It is the supervised learning approach you would come up with if you wanted to model a predictive modelling problem probabilistically.
- The probability of a class value given a value of an attribute is called the conditional probability. By multiplying the conditional probabilities together for each attribute for a given class value, we have a probability of a data instance belonging to that class. To make a prediction

we can calculate probabilities of the instance belonging to each class and select the class value with the highest probability.

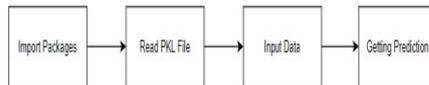
MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data
 output : getting accuracy

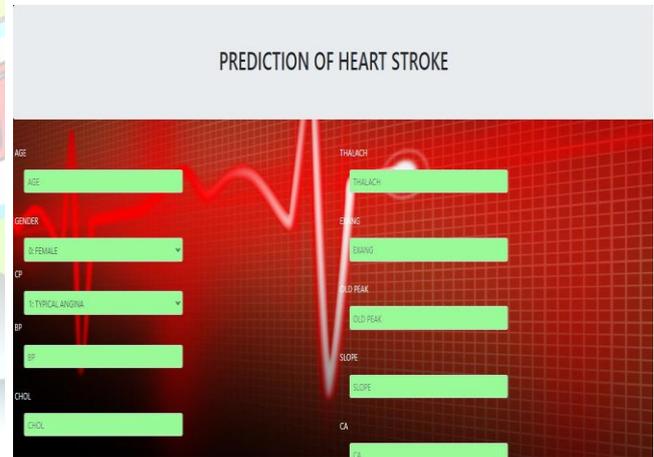
MODULE DIAGRAM



GIVEN INPUT EXPECTED OUTPUT

input : data values
 output : predicting output

SCREENSHOTS



CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out. This application can help to find the Prediction of Heart Attack.