**ABSTRACT**In today's fast paced world the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. researchers are focusing on designing smart systems to accurately diagnose them based on electronic health data, with the aid of machine learning algorithms. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression, Random Forest (RF) and KNN to predict and classify the patient with heart disease. Two classifiers such as support vector machine (SVM) with the best performance are selected for the classification in this method. Data preprocessing and feature selection steps were done before building the models. The models were evaluated based on accuracy, precision, recall, and F1-score. So, a quite significant amount of pressure has been lifted off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease. It is implemented in the.ipynb format.

**Objective of the Project:**The main objective of this research is to develop a heart prediction

system, the system can discover and extract hidden knowledge associated

with diseases from heart data set.

This system aims to exploit machine learning techniques on medical data

set to assist in the prediction of the heart disease.

Reduce the cost of medical tests.

To help avoid human biases.  
  
**Description:**

This project can save a lot of time for people dealing with misinformation and misguidance.

Technology Stack:

* Programming Language: Python
* Libraries: Seaborn, NumPy, Pandas, Sklearn, Matplotlib.
* Machine learning Models: Logistic Regression, Decision tree, Support Vector Machine, Random Forest, Linear regression, Kth Nearest Neighbors

**INTRODUCTION**  
  
Machine Learning is a way of Manipulating and extraction of implicit, previously known/ unknown, and potential useful information about data. Machine Learning is a very vast and diverse field, and its scope and implementation is increasing day by day. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset. We can use that knowledge in our project of HDPS as it will help a lot of people. Cardiovascular diseases are very common these days, they describe a range of conditions that could affect your heart. World health organization estimates that 17.9 million global deaths from (cardiovascular diseases) CVDs [2]. It is the primary reason of deaths in adults. Our project can help predict the number of people who are likely to be diagnosed with a heart disease by help of their medical history [6]. It recognizes who all are having any symptoms of heart disease such as chest pain or high blood pressure and can help in diagnosing disease with less medical tests and effective treatments, so that they can be cured accordingly. This project focuses on mainly three data mining techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 90.64%, for which is better than previous system where only one data mining technique is used. So, using more data mining techniques increased the HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression. By using this dataset, we predict whether the patient can have a heart disease or not. To predict this, we use 14 medical attributes of a patient and classify him if the patient is likely to have a heart disease. These medical attributes are trained under three algorithms: Logistic regression, KNN and Random Forest Classifier. Most efficient of these algorithms is KNN which gives us the accuracy of 80.70%. And finally we classify patients that are at risk of getting a heart disease or not and also this method is totally cost efficient.

**Methodology:**

This study aims to predict the probability of heart disease through computerized heart disease prediction, which can be beneficial for medical professionals and patients. To achieve this objective, we employed various machine learning algorithms on a dataset and present the results in this study report. To enhance the methodology, we plan to data cleaning (removing outliers checking distribution etc.), applying data transformation to sum input variables, Feature Selection and Reduction, eliminate irrelevant information, splitting data in 80:20 ratio randomly, training model on different algorithms, testing model on different algorithms, using various evaluation parameters comparing results.

**Literature review:**

Numbers of studies have been done that focus on the diagnosis of heart disease because most of the deaths occurred due to heart failure. Using the inbuilt implementation algorithm using uses some techniques of Neural Network and were just accurately able to predict whether the test patient is suffering from the given disease or not. Researchers used machine learning techniques for the prediction of heart disease some techniques are SVM support vector machine, naive Bayes, neural network, decision tree, and regression classifiers. Research is done by S. Prakash et al. in 2017 on heart disease prediction in which they compare two methods Optimality Criterion Feature Selection and rough set feature selection on information entropy. K. Polara j et al. compare different algorithms models for the prediction of heart diseases [19] and the result was that multiple linear Regression is better for the prediction of the risk of cardiovascular disease.

**References**

Sahoo, P. & Jeripothula, P. (2020) Heart Failure Prediction Using Machine Learning Techniques. SSRN Electronic Journal.

Chicco, D. & Jurman, G. (2020) Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision

Fitriyani N. L., Syafrudin M., Alfian G., Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system.

Zhenya Q., Zhang Z. A hybrid cost-sensitive ensemble for heart disease prediction. BMC Medical Informatics and Decision Making

C. Beyene and P. Kamat, “Survey on prediction and analysis the occurrence of heart disease using data mining techniques,” *International Journal of Pure and Applied Mathematics*

 Yuan X., Chen J., Zhang K., Wu Y., Yang T. A stable ai-based binary and multiple class heart disease prediction model for IoMT

V. V. Ramalingam, A. Dandapath, and M. K. Raja, “Heart disease prediction using machine learning techniques

**Business/Revenue Model**

**Subscription/Membership Model:** Users could pay a monthly or yearly subscription fee to access the heart disease prediction service. This could include unlimited access to the chatbot for consultations and predictions.

**Enterprise Solutions**: Offer customized versions of the chatbot for healthcare providers, insurance companies, or corporate wellness programs. These entities could pay for bulk access or licensing fees.

**Pay-per-Use Model:** Users could pay per prediction or consultation with the chatbot. This model could be suitable for those who don't require frequent predictions but still want access to the service on an as-needed basis

**API Keys Integration :** Hospitals would be provided with API keys upon subscription or purchase. These keys would allow them to integrate the chatbot seamlessly into their existing patient management systems, electronic health records to predict the heart disease.