

# Dr. B. R. Ambedkar National Institute Of Technology Jalandhar, Punjab



## Final Year Project Report

### Project - Heart Disease Prediction

Session: 2017-2021

#### **Project Mentor:**

Dr. Prashant Kumar

Assistant Professor

[Department of CSE]

#### **Submitted By:**

Gyanesh Kumar (17103033)

Sonal Keshri (17103083)

# Table Of Contents

1.	ABSTRACT	3
2.	PROBLEM STATEMENT	4-5
3.	FEASIBILITY STUDY REPORT	6-9
3.1.	Title	
3.2.	Abstract	
3.3.	Keywords	
3.4.	Description	
3.5.	Scope	
3.6.	Technical Feasibility	
3.7.	Risk Analysis	
3.8.	Work Plan Schedule	
3.9.	Deliverables	
4.	PRODUCT RESEARCH	10-14
4.1.	Gathering Data	
4.2.	Data Pre-Processing	
4.3.	Training & Testing the ML Model	
4.4.	Improvising, Hypertuning the Model	
4.5.	Conclusion	
5.	TOOLS & TECHNOLOGIES	14-15
5.1.	Frameworks & Libraries	
5.2.	Programming Languages	
5.3.	Architecture Patterns	
5.4.	Databases	
5.5.	Tools / Softwares	
6.	SOFTWARE WORKFLOW	16
7.	WORKING PROJECT DESCRIPTION	17-29
8.	REFERENCES	30

# 1. Abstract

Among all fatal diseases, heart attacks diseases are considered as the most prevalent. The number of people suffering from heart disease is on the rise (health topics, 2010). The report from the World Health Organization shows us that a large number of people die every year due to heart disease all over the world. Heart disease is also stated as one of the greatest killers in Africa.

One of the most important uses of data analysis is that it helps in keeping human bias away from medical conclusions with the help of proper statistical treatment. By use of data mining for exploratory analysis because of nontrivial information in large volumes of data. The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions for providing appropriate results and making effective decisions on data, some data mining techniques are used to better the experience and conclusion that have been given.

Heart predictor system will use the data mining knowledge to give a user-oriented approach to new and hidden patterns in the data. The knowledge which is implemented can be used by the healthcare experts to get better quality of service and to reduce the extent of adverse medicine effect.

## 2. Problem Statement

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expenses.

The overall objective of our work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease.

Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

### **Objective:**

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 a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

- Provides a new approach to concealed patterns in the data.
- Helps avoid human biases.
- To research & implement the most accurate classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests

The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

# 3. Feasibility Study Report

## 3.1 Title

Heart Disease Prediction System

## 3.2 Abstract

Heart predictor system will use the data mining knowledge to give a user-oriented approach to new and hidden patterns in the data. The knowledge which is implemented can be used by the healthcare experts to get better quality of service and to reduce the extent of adverse medicine effect.

## 3.3 Keywords

- **Machine Learning:**

Machine learning (ML) is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed.

- **Decision Tree:**

Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks.

- **Random Forest:**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting

the class that is the mode of the classes or mean/average prediction of the individual trees.

- **Support Vector Machines (SVMs):**

These are supervised machine learning algorithms used to find a hyperplane in an N-dimensional space (N – the number of features) that distinctly classifies the data points.

- **K Nearest Neighbor:**

The nearest neighbors algorithms are a simple, easy-to-implement set of supervised machine learning algorithms that can be used to solve both classification and regression problems.

### **3.4 Description 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 a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

- Provides a new approach to concealed patterns in the data.
- Helps avoid human biases.
- To research & implement the most accurate classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests

### **3.5 Project Scope**

The scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

This suggestion is promising as data modeling and analysis tools, e.g., data mining has the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

Using this model, we will be able to predict the presence/absence of given disease based on the parameters provided. Results obtained from different models are compared side by side based on the speed and accuracy of computation.

The results are not guaranteed to be 100% accurate as these are totally based on the dataset used to train the model.

### **3.6 Technical Feasibility**

- Period of Completion: At least 7-8 months.
- Processor: Intel® Core™ i5 processor or better.
- RAM and Memory: 8-16 GB of DDR3 RAM or better
- Python versions: 3.X
- Software / Tools Required: Jupyter Notebook (IPython), Android Studio, NodeJS
- Programming Languages: PYTHON, KOTLIN/JAVA, TYPESCRIPT

### **3.7 Risk Analysis**

- **Technical:**  
Android & Python Application Supporting Devices as mobiles, tablets, desktops, laptops and Machine Learning



- **Economic Feasibility:**

This project doesn't require much cost in development, it only requires cost for servers

- Internet connectivity is required for the system

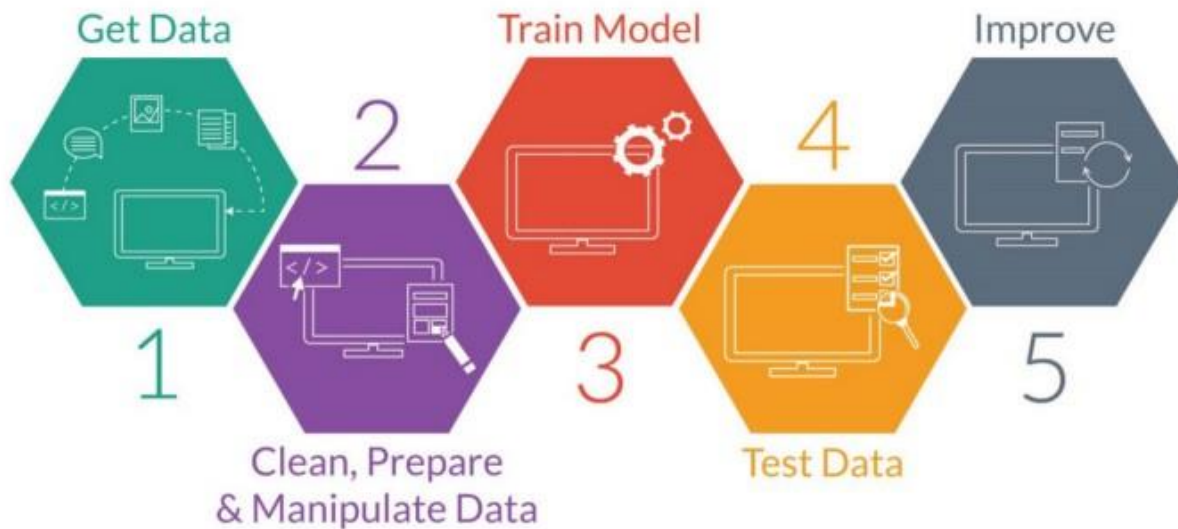
### **3.8 Work Plan/Schedule**

- |                                       |         |
|---------------------------------------|---------|
| ● Finalization of Project Idea        | 3 Weeks |
| ● Literature survey                   | 4 Weeks |
| ● Feasibility Study and Documentation | 2 Weeks |
| ● Requirement Gathering and Analysis  | 2 Weeks |
| ● Designing task & Modularizing       | 2 Weeks |
| ● Gathering Data & Processing         | 2 Weeks |
| ● Coding and Implementation           | 8 Weeks |
| ● Testing Task                        | 2 Weeks |
| ● Deployment                          | 1 Week  |

### **3.9 Deliverables**

- Product Research Report
- Android App
- Admin Panel
- ML Model
- Backend

## 4. PRODUCT RESEARCH



### 4.1 Gathering Data

Our Predictor (Y, Positive or Negative diagnosis of Heart Disease) is determined by 13 features (X)

The dataset has been taken from UCI Database. There are 14 columns in the dataset, where the `patient_id` column is a unique and random identifier. The remaining 13 features are described below.

1. `age` - age in years
2. `sex` - (1 = male; 0 = female)
3. `cp` - chest pain type
  - 0: Typical angina: chest pain related decrease blood supply to the heart
  - 1: Atypical angina: chest pain not related to heart
  - 2: Non-angina pain: typically esophageal spasms (non heart related)
  - 3: Asymptomatic: chest pain not showing signs of disease

4. **trestbps** - resting blood pressure (in mm Hg on admission to the hospital)  
anything above 130-140 is typically cause for concern
5. **chol** - serum cholesterol in mg/dl
  - o serum = LDL + HDL + .2 \* triglycerides
  - o above 200 is cause for concern
6. **fbs** - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
  - o '>126' mg/dL signals diabetes
7. **restecg** - resting electrocardiographic results
  - o 0: Nothing to note
  - o 1: ST-T Wave abnormality
    - can range from mild symptoms to severe problems
    - signals non-normal heart beat
  - o 2: Possible or definite left ventricular hypertrophy
    - Enlarged heart's main pumping chamber
8. **thalach** - maximum heart rate achieved
9. **exang** - exercise induced angina (1 = yes; 0 = no)
10. **oldpeak** - ST depression induced by exercise relative to rest looks at stress of heart during exercise unhealthy heart will stress more
11. **slope** - the slope of the peak exercise ST segment
  - o 0: Upsloping: better heart rate with exercise (uncommon)
  - o 1: Flat Sloping: minimal change (typical healthy heart)
  - o 2: Downsloping: signs of unhealthy heart
12. **ca** - number of major vessels (0-3) colored by fluoroscopy
  - o colored vessel means the doctor can see the blood passing through
  - o the more blood movement the better (no clots)
13. **thal** - thallium stress result
  - o 1,3: normal
  - o 6: fixed defect: used to be defect but ok now
  - o 7: reversible defect: no proper blood movement when exercising
14. **target** - have disease or not (1=yes, 0=no) (= the predicted attribute)

## 4.2 Data Pre-Processing

Datasets in a perfect world is a perfectly curated group of observations with no missing values or anomalies. However, this is not true. Real world data comes in all shapes and sizes. It can be messy, which means it needs to be clean and wrangled. Data cleaning is a necessary part in data science problems.

Machine learning models learn from data. It is crucial, however, that the data you feed them is specifically preprocessed and refined for the problem you want to solve. This includes data cleaning, preprocessing, feature engineering, and so on.

Let's say we have a column Gender, with values 1 for Male and 0 for Female. It needs to be converted into two columns with the value 1 where the column would be true and 0 where it will be false.

To get this done, we use the `get_dummies()` method from pandas. Next, we need to scale the dataset for which we will use the StandardScaler. The `fit_transform()` method of the scaler scales the data and we update the columns.

## 4.3 Training & Testing the ML Model

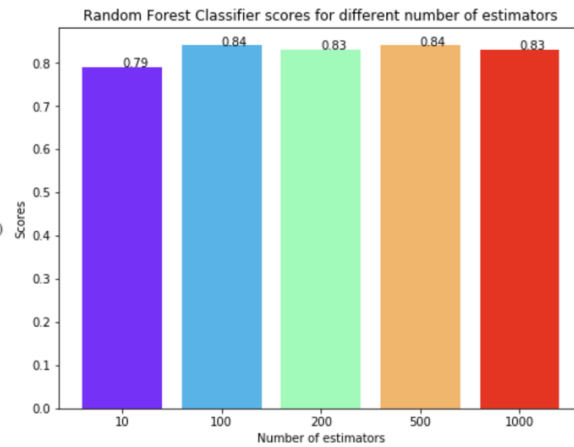
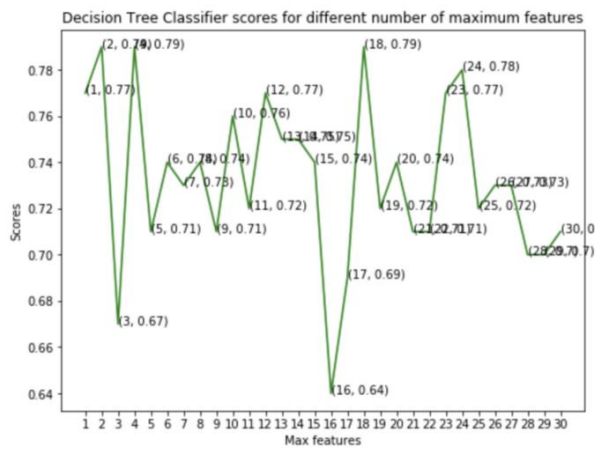
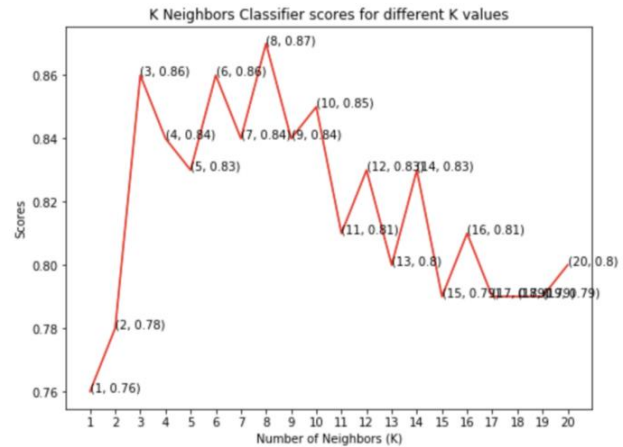
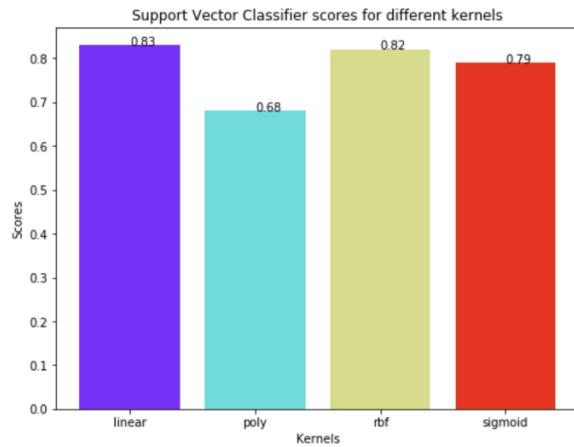
We trained various Classification Models on the Training set & see which yields the highest accuracy.

In this project, we took 4 algorithms and varied their various parameters and hypertuned them for the final model. We split the dataset into 67% training data and 33% testing data.

## 4.4 Improvising & Hypertuning the Model

The project involved analysis of the heart disease patient dataset with proper data processing. Then, 4 models were trained and tested with maximum scores as follows:

- K Neighbors Classifier: 87%
- Support Vector Classifier: 83%
- Decision Tree Classifier: 79%
- Random Forest Classifier: 84%



## 4.5 Conclusion

On studying the findings from the testing of our various ML Model, we found the following:

**K Nearest Neighbor Classifier scored the best score of 87% accuracy with 8 neighbors.**

# 5.Tools & Technologies

## 5.1 Framework & Libraries

- **NestJs** - Nest (NestJS) is a framework for building efficient, scalable NodeJS server-side applications. It uses progressive JavaScript, is built with and fully supports TypeScript(yet still enables developers to code in pure JavaScript) and combines elements of OOPd (Object Oriented Programming), FP (Functional Programming), and FRP (Functional Reactive Programming).
- **Flask** - Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions
- **ML Libraries** - sklearn, pandas

## **5.2 Programming Languages**

- Python
- Javascript
- Typescript
- Kotlin
- XML/JSON

## **5.3 Architecture Patterns Used**

- MVVM (Model-View-ViewModel) for Android App
- MVC(Model-View-Controller) for web backend

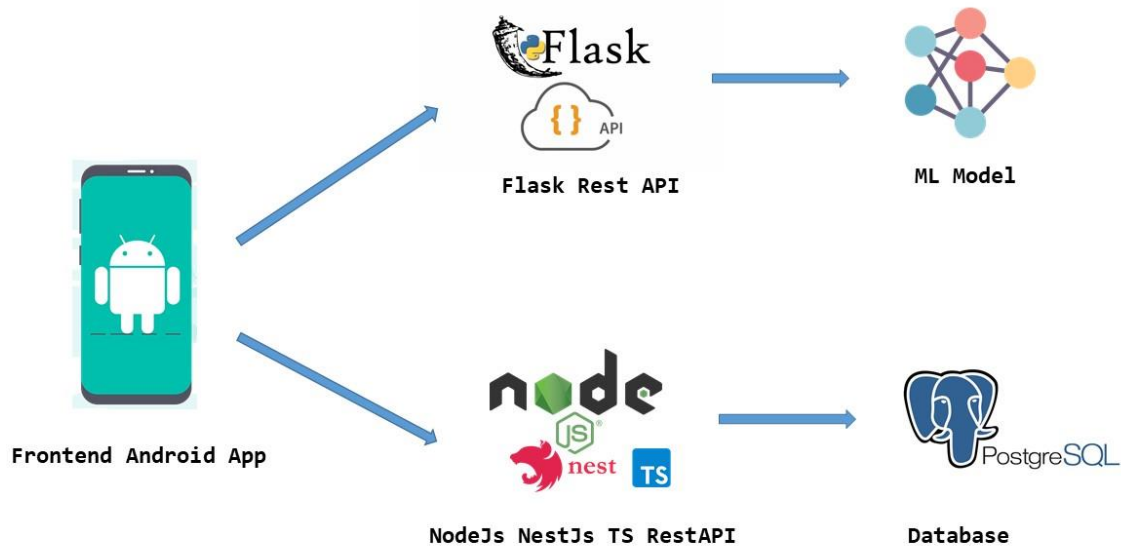
## **5.4 DataBase Used**

- Sqlite - Development Environment
- PostgreSQL - Production Environment

## **5.5 Tools**

- Jupyter Notebook
- VsCode
- Android Studio
- WebStorm

## 6. Software Workflow



- Android App Calls APIs of Python Flask Rest API for predicting the results
- Flask API internally uses our ML model to predict result
- For other Rest APIs Android App calls the NodeJs Backend
- Data on App can be added by accessing the admin panel of NodeJs Admin
- NodeJs Backend is connected to PostgreSQL Database



# 7. Working Project Description

## 7.1 Android App


### 7.1.1 Supported Languages



- To increase the accessibility of our app we have given the support for English as well as Hindi.
- During the first usage of the app or from clicking the globe icon we can launch the **“Choose Language”** screen and change the preferred language.



#### Choose Your Preferred Language

Please select your language

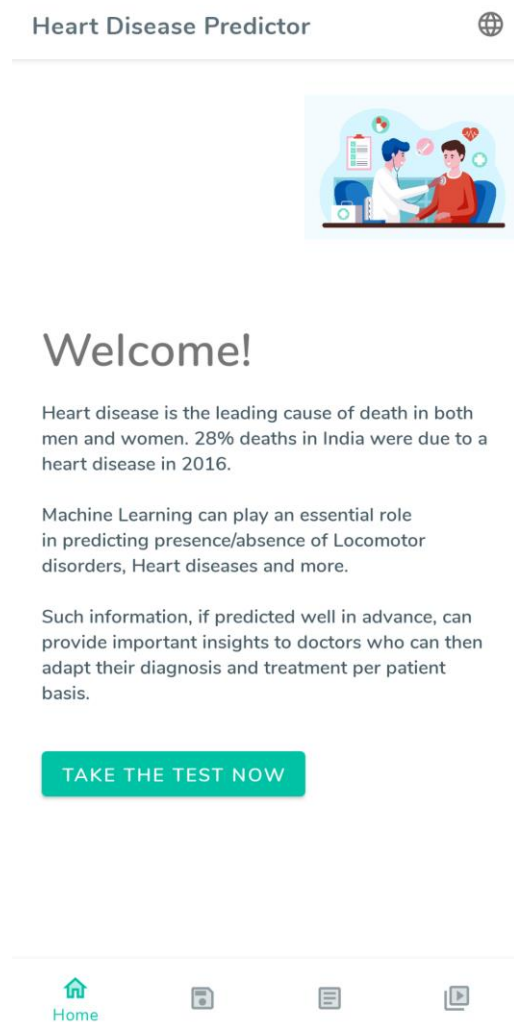
  
हिन्दी  
India

  
English  
United States 

CONTINUE


### 7.1.2 Home Screen

- On the home screen, a brief is shown about the app and the motivation behind it.
- Click the **“TAKE THE TEST NOW”** button launches the form screen for heart disease prediction



### 7.1.3 Prediction Test Screen

- Clicking the **“TAKE THE TEST NOW”** button launches the form screen for heart disease prediction
- After filling the various details of the person, we can click on the predict button to get the prediction result

 Heart Disease Test

Exercise Induced Angina

▼

Slope of Peak Exercise ST Segment

▼

No of Major Vessels colored (Flouroso...

▼

Thalium Stress Result

▼

Resting Blood Pressure (in mm Hg)



Serum Cholestoral (in mg/dl)

Maximum Heart Rate Achieved

ST Depression Induced by Exercise

PREDICT









### 7.1.4 Prediction Response Screen

Heart Disease Test	Heart Disease Test
Slope of Peak Exercise ST Segment Flatsloping: minimal change (typical healthy heart)	Slope of Peak Exercise ST Segment Flatsloping: minimal change (typical healthy heart)
No of Major Vessels colored (Flourosopy) 0	No of Major Vessels colored (Flourosopy) 3
 <p>Patient has no risk of a heart disease</p> <p>OK</p>	 <p>Patient has a significant risk of a heart disease!</p> <p>OK</p>
Maximum Heart Rate Achieved 0	Maximum Heart Rate Achieved 140
ST Depression Induced by Exercise 0	ST Depression Induced by Exercise 125
PREDICT	PREDICT

### 7.1.5 Report History Screen

We can see the report of previously generated reports by going to the reports section by clicking on the reports icon on the bottom navigation bar.

Reports	रिपोर्ट
Test Taken On 01:58 PM, 18 May 2021 Risk of Heart Disease Yes	परीक्षण दिवस 01:58 pm, 18 मई 2021 हृदय रोग का खतरा हाँ
Test Taken On 03:58 PM, 18 May 2021 Risk of Heart Disease Yes	परीक्षण दिवस 03:58 pm, 18 मई 2021 हृदय रोग का खतरा हाँ
Test Taken On 03:59 PM, 18 May 2021 Risk of Heart Disease Yes	परीक्षण दिवस 03:59 pm, 18 मई 2021 हृदय रोग का खतरा हाँ
Test Taken On 03:59 PM, 18 May 2021 Risk of Heart Disease No	परीक्षण दिवस 03:59 pm, 18 मई 2021 हृदय रोग का खतरा नहीं



### 7.1.6 Helpful Articles

If the admin wants to show users some important information / blog / article like symptoms of Heart Disease, It can be added from the admin panel and is shown on this screen.

#### Helpful Articles



**Don't Ignore These 15 Symptoms of Heart Disease**

**Heart disease - Symptoms and Causes**

Overview Heart disease describes a range of conditions that affect your heart. Heart diseases include: Blood vessel disease, such as c...

#### उपयोगी लेख



**हृदय रोग के लक्षण**

**हृदय रोग के लक्षण और उपचार**

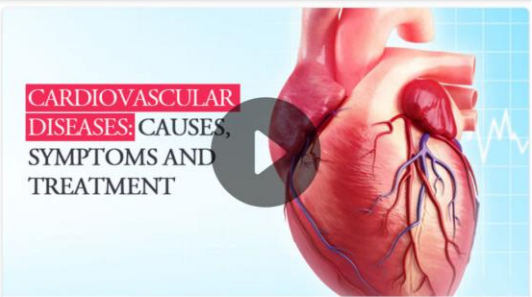
18/05/2021 Symptoms Of Heart Attack - येलण बताएंगेआप हो सकतेहहाट अटैक के शकार, इस तरह सेकरबचाव | Patrika News <https://www.patrika.com/health-n...>



### 7.1.7 Helpful Videos


If the admin wants to show some youtube video like symptoms of Heart Disease, It can be added from the admin panel and is shown on this screen.

#### Helpful Videos




Heart Disease - Causes, Symptoms and Treatment Options

#### उपयोगी वीडियो



हृदय रोग के कारण लक्षण और असरदार घरेलू उपचार



## 7.1.7 Article Detail Screen

← Heart disease - Symptoms and Causes



### Heart disease - Symptoms and Causes

Overview Heart disease describes a range of conditions that affect your heart. Heart diseases include: Blood vessel disease, such as coronary artery disease Heart rhythm problems (arrhythmias) Heart defects you're born with (congenital heart defects) Heart valve disease Disease of the heart muscle Heart infection Many forms of heart disease can be prevented or treated with healthy lifestyle choices. Products & Services Mayo Clinic Healthy Living Program Symptoms Heart disease symptoms depend on what type of heart disease you have. Symptoms of heart

← हृदय रोग के लक्षण और उपचार

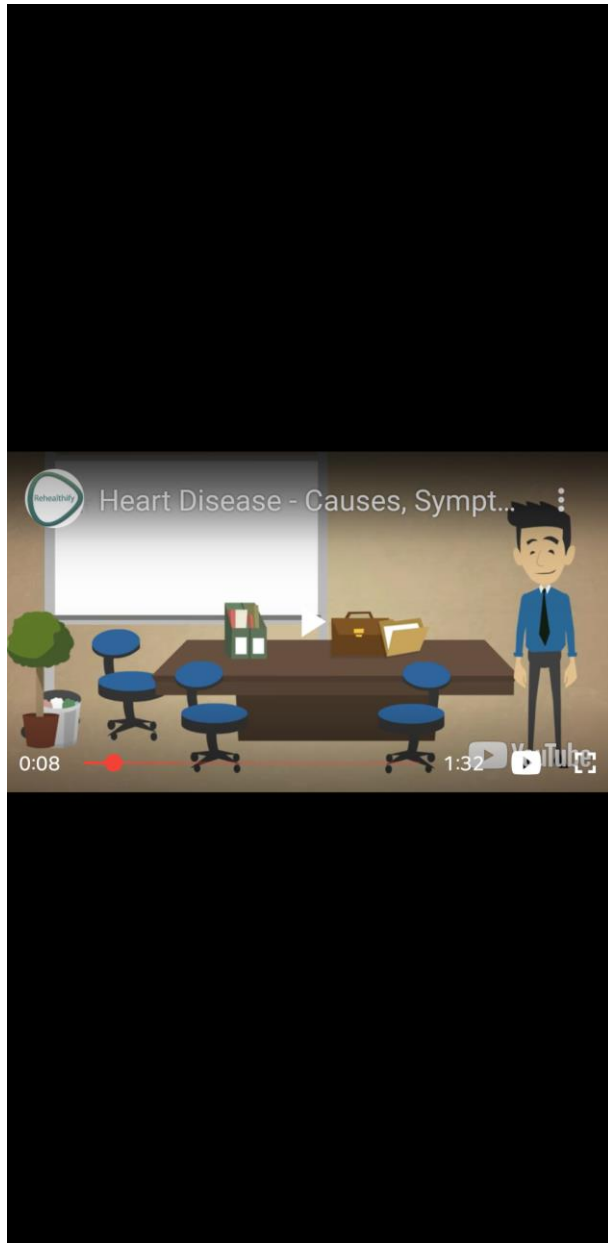


### हृदय रोग के लक्षण और उपचार

18/05/2021 Symptoms Of Heart Attack - ये लक्षण बताएंगे आप हो सकते हैं हार्ट अटैक के शिकार, इस तरह सेकरबचाव | Patrika News <https://www.patrika.com/health-news/symptoms-of-heart-attack-2341940/> 1/5 हर साल WHO वडहार्टडेके ज़रये लोग मरते हैं हृदय रोग के त जागकता फैलाने का कायकरता है। यिक दयाघात के लक्षण को जानना हर किसी के लिए जरूरी है। By: balram singh heart attack indications दिल हमारे शरीर का एक महत्वपूर्ण हिस्सा है। ये तो हम सभी जानते हैं। परंतु या हम अपने दिल के महत्वपूर्णता को जानते हुए भी इसका खयाल रखते हैं? शायद नहीं। “आज के आधुनिक लाइफटाइल और अनियमित आहार के कारण 30 से 40 साल के उम्र के लोग को दिल के रोग होने लग रहे हैं।

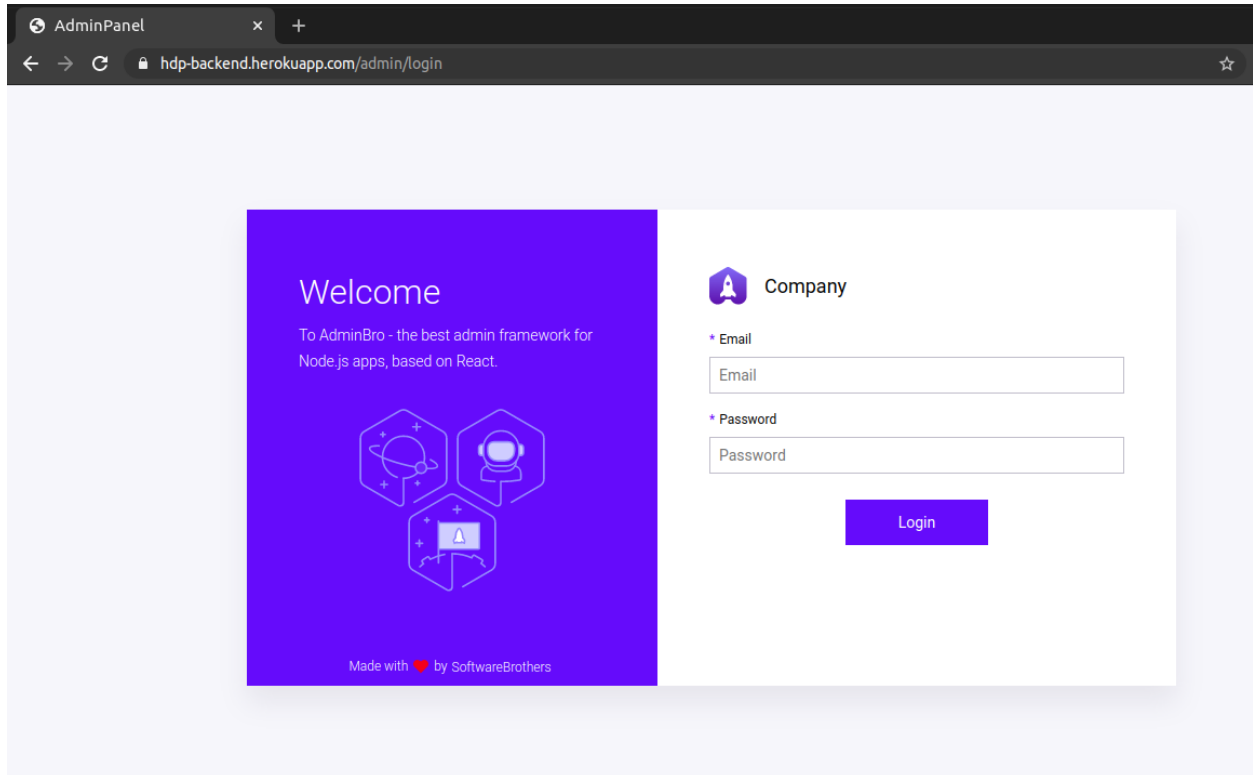


### 7.1.7 Video Detail Screen

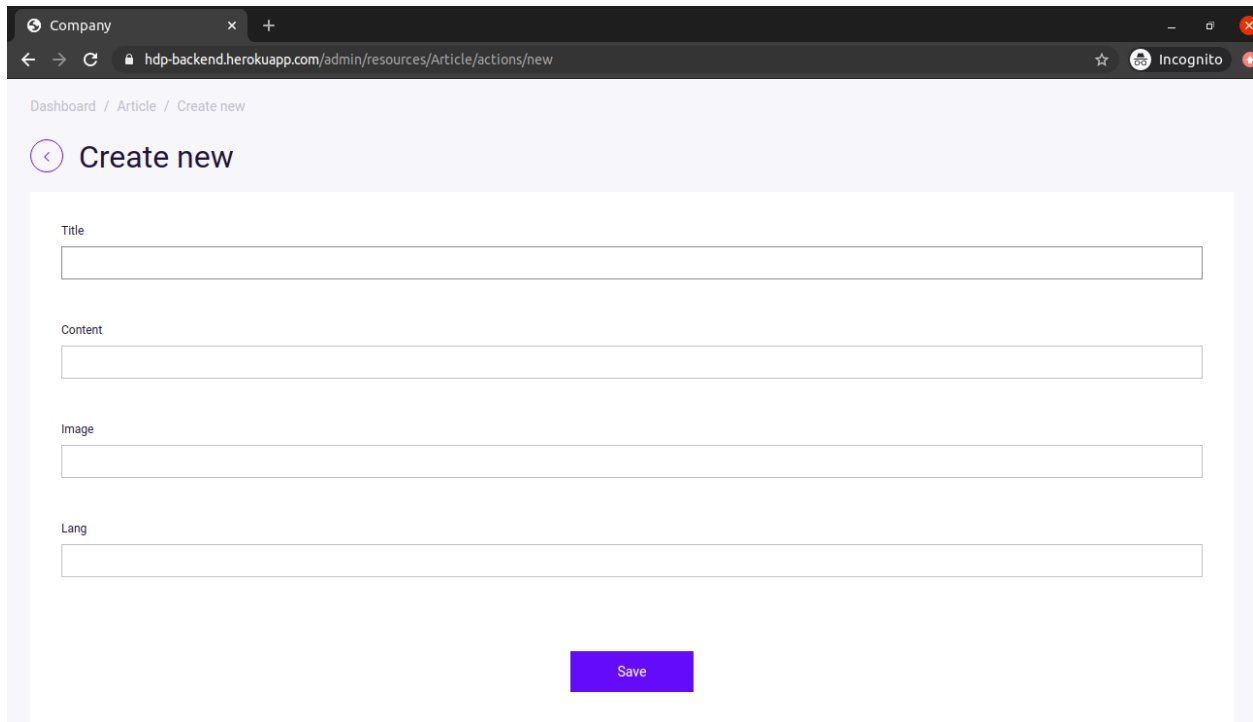
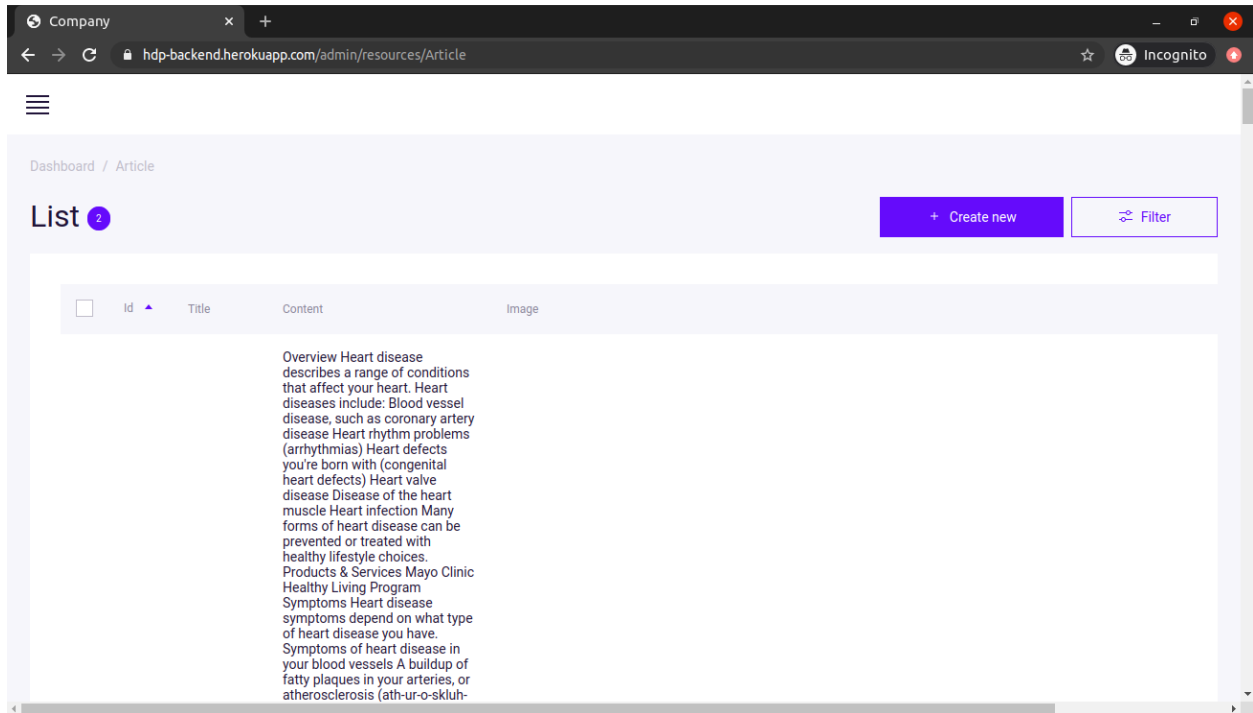


## 7.2 Admin Panel

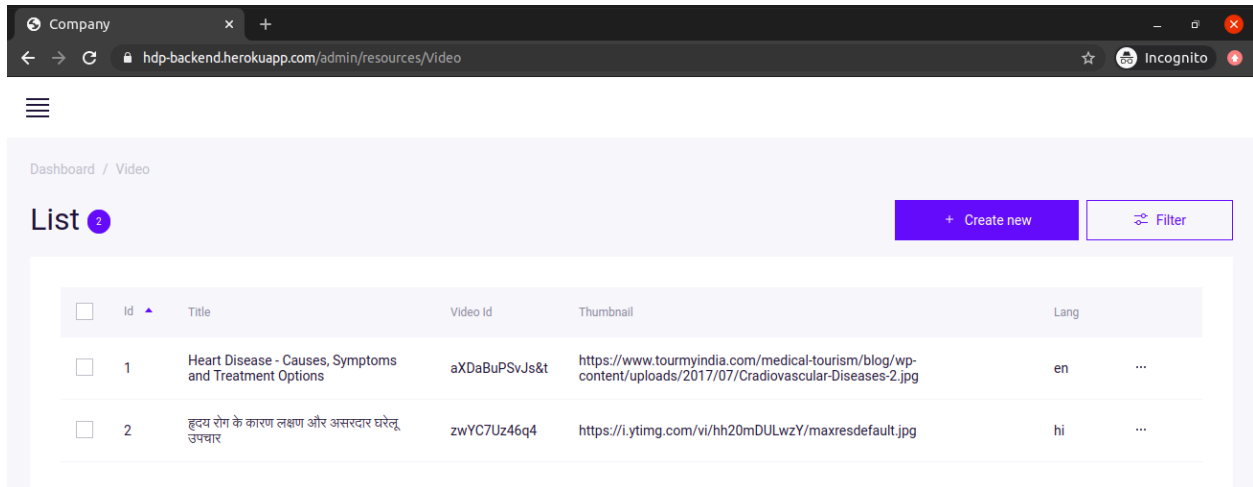
### 7.2.1 Login Screen



## 7.2.2 Add Article

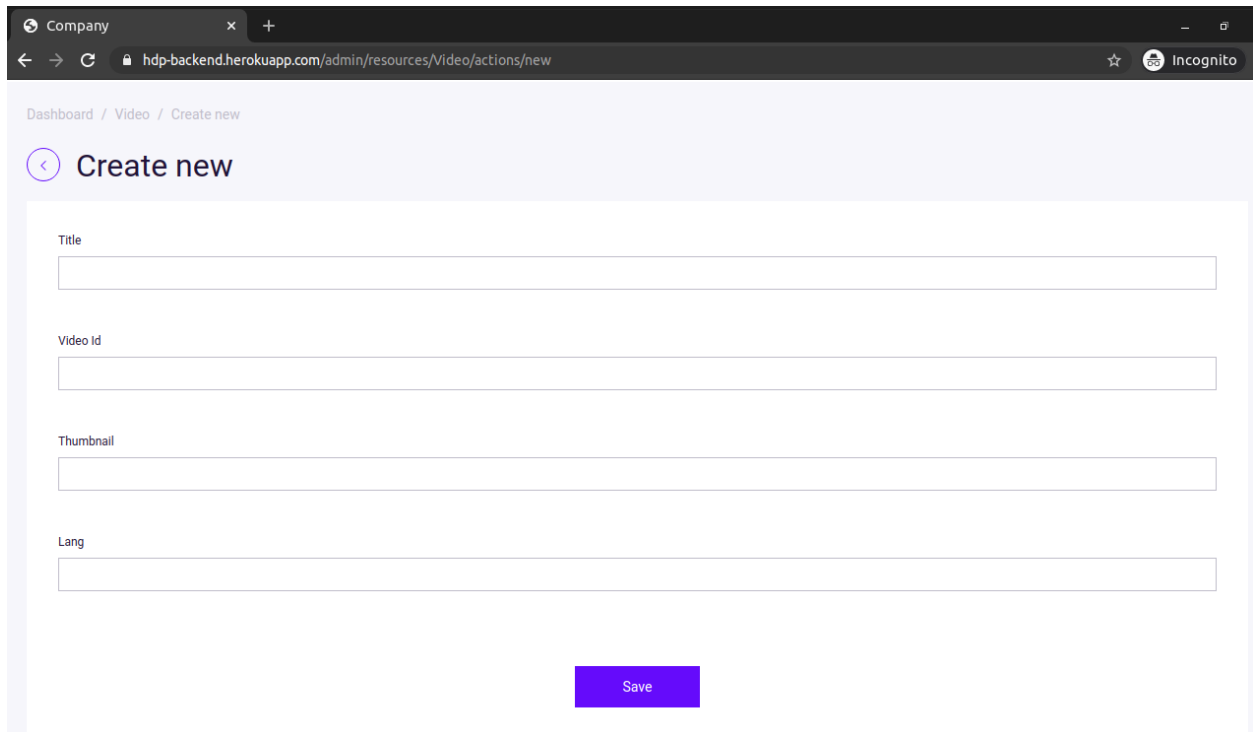


## 7.2.3 Add Video



The screenshot shows a web browser window with the URL `hdp-backend.herokuapp.com/admin/resources/Video`. The page title is "Company". The breadcrumb navigation is "Dashboard / Video". The main heading is "List" with a notification badge "2". There are two buttons: "+ Create new" and "Filter". Below is a table with columns: Id, Title, Video Id, Thumbnail, and Lang. The table contains two rows of data.

<input type="checkbox"/>	Id	Title	Video Id	Thumbnail	Lang	
<input type="checkbox"/>	1	Heart Disease - Causes, Symptoms and Treatment Options	aXDaBuPSvJs&t	<a href="https://www.tourmyindia.com/medical-tourism/blog/wp-content/uploads/2017/07/Cradiovascular-Diseases-2.jpg">https://www.tourmyindia.com/medical-tourism/blog/wp-content/uploads/2017/07/Cradiovascular-Diseases-2.jpg</a>	en	...
<input type="checkbox"/>	2	हृदय रोग के कारण लक्षण और असरदार घरेलू उपचार	zwYC7Uz46q4	<a href="https://i.ytimg.com/vi/hh20mDULwzY/maxresdefault.jpg">https://i.ytimg.com/vi/hh20mDULwzY/maxresdefault.jpg</a>	hi	...



The screenshot shows a web browser window with the URL `hdp-backend.herokuapp.com/admin/resources/Video/actions/new`. The page title is "Company". The breadcrumb navigation is "Dashboard / Video / Create new". The main heading is "Create new" with a back arrow. Below are four input fields labeled "Title", "Video Id", "Thumbnail", and "Lang". At the bottom is a "Save" button.

Title

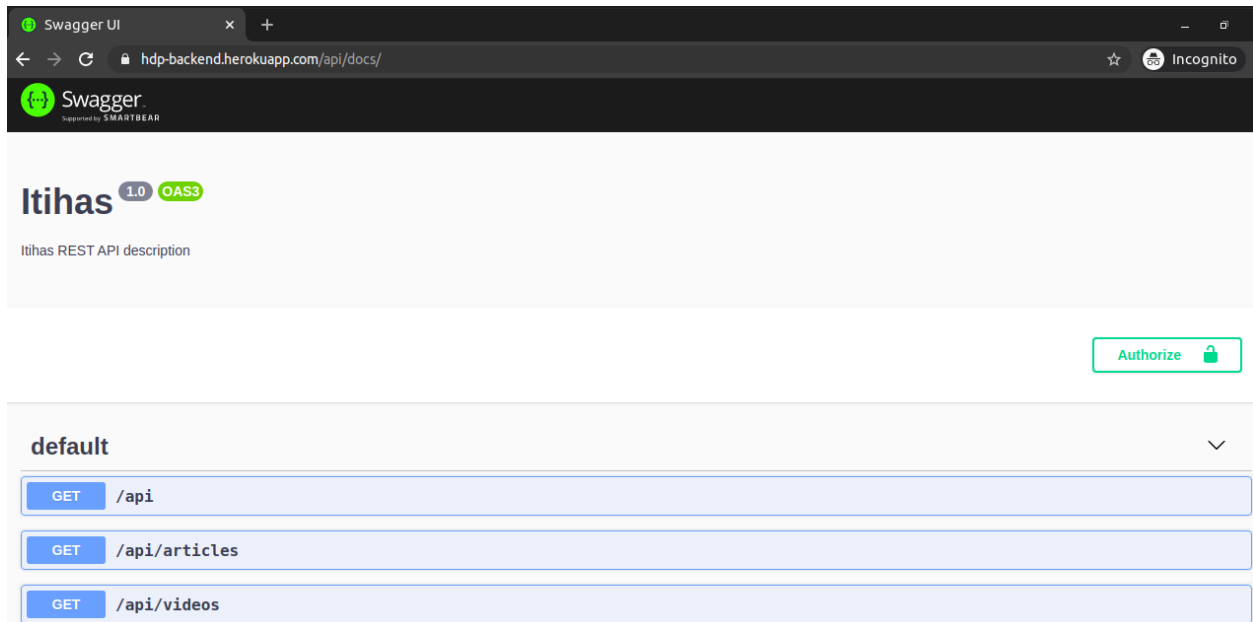
Video Id

Thumbnail

Lang

Save

## 7.2.4 Swagger API Docs



## 7.3 ML Model

We added two APIs to train and untrain the model on the fly as well

```
hdp-ml-predictor.herokuapp.com/train
```

```
{"message": "Model Trained"}
```

```
hdp-ml-predictor.herokuapp.com/untrain
```

```
{"message": "Model Untrained"}
```

## 8. References

- Heart Disease UCI Database - <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- Dheeru, D. and Karra Taniskidou, E. (2017). UCI machine learning repository
- Machine Learning Classification Models & Heart Diseases <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>
- NestJs - <https://docs.nestjs.com/>
- Flask - <https://flask.palletsprojects.com/en/2.0.x/>