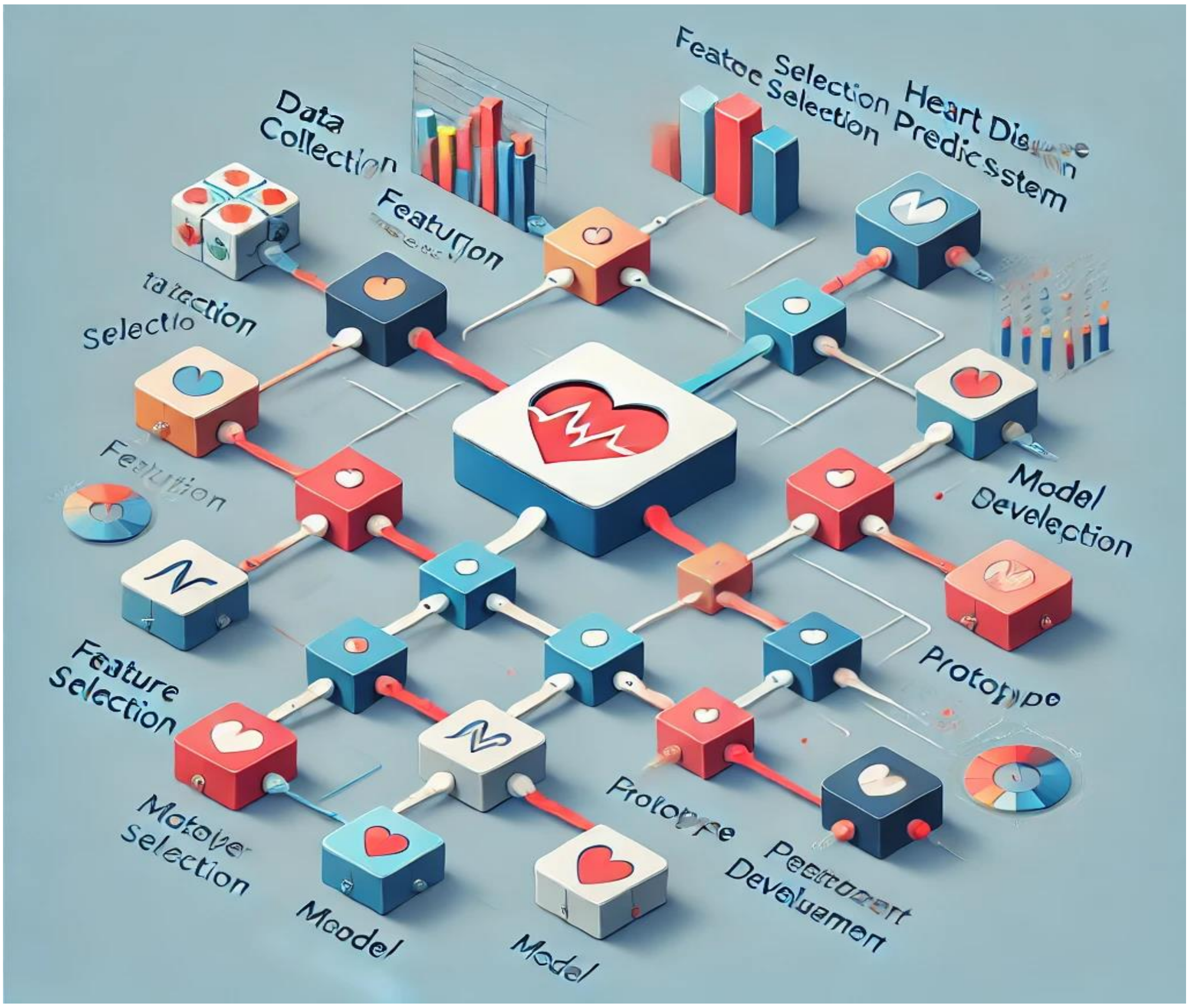
Heart Disease Prediction System

BATTALA RADHIKA

Step 1: Prototype Selection for Heart Disease Prediction System

Abstract

The rise of heart disease as a global health concern requires innovative solutions for early detection and prevention. In this report, I aim to apply Machine Learning techniques to predict the risk of heart disease in individuals based on medical data. This prototype will utilize classification algorithms to assess heart disease risk and provide actionable insights for healthcare providers and individuals. By leveraging machine learning, the system will not only offer early warning signs but also help healthcare professionals in personalized treatment planning, similar to how predictive models are used in industries like e-commerce for customer targeting and retention. The goal is to optimize patient outcomes and improve overall healthcare delivery.



Prototype Idea: Heart Disease Prediction System

A machine learning model that predicts the likelihood of heart disease in individuals based on health data (e.g., age, cholesterol levels, blood pressure, etc.) and provides actionable insights to users and healthcare providers.

Evaluation Criteria

a. Feasibility

Short-term Development:

- Data Availability: similar data from Kaggle.
- Technology: Accessible tools and frameworks (Python, T Scikit-learn) make it feasible to build this prototype in 6-12 months.
- Team Skill Requirements: Basic knowledge of data science, machine learning, and web development is sufficient.

Time Frame: A functional prototype could be developed and deployed within 1-2 years.

b. Viability

Long-term Relevance:

- Health Sector Demand: Chronic conditions like heart disease are among the leading causes of death worldwide. Predictive tools for early detection will remain valuable as healthcare increasingly focuses on prevention.
- AI Integration in Healthcare: Governments and organizations are adopting AI-based health tools to reduce costs and improve outcomes, ensuring long-term demand.
- Technological Evolution: As wearable devices and electronic health records (EHRs) become more prevalent, this product can integrate seamlessly for enhanced functionality.

c. Monetization

Direct Monetization:

- Subscription-based Model: Charge healthcare providers or hospitals a subscription fee for using the tool.
- Pay-per-Use Model: Individuals could pay for a one-time health risk assessment.
- Integration with Wearables: Partner with wearable device companies (e.g., Fitbit, Apple) to enhance their health tracking capabilities.

Market Potential:

- The global market for AI in healthcare is projected to grow significantly over the next two decades.
- By aligning with the increasing demand for preventive care, this product can achieve sustainable revenue streams.

Step 2: Prototype Development

Datasets

- https://www.kaggle.com/datasets/krishujeniya/heart-diseae
- Input Features: Age, sex, blood pressure, cholesterol, etc.
- Target: Presence or absence of heart disease (binary classification: 0 = no heart disease, 1 = heart disease).

Technology Stack

- Language: Python
- Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Steps for Code Implementation

1.Import Libraries

[3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sbn
```

2. Load and Explore the Dataset

[8]:
```
df=pd.read_csv("heart-disease.csv")
```

10]:
```
df.head(7)
```

10]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1

Columns

age: Age of the patient (in years)
sex: Sex of the patient (1 = male, 0 = female)
cp: Chest pain type (1-4)
trestbps: Resting blood pressure (in mm Hg on admission to the hospital)
chol: Serum cholesterol in mg/dl
fbs: Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
restecg: Resting electrocardiographic results (0-2)
thalach: Maximum heart rate achieved
exang: Exercise-induced angina (1 = yes; 0 = no)
oldpeak: ST depression induced by exercise relative to rest

[11]:
```
df.describe()
```

[11]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.313531	0
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.612277	0
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.000000	0
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.000000	1
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1

[17]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       303 non-null    int64  
 1   sex       303 non-null    int64  
 2   cp        303 non-null    int64  
 3   trestbps  303 non-null    int64  
 4   chol      303 non-null    int64  
 5   fbs       303 non-null    int64  
 6   restecg   303 non-null    int64  
 7   thalach   303 non-null    int64  
 8   exang     303 non-null    int64  
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64  
 11  ca        303 non-null    int64  
 12  thal      303 non-null    int64  
 13  target    303 non-null    int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

3.Dataframe mean

```
[12]: df.mean()
```

```
[12]: age          54.366337
      sex           0.683168
      cp            0.966997
      trestbps    131.623762
      chol        246.264026
      fbs           0.148515
      restecg       0.528053
      thalach     149.646865
      exang         0.326733
      oldpeak       1.039604
      slope         1.399340
      ca            0.729373
      thal          2.313531
      target        0.544554
      dtype: float64
```

4.Dataframe meadian

```
[18]: df.median()
```

```
[18]: age          55.0
      sex           1.0
      cp            1.0
      trestbps    130.0
      chol        240.0
      fbs           0.0
      restecg       1.0
      thalach     153.0
      exang         0.0
      oldpeak       0.8
      slope         1.0
      ca            0.0
      thal          2.0
      target        1.0
      dtype: float64
```

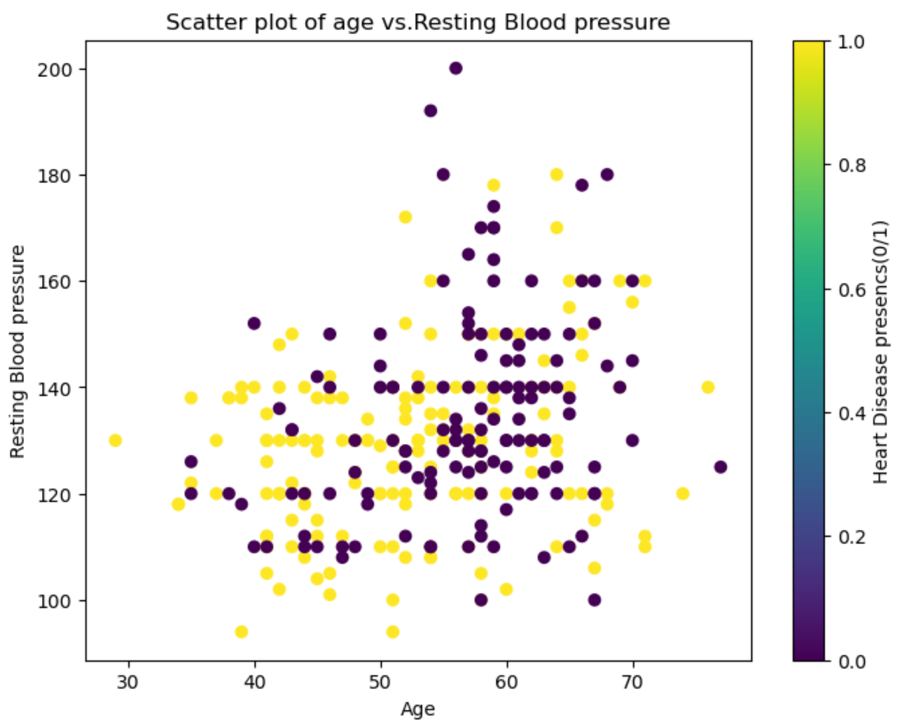
5.Dataframe Standard deviation

```
[29]: import numpy as np
      std_dev=np.std(df, axis=0)
      for i, std in enumerate(std_dev):
          print(f"Feature {i+1}:standard deviation= {std:.4f}")
```

```
Feature 1:standard deviation= 9.0671
Feature 2:standard deviation= 0.4652
Feature 3:standard deviation= 1.0303
Feature 4:standard deviation= 17.5092
Feature 5:standard deviation= 51.7452
Feature 6:standard deviation= 0.3556
Feature 7:standard deviation= 0.5250
Feature 8:standard deviation= 22.8673
Feature 9:standard deviation= 0.4690
Feature 10:standard deviation= 1.1592
Feature 11:standard deviation= 0.6152
Feature 12:standard deviation= 1.0209
Feature 13:standard deviation= 0.6113
Feature 14:standard deviation= 0.4980
```

6.Scatter Plot

```
import matplotlib.pyplot as plt
plt.figure(figsize=(8,6))
plt.scatter(df['age'],df['trestbps'],c=df['target'],cmap='viridis')
plt.xlabel('Age')
plt.ylabel('Resting Blood pressure')
plt.title('Scatter plot of age vs.Resting Blood pressure')
plt.colorbar(label='Heart Disease presencs(0/1)')
plt.show()
```



Step 3: Business Modelling (Short Version)

1. Value Proposition:

Early detection of heart disease using AI, enabling preventive care for individuals and resource optimization for healthcare providers.

2. Target Audience:

- Primary: Hospitals, clinics, insurance companies, individuals.
- Secondary: Wearable device companies, health tracking apps.

3. Revenue Streams:

- Subscription Model: Monthly/yearly fees for healthcare providers and individuals.
- Pay-Per-Use Model: One-time payment for predictions.
- Partnerships & Licensing: Collaborate with wearable device companies and insurers.
- Freemium Model: Basic risk assessments free, premium features paid.

4. Cost Structure:

- Development Costs: Model building, app/website development.
- Operational Costs: Cloud hosting, customer support.
- Compliance Costs: Regulatory compliance (e.g., HIPAA, GDPR).

  5. Marketing Strategy:
- Direct Outreach: Partnerships with hospitals and insurers.
- Digital Marketing: SEO, content marketing, and social media campaigns.
- Trade Shows: Showcase at healthcare tech events.
- Wearable Integration: Collaborate with fitness device companies.

5. Marketing Strategy:

- Outreach: Partnerships with hospitals and insurers.
- Digital Marketing: SEO, content marketing, and social media campaigns.
- Trade Shows: Showcase at healthcare tech events.
- Wearable Direct Integration: Collaborate with fitness device companies.

Step 4: Financial Modelling for Heart Disease Prediction Machine Learning Project

1. Financial Equation for Growth:

The general equation for profit over time is:

$$Y=X\times(1+r)t$$

Where:

- Y = Profit over time
- X = Price of the product (Rs. 250 for heart disease prediction)
- r = Growth rate (3.2% per period)
- t = Time interval (e.g., months or years)

Given:

- X = Rs. 250
- r = 3.2% (0.032)
- t = Time in months or years (we'll use months for simplicity)

So, the formula becomes:

Y=250×(1.032)t

This equation allows us to forecast the revenue or profit as the growth rate of customers or sales increases over time.

---

2. Profit Over Time (Example for 6 Months)

Let's calculate the profit for **t = 1 to 6 months** with the assumption of a consistent price of **Rs. 250 per prediction**.

Y=250×(1.032)t

For each month (t = 1 to 6):

- **Month 1 (t = 1):**

Y1=250×(1.032)1=250×1.032=258

- **Month 2 (t = 2):**

Y2=250×(1.032)2=250×1.065=266.25

- **Month 3 (t = 3):**

Y3=250×(1.032)3=250×1.098=274.50

- **Month 4 (t = 4):**

Y4=250×(1.032)4=250×1.132=283.00

- **Month 5 (t = 5):**

Y5=250×(1.032)5=250×1.167=291.75

- **Month 6 (t = 6):**

Y6=250×(1.032)6=250×1.202=300.50

So, over the first 6 months, the price per prediction grows steadily, following the 3.2% growth rate.

3. Profit Forecast with Sales

Let's assume that your **sales (X)** remain constant, and we focus only on how the growth rate (r) affects the price.

If we multiply these values by the number of predictions sold, we can calculate the total profit for each month. For example:

- **If 300 predictions are sold in the first month:**

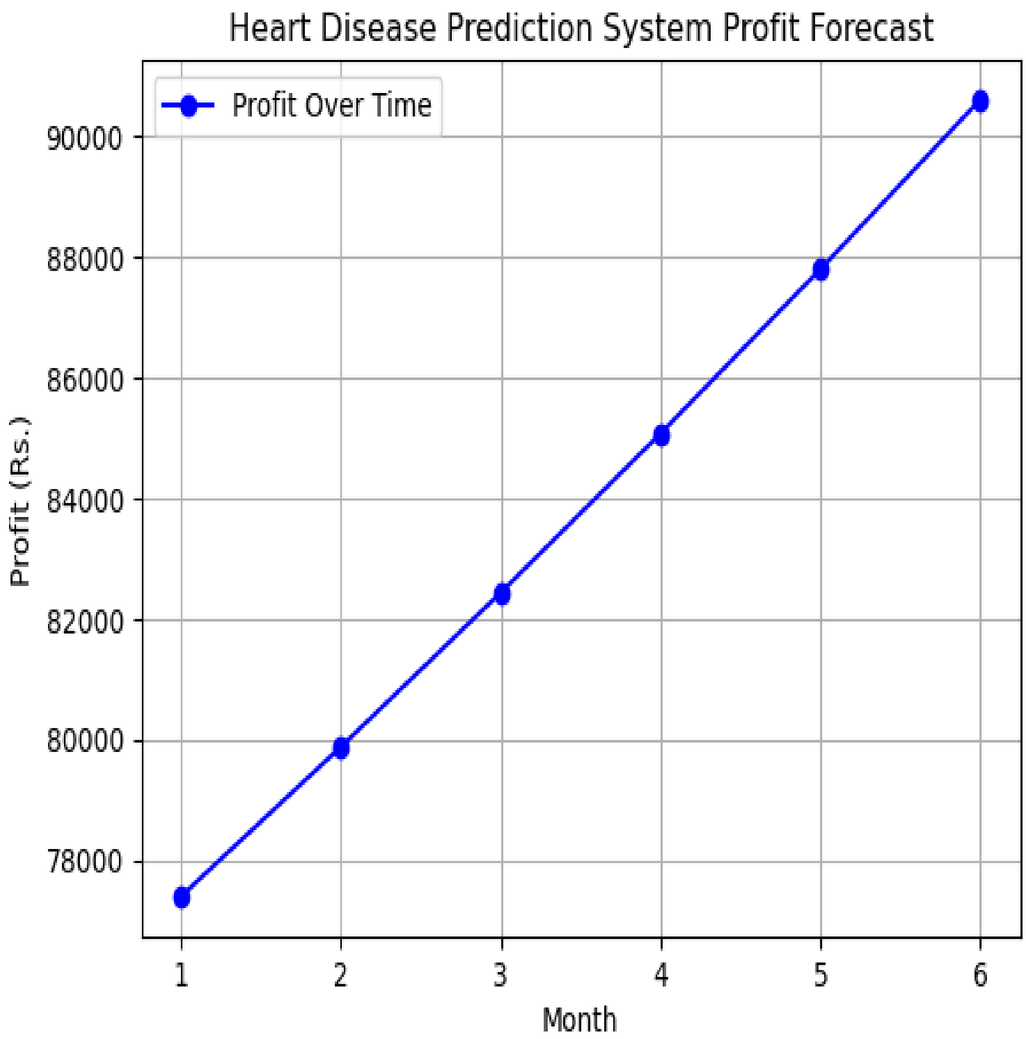
Y1=300×258=77,400

- **For the 2nd month (after growth rate increase):**

  $$Y2=300\times266.25=79,875$$

4. Graph of Profit Over Time

Now, let's visualize the profit over the 6-month period, with constant sales but growing price due to the 3.2% growth rate.



5. Conclusion:

The Heart Disease Prediction System leverages advanced machine learning techniques to provide early detection of heart disease, offering significant benefits for individuals and healthcare providers. By predicting potential risks, it enables proactive healthcare management and resource optimization. This system not only helps in preventing heart disease but also supports better decision-making in medical practices. With growing demand for preventive healthcare, this solution has the potential to revolutionize health management and improve patient outcomes.