Machine Learning Based Heart Attack Prediction Application

LOKESH R

05 | 12 | 2024

GITHUB LINK : https://github.com/Lokeshcham/FL_APP.git

Abstract

This project focuses on integrating advanced machine learning techniques to enhance heart attack prediction and raise health awareness. The primary goal is to utilize collaborative algorithms to create a personalized heart attack prediction system or application. By analyzing comprehensive patient profiles, including factors such as age, sex, cerebral palsy, blood pressure, cholesterol levels, fasting blood sugar, thallium stress test results, resting electrocardiogram readings, and oldpeak values, the system can accurately predict whether a patient is at risk of a heart attack or is symptom-free.

Beyond building a reliable prediction model, this project explores the development of a sustainable healthcare model to bring the heart attack prediction system to the market. The model aims to make the system accessible, promoting early detection and personalized care. By integrating this innovative solution into the healthcare industry, the project envisions a significant impact on improving patient outcomes and reducing the risk of heart-related issues.

1.0 Problem statement

Heart disease, particularly heart attacks, remains a leading cause of mortality worldwide, often due to delayed detection and insufficient personalized care. Current diagnostic methods do not always provide early warnings or individualized risk assessments, limiting the potential for timely intervention. There is a need for an advanced predictive system that can analyze diverse patient data and accurately forecast heart attack risks. The challenge lies in developing a robust machine learning model that not only identifies heart attack symptoms from detailed patient profiles but also integrates into a sustainable healthcare system, enabling widespread adoption and early preventive care.

2.0 Market/Customer/Business Requirements Evaluation

2.1Market Needs:

2.1.1 **Rising Incidence of Heart Disease:** With heart disease being one of the leading causes of death globally, the market demand for predictive health technologies is growing. The ability to accurately forecast heart attack risks could significantly improve preventive healthcare outcomes.

2.1.2 **Shift Towards Personalized Healthcare:** Consumers are increasingly seeking personalized medical solutions tailored to their unique profiles. A heart attack prediction system that leverages machine learning to analyze individualized patient data would meet this growing demand for customized healthcare services.

2.1.3 **Digital Health Integration:** The healthcare industry is moving towards digital and AI-driven solutions. A predictive system that integrates with existing digital health platforms and electronic medical records (EMRs) would address the need for innovation and technological integration in healthcare.

2.2 Customer Needs:

2.2.1 Early Detection and Prevention: Patients, particularly those at high risk of heart disease, need tools that can offer early warnings and risk assessments to prevent life-threatening cardiac events. This system should provide actionable insights and recommendations for improving health.

2.2.2 User-Friendly Interface: Customers, including patients and healthcare providers, require a system that is intuitive and easy to use, with clear and understandable results. Both mobile and web-based platforms would increase accessibility for users.

2.2.3 Affordability and Accessibility: A cost-effective solution is essential to ensure that the system is accessible to a wide range of users, including individuals in low-resource settings or those with limited access to healthcare.

2.3 Business Requirements:

2.3.1 Scalability: The heart attack prediction system must be scalable to accommodate a growing number of users as demand increases. The solution should be designed to handle diverse patient data across various healthcare settings.

2.3.2 Compliance with Healthcare Regulations: The system must adhere to medical data privacy regulations, such as HIPAA or GDPR, depending on the market. It should also meet healthcare standards for accuracy and reliability.

2.3.3 Partnerships and Collaborations: To bring the system to market successfully, partnerships with hospitals, insurance companies, and healthcare providers will be critical. Collaboration with these stakeholders can facilitate market penetration and integration into existing healthcare workflows.

2.3.4 Sustainable Business Model: A robust business model should be developed to ensure profitability and long-term sustainability. This could include subscription-based services for healthcare providers, direct-to-consumer options, or partnerships with insurance companies that offer the system as part of preventive care packages.

2.4 Technological Requirements:

2.4.1 Integration with Health Devices: The system should have the capability to integrate with wearables and health monitoring devices to collect real-time patient data, enhancing predictive accuracy.

2.4.2 Cloud-Based Data Storage and Processing: Given the volume and sensitivity of medical data, the system must leverage secure cloud-based infrastructure for data storage, processing, and real-time analytics.

2.4.3 Machine Learning and AI Algorithms: Advanced machine learning models will be needed to analyze patient data accurately. These algorithms should be trained on diverse datasets to ensure they are adaptable to a variety of demographic and clinical variables. Regular updates to the model will be necessary to maintain accuracy.

By evaluating these market, customer, and business requirements, the project can create a comprehensive and scalable heart attack prediction system that addresses the real-world needs of healthcare providers and patients while adhering to technological and regulatory standards.

3.0 Specifications and Characteristics of the Target:

3.1 Target Audience:

3.1.1 Patients at high risk of heart disease (e.g., those with a history of hypertension, high cholesterol, diabetes, or a sedentary lifestyle).

3.1.2 Healthcare providers, including doctors, cardiologists, and medical facilities looking to improve predictive diagnosis.

3.1.3 Health-conscious individuals seeking personalized health assessments for preventive care.

3.2 Demographic Characteristics:

3.2.1 Adults aged 35 and above, as this age group is typically at a higher risk of heart disease.

3.2.2 Both male and female users, as heart disease affects both genders, though risk factors may vary.

3.2.3 Geographic flexibility, targeting users globally, with a focus on regions with high heart disease prevalence.

3.3 Technological Characteristics:

3.3.1 Users with access to smartphones, tablets, or computers for web or mobile-based applications.

3.3.2 Integration with wearable health devices like fitness trackers to gather real-time data.

3.3.3 Users connected to healthcare systems that use electronic medical records (EMRs) for data input and tracking.

3.4 Behavioral Characteristics:

3.4.1 Individuals who are proactive about their health and prefer data-driven insights.

3.4.2 Patients willing to monitor their heart health regularly and make lifestyle adjustments based on the system's predictions.

3.4.3 Medical professionals seeking advanced tools to complement clinical evaluations.

3.5 Socioeconomic Characteristics:

3.5.1 Middle to upper-income groups who can afford digital health services, though accessibility for low-income users should be considered through cost-effective pricing models.

3.5.2 Users with varying levels of health literacy, requiring an intuitive and user-friendly interface with easily interpretable results.

These characteristics define the key segments that the heart attack prediction system will target, ensuring it meets the needs of both individual users and healthcare professionals.

4.0 External Search (online information sources/references/links)

Heart attack prediction systems using machine learning are gaining attention due to their potential to significantly improve early detection and prevention of cardiovascular diseases. Studies from sources like the Journal of Engineering and Applied Science and IEEE Xplore highlight how algorithms such as Random Forest, Support Vector Machines (SVM), and Decision Trees are being applied to clinical data like age, blood pressure, and cholesterol to accurately predict heart attack risks. These systems also tackle class imbalance in medical datasets using techniques like SMOTE, improving model performance and precision.

Furthermore, the integration of these predictive models with real-time data from wearable devices is revolutionizing healthcare by providing continuous monitoring of heart health. Sources like Frontiers in Cardiovascular Medicine discuss the use of IoT technologies alongside machine learning to create more robust and accurate prediction models. These innovations enable healthcare providers to offer more personalized, proactive care and enhance overall patient outcomes by predicting heart disease risks before symptoms manifest, making timely interventions possible(SpringerOpen).

5.0 Bench marking alternate products (comparison with existing products/services)

5.1 Existing Heart Attack Prediction Tools: Current heart disease prediction tools like *CardioRisk Calculator* and *HEART Pathway* rely on traditional risk assessment models (e.g., Framingham score) and are limited to predefined parameters like age, cholesterol, and smoking habits. While effective, these tools may not provide personalized predictions based on complex data interactions.

5.2 AI-based Systems: AI-powered solutions such as CLEW Medical and Aidoc leverage machine learning to predict heart disease using real-time patient data from wearables, ECGs, and other diagnostic inputs, providing more precise, personalized predictions compared to conventional methods. The use of real-time data and advanced algorithms makes these systems more adaptable and accurate, enabling early detection and interventions.

6.0 Applicable Patents:

6.1 Found patents related to community-driven platforms, picture recognition, and customized nutrition.

6.2 Ensuring compatibility with current developments while adding special features.

6.3 Risk reduction through innovation promotion and infringement avoidance.

7.0 Applicable Regulations:

7.1 Health Insurance Portability and Accountability Act (HIPAA): In the U.S., HIPAA regulates the privacy and security of patient health information, requiring systems to implement safeguards to protect sensitive data.

7.2 General Data Protection Regulation (GDPR): In Europe, GDPR governs the processing of personal data, ensuring users have control over their information, which is crucial for health-related applications.

7.3 Food and Drug Administration (FDA) Regulations: If the system is classified as a medical device, it must comply with FDA regulations for safety, efficacy, and quality assurance.

7.4 ISO/IEC 27001: This international standard provides a framework for information security management, applicable to ensure data protection in health technology.

These regulations ensure that heart attack prediction systems are compliant with data protection and patient safety standards.

8.0 Applicable Constraints (need for space, budget, expertise)

8.1 Space Requirements: The system may require dedicated server space for data storage, processing, and application deployment, particularly if it integrates with cloud services for real-time analytics.

8.2 Budget Constraints: Development costs can be significant, covering software development, machine learning model training, regulatory compliance, and potential partnerships with healthcare providers. Continuous funding will be needed for updates and maintenance.

8.3 Expertise: A multidisciplinary team is essential, including data scientists, software developers, healthcare professionals, and regulatory compliance experts, to ensure the system is robust and meets industry standards.

9.0 Business Model (Monetization Idea)

The heart attack prediction system can adopt a subscription-based model, offering tiered pricing for individuals, healthcare providers, and organizations. Individual users could pay a monthly fee for personalized health insights, while healthcare facilities could subscribe for bulk access to advanced analytics and reporting features. Additionally, partnerships with insurance companies could offer the service as part of wellness programs, potentially lowering premiums for users who actively engage with the system. An optional premium tier could include advanced features like real-time monitoring and personalized consultations with healthcare professionals.

10.0 Final Product Prototype Abstract:

The heart attack prediction system is a machine learning-based application designed to analyze individual patient data, such as age, blood pressure, cholesterol levels, and lifestyle factors, to assess the risk of heart attacks. By employing advanced algorithms, the system generates personalized risk assessments and recommendations for preventive measures. The prototype features a user-friendly interface for patients and healthcare providers, allowing easy access to health insights. Integration with wearable devices enables real-time monitoring and data collection, enhancing predictive accuracy and timely interventions.

Schematic Diagram:

10.1 Data Sources:

10.1.1 Wearable Devices (Heart rate, activity levels)

10.1.2 Patient Health Records (Medical history, lab results)

10.1.3 User Input (Lifestyle choices, symptoms)

10.2 Data Processing:

10.2.1 Data Cleaning

10.2.2 Feature Selection

10.2.3 Model Training (Machine Learning Algorithms)

10.3 Prediction Module:

10.3.1 Risk Assessment Output

10.3.2 Recommendations (Lifestyle changes, follow-up actions)

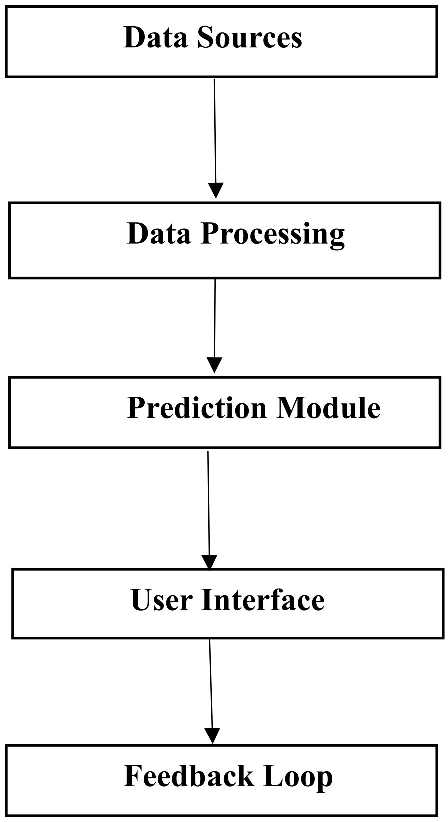
10.4 User Interface:

10.4.1 Dashboard for patients

10.4.2 Reports for healthcare providers

10.5 Feedback Loop:

10.5.1 Continuous Learning (Updating models with new data)


Data Sources
Data Processing
Prediction Module
User Interface
Feedback Loop

In [1]:
```python
# importing the library module to read the dataset
import pandas as pd
dataset = pd.DataFrame(pd.read_excel("1645792390_cep1_dataset.xlsx"))
```

In [2]:
```python
# printing the head of the dataset
dataset.head()
```

Out[2]:

	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

In [3]:
```python
# printing the shape of the dataset
dataset.shape
```

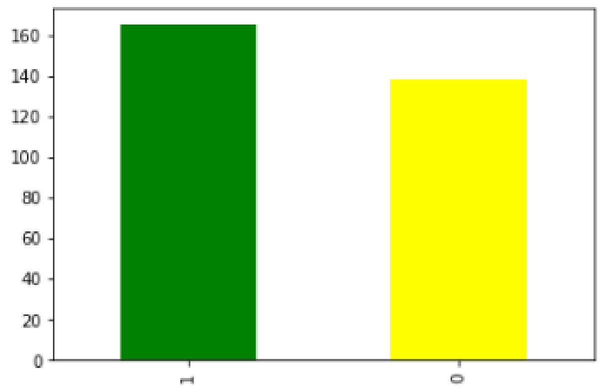
Out[3]:
```
(303, 14)
```

In [4]:
```python
# importing the required library modules
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```python
# printing the total count of categorical value of column Target
dataset['target'].value_counts()
```

Out[5]:
```
1    165
0    138
Name: target, dtype: int64
```

In [6]:
```python
# creating the var chart for the Target(categorical value)
dataset['target'].value_counts().plot(kind='bar',color=['green','yellow'])
plt.show()
```

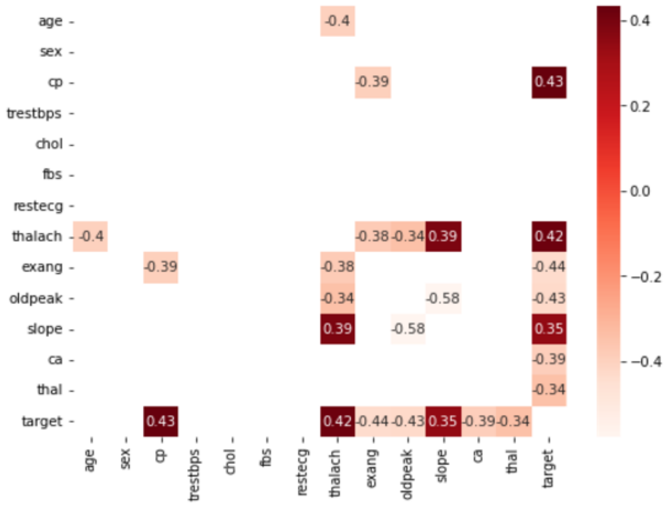


In [7]:
```python
# basic summary of the dataset loaded
dataset.info()
```

```
<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
```

In [11]:
```python
# plottingh the correlation
corr=dataset.corr()
thresh=0.3
kot=corr[((corr>=thresh)|(corr<=-thresh))&(corr!=1)]
plt.figure(figsize=(8,6))
sns.heatmap(kot,cmap='Reds',annot=True)
```

Out[11]: <AxesSubplot: >



In [12]:
```python
# counting the categorical values in sex column
dataset.sex.value_counts()
```

Out[12]:
```
1    207
0     96
Name: sex, dtype: int64
```

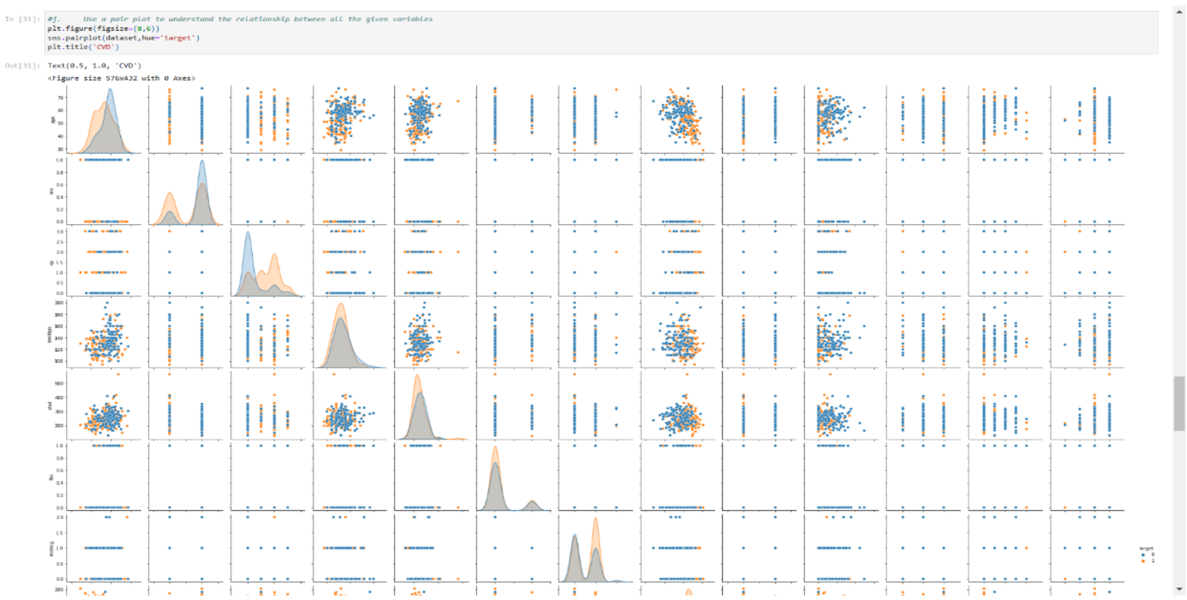
In [13]:
```python
#creating contingency table to compare sex with target
pd.crosstab(dataset.target,dataset.sex)
```

Out[13]:

sex	0	1
target		
0	24	114
1	72	93

In [14]:
```python
# 1=male,93 males as compared to 72 females are detected with CVD. So males are at a higher risk of CVD
```

In [15]:
```python
# createing plot of CVD against sex
pd.crosstab(dataset.target,dataset.sex).plot(kind='bar',figsize=(10,6),color=['yellow','green'])
```



In [33]:
```
#building the logistic regression model
from sklearn.linear_model import LogisticRegression
lr= LogisticRegression()
lr.fit(X_train,y_train)
```

Out[33]:
```
▾ LogisticRegression
LogisticRegression()
```

In [34]:
```
y_pred = lr.predict(X_test)
```

In [35]:
```
# checking the accuracy of the model
from sklearn.metrics import accuracy_score
score_lr = round(accuracy_score(y_pred,y_test)*100,2)
print("The accuracy score achieved using logistic regression is: "+ str(score_lr)+" %")
```

```
The accuracy score achieved using logistic regression is: 84.21 %
```

In [36]:
```
# fitting a stats logistic regression model
import statsmodels.api as sm
log_reg=sm.Logit(y_train,X_train).fit()
```

```
Optimization terminated successfully.
         Current function value: 0.343229
         Iterations 7
```

In [37]:
```
# displaying the summary reports
print(log_reg.summary())
```

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                 target   No. Observations:                  227
Model:                          Logit   Df Residuals:                      214
Method:                           MLE   Df Model:                           12
Date:                Wed, 06 Sep 2023   Pseudo R-squ.:                  0.5028
Time:                        12:26:38   Log-Likelihood:                -77.913
converged:                       True   LL-Null:                       -156.71
Covariance Type:            nonrobust   LLR p-value:                 1.627e-27
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
age            0.0176      0.022      0.803      0.422      -0.025       0.060
sex           -2.0263      0.531     -3.815      0.000      -3.067      -0.985
cp             0.8986      0.223      4.027      0.000       0.461       1.336
trestbps      -0.0065      0.011     -0.582      0.561      -0.028       0.015
chol          -0.0057      0.004     -1.363      0.173      -0.014       0.003
fbs           -0.6415      0.598     -1.072      0.284      -1.814       0.531
restecg        0.2590      0.402      0.645      0.519      -0.528       1.046
thalach        0.0321      0.010      3.280      0.001       0.013       0.051
exang         -0.8697      0.472     -1.843      0.065      -1.795       0.055
oldpeak       -0.5817      0.247     -2.356      0.018      -1.066      -0.098
slope          0.2910      0.428      0.680      0.497      -0.548       1.130
ca            -0.8349      0.222     -3.762      0.000      -1.270      -0.400
thal          -0.8006      0.327     -2.446      0.014      -1.442      -0.159
==============================================================================
```

In [38]:
```
from sklearn.ensemble import RandomForestClassifier
clf=RandomForestClassifier(criterion='gini',
                           max_depth=7,
                           n_estimators=200,
                           #min_samples_split=10,
                           random_state=5)
```

11.0 Conclusion:

In conclusion, the heart attack prediction system represents a significant advancement in personalized healthcare. By leveraging machine learning algorithms and real-time data integration, it provides accurate risk assessments and actionable insights for patients and healthcare providers. This proactive approach not only aims to reduce the incidence of heart attacks through early detection but also fosters a culture of health awareness and preventive care. With the potential for wide-scale adoption, the system can ultimately enhance patient outcomes and contribute to a more sustainable healthcare model.

12.0 References and Resources:

Heart Disease Prediction Techniques: This paper reviews various machine learning methods used for heart disease prediction, analyzing their effectiveness.

12.1 AI in Cardiovascular Health: This article discusses how AI and IoT can enhance cardiovascular health monitoring and prediction.

12.2 Machine Learning in Healthcare: A comprehensive review on the applications of machine learning in healthcare predictive modeling, including heart disease.