Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Heart Disease Prediction Using Machine Learning

[1] D.Syed Ali, [2] Sarranya S., [3] Nithya M., , [4] Dr.T.Kumanan

[1,4]*Professor* ,[2,3]*Students*

[1234]*Department of Computer Science and Engineering, Dr.M.G.R Educational And Research Institute, Chennai*

Abstract

A Web application built using the Flask framework, integrating user authentication, database management, and a predictive model for heart disease. Users can register, log in, and access a personalized dashboard displaying visualizations of age and gender distribution in a heart disease dataset. The application employs a logistic regression model trained on heart-related data to predict.Approximately 18 billion people die due to heart disease related problem over a year as per WHO.With increasing population, it gets further difficult to diagnose the disease. But in this growing technology world,Machine Learning techniques have accelerated the health sector by multiple researches. Thus, the objective of this paper is to build a ML model for heart disease prediction based on the related parameters. We have used adataset of UCI Heart disease prediction for this research, which consist of 13 different parameters related to Heart-disease. ML algorithms such as Random Forest, Support Vector Machine (SVM), Naïve Bayes and Decision tree have been used for developing the model. In this research we have also tried to find the correlations between the different attributes available in the dataset with the help of standard ML methods and then used them efficiently in predicting the chances of heart disease. Result shows that compared to other Machine learning techniques, Random Forest predicts with more accuracy in less time. This model can also be helpful to the medical practitioners at their clinic as decision support system.

Keywords: *Online Predicting System, Flask, Logistic Regression, Heart Disease Prediction*

1. Introduction

A Web application using the Flask framework, incorporating essential features such as user authentication, database management with SQLite, and a predictive model for heart disease. Users can seamlessly register, log in, and access a personalized dashboard that offers insightful visualizations of age and gender distribution within a heart disease dataset. Leveraging logistic regression, the application can predict the likelihood of heart disease based on user-inputted parameters. The integration of WTForms ensures secure and efficient form handling, while Plotly enhances the user experience with interactive data visualizations. This comprehensive implementation showcases a well-structured and functional web application tailored for heart disease prediction and user interaction. Furthermore, the code emphasizes best practices in security by incorporating CSRF protection and a secure password hashing mechanism. The user data is efficiently managed using SQLAlchemy, allowing seamless interaction with the SQLite database. The application not only provides a user-friendly interface for predicting heart disease but also facilitates user engagement through a streamlined registration and login process. The user dashboard offers a glimpse into the dataset's demographic distribution, fostering a better understanding of the underlying data. Overall, the code reflects a cohesive and well-implemented solution, combining predictive analytics, data visualization, and user authentication to deliver a robust and user-centric heart disease prediction web application.

2. System Model

Lots of research work have been undergone for assessment of the classification accuracies of different ML algorithms by using the Cleveland heart disease database which is uninhibitedly accessible at an online data mining repository of the UCI. Roughly 280 million individuals around the globe were experiencing diabetes as per a study directed by International Diabetes Federation (IDF). As a matter of fact, detection of type 2 diabetes at beginning phase isn't a simple undertaking, yet research done by the authors, in which data mining was used that it gives the best results, helped in the disclosure of information from the accessible data.

Machine Learning -Yu-Xuan Wang, et.al. have explored different applications that demonstrated the significance of the ML methods in various areas [9].A new technique for the designing of a working framework has been proposed by them.. The approach used the distinct machine learning procedures. After getting the proper result from the data miner, the whole information assembled from the structure was inspected. In light of the various tests, it was seen that proposed approach gave proficient results.

Algorithm - explored 8 unsupervised and 10 supervised learning algorithms. In their research, they showed an application work for the semi-supervised type learning algorithms. In industry method, it was seen that roughly 90%-95% applications utilized both the unsupervised and supervised machine learning procedures. Consequently, it was portrayed that the Machine Learning methods play an indispensable part in the planning of different novel applications for domains like medical services and industry.

3. Experimental System

The heart disease prediction Web Application exhibits several strengths, including user authentication, database integration, and a functional logistic regression model. However, certain aspects warrant attention to enhance its effectiveness and robustness. One notable concern is the limited scalability of the current model training approach. The code relies on a pre-loaded dataset, potentially hindering real-time updates and adaptation to evolving health data. Additionally, the absence of model evaluation metrics and interpretability features hinders the user's understanding of prediction reliability. Addressing these issues is crucial for ensuring the application's scalability, adaptability to changing datasets, and providing users with transparent insights into the reliability of the heart disease predictions.

3.1 Algorithm

1. User registers with a username, password, and email.
2. User information is securely stored in the database.
3. Registered users log in with their credentials.
4. Flask-Login manages user sessions and authenticates users.
5. Authenticated users access the home page with a welcome message.
6. Users navigate to the user dashboard to view age and gender distribution charts.
7. Users access the prediction page to input personal health details.
8. User input is processed, and the pre-trained model predicts the likelihood of heart disease.
9. Prediction result is displayed on the home page.
10. Users can log out, ending their session.

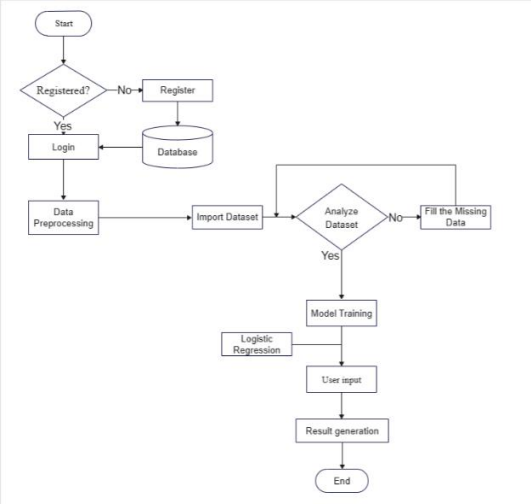

Figure 3.1 Architecture Diagram

3.2 Methodology

Problem Understanding and Dataset Exploration:Define the problem: Early prediction of heart disease using machine learning.Explore the "heart.csv" dataset to understand its structure and features. Identify the target variable and relevant features for the prediction.Set up a Flask web application with routes for different pages (index, login, register, user_dashboard, prediction_page).Implement user authentication using Flask-Login for secure access.User Authentication and Registration:Create User and LoginForm, RegisterForm classes using Flask-WTF for user authentication.Implement registration functionality to add new users to the database.Develop a login route to authenticate users and create user sessions.

Table 3.1 Attributes

S.No	Attributes	Desc	Mean Value
1	age	In years	54.434
2	Sex	Male, Female	0.696
3	Cp	Angina, abnang, notang, asympt	0.942
4	Trstbps	Resting Blood Pressure in mm hg	131.612
5	Chol	Serum Cholesterol in mg/dl	246
6	Fbs	Fasting blood sugar-1 if >120mg/dl, 0 if<120mg/dl	0.149
7	Restecg	Electrocardiographic Results	0.53

8	Thalach	Maximum Heart Rate observed	149.114
9	Exang	Extercise with angina has occurred	0.337
10	Oldpeak	ST depression induced through exercise	1.072
11	Slope	Slope of the ST segment	1.385
12	Thal	Number of major vessels ranging from 0-3 color by fluoroscopy	0.754
13	Ca	Heart status	2.34
14	target	Output Class	

Table 3.2 Sample Training Data

52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	156	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1

52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	1.3	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
46	1	0	120	249	0	0	144	0	0.8	2	0	2	0
54	1	0	122	286	0	0	116	1	3.2	1	2	3	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1

Table 3.2 Sample Testing Data

Dashboard Visualization: Create a user_dashboard route to display interactive charts using Plotly.Generate an age distribution bar chart and a gender distribution pie chart for insights.Develop a prediction_page route for users to input their data. Implement an HTML form to collect relevant patient information.

3.3 Modules

3.3.1 User Authentication Module:

This module, containing forms.py and routes/auth.py, manages user authentication within the application. It establishes forms for user registration and login, verifies user input, maintains user sessions, and regulates authentication-related routes including registration, login, and logout. In models.py, the User model is an integral part of the user data storage and retrieval process.

3.3.2 Data Collection Module:

The heart.csv file and data_preparation.py constitute the data collection module, which is in charge of loading, managing, and preparing the dataset needed to train the machine learning model. The format of the data is

guaranteed to be appropriate for training and assessment by this module. It is essential to preserving the dataset's quality and integrity.

3.3.3 Logistic Regression Module:

Logistic Regression is a troupe of unpruned classification. It gives amazing performance with concern to number of real-life problems, as it is non effective to noise in the dataset and risk of overfitting is also very less. In comparison to many other tree-based algorithms, it works faster than others and generally improves accuracy for testing and validation data. Logistic Regression are the aggregation of the predictions of individual decision tree algorithm. There are various choices to tune the performance of Logistic Regression when constructing a Logistic Regression.

3.3.4 Flask Configuration Module:

The Flask configuration module, represented by app.py, manages the overall configuration of the Flask application. It sets up the Flask app, configures the database URI, initializes Flask extensions (Flask-WTF and Flask-Login), and defines the main entry point for running the application. This module serves as the foundation for the entire Flask application.

3.3.5 Web Interface Module:

The application's primary pages are rendered by the web interface module, which also includes routes/index.py and related templates in the templates/ directory. It offers the user interface for the user dashboard, prediction page, and home page. In order to provide information, user input, and deliver the prediction results in an intuitive way, this interfaces with other modules.

3.3.6 Error Handling Module:

The error handling module controls the handling of errors that may arise during form submission, data processing, or prediction. When problems occur, it guarantees a smooth user experience by providing informative error messages.

3.3.7 Generating Charts Module:

The chart creation module, which is represented by routes/user_dashboard.py, is in charge of producing visuals depending on the dataset. It takes pertinent data out of the dataset and uses Plotly to create dynamic graphics. Following that, these charts are displayed on theuser dashboard, giving information about the distribution of age and gender in the heart data.

4. Result And Discussions

Flask web application for heart disease prediction, incorporating user authentication, registration, and a user dashboard with insightful visualizations. The logistic regression model trained on a heart disease dataset offers a basic yet functional prediction mechanism. When users input relevant health parameters on the prediction page, the application processes the data through the model and returns a result indicating whether the person is predicted to have a heart disease or not. Future improvements and evaluations could enhance the application's predictive accuracy, expand its features, and ensure its ethical and clinical relevance in the domain of healthcare. for a more comprehensive evaluation, it is important to assess the model's predictive performance metrics, such as accuracy, precision, recall, and F1-score. Including these metrics in the results would provide users with a clearer understanding of the model's reliability and effectiveness in making accurate predictions.Collaborating with healthcare professionals and researchers could further validate and improve the model, ensuring its alignment with medical standards.

5. Conclusion

In conclusion, the heart disease prediction web application seamlessly integrates web development, machine learning, and data visualization technologies. Utilizing Flask, it establishes a secure and interactive user interface,

with Flask-Login ensuring authenticated access to personalized dashboards. SQLAlchemy simplifies database management, while WTForms optimizes the user experience through efficient form handling. The core predictive functionality, driven by a logistic regression model, offers valuable insights into the likelihood of heart disease based on user-input parameters.However, for the application to maximize its potential, should be given to enhancing model scalability and adaptability to real-time data updates. While the current implementation effectively utilizes Plotly for data visualization, potential improvements involve exploring machine learning models, implementing real-time training, integrating additional evaluation metrics for enhanced prediction accuracy. The current implementation provides a robust foundation, but ongoing refinement and adaptation are crucial to ensure the application remains effective, scalable, and capable of delivering meaningful insights to users assessing their risk of heart disease.

6. Future Scope

Web-based heart disease prediction application, but there are several avenues for future development and enhancement. Firstly, considering the evolving field of machine learning, incorporating more advanced algorithms and models could significantly improve prediction accuracy. On the user interface side, the current implementation includes basic registration, login, and visualization features. To enhance user engagement, future iterations could introduce more interactive elements, such as personalized health insights based on predictions, historical trend tracking, or educational resources. Integration with external APIs for health-related information could provide users with more context and guidance.

References

[1] https://www.who.int/hrh/links/en/

[2] https://en.wikipedia.org/wiki/Machine_learning

[3] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp. 204-207, doi: 10.1109/ISCC.2017.8024530.

[4] Raparthi, M., Dodda, S. B., & Maruthi, S. (2023). Predictive Maintenance in IoT Devices using Time Series Analysis and Deep Learning. Dandao Xuebao/Journal of Ballistics, 35(3). https://doi.org/10.52783/dxjb.v35.113

[5] S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777531.

[6] C. Raju, E. Philipsy, S. Chacko, L. Padma Suresh and S. Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques," 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, 2018, pp. 253-255, doi: 10.1109/ICEDSS.2018.8544333.

[7] ZhiqiangGe, Zhihuan Song, Steven X. Ding, Biao Huang, “Data Mining and Analytics in the Process Industry: The Role of Machine Learning”, 2017 IEEE Translations and contentmining are permitted for academic research only, vol. 5, pp. 20590-20616, 2017.

[8] https://en.wikipedia.org/wiki/Support_vector_machine.

[9] https://archive.ics.uci.edu/ml/datasets/heart+disease.