Constructing Heart Disease Prediction Models Using Machine Learning

Author : Chayanan Jaknan
Advisor : Assistant Professor Dr. Matinee Kiewkanya

Abstract

This research developed a machine learning model to predict the risk of heart disease using patients' health data. Several algorithms, including Logistic Regression, Support Vector Machine, Random Forest, and K-Nearest Neighbor, were compared. Model performance was evaluated using Accuracy, Precision, Recall, and F1-score. Cross-validation and data scaling were applied to improve reliability. The best-performing model was then selected and implemented as a web application. The results showed that the SVM achieved the highest performance, with an F1-score of 90.95%, outperforming the other models. Furthermore, preprocessing techniques improved model accuracy.

Introduction

Heart disease remains a major global health problem and a leading cause of mortality. Early detection can significantly reduce severe complications and death rates. However, traditional diagnosis can be time-consuming.

This research applies machine learning techniques to develop a predictive model for heart disease risk. Machine learning can analyze large volumes of data and helps assess disease risk quickly and efficiently. These technologies can reduce the workload of medical personnel and enable patients to receive care more promptly. The proposed approach can serve as a guideline for developing medical developing future medical decision-support systems.

Data collection

13 features

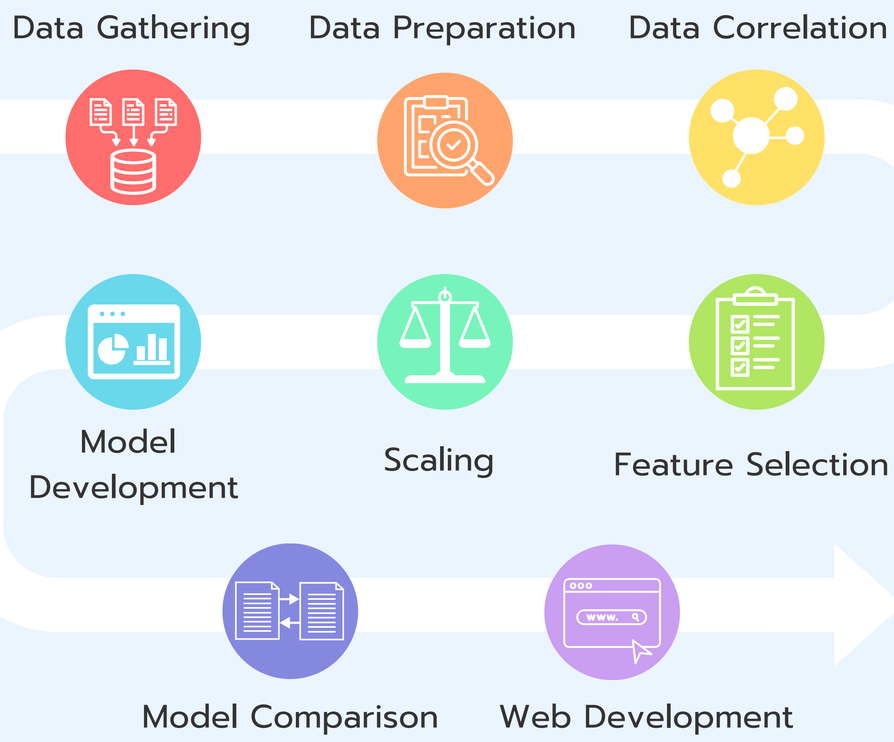
- Age
- Sex
- Rest ECG
- Chest Pain
- Max Heart Rate
- Exercise Angina
- Cholesterol (mg/dl)
- Fasting Blood Sugar
- Oldpeak
- ST Segment
- Fluoroscopy
- Thallium Stress
- Resting Blood Pressure

Class
- Heart Risk
- Nominal

Technology

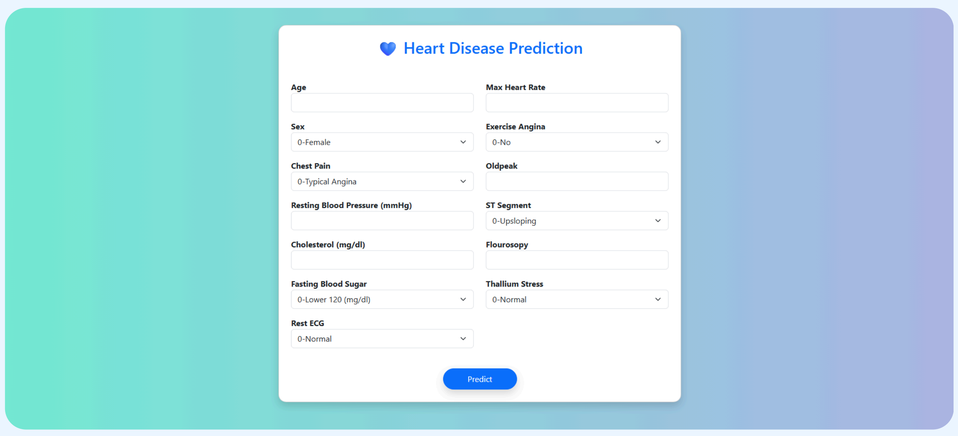


Methodology



Result

- Support Vector Machine (SVM) achieved the highest performance, with an F1-score of 90.95% and was implemented as a web application for practical use.
- K-Nearest Neighbor, Random Forest, and Logistic Regression achieved F1-scores of 89.33%, 89.08%, and 88.61%, respectively.
- Web application was developed using the SVM model to predict heart disease risk.



Conclusion

Machine learning can effectively predict the risk of heart disease using patient health data. Data preprocessing and model selection significantly improve predictive performance.

The developed model supports early screening and medical decision-making. This research can serve as a guideline for future studies. Future work may involve larger datasets and additional features to further enhance prediction accuracy.

Reference

[1]Amintra Chaipak , https://sixtygram.com/machine-learning/.
[2]safesiri , https://greedisgoods.com/correlation/.
[3]Yash S , https://intellipaat.com/blog/data-preprocessing/.

204499 Independent Study, Academic Year 2025
Department of Computer Science, Faculty of Science, Chiang Mai University