The Application Of Machine Learning In Predicting The Risk Of Heart Disease With Decision Tree Algorithm

Dicky Wahyudi¹, Ekatri Ayuningsih²
Politeknik LP3I Medan¹²

ARTICLE INFO

Article history:
Received : 20 May 2024
Revised : 18 June 2024
Accepted : 30 June 2024

Keywords:
Machine Learning,
Website,
Artificial Intelligence,
Decision Tree,
Classification

ABSTRACT

This research designs a web-based application that utilizes Machine Learning technology to predict the risk of possible heart disease with algorithms. This research aims to develop a web-based application that uses Machine Learning technology with the Decision Tree algorithm to predict the risk of heart disease. The research process begins with the collection of relevant medical data used in the training of Machine Learning models. Next, the web application is integrated with IBM Cloud services such as Watson Studio and IBM Cloud Object Storage to store user data and access configured Machine Learning algorithms for heart disease prediction. The data is processed and auto-configured using Machine Learning services from IBM Cloud, then the Machine Learning model is trained with Decision Tree algorithm using the processed dataset. As a result, a web application was successfully developed and integrated with IBM Cloud services, capable of providing heart disease prediction with sufficient accuracy. The application's simple and user-friendly interface allows users to easily access the heart disease prediction service without requiring in-depth technical knowledge. Thus, this research successfully created a solution that can assist in the early diagnosis of heart disease, provide direct benefits to the community, and expand access to heart health information widely. This application is built using the Flask framework and integrated with Machine Learning services from IBM Cloud. The use of the Decision Tree algorithm with an accuracy of 91.9% is based on a dataset that is processed and configured automatically using Machine Learning services from IBM Cloud.

This work is licensed under a Creative Commons Attribution 4.0 International License.

INTRODUCTION

Early detection of heart disease plays a crucial role in the prevention and management of the condition, which is the leading cause of death globally [1]. Machine Learning technology plays
an important role in the early detection of heart disease with its ability to analyze complex medical data quickly and accurately [2]. Using Machine Learning algorithms, individual information such as medical history and genetic factors can be incorporated into predictive models, enabling heart disease risk prediction tailored to each individual's characteristics. In addition, the development of predictive models based on historical patient data enables the system to provide more precise and personalized diagnosis and treatment recommendations. Thus, Machine Learning technology can improve early detection of heart disease by providing more accurate risk prediction, timely medical intervention, and overall effectiveness of heart disease prevention efforts [3].

This research will apply the concept of machine learning and decision tree algorithm to predict the risk of heart disease [4]. The data set used in this research comes from the CDC's annual survey data in 2022 of 400 thousand adults regarding their health status. This application is built using the Flask framework and integrated with Machine Learning services from IBM Cloud. The results of this research will display the results of the percentage probability of developing heart disease [5].

Developed a Machine Learning-based web application that can predict the risk of heart disease with a high level of accuracy [6]. Integration with IBM Cloud and Watson Studio services and the use of the Decision Tree algorithm in the Machine Learning model are the main focuses in this research [7]. With an application designed using Python Flask, HTML, and CSS, the goal of this research is to provide heart disease prediction services that are easily accessible to users without requiring in-depth technical knowledge. Through the use of Machine Learning technology, this research aims to provide a solution that can assist health practitioners in diagnosing heart disease early and provide quick and easy access for users regarding heart health information [8].

To solve a problem effectively, the steps that can be followed include clear identification of the problem, collection of relevant data, analysis of the data using methods such as Machine Learning, development of a solution based on the findings of the analysis, implementation of the solution, evaluation of the results, and monitoring the performance of the solution [9]. With this systematic approach, it is expected that problems can be solved effectively and efficiently, as well as ensuring that the solutions implemented have the desired impact in solving the problems at hand [10].

The main finding in this research is the development of a Machine Learning-based web application that is able to predict the risk of heart disease with a high level of accuracy using the Decision Tree algorithm. Integration with IBM Cloud services and Watson Studio allows the application to store and manage user media data and access Machine Learning algorithms that have been configured for heart disease prediction. The test results show that this web-based application can operate well in predicting heart disease using IBM services as a Machine Learning model. Thus, this research has successfully produced a solution that can assist health practitioners in diagnosing heart disease early and provide quick and easy access for users regarding heart health information.

The main conclusion of this research is that the development of Machine Learning-based web applications using the Decision Tree algorithm has been successful in predicting the risk of heart disease with a high level of accuracy [11]. Integration with IBM Cloud services and Watson Studio provides the ability to store user data and access Machine Learning algorithms that have been configured for heart disease prediction. This application provides direct benefits to the community by providing easily accessible and accurate heart disease prediction services, so that it can be used as a tool in early diagnosis and prevention of heart disease [12].

The recommendation for this research is to continue to develop and update this web application with the latest technology and data to improve the accuracy of heart disease
prediction. In addition, it is recommended to conduct further trials with larger samples to validate the accuracy of this application in various cases of heart disease. In addition, collaboration with health institutions and medical practitioners can help in integrating this application into clinical practice so that it can provide wider benefits to the community in efforts to prevent and manage heart disease [13].

METHODS

Data Collection
This research method, data collection is carried out by collecting relevant medical data which will be used in training Machine Learning models to predict the risk of heart disease. The medical data collected includes health information from 400 thousand adults in the CDC's annual survey in 2022. This data includes various factors that have the potential to influence the risk of heart disease, such as health history, lifestyle and genetic factors. This comprehensive data collection allows Machine Learning models to learn from a wide range of relevant information and improve the accuracy of heart disease risk predictions [3].

Integration With IBM Cloud Services And Watson Studio
The integration with IBM Cloud services and Watson Studio as well as the configuration of the Machine Learning algorithms used in this research played a key role in the development of the web application for predicting heart disease risk. By integrating the application with IBM Cloud services, especially Watson Studio and IBM Cloud Object Storage, the application can access the Machine Learning algorithms configured for heart disease prediction and store the necessary user data. The process of configuring the Machine Learning algorithms, particularly the use of the Decision Tree algorithm, allows the Machine Learning models to be trained with pre-processed datasets automatically, thus providing predictions with a high degree of accuracy. Thus, good integration with IBM Cloud services and proper configuration of Machine Learning algorithms enables the web application to provide accurate heart disease risk predictions that are useful in early diagnosis and prevention of heart disease [14].

Artificial Intelligence (AI)
Artificial Intelligence (AI), also known as artificial intelligence, is a branch of computer science that aims to develop systems and machines that can perform tasks that usually require human intelligence. The term Artificial intelligence was introduced by John McCarthy in 1956 at the Massachusetts Institute of Technology conference at the Dartmouth Conference attended by AI researchers in 1956. Artificial intelligence is a field of computer science that is indispensable in applying intelligent computers. Artificial intelligence emerged in the 1940s, although in ancient Egyptian times it can be known that this development existed. Artificial intelligence comes from the foreign language "Artificial Intelligent", meaning "intelligence" in Latin "Intelligo" which means "I understand". AI is a science that has several branches of knowledge, namely Natural Language Processing (NLP), Computer Vision, Navigation Systems and Robotics, Game Playing and Expert Systems. Three basic concepts of AI are Machine Learning, Deep Learning and Artificial Neural Networks. The purpose of studying and applying AI will be useful for humans, namely AI can be applied to programs or robots that can assist human activities as befits a human, it is hoped that with AI, the machine will become smarter than before and is expected to practically help humans solve complex problems such as intelligent calculators that can help humans do calculations quickly [15].
Machine Learning
Machine learning is a scientific field of artificial intelligence that deals with creating methods that can be programmed and learned from previous data. Machine learning, pattern recognition and data mining often refer to the same thing. The fields of probability and statistics and optimization depend as well. In data mining machine learning becomes an analytical tool. By definition, machine learning is the study of algorithms and statistical models used by computer systems to perform specific tasks without explicit instruction [16].

Algorithm Decision Tree
A Decision Tree is a data structure consisting of nodes and edges. Decision Tree algorithm is one of the machine learning algorithms used for classification and regression tasks. It works by dividing a dataset into smaller subsets based on certain features, until each subset has homogeneous data within a class or target value [17].

Python Programming
Python has become one of the most popular and widely used programming languages in various fields, including data analysis. Developed by Guido Van Rossum in 1990, Python offers flexibility and user-friendliness, making it the first choice for many data scientists, research and software developers [18].

Python Flask
Flask uses the python programming language and is part of the language's microframework suite. Flash itself has three dependencies, subsystems provided by Werkzeug, templates supported by Jinja2, and command-line integration with Clik. Phyton Flash was first released on April 1, 2010 by Armin Ronacher [19].

RESULTS AND DISCUSSION
The methods of collection data are used
1. Kaggle Dataset
   We will search and download heart disease-related datasets from the Kaggle platform. The selected dataset will contain relevant information such as the patient's medical history, risk factors, laboratory test results, and heart disease diagnosis. These datasets will be used as the basis for training and testing the prediction model.

2. Library research
   In addition to the Kaggle dataset, the research team will also utilize desk research to gain an in-depth understanding of heart disease risk factors, existing prediction techniques, and current approaches to prediction model development. The information obtained from the desk research will be used to enrich the understanding of heart disease and guide the model development process. After the data is collected, the following steps are to process and prepare the data before it is used in model training, namely data cleaning, normalization and dataset division.

Run Auto AI Experiment
After creating the IBM account and adding some services. On the new AutoAI experiment page, give a name to the AutoAI project. After creating the experiment, it will take you to the page to add the data source to the project. Click "Select from project" and then add the csv file click "Select Assets" to confirm the data source. After seeing that AutoAI is processing the data.
Next, explore the AutoAI settings by clicking "Experiment Settings". The following Machine Learning configuration definition can be seen in Figure 2 below.

![Figure 1. Machine Learning Configuration Defining View](image)

Figure 1. Machine Learning Configuration Defining View

To test the model, click the "Test" tab. Select a row from the data set and enter the data into the corresponding column, then click the "Predict" button at the bottom. The following Test model can be seen in Figure 3 below.

![Figure 2. Test Model](image)

Figure 2. Test Model

To validate the prediction, one can check the data file used to train the model. As seen, the model has predicted "I". For more information, it can also be seen in the Prediction Results feature which can be seen in Figure 4 below.
Figure 3. Prediction Results

The application has been successfully run and is ready for use. It can be seen the results of the screenshot by writing which contains Python and HTML code that will be made by writing into a prediction Flask Web that is integrated with IBM Service. The following Web Source Code Screenshot can be seen in Figure 4 below.

Figure 4. Source Code Screenshot Web Flask

The Web Flask interface that has been integrated into IBM Service uses Endpoints and API keys provided by IBM to connect Web Flask with services from IBM Cloud. The following
The hardware and software used in this research are laptop, IBM Watson Studio service, IBM Cloud Object Storage service, IBM Watson Machine Learning service, Flask, HTML, CSS, and JavaScript. The costs for IBM Watson Studio, IBM Cloud Object Storage and IBM Watson Machine Learning services may vary depending on the package and features selected as well as the subscription period (monthly or yearly). Internet access costs also depend on the service provider and the subscription plan chosen.

This study uses a dataset taken from the Kaggle Indicator of Heart Disease which is sourced from the CDC 2022 annual survey data related to the health status of more than 400 thousand adults, csv file type. Using several samples such as HeartDisease, BMI, Smoking, AlcoholDrinking, Stroke, PhysicalHealth, MentalHealth, DiffWalking, Sex, AgeCategory, Race, Diabetic, PhysicalActivity, GenHealth, SleepTime, Asthma, KidneyDiseas, and SkinCancer. The results of the dataset configuration with IBM on machine learning using the Snap Decision Tree Classifier algorithm with a high accuracy of 0.919 (holdout) or the percentage of truth reached 91.9% accuracy.
Integration of the web application with IBM services, particularly Watson Studio and IBM Cloud Object Storage, provides the ability to efficiently use machine learning technology to predict heart disease. These services allow the application to store and manage user media data and access machine learning algorithms that have been configured for heart disease prediction.

Figure 7. Resource List Penggunaan Service IBM

The web application designed using Python Flask, HTML, and CSS has the main function of predicting heart disease. With a simple and user-friendly interface, users can quickly access the heart disease prediction service without requiring in-depth technical knowledge.

Figure 8. Application of Coding
The test results show that this web-based application can operate properly in predicting heart disease using IBM service as a Machine Learning model.

Figure 9. Prediction Test Results on IBM

A pilot test was conducted to validate the performance and functionality of the web application in predicting heart disease. In this trial, the user enters medical data into the application form, and the application then performs heart disease prediction based on the data. The results of this trial show that the application is able to provide predictions with sufficient accuracy and can be used as a tool in early diagnosis and prevention of heart disease.

Figure 10. Web Application Test of Results

CONCLUSION

This research successfully produced a web-based application that uses Machine Learning technology with the Decision Tree algorithm to predict the risk of heart disease. This application can provide direct benefits to the community by providing easily accessible and accurate heart disease prediction services. Thus, this research contributes to efforts to early diagnosis and prevention of heart disease and expand access to heart health information to the wider community. This research uses the Decision Tree algorithm with an accuracy rate of
91.9%. With this high level of accuracy, this Machine Learning-based web application is able to predict the risk of heart disease with a good level of confidence.

REFERENCES


