

INTISARI

Penyakit jantung menjadi penyebab utama kematian di seluruh dunia, terutama menyerang kelompok usia produktif. Kemajuan teknologi telah mendorong perkembangan ilmu pengetahuan dan informasi dengan sangat pesat, mencakup berbagai bidang, termasuk dunia medis. Machine learning mampu mengolah data dan menghasilkan model yang dapat dimanfaatkan dalam berbagai aspek kehidupan. Tujuan dari penelitian ini yaitu untuk memvisualisasi dataset dan menentukan classifier terbaik untuk dapat mengidentifikasi hasil pada algoritma logistic regression, K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM). Dibatasi menggunakan metode pengembangan program melalui 3 tahapan yakni Preprocessing Dataset, Visualisasi Dataset, Klasifikasi Algoritma dengan menggunakan google colab. Klasifikasi machine learning untuk mendeteksi penyakit jantung dengan algoritma logistic regression, knn (k-nearest neighbours), dan svm (support vector machine) menghasilkan visualisasi data berupa grafik atribut penyakit jantung dan hasil dari klasifikasi algoritma menunjukkan bahwa model logistic regression memiliki accuracy sebesar 74%, presisi 1 sebesar 77%, presisi 0 sebesar 71%, recall 0 sebesar 67%, recall 1 sebesar 81%, Accuracy dari model K-Nearest Neighbors adalah 75%, presisi 0 78%, presisi 1 74%, recall 0 70%, recall 1 80% dan model Support Vector Machine (SVM) memiliki accuracy sebesar 75%, presisi 1 sebesar 72%, presisi 0 sebesar 80%, recall 0 sebesar 67%, recall 1 sebesar 84%.

Kata kunci : machine learning, Visualisasi Dataset, logistic regression, K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM)

ABSTRACT

Heart disease is the main cause of death throughout the world, especially affecting the productive age group. Technological advances have encouraged the development of science and information very rapidly, covering various fields, including the medical world. Machine learning is able to process data and produce models that can be used in various aspects of life. The aim of this research is to visualize datasets and determine the best classifier to be able to identify results in the logistic regression algorithm, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). Created using a program development method through 3 stages, namely Dataset Preprocessing, Dataset Visualization, Algorithm Classification using Google Collaboratory. Machine learning classification to detect heart disease using the logistic regression algorithm, knn (k-nearest neighbors), and svm (support vector machine) produces data visualization in the form of graphs of heart disease attributes and the results of the classification algorithm show that the logistic regression model has an accuracy of 74%, precision 1 of 77%, precision 0 of 71%, recall of 0 of 67%, recall of 1 of 81%, accuracy of the K-Nearest Neighbors model is 75%, precision 0 78%, precision 1 74%, recall 0 70%, recall 1 80% and the Support Vector Machine (SVM) model has an accuracy of 75%, precision 1 of 72%, precision 0 of 80%, recall of 0 of 67%, recall of 1 of 84%.

Keywords: machine learning, Dataset Visualization, logistic regression, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM)