ABSTRACT

EARLY DETECTION OF BREAST CANCER USING K-NEAREST NEIGHBOR (KNN) ALGORITHM AND DECISION TREE C-45

By:
Fahrurrozi
fahrurrozyghalib@gmail.com

Breast cancer is one of the types of cancer that has a significant impact on global health, especially in the female population, making early detection and disease management a priority. The purpose of the research was to compare the performance of the K-Nearest Neighbor (KNN) algorithm and Decision Tree-C45 in breast cancer classification and to evaluate the effectiveness of using Particle Swarm Optimization (PSO) in improving the performance of both algorithms. The data used was obtained from a dataset available on Kaggle, and this research employed the Knowledge Discovery in Database (KDD) method. The analysis was conducted using the RapidMiner tool, and the research results showed that the KNN model without PSO achieved an accuracy of 98.03%, but some positive instances went undetected. However, with the addition of PSO, there was a significant improvement in all evaluation metrics, making the KNN model with PSO the preferred choice with an accuracy of 99.78%, precision of 100%, and AUC of 1.000. Meanwhile, the Decision Tree-C45 model without PSO achieved an accuracy of 97.81%, but there was potential for improvement in some evaluation metrics. With the addition of PSO, Decision Tree-C45 showed a significant improvement in all evaluation metrics, with an accuracy of 99.34%, precision of 98.96%, and AUC of 0.997. Overall, this research emphasized the importance of comparing and optimizing algorithms in early breast cancer detection. The use of Particle Swarm Optimization (PSO) proved to significantly contribute to improving the performance of classification models. These results provided valuable insights into the development of more effective and efficient early breast cancer detection techniques.

Keywords: Breast Cancer, KNN, Decision Tree-C45, Particle