

ABSTRACT

Banyaknya data yang dimiliki oleh suatu sektor akan menjadi menumpuk jika tidak digunakan dengan tepat. Oleh karena itu data tersebut harus diolah dalam skala besar. Maka diperlukan suatu Teknik yang disebut data mining agar menjadi sebuah pengetahuan baru. Data mining dapat mengolah suatu data menjadi suatu pengetahuan dalam skala besar dengan cepat menggunakan metode klasifikasi, clustering dll. Bidang Kesehatan menjadi sector yang paling sering menggunakan hal tersebut terutama dalam mendiagnosa pasien. Penelitian ini akan membahas tentang kinerja dari tiga algoritma klasifikasi yaitu K-Nearest Neighbor, Decision Tree C4.5 dan Naïve Bayes. Tujuan dalam penelitian ini adalah membandingkan performa dari ketiga algoritma tersebut untuk memperoleh performa terbaik yang pada akhirnya akan digunakan untuk memprediksi suatu penyakit. Hasil yang didapatkan dari penelitian ini yaitu performa dari algoritma Decision tree C4.5 menjadi algoritma yang memiliki nilai overall accuracy terbaik dibandingkan algoritma K-Nearest Neighbor dan Naïve Bayes dengan nilai Overall Accuracy 100%. Algoritma Naïve Bayes menempati peringkat kedua terbaik dengan nilai accuracy 96.43% dan K-Nearest Neighbor menempati peringkat terakhir dengan nilai overall accuracy 64.29%. untuk perhitungan Accuracy menggunakan Kappa yaitu Overall Accuracy karena dataset yang di uji menggunakan multi-class.

ABSTRACT

The amount of data owned by a sector will pile up if it is not used properly. Therefore, the data must be processed on a large scale. Then we need a technique called data mining to become a new knowledge. Data mining can quickly process data into knowledge on a large scale using classification methods, clustering, etc. The Health sector is the sector that uses this most often, especially in diagnosing patients. This study will discuss the performance of three classification algorithms: K-Nearest Neighbor, Decision Tree C4.5, and Naïve Bayes. This study aims to compare the performance of the three algorithms to obtain the best performance, which will ultimately be used to predict disease. The results obtained from this study are the performance of the Decision tree C4.5 algorithm to be an algorithm with the best overall accuracy value compared to the K-Nearest Neighbor and Naïve Bayes algorithms with an Overall Accuracy value of 100%. The Naïve Bayes algorithm is ranked second best with an accuracy value of 96.43%, and K-Nearest Neighbors is ranked last with an overall accuracy value of 64.29%. The calculation of Accuracy using Kappa, namely Overall Accuracy, is because the dataset being tested uses multi-class.