

ABSTRACT

One form of utilization of information technology in health is the presence of electronic-based medical records. The presence of electronic-based medical record facilitates the storage of medical records data in effective and efficient manner. The growth of medical records data largely stored on the database has created a condition which is called ‘rich of data but poor of knowledge’.

Large sets of medical record data are yet to be used properly, making them into a collection of useless data. In carrying out the analysis of large medical records stored in the database, data mining technique is usually used. One of data mining tasks that can be used for medical records data extraction is classification. One classification method used in this research was Correlated Naive Bayes Classifier (C-NBC) algorithm and Naive Bayes Classifier (NBC).

The objective of this research is to confirm whether the accuracy level of Correlated Naive Bayes Classifier algorithm is higher than that of Naive Bayes Classifier algorithm in health, by using four datasets, namely Pima Indians Diabetes, Haberman, Indian Liver Patient, and Tyroid datasets. Based on the testing conducted, it shows that the accuracy level of Correlated Naive Bayes Classifier (C-NBC) algorithm obtains the best result, compared to Naive Bayes Classifier (NBC) algorithm on four datasets used. The accuracy level of Naive Bayes Classifier (C-NBC) algorithm on Pima Indians Diabetes dataset obtains 67.15%, the accuracy level of Correlated Naive Bayes Classifier (C-NBC) algorithm on Haberman dataset obtains 74.96%, the accuracy level of Naive Bayes Classifier (C-NBC) algorithm on Indian Liver patient dataset obtains 71.37%, and the accuracy level of Correlated Naive Bayes Classifier (C-NBC) algorithm on Tyroid dataset obtains 78.52%.

Keywords – C-NBC, NBC, Data Mining.

INTISARI

Salah satu pemanfaatan teknologi informasi dibidang kesehatan adalah adanya rekam medis berbasis elektronik. Dengan adanya rekam medis berbasis elektronik memudahkan penyimpanan data rekam medis secara efektif dan efisien. Pertumbuhan data rekam medis yang begitu besar yang tersimpan pada *database* telah menciptakan kondisi yang disebut *rich of data but poor of knowledge*.

Kumpulan data rekam medis yang begitu besar belum dimanfaatkan dengan sebaiknya, sehingga menjadi kumpulan data yang tidak berguna. Untuk melakukan analisa data rekam medis dengan jumlah yang besar tersimpan pada *database*, biasanya dapat digunakan teknik data mining. Salah satu tugas data mining yang bisa digunakan untuk ekstraksi data rekam medis adalah klasifikasi. Salah satu metode klasifikasi yang digunakan dalam penelitian ini adalah algoritme *Correlated Naive Bayes Classifier* (C-NBC) dan *Naive Bayes Classifier* (NBC).

Adapun tujuan dari peneilitian ini adalah melakukan konfirmasi apakah tingkat akurasi algoritme *Correlated Naive Bayes Classifier* lebih tinggi dibandingkan tingkat akurasi algoritme *Naive Bayes Classifier* di bidang kesehatan dengan menggunakan empat dataset yaitu Dataset *Pima Indian Diabetes*, Dataset *Haberman*, Dataset *Indian Liver Patient*, dan Dataset *Tyroid*. Berdasarkan pengujian yang telah dilakukan menunjukkan bahwa tingkat akurasi algoritme *Correlated Naive Bayes Classifier* (C-NBC) memperoleh hasil yang terbaik dibandingkan algoritme *Naive Bayes Classifier* (NBC) terhadap empat dataset yang digunakan. Tingkat akurasi algoritme *Correlated Naive Bayes Classifier* (C-NBC) pada Dataset *Pima Indian Diabetes* diperoleh sebesar 67,15%, tingkat akurasi algoritme *Correlated Naive Bayes Classifier* (C-NBC) pada dataset *Haberman* diperoleh sebesar 74,96%, tingkat akurasi algoritme *Correlated Naive Bayes Classifier* (C-NBC) pada Dataset *Indian Liver Patient* diperoleh sebesar 71,37 %, dan tingkat akurasi algoritme *Correlated Naive Bayes Classifier* (C-NBC) pada Dataset *Tyroid* diperoleh sebesar 78,52%.

Kata Kunci – C-NBC, NBC, Data Mining.