

ABSTRAK

Banyak penelitian yang melakukan *preprocessing* dengan menangani *missing data*. Akan tetapi sedikit sekali penelitian yang khusus membahas tentang *handling missing value*. *Missing value* hal biasa yang terjadi pada dataset dan kalau tidak ditangani secara maksimal akan mengakibatkan kurang signifikan pada akurasi hasil pengujian dan bahkan akan mengakibatkan hilangnya informasi berharga dari dataset yang ada. Tujuan dari penelitian ini adalah untuk mengetahui teknik *sequential methods* terbaik untuk menangani *missing value* yang diterapkan dengan algoritma *C4.5* dengan algoritma *Naïve Bayes* untuk mendapatkan akurasi terbaik dan signifikan dengan memperhatikan keterkaitan antara atribut yang satu dengan yang lain tanpa harus kehilangan informasi yang berharga dari dataset yang ada sehingga dataset tetap mencerminkan kondisi yang sebenarnya. Adapun manfaat dari penelitian ini yaitu dapat memberikan kontribusi dalam keilmuan terutama dalam bidang *data mining* khususnya untuk penanganan *missing value* dengan melakukan komparasi *sequential methods* pada *preprocessing data* sehingga akan dihasilkan akurasi yang baik dan signifikan dan tetap mencerminkan kondisi yang sebenarnya. Penelitian ini akan melakukan komparasi metode *handling missing value*. Sedangkan metode yang akan digunakan adalah *sequential methods*. *Sequential methods* adalah teknik untuk menangani *missing data* khususnya untuk *preprocessing*. *Assigning All Possible Values to a Missing Attribute Value* dan *Assigning All Possible Attribute Values Restricted to a Concept* mampu menghasilkan akurasi yang lebih baik dibanding teknik lainnya apabila diterapkan ke dalam algoritma *C4.5* dan *Naïve Bayes*. Algoritma *C4.5* menghasilkan akurasi sebesar 75.55 % dan *AUC* sebesar 0.814, sedangkan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 76.79 % dan *AUC* sebesar 0.836. Dari hasil penelitian di atas tidak mutlak yang terbaik, penelitian ini bisa dicoba algoritma klasifikasi yang lain, misalnya *Logistic Regression*, *K-Nearest Neighbors*, *Support Vector Machine* dan lain-lain atau dicoba dengan dataset yang lain.

Kata kunci : *Diabetes Mellitus*; *Handling Missing Value*; *Sequential Methods*, *C4.5*, *Naïve Bayes*

xii + 71 halaman; 5 gambar; 11 tabel; 8 lampiran

Bibliography: 23 (2005-2013)