

Abstrak

Missing value merupakan salah satu permasalahan yang umum terjadi dalam data. *Missing value* merupakan nilai data yang hilang. Penanganan *missing value* merupakan salah satu *task* dalam *data preprocessing*, bertujuan supaya data menjadi lebih berkualitas untuk dijadikan sebagai inputan bagi proses data *mining*. Salah satu cara penanganannya adalah dengan imputasi (*imputation*). Imputasi adalah yang secara otomatis akan mengganti *missing value* di dalam data dengan sejumlah nilai yang *plausible*.

Pada Tugas Akhir ini, penulis mengimplementasikan salah satu metode imputasi yaitu *Robust Least Squares Imputation with Principal Component* (RLSP). Tahapan proses imputasi dengan metode RLSP meliputi pemilihan *k nearest instance*, proses pembentukan *principal component* dan imputasi dengan menggunakan regresi kuantil berdasarkan *principal component* yang diperoleh pada tahap sebelumnya. Proses ini menghasilkan data hasil imputasi dan performansi dalam parameter *Normalized Root Mean Squared Error* (NRMSE).

Pengujian terhadap sistem imputasi dilakukan secara langsung berdasarkan hasil NRMSE dan tidak langsung melalui proses klasifikasi berdasarkan parameter *precision* dan *recall*. Proses klasifikasi dilakukan dengan *Naïve Bayes classifier* menggunakan *tools* WEKA. Tujuan dari pengujian tidak langsung ini adalah untuk mengetahui seberapa baik kualitas data hasil imputasi jika digunakan dalam proses data *mining* yaitu klasifikasi.

Kata kunci : *missing value*, imputasi, RLSP, NRMSE, klasifikasi, *Naïve Bayes*