Abstrak

Missing value merupakan kasus hilangnya nilai dari suatu data yang dapat terjadi pada suatu dataset. Ketika mencapai rate yang tinggi dari suatu dataset maka diperlukan penanganan lebih lanjut terhadap adanya kasus missing value tersebut. Penanganan yang dapat dilakukan adalah dengan melakukan imputasi, di mana nilai yang hilang dari suatu data diisi dengan suatu nilai tertentu. Salah satu metode imputasi yang dapat digunakan adalah Metode Hot Deck. Pada metode ini, nilai dari sebuah record pada dataset yang sama akan diambil untuk mengisi nilai pada record yang mengalami missing value. Pada tugas akhir ini, penentuan record vang menjadi pemberi nilai dilakukan dengan fungsi jarak Chebyshev. Squared Euclidean, Canberra, serta Random Hot Deck untuk dataset numerik. Sedangkan, pada dataset kategoris digunakan Hot Deck berbasis jarak yang berbeda. Pengukuran performansi dilakukan dengan mengukur RMSE pada dataset numerik serta mengukur precision, recall, dan f-measure pada dataset kategoris. Dari hasil pengujian yang dilakukan pada dataset numerik, Hot Deck berbasis Squared Euclidean tampak memberikan akurasi yang paling stabil, sedangkan Random Hot Deck menampilkan akurasi yang paling buruk. Pada dataset kategoris, akurasi yang dihasilkan oleh Random Hot Deck tidak mengahsilkan perbedaan yang signifikan dengan akurasi yang dihasilkan oleh Hot Deck berbasis jarak.

Kata Kunci: *missing value*, imputasi, hot deck.