

ABSTRAK

IMPLEMENTASI *K-NEAREST NEIGHBORS* (KNN) DAN FUNGSI KEANGGOTAAN FUZZY UNTUK IMPUTASI *MISSING VALUES* PADA KLASIFIKASI DATA

Oleh

REVITA SETIANINGSIH

Missing values merupakan suatu keadaan dimana terdapat nilai yang kosong atau hilang dalam sebuah dataset. Nilai yang hilang ini dapat terjadi karena kesalahan responden yang tidak menjawab semua pertanyaan dalam kuesioner, pengukuran dan eksperimen yang salah ketika input data secara manual, adanya beberapa data yang disensor, dan lainnya. Terdapat 3 cara untuk mengatasi *missing values* yaitu *listwise deletion*, *pairwise deletion*, dan imputasi. Pada penelitian ini dilakukan imputasi menggunakan KNN dan fungsi keanggotaan fuzzy untuk mengetahui apakah metode ini dapat menangani *missing values* dengan lebih baik dibanding dengan metode imputasi konvensional.

Fungsi keanggotaan fuzzy segitiga dan Gaussian digunakan dalam algoritma imputasi data untuk menghitung bobot nilai data tetangga terdekat sebelum digunakan selama proses imputasi. Adapun data penelitian pada penelitian ini menggunakan jenis data klasifikasi yang diperoleh dari *UCI Machine Learning Repository* yang berjumlah 5 dataset. Hasil yang diperoleh menunjukkan bahwa fungsi keanggotaan fuzzy memiliki nilai akurasi yang lebih tinggi dibandingkan dengan imputasi konvensional dimana fungsi keanggotaan fuzzy segitiga lebih unggul dari fungsi keanggotaan fuzzy Gaussian berdasarkan rata-rata akurasinya. Sehingga, fungsi keanggotaan fuzzy lebih baik untuk menangani *missing values* dalam meningkatkan akurasi.

Kata kunci: *Missing Values*, *K-Nearest Neighbor* (KNN), Fungsi Keanggotaan Fuzzy, Logika Fuzzy

ABSTRACT

IMPLEMENTATION OF K-NEAREST NEIGHBORS (KNN) AND FUZZY MEMBERSHIP FUNCTION FOR IMPUTATION OF MISSING VALUES IN DATA CLASSIFICATION

By

REVITA SETIANINGSIH

A missing value is a condition where there are any values that blank or missing in the dataset. This missing value can occur due to respondents' errors in not answering all questions in the questionnaire, incorrect measurements and experiments when entering data manually, the presence of some censored data, and others. There are 3 ways to overcome missing values, namely listwise deletion, pairwise deletion, and imputation. In this research, imputation was carried out using KNN and fuzzy membership functions to find out whether this method can handle missing values better than conventional imputation methods.

The triangular and Gaussian fuzzy membership functions are used in data imputation algorithms to calculate the weight for the data values of the nearest neighbor's before they are used during the imputation process. The research data in this study used the type of classification data obtained from the UCI Machine Learning Repository which amounted to 5 datasets. The results obtained show that the fuzzy membership function has a higher accuracy value compared to conventional imputation where the triangular fuzzy membership function is superior to the Gaussian fuzzy membership function based on its average accuracy. Thus, the fuzzy membership function is better for handling missing values in improving accuracy.

Keywords: Missing Values, K-Nearest Neighbor (KNN), Fuzzy Membership Function, Fuzzy Logic