

SELEKSI FITUR *GAIN RATIO* DAN MODEL KLASIFIKASI *K-NEAREST NEIGHBOR* PADA DIAGNOSIS PENYAKIT STROKE

Juhariyanto
(150411100017)

Pembimbing I : Mula'ab, S.Si., M.Kom.

Pembimbing II : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRAK

Kesehatan merupakan hal penting bagi kehidupan, terutama bagi manusia. Banyak penyakit yang menyerang manusia, salah satunya yaitu penyakit stroke. Penyakit stroke merupakan suatu gangguan fungsi sistem saraf yang dapat terjadi secara mendadak, disebabkan oleh gangguan pembuluh darah di otak. Banyaknya orang yang terkena stroke membuat dokter kesusahan untuk mendiagnosis pasien yang konsultasi maupun yang terkena stroke. Klasifikasi penyakit stroke secara cepat dan tepat dapat membantu dokter untuk mendiagnosis penyakit stroke. Klasifikasi penyakit stroke akan memprediksi penyakit yang diderita oleh pasien secara efektif dan efisien. Pada penelitian ini menggunakan model klasifikasi *K-Nearest Neighbor(KNN)* dengan seleksi fitur *Gain Ratio*. Metode *KNN* digunakan untuk klasifikasi penyakit stroke berdasarkan nilai *k* tetangga yang mana akan dihitung jarak antara data latih dan data uji, sedangkan *Gain Ratio* digunakan untuk meningkatkan akurasi dari metode *KNN* dengan cara mereduksi dimensi pada fitur. Data set yang digunakan yaitu data penyakit stroke dari situs <https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data> dengan jumlah data 43400 yang mana terdiri 10 fitur. Data dipencah menggunakan *K-fold Cross Validation* dengan menggunakan *k=10*. Hasil yang didapatkan dengan menggunakan metode seleksi fitur *Gain Ratio* dan model klasifikasi *KNN* yaitu dengan jumlah fitur = 1 dan *KNN(k=3)* yang menghasilkan akurasi 1, *recall* 1, *precision* 1, *error* 0, dan waktu komputasi selama 228 detik.

Kata Kunci : Stroke, Klasifikasi, *K-Nearest Neighbor*, *Gain Ratio*

GAIN RATIO FEATURE SELECTION AND CLASSIFICATION MODEL OF K-NEAREST NEIGHBOR IN STROKE DISEASE DIAGNOSIS

Juhariyanto
(150411100017)

Pembimbing I : Mula'ab, S.Si., M.Kom.

Pembimbing II : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRACT

Health is important for human being. Many disease attack human, one of which is a stroke. Stroke is a nervous system function that can occur suddenly, caused by a blood vessel disorder in the brain. A lot of people affected by a stroke . Quick and precise classification of stroke can help doctors to diagnose a stroke. The classification of stroke will predict the disease suffered by patients effectively and efficiently. This study used classification model of KNN and Gain Ratio feature selection. The KNN method is used to classify based on neighbor's k value, which calculated the distance between training data and test data, while the Gain Ratio is used to increase the akurasi of the KNN method by reducing the feature dimensions. The data set used is the stroke data from site <https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data> with 43.400 data which consist of 10 features. Data is split using K-fold Cross Validation by $K=10$. The results obtained using the Gain Ratio feature selection method and the classification model of knn are number of feature = 1 and knn($k=3$) which produce accuration 1, recall 1, precision 1, error 0, and computing time for 228 seconds.

Keywords : Stroke, Classification, K-Nearest Neighbor, Gain Ratio