

INTISARI

IMPLEMENTASI K-MEANS UNTUK KLASTERISASI PENYAKIT STROKE BERDASARKAN GEJALA DAN FAKTOR RISIKO

Oleh

ECHA DAMAYANTI

damayantiecha8@gmail.com

Stroke merupakan salah satu penyakit yang memiliki tingkat kematian dan kecacatan tinggi di seluruh dunia. Hal ini memerlukan penanganan secara khusus untuk mendeteksi secara dini. Jika hal ini tidak diatasi akan berdampak pada peningkatan kematian yang lebih tinggi. Deteksi dini dan pemahaman terhadap faktor risiko menjadi hal krusial dalam upaya pencegahan. Penelitian ini bertujuan untuk mengelompokkan pasien stroke berdasarkan gejala dan faktor risiko menggunakan algoritma K-Means Clustering. Dataset yang digunakan berasal dari platform Kaggle dengan total 5.110 data, mencakup atribut seperti usia, hipertensi, penyakit jantung, kadar glukosa, BMI, dan status merokok. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan (data cleaning, encoding, dan seleksi fitur), penerapan algoritma K-Means, serta evaluasi kluster dengan metode Elbow untuk menentukan jumlah kluster optimal. Hasil analisis menunjukkan bahwa nilai k terbaik adalah 4, dengan masing-masing kluster merepresentasikan tingkat risiko stroke yang berbeda, mulai dari rendah hingga tinggi. Kluster 0 dan 1 menunjukkan individu dengan risiko tinggi, Kluster 2 risiko sedang, dan Kluster 3 risiko rendah. Temuan ini menunjukkan bahwa K-Means dapat mengidentifikasi pola tersembunyi dalam data pasien dan memberikan landasan awal dalam penanganan preventif stroke secara lebih terarah. Implementasi ini juga mendukung pengambilan keputusan medis berbasis data untuk peningkatan layanan kesehatan.

Kata Kunci: *Stroke, K-Means Clustering, Data Mining, RapidMiner, Elbow Method.*