

## INTISARI

### **Implementasi Metode *K-Prototypes Clustering* untuk Mengatasi Data Campuran**

(Studi Kasus : Pasien di Klinik Medis Bangladesh Berdasarkan Faktor Risiko Stroke)

Daffa Alya Alodia

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Islam Indonesia

Stroke adalah suatu penyakit defisit neurologis yang disebabkan oleh perdarahan ataupun sumbatan dengan gejala dan tanda yang sesuai pada bagian otak yang terkena, yang dapat menimbulkan cacat atau kematian. Di Bangladesh, stroke merupakan salah satu penyebab utama kematian dan kecacatan. Prevalensi stroke di Bangladesh sebesar 1-2% dari penduduk yang berusia di atas 20 tahun. Tingginya angka kematian di Bangladesh yang disebabkan oleh stroke dipengaruhi oleh tingginya frekuensi faktor risiko aterosklerotik konvensional, terutama hipertensi dan diabetes melitus. Kematian akibat stroke mungkin terkait dengan tingkat keparahan stroke, keterlambatan diagnosis, dan kesenjangan perawatan stroke. Berdasarkan kondisi tersebut penting dilakukan pengelompokan pasien di klinik Bangladesh untuk mengetahui karakteristik risiko terkena penyakit stroke agar dapat dilakukan pencegahan. Data yang digunakan adalah data pasien di klinik medis Bangladesh yang diperoleh melalui *website Kaggle*. Data tersebut terdiri dari 11 variabel yaitu, jenis kelamin, usia, glukosa, hipertensi, *heart disease*, status pernikahan, jenis pekerjaan, jenis rumah, *body mass index* (BMI), status merokok, dan stroke. Pengelompokan dilakukan dengan menggunakan metode *K-Prototypes Clustering* yang merupakan perkembangan dari metode *K-Means* untuk menangani data campuran. Dari hasil analisis didapatkan dua *cluster* optimal. *Cluster* pertama terdiri dari 763 pasien, sedangkan *cluster 2* terdiri dari 4374 pasien dari klinik medis Bangladesh yang memiliki similaritas atau kemiripan karakteristik yang tinggi.

**Kata Kunci :** *Clustering, K-Prototypes, Stroke.*