

INTISARI

IMPLEMENTASI METODE *K-NEAREST NEIGHBOR* (KNN) DENGAN JARAK *HETEROGENEOUS EUCLIDEAN-OVERLAP METRIC* (HEOM) PADA KLASIFIKASI PENYAKIT KULIT

(Studi Kasus : Klasifikasi Penyakit Kulit Berdasarkan *Dermatology Data Set*
di *UC Irvine Machine Learning Repository*)

Hyra Maya Ayunindya

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

Penyakit kulit merupakan penyakit yang sering kali muncul dan dianggap sepele bagi sebagian orang apalagi jika dibandingkan dengan penyakit yang beresiko tinggi menyebabkan kematian, seperti kanker, HIV/AIDS, jantung, stroke dan lain sebagainya. Proses awal diagnosis adalah proses yang sangat penting dimana dokter akan menganalisis faktor-faktor sebelum menentukan diagnosis yang umumnya menjadi masalah yang sulit. Oleh sebab itu penelitian ini bertujuan untuk melakukan analisis klasifikasi jenis penyakit kulit menggunakan metode *K-Nearest Neighbor* (KNN). Data yang digunakan dalam penelitian ini adalah *dataset dermatology* yang diambil dari *website UC Irvine Machine Learning Repository*. Dikarenakan data yang digunakan pada penelitian ini adalah data dengan tipe data kategorik dan tipe data numerik, maka perhitungan jarak dapat dihitung dengan perhitungan jarak *Heterogeneous Euclidean-Overlap Metric* (HEOM). Dari data yang diperoleh, jumlah data pada setiap kelasnya memiliki selisih yang cukup besar yang mengindikasikan adanya ketidakseimbangan data (*data imbalance*), sehingga dilakukan *oversampling* menggunakan *Synthetic Minority Oversampling Technique* (SMOTE). Klasifikasi KNN dengan nilai k sebesar 3 sampai 13, didapatkan hasil bahwa model dengan performa terbaik adalah pada $k = 7$ dengan nilai *accuracy* sebesar 88%, nilai *recall* sebesar 0.86, nilai *precision* sebesar 0.92, dan nilai *G-mean* sebesar 0.85.

Kata Kunci: Penyakit kulit, SMOTE, KNN, HEOM