

## INTISARI

### **KLASIFIKASI PENYAKIT KANKER SERVIKS MENGGUNAKAN ALGORITMA FORWARD SELECTION SELECTION DAN K-NEAREST NEIGHBOR (KNN)**

**OLEH:  
DEDI ARBAIN  
2221210048**

Pengumpulan rekam medis masa lalu pasien kanker serviks di fasilitas kesehatan tidak dibarengi dengan proses ekstraksi pengetahuan dan informasi. Penggunaan teknik data mining dapat di implementasikan ke dalam sistem yang dapat memprediksi kanker serviks. Penelitian ini berfokus pada kumpulan data diagnosis medis pasien yang menjalani pap smear. Dataset terdiri dari 36 parameter variabel informasi, kebiasaan, dan rekam medis dari 858 pasien. Beberapa pasien mungkin menolak menjawab beberapa pertanyaan karena alasan privasi (nilai hilang). Dataset ini berisi empat variabel target tipe Boolean: Biopsy, Citology, Hinselmann dan Schiler. Setiap targetnya adalah data tentang hasil tes pasien, seperti terdeteksi adanya sel kanker di leher rahim pasien. Algoritma yang digunakan untuk klasifikasi kanker serviks adalah Forward Selection dan k-nearest neighbours (k-NN). Pengujian dilakukan terhadap algoritma Forward Selection dan k-nearest neighbours (k-NN) dengan menggunakan rumus matriks Confusion. Hasil pengujian terhadap algoritma yang digunakan sebanyak menunjukkan bahwa algoritma Forward Selection dan k-nearest neighbours (k-NN) mempunyai akurasi tertinggi sebesar 97,35% yang menggunakan variabel Hinselmann. Hal ini menunjukkan bahwa kinerja yang dicapai oleh setiap algoritma yang digunakan memungkinkan sistem prediksi kanker serviks digunakan untuk mendukung keputusan klinis pada pasien baru.

**Kata Kunci: Kanker Serviks, Forward Selection, K-NN, Klasifikasi, Performa, Matriks Confusion**

## ABSTRACT

# CLASSIFICATION OF CERVICAL CANCER USING ALGORITHM FORWARD SELECTION AND K-NEAREST NEIGHBOR (KNN)

By:  
DEDI ARBAIN

The collection of past medical records of cervical cancer patients in health facilities was not accompanied by a process of extracting knowledge and information. The use of data techniques mining could be implemented into a system that can predict cervical cancer. The research focused on collecting data on medical diagnoses of patients undergoing pap smears. The dataset consisted of 36 variable parameters of information, habits, and medical records of 858 patients. Some patients may refuse to answer some questions because of privacy reasons (missing values). This dataset contains four Boolean-type target variables: Biopsy, Cytology, Hinselmann, and Schiller. Each target is data about the patient's test results, such as detecting cancer cells in the patient's cervix. The algorithm used for Cervical Cancer Classification is Forward Selection and k-nearest neighbors (K-NN). The testing was carried out on the Forward Selection and k-nearest neighbors (K-NN) algorithms by using the Confusion matrix formula. Test results for the algorithm were used as much to show that the Forward Selection and K-nearest algorithm neighbors (K-NN) have the highest accuracy of 97.35% using Hinselmann variables. This showed that the performance achieved by each algorithm used allowed the cervical cancer prediction system to be used to support clinical decisions in new patients.

**Keywords:** Cervical Cancer, Forward Selection, K-NN, Classification, Performance, Matrix Confusion.