

ABSTRAK

Kanker serviks merupakan penyakit yang menyerang leher rahim wanita. Kanker serviks adalah penyebab kanker terbesar nomor dua bagi wanita di seluruh dunia. Menurut data terkini dari WHO (Organisasi Kesehatan Dunia) setiap tahunnya 500,000 wanita diseluruh dunia didiagnosa mengidap kanker serviks dan sekitar 300,000 meninggal karena kanker serviks. Angka kematian pada wanita diakibatkan kanker serviks selalu meningkat dari tahun ke tahun. Tingkat kematian wanita dengan kanker serviks sangatlah tinggi dan hampir 60% diantaranya datang dari kaum-kaum yang menengah dan menengah kebawah dikarenakan individu tidak memiliki dana untuk melakukan pengecekan atau deteksi dini.

Hinselmann, Schiller, Citology, dan Biopsy merupakan empat teknik *screening* untuk mendiagnosa ada tidaknya sel kanker pada leher rahim wanita. Pada penelitian ini dataset rekam medis pasien kanker akan dianalisa dengan menggunakan algoritma *Random Forest Tree* dan *KNN*. Kedua algoritma kemudian dibandingkan antara satu sama lain untuk menemukan model yang paling tepat diterapkan untuk mengetahui pola penderita kanker serviks dan memprediksi hasil *screening* dari pasien yang terduga kanker serviks.

Hasil dari penelitian ini yang diolah dengan menggunakan bahasa pemograman Python untuk 214 data uji dari 854 total data mendapatkan angka akurasi akhir 88.7% untuk algoritma *Random Forest* dan 90.6% untuk algoritma *KNN*. Dataset yang dipakai mempunyai empat target prediksi sehingga tergolong sebagai klasifikasi multilabel. *KNN* terbukti unggul dalam memprediksi klasifikasi multilabel dalam kasus deteksi pola pasien kanker serviks.

Kata kunci: Kanker Serviks, Klasifikasi, Random Forest Tree, KNN

ABSTRACT

Cervical cancer attacks the cell on patient cervix. Cervical cancer is ranked two as the cause of woman cancer in the whole world. From WHO (World Health Organization) data, 500.000 woman each year diagnosed with cervical cancer and 300.000 of those died. The mortality rate of cervical cancer is increasing throughout the year. The mortality rate of said disease can reach up to 60% and most of them are petitbourgeois or lower cause they can't afford to get early diagnosed.

Hinselmann, Schiller, Citology, and Biopsy are four screening technics to diagnose whether there is cancer cell on patient cervix. In this research, the patient medical history dataset will be analyzed by using Random Forest Tree and KNN algorithm. Both of the algorithms will be compared to find the most accurate model to be implemented in order to figure out the patterns on cervical cancer patient and predicting the patient screening result whether it's positive cervical cancer or negative.

The result of this research is processed by using programming code Python for 214 test data from 854 total data. The accuracy at the end is shown 88.7% for Random Forest and 90.6% for KNN. Dataset used have four target classification which is a multilabel classification. KNN was proven to be more advance on its prediction of multilabel classification in detecting the patterns of cervical cancer patient's case.

Keywords: Cervical Cancer, Classification, Random Forest Tree, KNN