

## INTISARI

# KLASIFIKASI PERKEMBANGAN KANKER SERVIKS MENGGUNAKAN *SUPPORT VECTOR MECHINE (SVM) DENGAN SMOTE DAN ADASYN*

(Studi Kasus : Perkembangan Kanker Serviks Pada Data GSE63514)

MAHARANI PURWASI

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

Berdasarkan data tahun 2012 yang bersumber dari *International Agency for Research on Cancer* salah satu jenis kanker yang paling banyak diderita perempuan adalah kanker serviks. Setiap tahun sekitar setengah juta perempuan di seluruh dunia didiagnosis menderita kanker serviks. Kanker serviks adalah kanker ganas yang dimulai di leher rahim dengan pertumbuhan sel yang tidak teratur, dan sel tersebut dapat menyerang organ biologis lainnya, baik dengan tumbuh langsung di jaringan yang berdekatan atau dengan berpindah ke sel lain. Seiring dengan kemajuan teknologi, dikembangkanlah bidang pengetahuan baru yaitu Bioinformatika. Penerapan Bioinformatika adalah penggunaan metode komputasi, matematika dan statistik untuk membantu memecahkan masalah biologis melalui analisis data ekspresi gen. Penelitian ini menggunakan data klasifikasi Ekspresi Gen perkembangan kanker serviks dengan kode series GSE63514. Dalam penelitian ini dilakukan analisis klasifikasi dengan metode SVM. Data yang digunakan terdapat ketidakseimbangan kelas data, untuk mengatasi permasalahan tersebut dilakukan *resampling* dengan *oversampling Synthetic Minority Oversampling Technique (SMOTE)* dan *Adaptive Synthetic (ADASYN)*. Penelitian ini menggunakan analisis klasifikasi metode SVM dengan 4 kernel yaitu kernel linear, kernel *polynomial*, kernel sigmoid dan kernel RBF. Untuk mendapatkan nilai akurasi terbaik dilakukanlah perbandingan antara nilai akurasi dari klasifikasi SVM data tidak seimbang, SVM dengan *oversampling SMOTE* dan SVM dengan *oversampling ADASYN*. Dari penelitian ini didapatkan akurasi tertinggi sebesar 92.31% yang merupakan nilai akurasi dari analisis klasifikasi menggunakan kernel linier pada data yang tidak seimbang. Sedangkan klasifikasi SVM dengan SMOTE memiliki performa yang baik, dimana SVM dengan SMOTE memiliki nilai sensitivitas yang tinggi sebesar 91.67% dan spesifisitas sebesar 95.83%. Dalam penelitian ini juga dilakukan prediksi dengan membangkitkan data *random* dari data set ekspresi gen perkembangan kanker serviks GSE63514, prediksi dilakukan dengan kernel linier. Gen yang memiliki nilai yang paling besar adalah gen X1558501\_at yaitu dengan bobot sebesar 0.023275656.

**Kata Kunci :** ADASYN, *Imbalance*, Kanker Serviks, SVM, SMOTE

## **ABSTRACT**

### **CERVICAL CANCER PROGRESSION CLASSIFICATION USING SUPPORT VECTOR MECHINE (SVM) WITH SMOTE AND ADASYN**

*(Case Study : Cervical Cancer Progression in GSE63514)*

MAHARANI PURWASI

*Department of Statistics, Faculty of Mathematics and Natural Sciences  
Islamic University of Indonesia*

*Based on 2012 data sourced from the International Agency for Research on Cancer, one of the types of cancer that most women suffer from is cervical cancer. Every year around half a million women worldwide are diagnosed with cervical cancer. Cervical cancer is a malignant cancer that begins in the cervix with irregular cell growth, and these cells can attack other biological organs, either by growing directly in adjacent tissue or by moving to other cells. Along with the advancement of technology, a new field of knowledge was developed, Bioinformatics. The application of Bioinformatics is the use of computational, mathematical and statistical methods to help solve biological problems through the analysis of gene expression data. This study uses gene expression classification data of cervical cancer development with series code GSE63514. In this study, classification analysis was carried out with the SVM method. The data used there is an imbalance of data classes, to overcome these problems resampling is done with oversampling Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic (ADASYN). This research uses SVM method classification analysis with 4 kernels, namely linear kernel, polynomial kernel, sigmoid kernel and RBF kernel. To get the best accuracy value, a comparison was made between the accuracy value of the SVM classification of unbalanced data, SVM with SMOTE oversampling and SVM with ADASYN oversampling. In this study, the highest accuracy of 92.31% was obtained which is the accuracy value of classification analysis using linear kernels on unbalanced data. While SVM classification with SMOTE has good performance, where SVM with SMOTE has a high sensitivity value of 91.67% and specificity of 95.83%. In this study, prediction was also carried out by generating random data from the GSE63514 cervical cancer development gene expression data set, prediction was carried out with a linear kernel. The gene that has the greatest value is the X1558501\_at gene, which is with a weight of 0.023275656.*

**Keywords:** ADASYN, *Imbalance*, *Cervical Cancer*, SMOTE, SVM