

APLIKASI METODE SUPPORT VECTOR MACHINE (SVM)

PADA DATA BIOINFORMATIKA

(Studi Kasus : Klasifikasi *Genes Expression Data Maize NAM*)

Ummi Maftuhatul Mu'inah

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Bioinformatika merupakan perpaduan informatika dan ilmu biologi yang mempelajari penerapan teknik komputasi untuk mengelola dan menganalisis informasi-informasi data biologi. Bioinformatika memicu peningkatan pemahaman manusia dalam bidang genomik yang terkait dengan penggunaan sekuens Deoxyribonucleic Acid (DNA). Database genom (ekspresi dan sekuens gen) berkembang sangat pesat dan berskala besar. Pada penelitian ini, data genom yang digunakan adalah nilai ekspresi gen pada dataset Maize NAM. Tujuan dari penelitian ini untuk mengklasifikasikan informasi nilai ekspresi gen berdimensi tinggi berdasarkan pada karakteristik organ apex, ear, root, shoot, dan tassel. Salah satu metode statistik yang dapat digunakan untuk klasifikasi adalah metode Support Vector Machine (SVM). Peneliti melakukan proses filtering dengan menghapus nilai nol dan uji ANOVA untuk mengurangi kompleksitas data. Berdasarkan penelitian diperoleh nilai akurasi sebesar 95% dan nilai AUC sebesar 0,989. Hal ini menunjukkan bahwa model SVM yang terbentuk dapat mengklasifikasikan nilai ekspresi gen galur jagung berdasarkan karakteristik organ dengan sangat baik.

Kata Kunci : Bioinformatika, Klasifikasi, Jagung, Support Vector Machine, AUC.

APPLICATION OF SUPPORT VECTOR MACHINE (SVM) IN BIOINFORMATICS

(Case Study: Classification of Genes Expression on Maize NAM)

Ummi Maftuhatul Mu'inah

Department of Statistics, Faculty Mathematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Bioinformatics is a combination of informatics and biology that studies the application of computational techniques to manage and analyze biological data information. Bioinformatics makes increased human understanding in the field of genomics related to the use of Deoxyribonucleic Acid (DNA) sequences. Genome database (expression and gene sequences) is growing very rapidly and on a large-scale. The genom data of this research was Maize NAM expression datasets. The purpose of this research was to classify the high-dimensional information on genes expression based on organs characteristics: apex, ear, root, shoot, and tassel. One of the statistical methods that can be used for classification is method of Support Vector Machine (SVM). The researcher conducted a filtering process by removing null value and ANOVA test to reduce data complexity. As the results, the accuracy value was 95% and AUC value was 0.989. It shows that SVM model can classify gene expression based on characteristics of the organs very well.

Keywords: *Bioinformatics, Classification, Maize, Support Vector Machine, AUC.*