

1. Pendahuluan

Latar Belakang

Seiring berkembangnya teknologi informasi jumlah data juga meningkat secara eksponensial dan ini menyebabkan para ilmuwan kewalahan dengan meningkatnya jumlah kebutuhan pemrosesan data yang timbul dari banjir data yang mengalir melalui hampir setiap bidang sains, seperti bioinformatika [1-2], biomedis [3-4], Cheminformatika [5], web [6] dan seterusnya. Bukanlah hal yang mudah untuk mengukur total volume data terstruktur dan tidak terstruktur yang menggunakan teknologi dan sistem berbasis mesin agar data-data tersebut dapat di analisis secara penuh [7]. Teknik implementasi yang efisien adalah kunci untuk memenuhi skalabilitas dan persyaratan kinerja yang diperlukan dalam analisis data ilmiah tersebut.

Pembelajaran mesin telah diteliti oleh banyak ilmuwan selama beretahun-tahun dan banyak metode penambangan data yang telah dikembangkan dan diterapkan. Support Vector Machine (SVM) adalah metode yang andal untuk melakukan klasifikasi dan regresi terutama dalam supervised machine learning [8]. Akan tetapi SVM memiliki masalah skalabilitas dalam waktu komputasi dan penggunaan memori seiring banyaknya training vector yang ada. Oleh karena itu banyak diusulkan Parallel Support Vector Machine (PSVM) untuk penambangan data yang besekala besar. PSVM membagi pekerjaan training dan klasifikasi data ke berbagai node yang berbeda, Hal ini dapat mempersingkat waktu komputasi dan mengurangi penggunaan memori.

Penulis menerapkan teknik *Sequential Minimal Optimization* (SMO) yang menggunakan *lagrange multipliers* menyelesaikan masalah quadratic programming (QP) yang muncul selama pelatihan SVM. Dalam melakukan dekomposisi SMO penulis memecah dataset kedalam beberapa subset dan menjalankan pelatihan SMO secara independent dan bergantian, proses ini dinamakan training lokal. setiap training SMO lokal akan memberikan nilai alpha (*lagrange multipliers*) lokal dan nilai bias lokal yang nantinya setiap nilai tersebut akan digabungkan menjadi nilai alpha global dan nilai bias global yang nantinya digunakan untuk membuat model klasifikasi SMO

Topik dan Batasannya

Pada penelitian tugas akhir kali ini akan dibangun sebuah sistem atau program SVM dengan menggunakan teknik SMO untuk menentukan label kelas kanker pada training data yang berupa data *microarray*. Penelitian ini berfokus dalam menguji konsep PSVM dengan dekomposisi SMO dan membandingkan hasil akurasi dan performan dekomposisi SMO dengan non-dekomposisi SMO.

Tujuan

Dari penjelasan latar belakang dan perumusan masalah dapat dihasilkan tujuan sebagai berikut:

- Membuat PSVM dan SVM dengan menggunakan teknik SMO untuk mendeteksi dan mengklasifikasi kanker menggunakan data *microarray*
- Menguji konsep PSVM dengan membuat program pelatihan SVM/SMO dengan dekomposisi dataset
- Membandingkan akurasi dan performan dari hasil pelatihan dekomposisi SMO dengan non-dekomposisi SMO menggunakan 4 dataset dan 2 kernel sebagai parameter pengujian.

Organisasi Tulisan

Penulisan tugas akhir ini mempunyai organisasi tulisan sebagai berikut. Bagian 2 adalah penelitian yang digunakan sebagai acuan dan penjabaran dari metode yang digunakan. Bagian 3 adalah detail dari sistem yang dibangun. Pada bagian 4 diberikan hasil eksperimen dan analisis dari sistem. Dan pada bagian 5, dijelaskan kesimpulan dari penelitian tugas akhir ini