

IMPLEMENTASI *EFFICIENT-SOM* PADA *TEXT CLUSTERING* MENGGUNAKAN REDUKSI DIMENSI

Nama : Toyibatu Lailiya
NRP : 09.04.111.00036
Program Study : Teknik Informatika
Dosen Pembimbing : Dr. Indah Agustien S. M.Kom
: Mulaab M. S.Si, M.Kom

ABSTRAK

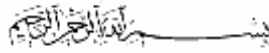
Data berukuran besar yang sudah disimpan jarang digunakan secara optimal karena manusia seringkali tidak memiliki waktu dan kemampuan yang cukup untuk mengelolanya. Pengelompokan data khususnya data teks yang mempunyai skala besar juga menemui beberapa hambatan, salah satunya adalah diperlukannya waktu komputasi yang besar pula. Penelitian ini bertujuan untuk mengimplementasikan metode *Self-Organizing Maps* (SOM) dengan menerapkan reduksi dimensi sehingga dapat mengurangi waktu komputasi.

Metode ini diterapkan untuk mengelompokan data tugas akhir mahasiswa Teknik Informatika Universitas Trunojoyo Madura. Dalam metode yang diusulkan, analisis morfologi dilakukan pada teks abstrak tugas akhir mahasiswa untuk menghasilkan vektor input dengan unsur term dari tugas akhir tersebut. Adanya elemen vector yang jarang, memungkinkan terjadinya reduksi dimensi. Reduksi dimensi dilakukan dengan menghilangkan matriks jarang atau kolom yang banyak mengandung nilai 0.

Dari percobaan yang dilakukan, diperoleh hasil bahwa optimum cluster untuk data abstrak berada pada scenario 1 dengan jumlah term setelah direduksi sebanyak 150 term sehingga menghasilkan nilai rata-rata SSE = 0.01117. Sedangkan untuk data laporan optimum cluster berada pada scenario 2 dengan jumlah term setelah direduksi sebanyak 346 term sehingga menghasilkan nilai rata-rata SSE = 0.0085 detik.

Kata kunci : *clustering, text, self-organizing map, reduksi dimensi.*

KATA PENGANTAR



Segala puji bagi Allah, karena atas Rahmat dan Hidayah-Nya penulis dapat menyelesaikan Tugas Akhir dengan judul:

Implementasi *Efficient-SOM* Pada *Text Clustering* Menggunakan Reduksi Dimensi

Dalam penyusunan Tugas Akhir ini, penulis menyadari bahwa semua ini tidak lepas dari peran berbagai pihak yang telah dengan tulus senantiasa memberikan dukungan, bimbingan, dan motivasi, hingga akhirnya Tugas Akhir ini dapat terselesaikan. Oleh sebab itu, pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ibu, Bapak dan Mas yang tak pernah terputus mencurahkan kasih sayangnya dan tak pernah berhenti mendukungku.
2. Dr. Indah Agustien S. M. Kom selaku dosen pembimbing I dan Mulaab M. S.Si, M. Kom selaku dosen pembimbing II. Terimakasih karena ibu dan bapak telah banyak meluangkan waktu untuk membimbing dan memberikan arahan kepada penulis, sehingga mampu terselesaikannya Tugas Akhir ini.
3. Okie, Umie dan Nurdiah yang selama ini telah menemaniku dan ada buatku.
4. Mak Ka dan teman-teman serta adik-adik kosan yang selama ini menemani dan sering menyemangati dengan menanyakan “kapan lulus?”
5. Mas Lukman, Mas Hasan, Mas Okfan, Mas Fajar, Mas Chipul, Mas Faza, Mas Wahyu, Mas Darul dan Sholeh yang telah mengajari banyak hal dan memberi semangat hingga Tugas Akhir ini akhirnya mampu terselesaikan juga.
6. Bapak dan Ibu Dosen Fakultas Teknik yang telah memberikan ilmunya selama proses perkuliahan berlangsung.
7. Seluruh teman satu angkatan yang tergabung dalam keluarga SOLITE '09. Terimakasih telah menjadi teman seperjuangan

yang benar-benar SOLID. Senang bisa menjadi bagian dari kalian.

8. Dan untuk semua pihak yang tidak dapat disebutkan satu persatu yang turut berperan dalam Tugas Akhir ini.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan. Semoga Tugas Akhir ini dapat bermanfaat dan dapat dikembangkan dengan lebih baik lagi nantinya.

Bangkalan, Juli 2013

Toyibatu Lailiya

DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
HALAMAN MOTTO	iv
HALAMAN PERSEMBAHAN.....	v
ABSTRAK.....	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
DAFTAR ISTILAH	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	2
1.4 Batasan Masalah.....	2
1.5 Metodologi Penelitian.....	2
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Sebelumnya	5
2.2 Jaringan Syaraf Tiruan	5
2.3 <i>Self Organizing Map</i>	6
2.3.1 Arsitektur	6
2.3.2 Algoritma	7
2.3.2 Aplikasi	9
2.4 Speed-Up SOM Melalui Reduksi Dimensi	12
2.4.1 Reduksi Dimensi Berdasarkan Karakteristik Data	12
BAB III PERANCANGAN SISTEM	15
3.1 Desain Perancangan Sistem	15
3.2 Pengambilan Data	17
3.3 Preprosesing	19
3.4 Reduksi Dimensi Menghilangkan <i>Sparse Part</i>	20
3.5 <i>Clustering Using SOM</i>	20
3.6 Mengevaluasi Hasil Cluster Dengan SSE	21
BAB IV IMPLEMENTASI SISTEM	23
4.1 Lingkungan Uji Coba	23

4.2 Data Uji Coba	23
4.3 Data Hasil Preprosesing	24
4.4 Data Hasil Reduksi Dimensi	25
4.5 <i>Learning</i> Dan Ujicoba	27
4.5.1 Data Abstrak	27
4.5.2 Data Laporan	30
4.6 Waktu Komputasi	32
4.7 Evaluasi Hasil Cluster	32
BAB V KESIMPULAN	35
5.1 Kesimpulan	35
5.2 Saran	36
DAFTAR PUSTAKA	37
LAMPIRAN	39
RIWAYAT HIDUP	55

DAFTAR GAMBAR

	Hal.
Gambar 2.1 Arsitektur SOM	7
Gambar 2.2 Array Linier dalam Unit Cluster	7
Gambar 2.3 Neighborhood Grid Rectangular	8
Gambar 2.4 Neighborhood Grid Hexagonal	8
Gambar 2.5 Contoh Reduksi Dimensi	13
Gambar 3.1 <i>Outline</i> Dari Sistem Yang Akan Dibuat	15
Gambar 3.2 Diagram Alir Dari Sistem Yang Dikerjakan	16
Gambar 3.3 Ilustrasi Reduksi Dimensi	20
Gambar 3.4 Diagram Alir SOM	21
Gambar 4.1 Tampilan Daftar File Data Training	24
Gambar 4.2 Tampilan Daftar File Data Testing	24
Gambar 4.3 Data TF Sebelum Proses Reduksi Dimensi	26
Gambar 4.4 Data TF Setelah Proses Reduksi Dimensi	26

DAFTAR TABEL

	Hal.
Tabel 3.1 Data Training	23
Tabel 3.2 Data Testing	23
Tabel 2.1 Lingkungan Ujicoba	23
Tabel 4.2 Data Ujicoba	23
Tabel 4.3 Hasil Uji coba dengan skenario 1 untuk data abstrak.....	27
Tabel 4.4 Hasil Uji coba dengan scenario 2 untuk data abstrak.....	28
Tabel 4.5 Hasil Uji coba dengan scenario 3 untuk data abstrak.....	29
Tabel 4.6 Hasil Uji coba dengan skenario 1 untuk data laporan.....	30
Tabel 4.7 Hasil Uji coba dengan skenario 2 untuk data laporan.....	31
Tabel 4.8 Hasil Uji coba dengan skenario 3 untuk data laporan.....	31
Tabel 4.9 Lama KOmputasi Proses Ujicoba	53
Tabel 4.10 Nilai SSE Pada Data Abstrak	50
Tabel 4.11 Nilai SSE Pada Data Laporan	53

DAFTAR ISTILAH

<i>Clustering</i>	Pengelompokan data sesuai dengan karakteristik masing-masing
<i>Reduksi</i>	Pengurangan dimensi data baik itu data statistic maupun data non statistic
<i>Dimensi</i>	
<i>Preprosesing</i>	Tahapan awal yang dilakukan dalam pengolahan data pada suatu penelitian
<i>Self-Organizing Map</i>	Metode yang bertujuan untuk mengelompokkan suatu vektor-vektor input berdasarkan bagaimana mereka mengelompok sesuai dengan karakteristik inputnya.
<i>Euclidean Distance</i>	Penghitungan jarak dari dua buah titik dalam <i>Euclidean space</i> .

DAFTAR LAMPIRAN

Lampiran A Data Term.....	39
Lampiran B Alur Implementasi	49