

**IMPLEMENTASI ALGORITMA *PRO-KMEANS* DALAM
PENGELOMPOKKAN BADAK SUMATERA**

SKRIPSI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun oleh :

SONY NOVIANTO

NIM. 0710963004

**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
PROGRAM STUDI TEKNIK INFORMATIKA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2013**

LEMBAR PERSETUJUAN
IMPLEMENTASI ALGORITMA *PRO-KMEANS* DALAM
PENGELOMPOKKAN BADAK SUMATERA

SKRIPSI



Disusun oleh :

SONY NOVIANTO

NIM. 0710963004

Telah diperiksa dan disetujui oleh :

Pembimbing I,

Pembimbing II,

Dian Eka R , S.Si., M.Kom
NIP. 197306192002122001

Edy Santoso, S.Si., M.Kom
NIP. 197404142003121004

LEMBAR PENGESAHAN
IMPLEMENTASI ALGORITMA PRO-KMEANS DALAM
PENGELOMPOKKAN BADAK SUMATERA

SKRIPSI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun oleh :

SONY NOVIANTO

NIM. 0710963004

Skripsi ini telah diuji dan dinyatakan lulus tanggal 23 Januari 2013

Penguji I,

Nurul Hidayat, S.Pd., M.Sc
NIP. 196804302002121001

Penguji II,

Barlian Henryranu P, ST, MT.
NIK. 82102406110254

Penguji III,

Ahmad Afif Supianto, S.Si., M.Kom.

Mengetahui
Ketua Program Studi Teknik Informatika

Drs. Marji, M.Si.
NIP. 196708011992031001

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang telah saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku. (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, Januari 2013

Mahasiswa,

Sony Novianto

NIM 0710963004

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan segala Rahmat, Karunia dan Hidayah-Nya sehingga Penulis dapat menyelesaikan skripsi dengan judul: **"IMPLEMENTASI ALGORITMA PRO-KMEANS DALAM PENGELOMPOKKAN BADAK SUMATERA"**

Skripsi ini diajukan sebagai syarat ujian seminar skripsi dalam rangka untuk memperoleh gelar Sarjana Komputer di Program Teknologi Informasi dan Ilmu Komputer, Program Studi Informatika/Ilmu Komputer, Universitas Brawijaya Malang. Atas terselesaikannya skripsi ini, Penulis mengucapkan terima kasih kepada:

1. Dian Eka R, S.Si., M.Kom. selaku Dosen Pembimbing Skripsi I.
2. Edy Santoso, S.Si., M.Kom. selaku Dosen Pembimbing Skripsi II.
3. Drs. Marji, MT. selaku Ketua Program Studi Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
4. Ir. Sutrisno, MT, selaku Ketua Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
5. Djoko Pramono, ST. selaku Dosen Penasehat Akademik.
6. Segenap Bapak dan Ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada Penulis selama menempuh pendidikan di Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
7. Segenap staf dan karyawan di Program Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya yang telah banyak membantu Penulis dalam pelaksanaan penyusunan skripsi ini.
8. Orang tua Penulis dan saudara-saudaraku atas segala dukungan materi dan doa restunya kepada Penulis.
9. Rekan-rekan Program Studi Informatika/Ilmu Komputer yang telah memberikan dukungannya kepada penulis.
11. Dota, yang telah memberi kreativitas dan menjalani malam-malam bersama skripsi yang telah disusun ini.

12. Novieta Putri R, yang telah memberi dukungan baik dalam bentuk material maupun non material demi terselesaikannya skripsi ini.
13. Semua pihak yang telah membantu terselesaikannya skripsi ini yang tidak dapat kami sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini tentunya tidak terlepas dari berbagai kekurangan dan kesalahan. Oleh karena itu, segala kritik dan saran yang bersifat membangun sangat Penulis harapkan dari berbagai pihak demi penyempurnaan penulisan skripsi ini.

Akhirnya penulis berharap agar skripsi ini dapat memberikan sumbangan dan manfaat bagi semua pihak yang berkepentingan.

Malang, Januari 2013



Sony Novianto
Penulis

ABSTRAK

Sony Novianto. 2013. : Implementasi Algoritma Pro-Kmeans dalam Pengelompokan Badak Sumatera.

Dosen Pembimbing : Dian Eka R, S.Si., M.Kom. dan Edy Santoso, S.Si., M.Kom.

Indonesia merupakan negara hutan hujan tropis yang memiliki keanekaragaman hayati yang tinggi dan dikenal sebagai salah satu *Megabiodiversity Country*. Pulau Sumatera salah satu pulau di Indonesia yang dapat menjadi gambaran kekayaan daerah tropis di Asia. Hal tersebut bisa dilihat dari flora dan fauna yang ada di pulau ini, yaitu satu diantaranya adalah badak Sumatera (*Dicerorhinus sumatrensis*). Kendala yang sering ditemukan dalam proses pengelompokan badak Sumatera yaitu sulitnya menentukan jenis badak. Hal ini dikarenakan sulitnya menemukan data DNA badak yang hampir punah. Pada penelitian ini akan diangkat suatu kasus yaitu sistem implementasi algoritma Pro-Kmeans dalam pengelompokan badak Sumatera. Algoritma Pro-Kmeans dapat mengelompokkan (segmentasi) data yang mempunyai atribut dan mempunyai jumlah data yang sesuai dengan jenis badak sumatera, sehingga dapat dimanfaatkan dalam sistem pengelompokan badak Sumatera. Berdasarkan hasil analisis terhadap algoritma Pro-Kmeans untuk sistem pengelompokan badak Sumatera, maka kesimpulan yang dapat diambil adalah algoritma Pro-Kmeans tepat untuk mengelompokkan badak Sumatera berdasarkan data nilai uji akurasi yang kecil.

Kata kunci : Algoritma Pro-Kmeans ,Badak Sumatera.

ABSTRACT

Sony Novianto. 2013.: Implementation Of Pro-Kmeans Algorithm In classifying Dicerorhinus Sumatrensis.

Advisor : Dian Eka R, S.Si., M.Kom. dan Edy Santoso, S.Si., M.Kom.

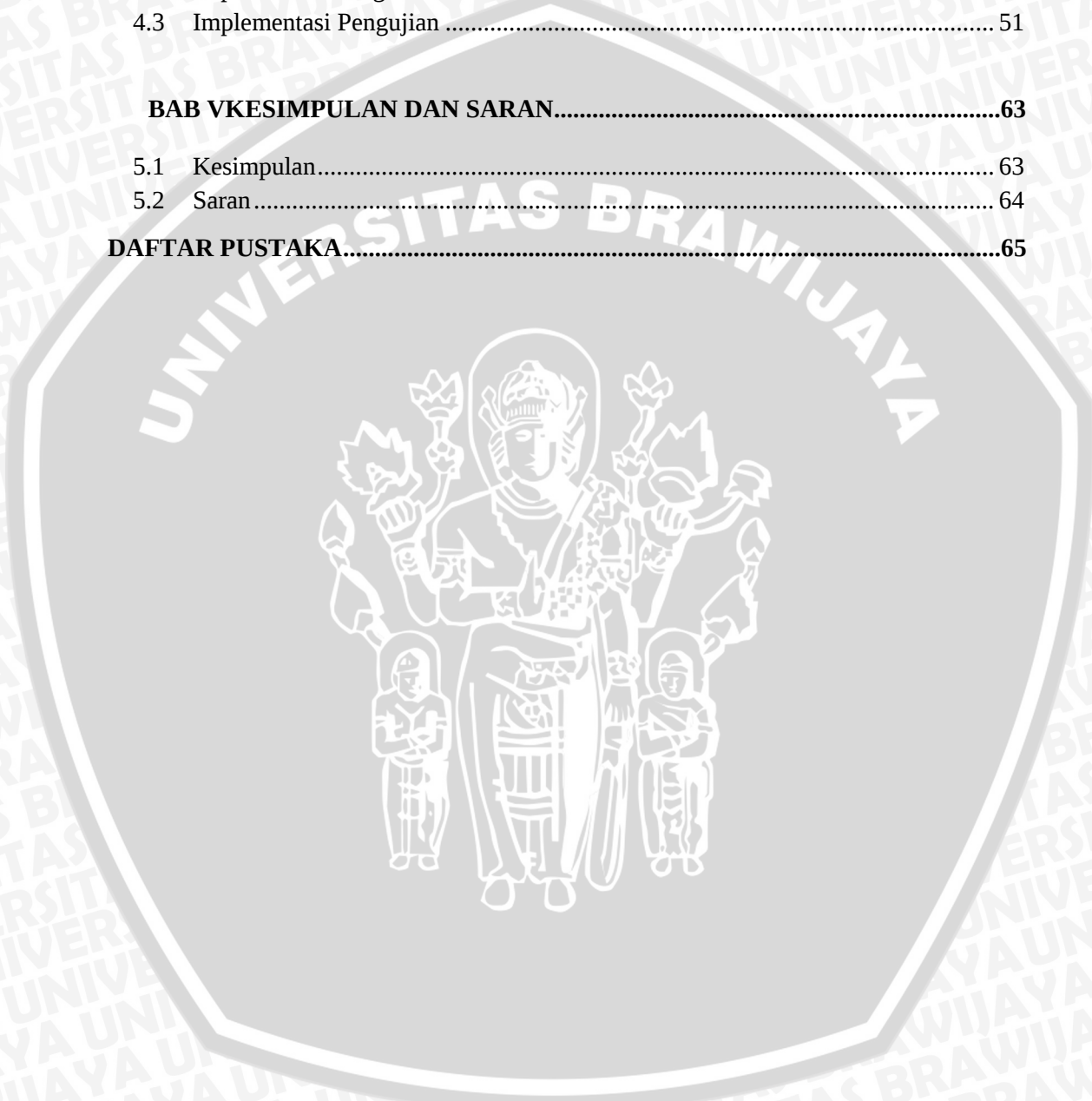
Indonesia is a tropical rain forest with high biodiversity and is known as one of the megabiodiversity Country. Pulau one of the islands of Sumatra in Indonesia, which could be a picture of the wealth of tropical regions in Asia. It can be seen from the flora and fauna on the island, one of which is the Sumatran rhino (*Dicerorhinus sumatrensis*). Constraints are often found in the process of grouping Sumatran rhinos are difficult to determine types of rhinos. This is due to the difficulty of finding DNA data the endangered rhino. In this study, a case that will be appointed system Pro-Kmeans algorithm implementation in grouping Sumatran rhino. Pro-Kmeans algorithm to group (segmentation) data that have attributes and have the amount of data in accordance with the type of the Sumatran rhino, so it can be used in clustering systems Sumatran rhino. Based on the analysis of the Pro-Kmeans algorithm for clustering systems Sumatran rhino, the conclusions that can be drawn is Pro-Kmeans algorithm appropriate to categorize the data Sumatran rhinos by its small value of the test accuracy.

Keywords: *Pro-Kmeans Algorithm ,Dicerorhinus Sumatrensis.*

DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS SKRIPSI	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xii
1 BAB 1PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan dan Manfaat.....	3
1.4 Batasan Masalah.....	4
1.5 Metodologi Pengerjaan Skripsi	4
1.6 Sistematika Penulisan.....	6
BAB IITINJAUAN PUSTAKA.....	7
2.1 Klasifikasi.....	9
2.2 Penyebaran	8
2.3 DNA	9
2.4 Matriks Skor	9
2.5 Algoritma Smith-Waterman	10
2.6 Algoritma Pro-Kmeans.....	10
2.7 Error Ratio.....	13
BAB IIIMETODE DAN PERANCANGAN SISTEM.....	15
3.1 Perancangan Sistem.....	15
3.2 Perancangan Penelitian.....	17
3.3 Implementasi Pada Data	27

3.4	Bahan Pengujian.....	33
3.5	PerancanganAntar Muka	34
4	BAB IVIMPLEMENTASI DAN PEMBAHASAN.....	35
4.1	Lingkungan Implementasi.....	35
4.2	Implementasi Program.....	35
4.3	Implementasi Pengujian	51
	BAB VKESIMPULAN DAN SARAN.....	63
5.1	Kesimpulan.....	63
5.2	Saran.....	64
	DAFTAR PUSTAKA.....	65



DAFTAR GAMBAR

Gambar 2.1 Dicerorhinus Sumatrensis	7
Gambar 2.2 Peta Penyebaran Badak Sumatera.....	9
Gambar 2.3 Matriks Skor DNA	10
Gambar 3.1 Bagan Alur Tahapan Penelitian	15
Gambar 3.2 Flowchart Perancangan Sistem	16
Gambar 3.3 Flowchart Proses Partisi Data	17
Gambar 3.4 Flowchart Proses Similarity Score Antar Pasangan DNA dalam 1 Cluster.....	18
Gambar 3.5 Flowchart Proses SumScore.....	19
Gambar 3.6 Flowchart Proses Hitung Score.....	20
Gambar 3.7 Flowchart Proses Hitung $g(n)$	21
Gambar 3.8 Flowchart Proses Pembentukan Centroid Tiap Cluster	22
Gambar 3.9 Flowchart Proses Pembentukan Cluster Menggunakan Algoritma Pro-Kmeans	23
Gambar 3.10 Flowchart Proses Perhitungan Similarity Dataset i Terhadap Centroid j	24
Gambar 3.11 Flowchart Proses Pembentukan Cluster.....	25
Gambar 3.12 Flowchart Proses Perhitungan $f(v)$	26
Gambar 3.13 Rancangan Aplikasi	34
Gambar 4.1 Source Code Load Data	35
Gambar 4.2 Source Code Proses Random Data ke dalam k Cluster	38
Gambar 4.3 Source Code Similarity Score Antar Pasangan DNA Pada Satu Cluster Menggunakan Algoritma Smith-Waterman	40
Gambar 4.4 Source Code SumScore Tiap-Tiap DNA.....	43
Gambar 4.5 Source Code Centroid Tiap-Tiap Cluster.....	44
Gambar 4.6 Source Code Pembentukan Cluster Menggunakan Algoritma Smith-Waterman dan Analisa Cluster Menggunakan Perhitungan $f(v)$	45

Gambar 4.7 Source Code Implementasi Perulangan Iterasi	47
Gambar 4.8 Tampilan Proses Dataset	48
Gambar 4.9 Tampilan Proses Similarity Score	49
Gambar 4.10 Tampilan Proses SumScore dan Penentuan Centroid	50
Gambar 4.11 Tampilan Proses <i>Similarity</i> Dengan Pengurangan g(n) dan Penentuan Centroid Baru.....	50
Gambar 4.12 Tampilan Proses Hasil Akhir Clustering	51
Gambar 4.13 Grafik Akurasi Uji Wilayah.....	52
Gambar 4.14 Grafik Akurasi Uji Po	54
Gambar 4.15 Grafik Akurasi Uji Pe	55
Gambar 4.16 Akurasi Badak Sumatera wilayah Aceh.....	57
Gambar 4.17 Akurasi Badak Sumatera wilayah Jambi	58
Gambar 4.18 Akurasi Badak Sumatera wilayah Sumatera Selatan	59
Gambar 4.19 Akurasi Badak Sumatera wilayah Lampung.....	60



DAFTAR TABEL

Tabel 3.1 Tabel Data Badak Sumatera	27
Tabel 3.2 Perhitungan Similarity Score	27
Tabel 3.3 Proses Perhitungan Similarity Value dengan $\sum g(n)$	29
Tabel 3.4 Perhitungan Similarity Score	29
Tabel 3.5 Proses Perhitungan Similarity Value dengan $\sum g(n)$	30
Tabel 3.6 Perhitungan Similarity Score	31
Tabel 3.7 Proses Perhitungan Similarity Value dengan $\sum g(n)$	32
Tabel 3.8 Pelabelan Data	33
Tabel 3.9 Hasil Clustering	33
Tabel 3.10 Kriteria Pengujian	34
Tabel 4.1 Akurasi Uji Wilayah Menggunakan Pro-Kmeans	52
Tabel 4.2 Akurasi Uji Po Menggunakan Pro-Kmeans.....	53
Tabel 4.3 Akurasi Uji Peo Menggunakan Pro-Kmeans.....	55
Tabel 4.4 Wilayah Aceh	57
Tabel 4.5Wilayah Jambi	58
Tabel 4.6Wilayah Sumatera Selatan	60
Tabel 4.7Wilayah Lampung.....	61