

PERBANDINGAN KINERJA ALGORITMA *SYSTEMATIC CLUSTERING* DAN *GREEDY K-MEMBER* PADA MODEL *K-ANONYMITY* YANG MENGGUNAKAN DUA ATRIBUT SENSITIF



**NASIHA SAKINA
5235134389**

**Skripsi ini Ditulis untuk Memenuhi Sebagian Persyaratan dalam Memperoleh
Gelar Sarjana**


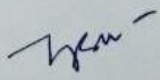
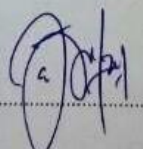
**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA
DAN KOMPUTER
FAKULTAS TEKNIK
UNIVERSITAS NEGERI JAKARTA**

2017

HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
<u>Widodo, M.Kom</u> (Dosen Pembimbing I)		23-2-17
<u>Bambang Prasetya Adhi, M.Kom</u> (Dosen Pembimbing II)		24-2-17

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
<u>Lipur Sugivanta., Ph. D</u> (Ketua Penguji)		24-2-17
<u>Drs. Bachren Zaini, M.Pd</u> (Sekretaris Penguji)		23-2-17
<u>ZE. Ferdi Fauzan Putra., M.Pd.T</u> (Dosen Ahli)		23/2/2017

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Karya tulis skripsi saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan dan penelitian saya sendiri dengan arahan dosen pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 22 Februari 2017

Yang membuat pernyataan



Nasiha Sakina

5235134389

ABSTRAK

Nasiha Sakina, Perbandingan Kinerja Algoritma *Systematic Clustering* dan *Greedy K-Member* pada Model *K-Anonymity* yang Menggunakan Dua Atribut *Sensitif*. Skripsi. Jakarta, Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta, 2017. Dosen Pembimbing: Widodo, M.Kom dan Bambang P. Adhi, M.Kom.

Latar belakang dari penelitian ini adalah adanya perusahaan atau organisasi yang harus menyebarkan informasi tanpa harus diketahui informasi aslinya, sehingga dilakukan penganoniman data. Model yang digunakan adalah *K-Anonymity* dan algoritma yang akan digunakan adalah Algoritma *Systematic Clustering* dan Algoritma *Greedy K-Member*. Metode yang digunakan adalah rekayasa teknik dimana nantinya akan dihitung algoritma mana yang memiliki *information loss* paling sedikit untuk mengetahui algoritma mana yang lebih baik. Tujuan dari penelitian ini ialah diharapkan perusahaan atau organisasi dapat memberikan informasi mengenai perusahaannya tapi tidak secara detail sehingga masih tetap aman. Data pada penelitian ini di ambil dari *UCI Machine Learning Repository*. Dataset yang digunakan adalah data set *Adult*. Hasil dari penelitian ini adalah nilai *information loss* terendah untuk *Systematic Clustering* adalah 11843,9 dan untuk *Greedy K-Member* adalah 13950,83. Sedangkan nilai *information loss* terbesar untuk *Systematic Clustering* adalah 14705,2 dan untuk *Greedy K-Member* adalah 13962,09. Kesimpulan dari penelitian ini adalah Algoritma *Systematic Clustering* lebih baik dibandingkan dengan Algoritma *Greedy K-Member*.

Kata kunci: Model *K-Anonymity*, Algoritma *Systematic Clustering*, Algoritma *Greedy K-Member*, *Information Loss*

ABSTRACT

Nasiha Sakina, *Performance Comparison Of Algorithm Systematic And Greedy Clustering K-Member Of Model K-Anonymity Which Uses Two Sensitive Attributes*. Essay. Jakarta, Education of Informatics and Computer Engineering Study Program, Faculty of Engineering, Universitas Negeri Jakarta, 2017. Supervisor: Widodo, M.Kom and Bambang P. Adhi, M.Kom.

The background of this research is the existence of companies or organizations which need to share information without revealing the original version, so anonymization of the data is required. K anonymity has been used as a method of this research, meanwhile systematic clustering algorithm and greedy K member algorithm as the algorithm. The K algorithm method is an engineering which will be calculated to decide which one of the algorithm that has the least information loss, so the best algorithm will be known. The purpose of this research is hopefully the company or organization could give information about themselves narrowly, so it's still safe. The data in this research was taken from UCI machine learning repository. Adult data set has been used as data set. The result of this research is 11843,9 as the least value of information loss for systematic clustering and 13950,83 for the greedy K member. Meanwhile the biggest value of information loss for systematic clustering is 14705,2 and 13962,09 for greedy K member. The conclusion of this research is Systematic Clustering Algorithm is better than Greedy K-Member Algorithm.

Keywords: Model K-Anonymity, Algorithm Systematic Clustering, Algorithm Greedy K-Member, Information Loss

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT Yang Maha Pengasih dan Maha Penyayang. Atas rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Kinerja Algoritma *Systematic Clustering* Dan *Greedy K-Member* Pada Model *K-Anonymity* Yang Menggunakan Dua Atribut Sensitif”. Skripsi ini disusun sebagai salah satu syarat dalam memperoleh gelar Sarjana Pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Jakarta.

Selama penyusunan skripsi, banyak bantuan yang penulis terima dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Allah SWT, yang telah memberikan segala kemudahan dalam menyelesaikan skripsi ini.
2. Ibu Yuliatri Sastrawijaya, M.Pd., selaku Ketua Program Studi Pendidikan Teknik Informatika dan Komputer, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Jakarta.
3. Bapak Widodo, S.Kom, M.Kom., selaku Dosen Pembimbing I.
4. Bapak Bambang Prasetya Adhi, M.Kom selaku Dosen Pembimbing II.
5. Bapak Prasetyo Wibowo Yunanto, M.Eng, selaku pembimbing akademik.
6. Seluruh Staf Administrasi Tata Usaha Jurusan Teknik Elektro, Pustikom UNJ, dan Kantor Wakil Rektor 2 Universitas Negeri Jakarta.
7. Bapak Ridwan Hasballah dan Ibu Maidarwati, selaku kedua orang tua penulis. Terima kasih sudah membesarkan, mendidik, merawat, dan mendoakan aku sehingga dapat menyelesaikan skripsi ini.
8. Adik-adik tersayang Maghvira Sakina dan Misarah Sakina serta seluruh keluarga yang tidak disebutkan satu persatu diskripsi ini. Terima kasih atas doa dan dukungannya.
9. Reza Ridwansyah dan M. Aldy Azmi sebagai teman sepayung skripsi.

Terima kasih telah membantu dalam menyelesaikan skripsi ini.

10. Teman-teman seperjuangan PTIK angkatan 2013, Cantik Sejagat, Pejuang Jannah-Nya, Keluarga Advokasi BEMFT 2016, Kak Umi, Kak Icha, dan Hilma yang telah memberikan motivasi, semangat, dan doanya dalam penulisan skripsi ini.
11. Seluruh pihak yang telah mendukung yang tidak bisa disebutkan satu persatu demi terselesaikannya skripsi ini dengan baik dan lancar. Semoga Tuhan membalas semua kebaikan yang telah diberikan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, karenanya penulis mohon maaf apabila terdapat kekurangan kesalahan baik dari isi maupun tulisan. Akhir kata penulis berharap semoga skripsi ini bermanfaat bagi yang membacanya.

Jakarta, 22 Februari 2017

Nasiha Sakina

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	
1.1. Latar Belakang Masalah	1
1.2. Identifikasi Masalah	5
1.3. Pembatasan Masalah	5
1.4. Perumusan Masalah	6
1.5. Tujuan Penelitian	6
1.6. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	
2.1 Teori	7
2.1.1. Privasi	7
2.1.2. Data.....	7
2.1.3. Model.....	8
2.1.4. Anonim	9
2.1.5. Algoritma	9
2.1.6. Privacy Preserving Data Publishing	11
2.1.6.1. Pertubative	11
2.1.6.2. Non-Pertubative	11
2.1.7. Atribut	14
2.1.7.1. Explicit Identifier	14

2.1.7.2. Quasi Identifier.....	14
2.1.7.3 Sensitive Attribute.....	15
2.1.7.4. Non-sensitive Attribute.....	15
2.1.8. K-Anonymity	15
2.1.9. <i>Systematic Clustering</i>	17
2.1.10. <i>Greedy K-Member</i>	18
2.1.11. <i>Information Loss</i>	19
2.2. Metode dan Proses Penelitian.....	20
2.2.1. <i>Achieving K-Anonymity Privacy Protection Using Generalization and Suppression</i> (Latanya Sweeney).....	20
2.2.2. <i>Efficient Systematic Clustering Method for K-Anonymization</i> (Md. Enamul Kabir, Hua Wang, dan Elisa Bertino).....	21
2.2.3. <i>Efficient K-Anonymization Using Clustering Techniques</i> (Ji-Won Byun, Ashish Kamra, Elisa Bertino, dan Ninghui Li)....	22
2.3. Konsep dan Prosedur Penelitian	23

BAB III METODELOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian	26
3.2. Alat dan Bahan Penelitian	26
3.3. Diagram Alir Penelitian	28
3.4. Teknik dan Proses Pengumpulan Data	31
3.5. Teknik Analisis Data	30

BAB IV HASIL PENELITIAN

4.1. Deskripsi Hasil Penelitian.....	32
4.1.1. Menarik Data	32
4.1.2. Menormalkan Data	33
4.1.3. <i>Clustering</i> dengan Algoritma <i>Systematic Clustering</i>	34
4.1.4. <i>Clustering</i> dengan Algoritma <i>Greedy K-Member</i>	34
4.1.5. <i>K-Anonymity</i>	35
4.2. Analisis Data Penelitian	38
4.3. Pembahasan.....	41
4.4 Aplikasi Hasil penelitian.....	43

BAB V KESIMPULAN, DAN SARAN

5.1. Kesimpulan 44

5.2. Saran 44

DAFTAR PUSTAKA 45

LAMPIRAN..... 46

DAFTAR RIWAYAT HIDUP 70

DAFTAR TABEL

	Halaman
Tabel 1.1. Data Diri Pasien di Rumah Sakit	1
Tabel 1.2 Data Diri Pasien di Masyarakat	2
Tabel 1.3 Data Diri Pasien yang Telah di Generalisasi	3
Tabel 2.1 Tabel Data Masyarakat yang Telah di Suppresi dengan <i>Global Recording</i>	13
Tabel 2.2 Tabel Data Masyarakat yang Telah di Suppresi dengan <i>Local Recording</i>	14
Tabel 3.1 Spesifikasi Laptop yang Digunakan	6
Tabel 4.1. <i>Cluster</i> pada <i>Systematic Clustering</i>	34
Tabel 4.2 <i>Cluster</i> pada <i>Greedy K-Member</i>	35
Tabel 4.3 Generalisasi pada <i>Systematic Clustering</i>	37
Tabel 4.4 Generalisasi pada <i>Greedy K-Member</i>	38
Tabel 4.5 Hasil <i>Systematic Clustering</i>	38
Tabel 4.6 Hasil <i>Greedy K-Member</i>	38

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Hubungan Antara Fase dan Peranan di PPDP.....	11
Gambar 2.2. <i>Taxonomy Tree</i> Negara	12
Gambar 2.3. <i>Taxonomy Tree Zip Code</i>	24
Gambar 2.4. Bagan Kerangka Berpikir.....	25
Gambar 3.1. Diagram Alir Penelitian	29
Gambar 3.2. Data Set <i>Adult</i> pada <i>UCI Machine Learning</i>	30
Gambar 4.1. Data Set <i>Adult</i> pada <i>UCI Machine Learning</i>	32
Gambar 4.2. Data Set <i>Adult</i> pada Microsoft Excel dengan Format CSV.....	33
Gambar 4.3. Data Set <i>Adult</i> pada <i>Database</i>	33
Gambar 4.4. <i>Taxonomy Tree</i> untuk <i>Workclass</i>	36
Gambar 4.5. <i>Taxonomy Tree</i> untuk <i>Marital Status</i>	36
Gambar 4.6. Grafik Perbandingan <i>Information Loss</i> Algoritma <i>Systematic Clustering</i> dan <i>Greedy K-Member</i>	38
Gambar 4.7. Grafik Perbandingan <i>Execution Time</i> Algoritma <i>Systematic Clustering</i> dan <i>Greedy K-Member</i>	40
Gambar 4.8. Grafik Perbandingan <i>Information Loss</i> Maksimal dan Minimal Algoritma <i>Systematic Clustering</i> dan <i>Greedy K-Member</i>	41
Gambar 4.9. Grafik Perbandingan <i>Execution Time</i> Maksimal dan Minimal Algoritma <i>Systematic Clustering</i> dan <i>Greedy K-Member</i>	42

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. <i>Pseudocode Systematic Clustering</i>	50
Lampiran 2. <i>Pseudocode Greedy K-Member</i>	51
Lampiran 3. <i>Sourcecode Systematic Clustering</i>	52
Lampiran 4. <i>Sourcecode Greedy K-Member</i>	59