PERBANDINGAN KINERJA ALGORITMA *SYSTEMATIC CLUSTERING* DAN *ONE PASS K-MEANS* PADA MODEL *K-ANONIMITY* DATA

SKRIPSI

REZA RIDWANSYAH
5235134442

Skripsi ini Ditulis untuk Memenuhi Sebagai Persyaratan
dalam Memperoleh Gelar Sarjana Pendidikan

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

2017

HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
Widodo, M.Kom (Dosen Pembimbing I)	.....	17-05-2017
Bambang P. Adhi, M.Kom (Dosen Pembimbing II)	.....	17-05-2017

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Dr. Yuliatri Sastrawijaya, M.Pd. (Ketua Penguji)	.....	13-01-2017
Z.E. Ferdi Fauzan Putra, S.Pd., M.Pd.T (Sekretaris)	.....	13-01-2017
Prof. Dr. Ir. Ivan Hanafi, MPd. (Dosen Ahli)	.....	12-01-2017

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Karya tulis skripsi saya yang berjudul Perbandingan Kinerja Algoritma *Systematic Clustering* dan *One Pass K-Means* Dengan Model *K-Anonimity Data* adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di perguruan tinggi lain.
2. Karya tulis yang berjudul Perbandingan Kinerja Algoritma *Systematic Clustering* dan *One Pass K-Means* Dengan Model *K-Anonimity Data* adalah murni gagasan, rumusan, dan penelitian saya sendiri dengan arahan dosen pembimbing.
3. Dalam karya tulis, 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 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, 9 Januari 2017

Yang membuat pernyataan

Reza Ridwansyah

5235134442

iii

KATA PENGANTAR

Puji dan syukur saya ucapkan kehadirat Tuhan Yang Maha Esa, yang dengan kehendaknya memberikan saya izin untuk menyelesaikan skripsi ini. Skripsi ini disusun sebagai persyaratan untuk meraih gelar Sarjana Pendidikan Teknik Informatika dan Komputer pada Fakultas Teknik, Universitas Negeri Jakarta.

Dalam menyelesaikan skripsi ini penulis telah mencurahkan segala kemampuan dan penulis menyadari akan kemampuan dan keterbatasan yang dimiliki. Skripsi ini tidak dapat terwujud dengan baik tanpa adanya bimbingan, dorongan, saran-saran, dan bantuan dari berbagai pihak. Oleh sebab itu pada kesempatan ini saya ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Widodo, M.Kom selaku pembimbing I yang telah memberikan waktu, motivasi, arahan dan kepercayaan kepada penulis dalam menyelesaikan skripsi ini.
2. Bapak Bambang P. Adhi, M.Kom., selaku pembimbing II. Guru sekaligus teman yang baik dan sabar
3. Ibu Dr. Yuliatri Sastrawidjaya, M.Pd., selaku Ketua Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta
4. Prasetyo Wibowo Yunanto, ST., M.Eng, selaku pembimbing akademik penulis.
5. Seluruh dosen dan staf Tata Usaha Program Studi Pendidikan Teknik Informatika dan Komputer yang selalu membantu menyediakan informasi dan membantu proses administrasi skripsi
6. Keluarga Alim Raja dan Keluarga Husein yang selalu memberikan semangat, kekuatan, dan doa yang tulus dalam pengerjaan skripsi penulis
7. Keluarga PTIK 2013 dan University Of Jagad Raya, terutama Akbar Jaya, Emas Arum, Muhammad Rayhan dan Satria Luqman yang senantiasa memberikan bantuan yang tak ternilai harganya sampai penulis bisa menyelesaikan skripsinya.

8. Mas Bimo sebagai laboran PTIK yang membantu penyediaan sarana dan prasarana laboratorium PTIK untuk pengerjaan skripsi
9. Mahsadini Putri R. Partner, teman , keluarga yang hebat.
10. Semua pihak yang secara langsung maupun tidak langsung membantu proses penyelesaian skripsi ini.

Saya menyadari bahwa skripsi masih jauh dari sempurna, karenanya saya mengharapkan kritik dan saran yang membangun untuk perbaikan yang lebih baik lagi di masa yang akan datang. Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat dan berguna bagi pembaca serta dapat mendukung kemajuan ilmu pengetahuan khususnya di bidang pendidikan.

Jakarta, 9 Januari 2017
Penulis,

Reza Ridwansyah
5235134442

ABSTRAK

Reza Ridwansyah, *Perbandingan Algoritma Systematic Clustering dengan One Pass K-Means Pada Model K-Anonimity*. 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.

Penelitian ini bertujuan untuk membandingkan model *K-Anonimity* dengan algoritma *Systematic Clustering* dan *One Pass K-Means.* Model ini dapat mengatasi masalah privasi data pada data sensitif yang dipublikasikan dan berbentuk *microdata.* Metode penelitian yang digunakan adalah eksperimen labratorium. Setelah kedua algoritma dibangun, kedua algoritma tersebut diuji performanya dengan menggunakan *Information Loss Matrix.* Penelitian ini dilakukan di Laboratorium Multimedia Teknik Elektro Universitas Negeri Jakarta pada semester ganjil (105) tahun ajaran 2016/2017. Penelitian ini dilakukan dengan mengumpulkan data kemudian menormalkannya dengan bentuk CSV(*Comma Separated View).* Data yang digunakan untuk penelitian merupakan *dataset 'adult'* pada tahun 1994 yang diambil dari UCI *Machine Learning.* Hasil perhitungan *information loss* dari algoritma *One Pass K-Means* diperoleh nilai terendah sebesar 20160.3 dan nilai tertinggi sebesar 20173.3, sedangkan dari algoritma *Systematic Clustering* diperoleh nilai terendah sebesar 8554.20 dan tertinggi sebesar 15490.846. Kesimpulan dari penelitian ini adalah algoritma *Systematic Clustering* lebih baik dari *One Pass K-Means* dalam membangun model *K-Anonimity.*

Kata kunci: K-Anonimity, *Information Loss, Systematic Clustering, One Pass K-Means.*

ABSTRACT

Reza Ridwansyah, Systematic Clustering Algorithm with One Pass K Means Comparative on K-Anonimity Data Model. Thesis. Jakarta, the Education of Informatics Engineering and computers, Faculty of engineering, State University of Jakarta, 2017. Supervising Lecture: Widodo, M. Kom and Bambang p. Adhi, M. Kom.

This research aimed to compare the K-Anonimity Model with Systematic Clustering algorithm and One Pass K-Means. This model can resolve privacy concerns data on sensitive data that are published with microdata form. After the algorithm was built, the two model will be evaluated the performance with Information Loss Matrix / Information Loss. The method used in the research was a laboratory experiment. This research was conducted in the laboratory of Multimedia in electrical engineering State University of Jakarta in odd semester (105) 2016/2017 period. This research was done by collecting data and then data was normalized with CSV (Comma Separated View) format. The dataset used for the research was a dataset 'adult' in 1994 taken from the UCI Machine Learning. The result of information loss calculation algorithm K-Means One Pass obtained the lowest value was 20160.3 and the highest value was 20173.3, The lowest Information Loss value of Systematic Clustering was 8554.20 and the highest values was 15490,846. So, it can be concluded that Systematic Clustering algorithms are better than One Pass K-Means in the build K-Anonimity Model .

Keywords: K-Anonimity, Information Loss, Systematic Clustering, One Pass K-Means.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pada umumnya sebuah instansi memiliki data-data yang perlu dipublikasikan ke khalayak umum dalam selang waktu tertentu. Data tersebut ada yang sudah diolah menjadi informasi dalam bentuk grafis sehingga mudah dibaca, ada juga yang berupa data mentah berisi tabel-tabel lengkap dengan informasi yang ada di dalamnya. Dalam data yang dipublikasikan kepada masyarakat umum tersebut ada kemungkinan terdapat data yang sifatnya sensitif. Maksud dari sensitif adalah bila data ini dipublikasikan ke luar, orang atau instansi tertentu pemilik data tersebut akan merasa malu karena data yang dipublikasikan bersifat aib. Misalnya bila sebuah rumah sakit mempublikasikan data penyakit pasien, ternyata penyakit yang dimiliki oleh pasien tersebut merupakan penyakit yang bersifat aib seperti HIV atau Sifilis. Data tersebut akan menjadi informasi yang kurang baik, karena orang yang memiliki data itu akan merasa malu bila informasinya diketahui banyak orang.

Dalam paper dari Sweeney *K-Anonymity: A Model For Protecting Privacy* pada tahun 2002, ada beberapa cara yang dilakukan untuk meminimalisir kemungkinan data sensitif tersebut diketahui siapa pemiliknya. Cara paling dasar adalah menghapus identitas dari pemilik data tersebut, dengan menghapus identitas dari pemilik data maka kita tidak akan mengetahui siapakah pemilik data tersebut, untuk contohnya kita bisa melihat di Tabel 1.1 dan Tabel 1.2.

Tabel 1.1 Tabel Data Pasien Sebelum Penghapusan Identitas

Rec.No	Name	Age	Gender	Zip Code	Disease
1	George	35	M	302023	HIV
2	Barbara	31	F	302025	Stomach Cancer
3	Charles	29	M	302020	Bronchitis
4	Esra	33	F	302022	Pneumonia
5	Febi	24	M	302018	Stomach Cancer
6	Mike	30	M	302020	Flu
7	Peter	27	M	302020	Pneumonia
8	Polat	25	M	302018	Stomach Cancer
9	Jessica	21	F	302018	Stomach Cancer
10	Jack	26	M	302019	Gastiris

Sumber: K-anonymity: a model for protecting privacy *hal 9 (Dengan Penyesuaian)*

Tabel 1.2 Tabel Data Pasien Setelah Penghapusan Identitas

Rec.No	Age	Gender	Zip Code	Disease
1	35	M	302023	HIV
2	31	F	302025	Stomach Cancer
3	29	M	302020	Bronchitis
4	33	F	302022	Pneumonia
5	24	M	302018	Stomach Cancer
6	30	M	302020	Flu
7	27	M	302020	Pneumonia
8	25	M	302018	Stomach Cancer
9	21	F	302018	Stomach Cancer
10	26	M	302019	Gastiris

Sumber: K-anonymity: a model for protecting privacy *hal 9 (Dengan Penyesuaian)*

Dari dua tabel tersebut, kita bisa melihat bahwa dengan menghapus identitas dari Tabel 1.1 dan mengubahnya menjadi seperti Tabel 1.2 telihat sudah cukup untuk menyembunyikan identitas pemilik data sensitif. Data tidak diketahui siapa pemiliknya sehingga privasi data tidak perlu dikhawatirkan lagi namun cara ini masih memiliki kelemahan, yaitu adanya kemungkinan data eksternal yang dapat dicocokan dengan data yang sudah ada. Misalnya Tabel 1.2 sudah dipublikasikan.

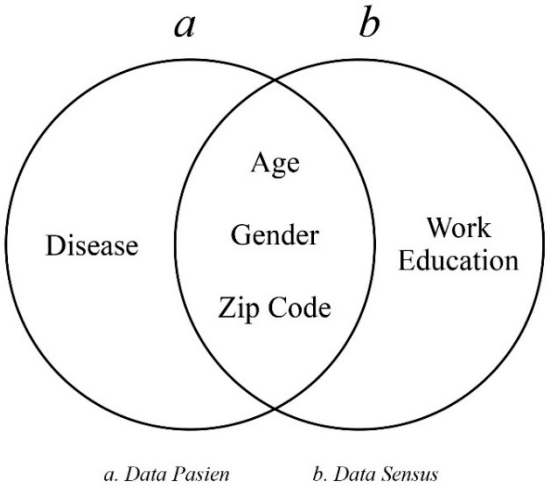
Tidak ada identitas yang terbawa didalamnya, namun di luar itu ada tabel lain. Misal dalam kasus ini terdapat data sensus penduduk yang bisa diakses secara bebas data tersebut menjadi data eksternal.

Tabel 1.3 Data Sensus

Rec.No	Name	Age	Gender	Zip Code	Work	Education
1	George	35	M	302023	Programmer	Bachelor
2	Barbara	31	F	302025	Nurse	High-School
3	Charles	29	M	302020	Hacker	High-School
4	Esra	33	F	302022	Lecture	Professor
5	Febi	24	M	302018	Teacher	Bachelor
6	Mike	30	M	302020	Doctor	Bachelor
7	Peter	27	M	302020	Driver	High-School
8	Polat	25	M	302018	Wrestler	High-School
9	Jessica	21	F	302018	Boxer	High-School
10	Jack	26	M	302019	Postman	High-School

Sumber: K-anonymity: a model for protecting privacy *hal 9 (Dengan Penyesuaian)*

Dengan dua tabel tersebut (Tabel 1.1 dan Tabel 1.2.), jika kita menghubungkannya akan terbentuk sebuah irisan yang terdiri dari atribut dua tabel tersebut (Gambar 1.1.), atribut yang beririsan adalah *Age* ,*Gender* dan *Zip Code,* dari irisan yang terbentuk tersebut kita dapat menebak data pasien dengan mencocokan atribut yang beririsan dari Tabel 1.3 dan Tabel 1.2. Dari hasil pencocokan tersebut kita dapat mengetahui siapa pemilik data sensitif. Misal kita akan mencari orang yang memiliki penyakit HIV. Kita dapat melakukan pencocokan data dengan mencari orang yang memiliki atribut *age* 35, *sex* laki-laki dan *zip code* 302023 di Tabel 1.2 dan Tabel 1.3. Setelah dilakukan pencarian dan dilakukan pencocokan data antara Tabel 1.2 dan Tabel 1.3 ditemukan bahwa pemilik data dengan atribut yang ditentukan adalah George.



Gambar 1.1. Irisan Data Pasien dan Data Sensus

Cara selanjutnya adalah dengan menghapus data sensitif tersebut sehingga penerbit data rumah sakit tidak perlu lagi mengkhawatirkan privasi data sensitif yang dipublikasikan. Untuk lebih jelasnya bisa dilihat Tabel 1.4

Tabel 1.4 Tabel Data Pasien Dengan Penghapusan Atribut Sensitif

Rec.No	Name	Age	Gender	Zip Code
1	George	35	M	302023
2	Barbara	31	F	302025
3	Charles	29	M	302020
4	Esra	33	F	302022
5	Febi	24	M	302018
6	Mike	30	M	302020
7	Peter	27	M	302020
8	Polat	25	M	302018
9	Jessica	21	F	302018
10	Jack	26	M	302019

Sumber: K-anonymity: a model for protecting privacy *hal 9 (Dengan Penyesuaian)*

Dengan menghapus atribut sensitif, privasi data sensitif akan terjaga. Masalah yang muncul adalah esensi data akan hilang, karena data tidak layak

digunakan lagi. Misal data yang dipublikasikan tersebut akan menjadi bahan data sebuah penelitian medis. Data penyakit merupakan data kunci yang sangat menentukan hasil dari penelitian tersebut, dengan hilangnya atribut penyakit maka penelitian tersebut menjadi gagal. Hal ini juga berlaku dalam *data mining.* Hilangnya atribut penyakit sebagai atribut yang akan diteliti menyebabkan data tidak bisa diolah.

Dalam tulisan Sweeney *K-Anonymity: A Model For Protecting Privacy* permasalahan tersebutt dipecahkan menggunakan metode yang terdapat dalam pendekatan *Privacy-Preserving Data Mining* dengan menggunakan model *K-Anonimity*. Teknik lain yang terkenal dalam pendekatan tersebut diantaranya *Additive-Noise-based Perturbation* dan *Cryptography-based Perturbation.*

Cara yang digunakan dalam *Additive-Noise-based Perturbation* dan *Cryptography-based Perturbationn* menekankan pada pengacakan isi data dengan pola tertentu. Hal ini menyebabkan data dihasilkan akan sulit untuk digunakan kembali. Karena data yang akan dihasilkan akan sulit untuk dikenali. Untuk melihat contohnya kita bisa melihat pada Tabel 1.5.

Tabel 1.5 Data yang Diberikan Noise

Age	Gender	Zip Code	Disease
9x	Sdasja	12793871	Asjdhas
Y7	Jdcjhs	12793871	jkaskndk
X9	Ajdkd	Hsjd7647	Ajkasdhk
7x	Sjdhjk	Sdas273ks	Asjdhaih

Dari data yang ada di Tabel 1.5 kita bisa melihat bahwa, isi dari data sudah tidak kenali lagi, sehingga identitas dari pemilik data menjadi aman. Teknik ini memiliki kekurangan, yaitu data yang dipublikasikan menjadi tidak berarti lagi.

Hal ini terjadi karena data sudah kehilangan makna dalam jumlah besar dan tidak bisa diolah menjadi informasi karena *noise* yang diberikan terlalu besar.

Cara pemberian *noise* pada data merupakan cara yang paling dekat dengan tujuan, namun besarnya *noise* harus efisien. Maksudnya adalah walaupun data tersebut diberikan *noise*, data tersebut harus masih bisa digunakan dan dikenali. Oleh karena itu diperlakukan teknik untuk melakukan privasi data yang bisa menjaga privasi data sensitif dengan kehilangan makna data yang minim.

Teknik yang paling tepat untuk melakukan privasi data dalam kasus ini adalah model *K-Anonimity* data. Teknik ini dipilih karena teknik ini memiliki kecenderungan untuk menghilangkan makna data seminim mungkin agar data yang dihasilkan dari model *K-Anonimity* ini masih bisa dipakai**.**

Tabel 1.6 Contoh Model *K-Anonimity*

Group	Job	Sex	Age	Disease
1	Professional	Male	[35-40]	Hepatitis
1	Professional	Male	[35-40]	Hepatitis
1	Professional	Male	[35-40]	Flu
2	Artist	Gender	[30-35]	HIV
2	Artist	Gender	[30-35]	HIV
2	Artist	Gender	[30-35]	HIV

Sumber: A methodology for identifying and improving occupant behavior in residential buildings.

Bisa dilihat pada Tabel 1.6**,** data yang dianonimkan dengan model *K-Anonimity* masih dapat digunakan namun privasi dari data tetap terjaga karena akan sulit untuk menebak kepemilikan dari data sensitif. Misal bila kita melihat di kolom *Job.* Data yang ada dalam kolom tersebut bersifat umum, contohnya *Professional. Professional* bisa berarti pekerjaan apa saja yang bersifat *Professional* begitu juga *Artist.* Musisi bisa kita sebut sebagai *Artist.* Bila kita melihat pada kolom *Age* kita bisa melihat bahwa umur dari pemilik data baris pertama berkisar pada 35-40 yang artinya ada 6 kemungkinan umur yang bisa

dimiliki. Dengan menggunakan model *K-Anonimity* data yang diberikan *noise* tidak kehilangan makna secara keseluruhan.

Untuk membangun model *K-Anonimity* ada dua algoritma yang bisa digunakan, yaitu *Systematic Clustering* dan *One Pass K-Means*. Kedua algoritma tersebut dapat membangun model *K-Anonimity* dengan cara yang berbeda. *Systematic clustering* menggunakan metode yang terstruktur dengan mengambil data dengan cara *clustering* yang kaku. Sedangkan untuk *K-Anonimity* memiliki metode untuk menganonimkan data berdasarkan kedekatan data. Dengan cara yang berbeda ini maka performa dari algoritma ini perlu dianalisis, apakah ada perbedaan performa untuk membangun model yang sama.

Performa yang terpenting adalah besarnya makna data yang hilang. Algoritma yang baik akan menghasilkan data yang anonim namun tetap meminimalisir makna data yang hilang.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah dikemukakan, dapat diidentifikasi beberapa permasalahan sebagai berikut:

1. Model yang ditawarkan sebelumnya belum bisa mengatasi masalah publikasi data sensitif
2. Sulitnya mempublikasikan data yang mengandung data sensitif tanpa menyebabkan data sensitif itu diketahui siapa pemiliknya
3. Makna data yang hilang, diakibatkan data yang semakin anonim.

1.3. Batasan Masalah

Untuk memfokuskan pembahasan dalam penelitian ini penulis membatasi masalah pembahasan adalah:

1. Algoritma yang digunakan adalah algoritma *Systematic Clustering* dan *One Pass K-Means*
2. Untuk menghitung hilangnya makna data yang hilang dari tiap algoritma menggunakan *Information Loss Matrix.*

1.4. Perumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah:

Bagaimana perbandingan kinerja algoritma *One Pass K-Means* dan *Systematic Clustering* pada model *K-Anonimity*?

1.5. Tujuan Penelitian

Berdasarkan perumusan masalah yang telah dipaparkan maka tujuan penelitian ini adalah untuk memperoleh algoritma terbaik dengan indikator *information loss matrix* yang kecil.

1.6. Manfaat Penelitian

Adapun manfaat dari penelitian ini mengandung dua manfaat, yaitu manfaat teoritis dan juga manfaat praktis.

1. Manfaat Teoritis

   Memberikan informasi kepada para pembaca dalam membangun model *K-Anonimity* data dengan algoritma *Systematic Clustering* dan algoritma *One Pass K-Means.*

2. Manfaat Praktis

   Dapat mempublikasikan data dengan menjaga privasi data yang bersifat sensitif dengan meminimalkan makna informasi yang hilang dari data tersebut. Sehingga data tersebut masih layak pakai.

BAB II
TINJAUAN PUSTAKA

2.1. Teori

2.1.1. Model

Menurut Simamarta (1983:ix – xii), model adalah abstraksi dari sistem sebenarnya, dalam gambaran yang lebih sederhana serta mempunyai tingkat prosentase yang bersifat menyeluruh, atau model adalah abstraksi dari realitas dengan hanya memusatkan perhatian pada beberapa sifat dari kehidupan sebenarnya. Menurut Gordon (1994:204), model adalah suatu kerangka utama informasi sistem yang dikumpulkan untuk mempelajari sistem tersebut.

Menurut Ackoff, dkk (1962:108), model dapat dipandang dari tiga jenis kata yaitu sebagai kata benda, kata sifat dan kata kerja. Sebagai kata benda, model berarti representasi atau gambaran, sebagai kata sifat model adalah ideal, contoh, teladan dan sebagai kata kerja model adalah memperagakan, mempertunjukkan.

Dari tiga pendapat tersebut, dapat ditarik bahwa model adalah suatu representasi yang memadai dari suatu sistem dan dikatakan memadai jika telah sesuai dengan tujuan sistem tersebut.

2.1.2. Data

Menurut Arikunto (2006:229), data merupakan segala fakta dan angka yang dapat dijadikan bahan untuk menyusun suatu informasi, sedangkan informasi adalah hasil pengolahan data yang dipakai untuk suatu keperluan.

Menurut Irmansyah (2003:14), secara umum, pengertian data dapat didefinisikan sebagai nilai (*value*) yang merepresentasikan deskripsi dari suatu objek atau peristiwa. Data dibentuk dari data mentah (*raw data*) yang berupa angka, karakter, gambar, atau bentuk lainnya. Data adalah bentuk jamak dari

datum. Data merupakan keterangan-keterangan tentang suatu hal, dapat berupa sesuatu yang punya makna. Data dapat diartikan sebagai sesuatu yang diketahui atau yang dianggap atau anggapan.

Menurut Inmon (2005: 493), data adalah kumpulan dari fakta, konsep, atau instruksi pada penyimpanan yang digunakan untuk komunikasi, perbaikan dan diproses secara otomatis yang mempresentasikan informasi yang dapat dimengerti oleh manusia.

Jadi, dapat disimpulkan bahwa data merupakan sejumlah informasi yang dapat memberikan gambaran tentang suatu keadaan, atau masalah baik yang berbentuk angka-angka maupun yang berbentuk kategori atau keterangan. Data bisa juga didefinisikan sebagai sekumpulan informasi atau nilai yang diperoleh dari pengamatan (observasi) suatu objek.

Data untuk publikasi dibedakan menjadi dua tipe. Yaitu *microdata* dan *macrodata*.

1. Microdata

Data yang disajikan dalam bentuk tabel yang mengandung data identitas individu atau kelompok.

Tabel 2.1. *Microdata*

Rec.No	*Name*	*Age*	*Gender*	*Zip Code*	*Disease*
1	George	35	M	302023	Flu
2	Barbara	31	F	302025	Stomach Cancer
3	Charles	29	M	302020	Bronchitis
4	Esra	33	F	302022	Pneumonia
5	Febi	24	M	302018	Stomach Cancer

Sumber : An Efficient Clustering Method for k-Anonymization 2007

2. Macrodata

Data yang sudah diagregasikan, sehingga tidak berkaitan secara langsung dengan identitas individu atau kelompok.

Tabel 2.2. *Macrodata*

	Hypertension	Obesity	Chest Pain	Short Breath	Tot
M	1	2	2	1	6
F	1	2	0	2	5
Tot	2	4	2	3	11

(a) number of respondents with a disease

	Hypertension	Obesity	Chest Pain	Short Breath	Tot
M	9.1	18.2	18.2	9.1	54.6
F	9.1	18.2	0	18.2	45.4
Tot	18.2	36.4	18.2	27.2	100

(b) percentage of respondents with a disease

Sumber : An Efficient Clustering Method for k-Anonymization 2007

2.1.3. Privasi

Menurut Altman (1975 : 221), privasi adalah proses pengontrolan yang selektif terhadap akses kepada diri sendiri dan akses kepada orang lain. Sedangkan menurut Garner (1999: 74-75), privasi adalah hak untuk sendiri, hak seseorang untuk bebas dari keterbukaan publik. Aturan “hak privasi” merupakan aturan umum meliputi berbagai hak yang diakui dan melekat dalam konsep kebebasan untuk diperintah, dan beberapa hak yang mencegah pemerintah mencampuri urusan setiap individu untuk berhubungan dan berinteraksi, kebebasan individu, untuk membuat pilihan hidup yang diterapkan pada dirinya, keluarganya, dan hubungannya dengan individu lain.

Berdasarkan pemaparan di atas, dapat disimpulkan bahwa privasi merupakan hak dasar yang dimiliki setiap individu untuk menyimpan sesuatu yang bersifat pribadi dan tidak boleh diketahui orang lain yang dijamin oleh negara dan diatur dalam peraturan perundang-undangan yang berlaku.

2.1.4. Anonim

Menurut kamus bahasa Indonesia (KBBI) makna anonimitas adalah sebagai berikut.

(1) tanpa nama; tidak beridentitas; awanama;

(2) [Sos]tidak ada penandatangannya,

Anonimitas data adalah, suatu keadaan dimana data tidak diketahui siapa pemiliknya. Data tersebut tidak memiliki identitas yang jelas untuk menunjukan siapa pemilik datanya. Identitas bisa ditunjukan dari atribut data tersebut.

2.1.5. Algoritma

Menurut Rinaldi Munir (2011:10), para ahli bahasa menemukan kata algorism berasal dari nama cendikiawan muslim yang terkenal yaitu Abu Ja'far Muhammad Ibnu Musa Al-Khuwarijmi (Al-Khuwarijmi dibaca oleh orang Barat menjadi algorism) dalam bukunya yang berjudul Kitab Aljabar Wal-Muqabala, yang artinya *"Buku Pemugaran dan Pengurangan"* (*The book of restoration and reduction*). Dari judul buku itu kita memperoleh kata *"aljabar"* (*algebra*). Perubahan dari kata algorism menjadi algorithm muncul karena kata algorism sering dikelirukan dengan arithmetic sehingga akhiran –sm berubah menjadi –thm.

Menurut Cormen (2009:5), algoritma adalah prosedur komputasi yang mengambil beberapa nilai atau kumpulan nilai sebagai *input* kemudian di proses sebagai *output* sehingga algoritma merupakan urutan langkah komputasi yang mengubah *input* menjadi *output*.

Para ahli matematika meyakini bahwa kata algorism berasal dari nama penulis buku berkebangsaan Arab yang terkenal yaitu Abu Ja'far Muhammad Ibnu Musa

Al-Khwarizmi (770-840 M), orang barat melafalkan Al-Khwarizmi sebagai Algorism. Pengertian algoritma adalah logika, metode, dan tahapan (urutan) sistematis yang digunakan untuk memecahkan suatu permasalahan. Algoritma dapat juga diartikan sebagai urutan langkah secara sistematis dan logis.

Algoritma berusaha melakukan langkah-langkah seefisien mungkin untuk mencapai tujuan semaksimal mungkin. Algoritma sebenarnya implementasi dari kehidupan sehari-hari misalnya algoritma *stack* dan algoritma *queue* yang merupakan implementasi dari antrian dan tumpukan yang terjadi dalam aktifitas sehari-hari.

2.1.6. Privacy Preserving Data Publishing

Privacy Preserving Data Publishing berasal dari 4 kata bahasa Inggris. Yaitu *Privacy, Preserving , Data* dan *Publishing*. *Privacy* dalam bahasa Indonesia dapat diartikan sebagai kerahasiaan pribadi sedangkan *Preserving* berarti melestarikan. Kita juga dapat menganggap *preserving* sebagai menjaga lalu *Publishing* bisa kita artikan sebagai menerbitkan. Sehingga *Privacy Preserving Data Publishing* berarti menjaga kerahasiaan data yang diterbitkan. *Privacy Preserving Data Publishing* (PPDP) merupakan bagian dalam *Data Mining*. Menurut Sweeney (2002: 4), PPDP merupakan cabang ilmu yang menjawab persoalan lembaga yang kesulitan mempublikasikan data yang bersifat rahasia kepada masyarakat, contohnya rumah sakit dan lembaga pemerintahan.

Privacy Preserving Data Publishing merupakan pengembangan model dan algoritma yang mengusulkan sejumlah teknik untuk melakukan tugas-tugas *data mining* menggunakan teknik penganoniman data. Teknik tersebut secara umum terbagi ke dalam beberapa kategori: teknik modifikasi data, metode kriptografi,

dan protokol untuk berbagi data, teknik statistik untuk pengungkapan dan kontrol inferensi, metode *query* audit, pengacakan, dan teknik berbasis gangguan.

Kerahasiaan data menjadi hal yang sangat penting. Karena ada data yang merupakan data sensitif, misal data penyakit pasien. Dalam kasus penyakit pasien, data yang akan dipublikasi merupakan data yang mengandung data personal. Artinya dalam data tersebut akan mengandung identitas orang, lebih spesifiknya adalah identitas orang yang memiliki penyakit. Yang menjadi masalah adalah, tidak semua orang mau diketahui penyakitnya, apalagi bila penyakit tersebut merupakan penyakit yang bersifat aib seperti HIV. Maka diperlukan adanya upaya untuk menjaga privasi data. Untuk menjaga privasi data, maka data harus dibuat menjadi anonim.

Ada dua cara yang umum untuk melakukan privasi data yaitu

2.1.6.1. Perturbative

Teknik *Perturbative* pertama kali dikenalkan oleh Aggarwal dan Srikant (2002) menurutnya *pertubative* merupakan salah satu pendekatan umum dalam *privacy preserving data mining*, dimana *dataset* yang dirilis diberikan *noise* dan hasilnya digunakan untuk melakukan analisis data. Lokesh Patel dan Ravindra Gupta (2013:162) membuat penggolongan teknik *pertubative* menjadi dua yaitu:

a. Aditive Pertubation

Penambahan *noise* dengan menambahkan nilai asli dengan nilai palsu secara acak maupun tidak agar nilai aslinya tidak terlihat. Contoh: Mrs. X yang berumur 35 tahun jika tidak ingin diketahui umurnya ketika dipublikasikan datanya dapat dilakukan penambahan *noise* diawali dengan

Mrs. X menentukan angka acak yaitu 21, kemudian Mrs. X melakukan penambahan noisenya yaitu 35+21= 56.

b. Matrix Multiplicative Perturbation

Matrix Multiplicative Pertubation adalah teknik privasi data dengan melakukan *noise* melalui perkalian *matrix*. *Matrix* yang digunakan juga merupakan *matrix* yang bisa dipilih acak atau tidak.

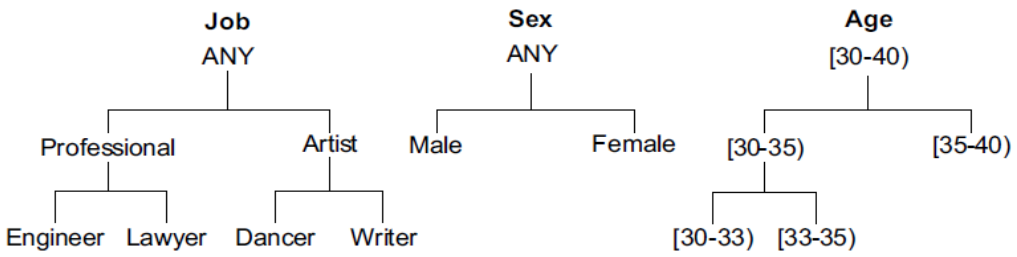
2.1.6.2. Non Pertubative

Menurut Pierangela Samarati dan Sweeney (2002:4), *Non-Pertubative* merupakan teknik penyembunyian data tanpa mengubah esensi dari data tersebut. Teknik yang termasuk golongan *Non-Perturbative* memiliki ciri khas penganoniman data, yaitu dengan mengganti nilai dari deskripsi secara spesifik atau disebut dengan *Quasi-Identifier* (QID). Penggantian nilai dilakukan dengan membuat *taxonomy tree*, yaitu dengan mencari induk atau hal umum dari data tersebut. Teknik yang tergolong dalam *Non-perturbative* ialah:

2.1.6.2.1. Generalization

Generalization merupakan teknik untuk menjadikan data lebih umum. Cara yang paling umum digunakan adalah dengan menggunakan *Taxonomy Tree*. *Generalization* dibangun dengan *Tree*, sehingga *Generalization* menjadi berlevel. Penggunaan *Generalization* umumnya untuk sebuah atribut yang berbentuk nama seperti negara, pekerjaan , umur dan ras. Misal dalam sebuah data terdapat 6 negara. USA, Perancis, Kanada, Indonesia, Jerman dan Jepang. Hal pertama yang perlu dilakukan dalam *Generalization* adalah dengan melihat apakah 6 negara tersebut sama? Bila sama, maka tidak perlu dilakukan *Generalization*.

Tahap selanjutnya untuk melakukan *Generalization* dengan *Tree* harus dikelompokkan terlebih dahulu, masuk ke dalam benua apakah negara yang akan digeneralisasi. USA dan Kanada masuk ke dalam benua Amerika, sedangkan Jerman dan Perancis masuk ke dalam benua Eropa lalu Jepang dan Indonesia masuk ke dalam benua Asia. Setelah digeneralisasi di dalam level benua, maka dilakukan pengecekan kembali. Apakah 6 negara tadi memiliki benua yang sama ataukah berbeda. Bila memiliki benua berbeda, maka 6 negara tersebut akan digeneralisasi menjadi dunia namun bila benua nya sama akan digeneralisasi menjadi nama benua nya misal Asia. *Generalization* juga memiliki bentuk lain, untuk lebih jelasnya kita bisa melihat Gambar 2.1.



Sumber : Privacy Preserving Data Publishing : A survey of Recent Development

Gambar 2.1. *Generalization*

Untuk data yang berbentuk pekerjaan, kita dapat melakukan hal yang sama dengan negara yaitu dengan membuat kategori. Sedangkan untuk kelamin yang hanya memiliki dua pilihan saja, bisa langsung menggeneralisasikan menjadi *any* atau kelamin dan untuk data yang berupa angka, maka yang dilakukan adalah dengan membuat *range* untuk angka tersebut.

2.1.6.2.2. Supression

Supression adalah teknik untuk menjadikan data lebih umum dengan mengganti nilai atribut/karakter dengan atribut tertentu yang berakibat menutupi data sebenarnya, misalnya dengan '*'. Dengan menggunakan *Supression* maka sebuah atribut tidak langsung berganti nilai secara keseluruhan melainkan berganti sesuai karakter. *Supression* digunakan untuk atribut yang merupakan suatu nomor berstandar. Maksudnya berstandar adalah memiliki aturan tertentu seperti panjang karakter atau aturan kode nomor. Contoh dari atribut ini adalah nomor identitas, kodepos atau nomor telpon rumah. Misal dalam data sebuah universitas akan digeneralisasi nomor registrasi mahasiswa. Nomor registrasi mahasiswa itu adalah 5235134442, 523513432, 5235134441 dan 5235134443. Maka tahap pertama yang akan dilakukan adalah mengecek nomor apakah 4 nomor registrasi tersebut identik? Bila tidak, maka nomor yang paling belakang akan di *Supression* menjadi 523513444*, 52351343*, 523513444* dan 523513444*. Setelah itu cek lagi, apakah 4 nomor registrasi tersebut sudah sama? Bila belum masih sama. Maka nomor registrasi akan di *Supression* lagi menjadi 5235134**, 5235134**, 5235134** dan 5235134**. Ternyata sudah sama. Maka hasil ini akan menjadi hasil akhir.

Tabel 2.3. *Supression* Nomor Registrasi

Tahap 2	52351344**	52351344**	52351344**	52351344**
Tahap 1	523513444*	523513443*	523513444*	523513444*
Tahap 0	5235134442	5235134432	5235134441	5235134443

Suppression juga disebut menutup data, maksudnya adalah satu atau beberapa nilai yang terdapat pada atribut tertentu ditutup atau diganti dengan

suatu lambang, contohnya: * atau # sesuai dengan kondisinya. Jika satu karakter saja yang di tutup, maka dapat membuat suatu tabel tidak dapat terhubung ke tabel lainnya, sehingga satu *Suppression* dianggap sudah cukup. Hal tersebut tidak berlaku jika tabel-tabel yang ada dapat saling terhubung setelah melakukan *Supression*, maka nilai sebelumnya harus ditutup sampai tabel-tabel yang ada tidak dapat terhubung. Dalam pengaplikasiannya, baik *Supression* maupun *Generalization,* ada dua cara yang umum digunakan, yaitu *Global Recording* dan *Local Recording.*

a. *Global Recording*

Global recording merupakan teknik untuk memetakan seluruh nilai atribut kategori *Quasi-Identifier* menjadi nilai yang lebih umum dalam hirarki domain *Generalization* dan *Supression*. Contohnya dapat dilihat pada Tabel 2.4. yang menyajikan contoh *Global Recording* untuk atribut *ZipCode*, *Gender, Age*, dan *Education*.

Tabel 2.4 Bentuk *Supression* dan *Generalization* Dengan *Global Recording*

ZipCode	Gender	Age	Education	Disease
435*	Person	[21-40]	Educated	Flue
435*	Person	[21-40]	Educated	Cancer
435*	Person	[21-40]	Educated	HIV+
435*	Person	[21-40]	Educated	Diabetes
435*	Person	[21-40]	Educated	Diabetes

Sumber: Data Penelitian Enamul Kabir, Springer © 2011 (Dengan Penyesuaian)

Dalam contoh Tabel 2.4. kita *melihat* ada dua teknik yang dipakai, yaitu *supression* dan *generalization.* Dua-duanya dipakai secara *global*, artinya semua atribut kecuali yang merupakan atribut sensisitif yang berada dalam tabel tersebut memiliki nilai yang sama.

b. *Local Recording*

Agak sedikit berbeda dengan *global recording* yang mengeneralisasikan semua data yang ada didalam tabel. *Local Recording* adalah teknik melakukan *generalization* dan *supression* untuk setiap kelompok *quasi identifier* yang berbeda. Contohnya dapat dilihat pada Tabel 2.5. yang menyajikan contoh *local recording* untuk atribut *ZipCode*, *Gender*, dan *Age*.

Tabel 2.5. Bentuk *Supression* dan *Generalization* Dengan *Local Recording*

ZipCode	Gender	Age	Education	Disease
435*	Male	[21-30]	Educated	Flue
435*	Male	[21-30]	Educated	Cancer
435*	Male	[21-30]	Educated	HIV+
43**	Person	[31-40]	Educated	Diabetes
43**	Person	[31-40]	Educated	Diabetes
43**	Person	[31-40]	Educated	Diabetes

Sumber: Data Penelitian Enamul Kabir, Springer © 2011 (Dengan Penyesuaian)

Kita bisa melihat perbedaan Tabel 2.4. dengan Tabel 2.5. Dalam Tabel 2.5, kita melihat *Supression* dan *Generalization* dilakukan dalam kelompok teretentu. Sehingga hasilnya juga berbeda. Bila kita melihat contoh diatas, setiap 3 baris data memiliki *Quasi Identifier* yang sama.

2.1.7. K-Anonimity

K-Anonimity Merupakan teknik *Privacy Preserving Data Publishing* yang bersifat *Non-Pertubative*. Artinya data yang akan dianonimkan tidak kehilangan makna secara keseluruhan. Untuk mendapatkan model *K-Anonymity*, ada berbagai teknik diklasifikasikan menjadi dua jenis: *Generalization* dan *Supression*. Teknik *Generalization* membangun sistem hirarki nilai-nilai dari sebuah domain nilai atribut berdasarkan generalitas nilai-nilai dan menggantikan nilai-nilai tertentu dengan yang lebih umum. Teknik ini diklasifikasikan menjadi dua tingkat *Generalization*: atribut dan sel tingkat. Tingkat atribut menggantikan nilai domain

saat ini dari atribut *Quasi Identifier* dengan yang lebih umum. Misalnya, domain usia atribut dipetakan dari tahun ke interval 10 tahun. Tingkat sel hanya menggantikan nilai-nilai saat ini dari beberapa sel penting dengan yang lebih umum. Kedua tingkat mendapatkan model *K-Anonymity*.

Menurut Sweeney (2002:8), *K-Anonimity* merupakan perbaikan dari algoritma sebelumnya *K-Map. K-Anonimity* dibuat untuk melindungi privasi data dari adanya kemungkinan pencocokan dengan data *external*. Sebuah tabel data dikatakan memenuhi kondisi *K-Anonymity* jika masing-masing r*ecord* pada grup *Quasi Identifier* data anonim, tidak bisa dibedakan dengan minimal K-1 data lain dalam grup tersebut. Sebuah tabel memenuhi *property k-anonymity* jika ukuran *class equivalence* atau grup *Quasi Identifier*-nya sejumlah k atau lebih.

2.1.8. Jenis Atribut *Microdata*

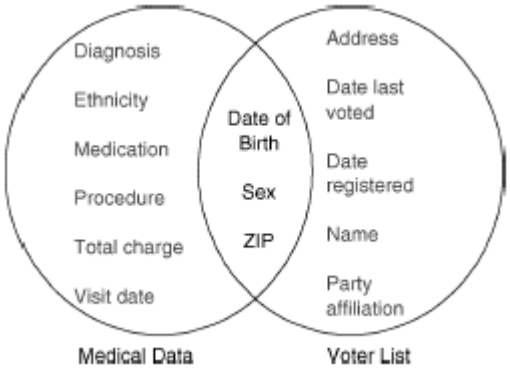
Menurut Charu C. Aggarwal (2015:668), dalam *Privacy Preserving Data Publishing* terdapat beberapa macam atribut yang membedakan satu atribut dengan atribut lainnya. Atribut itu adalah *Explicit Identifier*, *Quasi Identifier*, *Sensitive Attributes,* dan *Non-Sensitive Attributes*.

a. Explicit Identifier

Explicit Identifier merupakan atribut atau tipe data yang secara eksplisit menyatakan identitas dari suatu individu, karena atribut tersebut hampir secara keseluruhan dihilangkan dalam proses sanitasi data, sehingga data tersebut tidak relevan untuk dilakukan penelitian dalam algoritma privasi maksudnya adalah *Expicit Identifier* merupakan atribut yang secara eksplisit menyatakan identitas dari suatu individu, misalnya nomor registrasi, nama, NIP dan NIK (Nomor Induk Kependudukan).

b. Quasi Identifier

Pseudo Identifier atau *Quasi Identifier* merupakan atribut yang tidak menunjukkan identitas seseorang secara eksplisit dalam sebuah isolasi, namun atribut tersebut dapat dilakukan penggabungan dengan beberapa atribut lain. Jika atribut tersebut dilakukan penggabungan atau *join* dengan informasi yang beredar dimasyarakat, maka dapat membentuk suatu identifikasi yang berpotensi untuk menunjuk identitas tertentu. Serangan yang dilakukan dengan melakukan *join* informasi pada *Quasi Identifier* dengan informasi yang beredar dimasyarakat sering disebut sebagai *Linkage Attack* atau serangan bertautan.



Sumber : Privacy Preserving Data Publishing : A Survey of Recent Development

Gambar 2.2. *Linkage Attack*

Kita dapat mengatakan bahwa *Quasi Identifier* merupakan gabungan beberapa atribut yang bisa berpotensi untuk menunjuk identitas tertentu jika dilakukan *join* dengan data lain.

c. Sensitive Attribute

Atribut sensitif merupakan bagian dari atribut *Quasi-identifier* dan merupakan atribut yang dianggap pribadi dan bersifat rahasia bagi semua orang. Jika atribut sensitif diketahui oleh orang lain maka akan

menyebabkan pemilik data menjadi malu.seperti data mengenai penyakit, data tindak kriminal seseorang, lama hukuman, status disabilitas seseorang, bahkan gaji juga bisa dianggap sebagai atribut sensitif bagi orang tertentu.

d. Non-sensitive Attribute

Non-sensitive Attribute merupakan atribut yang tidak termasuk ke dalam tiga kategori sebelumnya.

2.1.9. *Systematic* Clustering

Menurut Md. Enamul Kabir, dkk (2011:57), algoritma *Systematic Clustering* merupakan algoritma yang tergolong dalam algoritma *clustering*. Algoritma tersebut dirancang untuk memperbaiki algoritma *clustering* sebelumnya yaitu algoritma pada model *K-Anonymity*. Pada model *K-Anonimity* tidak ada batasan yang ketat untuk menjaga jumlah data yang ada dalam sebuah *cluster*. Hal ini menyebabkan besarnya *Information Loss* karena proses *Supression* dan *Generalization* yang cenderung akan menganonimkan data pada level tertinggi.

Systematic Clustering merupakan algoritma yang menggunakan konsep *K-Anonimity*. *Systematic Clustering* bekerja dengan membuat *cluster*/grup yang jumlah anggota nya sebanyak K yang ada. Data yang tersisa kemudian dimasukan kedalam *cluster*/grup yang memiliki *Information Loss* terkecil bila data yang tersisa tersebut masuk kedalam *cluster*/grup tersebut. Pada kebanyakan kasus, bila data sudah diurutkan maka data yang tersisa akan masuk kedalam *cluster* terdekat karena *cluster* terdekat cenderung memiliki data yang berdekatan.

Algoritma dari *Systematic Clustering* adalah :

1. Urutkan data terlebih dahulu berdasarkan *Quasi Identifier* yang berbentuk data numerik dan kontinu.

2. Buatlah sebuah *cluster*, ambil sebanyak K data dari data urutan terkecil.
3. Lakukan *Generalization* atau *Supression* pada data dalam *cluster* yang sama.
4. Ulangi sampai data habis.
5. Bila ada data sisa, maka masukkan data tersebut ke dalam *cluster* terdekat atau *cluster* yang memiliki *Information Loss* terkecil.
6. Selesai.

Pengurutan data bertujuan untuk membuat *Information Loss* dari data semakin kecil dengan cara mendekatkan data.

2.1.10. One Pass K-Means

One Pass K-Means adalah modifikasi dari *K-Means*. *K-Means* merupakan salah satu algoritma *cluster*ing. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok.

Menurut Meng Cheng Wei (2007:6), *One Pass K-Means* merupakan turunan dari *K-Means*. Perbedaan mendasar yang ada dalam *One Pass K-Means* adalah K-Means hanya melakukan satu iterasi. Berbeda dengan *K-Means* yang melakukan iterasi sampai data stabil. *One Pass K-Means* dibuat untuk melengkapi *Greedy K-Member* yang menganonimkan data dengan cara *Global Recording. One Pass K-Means* melakukannya dengan cara *Local Recording.* Secara umum dalam algoritma *One Pass K-Means.* Terdapat dua proses, pertama adalah proses *clustering* kemudian proses penyesuaian. Secara detil untuk melakukan algoritma *One Pass K-Means* adalah sebagai berikut:

1. Tentukan *centroid* dari kumpulan data.
2. Cocokan jarak antara *centroid* dengan data satu persatu
3. Masukan data ke dalam jarak terdekat.
4. Ulangi langkah ketiga sampai semua data masuk kedalam centroid. Data ini akan menjadi *cluster.*

5. Hitung anggota tiap *centroid*. Bila masih ada *centroid* data yang anggotanya kurang dari K yang ditentukan, masukan K tersebut ke dalam *cluster* yang sudah memenuhi K.
6. Ulangi langkah ke 5 sampai tiap *cluster* memnuhi >=K.

2.1.11. Information Loss

Setelah data diolah dengan *Supression* dan *Generalization*, efek samping yang pasti terjadi adalah adanya *Information Loss* atau dalam bahasa Indonesia diartikan sebagai informasi yang hilang.

Menurut Byun (2007:5), *Information Loss* digunakan untuk mengukur kuantitas dari informasi yang hilang selama melakukan proses *K-Anonimity*. Sedangkan menurut Jun-Lin Lin dan Meng-Cheng Wei (2008:47), *Information Loss Metric* atau *Information Loss* merupakan besaran dalam mengukur banyaknya informasi yang hilang akibat penerapan teknik *Generalization* dan *Supression* untuk menjaga privasi data.

Untuk menghitung banyaknya *Information Loss*, dalam tulisan ini akan menggunakan teknik perhitungan *Information Loss* oleh Byun (2007:9).

$$IL(\Omega) = |\ \Omega\ | \cdot \left(\sum_{i=1}^{r} \frac{N_{i_{max}} - N_{i_{min}}}{\eta_{N_{i_{max}}} - \eta_{N_{i_{min}}}} + \sum_{j=1}^{s} \frac{H(\Lambda(\cup_{C_j}))}{H(\tau_{C_j})} \right)$$

Ω adalah banyaknya data dalam suatu *cluster*, N adalah *Quasi Identifier numerical*. Sedangkan C adalah *Quasi Identifier* kategorial. Sehingga

taksonomi tertinggi dari pohon τ .Untuk lebih memahaminya maka digunakan contoh di Tabel 2.5 dan Tabel 2.6.

Tabel 2.6 Data *K-Anonimity* Pasien

Zipcode	Gender	Age	Education	Disease
4350	Male	24	9th	Flue
4351	Male	25	10th	Cancer
4352	Male	26	9th	Hiv +
4350	Male	35	9th	Diabetes
4350	Female	40	10th	Diabetes
4350	Female	38	11th	Diabetes

Sumber : Efficient systematic clustering method for k-anonymization

Tabel 2.7 Data Hasil *K-Anonimity*

Zipcode	Gender	Age	Education	Disease
435*	Person	21-30	Educated	Flue
435*	Person	21-30	Educated	Cancer
435*	Person	21-30	Educated	Hiv +
435*	Person	31-40	Educated	Diabetes
435*	Person	31-40	Educated	Diabetes
435*	Person	31-40	Educated	Diabetes

Sumber : Efficient systematic clustering method for k-anonymization

Tabel 2.5 merupakan tabel data pasien yang belum dianonimkan dengan *K-Anonimity*, sedangkan Tabel 2.6 merupakan tabel yang sudah dianonimkan dengan *Generalization* dan *Supression*. Yang menjadi data sensitif adalah penyakit. Sehingga hasil dari perhitungannya.

3 Adalah Ω yang menyatakan banyaknya jumlah data dalam sebuah *cluster*. 26 – 24 merupakan operasi dari

yang digunakan dalam penelitian ini adalah *One Pass K-Means* dan *Greedy K-Member*. Dalam penelitian ini berfokus pada kekurangan *Greedy K-Member* yang melakukan *clustering* dengan cara *Global Recording* yang mengakibatkan besarnya *Information Loss*. Dengan menggunakan *One Pass K-Means, K-Anonimity* dapat dilakukan secara *Local Recording.* Hasil dari penelitian ini adalah *One Pass K-Means* lebih baik daripada *Greedy K-Member*

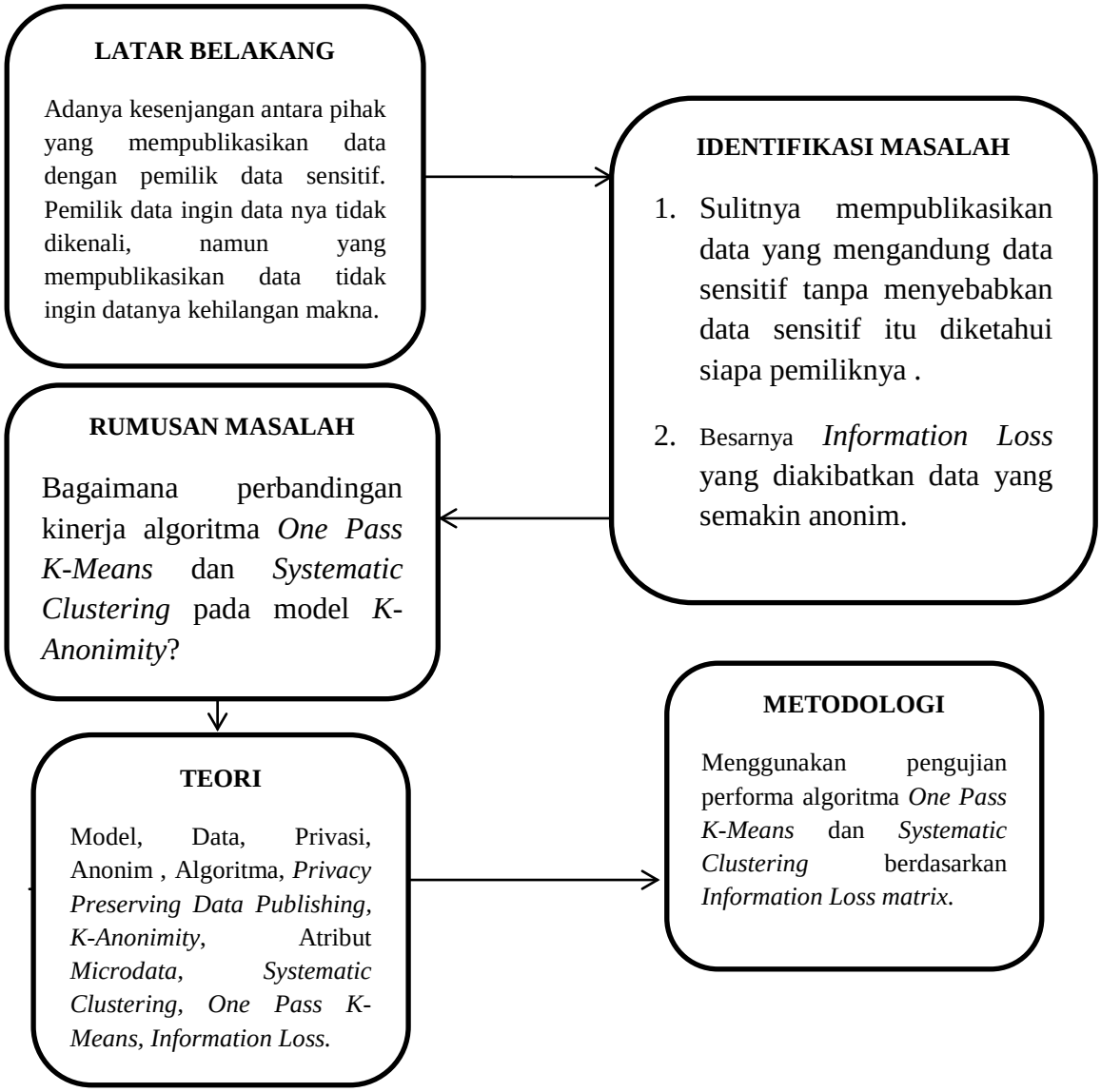
3. ***Efficient Systematic Clustering Method for K-Anonymization*** oleh Md. Enamul Kabir, Hua Wang, dan Elisa Bertino. Penelitian tersebut diperkenalkan oleh Md. Enamul Kabir, Hua Wang, dan Elisa Bertino pada tahun 2009 dan diterbitkan dalam jurnal pada tahun 2011. Model yang digunakan dalam penelitian ini adalah *Systematic Clustering*, yaitu model yang mengelompokkan beberapa *record* menjadi satu *cluster* yang memiliki kesamaan atribut, dan dapat dibedakan dengan *cluster* lainnya. Dalam penelitian ini fokus yang ditekankan adalah besarnya *Information Loss* yang diakibatkan oleh banyaknya anggota dalam sebuah *cluster*. Hal ini terjadi karena tidak ada batasan kuat yang mengatur soal banyaknya anggota dalam *cluster*. Dalam tulisan ini diperkenalkan *Systematic Clustering* yang memiliki kelebihan dibagian *clustering. Systematic Cluster*ing berupaya untuk membuat *cluster* memiliki *Information Loss* yang kecil dengan memaksimalkan kemiripan data yang ada dalam *cluster* tersebut.
4. ***Detecting Achieving K-Anonymity Privacy Protection Using Generalization and Suppression*** oleh Latanya Sweeney. Penelitian tersebut dilakukan oleh Latanya Sweeney pada tahun 2002, merupakan pengembangan penelitian

disertasinya pada tahun 1998. Metode yang digunakan dalam penelitian tersebut adalah metode *Preferred Minimal Generalization Algorithm (MinGen)*. Metode tersebut menggabungkan kedua teknik, yaitu teknik *Generalization* untuk mengganti (merekam) nilai dengan tingkat spesifikasi yang rendah, dan teknik *Supression* untuk tidak mengubah nilai informasi secara keseluruhan. Metode tersebut digunakan untuk menemukan *Generalization* minimal dari sebuah tabel menggunakan model *K-Anonymity* dengan distorsi yang minimum.

2.3. Prosedur

Berdasarkan kajian teori yang telah dijelaskan, perlu adanya perlindungan privasi data sensitif ketika data tersebut akan dipublikasikan, karena data sensitif dapat menimbulkan rasa malu kepada pemiliknya apabila data tersebut diketahui masyarakat umum. Dengan menghapus *Explicit Identifier*nya saja belum cukup, karena masih bisa diserang dengan *linked attack* yaitu menyerang data dengan mencocokan dengan data eksternal. Cara *pertubative* dianggap kurang layak, karena esensi data akan hilang lewat *noise* yang diberikan. Oleh karena itu, dipilihlah cara *non perturbative*, dengan model *K-Anonimity*. Ada dua algorima yang umum digunakan dalam model *K-Anonimity*, yaitu *One Pass K-Means* dan *Systematic Clustering*. Perlu dilakukan penelitian yang lebih lanjut untuk menentukan mana yang lebih baik diantara keduanya. Maka dilakukan percobaan untuk mengujinya, yaitu dengan menghitung *Information Loss* dari kedua algoritma tersebut. Dengan menghitung *Information Loss* kita dapat melihat sebanyak apa makna informasi yang hilang setelah data dianonimkan dan kita

juga perlu menghitung efisiensi dari dua algoritma tersebut dengan menghitung berapa waktu yang dibutuhkan untuk menyelesaikan tugasnya.



Gambar 2.3. Kerangka Berpikir

BAB III

METODE PENELITIAN

3.1. Tempat dan Waktu Penelitian

1. **Tempat :** Laboratorium Multimedia Gedung L2 Teknik Elektro lantai 3, Fakultas Teknik. Universitas Negeri Jakarta.
2. **Waktu :** 10 Agustus – 30 Oktober 2016

3.2. Alat dan Bahan Penelitian

Untuk menunjang penelitian ini, maka diperlukan alat dan bahan yang akan digunakan dalam penelitian. Alat dan bahan yang akan dipakai adalah :

1. **Laptop**

Pada penelitian ini, digunkan laptop Lenovo Y410P dengan spesifikasi :

Tabel 3.1. Spesifikasi Laptop

Processor	4th generation Intel® Core™ i7-4700MQ (2.40GHz 1600MHz 6MB)
Operating System	Windows 8.1 64
Display/Resolution	14.0" HD+ Glossy LED Backlit with integrated camera (1600x900)
Graphics	NVIDIA® GeForce® GT 755M 2GB
Memory	8GB PC3-12800 DDR3L SDRAM 1600 MHz
Hard Disk Drive	1TB 5400 RPM
Sound	JBL® designed speakers supporting Dolby Home Theatre v4 audio certification for immersive sound

Integrated Communications	• Intel Centrino Wireless N-2230 - 802.11b/g/n Wi-Fi connectivity • Bluetooth 4.0 • 1 GB LAN
Connectors	• 2x USB 2.0 (one always on) • 1x USB 3.0 SuperSpeed • 6-in-1 card reader (SD, SDHC, SDXC, MMC, MS, MS-Pro) • Headphone • Microphone • HDMI-out • VGA port (15-pin)
Camera	HD 720P camera
Battery	Up to 5 hrs
Weight	5.5 lbs
Dimensions	13.77 x 9.64 x 0.59-1.29 inches

Sumber : Shop.lenovo.co.id

2. Internet

Internet digunakan sebagai penunjang untuk melakukan koneksi ke *database* UCI.

3. Dataset "*Adult*" dari UCI *Machine Learning Repository*

UCI *machine learning repository* merupakan sebuah situs web yang menyediakan *dataset* untuk keperluan *machine learning*, *data mining* atau pengolahan algoritma *big data*. UCI dibangun sebagai arsip ftp pada tahun 1987 oleh David Aha. Sejak saat itu banyak yang menggunakan UCI sebagai sumber utama *dataset* untuk melakukan penelitian algoritma atau mempelajari data yang ada didalam *dataset* tersebut untuk menemukan pola-pola yang tidak terlihat. UCI sudah

dikutip lebih dari 10000 kali dalam paper yang berkaitan dengan ilmu komputer.

Oleh karena itu dipilihlah *dataset "adult"* dari UCI *Machine Learning*, sebagai *dataset* untuk membangun model *Systematic Clustering* dan *One Pass K-Means*. Data ini merupakan hasil ekstraksi data sensus penduduk *United States Census Bureau*. Data ini bisa diakses dari http://www.census.gov/ yang disumbangkan oleh Ronny Kohavi and Barry Becker. *Dataset* ini pernah digunakan untuk beberapa penelitian misalnya "*Rakesh Agrawal and Ramakrishnan ikant and Dilys Thomas. Privacy Preserving OLAP. SIGMOD Conference.*" pada tahun 2005 atau "*Bianca Zadrozny. Learning and evaluating classifiers under sample selection bias. ICML"* pada tahun 2004. Selain itu masih banyak lagi paper yang menggunakan *dataset* ini sebagai sumber data penelitian. Untuk lebih lengkapnya bisa dicek pada tautan ini http://archive.ics.uci.edu/ml/datasets/Adult

Penggunaan *dataset adult* didasari oleh bentuk *dataset* ini yang berupa *microdata.* Bentuk ini merupakan bentuk yang paling cocok untuk *privacy preserving data publishing.*

3.3. Diagram Alir Penelitian

Berikut adalah diagram alir penelitian untuk model k-anonimity dengan algoritma *Systematic Clustering* dan *One Pass K-Means*

Gambar 3.1. Diagram Alir Penelitian

Dalam penelitian ini akan dibangun dua model *K-Anonimity* dengan dua algoritma yang berbeda. Untuk melakukan hal tersebut maka harus ada data yang akan digunakan untuk membentuk model *K-Anonimity*. Maka dipilihlah *dataset adult* yang berbentuk *microdata*, setelah data tersebut ditarik maka selanjutnya yang dilakukan adalah melakukan normalisasi data. Normalisasi disini adalah membuat format **.data** menjadi **.csv** yang dapat dipahami oleh DBMS MySQL. Di dalam normalisasi data, data juga diurutkan terlebih dahulu. Selain itu data yang memiliki nilai kosong juga dihapus. Setelah data siap untuk diolah, maka selanjunya data dibuat model *K-Anonimitynya.* Untuk membandingkan dua algoritma tersebut maka digunakanlah *information loss* untuk mengukur seberapa besar informasi yang hilang dari sebuah tabel. Tahap ini dilakukan sebanyak 7 kali dengan jumlah K yang berbeda-beda.

3.4. Teknik dan Prosedur Pengumpulan Data

Pengumpulan data untuk penelitian ini menggunakan data dari UCI *Machine Learning.* Data yang diambil memiliki format **.data** yang berbentuk seperti di bawah ini :

53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K

28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K

Data ini merupakan standar data yang ada di UCI. Namun data dengan format seperti ini tidak bisa dibaca oleh DBMS. Oleh karena itu, data di rubah menjadi berbentuk csv. Data dirubah menjadi bentuk csv, karena DBMS dapat mengenali

data dengan format ini. Ketika data sudah dalam bentuk csv kemudian data dinormalkan lagi dengan menghapus baris yang memiliki *missing values* pada atributnya. Hal ini dilakukan agar proses *clustering* dapat berjalan sesuai dengan yang direncanakan. Setelah data disaring, kemudian data diurutkan berdasarkan umur. Hal ini akan mempermudah algoritma untuk melakukan *clustering*. Data yang sudah selesai akan berbentuk seperti Gambar 3.2.

id_coba	kelamin	umur	negara	pekerjaan	cluster
1	Female	17	United-States	Sales	
2	Male	17	United-States	Other-service	
3	Male	17	United-States	Other-service	
4	Male	17	Mexico	Other-service	
5	Male	17	United-States	Other-service	
6	Male	17	United-States	Handlers-cleaners	
7	Male	17	United-States	Other-service	
8	Male	17	United-States	Sales	
9	Female	17	United-States	Other-service	
10	Male	17	United-States	Farming-fishing	

Gambar 3.2. Data Mentah CSV

Data dalam bentuk .csv ini kemudian dimasukan ke dalam DBMS MySQL dan formatnya berubah menjadi .sql

	id_coba	kelamin	umur	negara	pekerjaan	cluster
Edit Copy Delete	1	Female	17	United-States	Sales	0
Edit Copy Delete	2	Male	17	United-States	Other-service	0
Edit Copy Delete	3	Male	17	United-States	Other-service	0
Edit Copy Delete	4	Male	17	Mexico	Other-service	0
Edit Copy Delete	5	Male	17	United-States	Other-service	0
Edit Copy Delete	6	Male	17	United-States	Handlers-cleaners	0
Edit Copy Delete	7	Male	17	United-States	Other-service	0
Edit Copy Delete	8	Male	17	United-States	Sales	0

Gambar 3.3. Data Dalam Bentuk SQL

Data dalam bentuk .sql inilah yang dapat dikenali oleh *driver* JDBC dari Java. Data ini kemudian akan dilakukan *clustering* dengan menggunakan *Systematic Clustering* dan *One Pass K-Means*.

3.5. Teknik Analisis Data

Untuk menentukan mana yang lebih baik perlu adanya parameter untuk dianalisis. Parameter yang akan dianalisis adalah *Information Loss* dan *Running Time Program*. *Information Loss* akan dihitung sebanyak 7 kali kemudian jumlah *Information Loss* yang didapat dari hasil penelitian dibandingkan.

Tabel 3.2. Form Perbandingan

Jumlah K	Systematic Clustering	One Pass K-Means
3		
4		
5		
6		
7		
8		
9		

Untuk melakukan perhitungan *Information Loss.* Kita menggunakan rumus yang digunakan Pawan R. Bhaladhare pada tulisannya yang berjudul *Novel Approaches for Privacy Preserving Data Mining in k-Anonymity Model.*

$$IL(\Omega) = \mid \Omega \mid . \left(\sum_{i=1}^{r} \frac{N_{i_{max}} - N_{i_{min}}}{\eta N_{i_{max}} - \eta N_{i_{min}}} + \sum_{j=1}^{s} \frac{H(\Lambda(\cup C_j))}{H(\tau C_j)} \right)$$

Sebagai acuan untuk melakukan perhitungan *Information Loss* pada atribut non numerik maka dibuatlah daftar taksonomi untuk atribut yang akan diteliti.

Tabel 3.3. Acuan Taksonomi Untuk *Information Loss*

Atribut	**Cabang**	**Nilai taksonomi**
Kelamin	Female, Male	0
	Kelamin	1
Negara	Thailand, Vietnam, USA, German dan lain-lain	0
	America, Eropa, Africa dan nama benua lainnya	0,5
	Dunia	1

Selanjutnya untuk

Tabel 3.4. Form Untuk Perhitungan *Running Time*

Jumlah K	Systematic Clustering	One Pass K-Means
3		
4		
5		
6		
7		
8		
9		

Kemudian dari dua parameter tersebut akan dianalisis mana algoritma yang lebih baik. Sama dengan *Information Loss, running time* juga dibuat grafiknya untuk mempermudah analisis. Selain itu dari dua grafik ini kita juga bisa menemukan hal lain selain menentukan mana algoritma yang lebih baik.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Deskripsi Hasil Penelitian

4.1.1 Model Yang Dihasilkan

Produk yang dihasilkan merupakan model algoritma *K-Anonimity* dengan menggunakan algoritma *Systematic Clustering* dan *One Pass K-Means*.

Tabel 4.1. Data Mentah

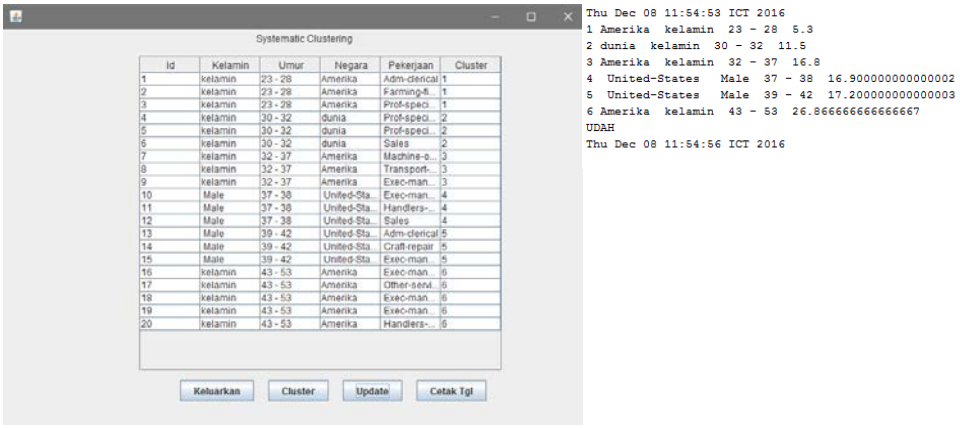
id_coba	kelamin	umur	negara	pekerjaan	cluster
1	Female	23	United-States	Adm-clerical	0
2	Male	25	United-States	Farming-fishing	0
3	Female	28	Cuba	Prof-specialty	0
4	Male	30	India	Prof-specialty	0
5	Female	31	United-States	Prof-specialty	0
6	Male	32	United-States	Sales	0
7	Male	32	United-States	Machine-op-inspct	0
8	Male	34	Mexico	Transport-moving	0
9	Female	37	United-States	Exec-managerial	0
10	Male	37	United-States	Exec-managerial	0
11	Male	38	United-States	Handlers-cleaners	0
12	Male	38	United-States	Sales	0
13	Male	39	United-States	Adm-clerical	0
14	Male	40	United-States	Craft-repair	0
15	Male	42	United-States	Exec-managerial	0
16	Female	43	United-States	Exec-managerial	0
17	Female	49	Jamaica	Other-service	0
18	Male	50	United-States	Exec-managerial	0
19	Male	52	United-States	Exec-managerial	0
20	Male	53	United-States	Handlers-cleaners	0

Hal pertama yang dilakukan adalah menarik data dari *dataset adult*. Setelah *dataset* tersebut berhasil ditarik, maka data dinormalkan menjadi .csv. Untuk melihat contoh cara kerjanya maka kita akan lihat contoh untuk 20 *record* data dengan menggunakan data yang ada pada Tabel 4.1.

Setiap data akan dibuat model *K-Anonimity* dengan tiap algoritma, masing-masing algoritma menggunakan 3-K, yaitu K-3, K-4 dan K-5.

1. Algoritma *Systematic Clustering*

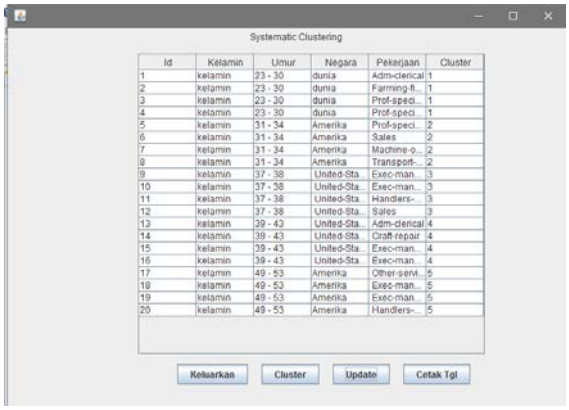
a. K-3



Gambar 4.1. K-3 Pada *Systematic Clustering*

Gambar 4.1 merupakan hasil dari model *K-Anonimity* dengan K-3. Maksud dari K-3 adalah, minimal anggota *cluster* adalah 3 namun diperbolehkan lebih dari kriteria minimum misal 4 atau 5. Kita bisa melihat di dalam *console* (gambar bagian kanan) *Information Loss* yang didapat adalah 26,866. Kita dapat menggunakan waktu *start* dan akhir untuk menghitung berapa waktu eksekusinya. Dalam kasus ini untuk generalisasinya membutuhkan waktu selama 3 detik.

b. K-4

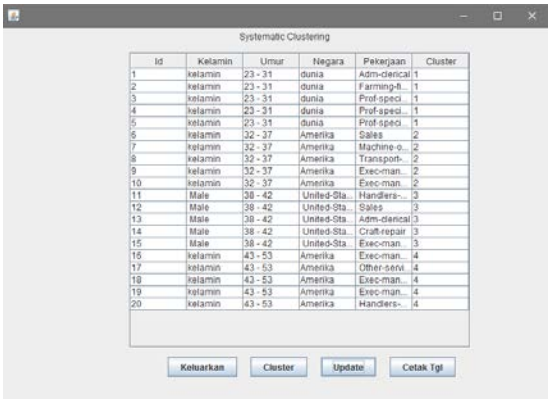



```
Thu Dec 08 11:48:42 ICT 2016
1 dunia  kelamin  23 - 30  8.933333333333334
2 Amerika  kelamin  31 - 34  15.733333333333334
3  United-States  kelamin  37 - 38  19.866666666666667
4  United-States  kelamin  39 - 43  24.4
5 Amerika  kelamin  49 - 53  31.333333333333332
UDAH
Thu Dec 08 11:48:47 ICT 2016
```

Gambar 4.2. K-4 Pada *Systematic Clustering*

Selanjutnya untuk K-4 kita dapat melihatnya pada Gambar 4.2. Untuk K-4 membutuhkan waktu generalisasi selama 5 detik dan *Information Loss* sebesar 31.33.

c. K-5



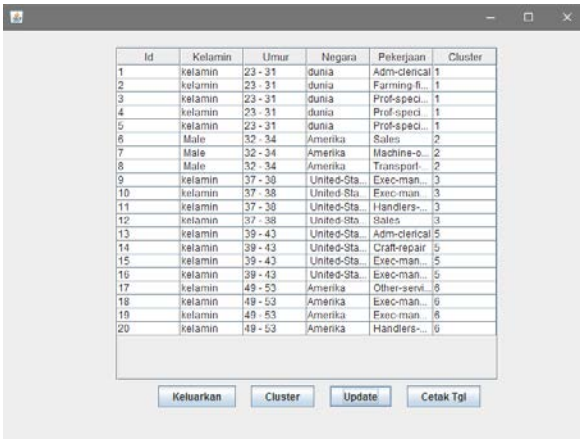

```
Thu Dec 08 11:45:00 ICT 2016
1 dunia  kelamin  23 - 31  11.333333333333332
2 Amerika  kelamin  32 - 37  20.166666666666664
3  United-States  Male  38 - 42  20.833333333333332
4 Amerika  kelamin  43 - 53  30.5
UDAH
Thu Dec 08 11:45:08 ICT 2016
```

Gambar 4.3. K-5 Pada *Systematic Clustering*

Selanjutnya untuk K-5 kita dapat melihatnya pada Gambar 4.3. Untuk K-5 membutuhkan waktu generalisasi selama 8 detik dan *Information Loss* sebesar 30.5.

2. Algoritma *One Pass K-Means*

a. K-3

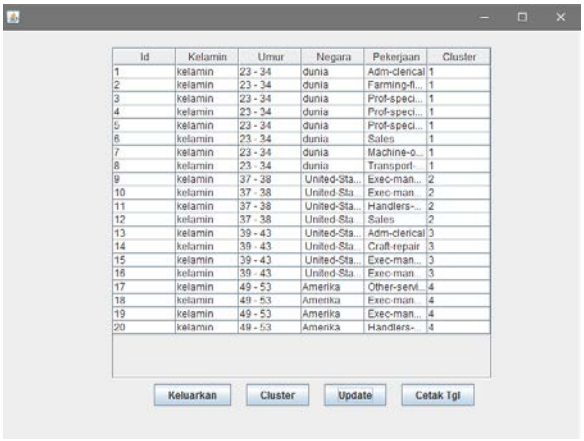



```
selesaiThu Dec 08 11:39:04 ICT 2016
1 dunia  kelamin  23 - 31  11.333333333333332
2 Amerika  Male  32 - 34  13.333333333333332
3  United-States  kelamin  37 - 38  17.4666666666666
5  United-States  kelamin  39 - 43  22.0
6 Amerika  kelamin  49 - 53  28.933333333333334
UDAH
Thu Dec 08 11:39:09 ICT 2016
```

Gambar 4.4. K-3 Pada *One Pass K-Means*

Gambar 4.4 merupakan hasil produk untuk *One Pass K-Means* dengan K-3. Dalam gambar ini kita bisa melihat perbedaan mencolok antara *Systematic Clustering* dan *One Pass K-Means*. Dalam *One Pass K-Means* banyak anggota dalam *cluster*nya cendung bervariasi, berbeda dengan *Systematic Clustering* yang cenderung kaku. Dalam K-3 didapat *Information Loss* sebanyak 28,934 dengan waktu generalisasi 5 detik.

b. K-4



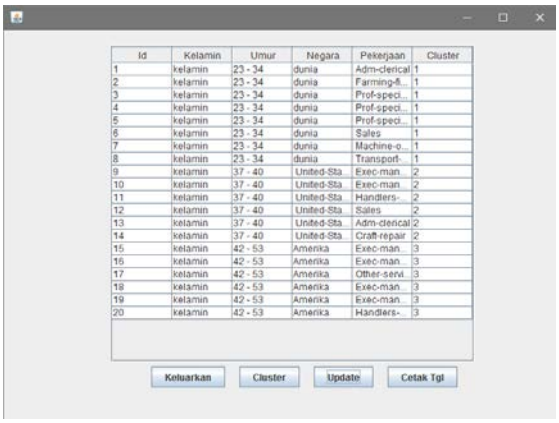

```
selesaiThu Dec 08 11:17:03 ICT 2016
Thu Dec 08 11:17:06 ICT 2016
1 dunia  kelamin  23 - 34  18.933333333333334
2  United-States  kelamin  37 - 38  23.066666666666666
3  United-States  kelamin  39 - 43  27.6
4 Amerika  kelamin  49 - 53  34.53333333333333
UDAH
Thu Dec 08 11:17:10 ICT 2016
```

Gambar 4.5. K-4 Pada *One Pass K-Means*

Gambar 4.5 menunjukan hasil dari model *K-Anonimity* menggunakan *One Pass K-Means* dengan K-4 yang berarti minimal dari anggota *cluster* adalah 4.

Pada algoritma ini kita mendapat *Information Loss* sebanyak 34,53 dan waktu ekesekusi generalisasi selama 7 detik.

c. K-5




```
selesaiThu Dec 08 11:14:52 ICT 2016
1 dunia  kelamin  23 - 34  18.933333333333334
2  United-States  kelamin  37 - 40  25.533333333333335
3 Amerika  kelamin  42 - 53  37.33333333333336
UDAH
Thu Dec 08 11:14:59 ICT 2016
```

Gambar 4.6. K-5 Pada *One Pass K-Means*

Gambar 4.6 menunjukan hasil dari algoritma *One Pass K-Means* dengan K-5 yang berarti minimal dari anggota *cluster* adalah 5. Pada algoritma ini kita mendapat *Information Loss* sebanyak 37,33 dan waktu eksekusi generalisasi selama 7 detik.

4.2 Analisis Hasil Penelitian

Dari algoritma yang sudah dibangun. Tahap selanjutnya adalah menguji algoritma tersebut dengan K yang variatif. K yang akan diuji pada masing-masing algoritma ada K-3 sampai K-11. Setelah melakukan penelitian dengan memberikan K 3-11 pada masing-masing algoritma didapatkan hasil untuk *One Pass K-Means* dengan 10163 data hasilnya sebagai berikut :

Tabel 4.2. Hasil *One Pass K-Means*

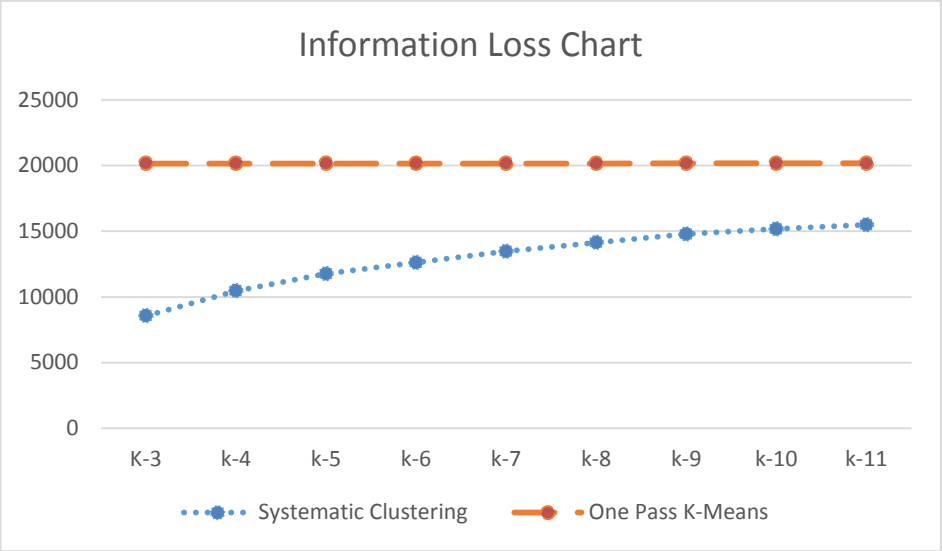
K	Mulai *Cluster*	Selesai *Cluster*	Mulai Proses	Selesai Proses	Total	*Information Loss*
3	10:31:19	10:47	10:51:39	11:08:14	0:32:16	20160.3
4	11:14:36	11:26:48	12:06:48	12:25:37	0:31:01	20160.3
5	13:28:04	13:39:42	13:52:29	14:07:58	0:27:07	20160.3
6	14:13:14	14:25:23	14:27:28	14:43:46	0:28:27	20160.6
7	14:50:17	15:02:27	15:08:15	15:23:06	0:27:01	20159.8
8	13:56:04	14:07:55	14:08:03	14:23:23	0:27:11	20162.7
9	14:31:58	14:41:55	14:42:30	15:04:44	0:32:11	20177.5
10	19:50:36	20:01:47	20:02:03	20:19:23	0:28:31	20176.5
11	20:25:25	20:36:41	20:41:20	20:55:41	0:25:37	20176.4

Sedangkan untuk *Systematic Clustering* hasilnya adalah sebagai berikut

Tabel 4.3. Hasil *Systematic Clustering*

K	Mulai *Cluster*	Selesai *Cluster*	Mulai Proses	Selesai Proses	Total	*Information Loss*
3	20:20:29	20:31:22	20:36:29	20:45:39	0:20:03	8554.20
4	13:31:17	13:42:44	13:53:31	14:02:57	0:20:53	10462.868
5	08:10:18	08:19:48	08:20:03	08:29:18	0:18:45	11767.167
6	08:34:22	08:44:31	08:45:13	08:54:23	0:19:19	12620.131
7	08:57:41	09:06:59	09:07:08	09:17:26	0:19:36	13472.241
8	09:22:03	09:31:28	09:31:55	09:44:41	0:22:11	14141.413
9	09:48:23	09:57:54	09:58:11	10:08:44	0:20:04	14789.539
10	10:12:48	10:22:02	10:22:08	10:33:34	0:20:40	15178.180
11	08:53:56	09:05:18	10:44:41	10:57:23	0:18:04	15490.846

Bila disajikan dalam bentuk grafik, hasil yang didapat untuk perbandingan *Information Loss* antara *Systematic Clustering* dan *One Pass K-Means* akan menghasilkan grafik seperti pada Gambar 4.7



Gambar 4.7. Grafik Perbandingan *Information Loss Systematic Clustering* dan *One Pass K-Means*

Dalam grafik ini kita dapat melihat garis titik-titik pendek (*Systematic Clustering*) berada di bawah garis titik-titik panjang (*One Pass K-Means*). Ini menandakan *Information Loss* dari *Systematic Clustering* lebih sedikit dibanding *One Pass K-Means*. *Systematic Clustering* melakukan *Clustering* dengan lebih baik karena *Systematic Clustering* melakukan *clustering* dengan cara yang teratur dan menjaga sebisa mungkin untuk satu cluster memiliki kedekatan yang besar atau meminimalkan jumlah anggota dalam sebuah *cluster*. Setelah data diurutkan, data dimasukan kedalam *cluster* satu persatu sesuai dengan K yang ditentukan. Untuk menganalisisnya maka diambil contoh 10 data teratas yang ada pada *database*. Data teratas ini adalah contoh yang bisa merepresentasikan seperti apa data yang ada. Untuk melihatnya kita bisa melihat pada Tabel 4.4

Tabel 4.4. Contoh Data Teratas

id_coba	kelamin	umur	negara	Pekerjaan
1	Female	17	United-States	Sales
2	Male	17	United-States	Other-service
3	Male	17	United-States	Other-service
4	Male	17	Mexico	Other-service
5	Male	17	United-States	Other-service
6	Male	17	United-States	Handlers-cleaners
7	Male	17	United-States	Other-service
8	Male	17	United-States	Sales
9	Female	17	United-States	Other-service
10	Male	17	United-States	Farming-fishing

Dengan tabel yang ada diatas. Dengan K = 3, maka hasil yang didapat untuk Model *K-Anonimity* dengan algoritma *One Pass K-Means* adala

Tabel 4.5. *One Pass K-Means* Pada Data Teratas

id_coba	kelamin	umur	negara	Pekerjaan
1	Gender	17	America	Sales
2	Gender	17	America	Other-service
3	Gender	17	America	Other-service
4	Gender	17	America	Other-service
5	Gender	17	America	Other-service
6	Gender	17	America	Handlers-cleaners
7	Gender	17	America	Other-service
8	Gender	17	America	Sales
9	Gender	17	America	Other-service
10	Gender	17	America	Farming-fishing

Sedangkan pada data yang sama dengan data untuk *One Pass K-Means.* Hasil yang didapat untuk model *K-Anonimity* dengan algoritma *Systematic Clustering* untuk K-3 bisa dilihat pada Tabel 4.6. Pada tabel tersebut terlihat perbedaan hasil yang cukup signifikan dengan algoritma *One Pass K-Means*

Tabel 4.6. *Systematic Clustering* Pada Data Teratas

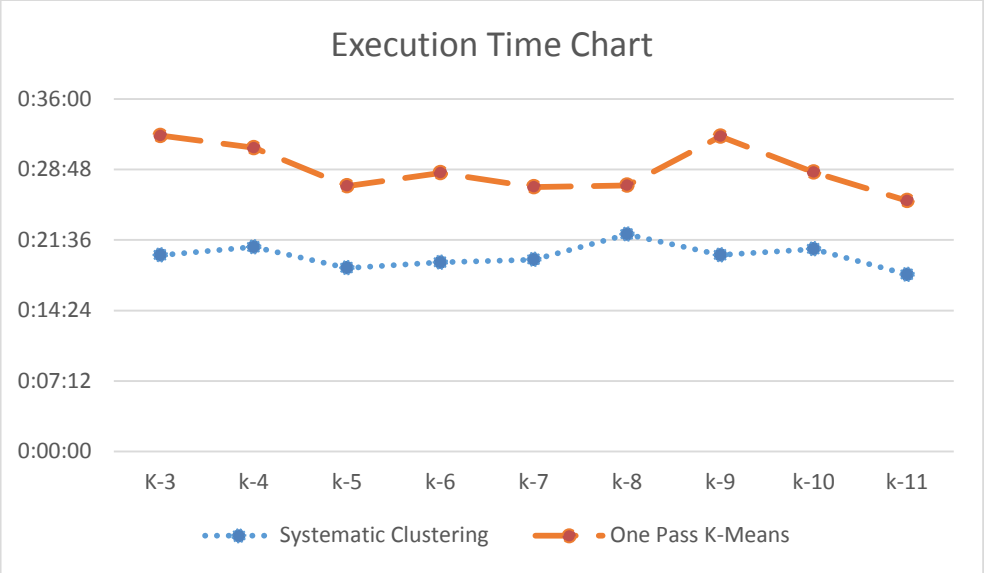
id_coba	kelamin	umur	Negara	Pekerjaan
1	Gender	17	United-States	Sales
2	Gender	17	United-States	Other-service
3	Gender	17	United-States	Other-service
4	Male	17	America	Other-service
5	Male	17	America	Other-service
6	Male	17	America	Handlers-cleaners
7	Gender	17	United-States	Other-service
8	Gender	17	United-States	Sales
9	Gender	17	United-States	Other-service
10	Gender	17	United-States	Farming-fishing

Dari dua tabel tersebut (Tabel 4.6 dan Tabel 4.5) sudah terlihat jelas, bahwa *One Pass K-Means*, menghasilkan generalisasi pada taksonomi tingkat maksimal di atribut kelamin dan negara pada cluster. Ini lah yang menyebabkan *One Pass K-Means* memiliki *Information Loss* yang tinggi yaitu bila dengan rata-rata *Information Loss* sebesar 20166.04. Hal ini terjadi karena *One Pass K-Means* membandingkan tingkat kedekatan *cluster* sesuai umur. Dengan data yang besar (10169) data, pastinya banyak yang memiliki umur yang sama, bila banyak yang memiliki umur yang sama maka umur yang sama itu akan menjadi satu *cluster* tersendiri. Semakin banyak yang memiliki umur yang sama maka akan menjadikan cluster memiliki anggota yang besar. Inilah penyebab utama *One Pass K-Means* memiliki *Information Loss* yang besar dibandingkan dengan *Systematic Clustering* pada 10169 data karena itu pula jumlah *Information Loss* dari *One Pass K-Means* konstan, karena berapapun K nya akan dilakukan *clustering* sesuai umurnya.

Kemudian pada *Systematic Clustering* terlihat bahwa K-9 merupakan titik jenuh. Maksudnya adalah ketika K nya sudah lebih besar dari 9 hasil *Information Loss* nya akan berbeda tipis dari K-9 dengan kata lain perkembangan *Information*

Loss nya melambat. Sedangkan untuk *One Pass K-Means* titik jenuhnya adalah K-3.

Selanjutnya yang akan dibahas adalah dari segi *execution time*. Ini merupakan hal yang penting. Karena dari sini kita juga dapat menyatakan bahwa algoritma tersebut efisien atau tidak.



Gambar 4.8. Grafik Perbandingan *Execution Time Systematic Clustering* dan *One Pass K-Means*

Dengan membaca kembali tentang prosedur yang dilakukan *One Pass K-Means.* Kita dapat menemukan kekurangan pada *One Pass K-Means* yaitu proses membandingkan data dengan seluruh *centroid.* Bila data yang diolah merupakan data besar, maka *centroid* yang akan dibandingkan juga akan semakin banyak. Selain itu untuk melakukan *One Pass K-Means* juga melakukan *query* pada jumlah yang besar yaitu sesuai *cluster*. *Systematic Clustering* yang lebih sederhana akhirnya mendapatkan waktu yang lebih cepat daripada *One Pass K-Means.*

Dari analisis yang didapat. Kelemahan *One Pass K-Means* adalah karena *range* umur yang pendek namun data besar. Oleh karena itu dilakukan pengujian kembali pada data kecil sebesar 100 data. Hal ini dilakukan karena *range* umur yang hanya berjumlah 62 (18-80) maka dilakukan pengujian sebesar dua kali *range*. Hasil yang didapat adalah.

Tabel 4.7. *Systematic Clustering* Pada Data Kecil

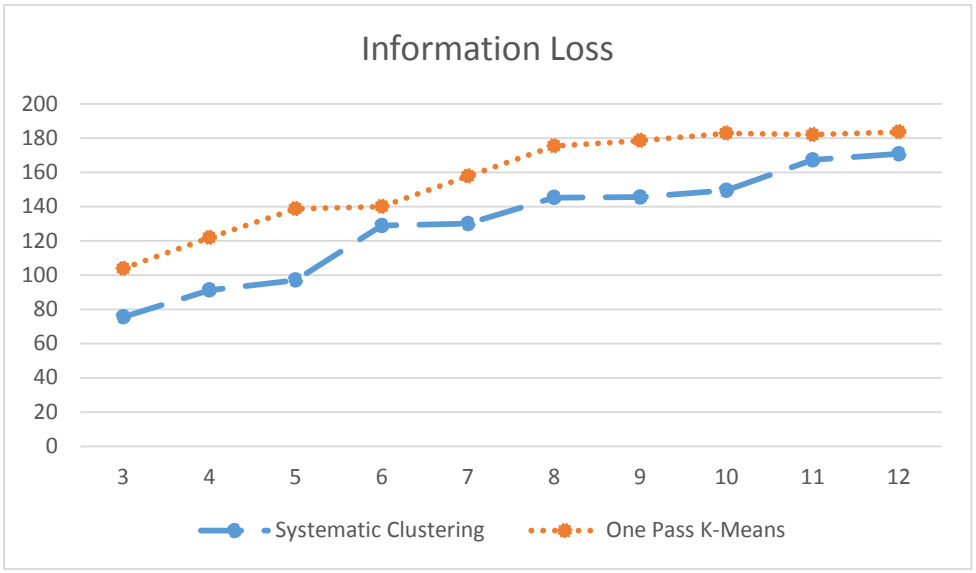
K	Mulai	Selesai	Mulai	Selesai	IL
3	22:31:49	22:31:55	22:31:57	22:32:03	75.49
4	23:01:34	23:01:41	23:01:46	23:01:51	91.27
5	23:03:29	23:03:35	23:03:46	23:03:51	97.00
6	23:05:23	23:05:30	23:05:31	23:05:37	128.91
7	23:09:13	23:09:19	23:09:21	23:09:27	130.01
8	23:13:53	23:14:01	23:14:02	23:14:08	145.31
9	23:16:41	23:16:48	23:16:49	23:16:54	145.54
10	23:18:46	23:18:53	23:19:07	23:19:14	149.50
11	23:20:40	23:20:47	23:20:49	23:20:55	167.24
12	23:22:41	23:22:48	23:22:50	23:22:56	170.78

Sedangkan Untuk *One Pass K-Means* hasil yang didapat adalah

Tabel 4.7. *One Pass K-Means* Pada Data Kecil

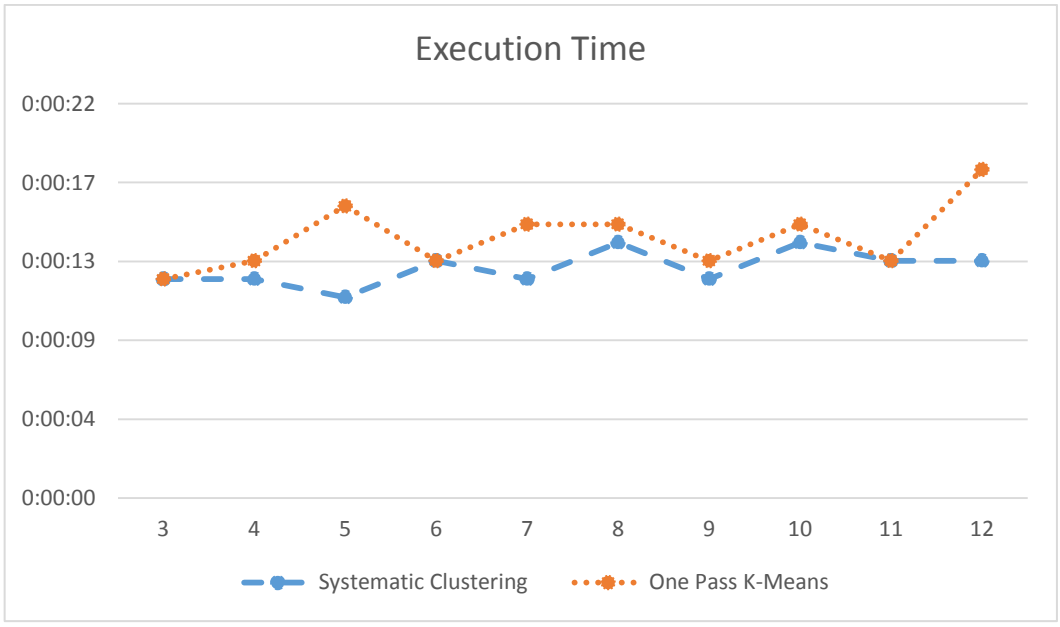
K	Mulai	Selesai	Mulai	Selesai	IL
3	22:22:24	22:22:30	22:22:32	22:22:38	103.82
4	23:25:02	23:25:09	23:25:11	23:25:17	121.90
5	23:31:42	23:31:52	23:31:53	23:31:59	138.68
6	23:34:03	23:34:09	23:34:23	23:34:28	137.56
7	23:36:47	23:36:54	23:36:56	23:37:03	157.80
8	23:40:08	23:40:17	23:40:19	23:40:24	175.36
9	23:44:46	23:44:52	23:44:59	23:45:05	178.55
10	23:49:28	23:49:35	23:49:37	23:49:42	182.84
11	23:56:39	23:56:46	23:56:48	23:56:53	168.35
12	00:01:10	00:01:16	00:01:18	00:01:30	183.65

Hasil ini kemudian dibuat grafiknya seperti yang dibuat pada Gambar 4.7 dan 4.8. Hasil dari grafik ini bisa dilihat pada Gambar 4.9 dan Gambar 4.10



Gambar 4.9. Grafik Perbandingan *Information Loss Time Systematic Clustering* dan *One Pass K-Means* Pada 100 Data

Dari gambar ini kita melihat bahwa *Systematic Clustering* memiliki *Information Loss* yang lebih kecil dari *One Pass K-Means* namun kali ini *Gap* yang dihasilkan tidak sejauh Gambar 4.7.



Gambar 4.10. Grafik Perbandingan *Execution Time Loss Time Systematic Clustering* dan *One Pass K-Means* Pada 100 Data

Hal menarik juga ditemukan pada waktu ekeskusi. Pada data kecil, *One Pass K-Means* dan *Systematic Clustering* mendapatkan waktu eksekusi yang berbeda tipis. Di titik K 4,8,9,10 *One Pass K-Means dan Systematic Clustering* hanya berbeda 1 detik.

Selanjutnya, dalam penelitian algoritma. Perlu adanya analisis *Big-O*. Untuk melakukannya, pertama yang kita hitung adalah proses dalam *clustering*.

```
    for(int
i=0;i<jumlahbaris;i++){
    if((i+1)%k==0){
    for(int  j=i;j>=(i-(k-
1));j--){
                }
            }
    }
```

Kita abaikan operasi yang ada didalam *looping* karena itu tidak akan dihitung dalam *Big-O*. Jumlah baris adalah jumlah *record.* Kita beri jumlah baris nilai

```
    for (int i = 0; i <= jumlahbari; i++) {
    if                                          (new
FormUpdaterScnu().cariBanyakDataDalamCluster(i) > 0) {
    for (int a = new FormUpdaterScnu().selectMin(i); a <
new FormUpdaterScnu().selectMax(i); a++) {}
    for (int a = new FormUpdaterScnu().selectMin(i); a <
new FormUpdaterScnu().selectMax(i); a++) {}
    for (int a = new FormUpdaterScnu().selectMin(i); a <
new FormUpdaterScnu().selectMax(i); a++) {}
    for (int a = new FormUpdaterScnu().selectMin(i); a <
new FormUpdaterScnu().selectMax(i); a++) {}
```

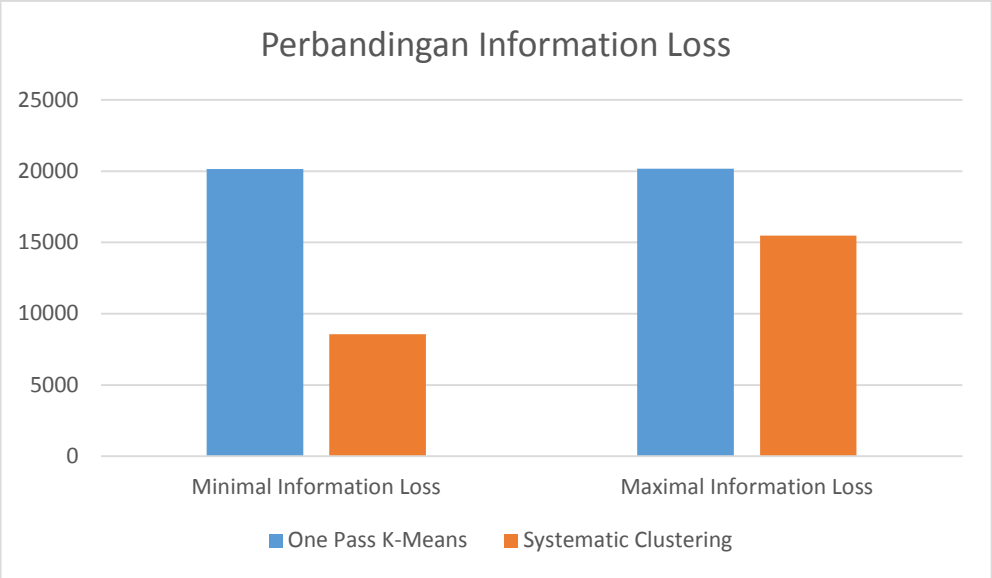
For yang pertama bertujuan untuk *looping* sebanyak cluster yang ada. Oleh karena itu kita beri nilai *for* pertama sebagai

$$T_1(n) + T_2(n) \in O(f(n)) + O(g(n)) = O(\max(f(n), g(n))$$

$$O(cf(n)) = O(f(n)), c \text{ adalah konstanta}$$

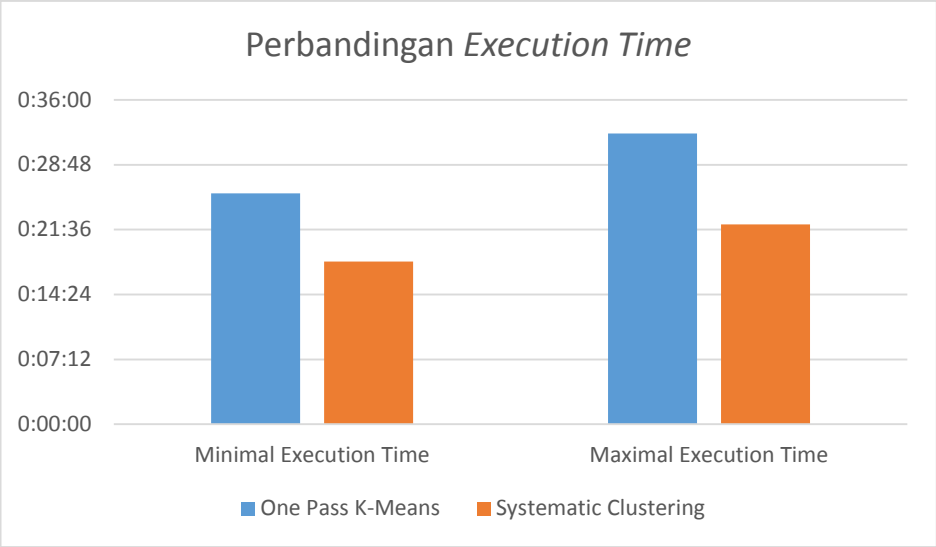
```
    for (int i = 0; i < jumlahbaris;
i++) {
                for (int x = 0; x <
jumlahbaris; x += k) {
     }
    }
```

Sekilas kita dapat melihat bahwa *looping* dilakukan sebanyak dua kali sebanyak variabel yang sama. Oleh karena itu bila jumlah baris adalah



Gambar 4.11. Grafik Perbandingan Nilai Maksimal dan Minimal *Information Loss* Pada Algoritma *Systematic Clustering* dan *One Pass K-Means*

Pada algoritma *Systematic Clustering* memiliki titik jenuh pada K-9, yang berarti pada K-9 pertumbuhan *Information Loss* mulai rendah. *Big-O* yang dihasilkan pada algoritma ini adalah



Gambar 4.12. Grafik Perbandingan Nilai Maksimal dan Minimal *Information Loss Systematic Clustering* dan *One Pass K-Means*

Secara persentase hasil *Information Loss* minimal dari *One Pass K-Means* 135.67% lebih besar dari algoritma *Systematic Clustering* sedangkan untuk *Information Loss* maksimal dari *One Pass K-Means* 30.2% lebih besar daripada algoritma *Systematic Clustering.* Yang menandakan algoritma *Systematic Clustering* lebih baik daripada *One Pass K-Means* dari segi *Information Loss.*

4.4 Aplikasi Hasil Penelitian

Banyak sekali kegunaan dari model *K-Anonimity* dengan *One Pass K-Means* dan *Systematic Clustering.* Untuk pengaplikasian pada kehidupan sehari-hari dapat dilakukan pada :

1. **Rumah Sakit**

Data pasien yang harus dipublikasikan ke khalayak umum merupakan data yang berisi data sensitif. Ini akan menimbulkan perasaan tidak enak kepada pasien yang data pribadinya dapat tersebar. Walaupun sudah menghapus identitas pasien,

data masih dapat diketahui dengan menggabungkan dua tabel untuk dikaitkan. Dengan menggunakan model *K-Anonimity* kita dapat meminimalkan peluang adanya menggabungkan dua tabel untuk dikaitkan. Dengan begitu identitas dari pemilik data tidak akan mudah untuk diketahui.

2. Kepolisian

Dengan adanya model ini, kepolisian dapat meminimalkan kebocoran informasi yang ada tentang para mantan tersangka atau narapidana. Ini sangat berguna dalam menjaga kerahasiaan data bila akan dilakukan mining data atau publikasi. Selain untuk *publishing* ada pula *privacy preserving data mining.*

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat diambil kesimpulan sebagai berikut:

1. Algoritma *Systematic Clustering* memiliki *information lo*ss yang lebih rendah.
2. Algoritma *Systematic Clustering* memiliki waktu eksekusi yang lebih cepat.
3. Algoritma *Systematic Clustering* memiliki titik jenuh pada K-9 sedangkan *One Pass K-Means* pada K-3.

5.2. Saran

Penelitian ini perlu dilakukan pada data yang lebih variatif lagi. Misal mengui pada 1000,2000,3000,4000 data. Hal ini untuk memperkuat hasil penelitian

DAFTAR PUSTAKA

A. Garner, Bryan. (1999). *Black's Law Dictionary* seventh Edition. St. Paul Minn, New York.

Ackoff, Russell L.(1962). *Scientific Method Optimizing Applied Research. Decisions*. New York and London: John Wiley & Sons, Inc.

Aggarwal, Charu C. (2015). *Data Mining The Text Book,* IBM T.J. Watson Research Center Yorktown Heights, New York. USA

Altman, Irwin. (1975). *The Environment and Social Behaviour : Privacy, Personal, Space, Teritory and Crowding, Monterey, Brooks/ Cole,* California.

Arikunto S. (2006). *Prosedur Penelitian Suatu Pendekatan Praktik,* Ed Revisi VI,. Penerbit PT Rineka Cipta, Jakarta.

B. C. M. Fung, K. Wang, dan P. S, Yu. (2005). Top-down specialization for information and privacy preservation. *International Conference on Data Engineering*. 21:205–216

B, Pawan R dan Jinwala, Devesh C. (2016). Novel Approaches for Privacy Preserving Data Mining in k-Anonymity Model. *Journal of Information Science and Engineering*. 32:63-78.

Cai, Li-xia. (2010). Additive and multiplicative perturbation bounds for the Moore–Penrose inverse. *Linear Algebra and its Applications.* 434 :480-489

Cormen, Thomas H. (2009). *Introduction to Algorithms,*. London. The MIT Press.

Dalenius, T.(1986). Finding a needle in a haystack – or identifying anonymous census record. *Journal of Official Statistics*, 2:329-336.

Fung, Benjamnin. (2010) Privacy-Preserving Data Publishing: A Survey of Recent Developments. *ACM Computing Surveys*, 42:4-14.

Fung,Benjamin.,dkk. (2011). *Privacy-Preserving Data Publishing Concepts and Techniques.* Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. United States of America

Enamul Kabir, Md. (2011). Efficient Systematic Clustering Method For K-Anonymization. *Acta Informatica*. Vol. 10. pp 51–66

Gordon, Sheldon P. (1994). *Contempo-rary Statistics: A Computer Approach.* New York: McGraw-Hill, Inc.

Inmon, William H. (2005). *Building The Data Warehouse*. Indianapolis :Wiley Publishing, Inc.

Irmansyah, Faried. (2003). *Pengantar Database*, http://www.ilmukomputer.com/. Diakses pada 20 September 2016.

Iskandar. (2008). *Metodologi Penelitian Pendidikan dan Sosial (Kuantitaif dan Kualitatif)*. Jakarta: Gaung Persada Group

Ji-Won Byun., dkk. (2007). *Efficient K-Anonymization Using Clustering Techniques.* Springer-Verlag Berlin Heidelberg.

Kamus Besar Bahasa Indonesia. http://kbbi.web.id/ (diakses tanggal 28 Agustus 2016).

Lin, Jun-Lin dan Wei, Meng-Cheng.(2008). *An Efficient Clustering Method for k-Anonymization. Department of Information Management Yuan Ze University Chung-Li,* Taiwan dan Department of Information Management Yuan Ze University Chung-Li, Taiwan

Munir, Rinaldi. (2011). *Algoritma dan Pemrograman dalam Bahasa Pascal dan C*. Informatika, Bandung

Simamarta, D.A. (1983). *Operation Research – Sebuah Pengantar,* Jakarta. PT. Gramedia Pustaka Utama.

Sweeney, L. (1998). *Generalizing Data to Provide Anonymity when Disclosing Information*. Massachusetts Institute of Technology, United States of America

Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10:571 - 588.

Tsai, C.-Y. dan Chiu A .(2007). k-anonymity clustering method for effective data privacy preservation. Pada *Third International Conference on Advanced Data Mining and Applications (ADMA)*

LAMPIRAN 1
Dataset Adult

39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
37, Private, 284582, Masters, 14, Married-civ-spouse, Exec-managerial, Wife, White, Female, 0, 0, 40, United-States, <=50K
49, Private, 160187, 9th, 5, Married-spouse-absent, Other-service, Not-in-family, Black, Female, 0, 0, 16, Jamaica, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 45, United-States, >50K

LAMPIRAN 2

Source Code

Ini merupakan *Source Code* dari aplikasi yang digunakan sebagai pendukung penelitian.

```
/*
 * To change this license header, choose License Headers in Project Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package skripsialpha;


import javax.swing.JOptionPane;
import javax.swing.table.DefaultTableModel;



/**
 *
 * @author rezar_4
 */
public class FormScNew extends javax.swing.JFrame {


    /**
     * Creates new form Form
     */
    public FormScNew() {
        initComponents();
```

```
}

private void retrieve() {
    DefaultTableModel dm = new FormUpdaterSc().getData();
    jTable1.setModel(dm);
}


/**
 * This method is called from within the constructor to initialize the form.
 * WARNING: Do NOT modify this code. The content of this method is always
 * regenerated by the Form Editor.
 */
@SuppressWarnings("unchecked")
// <editor-fold defaultstate="collapsed" desc="Generated Code">
private void initComponents() {

    jScrollPane1 = new javax.swing.JScrollPane();
    jTable1 = new javax.swing.JTable();
    jButton1 = new javax.swing.JButton();
    jButton2 = new javax.swing.JButton();
    update = new javax.swing.JButton();

setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
```

```
jTable1.setModel(new javax.swing.table.DefaultTableModel(
    new Object [][] {
        {null, null, null, null, null},
        {null, null, null, null, null},
        {null, null, null, null, null},
        {null, null, null, null, null}
    },
    new String [] {
        "Id", "Kelamin", "Umur", "Negara", "Penyakit"
    }
));
jScrollPane1.setViewportView(jTable1);

jButton1.setText("Keluarkan");
jButton1.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        jButton1ActionPerformed(evt);
    }
});

jButton2.setText("Proses");
jButton2.addMouseListener(new java.awt.event.MouseAdapter() {
    public void mouseClicked(java.awt.event.MouseEvent evt) {
        jButton2MouseClicked(evt);
    }
});
```

```
        update.setText("Update");
        update.addMouseListener(new java.awt.event.MouseAdapter() {
            public void mouseClicked(java.awt.event.MouseEvent evt) {
                updateMouseClicked(evt);
            }
        });
        update.addActionListener(new java.awt.event.ActionListener() {
            public void actionPerformed(java.awt.event.ActionEvent evt) {
                updateActionPerformed(evt);
            }
        });

        javax.swing.GroupLayout layout = new
javax.swing.GroupLayout(getContentPane());
        getContentPane().setLayout(layout);
        layout.setHorizontalGroup(

layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
            .addGroup(layout.createSequentialGroup()
                .addContainerGap(134, Short.MAX_VALUE)

.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEA
DING)
                    .addGroup(javax.swing.GroupLayout.Alignment.TRAILING,
layout.createSequentialGroup()
                        .addComponent(jScrollPane1,
javax.swing.GroupLayout.PREFERRED_SIZE,
```

```
javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.PREFERRED_SIZE)

                .addGap(105, 105, 105))

            .addGroup(javax.swing.GroupLayout.Alignment.TRAILING,
layout.createSequentialGroup()

                .addComponent(jButton1)

                .addGap(18, 18, 18)

                .addComponent(jButton2)

                .addGap(18, 18, 18)

                .addComponent(update)

                .addGap(221, 221, 221))))

    );

    layout.setVerticalGroup(


layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)

        .addGroup(layout.createSequentialGroup()

            .addGap(20, 20, 20)

            .addComponent(jScrollPane1,
javax.swing.GroupLayout.PREFERRED_SIZE, 304,
javax.swing.GroupLayout.PREFERRED_SIZE)

            .addGap(18, 18, 18)


.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.BAS
ELINE)

                .addComponent(jButton1)

                .addComponent(jButton2)

                .addComponent(update))

            .addContainerGap(120, Short.MAX_VALUE))

    );
```

```
        pack();
    }// </editor-fold>

    private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
        // TODO add your handling code here:
        retrieve();
    }

    private void jButton2MouseClicked(java.awt.event.MouseEvent evt) {
        // TODO add your handling code here:
        int jumlahbaris = jTable1.getRowCount();
        int point=1;
        double informationloss=0;
        int pembagi = new FormUpdaterScNew().selectMaxUmur()-new
FormUpdaterScNew().selectMinUmur();

        for(int i=0;i<jumlahbaris;i++){
            if((i+1) % 3==0 && i!=0){
                int id = Integer.parseInt(jTable1.getValueAt(i, 0).toString());
                String kelamin = "kelamin";
                int pbsatu;
                int pbdua;
                double ilneg;
                String umur;
                String negara;
```

```
String negara1;

String negara2;

String negfix;


int panjang_kode=jTable1.getValueAt(i, 3).toString().length();

   if(i<=3){

      umur = jTable1.getValueAt(i-2, 2).toString() + " - " +
jTable1.getValueAt(i, 2).toString();

   }else {

      umur = jTable1.getValueAt(i-3, 2).toString() + " - " +
jTable1.getValueAt(i, 2).toString();

   }

pbsatu = Integer.parseInt(jTable1.getValueAt(i, 2).toString())-
Integer.parseInt(jTable1.getValueAt(i-2, 2).toString());

pbdua = Integer.parseInt(jTable1.getValueAt(i-2, 2).toString());



negara = jTable1.getValueAt(i, 3).toString();

negara1 = jTable1.getValueAt(i-1, 3).toString();

negara2 = jTable1.getValueAt(i-2, 3).toString();


if(negara.equals(negara1) && negara.equals(negara2) &&
negara1.equals(negara2)){

   negfix=negara;

   ilneg = 0;

}else {

   if(" United-States".equals(negara) || " Cuba".equals(negara) || "
Mexico".equals(negara) || " Puerto-Rico".equals(negara)){
```

```
            negara="america";
        }else if(" India".equals(negara) || " South".equals(negara)){
            negara="asia";
        }

        if(" United-States".equals(negara1) || " Cuba".equals(negara1) || "
Mexico".equals(negara1) || " Puerto-Rico".equals(negara1)){
            negara1="america";
        }else if(" India".equals(negara1) || " South".equals(negara1)){
            negara1="asia";
        }

        if(" United-States".equals(negara2) || " Cuba".equals(negara2) || "
Mexico".equals(negara2) || " Puerto-Rico".equals(negara2)){
            negara2="america";
        }else if(" India".equals(negara2) || " South".equals(negara2)){
            negara2="asia";
        }

        if(negara==negara1 && negara==negara2 &&
negara1==negara2){
            negfix=negara;
            ilneg = 0.6;
        }else{
            negfix="dunia";
            ilneg = 1;
         }
```

```
            }


              String penyakit = jTable1.getValueAt(i, 4).toString();
              informationloss = informationloss + (3*(pbsatu/pembagi + 1 + ilneg));




              System.out.print(" " +id+" ");
              System.out.print(" "+kelamin+" ");
              System.out.print(" "+ negfix + " ");
              System.out.print(" "+ umur + " ");
              System.out.print(" "+penyakit);
              System.out.print(" "+ informationloss + " ");
              System.out.println();
            }



            //new BusUpdater().update(jTextField1.getText(), jTextField2.getText(), t,
j)
            //new FormUpdater().update(id, kelamin, umur, kode_pos, penyakit);
        }
    }

    private void updateActionPerformed(java.awt.event.ActionEvent evt) {
```

```
        // TODO add your handling code here:
    }

    private void updateMouseClicked(java.awt.event.MouseEvent evt) {
        // TODO add your handling code here:
        int jumlahBaris = jTable1.getRowCount();
        System.out.println(jumlahBaris);

        for(int u=1;u<=jumlahBaris;u++){
            if(jTable1.getValueAt(u-1, 3).toString().length()<6){
                new FormUpdater().updatekode(u,jTable1.getValueAt(u-1, 3).toString()
+ "0" );
            }
        }
        System.out.println(jTable1.getValueAt(0, 3).toString().length());

    }

    /**
     * @param args the command line arguments
     */
    public static void main(String args[]) {
        /* Set the Nimbus look and feel */
        //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code
(optional) ">
        /* If Nimbus (introduced in Java SE 6) is not available, stay with the default
look and feel.
```

```
        * For details see
http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html
        */
       try {
          for (javax.swing.UIManager.LookAndFeelInfo info :
javax.swing.UIManager.getInstalledLookAndFeels()) {
             if ("Nimbus".equals(info.getName())) {
                javax.swing.UIManager.setLookAndFeel(info.getClassName());
                break;
             }
          }
       } catch (ClassNotFoundException ex) {
java.util.logging.Logger.getLogger(Form.class.getName()).log(java.util.logging.L
evel.SEVERE, null, ex);
       } catch (InstantiationException ex) {
java.util.logging.Logger.getLogger(Form.class.getName()).log(java.util.logging.L
evel.SEVERE, null, ex);
       } catch (IllegalAccessException ex) {
java.util.logging.Logger.getLogger(Form.class.getName()).log(java.util.logging.L
evel.SEVERE, null, ex);
       } catch (javax.swing.UnsupportedLookAndFeelException ex) {
java.util.logging.Logger.getLogger(Form.class.getName()).log(java.util.logging.L
evel.SEVERE, null, ex);
       }
       //</editor-fold>
```

```
        /* Create and display the form */
        java.awt.EventQueue.invokeLater(new Runnable() {
            public void run() {
                new Form().setVisible(true);
            }
        });
    }

    // Variables declaration - do not modify
    private javax.swing.JButton jButton1;
    private javax.swing.JButton jButton2;
    private javax.swing.JScrollPane jScrollPane1;
    private javax.swing.JTable jTable1;
    private javax.swing.JButton update;
    // End of variables declaration

}

/*
 * To change this license header, choose License Headers in Project Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */
package skripsialpha;

import java.sql.Connection;
```

```
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.sql.Statement;
import javax.swing.table.DefaultTableModel;

/**
 *
 * @author rezar_4
 */
public class FormUpdaterScNew {

    String consString = "jdbc:mysql://localhost:3306/fake";
    String username = "root";
    String password = "";

    public Boolean add(String kelamin, int umur, String kode_pos, String
jenis_kelamin) {
        String sql = "INSERT INTO coba(kelamin,umur,kode_pos,jenis_kelamin)
VALUES ('" + kelamin + "','" + umur + "','" + kode_pos + "','" + jenis_kelamin
+ "')";
        try {
            Connection con = DriverManager.getConnection(consString, username,
password);
            Statement s = con.prepareStatement(sql);
            s.execute(sql);
            return true;
```

```
        } catch (Exception ex) {
            ex.printStackTrace();
        }
        return false;
    }

    public DefaultTableModel getData() {
        //add column
        DefaultTableModel dm = new DefaultTableModel();
        dm.addColumn("Id");
        dm.addColumn("Kelamin");
        dm.addColumn("Umur");
        dm.addColumn("Negara");
        dm.addColumn("Pekerjaan");

        String sql = "select id_coba,kelamin,umur, negara,pekerjaan from sc;";
        try {
            Connection con = DriverManager.getConnection(consString, username,
password);
            Statement s = con.prepareStatement(sql);
            ResultSet rs = s.executeQuery(sql);
            //looping
            while (rs.next()) {
                String id_coba = rs.getString(1);
                String kelamin = rs.getString(2);
                String umur = rs.getString(3);
```

```
            String negara = rs.getString(4);
            String pekerjaan = rs.getString(5);
            dm.addRow(new String[]{id_coba, kelamin, umur, negara, pekerjaan});
        }
        return dm;
    } catch (Exception ex) {
        ex.printStackTrace();

    }
    return null;
  }

  public Boolean update(int id_coba,String kelamin, int umur, String negara,
String penyakit) {
    String sql = "UPDATE sc SET kelamin='" + kelamin + "',umur='" + umur +
"',negara='" + negara + "',penyakit='" + penyakit + "' WHERE id_coba='" +
id_coba + "'";
    try {
        Connection con = DriverManager.getConnection(consString, username,
password);
        Statement s = con.prepareStatement(sql);
        s.execute(sql);
        return true;
    } catch (SQLException ex) {
        ex.printStackTrace();
        return false;
    }
```

```
    }


    public Boolean updatekode(int id_coba, String kode_pos) {
        String sql = "UPDATE coba SET kode_pos='" + kode_pos + "' WHERE
id_coba='" + id_coba + "'";
        try {
            Connection con = DriverManager.getConnection(consString, username,
password);
            Statement s = con.prepareStatement(sql);
            s.execute(sql);
            return true;
        } catch (SQLException ex) {
            ex.printStackTrace();
            return false;
        }
    }
    public int selectMinUmur() {
        String sql = "select min(umur) from sc;";
        int id_coba = 0;
        try {
            Connection con = DriverManager.getConnection(consString, username,
password);
            Statement s = con.prepareStatement(sql);
            ResultSet rs = s.executeQuery(sql);
//            looping
            while (rs.next()) {
                id_coba = rs.getInt(1);
```

```
            }
            s.close();
            con.close();
            return id_coba;
        } catch (Exception ex) {
            ex.printStackTrace();
            System.err.println("Got an exception! ");
            return 0;
        }
    }
    public int selectMaxUmur() {
        String sql = "select max(umur) from sc;";
        int id_coba = 0;
        try {
            Connection con = DriverManager.getConnection(consString, username,
password);
            Statement s = con.prepareStatement(sql);
            ResultSet rs = s.executeQuery(sql);
//            looping
            while (rs.next()) {
                id_coba = rs.getInt(1);
            }
            s.close();
            con.close();
            return id_coba;
        } catch (Exception ex) {
```

```
            ex.printStackTrace();
            System.err.println("Got an exception! ");
            return 0;
        }
    }
}
```

DAFTAR RIWAYAT HIDUP

Penulis bernama Reza Ridwansyah, lahir di Lampung 10 Juni 1995. Anak tunggal dari pasangan Idrus Suwito dan Elia Roseli. Riwayat pendidikan formal yang pernah ditempuh oleh peneliti, Pendidikan Dasar di SDIT Al-Fidaa, Bekasi (2001-2006), Pendidikan Menengah di SMPN 3 Tambun Selatan, Jakarta (2007-2009), Pendidikan Tingkat Atas di SMAN 3 Tambun Selatan, Bekasi (2010-2012) dan melanjutkan ke jenjang universitas di Universitas Negeri Jakarta, Jakarta Timur (2013-2017), Fakultas Teknik, Program Studi Pendidikan Teknik Informatika dan Komputer dengan Konsentrasi Peminatan Rekayasa Perangkat Lunak pada tahun 2015 dan lulus pada tahun 2017.

Kegiatan yang telah diikuti selama kuliah di Universitas Negeri Jakarta adalah Praktek Kerja Lapangan di PUSTIKOM UNJ tahun 2016 dan Kuliah Kerja Nyata di Desa Banggala Mulya pada tahun 2016. Kegiatan selama kuliah di Universitas Negeri Jakarta antara lain Program Praktek Kerja Lapangan (PKM) di SMKN 1 Bekasi (2016).