

SKRIPSI

ANALISIS ALGORITMA *DATAFLY* DAN *GREEDY*
K-MEMBER CLUSTERING DALAM MENJAGA PRIVASI
DATA



Edrick Ezra Zacharia

NPM: 2015730036

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2019

UNDERGRADUATE THESIS

**DATAFLY AND GREEDY K-MEMBER CLUSTERING
ALGORITHM ANALYSIS TO PRESERVE DATA PRIVACY**



Edrick Ezra Zacharia

NPM: 2015730036

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2019**

LEMBAR PENGESAHAN

**ANALISIS ALGORITMA *DATAFLY* DAN *GREEDY*
K-MEMBER CLUSTERING DALAM MENJAGA PRIVASI
DATA**

Edrick Ezra Zacharia

NPM: 2015730036

Bandung, 16 Mei 2019

Menyetujui,

Pembimbing

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

Anggota Tim Penguji

Husnul Hakim, M.T.

Luciana Abednego, M.T.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS ALGORITMA *DATAFLY* DAN *GREEDY K-MEMBER CLUSTERING* DALAM MENJAGA PRIVASI DATA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 16 Mei 2019

Meterai Rp. 6000

Edrick Ezra Zacharia
NPM: 2015730036

ABSTRAK

Privasi pada dasarnya adalah kemampuan seseorang dalam mengatur informasi pribadinya. Informasi-informasi pribadi bisa disebut juga dengan istilah *personally identifiable information* atau PII. Pada zaman ini, *data mining* merupakan suatu aktivitas yang menguntungkan, tetapi dengan melakukan *data mining* juga dapat membuka informasi pribadi. Tujuan dari penelitian ini adalah dapat merilis sebuah data agar dapat digunakan untuk *data mining*, tetapi tidak melanggar privasi, dengan menggunakan teknik anonimisasi seperti algoritma *k-anonymity*. Algoritma *k-anonymity* yang akan digunakan pada penelitian ini adalah algoritma *Datafly* dan algoritma *Greedy k-member Clustering*.

Perangkat lunak dibangun untuk mengimplementasikan algoritma *k-anonymity*, yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Pengujian juga dilakukan dalam rangka mendapatkan relasi antara kualitas, nilai k , tipe dan banyaknya atribut yang digunakan, dan waktu eksekusi algoritma *k-anonymity* yang diimplementasikan.

Berdasarkan hasil pengujian, algoritma *Datafly* dan algoritma *Greedy k-member Clustering* menunjukkan hasil yang berbeda. Algoritma *Datafly* memiliki kualitas dan waktu eksekusi yang rendah bila nilai k semakin tinggi, tetapi jumlah entri pada hasil anonimisasi menjadi lebih sedikit dibanding dengan jumlah entri pada tabel privat. Di sisi lain, algoritma *Greedy k-member Clustering* memiliki kualitas dan waktu eksekusi yang tinggi bila nilai k semakin tinggi, tetapi entri pada hasil anonimisasi sama dengan jumlah entri pada tabel privat. Pada akhirnya, kualitas dapat diatur dengan menyesuaikan nilai k dengan algoritma yang digunakan atau menyesuaikan jumlah atribut yang digunakan. Nilai k dan banyaknya atribut yang digunakan juga dapat berpengaruh pada waktu eksekusi algoritma *k-anonymity* yang diimplementasikan.

Kata-kata kunci: Privasi, *Privacy-Preserving Data Mining*, *k-Anonymity*, *Datafly*, *Greedy k-member Clustering*

ABSTRACT

Privacy is basically a person's ability to manage his personal information. Personal information can also be referred to as personally identifiable information or PII. In this era, data mining is a profitable activity, but by doing data mining it can also disclose personal information. The purpose of this study is to be able to release a data so that it can be used to data mining, but does not violate privacy, using anonymization techniques such as the k-anonymity algorithm. The k-anonymity algorithm that will be used in this study is the Datafly algorithm and Greedy k-member Clustering algorithm.

A software was built to implement k-anonymity algorithms, that are Datafly algorithm and Greedy k-member Clustering algorithm. Tests are also done in order to get relations between quality, k value, type and number of attributes used, and execution time of implemented k-anonymity algorithms.

Based on test results, Datafly algorithm and Greedy k-member Clustering algorithm show different results. The Datafly algorithm has low quality and execution time the more higher k value is used, but the number of entries in anonymization result is less than the number of entries in private table. The Greedy k-member Clustering has high quality and execution time the more higher k value is used, but the number of entries in anonymization result is equals to the number of entries in private table. In the end, quality can be adjusted by changing k value with the algorithm used or changing the number of attributes used. k value and number of attributes used can also affect the execution time of implemented k-anonymity algorithms.

Keywords: Privacy, Privacy-Preserving Data Mining, k-Anonymity, Datafly, Greedy k-member Clustering

Dipersembahkan untuk Tuhan YME, keluarga tercinta, Ibu Mariskha sebagai dosen pembimbing, teman-teman yang telah berperan dalam pembuatan skripsi ini, dan diri sendiri

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas karunia-Nya, penulis dapat menyelesaikan penyusunan penelitian yang berjudul "Analisis Algoritma *Datafly* dan *Greedy k-member Clustering* dalam Menjaga Privasi Data". Penulisan penelitian ini diajukan untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika Universitas Katolik Parahyangan. Pada penyusunan dan penulisan penelitian ini, penulis menyadari bahwa penyusunan dan penulisan penelitian ini juga tidak terlepas dari bantuan berbagai pihak, baik langsung maupun tidak langsung. Secara khusus, penulis ingin berterima kasih kepada:

1. Keluarga yang selalu memberikan dukungan kepada penulis baik berupa doa maupun bimbingan moral agar dapat menyelesaikan penyusunan penelitian ini.
2. Ibu Mariskha Tri Adithia, S.Si., M.Sc., PDEng. selaku dosen pembimbing yang telah membimbing penulis dan memberikan dukungan dan bantuan kepada penulis dalam proses penyusunan penelitian ini.
3. Bapak Husnul Hakim, S.Kom., M.T., dan Ibu Luciana Abednego, S.Kom., M.T. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga penelitian ini menjadi lebih baik.
4. Sahabat baik penulis dari SMA Santa Maria 3 Cimahi, yaitu Vita, Irene, Friska, Bram, Anes, Aye, Geo, Sasha, Windy, dan Janice, yang selalu memberikan dukungan dan menyediakan tempat untuk *refreshing* ketika penulis jenuh mengerjakan penelitian ini.
5. Teman-teman sesama pejuang penelitian, khususnya kepada Matthew, Himawan, Irvan, Thoby, Felicia, dan Yudhis, serta teman-teman dari Jurusan Teknik Informatika UNPAR khususnya angkatan 2015 yang tidak dapat disebutkan satu persatu, yang telah memberi semangat kepada penulis dalam penyusunan penelitian ini.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf jika terdapat kesalahan. Penulis juga mengharapkan kritik dan saran yang membangun untuk menyempurnakan penelitian ini. Semoga penelitian ini dapat memberi informasi yang bermanfaat dan menjadi inspirasi untuk penelitian-penelitian berikutnya.

Bandung, Mei 2019

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	3
1.6 Sistematika Pembahasan	3
2 DASAR TEORI	5
2.1 Privasi	5
2.2 <i>Data Mining</i>	6
2.3 <i>k-Anonymity</i>	7
2.3.1 Algoritma <i>Datafly</i>	8
2.3.2 Algoritma <i>Greedy k-member Clustering</i>	9
2.4 Penyimpanan dan Pembacaan Data	12
2.4.1 JSON	12
2.4.2 CSV	13
2.4.3 RegEx	13
3 ANALISIS	15
3.1 Analisis Masalah	15
3.1.1 <i>Personally Identifiable Information</i>	15
3.1.2 <i>k-Anonymity</i> untuk Menganonimisasi Data	15
3.1.3 <i>Total Information Loss</i>	16
3.2 Studi Kasus	16
3.2.1 Algoritma <i>Datafly</i>	17
3.2.2 Algoritma <i>Greedy k-member Clustering</i>	18
3.3 Gambaran Umum Perangkat Lunak	20
3.3.1 Diagram Aktivitas Perangkat Lunak	21
3.3.2 Diagram Kelas Perangkat Lunak	23
4 PERANCANGAN	27
4.1 Perancangan Antarmuka	27
4.2 Diagram Kelas Lengkap	29
4.3 Masukan Perangkat Lunak	51

5	IMPLEMENTASI DAN EKSPERIMEN	61
5.1	Implementasi Antarmuka	61
5.2	Pengujian	65
5.2.1	Pengujian Fungsional	66
5.2.2	Pengujian Eksperimental	67
6	KESIMPULAN DAN SARAN	77
6.1	Kesimpulan	77
6.2	Saran	78
	DAFTAR REFERENSI	79
A	KODE PROGRAM	81
A.1	<i>Package anonymization_algorithm</i>	81
A.1.1	Kelas AnonymizationAlgorithm	81
A.1.2	Kelas AnonymizationInfo	83
A.1.3	Kelas AATuple	86
A.1.4	Kelas AACluster	89
A.1.5	Kelas AAPrinter	94
A.2	<i>Package domain_generalization_hierarchy</i>	96
A.2.1	Kelas DomainGeneralizationHierarchy	96
A.2.2	Kelas Tree	98
A.2.3	Kelas Node	100
A.3	<i>Package datafly</i>	103
A.3.1	Kelas Datafly	103
A.3.2	Kelas FrequencyRow	105
A.4	<i>Package greedy_k_member_clustering</i>	106
A.4.1	Kelas GreedyKMemberClustering	106
A.4.2	Kelas GKMCCluster	108
A.4.3	Kelas GKMCTuple	109
A.4.4	Kelas GKMCNumberDomain	110
A.4.5	Kelas GKMCDateDomain	112
A.5	<i>Package utils</i>	112
A.5.1	Kelas CsvParser	112
A.5.2	Kelas CsvRow	114
A.5.3	Kelas CsvWriter	116
A.5.4	Kelas Pair	117
A.5.5	Kelas DateHelper	118
A.5.6	Kelas StringRegex	118
A.5.7	Kelas Logger	119
A.5.8	Kelas Constants	122
A.5.9	Kelas FXMLHelper	122
A.6	<i>Package ui</i>	125
A.6.1	Kelas Main	125
A.7	<i>Package controllers</i>	125
A.7.1	Kelas MainMenuController	125
A.7.2	Kelas AnonymizeDialogController	138
A.7.3	Kelas LoggerController	139

DAFTAR GAMBAR

2.1	Hierarki generalisasi untuk atribut umur (z_0) pada Tabel 1.2	7
3.1	DGH yang digunakan untuk menggeneralisasi Tabel 3.1	17
3.5	Diagram kelas perangkat lunak anonimisasi data	23
3.2	Diagram aktivitas perangkat lunak anonimisasi data	25
3.3	Diagram aktivitas proses <i>clustering</i> data menggunakan perangkat lunak lain	26
3.4	Diagram aktivitas perangkat lunak analisis	26
4.1	Prototipe tampilan informasi data	28
4.2	Prototipe tampilan dialog anonimisasi	28
4.3	Prototipe tampilan hasil anonimisasi	28
4.4	Prototipe tampilan informasi tabel	29
4.5	Prototipe tampilan perbandingan nilai atribut pada tiap tabel	30
4.10	Diagram kelas pada <i>package greedy_k_member_clustering</i> pada perangkat lunak anonimisasi data.	38
4.13	Diagram kelas singkat dan hubungan antara kelas-kelasnya pada perangkat lunak analisis.	49
4.6	Diagram kelas singkat dan hubungan antara kelas-kelasnya pada perangkat lunak anonimisasi data.	54
4.7	Diagram kelas pada <i>package anonymization_algorithm</i> pada perangkat lunak anonimisasi data.	55
4.8	Diagram kelas pada <i>package domain_generalization_hierarchy</i> pada perangkat lunak anonimisasi data.	56
4.9	Diagram kelas pada <i>package datafly</i> pada perangkat lunak anonimisasi data.	56
4.11	Diagram kelas pada <i>package utils</i> pada perangkat lunak anonimisasi data.	57
4.12	Diagram kelas pada <i>package ui</i> pada perangkat lunak anonimisasi data.	58
4.14	Diagram kelas pada <i>package ui</i> pada perangkat lunak analisis.	59
5.1	Tampilan utama perangkat lunak anonimisasi data	61
5.2	Tampilan utama perangkat lunak anonimisasi data setelah pengguna memasukkan data ke perangkat lunak	62
5.3	Dialog untuk memasukkan nilai k sebelum melakukan anonimisasi data	62
5.4	Tampilan hasil anonimisasi perangkat lunak anonimisasi data setelah perangkat lunak selesai mengeksekusi algoritma anonimisasi data	63
5.5	Tampilan <i>console log</i> pada perangkat lunak anonimisasi data	64
5.6	Tampilan utama perangkat lunak analisis	64
5.7	Dialog untuk memasukkan sebuah tabel privat atau tabel hasil anonimisasi data	64
5.8	Tampilan utama perangkat lunak analisis setelah pengguna memasukkan suatu data ke perangkat lunak	65
5.9	Tampilan perbandingan atribut perangkat lunak analisis	65
5.10	Hasil anonimisasi data yang dilakukan oleh perangkat lunak anonimisasi data menggunakan algoritma <i>Datafly</i>	66

5.11 Hasil anonimisasi data yang dilakukan oleh perangkat lunak anonimisasi data menggunakan algoritma <i>Greedy k-member Clustering</i>	66
5.12 Grafik perbandingan antara nilai k dengan <i>information loss</i> pada pengujian eksperimental dengan menggunakan semua atribut	69
5.13 Grafik perbandingan antara nilai k dengan waktu eksekusi pada pengujian eksperimental dengan menggunakan semua atribut	69
5.14 Grafik perbandingan antara nilai k dengan banyaknya <i>tuples</i> dengan <i>cluster</i> yang sama dengan <i>tuples</i> pada tabel privat pada pengujian eksperimental dengan menggunakan semua atribut	70
5.15 Grafik perbandingan antara nilai k dengan <i>information loss</i> pada pengujian eksperimental dengan menggunakan atribut yang bertipe numerikal atau tanggal	72
5.16 Grafik perbandingan antara nilai k dengan waktu eksekusi pada pengujian eksperimental dengan menggunakan atribut yang bertipe numerikal atau tanggal	72
5.17 Grafik perbandingan antara nilai k dengan banyaknya <i>tuples</i> dengan <i>cluster</i> yang sama dengan <i>tuples</i> pada tabel privat pada pengujian eksperimental dengan menggunakan atribut yang bertipe numerikal atau tanggal	73
5.18 Grafik perbandingan antara nilai k dengan <i>information loss</i> pada pengujian eksperimental dengan menggunakan atribut yang bertipe kategorikal	75
5.19 Grafik perbandingan antara nilai k dengan waktu eksekusi pada pengujian eksperimental dengan menggunakan atribut yang bertipe kategorikal	75
5.20 Grafik perbandingan antara nilai k dengan banyaknya <i>tuples</i> dengan <i>cluster</i> yang sama dengan <i>tuples</i> pada tabel privat pada pengujian eksperimental dengan menggunakan atribut yang bertipe kategorikal	76

DAFTAR TABEL

1.1	Contoh <i>database</i> pasien kesehatan.	1
1.2	Contoh <i>database</i> pasien kesehatan yang bersifat <i>3-anonymized</i>	2
3.1	Tabel daftar orang dan penyakit yang dideritanya yang digunakan sebagai contoh kasus	17
3.2	Tabel frekuensi <i>freq</i> yang didapatkan dari langkah 2 Algoritma 1	18
3.3	Tabel frekuensi <i>freq</i> yang didapatkan setelah melakukan satu kali iterasi dalam perulangan langkah 3 Algoritma 1	18
3.4	MGT yang dihasilkan dari tabel frekuensi <i>freq</i>	19
3.5	Tabel kasus dengan sedikit modifikasi	19
3.6	Tabel hasil yang dibangun dari himpunan <i>cluster result</i>	20
3.7	Tabel hasil yang dibangun dari himpunan <i>cluster result</i> yang sudah digeneralisasikan	20
5.1	Sampel tabel privat dengan menggunakan semua atribut, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	68
5.2	Sampel hasil anonimisasi algoritma <i>Datafly</i> dengan menggunakan semua atribut dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	68
5.3	Sampel hasil anonimisasi algoritma <i>Greedy k-member Clustering</i> dengan menggunakan semua atribut dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	69
5.4	Sampel tabel privat dengan menggunakan atribut yang bertipe numerikal atau tanggal, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	71
5.5	Sampel hasil anonimisasi algoritma <i>Datafly</i> dengan menggunakan atribut yang bertipe numerikal atau tanggal dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	71
5.6	Sampel hasil anonimisasi algoritma <i>Greedy k-member Clustering</i> dengan menggunakan atribut yang bertipe numerikal atau tanggal dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	71
5.7	Sampel tabel privat dengan menggunakan atribut yang bertipe kategorikal, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	74
5.8	Sampel hasil anonimisasi algoritma <i>Datafly</i> dengan menggunakan atribut yang bertipe kategorikal dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	74
5.9	Sampel hasil anonimisasi algoritma <i>Greedy k-member Clustering</i> dengan menggunakan atribut yang bertipe kategorikal dan $k = 20$, yang sudah dilakukan <i>clustering</i> menggunakan Weka.	74

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Data mining atau penambangan data pada dasarnya adalah salah satu kegiatan yang bertujuan untuk mendapatkan informasi dari sebuah data [1]. Informasi yang didapat umumnya berguna untuk analisis, penelitian, dan sebagainya. Tetapi, data yang digunakan untuk melakukan *data mining* mungkin terdapat data pribadi milik seseorang. Bila data tersebut digunakan untuk *data mining*, maka ada kemungkinan data pribadi dalam data tersebut terbuka. Maka dari itu, perlu adanya pendekatan agar privasi data dapat terjaga walaupun dilakukan *data mining* terhadap data tersebut.

Pada dasarnya, privasi merupakan kemampuan seseorang dalam mengatur data pribadinya. Privasi melindungi data pribadi tersebut dengan mengatur bagaimana informasi tersebut diperoleh, disimpan, digunakan, dan pada akhirnya dihapus [2]. *Privacy preserving data mining* merupakan salah satu cara untuk melindungi data pribadi pada sebuah data agar data tersebut bisa digunakan untuk kepentingan analisis dan penelitian, tetapi tidak ada informasi pribadi yang terbuka dari data tersebut [3].

Terdapat dua pendekatan agar tujuan dari *privacy preserving data mining* tercapai, yaitu pendekatan kriptografi dan pendekatan anonimisasi. Pada pendekatan kriptografi, data yang ingin diberikan akan dienkripsi terlebih dahulu, sehingga tidak ada informasi pribadi yang terbuka. Tetapi, jika dimisalkan pada kegiatan analisis dan penelitian dibutuhkan informasi tentang umur, sedangkan informasi tentang umur terenkripsi menggunakan pendekatan ini, maka data tersebut tidak bisa digunakan untuk kegiatan tersebut. Pendekatan lainnya adalah pendekatan anonimisasi, yaitu data yang ingin diberikan dianonimkan terlebih dahulu. Dengan kata lain, informasi dalam data tersebut digeneralisasikan atau dihilangkan sebagian. Salah satu model yang umum digunakan dalam pendekatan ini adalah *k-anonymity*. *k-anonymity* secara garis besar adalah suatu metode perlindungan privasi dengan cara membuat suatu entri dalam data menjadi tidak dikenali dengan $k - 1$ buah entri lainnya [4]. Hasil anonimisasi menggunakan *k-anonymity* dapat dilihat pada Tabel 1.2, di mana data tersebut adalah hasil anonimisasi dari Tabel 1.1.

Tabel 1.1: Contoh *database* pasien kesehatan.

Nama	Umur	Jenis Kelamin	Kode Pos	Penyakit
Tyson	20	Pria	443350	Diabetes
Weber	22	Pria	443351	Kanker
Tim	24	Pria	443352	Flu
Levine	32	Pria	443350	Hepatitis
Ocean	34	Wanita	443350	Hepatitis
Bob	38	Wanita	443350	Hepatitis

Pada penelitian ini, dibangun sebuah perangkat lunak yang mengimplementasikan dua buah algoritma *k-anonymity*, yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Dengan adanya perangkat lunak tersebut, informasi pribadi seseorang yang terdapat dalam suatu data tetap

Tabel 1.2: Contoh *database* pasien kesehatan yang bersifat *3-anonymized*.

ID	Umur	Jenis Kelamin	Kode Pos	Penyakit
1	[20-29]	Orang	44335*	Diabetes
2	[20-29]	Orang	44335*	Kanker
3	[20-29]	Orang	44335*	Flu
4	[30-39]	Orang	443350	Hepatitis
5	[30-39]	Orang	443350	Hepatitis
6	[30-39]	Orang	443350	Hepatitis

terlindungi walaupun data tersebut digunakan untuk kepentingan analisis dan penelitian. Perangkat lunak yang dibangun dapat menerima masukan tabel data yang ingin dianonimisasi menggunakan algoritma *k-anonymity*. Tabel data tersebut memiliki baris-baris yang merepresentasikan sebuah *record*. Keluaran dari perangkat lunak ini adalah hasil anonimisasi algoritma *k-anonymity* dari tabel data yang diberikan.

1.2 Rumusan Masalah

Rumusan masalah yang muncul berdasarkan deskripsi dan latar belakang yang sudah dibahas adalah sebagai berikut:

1. Bagaimana cara kerja kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*, dalam melindungi privasi?
2. Bagaimana membangun perangkat lunak yang mengimplementasikan kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*?
3. Bagaimana menguji hasil implementasi kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*, menggunakan teknik *data mining* terhadap data yang sudah dianonimisasi?

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mempelajari metode kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*, dalam melindungi privasi.
2. Mengimplementasikan perangkat lunak yang mengimplementasikan kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*.
3. Menguji hasil anonimisasi yang didapat dari kedua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan *Greedy k-member Clustering*, dengan menerapkan teknik *data mining* pada hasil anonimisasi tersebut.

1.4 Batasan Masalah

Batasan-batasan masalah untuk penelitian ini adalah sebagai berikut:

1. Perangkat lunak yang dibuat hanya dapat menerima masukan yang lengkap dan valid, karena perangkat lunak yang dibangun hanya sebagai alat untuk membantu penelitian ini, sehingga perangkat lunak tidak dibangun untuk digunakan secara umum.

2. Jumlah *record* maksimal untuk dianonimisasi adalah 1000 buah *record*, karena terdapat algoritma dengan kompleksitas yang tinggi, yaitu algoritma *Greedy k-member Clustering*.

1.5 Metodologi

Metodologi yang digunakan dalam penyusunan penelitian ini adalah:

1. Melakukan studi literatur mengenai dasar-dasar privasi data.
2. Melakukan studi literatur mengenai cara kerja *k-anonymity*.
3. Melakukan studi literatur mengenai teknik *data mining*.
4. Melakukan analisis masalah dan pengujian dengan kasus sederhana untuk mendapatkan gambaran tentang proses anonimisasi menggunakan *k-anonymity*.
5. Melakukan perancangan perangkat lunak yang dibuat.
6. Mengimplementasikan perangkat lunak yang sudah dirancang.
7. Melakukan pengujian eksperimental pada sebuah data menggunakan teknik *data mining*.
8. Melakukan analisis hasil dari eksperimen yang telah dilakukan.
9. Menarik kesimpulan berdasarkan hasil eksperimen yang telah dilakukan.

1.6 Sistematika Pembahasan

Laporan penelitian tersusun ke dalam enam bab secara sistematis sebagai berikut:

- Bab 1 Pendahuluan
Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- Bab 2 Dasar Teori
Berisi dasar teori tentang dasar-dasar dari privasi, *k-anonymity* berbasis *clustering*, dan metode dalam *data mining*.
- Bab 3 Analisis Masalah
Berisi analisis masalah, studi kasus, dan diagram aliran proses.
- Bab 4 Perancangan
Berisi perancangan perangkat lunak yang dibangun, meliputi perancangan antarmuka dan diagram kelas yang lengkap.
- Bab 5 Implementasi dan Pengujian
Berisi implementasi antarmuka perangkat lunak, pengujian fungsional, pengujian eksperimental, dan kesimpulan dari pengujian.
- Bab 6 Kesimpulan dan Saran
Berisi kesimpulan dari awal hingga akhir penelitian dan saran untuk pengembangan selanjutnya.