JURNAL PINTER
VOL 4. NO. 1 JUNI 2020
E - ISSN: 2597 - 4475
http://doi.org.10.21009/pinter.4.1.10

ANALISIS MODEL *L-DIVERSITY* DENGAN ALGORITMA *SYSTEMATIC CLUSTERING* DAN *DATAFLY*

Shafa Sya'airillah[1], Widodo., M.Kom[2], Bambang Prasetya A., S.Pd., M.Kom[3]

[1] Mahasiswa Prodi PendidikanTeknik Informatika dan Komputer, Teknik Elektro, FT – UNJ
[2,3] Dosen Prodi Pendidikan Teknik Informatika dan Komputer, Teknik Elektro, FT – UNJ
[1] syaairillahshafa@gmail.com, [2] widodo@unj.ac.id, [3] bambangpadhi@unj.ac.id

Abstrak

Penelitian ini dilatar belakangi oleh teknik anonimitas data yang terdapat pada Privacy Preserving Data Publishing. Sehingga data yang ingin dipublikasikan bersifat anonim, tanpa mengungkap informasi yang sebenarnya. Metode penelitian yang digunakan pada penelitian ini adalah rekayasa teknik dengan cara menghitung nilai information loss yang dihasilkan pada masing-masing algoritma, kemudian membandingkannya. Model yang digunakan pada penelitian ini adalah l-Diversity. Algoritma yang digunakan adalah algoritma Systematic Clustering dan algoritma Datafly. Data yang digunakan adalah dataset 'Adult' yang diunduh dari repositori UCI Machine Learning. Sampel yang digunakan dari dataset 'Adult' ini adalah sebanyak 2000 tuple. Nilai information loss tertinggi yang dihasilkan algoritma Systematic Clustering adalah 475673.19, sedangkan nilai information loss tertinggi dari algoritma Datafly adalah 46298.00. Kemudian, untuk nilai information loss terendah yang dihasilkan algoritma Systematic Clustering adalah 22364.79, sedangkan nilai information loss terendah dari algoritma Datafly adalah 36659.00. Algoritma dengan tingkat information loss paling kecil dianggap sebagai algoritma yang paling baik dalam membangun model l-Diversity di antara kedua algoritma yang diuji. Hasil pengujian menyatakan bahwa algoritma Systematic Clustering adalah algoritma yang paling baik dalam membangun model l-Diversity di antara algoritma Systematic Clustering dan Datafly.

Kata kunci : ***k-Anonymity, Homogeneity Attack, Background Knowledge Attack, l-Diversity, Systematic Clustering, Datafly, Information Loss* (IL).**

1. Pendahuluan

Dalam kehidupan manusia saat ini tidak lepas dari data. Baik organisasi, instansi, perusahaan, bahkan per individu manusia itu sendiri memiliki sekumpulan dari data. Namun pada hakikatnya, data mentah tersebut tentunya memiliki atribut atau muatan bernilai sensitif, maksudnya berisi atribut yang tidak semua orang bahkan tidak sama sekali boleh mengetahuinya.

Teknik yang memungkinkan seseorang untuk mempublikasikan atau membagikan data yang bersifat anonim dengan tujuan melindungi agar tidak terungkapnya atribut sensitif atau data pribadi milik orang lain disebut dengan *Privacy Preserving Data Publishing* (PPDP). Menurut Kabir, dkk. (2010: 94) salah satu konsep penting dalam PPDP adalah data *anonymization* atau bisa dikenal dengan *anonymity*.

Salah satu pendekatan yang biasa digunakan dalam privasi data adalah *k-Anonymity* yang diusulkan oleh Samarati pada tahun 2001 dan Sweeney pada tahun 2002. Walaupun begitu, dengan model *k-Anonymity* ini masih terdapat kelemahan dimana data sensitif yang ada bisa saja terungkap apabila terjadi *homogeneity attack* dan *background knowledge attack* (Kabir, dkk., 2010: 94, diacu dari Machanavajjhala, 2006).

Karena permasalahan yang ada pada *k-Anonymity* itulah, lahir sebuah model *l-Diversity* yang diperkenalkan oleh Machanavajjhala, dkk. pada tahun 2007 yang diharapkan dapat mengatasi permasalahan tersebut. Dalam membangun model *l-Diversity* ini, dalam penelitian ini penulis menggunakan algoritma *Systematic Clustering* dan *Datafly*.

Algoritma *Systematic Clustering* akan menggunakan metode dengan cara mengelompokkan data menjadi beberapa cluster. Dimana akan dipilih *record r* secara acak sebagai acuan dalam membentuk *cluster*. Kemudian data akan dikelompokkan berdasarkan perbandingan jarak terdekat dengan *cluster* yang sudah ditentukan tersebut. Kemudian, untuk algoritma *Datafly* sendiri menggunakan prinsip dengan cara menggeneralisasi secara umum, mensubtitusi, serta menghapus informasi secara tepat tanpa takut kehilangan banyak detil yang telah ditemukan dari data tersebut.

Pada penelitian ini akan dibandingkan kinerja dari dua algoritma tersebut berdasarkan tingkat *information loss* dan *running time*-nya. Algoritma yang menghasilkan tingkat *information loss* yang lebih rendah dianggap sebagai algoritma yang lebih baik dalam membangun model *l-Diversity*.

2. Dasar Teori

Model *l-Diversity* adalah salah satu model yang dihasilkan dari teknik anonimitas data yang terdapat pada PPDP (*Privacy Preserving Data Publishing*). Dimana pada model *l-Diversity* ini dapat mengatasi kelemahan yang terdapat pada model *k-Anonymity*, yaitu *homogeneity attack* dan *background knowledge attack*.

2.1. Data

Menurut Hermansyah dan Nurhayati (2012: 14), diacu dalam Edhy Sutanta (2004: 18) mendefinisikan data sebagai bahan keterangan tentang kejadian nyata atau fakta-fakta yang dirumuskan dalam sekelompok lambang tertentu yang tidak acak yang meunjukkan jumlah, tindakan, atau hal. Jenis-jenis atribut atau data menurut C. M. Fung, dkk. (2011: 7) yaitu: *explicit identifier, quasi-dentifier, sensitive identifier*, dan *non-sensitive identifier*.

2.2. Privasi

Kemudian menurut Altman (1975: 221), privasi adalah proses pengontrolan yang selektif terhadap akses pada diri sendiri dan akses kepada diri sendiri dan akses kepada orang lain.

2.3. *Privacy Preserving Data Publishing*

Privacy Preserving Data Publishing (PPDP) menurut Charlie & Zaman (2014: 129), adalah salah satu cara untuk memungkinkan seseorang untuk membagikan data anonim untuk memastikan perlindungan dari terungkapnya identitas individu.

2.4. Anonim

Dalam Kamus Besar Bahasa Indonesia (KBBI), anonimitas diartikan sebagai [n] hal tidak ada nama. Untuk definisi anonim sendiri diartikan dalam kamus KBBI sebagai:

1) tanpa nama; tidak beridentitas; awanama;
2) [Sos] tidak ada penandatangannya

2.5. 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.

2.6. *K-Anonymity*

Model *k-Anonymity* ini dikenalkan oleh Samarati (2001) dan Sweeney (2002). Menurut Sweeney (2002), model *k-Anonymity* ini merupakan perbaikan dari model sebelumnya, yaitu *k-Map*.

Suatu tabel dikatakan memenuhi kondisi *k-Anonymity* apabila masing-masing *record* pada sebuah grup quasi-identifier data anonim, tidak bisa dibedakan dengan syarat minimal *k* - 1 dengan *record* lainnya dalam grup quasi-identifier tersebut.

2.7. *L-Diversity*

Model ini diperkenalkan oleh Machanavajjhala, dkk. (2007) dari *Department of Computer Science, Cornell University*. Model ini sebenarnya memiliki kesamaan dengan model *p-sensitive k-Anonymity*. Model ini mengharuskan setiap grup *quasi-identifier* setidaknya memiliki nilai atribut sensitif yang *l* "*well-represented*".

2.8. Algoritma

Pengertian algoritma menurut Wahid (2004: 2) adalah urutan langkah-langkah yang menyatakan dengan jelas dan tidak rancu untuk memecahkan suatu masalah (jika ada pemecahannya) dalam rentang waktu tertentu.

2.9. *Systematic Clustering*

Tahap *clustering* menggunakan algoritma *Systematic Clustering* adalah sebagai berikut: langkah pertama yang dilakukan yaitu mengurutkan data berdasarkan *quasi-identifier*, kemudian buatlah *cluster* pertama dengan mengambil data sebanyak *k* dari data urutan terkecil. Selanjutnya adalah proses melakukan generalisasi dan supresi pada data dalam satu *cluster* yang sama. Lakukan langkah ketiga tersebut sampai data habis. Apabila ada data yang tersisa, masukkan data tersebut ke dalam *cluster* terdekat yang memiliki tingkat *information loss* paling kecil. Apabila tabel yang dihasilkan belum memenuhi kriteria *l-Diversity*, maka akan dilakukan proses *clustering* kembali agar nilai atribut sensitif yang terdapat pada suatu *cluster* memenuhi sebanyak *l*. Dimana nilai *l* harus lebih atau sama dengan dua.

2.10. *Datafly*

Proses menganonimkan data dengan menggunakan algoritma *Datafly* adalah sebagai berikut, yakni yang pertama adalah *record* akan dikelompokkan (dihitung frekuensi kemunculan data) terlebih dahulu berdasarkan *quasi-identifier* yang memiliki *value* yang sama. Jumlah anggota perkelompok data akan ditentukan berdasarkan parameter kondisi *k* dan *l*. Jika tidak memenuhi kondisi *k* dan *l*, maka akan dilakukan proses generalisasi secara global (umum) terhadap salah satu atribut *quasi-identifier* yang memiliki nilai distinct terbesar. Kemudian akan dihitung kembali frekuensi perkelompok data berdasarkan QID setelah dilakukan generalisasi tersebut. *Record* yang tidak termasuk dalam salah satu di antara *cluster* atau kelompok data yang ada (biasa disebut dengan *outlier*), akan otomatis dieleminasi atau dihapus dari database.

2.11. *Information Loss*

Menurut Zhi-ting Yu, dkk. (2015: 87), *information loss* ini digunakan untuk mengukur perbedaan antara dataset yang asli dengan dataset yang telah dianonimisasi. Jadi, tujuan dihitungnya nilai *information loss* ini adalah untuk mengetahui seberapa banyak informasi yang hilang selama proses *k-Anonymization*. Rumus untuk menghitung *information loss* ini diacu dari teknik yang digunakan oleh Byun, dkk. (2007) yaitu:

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

Keterangan:
Ω : jumlah data yang terdapat dalam suatu *cluster*
Ƞ : kumpulan *record*
N : *quasi-identifier* yang bersifat numerikal (N_1, N_2,..., N_r)
$N_{i\ max}$: nilai maksimal (bersifat numerik) yang terdapat pada *cluster*
$N_{i\ min}$: nilai minimum (bersifat numerik) yang terdapat pada *cluster*
C : *quasi-identifier* yang bersifat kategorial (C_1, C_2,..., N_s)
($\cup C_j$) : *taxonomy tree* (pohon kategorial)
$\tau(\cup C_j)$: nilai taksonomi terendah dari suatu akar *taxonomy tree*
H(τ) : taksonomi tertinggi pada suatu *taxonomy tree*

3. Metodologi

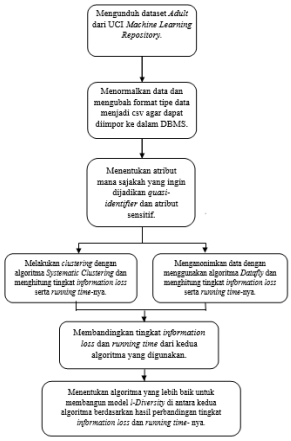
3.1 Tempat dan Waktu Penelitian

Penelitian dilakukan di Laboratorium Komputer Program Studi Pend. TIK Gedung L2 Lantai 2, Fakultas Teknik. Universitas Negeri Jakarta. Selama dua semester, yakni semester 108 (TA. 2017/2018) dan semester 109 (TA. 2018/2019).

3.2 Alat dan Bahan Penelitian

Perangkat keras (*hardware*) yang digunakan adalah laptop ASUS A456UR (Intel® Core™ i5 7200U Processor, OS Windows 10 Pro 64-bit, 4 GB SDRAM, Storage 1TB 5400RPM SSH). Adapun perangkat lunak (*software*) yang digunakan adalah Netbeans IDE 8.1, XAMPP Versi 3.2.2, dan Ms. Office 2016. Tidak lupa koneksi jaringan internet dan dataset *Adult* dari repositori UCI *Machine Learning* yang akan dijadikan sampel data pada penelitian ini.

3.3 Diagram Alir Penelitian



Gambar 0.1. Diagram Alir Penelitian

Pada penelitian ini akan dibangun dua model *l-Diversity* dengan dua jenis algoritma yang berbeda, yakni algoritma *Systematic Clustering* dan *Datafly*. Dalam membangun model *l-Diversity* ini penulis menggunakan dataset *Adult* dari UCI *Machine Learning*. Data yang akan diuji hanya diambil 2.000 *record* sebagai sampel. Hal pertama yang dilakukan adalah melakukan normalisasi data. Setelah selesai dalam proses normalisasi data, adalah menentukan atribut mana saja yang akan dijadikan atribut sensitif dan *quasi-identifier*.

Pada penelitian ini, penulis memilih atribut *Marital-Status* sebagai atribut sensitif. Kemudian untuk *quasi-identifier*-nya adalah atribut *Age*, *Gender*, *Education*, *Occupation*, dan *Native Country*. Setelah ditentukan atribut sensitif dan *quasi-identifier*-nya, maka dataset akan diimpor ke dalam DBMS dan dianggap sudah siap diolah untuk tahap *clustering*.

Tahap analisis dan perhitungan *information loss* ini dilakukan sebanyak empat kali dengan nilai *k* yang berbeda-beda, yakni dimulai dari *k* = 3 sampai dengan *k* = 6. Langkah terakhir dalam proses penelitian ini adalah membandingkan nilai dari *information loss* serta *running time* yang dihasilkan dari masing-masing algoritma. Algoritma yang memiliki tingkat information loss paling rendah dan running time paling cepat adalah algoritma yang lebih baik dalam membangun model *l-Diversity*.

4. Hasil dan Analisis

Berikut hasil penelitian dari perhitungan tingkat *information loss* dan *running time* yang dihasilkan dari kedua algoritma yang diuji.

4.1 Tabel Hasil *Output l-Diversity* dengan Algoritma *Systematic Clustering*

Tabel 0.1. *l-Diversity* dengan *k* =4; *l* =3

Age	Gender	Education	Occupation	Native Country	Marital Status	Cluster
17-17	Gender	Educated	Worked	World	Never-married	1
17-	Gender	Educated	Work	World	Never-	1

17			ed		married	
17-17	Gender	Educated	Worked	World	Never-married	1
17-17	Gender	Educated	Worked	World	Never-married	1
17-17	Gender	Educated	Worked	World	Never-married	1
17-17	Gender	Educated	Worked	World	Never-married	1
17-17	Gender	Educated	Worked	World	Never-married	1
17-17	Gender	Educated	Worked	World	Married-civ-spouse	1
17-18	Gender	Educated	Worked	World	Married-civ-spouse	1
17-18	Gender	Educated	Worked	World	Never-married	1
17-18	Gender	Educated	Worked	World	Never-married	1
17-18	Gender	Educated	Worked	World	Never-married	1
17-18	Gender	Educated	Worked	World	Never-married	1
17-18	Gender	Educated	Worked	World	Divorced	1
18-18	Gender	Educated	Worked	World	Never-married	2
18-18	Gender	Educated	Worked	World	Never-married	2
18-18	Gender	Educated	Worked	World	Married-civ-spouse	2
18-18	Gender	Educated	Worked	World	Never-married	2
18-18	Gender	Educated	Worked	World	Married-civ-spouse	2
18-18	Gender	Educated	Worked	World	Never-married	2
18-18	Gender	Educated	Worked	World	Widowed	2

Maksud k = 4 disini yakni jumlah minimal anggota dalam suatu cluster adalah 4 (4 *record* atau 4 Id). Sedangkan parameter kedua yakni l = 3 adalah jumlah minimal *distinct* (perbedaan atau jenis) dari atribut sensitif adalah 3. Proses *looping* penambahan anggota dalam suatu *cluster* akan terus berlanjut hingga *distinct* (*l*) dalam suatu *cluster* tersebut telah terpenuhi sesuai parameter yang diinput atau sejumlah sama dengan *k*.

4.2 Tabel Hasil *Output l-Diversity* dengan Algoritma *Datafly*

Tabel 0.2. *l-Diversity* dengan k =3; l =3

Age	Gender	Education	Occupation	Native Country	Marital Status	Cluster
17-90	Male	11th	Sales	United-States	Never-married	1
17-	Male	11th	Sales	United-	Divorce	1
90				States	d	
17-90	Male	11th	Sales	United-States	Divorced	1
17-90	Male	11th	Sales	United-States	Divorced	1
17-90	Male	11th	Sales	United-States	Married-civ-spouse	1
17-90	Male	10th	Craft-repair	United-States	Never-married	2
17-90	Male	10th	Craft-repair	United-States	Divorced	2
17-90	Male	10th	Craft-repair	United-States	Separated	2
17-90	Male	10th	Craft-repair	United-States	Never-married	2
17-90	Male	10th	Craft-repair	United-States	Separated	2
17-90	Male	10th	Craft-repair	United-States	Separated	2
17-90	Male	10th	Craft-repair	United-States	Married-spouse-absent	2
17-90	Male	10th	Craft-repair	United-States	Separated	2
17-90	Male	10th	Craft-repair	United-States	Married-civ-spouse	2
17-90	Male	10th	Craft-repair	United-States	Divorced	2
17-90	Male	10th	Craft-repair	United-States	Married-civ-spouse	2
17-90	Male	10th	Craft-repair	United-States	Never-married	2
17-90	Male	10th	Craft-repair	United-States	Never-married	2
17-90	Male	10th	Craft-repair	United-States	Married-civ-spouse	2
17-90	Male	10th	Craft-repair	United-States	Widowed	2

Maksud k = 3 disini yakni jumlah minimal anggota dalam suatu *cluster* adalah 3 (3 *record* atau 3 Id). Sedangkan parameter kedua yakni l = 3 adalah jumlah minimal *distinct* (perbedaan atau jenis) dari atribut sensitif adalah 3. Atribut *quasi-identifier* yang digeneralisasi hanya atribut *Age* saja. Dan proses generalisasi dilakukan secara *global recoding*.

Setelah itu, data akan dikelompokkan berdasarkan *quasi-identifier* dan dihitung *occurs*-nya (frekuensi data) per kelompok kuasi. Jika suatu kelompok *quasi-identifier* memiliki *occurs* yang

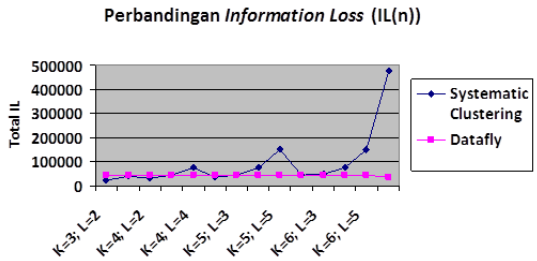
kurang dari parameter *k* dan atribut sensitifnya (*marital status*) kurang dari parameter *l* yang diinputkan pada program, maka otomatis satu kelompok data *quasi-identifier* tersebut akan dihapus dari database karena terhitung sebagai *outlier*.

4.3 Tabel Perbandingan Hasil *Information Loss* dan *Running Time* dari Kedua Algoritma

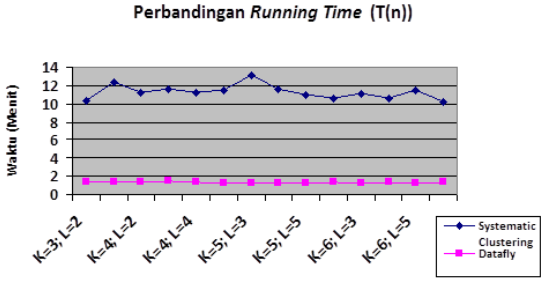
Tabel 0.3 Perbandingan *Information Loss* dan *Running Time* dari Kedua Algoritma

No	*k* (*n*)	*l* (*n*)	*Information Loss*		*Running Time*	
			IL (Ω) SC	IL (Ω) DF	T(*n*) SC	T(*n*) DF
1.	3	2	22364.79	46298.00	10:31	1:42
		3	41759.62	46033.00	12:39	1:34
2.	4	2	29175.86	46127.00	11:23	1:35
		3	42150.90	45961.00	11:55	1:52
		4	76687.20	45382.00	11:18	1:36
3.	5	2	36961.05	45935.00	11:44	1:31
		3	44752.79	45849.00	13:18	1:23
		4	76463.83	45334.00	11:58	1:26
		5	150873.79	43746.00	11:02	1:31
4.	6	2	44872.24	45485.00	10:55	1:36
		3	49713.35	45449.00	11:13	1:30
		4	76131.02	45134.00	10:57	1:35
		5	150873.79	43721.00	11:46	1:28
		6	475673.19	36659.00	10:26	1:33

4.4 Grafik Per Perbandingan Hasil *Information Loss* dan *Running Time* dari Kedua Algoritma



Gambar 4.2. Grafik Perbandingan IL dari Kedua Algoritma



Gambar 4.2. Grafik Perbandingan *Running Time* dari Kedua Algoritma

Berdasarkan perbandingan tingkat *information loss* dari kedua algoritma tersebut, maka dapat disimpulkan bahwa algoritma *Systematic Clustering* adalah algoritma yang lebih baik dalam membangun model *l-Diversity* dibandingkan algoritma *Datafly*. Hal itu dikarenakan total information loss minimum yang dihasilkan algoritma *Systematic Clustering* cenderung lebih rendah dibandingkan total *information loss* minimum yang dihasilkan oleh algoritma *Datafly*, yakni sebesar 22364.79.

Namun dari segi tingkat efektivitas algoritma, maka dapat disimpulkan bahwa algoritma *Datafly* adalah algoritma yang lebih baik dalam membangun model *l-Diversity* dibandingkan algoritma *Systematic Clustering*. Hal ini terlihat dari rata-rata waktu yang dibutuhkan algoritma *Datafly* yang sangat singkat dalam sekali eksekusi program, yakni hanya sekitar satu menit. Sedangkan algoritma *Systematic Clustering* membutuhkan waktu rata-rata sekitar sepuluh menit per sekali eksekusi program.

4.5 Kelemahan pada Model *l-Diversity*

Diacu pada paper milik Chourasia, dkk. (2017: 803-804) dan Jayabalan, M., dkk. (2017:176) seperti halnya pada *k-Anonymity*, model *l-Diversity* ini dianggap masih terdapat dua kelemahan, yaitu *Skewness Attack* dan *Similarity Attack*. *Skewness attack* ini adalah suatu kondisi dimana suatu frekuensi atribut sensitif dalam satu *quasi-identifier* terlihat kesenjangan (perbedaan) yang cukup signifikan (terutama apabila kondisi l = 2), sehingga atribut sensitif dari suatu cluster akan terlihat lebih condong ke salah satu jenis atribut sensitif. Kemudian untuk *similarity attack* adalah suatu kondisi dimana atribut-atribut sensitif dalam suatu *cluster* terlihat berbeda, namun memiliki maksud/makna yang sama.

Pada penelitian ini, data sudah disederhanakan sedemikian rupa agar distribusinya terbagi merata berdasarkan kelompok umur, sehingga kemungkinan terjadinya *skewness attack* sangat kecil. Kemudian untuk contoh dari *similarity attack* dapat dilihat pada tabel 4.4 di bawah ini:

Tabel 0.4 Contoh *Similarity Attack* (*k* =4; *l* =2)

Age	Gender	Education	Occupation	Native Country	Marital Status	Cluster
17-90	Female	10th	Sales	United-States	Married-civ-spouse	4
17-90	Female	10th	Sales	United-States	Separated	4
17-90	Female	10th	Sales	United-States	Separated	4
17-90	Female	10th	Sales	United-States	Divorced	4
17-90	Female	10th	Sales	United-States	Separated	4

Jika dilihat berdasarkan atribut sensitifnya, secara semantik (makna) dari *married-civ-spouse*, *separated*, dan *divorced* memiliki maksud/arti yang sama, yakni pasangan suami-istri yang (sudah/sedang) berpisah. Sehingga atribut sensitif milik individu masih dapat ditebak berdasarkan kesamaan makna/arti dari atribut sensitif tersebut.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa algoritma yang lebih baik dalam membangun model *l-Diversity* adalah algoritma *Systematic Clustering*. Namun dari segi efektivitas, algoritma *Datafly* adalah algoritma yang lebih baik dalam membangun model *l-Diversity*. Pada model *l-Diversity* ini masih terdapat dua kelemahan, yaitu: yang pertama adalah *skewness attack* dan *similiarity attack*.

5.2 Saran

1. Pelu dilakukan penelitian dengan menggunakan algoritma lain, sehingga dapat dibandingkan tingkat *information loss* dan *running time*-nya.
2. Perlu dilakukan pada jumlah *tuple* atau *record* dan atribut sensitif yang lebih banyak, sehingga hasil penelitian yang didapatkan lebih variatif.
3. Perlu dilakukan penelitian dengan membandingkan model *l-Diversity* yang lain, seperti: *Entropy l-Diversity*, *Recursive l-Diversity*, *atau Multiple-Attribute l-Diversity*. Sehingga hasil penelitian lebih variatif.
4. Perlu dilakukan penelitian lanjutan menggunakan model yang lain untuk mengatasi *similarity attack* dan *skewness attack* yang dihasilkan oleh model l-Diversity, yakni *t-Closeness* dan Anatomi.

Daftar Pustaka:

Chourasia, Uday; Sadhwani, Divya; & Silakari, Sanjay. (2017). *Preserving Privacy during Big Data Publishing using K-Anonymity Model – A Survey*. International Journal of Advanced Research in Computer Science May-June 2017, 8(5): 801-810.

Departemen Pendidikan dan Kebudayaan. (2012). *Kamus Besar Bahasa Indonesia* (KBBI). https://www.kbbi.web.id/. Diakses tanggal 17

Fung, Benjamin C.M.; Wang, Ke; Wai-C.F., Ada; and S.Yu, Philip. (2011). *Introduction to Privacy-Preserving Data Publishing*. USA: Chapman & Hall/CRC (an imprint of Taylor & Francis Group)

Jayabalan, M., Rajendran, K., dan Rana, M.E. (2017). *A Study on k-anonymity, l-diversity, and t-closeness Techniques focusing Medical Data*. International Journal of Computer Science and Network Security 17(12): 172-177, December 2017.

Kabir, Md. E.; Wang, Hua; Bertino, Elisa; & Chi, Yunxiang. (2010). *Systematic Clustering Method for l-Diversity Model*. In: Twenty Australian Database Conference (ADC 2010), Brisbane, Australia, January 2010, 18 January 2010-22 January 2010.

Machanavajjhala, A.; Kifer, D.; Gehrke, J.; & Venkitasubramaniam, M. (2007). *l-Diversity: Privacy beyond k-Anonymity*. ACM Transactions on Knowledge Discovery from Data 1(1) Article 3 (March 2007), 52 pages.

Ridwansyah, Reza. 2017. *Perbandingan Kinerja Algoritma Systematic Clustering dan One Pass K-Means pada Model k-Anonymity Data* [skripsi]. Jakarta: Fakultas Teknik, Universitas Negeri Jakarta.

Sakina, Nasiha. 2017. *Perbandingan Kinerja Algoritma Systematic Clustering dan Greedy k-Member pada Model k-Anonymity yang Menggunakan Dua Atribut Sensitif* [skripsi]. Jakarta: Fakultas Teknik, Universitas Negeri Jakarta.

Sweeney, Latanya. (1998). *Datafly: a System for Providing Anonymity in Medical Data*. In: Lin T.Y., Qian S. (eds) Database Security XI. IFIP Advances in Information and Communication Technology. Springer, Boston, MA.

Wahid, Fathul. (2004). *Dasar-dasar Algoritma dan Pemrograman*. Yogyakarta: Penerbit ANDI.

Yu, Z.; Qian, Q.; Lin, C.Y.; and Hung, C.L. (2015). *High Performance Datafly based Anonymity Algorithm and Its L-Diversity*. International Journal of Grid and High Performance Computing, 7(3): 85-100.