

BAB IV

HASIL PENELITIAN

4.1. Deskripsi Hasil Penelitian

4.1.1 Pengumpulan Data

Data yang digunakan dalam penelitian bersumber dari UCI *Machine Learning Repository*, yaitu menggunakan *dataset Adult*. *Adult dataset* merupakan hasil sensus yang dilakukan oleh *US Census Bureau* pada tahun 1996, yang terdiri dari 14 atribut dengan enam atribut *continuous* dan delapan atribut kategorial dalam 48842 *field dataset*. Format data *Adult dataset* terdiri dari tupel yang ditulis dalam satu baris yang dipisahkan dengan tanda koma dan diakhiri dengan tanda titik.

4.1.2 Penyortiran dan Konversi Data

Format data *Adult dataset* dikonversi menjadi format *Comma-Separated Value (CSV)* menggunakan *Microsoft Excel* untuk disimpan ke dalam basis data. *Field* pada *Adult dataset* memiliki *missing attribute* yang dilambangkan dengan tanda tanya (?) sebesar 7% atau 3620 *field* pada Gambar 4.1.

```
40, Private, 121772, Assoc-voc, 11, Married-civ-spouse, Craft-repair, Husband,
Asian-Pac-Islander, Male, 0, 0, 40, ?, >50K
54, ?, 180211, Some-college, 10, Married-civ-spouse, ?, Husband, Asian-Pac-
Islander, Male, 0, 0, 60, South, >50K
31, Private, 84154, Some-college, 10, Married-civ-spouse, Sales, Husband, White,
Male, 0, 0, 38, ?, >50K
18, Private, 226956, HS-grad, 9, Never-married, Other-service, Own-child, White,
Female, 0, 0, 30, ?, <=50K
32, ?, 293936, 7th-8th, 4, Married-spouse-absent, ?, Not-in-family, White, Male,
0, 0, 40, ?, <=50K
25, ?, 200681, Some-college, 10, Never-married, ?, Own-child, White, Male, 0,
67, ?, 212759, 10th, 6, Married-civ-spouse, ?, Husband, White, Male, 0, 0, 2,
United-States, <=50K
```

Gambar 4.1 Missing Atribut Dataset Adult UCI Machine Learning

Langkah selanjutnya adalah dilakukan penyortiran pada *dataset* dengan membuang *field* yang memiliki *missing attribute* menggunakan bahasa pemrograman *Java*. Atribut pada *dataset* penelitian diberi batasan, yaitu

digolongkan menjadi atribut *Quasi-Identifier* dan *Sensitive*. Merujuk pada penelitian Traian Marius, dkk. yang berjudul *Privacy Protection: p-Sensitive k-Anonymity Property* dan penelitian Pawan R. Bhaladhare dan Devesh C. Jinwala yang berjudul *Novel Approaches for Privacy Preserving Data Mining*, atribut yang termasuk ke dalam *Quasi-Identifier* pada penelitian adalah: *Age*, *Workclass*, dan *Race*, sedangkan untuk atribut sensitifnya adalah *Marital-status*. Tabel 4.1 merupakan hasil penetapan atribut *Quasi-Identifier* dan sensitif data penelitian.

Tabel 4.1 Contoh Atribut *Quasi-Identifier* dan *Sensitive* pada Data Penelitian

<i>ID</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>	<i>Marital-status</i>
1	39	<i>State-gov</i>	<i>White</i>	<i>Never-married</i>
2	50	<i>Self-emp-not-inc</i>	<i>White</i>	<i>Married-civ-spouse</i>
3	38	<i>Private</i>	<i>White</i>	<i>Divorced</i>
4	53	<i>Private</i>	<i>Black</i>	<i>Married-civ-spouse</i>
5	28	<i>Private</i>	<i>Black</i>	<i>Married-civ-spouse</i>
6	37	<i>Private</i>	<i>White</i>	<i>Married-civ-spouse</i>
7	49	<i>Private</i>	<i>Black</i>	<i>Married-spouse-absent</i>
8	52	<i>Self-emp-not-inc</i>	<i>White</i>	<i>Married-civ-spouse</i>
9	31	<i>Private</i>	<i>White</i>	<i>Never-married</i>
10	42	<i>Private</i>	<i>White</i>	<i>Married-civ-spouse</i>

4.1.3 Clustering data dengan Systematic Clustering

Model yang diterapkan adalah model Anatomi, sedangkan algoritma yang diterapkan adalah *Systematic Clustering*. Penerapan algoritma *Systematic Clustering* bertujuan untuk memperkecil nilai *Information Loss*, karena dilakukan *clustering* berdasarkan kedekatan atau kesamaan antar atribut. Tahapan penerapan model Anatomi diawali dengan pengurutan dan *clustering dataset* pada Tabel 4.1 dengan algoritma *Systematic Clustering*. Hasil *clustering* disimpan pada tabel yang berbeda dengan tabel asli. *Dataset* penelitian diurutkan berdasarkan atribut *Age* secara *ascending* (data terkecil ke terbesar) seperti pada Tabel 4.2.

Tabel 4.2 Hasil *Clustering Dataset* Penelitian dengan *Systematic Clustering*

<i>ID</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>	<i>Marital-status</i>
195	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>
243	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>
251	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>
312	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>
345	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>
392	17	<i>Private</i>	<i>White</i>	<i>Never-married</i>

4.1.4 *Penganoniman data dengan model k-Anonymity*

Model Anatomi dalam penelitian diterapkan setelah menerapkan model *k-Anonymity* dan *p-Sensitive k-Anonymity*. Syarat untuk menerapkan model *k-Anonymity* adalah minimal nilai ($k-1$) dalam *dataset*. Nilai k yang ditetapkan dalam penelitian ini adalah $3 \leq k \leq 10$. Teknik yang digunakan untuk membentuk data menjadi model *k-Anonymity* adalah teknik Generalisasi dan Supresi sesuai *Taxonomy Tree*. Proses generalisasi dan supresi dalam penelitian menggunakan *Taxonomy Tree* yang mengacu pada Gambar 3.2 dan Gambar 3.3.

Langkah selanjutnya adalah menerapkan model *k-Anonymity* menggunakan teknik generalisasi dan supresi. Teknik tersebut dilakukan dengan dua cara, yaitu *Local Recording* pada atribut *Workclass* dan *Age*, dan *Global Recording* pada atribut *Race*. Tabel yang digunakan dalam penerapan model *k-Anonymity* mengacu pada tabel hasil *clustering* pada Tabel 4.2. Hasil penelitian menggunakan model *k-Anonymity* dapat dilihat pada Tabel 4.3.

Tabel 4.3 Hasil Penerapan Model *k-Anonymity* pada Data Penelitian

<i>ID</i>	<i>Cluster</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>	<i>Marital-status</i>
195	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
243	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
251	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
312	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
345	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
392	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>

Model yang diterapkan pada Tabel 4.3 adalah model *6-Anonymity*, yaitu dalam satu *cluster* terdapat enam *field* yang memiliki kedekatan antar atribut. Pada Tabel 4.3, atribut sensitif *Marital-status* hanya terdapat satu kategori, yaitu *Never-married*. Hal tersebut memungkinkan terjadinya *Homogeneity Attack* atau penyerangan data pada cluster yang memiliki atribut sensitif yang sama, meskipun atribut *Quasi-identifier* telah digeneralisasi, sehingga orang yang membaca Tabel 4.3 akan mengetahui pemilik dari atribut sensitif tersebut.

4.1.5 Penganoniman data dengan model *p-Sensitive k-Anonymity*

Untuk mengatasi *Homogeneity Attack*, diperlukan pengecekan keberagaman data pada atribut sensitif dalam data penelitian dengan menerapkan model *p-Sensitive k-Anonymity*. Syarat untuk mencapai model *p-Sensitive k-Anonymity* adalah data pada atribut sensitif memiliki keberagaman data $2 \leq p \leq k$. Penerapan model *p-Sensitive k-Anonymity* dalam penelitian diawali dengan penetapan nilai p yaitu $2 \leq p \leq 3$. Langkah selanjutnya adalah mengambil data pada tabel hasil model *k-Anonymity* untuk mengecek keberagaman dari jenis atribut sensitif pada Tabel 4.2. Tabel 4.4 merupakan hasil pengecekan keberagaman atribut sensitif Tabel 4.2.

Tabel 4.4 Hasil Pengecekan nilai p Model *k-Anonymity* Data Penelitian

<i>ID</i>	<i>Cluster</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>	<i>Marital-status</i>
2472	28	18	<i>Workclass</i>	<i>Race</i>	<i>Divorced</i>
7053	11	17	<i>Private</i>	<i>Race</i>	<i>Married-civ-spouse</i>
195	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
243	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
251	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
312	1	17	<i>Private</i>	<i>Race</i>	<i>Never-married</i>

Proses pengecekan dilakukan dengan melihat banyaknya atribut sensitif yang berbeda dalam satu *cluster*, mengacu pada nilai p . Jika keberagaman atribut sensitif dalam satu *cluster* belum memenuhi nilai p , maka atribut yang sama dalam satu

cluster akan ditukar dengan atribut yang berbeda yang diperoleh dari *cluster* lain sampai mencapai nilai p , seperti pada Tabel 4.4. Langkah selanjutnya adalah pengambilan data dari tabel hasil *clustering* (pada Tabel 4.2) menggunakan *ID* tiap *field*, kemudian dicocokkan dengan *ID* pada hasil penerapan model p -Sensitive k -Anonymity, dan *ID* yang cocok akan disimpan pada tabel yang berbeda.

Tabel 4.5 Hasil Generalisasi Model p -Sensitive k -Anonymity Data Penelitian

<i>ID</i>	<i>Cluster</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>	<i>Marital-status</i>
2472	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Divorced</i>
7053	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Married-civ-spouse</i>
195	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
243	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
251	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Never-married</i>
312	1	[17-18]	<i>Private</i>	<i>Race</i>	<i>Never-married</i>

Pengambilan data dari Tabel 4.2 bertujuan untuk memperkecil kenaikan nilai *Information Loss* dari model k -Anonymity menjadi model p -Sensitive k -Anonymity. Data pada Tabel 4.4 dilakukan generalisasi sesuai dengan teknik generalisasi yang dilakukan pada model k -Anonymity. Hasil generalisasi data penelitian menggunakan model p -Sensitive k -Anonymity dapat dilihat pada Tabel 4.5. Model yang diterapkan pada Tabel 4.5 adalah model 3-Sensitive 6-Anonymity, yaitu dalam satu *cluster* terdapat enam *field* yang memiliki kedekatan antar atribut dan terdapat minimal dua atau tiga jenis yang berbeda pada atribut sensitif. Pada Tabel 4.5 data diurutkan berdasarkan atribut sensitif *Marital-status* untuk mempermudah pengecekan atribut sensitif yang berbeda pada tiap *cluster*.

4.1.6 Penganoniman Data dengan Model Anatomi

Tahapan akhir penelitian adalah menerapkan model Anatomi. Langkah awal penerapan model Anatomi adalah menarik isi dari Tabel 4.4, kemudian dilakukan generalisasi dan supresi. Proses generalisasi dan supresi atribut *Quasi-Identifier*

pada model Anatomi dilakukan dengan *Local Recording*, untuk memenuhi syarat suatu atribut dikatakan sebagai *Quasi-Identifier* yaitu dalam satu *cluster* data pada tiap atribut seragam dan tidak dapat menunjukkan identitas seseorang secara eksplisit. Langkah selanjutnya adalah menyimpan data hasil generalisasi dan supresi pada tabel yang berbeda, kemudian dipecah menjadi dua tabel, yaitu *Quasi-Identifier Table (QIT)* dan *Sensitive Table (ST)*.

Isi data penelitian pada *QIT* mengikuti format atribut pada tabel sebelumnya yang terdiri dari atribut *Explicit Identifier (ID dan Cluster)* dan atribut *Quasi-Identifier (Age, Workclass, dan Race)*, namun tidak mencantumkan atribut sensitif. Atribut sensitif akan ditampilkan dalam *ST* bersama dengan atribut *Cluster* dan *Count*. Atribut *Cluster* dalam *ST* berfungsi sebagai *link* antara tabel *QIT* dengan *ST*, sedangkan *Count* merupakan banyaknya satu atribut sensitif dalam satu *cluster*. *QIT* terbentuk dari penarikan data pada tabel hasil pengecekan nilai p pada model p -*Sensitive k-Anonymity* untuk mengambil data atribut *Explicit Identifier* dan *Quasi-Identifier* dalam Tabel 4.4, kemudian disimpan pada tabel *QIT*.

Tabel 4.6 Hasil Penerapan *QIT* model Anatomi Data Penelitian

<i>ID</i>	<i>Cluster</i>	<i>Age</i>	<i>Workclass</i>	<i>Race</i>
2472	1	[17-18]	<i>Private</i>	<i>White</i>
7053	1	[17-18]	<i>Private</i>	<i>White</i>
312	1	[17-18]	<i>Private</i>	<i>White</i>
195	1	[17-18]	<i>Private</i>	<i>White</i>
243	1	[17-18]	<i>Private</i>	<i>White</i>
251	1	[17-18]	<i>Private</i>	<i>White</i>

Tabel 4.7 Hasil Penerapan *ST* Model Anatomi Data Penelitian

<i>Group-ID</i>	<i>Disease</i>	<i>Count</i>
1	<i>Divorced</i>	1
1	<i>Married-civ-spouse</i>	1
1	<i>Never-married</i>	4

Proses pembentukan *ST* dilakukan dengan mengecek atribut sensitif yang berbeda dalam satu *cluster*, kemudian disimpan dalam *ST*. Langkah terakhir dalam model Anatomi adalah mengecek banyaknya satu jenis atribut sensitif pada satu *cluster* menggunakan *query* pada basis data, kemudian hasil perhitungannya disimpan pada atribut *Count*. Hasil penerapan data penelitian menggunakan model Anatomi dapat dilihat pada Tabel 4.6 dan Tabel 4.7.

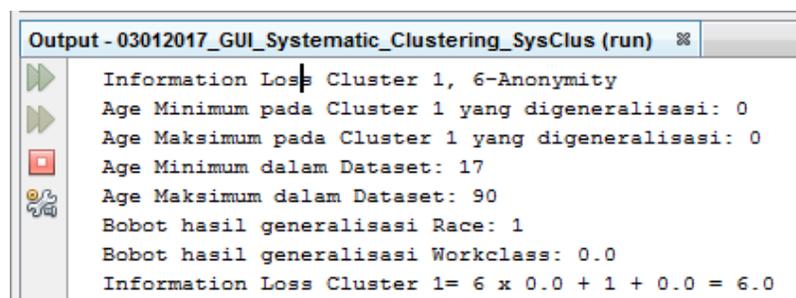
4.2. Analisis Data Penelitian

Hasil penerapan model Anatomi akan dibandingkan dengan hasil pengukuran banyaknya informasi yang hilang pada penganoniman data model *p-Sensitive k-Anonymity* dengan generalisasi dan penganoniman data dengan model Anatomi menggunakan *Information Loss Metric*. Perhitungan *Information Loss Metric* mengacu pada penelitian Ji-Won Byun, dkk. atau menggunakan formula pada persamaan (2.1) dan (2.2). Model *p-Sensitive k-Anonymity* dan model Anatomi akan diberikan dua perlakuan yaitu batasan nilai *k* dan nilai *p* yang telah ditetapkan dalam penelitian. Pada perlakuan pertama, kedua model dilakukan pengulangan sebanyak delapan kali pada model *k-Anonymity* mengikuti nilai *k* yaitu $3 \leq k \leq 10$. Perlakuan kedua yang diberikan pada kedua model tersebut adalah perulangan sebanyak dua kali pada model *p-Sensitive k-Anonymity* mengikuti nilai *p* yaitu $2 \leq p \leq 3$. Langkah awal pada teknik analisis *Information Loss* adalah menetapkan bobot pada atribut *Quasi-Identifier* kategorial (*Race* dan *Workclass*) dan numerik (*Age*), mengacu pada Gambar 3.2 dan Gambar 3.3.

Data dalam sebuah *cluster* pada atribut *Race* ditetapkan secara langsung berbobot 1, karena generalisasi yang dilakukan pada atribut tersebut tergolong *Global Recording*. Proses generalisasi pada *Quasi-Identifier Workclass* terdapat

dua tahapan generalisasi, jika digeneralisasi pada *node* pertama diberi bobot 0,5, jika digeneralisasi pada *node* tertinggi diberi bobot 1, namun jika tidak mengalami generalisasi maka berbobot 0. Pembobotan pada atribut *Age* bergantung pada selisih umur tertinggi dan terendah dalam satu *cluster* dibagi dengan selisih umur tertinggi dan terendah data secara keseluruhan, kemudian dilakukan perhitungan *Information Loss* tiap *cluster* (IL_e).

Rumus yang digunakan untuk menghitung IL_e mengacu pada persamaan (2.1). IL_e dihitung pada saat *field* terakhir pada satu *cluster* selesai dilakukan generalisasi dan supresi. Perhitungan IL_e dalam penelitian ini dilakukan bersamaan dengan proses generalisasi dan supresi dalam bahasa pemrograman Java. Perhitungan IL_e dilakukan untuk mengecek besarnya *Information Loss* yang dihasilkan oleh proses penganoniman data menggunakan teknik generalisasi dan supresi dalam satu *cluster*.



```

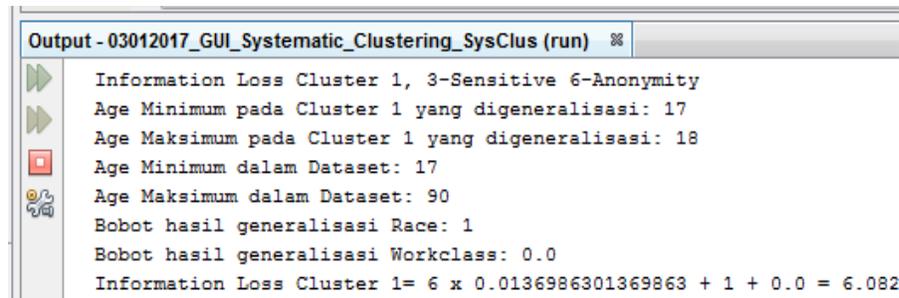
Output - 03012017_GUI_Systematic_Clustering_SysClus (run)
Information Loss Cluster 1, 6-Anonymity
Age Minimum pada Cluster 1 yang digeneralisasi: 0
Age Maksimum pada Cluster 1 yang digeneralisasi: 0
Age Minimum dalam Dataset: 17
Age Maksimum dalam Dataset: 90
Bobot hasil generalisasi Race: 1
Bobot hasil generalisasi Workclass: 0.0
Information Loss Cluster 1= 6 x 0.0 + 1 + 0.0 = 6.0

```

Gambar 4.2 Analisis *Information Loss* Tabel 4.3

Gambar 4.2 merupakan perhitungan IL_e pada *cluster* pertama pada model 6-*Anonymity*. Tahapan pertama dalam perhitungan IL_e adalah menentukan besarnya nilai tiap variabel dalam rumus tersebut yang terdapat dalam model tersebut, yaitu atribut *Age* sebagai *Quasi-Identifier* numerik, *Race* dan *Workclass* sebagai *Quasi-Identifier* kategorial. Nilai *Information Loss* pada *Age* didapat berdasarkan selisih *Age* maksimum dengan *Age* minimum dalam satu *cluster*, kemudian dibagi dengan

selisih *Age* maksimum dengan *Age* minimum dalam *dataset*. Mengacu pada Gambar 4.2, *Age* minimum dan maksimum dalam *cluster* bernilai 0 karena atribut tersebut tidak mengalami generalisasi. Nilai generalisasi *Race* adalah 1 karena mengalami generalisasi, sedangkan nilai generalisasi *Workclass* bernilai 0 karena tidak mengalami generalisasi.

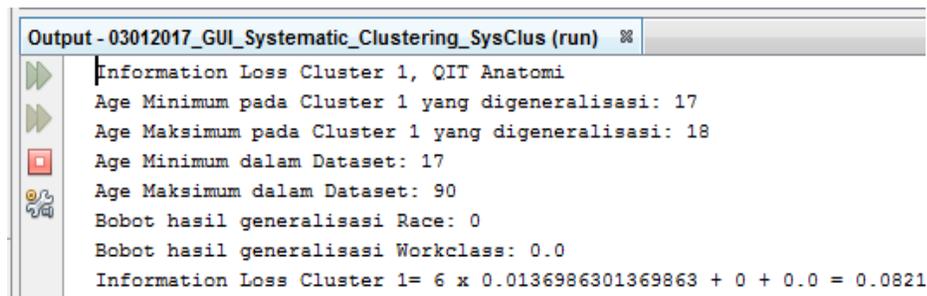


```

Output - 03012017_GUI_Systematic_Clustering_SysClus (run) %
Information Loss Cluster 1, 3-Sensitive 6-Anonymity
Age Minimum pada Cluster 1 yang digeneralisasi: 17
Age Maksimum pada Cluster 1 yang digeneralisasi: 18
Age Minimum dalam Dataset: 17
Age Maksimum dalam Dataset: 90
Bobot hasil generalisasi Race: 1
Bobot hasil generalisasi Workclass: 0.0
Information Loss Cluster 1= 6 x 0.0136986301369863 + 1 + 0.0 = 6.082

```

Gambar 4.3 Analisis *Information Loss* Tabel 4.5



```

Output - 03012017_GUI_Systematic_Clustering_SysClus (run) %
Information Loss Cluster 1, QIT Anatomi
Age Minimum pada Cluster 1 yang digeneralisasi: 17
Age Maksimum pada Cluster 1 yang digeneralisasi: 18
Age Minimum dalam Dataset: 17
Age Maksimum dalam Dataset: 90
Bobot hasil generalisasi Race: 0
Bobot hasil generalisasi Workclass: 0.0
Information Loss Cluster 1= 6 x 0.0136986301369863 + 0 + 0.0 = 0.0821

```

Gambar 4.4 Analisis *Information Loss* Tabel 4.6

Gambar 4.3 merupakan proses perhitungan IL_e pada *cluster* pertama pada model *3-Sensitive 6-Anonymity*. Atribut *Age* pada Gambar 4.3 mengalami generalisasi, karena terdapat umur yang berbeda dalam satu *cluster*, yaitu 17 dan 18, sehingga angka 17 dimasukkan ke dalam variabel *Age* minimum *cluster*, sedangkan 18 masuk ke dalam variabel *Age* maksimum *cluster* dan berlaku pada penetapan *Age* minimum *cluster* dan *Age* maksimum *cluster* pada Gambar 4.4. Nilai generalisasi *Workclass* bernilai 0 karena tidak mengalami generalisasi dan berlaku juga pada Nilai generalisasi *Workclass* pada Gambar 4.4. Nilai generalisasi *Race*

adalah 1 karena mengalami generalisasi, sedangkan pada Gambar 4.4, nilai generalisasi *Race* yang ditetapkan adalah 0 karena tidak mengalami generalisasi.

Hasil perhitungan IL_e dijumlahkan dengan IL_e lain menjadi *Information Loss Total* (IL_{AT}), mengacu pada persamaan (2.2). Hasil IL_{AT} dalam penelitian dibulatkan menjadi tiga angka dibelakang koma, untuk mempermudah penyajian hasil analisis data penelitian. Perhitungan IL_{AT} dilakukan pertama kali pada data model *k-Anonymity*, seperti pada Gambar 4.2. Perhitungan tersebut berguna sebagai patokan dalam membandingkan hasil IL_{AT} pada kedua model. Perhitungan selanjutnya dilakukan pada data model *p-Sensitive k-Anonymity*, seperti pada Gambar 4.3. Analisis perhitungan IL_{AT} yang dilakukan pada data yang menerapkan model Anatomi berbeda dengan model sebelumnya, karena perhitungan dilakukan pada tabel QIT yang berisi atribut *Quasi-Identifier* data, seperti pada Gambar 4.4. Hasil IL_{AT} kedua model tersebut dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil Perbandingan IL_{AT} Model 2-Sensitive k-Anonymity

<i>k</i>	<i>k-Anonymity</i>	<i>p-Sensitive k-Anonymity</i>	Anatomi
3	18.464,014	18.731,384	6.131,384
4	19.700,904	19.954,986	7.354,986
5	20.644,110	20.822,123	8.222,123
6	21.341,014	21.418,438	8.818,438
7	21.985,370	22.111,274	9.511,274
8	22.463,014	22.491,507	9.891,507
9	22.881,514	22.939,151	10.339,151
10	23.149,041	23.146,507	10.546,507
Rata-Rata	21.328,623	21.451.921	8.851,921

Tabel 4.8 dan Tabel 4.9 menyajikan nilai IL_{AT} hasil generalisasi ketiga model, yaitu *k-Anonymity*, *p-Sensitive k-Anonymity*, dan Anatomi, dengan perlakuan nilai $k=3 \leq k \leq 10$ dan nilai $p=2 \leq p \leq 3$. Berdasarkan Tabel 4.8 dan Tabel 4.9 secara keseluruhan dengan diberikannya perlakuan nilai *k* dan *p* yang berbeda pada model

p-Sensitive k-Anonymity dan Anatomi menunjukkan bahwa terjadi kenaikan yang signifikan. Sebagai contoh, jika diterapkan nilai $p=2$ dan $k=3$ pada model Anatomi menghasilkan nilai $IL_{AT}= 6.131,384$ mengalami kenaikan sebesar 888,349 menjadi 7.019,733 dengan nilai $p=3$ dan $k=3$. Jika ditinjau dari nilai k pada Tabel 4.8 dan Tabel 4.9, besarnya nilai k yang digunakan berpengaruh signifikan pada kenaikan IL_{AT} hasil generalisasi pada tiap model. Sebagai contoh pada Tabel 4.9, jika diterapkan nilai $k=3$ pada model Anatomi menghasilkan nilai $IL_{AT} = 7.019,733$ mengalami kenaikan sebesar 1.843,527 menjadi 8.863,260 pada saat nilai $k=4$.

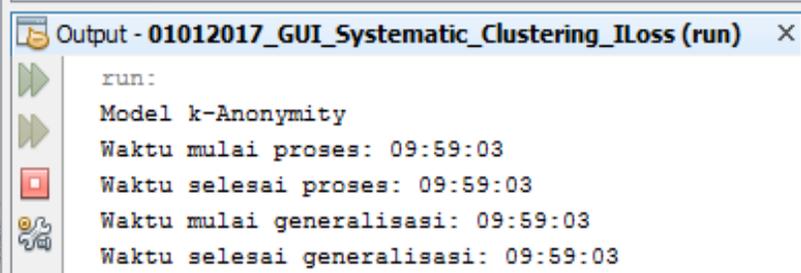
Tabel 4.9 Hasil Perbandingan IL_{AT} Model 3-Sensitive *k-Anonymity*

k	<i>k-Anonymity</i>	<i>p-Sensitive k-Anonymity</i>	Anatomi
3	18.464,014	19.619,733	7.019,733
4	19.700,904	21.463,260	8.863,260
5	20.644,110	21.848,699	9.248,699
6	21.341,014	21.992,219	9.392,219
7	21.985,370	22.376,171	9.776,171
8	22.463,014	22.782,082	10.182,082
9	22.881,514	23.042,034	10.442,034
10	23.149,041	23.332,740	10.732,740
Rata-Rata	21.328,623	22.057,117	9.457,117

Mengacu pada nilai IL_{AT} hasil generalisasi data model *k-Anonymity* pada Tabel 4.8 dan Tabel 4.9, terjadi kenaikan nilai IL_{AT} pada hasil generalisasi data model *p-Sensitive k-Anonymity*, sedangkan nilai IL_{AT} pada hasil generalisasi data model Anatomi mengalami penurunan yang signifikan, yaitu setengah dari nilai IL_{AT} hasil generalisasi data model *k-Anonymity*. Sebagai contoh, pada Tabel 4.9 nilai IL_{AT} model *k-Anonymity* saat $p=3$ dan $k=3$ adalah 18.464,014, mengalami kenaikan nilai IL_{AT} sebesar 1.155,719 menghasilkan nilai $IL_{AT} = 19.619,733$ pada hasil generalisasi data model *p-Sensitive k-Anonymity*, namun mengalami penurunan nilai IL_{AT} sebesar 11.444,281 menjadi 7.019,733 pada model Anatomi.

Mengacu pada Tabel 4.8 dan Tabel 4.9, dapat dikatakan bahwa kenaikan nilai k dan p berpengaruh pada kenaikan nilai IL_{AT} hasil generalisasi data. Hal tersebut beralasan karena nilai IL_{AT} hasil generalisasi data model p -Sensitive k -Anonymity akan mengalami kenaikan yang signifikan pada nilai p yang ditetapkan semakin besar, namun hal tersebut tidak berlaku pada kenaikan nilai IL_{AT} pada model Anatomi. Hasil generalisasi data model Anatomi dengan perlakuan nilai k dan p yang berbeda menunjukkan kenaikan nilai IL_{AT} yang fluktuatif, karena kenaikan IL_{AT} pada model Anatomi mengalami perlambatan pada nilai $k=6$.

Teknik analisis kedua yang digunakan adalah membandingkan waktu eksekusi (*Execution Time*) model p -Sensitive k -Anonymity dengan model Anatomi. Indikator yang digunakan dalam mengukur *Execution Time* adalah waktu proses dan waktu generalisasi dalam kedua model tersebut. Waktu proses diukur berdasarkan waktu yang dibutuhkan untuk mencapai model tersebut, sedangkan waktu generalisasi merupakan waktu yang dibutuhkan data hasil penerapan model tersebut untuk menjadi data anonim.



```

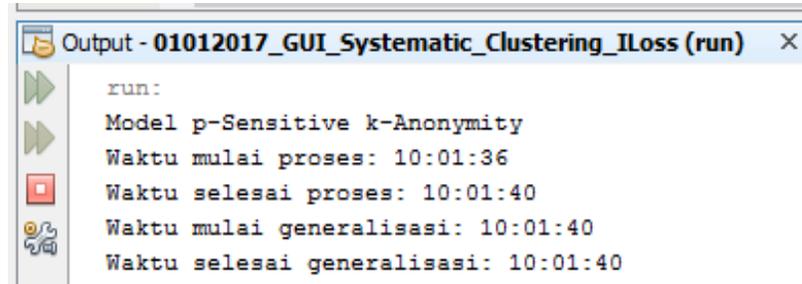
Output - 01012017_GUI_Systematic_Clustering_ILoss (run) x
run:
Model k-Anonymity
Waktu mulai proses: 09:59:03
Waktu selesai proses: 09:59:03
Waktu mulai generalisasi: 09:59:03
Waktu selesai generalisasi: 09:59:03

```

Gambar 4.5 Execution Time Model k -Anonymity

Gambar 4.5 dan Gambar 4.6 secara berturut-turut menunjukkan waktu eksekusi yang dibutuhkan dalam menerapkan model k -Anonymity dan p -Sensitive k -Anonymity sampai membentuk data anonim. Model k -Anonymity merupakan

dasar dalam pengembangan model *p-Sensitive k-Anonymity* dan model Anatomi. Hal tersebut berindikasi pada penambahan waktu eksekusi kedua model tersebut.



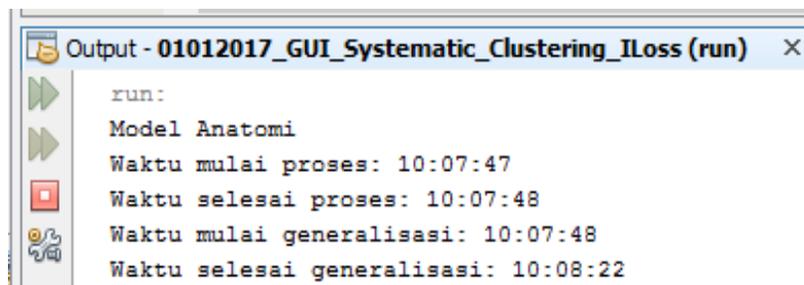
```

run:
Model p-Sensitive k-Anonymity
Waktu mulai proses: 10:01:36
Waktu selesai proses: 10:01:40
Waktu mulai generalisasi: 10:01:40
Waktu selesai generalisasi: 10:01:40

```

Gambar 4.6 Execution Time Model *p-Sensitive k-Anonymity*

Waktu proses pada Gambar 4.5 diukur berdasarkan lamanya waktu yang dibutuhkan untuk mencapai nilai k atau membentuk suatu *cluster* dengan *field* berjumlah k , sedangkan waktu generalisasi merupakan waktu yang dibutuhkan data model tersebut menjadi data anonim. Pada Gambar 4.6, waktu proses merupakan waktu yang dibutuhkan untuk mengecek keberagaman jenis atribut sensitif atau telah mencapai nilai p pada data model *k-Anonymity* setelah melalui proses penganoniman data, sedangkan waktu generalisasi merupakan waktu yang dibutuhkan data model tersebut menjadi data anonim.



```

run:
Model Anatomi
Waktu mulai proses: 10:07:47
Waktu selesai proses: 10:07:48
Waktu mulai generalisasi: 10:07:48
Waktu selesai generalisasi: 10:08:22

```

Gambar 4.7 Execution Time Model Anatomi

Berbeda halnya dengan Gambar 4.5 dan Gambar 4.6, waktu generalisasi pada Gambar 4.7 dihitung berdasarkan waktu yang dibutuhkan untuk mengubah tabel hasil pengecekan keberagaman data pada model *p-Sensitive k-Anonymity* menjadi tabel QIT, sedangkan waktu proses merupakan waktu yang dibutuhkan untuk

mengecek jenis pada atribut sensitif dalam satu *cluster* pada model *p-Sensitive k-Anonymity* serta menghitung kemunculan satu jenis atribut dalam *cluster* dan disimpan dalam tabel ST.

Perhitungan waktu proses dan waktu generalisasi pada ketiga model menggunakan bahasa pemrograman Java. Perhitungan waktu proses didapat dari selisih waktu yang diperoleh dari waktu selesai proses dengan waktu mulai proses, sedangkan waktu generalisasi merupakan selisih waktu yang diperoleh dari waktu selesai generalisasi dengan waktu mulai generalisasi.

Tabel 4.10 Perbandingan *Execution Time* Model *p-Sensitive k-Anonymity* dan Model Anatomi

Model	Waktu Proses	Waktu Generalisasi
<i>p-Sensitive k-Anonymity</i>	4	1
Anatomi	34	1

Hasil *Execution Time* yang akan dianalisis pada penelitian adalah model *p-Sensitive k-Anonymity* dengan model Anatomi yang dapat dilihat pada Tabel 4.10. Tabel 4.10 menunjukkan perbedaan *Execution Time* antara model *p-Sensitive k-Anonymity* dengan model Anatomi dalam satuan detik, namun perhitungan tersebut tidak termasuk waktu penyimpanan hasil penerapan kedua model dalam basis data. Berdasarkan Tabel 4.10, waktu proses pada model *p-Sensitive k-Anonymity* lebih cepat 30 detik dibandingkan dengan waktu proses model Anatomi.

Jika ditinjau dari waktu generalisasi, model *p-Sensitive k-Anonymity* sama dengan waktu generalisasi model Anatomi. Apabila waktu proses dan waktu generalisasi pada kedua model digabungkan, maka didapatkan *Execution Time* dari model *p-Sensitive k-Anonymity* lebih cepat 30 menit terhadap model Anatomi, sehingga dibutuhkan waktu yang lebih lama untuk menerapkan model Anatomi.

4.3. Pembahasan

Berdasarkan analisis hasil penelitian dengan mengukur *Information Loss* pada model penganoniman data dengan proses generalisasi model *k-Anonymity*, *p-Sensitive k-Anonymity*, dan Anatomi didapatkan kenaikan nilai k dan p yang signifikan terhadap kenaikan nilai IL_{AT} . Hal tersebut disebabkan oleh penerapan algoritma *Systematic Clustering* dengan mengurutkan data berdasarkan kedekatan atau kesamaan yang dimiliki antar data tiap atribut.

Mengacu pada persamaan (2.1), diketahui bahwa variabel yang berpengaruh tinggi terhadap kenaikan nilai IL_{AT} pada generalisasi data adalah banyaknya *field* dalam satu *cluster* (k) yang dilambangkan dengan e . Kenaikan IL_{AT} disebabkan adanya perkalian antara e dengan variabel generalisasi lain pada persamaan (2.1). Pada model *k-Anonymity*, e mengacu pada nilai k ; sedangkan pada model *p-Sensitive k-Anonymity* e mengacu pada nilai k dan p (keberagaman atribut sensitif).

Merujuk pada hasil penelitian Traian Marius Truta dan Bindu Vinay, yaitu semakin besar nilai k dan p yang ditetapkan, maka semakin kecil kemungkinan dilakukan *attribute disclosure* atau pengungkapan kembali data pada atribut *Quasi-Identifier* yang telah dianonimkan, berimplikasi pada semakin tinggi tingkat privasi data yang dihasilkan. Jika nilai k dan p diperkecil untuk menghasilkan *Information Loss* yang rendah akan mempengaruhi penurunan tingkat privasi data.

Jika ditinjau dari nilai IL_{AT} model *k-Anonymity*, terjadi kenaikan yang signifikan pada nilai IL_{AT} model *p-Sensitive k-Anonymity*, karena terdapat proses pengecekan keragaman atribut sensitif dalam satu cluster (nilai p). Hal tersebut menyebabkan pertukaran *field* antar *cluster* untuk memenuhi nilai p , sehingga didalam *cluster* terdapat beberapa data dalam atribut memiliki perbedaan yang

memicu kenaikan nilai IL_{AT} . Berbeda dengan model *p-Sensitive k-Anonymity*, kenaikan nilai IL_{AT} yang diperoleh dari penerapan model Anatomi bersifat fluktuatif, artinya bahwa kenaikan nilai k dan p yang diterapkan dalam model tersebut tidak diikuti dengan besarnya kenaikan yang signifikan pada nilai IL_{AT} .

Model Anatomi menggunakan teknik pemecahan tabel data menjadi dua, yaitu menjadi tabel *QIT* (berisi *Explicit Identifier* dan *Quasi-Identifier*) dan *ST* (berisi data sensitif). Teknik generalisasi yang digunakan dalam model Anatomi adalah *Local Recording* (generalisasi pada atribut dalam satu *cluster*) pada atribut tabel *QIT*. Jika nilai k dan p diperbesar dan dilakukan penganoniman data menggunakan model Anatomi, maka kenaikan *Information Loss* pada data semakin kecil serta menghasilkan tingkat privasi data yang tinggi.

Untuk mendapatkan nilai *Information Loss* yang minimum dan privasi data tingkat tinggi dapat diterapkan model Anatomi dan algoritma *Systematic Clustering* dalam menganonimkan data hasil pengecekan keberagaman atribut sensitif model *p-Sensitive k-Anonymity*, karena besaran nilai k dan p yang ditetapkan tidak mempengaruhi besarnya kenaikan nilai *Information Loss* yang dihasilkan dan menghasilkan privasi data tingkat tinggi.

Berdasarkan hasil analisis *Execution Time*, waktu generalisasi pada model Anatomi sama dengan model *p-Sensitive k-Anonymity*, karena proses penganoniman yang dilakukan dengan teknik generalisasi dan supresi. Jika ditinjau dari waktu proses yang dibutuhkan model Anatomi lebih lama dibandingkan model *p-Sensitive k-Anonymity*, karena tahapan proses dalam model Anatomi memecah tabel data menjadi dua, yaitu *QIT* dan *ST*, sedangkan pada model *p-Sensitive k-Anonymity* terdapat proses *sorting* pada atribut sensitif untuk mencapai nilai p .

Diperlukan waktu yang lebih lama untuk menerapkan model Anatomi dibandingkan dengan penerapan model *p-Sensitive k-Anonymity*. Perlambatan waktu yang terjadi pada proses penerapan model *k-Anonymity* menjadi model Anatomi disebabkan oleh pemecahan tabel QIT menjadi tabel ST. Untuk membentuk tabel ST diperlukan penarikan data secara langsung dari basis data untuk mengecek jenis atribut sensitif dalam satu *cluster* dan jumlah satu jenis atribut sensitif dalam *cluster*.

Hasil yang diperoleh dalam penelitian belum mencapai hasil yang maksimum, karena terdapat beberapa keterbatasan, yaitu data anonim yang dihasilkan model *p-Sensitive k-Anonymity* dengan nilai $k=3$ dan $k=4$ masih ditemukan keberagaman atribut sensitif $p=1$ atau belum beragam pada beberapa *cluster*. Hal tersebut terjadi karena pada proses penyortiran data dilakukan secara acak, yang mengakibatkan proporsi jumlah satu jenis data pada atribut sensitif tidak seimbang. Diperlukan teknik penyortiran data menggunakan metode *sampling* untuk mengecek proporsi jumlah satu jenis data pada atribut sensitif untuk mendapatkan nilai p yang lebih tinggi.

4.4. Aplikasi Hasil Penelitian

Hasil penelitian dapat digunakan sebagai panduan untuk melakukan publikasi data. Publikasi data dapat dilakukan setelah data yang memiliki atribut bersifat sensitif dilakukan penganoniman tanpa mengurangi inti dari informasi yang terdapat dalam data tersebut. Penganoniman data yang dilakukan dalam penelitian dapat diaplikasikan menjadi sebuah program menggunakan berbagai bahasa pemrograman untuk memproses data yang akan dipublikasikan menjadi data

anonim secara otomatis. Beberapa instansi atau lembaga yang memerlukan penganoniman data untuk dipublikasikan, yaitu:

1. Rumah Sakit

Sebagian besar data pasien yang perlu dipublikasikan ke masyarakat tergolong data yang bersifat sensitif. Jika data tersebut akan dipublikasikan dapat menimbulkan kekecewaan pasien karena data pribadinya disebar, diperlukan teknik penganoniman data untuk meminimalisir peluang serangan pada data. Manfaat lain yang dapat diperoleh adalah data anonim yang dihasilkan dapat dimengerti oleh khalayak umum karena makna dari data tersebut masih jelas dan mencegah identitas pemilik data tidak akan mudah untuk diketahui.

2. Kepolisian

Lembaga kepolisian dianggap perlu mempublikasikan data untuk diketahui khalayak umum sebagai contoh daftar kejahatan yang terjadi dalam kurun waktu tertentu. Publikasi data bertujuan untuk mengedukasi masyarakat untuk waspada terhadap tindakan kejahatan yang terjadi di lingkungan tempat tinggalnya. Dengan diterapkannya hasil penelitian ini, kepolisian dapat meminimalkan kebocoran informasi yang bersifat sensitif.

3. Asuransi

Perusahaan asuransi dianggap perlu untuk mempublikasikan datanya agar diketahui khalayak umum seperti asuransi kesehatan. Data yang perlu dipublikasikan adalah data mengenai data klaim anggota asuransi beserta riwayat penyakit yang ditanggung oleh asuransi dalam kurun waktu tertentu. Hal tersebut penting sebagai bentuk pertanggung jawaban perusahaan terhadap anggotanya.