

PRIVACY PRESERVING DATA MINING METODE K-ANONYMITY DENGAN TEKNIK CLUSTERING

APSARI AYUSYA CANTIKA—2016730012

1 Deskripsi

Seiring perkembangan zaman, teknologi semakin maju dan berkembang. Hal ini membuat data yang dirilis untuk kebutuhan *data mining* semakin banyak. Data yang akan dirilis biasanya memiliki privasi. Privasi merupakan informasi yang dapat dipakai untuk membedakan identitas satu individu dengan individu lainnya. Contohnya pada data penduduk, data seperti NIK dan nama lengkap merupakan privasi. Jika data tersebut langsung dirilis untuk proses *data mining*, maka privasi akan terbuka dan tidak terlindungi. Privasi harus dilindungi agar tidak disalahgunakan oleh pihak tidak bertanggungjawab.

Perlindungan privasi dapat dilakukan dengan *privacy preserving data mining*. *Privacy preserving data mining* merupakan bagian dari *data mining* yang bertanggungjawab atas perlindungan privasi dalam *data mining*. Metode untuk *privacy preserving data mining* dapat diklasifikasikan menjadi empat kategori, yaitu metode randomisasi, metode anonimisasi, *distributed privacy preservation*, dan menurunkan efektivitas dari hasil *data mining*.

Metode yang akan digunakan adalah metode anonimisasi *k-anonymity*. Metode ini membagi data ke dalam kelompok berukuran k dan membuat data dalam suatu kelompok tidak bisa dibedakan dengan $k - 1$ data lainnya di kelompok yang sama. *K-anonymity* dapat dilakukan dengan beberapa teknik seperti generalisasi, semantik, *clustering*, dan lain-lain.

K-anonymity dengan teknik *clustering* mengimplementasi algoritma *clustering* untuk menganonimisasi. Teknik *clustering* akan membagi data ke dalam beberapa *cluster*. Data yang telah terbagi akan digeneralisasi. Kemudian, proses ini akan menghasilkan data yang telah teranonimisasi.

Algoritma *clustering* yang akan digunakan adalah *One Time Pass k-Means* (OKA) dan *Clustering using Representative* (CURE). Algoritma OKA merupakan improvisasi Algoritma k-Means. Pada Algoritma OKA, perulangan hanya dilakukan satu kali. Algoritma OKA terdiri dari dua tahapan, yaitu *clustering* dan *adjustment*. Sedangkan Algoritma CURE menggunakan representatif data. Algoritma CURE mengambil sebagian data sebagai representatif untuk membuat *cluster*.

Pada skripsi ini, akan dibuat sebuah perangkat lunak yang dapat menjaga privasi data dengan metode *k-anonymity* menggunakan teknik *clustering*. Perangkat lunak akan diimplementasikan menggunakan Algoritma *One Time Pass K-means* (OKA) dan *Clustering using Representative* (CURE).

2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada skripsi ini adalah:

- Bagaimana cara kerja Algoritma OKA dan CURE untuk anonimisasi?
- Bagaimana cara mengimplementasikan Algoritma OKA dan CURE untuk anonimisasi?
- Bagaimana cara mengukur performa Algoritma OKA dan CURE untuk anonimisasi?

3 Tujuan

Tujuan yang ingin dicapai dari skripsi ini:

- Mempelajari Algoritma OKA dan CURE
- Membangun perangkat lunak yang mengimplementasikan Algoritma OKA dan CURE
- Mengukur performa Algoritma OKA dan CURE dengan teknik *Data Mining*

4 Deskripsi Perangkat Lunak

Perangkat lunak akhir yang akan dibuat memiliki fitur minimal sebagai berikut:

- Pengguna dapat memasukan data yang akan dianonimisasi
- Perangkat lunak dapat menganonimisasi data
- Pengguna dapat mengatur nilai k yang merupakan jumlah kelompok data
- Pengguna dapat melihat hasil anonimisasi

5 Detail Pengerjaan Skripsi

Bagian-bagian pengerjaan skripsi ini adalah sebagai berikut :

1. Melakukan studi literatur mengenai *privacy preserving*, metode *k-Anonymity* dengan teknik *clustering*, Algoritma OKA, Algoritma CURE, dan teknik *Data Mining*
2. Analisis masalah perangkat lunak yang akan dibangun
3. Merancang perangkat lunak yang akan dibangun
4. Membangun perangkat lunak yang mengimplementasikan Algoritma OKA dan CURE
5. Melakukan pengujian fungsional dan eksperimental
6. Analisis hasil pengujian dan mengambil kesimpulan
7. Menulis dokumen skripsi

6 Rencana Kerja

Rincian capaian yang direncanakan di Skripsi 1 adalah sebagai berikut:

1. Melakukan studi literatur mengenai *privacy preserving*, metode *k-Anonymity* dengan teknik *clustering*, Algoritma OKA, Algoritma CURE, dan teknik *Data Mining*
2. Analisis masalah *privacy preserving data mining* metode *k-Anonymity* dengan teknik *clustering*
3. Menulis dokumen skripsi

Sedangkan yang akan diselesaikan di Skripsi 2 adalah sebagai berikut:

1. Merancang perangkat lunak yang akan dibangun
2. Membangun perangkat lunak yang mengimplementasikan Algoritma OKA dan CURE
3. Melakukan pengujian fungsional dan eksperimental
4. Analisis hasil pengujian dan mengambil kesimpulan
5. Menulis dokumen skripsi

Bandung, 27/08/2019

Apsari Ayusya Cantika

Menyetujui,

Nama: _____
Pembimbing Tunggal