Privacy Preserving Data Mining dengan Metode Randomization

Chris Eldon–2016730073

1 Deskripsi

Penambangan data semakin gencar dilakukan bahkan terus menerus dilakukan demi mendapat keuntungan berupa informasi, oleh sebab itu banyak data-data orang lain yang dimiliki oleh pihak lain. Tetapi karena banyaknya data-data yang digunakan atau diambil dari banyak orang, banyak juga privasi seseorang yang terdapat pada data tersebut. Hal ini bisa berbahaya bagi orang tersebut karena privasinya bisa saja disalahgunakan untuk mencelakakan orang tersebut.

Dengan semakin banyaknya penambangan data yang dilakukan dan data-data yang digunakan juga semakin banyak, semakin banyak juga privasi-privasi di dalam data yang digunakan untuk penambangan data. Akhirnya data-data privasi yang sensitif sangat berpotensi bocor kepada orang yang tidak bertanggung jawab dan disalahgunakan. Oleh karena itu perlu adanya suatu cara untuk mencegah kebocoran privasi pada proses penambangan data, menjaga privasi pada data-data tersebut. Istilah untuk hal tersebut adalah *Privacy Preserving Data Mining*.

Salah satu cara untuk melakukan *Privacy Preserving Data Mining* adalah dengan memodifikasi data yang ada sebelum diberikan kepada pihak lain. Ada macam-macam teknik dan algoritma yang bertujuan memodifikasi data untuk *Privacy Preserving Data Mining* yang bisa dibagi menjadi dua jenis yaitu *Perturbation Approach* dan *Anonymization Approach*.

Perturbation Approach adalah pendekatan untuk *Privacy Preserving Data Mining* dengan cara mengacaukan data yang ada, tetapi hasil data yang dikacaukan masih tetap bisa dilakukan penambangan data. *Perturbation Approach* bisa dibagi menjadi dua jenis yaitu *Value-based Perturbation Techniques* dan *Multi-Dimensional Perturbation*. *Value-based Perturbation Techniques* adalah teknik yang bekerja dengan cara menyisipkan *random noise* pada data. *Value-based Perturbation Techniques* bisa dibagi menjadi dua jenis lagi yaitu *Random Noise Addition* dan *Randomized Response*

Pada skripsi ini, akan dibuat sebuah perangkat lunak yang dapat memproses data-data yang akan ditambang menjadi data-data yang telah dikacaukan dengan metode *Randomization* sehingga tidak mengandung privasi tetapi masih dapat digunakan untuk dilakukan penambangan data.

Dari berbagai macam teknik dengan metode *Randomization* yang ada, dipilih satu teknik yaitu *Random Noise Addition* untuk diimplementasikan menjadi perangkat lunak.

2 Rumusan Masalah

- Bagaimana cara kerja dari teknik *Random Noise Addition* untuk *Privacy Preserving Data Mining*?
- Bagaimana implementasi dari teknik *Random Noise Addition* pada perangkat lunak?
- Apakah teknik *Random Noise Addition* bagus dalam arti seimbang antara *Privacy Loss* dengan *Information Loss*?

3 Tujuan

- Mempelajari cara kerja dari teknik *Random Noise Addition* untuk *Privacy Preserving Data Mining*
- Mengimplementasikan teknik *Random Noise Addition* pada perangkat lunak
- Menganalisa *Privacy Loss* dan *Information Loss* yang dihasilkan dari perangkat lunak yang dibuat dengan teknik *Random Noise Addition*

4 Deskripsi Perangkat Lunak

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

- Pengguna dapat memasukkan data mentah yang ingin dilakukan penambangan data (input)
- Pengguna dapat mengatur berbagai pengaturan untuk teknik *Random Noise Addition*
- Pengguna dapat mengatur jalannya proses perangkat lunak: memulai proses teknik *Random Noise Addition* dan menghentikan perangkat lunak di tengah proses
- Pengguna dapat mendapatkan data yang telah diproses dengan teknik *Random Noise Addition* (output)

5 Detail Pengerjaan Skripsi

Bagian-bagian pekerjaan skripsi ini adalah sebagai berikut :

1. Mempelajari dasar-dasar privasi data
2. Mempelajari teknik penambangan data yang akan digunakan
3. Mempelajari teknik *Random Noise Addition* untuk *Privacy Preserving Data Mining*
4. Membangun perangkat lunak yang mengimplementasikan teknik *Random Noise Addition*
5. Memodifikasi algoritma penambangan data *distribution-based* yang dipakai agar bisa menggunakan data yang telah dimodifikasi oleh teknik *Random Noise Addition*
6. Menguji perangkat lunak dengan menggunakan real data
7. Menerapkan teknik penambangan data untuk menganalisis hasil dari metode *Randomization* pada data
8. Melakukan analisa terhadap hasil pengujian
9. Menulis dokumen skripsi

6 Rencana Kerja

Rincian capaian yang direncanakan di Skripsi 1 adalah sebagai berikut:

1. Mempelajari dasar-dasar privasi data
2. Mempelajari teknik penambangan data yang akan digunakan

3. Mempelajari teknik *Random Noise Addition* untuk *Privacy Preserving Data Mining*
4. Menulis dokumen skripsi bab 1, 2, dan 3

Sedangkan yang akan diselesaikan di Skripsi 2 adalah sebagai berikut:

1. Membangun perangkat lunak yang mengimplementasikan teknik *Random Noise Addition*
2. Memodifikasi algoritma penambangan data *distribution-based* yang dipakai agar bisa menggunakan data yang telah dimodifikasi oleh teknik *Random Noise Addition*
3. Menguji perangkat lunak dengan menggunakan real data
4. Menerapkan teknik penambangan data untuk menganalisis hasil dari metode *Randomization* pada data
5. Melakukan analisa terhadap hasil pengujian
6. Menulis dokumen skripsi bab 4, 5, dan 6

Bandung, 23/08/2019

Chris Eldon

Menyetujui,

Nama: _____

Pembimbing Tunggal