Privacy Preserving Data Mining dengan Metode Randomization

Chris Eldon–2016730073

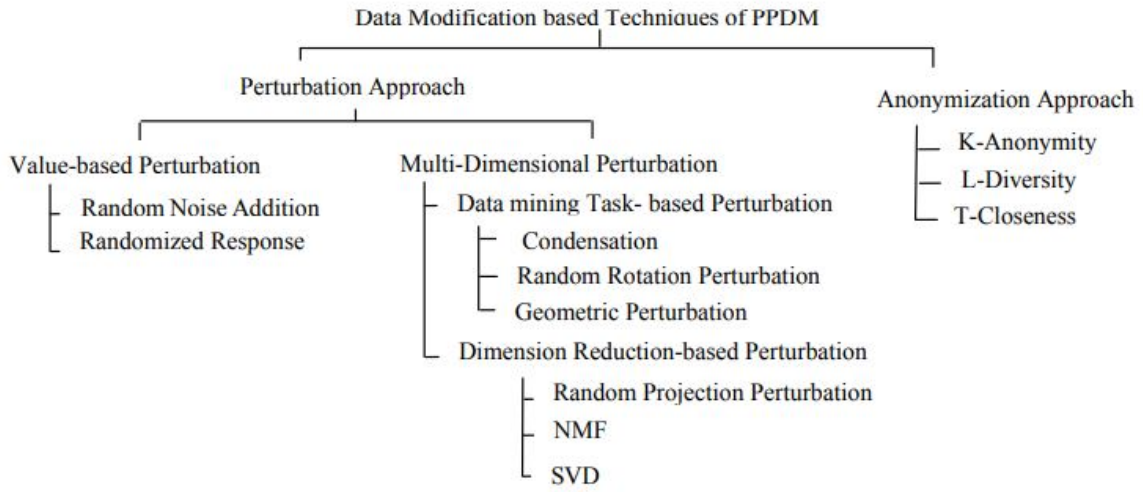
1 Deskripsi

Dengan semakin banyaknya penambangan data yang dilakukan dan data-data yang digunakan juga semakin banyak, semakin banyak juga privasi-privasi di dalam data tersebut yang tersebar kepada pihak yang melakukan penambangan data. Data-data privasi yang sensitif sangat berpotensi bocor kepada pihak 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 ditambang. *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. Sedangkan terdapat dua jenis teknik Multi-Dimensional Perturbation yaitu *Data mining Task-based Perturbation* dan *Dimension Reduction-based Perturbation*. *Data mining Task-based Perturbation* adalah teknik yang bekerja dengan cara memodifikasi data sehingga properti yang bertahan pada data yang telah dimodifikasi spesifik hanya properti yang digunakan oleh suatu teknik penambangan data tertentu.

Dari berbagai macam teknik memodifikasi data untuk *Privacy Preserving Data Mining* yang dapat dilihat pada Gambar 1, terdapat empat teknik yang menggunakan metode *Randomization* yaitu *Random Noise Addition*, *Randomized Response*, *Random Rotation Perturbation*, dan *Random Projection Perturbation*.



Gambar 1: Berbagai macam teknik memodifikasi data untuk *Privacy Preserving Data Mining*

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

Dari berbagai macam teknik dengan metode *Randomization* yang ada, dipilih dua buah teknik yaitu *Random Noise Addition* dan *Random Rotation Perturbation* untuk diimplementasikan pada perangkat lunak.

2 Rumusan Masalah

- Bagaimana cara kerja dari teknik *Random Noise Addition* dan *Random Rotation Perturbation* untuk *Privacy Preserving Data Mining*?
- Bagaimana implementasi dari teknik *Random Noise Addition* dan *Random Rotation Perturbation* pada perangkat lunak?
- Bagaimana perbandingan antara hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*?

3 Tujuan

- Mempelajari cara kerja dari teknik *Random Noise Addition* dan *Random Rotation Perturbation* untuk *Privacy Preserving Data Mining*
- Mengimplementasikan teknik *Random Noise Addition* dan *Random Rotation Perturbation* pada perangkat lunak
- Melakukan analisis dan pengujian untuk membandingkan dan mengukur hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*

4 Deskripsi Perangkat Lunak

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

- Pengguna dapat memasukkan data mentah yang akan ditambang (input)
- Pengguna dapat memilih teknik mana yang ingin digunakan antara teknik *Random Noise Addition* atau *Random Rotation Perturbation*
- Pengguna dapat mengatur berbagai pengaturan seperti parameter untuk teknik *Random Noise Addition* maupun *Random Rotation Perturbation*
- Perangkat lunak dapat memodifikasi data mentah yang dimasukkan dengan menggunakan teknik *Random Noise Addition* dan *Random Rotation Perturbation*
- Pengguna dapat memperoleh data yang telah diproses dengan teknik *Random Noise Addition* atau *Random Rotation Perturbation* (output)

5 Detail Pengerjaan Skripsi

Bagian-bagian pekerjaan skripsi ini adalah sebagai berikut :

1. Mempelajari dasar-dasar privasi data
2. Mempelajari teknik *Random Noise Addition* dan *Random Rotation Perturbation* untuk *Privacy Preserving Data Mining*
3. Mempelajari teknik penambangan data yang akan digunakan
4. Melakukan analisis terhadap teknik *Random Noise Addition* dan *Random Rotation Perturbation* serta bagaimana penerapannya dengan teknik penambangan data yang akan digunakan
5. Melakukan perancangan perangkat lunak yang mengimplementasikan teknik *Random Noise Addition* dan *Random Rotation Perturbation*
6. Membangun perangkat lunak yang mengimplementasikan teknik *Random Noise Addition* dan *Random Rotation Perturbation*
7. Menguji perangkat lunak secara fungsional dan eksperimental dengan menggunakan *real* data
8. Menerapkan teknik penambangan data terhadap data yang telah diproses untuk menganalisis hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*
9. Melakukan analisis dan pengujian untuk membandingkan dan mengukur hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*
10. 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 *Random Noise Addition* dan *Random Rotation Perturbation* untuk *Privacy Preserving Data Mining*
3. Mempelajari teknik penambangan data yang akan digunakan
4. Melakukan analisis terhadap teknik *Random Noise Addition* dan *Random Rotation Perturbation* serta bagaimana penerapannya dengan teknik penambangan data yang akan digunakan
5. Menulis dokumen skripsi bab 1, 2, dan 3

Sedangkan yang akan diselesaikan di Skripsi 2 adalah sebagai berikut:

1. Melakukan perancangan perangkat lunak yang mengimplementasikan teknik *Random Noise Addition* dan *Random Rotation Perturbation*
2. Membangun perangkat lunak yang mengimplementasikan teknik *Random Noise Addition* dan *Random Rotation Perturbation*
3. Menguji perangkat lunak secara fungsional dan eksperimental dengan menggunakan *real* data

4. Menerapkan teknik penambangan data terhadap data yang telah diproses untuk menganalisis hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*
5. Melakukan analisis dan pengujian untuk membandingkan dan mengukur hasil dari teknik *Random Noise Addition* dan *Random Rotation Perturbation*
6. Menulis dokumen skripsi bab 4, 5, dan 6

Bandung, 27/08/2019

Chris Eldon

Menyetujui,

Nama: _____
Pembimbing Tunggal