

SKRIPSI

**ANALISIS TEKNIK *RANDOM ROTATION PERTURBATION*
DAN *RANDOM PROJECTION PERTURBATION* DALAM
MENGACAK DATA UNTUK PENAMBANGAN DATA**



Chris Eldon

NPM: 2016730073

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2020**

UNDERGRADUATE THESIS

**ANALYSIS OF RANDOM ROTATION PERTURBATION AND
RANDOM PROJECTION PERTURBATION TECHNIQUES IN
RANDOMIZING DATA FOR DATA MINING**



Chris Eldon

NPM: 2016730073

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2020**

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

**ANALISIS TEKNIK *RANDOM ROTATION PERTURBATION* DAN
RANDOM PROJECTION PERTURBATION DALAM MENGACAK DATA
UNTUK PENAMBANGAN DATA**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 12 Juni 2020



Chris Eldon
NPM: 2016730073

LEMBAR PENGESAHAN

ANALISIS TEKNIK *RANDOM ROTATION PERTURBATION* DAN *RANDOM PROJECTION PERTURBATION* DALAM MENGACAK DATA UNTUK PENAMBANGAN DATA

Chris Eldon

NPM: 2016730073

Bandung, 12 Juni 2020

Menyetujui,

Pembimbing

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

Anggota Tim Penguji

Rosa De Lima, M.T.

Luciana Abednego, M.T.

Mengetahui,

Ketua Program Studi

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

***ANALISIS TEKNIK *RANDOM ROTATION PERTURBATION* DAN
RANDOM PROJECTION PERTURBATION DALAM MENGACAK DATA
UNTUK PENAMBANGAN DATA***

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 12 Juni 2020

Chris Eldon
NPM: 2016730073

ABSTRAK

Pada proses penambangan data, data yang digunakan seringkali diberikan kepada pihak lain dan ada kemungkinan privasi di dalam data tersebut tersebar kepada pihak yang tidak berhak. Data privasi tersebut dapat tersebar kepada pihak yang tidak bertanggung jawab dan disalahgunakan. Dalam menghindari hal tersebut, *privacy-preserving data mining* perlu dilakukan. Privasi dapat diartikan sebagai sebuah informasi personal seseorang yang dapat mengidentifikasi sesuatu hal pada orang tersebut. Salah satu cara untuk melakukan *privacy-preserving data mining* adalah mengacak data menggunakan metode *Randomization*. Metode *Randomization* bekerja dengan cara mengacak data tetapi data tersebut masih dapat digunakan untuk penambangan data. Pada penelitian ini dibangun sebuah perangkat lunak yang mengimplementasikan 2 buah teknik yang menggunakan metode *Randomization* yaitu teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*.

Pengujian dilakukan dengan menerapkan penambangan data klasifikasi dengan algoritma *k-nearest neighbors* dan penambangan data *clustering* dengan algoritma *k-means* masing-masing untuk menghitung akurasi model dan kemiripan hasil *cluster*. Berdasarkan hasil pengujian, model penambangan data yang dilatih dengan *dataset* asli dan *dataset* yang telah diacak dengan teknik *Random Rotation Perturbation* atau *Random Projection Perturbation* memiliki kualitas yang sama atau sangat mirip. Kedua teknik tersebut hanya dapat digunakan untuk data yang bersifat numerik dan khususnya untuk teknik *Random Projection Perturbation* hanya dapat digunakan untuk data yang memenuhi syarat teknik tersebut yaitu jumlah fitur pada data harus cukup banyak.

Kata-kata kunci: Privasi, *privacy-preserving data mining*, *Randomization*, *Random Rotation Perturbation*, *Random Projection Perturbation*, penambangan data, klasifikasi, *clustering*, *k-nearest neighbors*, *k-means*

ABSTRACT

In the data mining process, the data used is often given to other parties and there is the possibility of privacy in the data being spread to unauthorized parties. The privacy data can be spread to irresponsible and misused parties. To avoid this, privacy-preserving data mining needs to be done. Privacy can be interpreted as a person's personal information that can identify something about that person. One way to do privacy-preserving data mining is to randomize data using the Randomization method. The Randomization method works by randomizing data but the data can still be used for data mining. In this research, a software will be built that implements 2 techniques that use the Randomization method, namely the Random Rotation Perturbation technique and the Random Projection Perturbation.

The test will be carried out by applying classification data mining with the k-nearest neighbors algorithm and clustering data mining with the k-means algorithm respectively to calculate the accuracy of the model and the similarity of the cluster results. Based on the test results, the data mining model that is trained with the original dataset and the dataset that has been randomized with the Random Rotation Perturbation technique or Random Projection Perturbation has the same or very similar quality. Both of these techniques can only be used for numerical data and specifically for the Random Projection Perturbation technique can only be used for data that meet the technical requirement that is the number of features in the data must be quite a lot.

Keywords: Privacy, privacy-preserving data mining, Randomization, Random Rotation Perturbation, Random Projection Perturbation, data mining, classification, clustering, k-nearest neighbors, k-means

*Dipersembahkan untuk Tuhan YME, keluarga tercinta, Ibu
Mariskha sebagai dosen pembimbing, teman-teman yang telah
membantu dan memberi semangat dalam penyusunan skripsi ini,
dan diri sendiri*

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas karunia-Nya, penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Analisis Teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* dalam Mengacak Data untuk Penambangan Data". Selama penyusunan skripsi ini, penulis menghadapi banyak kendala dan berbagai masalah. Penulis menyadari bahwa penyusunan skripsi ini juga tidak terlepas dari bantuan berbagai pihak, baik langsung maupun tidak langsung. Secara khusus, penulis ingin berterima kasih kepada:

1. Keluarga yang selalu memberikan dukungan kepada penulis baik berupa doa atau dukungan mental serta materiil.
2. Ibu Mariskha Tri Adithia, S.Si., M.Sc., PDEng. selaku dosen pembimbing yang telah membimbing penulis dan memberikan dukungan maupun bantuan kepada penulis dalam proses penyusunan skripsi ini.
3. Ibu Rosa De Lima, M.Kom. dan Ibu Luciana Abednego, M.T. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga penelitian ini menjadi lebih baik.
4. Teman-teman penulis di jurusan Teknik Informatika UNPAR angkatan 2016 yang telah menemani penulis dalam menyelesaikan perkuliahan dari awal semester sampai akhir semester penulis selesai menyusun skripsi ini serta teman-teman di luar jurusan yang tidak dapat disebutkan satu persatu.
5. Teman seperjuangan skripsi, khususnya Kevin Arnold dan Apsari Ayusya Cantika yang berdosan pembimbing sama dengan penulis, bimbingan bersama, saling membantu, dan saling memberikan dukungan satu sama lain selama menyusun skripsi ini.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis memohon maaf jika terdapat kesalahan. Penulis juga mengharapkan kritik dan saran yang membangun untuk menyempurnakan penelitian ini. Semoga penelitian ini dapat memberi informasi yang bermanfaat dan menjadi inspirasi untuk penelitian-penelitian berikutnya.

Bandung, Juni 2020

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL	xxiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	3
1.6 Sistematika Pembahasan	3
2 DASAR TEORI	5
2.1 Privasi Data	5
2.1.1 Privasi	5
2.1.2 <i>Personally Identifiable Information</i>	6
2.2 Penambangan Data	6
2.2.1 Klasifikasi	7
2.2.2 <i>Clustering</i>	8
2.3 <i>Privacy Preserving Data Mining</i>	9
2.4 Metode <i>Randomization</i>	10
2.4.1 <i>Random Noise Addition</i>	11
2.4.2 <i>Random Rotation Perturbation</i>	12
2.4.3 <i>Random Projection Perturbation</i>	13
2.5 Bahasa Pemograman Python	15
3 ANALISIS	17
3.1 Analisis Masalah	17
3.1.1 Implementasi Metode <i>Randomization</i>	17
3.1.2 Pengujian Dengan Penambangan Data	19
3.2 Studi Kasus	21
3.2.1 <i>Random Rotation Perturbation</i>	21
3.2.2 <i>Random Projection Perturbation</i>	23
3.3 Gambaran Umum Perangkat Lunak	25
3.3.1 Diagram Aktivitas	25
3.3.2 Diagram Kelas	27
4 PERANCANGAN PERANGKAT LUNAK	31
4.1 Perancangan Antarmuka	31

4.2	Perancangan Kelas	33
4.2.1	Diagram <i>Package</i>	33
4.2.2	Diagram Kelas pada <i>Package Perturbation</i>	36
4.2.3	Diagram Kelas pada <i>Package Matrix</i>	39
4.2.4	Diagram Kelas pada <i>Package Preprocessor</i>	42
4.3	Masukan dan Keluaran Perangkat Lunak	45
5	IMPLEMENTASI DAN PENGUJIAN	47
5.1	Implementasi Antarmuka	47
5.1.1	Masukan dan Pengaturan	47
5.1.2	Deskripsi <i>Dataset</i>	56
5.1.3	Deskripsi Hasil Pengacakan	58
5.2	Pengujian Fungsional	59
5.2.1	Teknik <i>Random Rotation Perturbation</i>	59
5.2.2	Teknik <i>Random Projection Perturbation</i>	62
5.3	Pengujian Eksperimental	65
5.3.1	Properti Data	66
5.3.2	Penambahan Data Klasifikasi	69
5.3.3	Penambahan Data <i>Clustering</i>	73
5.3.4	Kesimpulan Akhir	80
6	KESIMPULAN DAN SARAN	81
6.1	Kesimpulan	81
6.2	Saran	82
	DAFTAR REFERENSI	83
A	KODE PERANGKAT LUNAK <i>Randomization</i>	85
A.1	<i>Package Perturbation</i>	85
A.1.1	Kelas <i>Perturbation</i>	85
A.1.2	Kelas <i>RandomRotationPerturbation</i>	85
A.1.3	Kelas <i>RandomProjectionPerturbation</i>	86
A.2	<i>Package Matrix</i>	86
A.2.1	Kelas <i>Matrix</i>	86
A.2.2	Kelas <i>RandomRotationMatrix</i>	87
A.2.3	Kelas <i>RandomTranslationMatrix</i>	87
A.2.4	Kelas <i>RandomProjectionMatrix</i>	87
A.3	<i>Package Preprocessor</i>	87
A.3.1	Kelas <i>CSVPreprocessor</i>	87
A.3.2	Kelas <i>ProjectionMatrixPreprocessor</i>	88
A.3.3	Kelas <i>RotationMatrixPreprocessor</i>	88
A.4	<i>Package View</i>	89
A.4.1	Dokumen <i>main_menu.py</i>	89
A.4.2	Dokumen <i>randomization_app.py</i>	94
A.4.3	Dokumen <i>randomization.kv</i>	95
B	KODE PERANGKAT LUNAK PENGUJIAN	99
B.1	Pengujian Penambahan Data Klasifikasi	99
B.1.1	<i>Random Rotation Perturbation</i>	99
B.1.2	<i>Random Projection Perturbation</i>	100
B.2	Pengujian Penambahan Data <i>Clustering</i>	102
B.2.1	<i>Random Rotation Perturbation</i>	102

B.2.2	<i>Random Projection Perturbation</i>	104
B.3	Pengujian Lainnya	105

DAFTAR GAMBAR

1.1	Berbagai macam teknik modifikasi data untuk <i>privacy preserving data mining</i>	1
2.1	Ilustrasi penjelasan teknik <i>Random Rotation Perturbation</i> pada bidang Euclidean .	12
3.1	Diagram aktivitas perangkat lunak <i>Randomization</i>	26
3.2	Diagram kelas perangkat lunak <i>Randomization</i>	28
4.1	Halaman saat memilih teknik <i>Random Rotation Perturbation</i>	32
4.2	Halaman saat memilih teknik <i>Random Projection Perturbation</i>	32
4.3	<i>Popup</i> yang ditampilkan apabila pengacakan sukses dilakukan	33
4.4	<i>Popup</i> yang ditampilkan apabila pengacakan gagal dilakukan	34
4.5	<i>Popup</i> yang ditampilkan apabila persyaratan pada <i>dataset</i> tidak terpenuhi	34
4.6	Diagram <i>package</i> perangkat lunak	35
4.7	Diagram kelas pada <i>package Perturbation</i>	36
4.8	Diagram kelas pada <i>package Matrix</i>	39
4.9	Diagram kelas pada <i>package Preprocessor</i>	42
5.1	Tampilan perangkat lunak yang pertama ditampilkan saat perangkat lunak baru dibuka	48
5.2	Bagian-bagian pada antarmuka perangkat lunak	48
5.3	Bagian antarmuka masukan dan pengaturan perangkat lunak	49
5.4	Jendela untuk memilih <i>dataset</i> yang berupa dokumen CSV	49
5.5	Tampilan <i>popup</i> yang ditampilkan saat proses memuat dokumen berlangsung	50
5.6	Tampilan antarmuka setelah sebuah dokumen dipilih	51
5.7	Tampilan antarmuka saat pengguna memilih teknik	52
5.8	Tampilan antarmuka saat perangkat lunak membuat dan menyimpan matriks	53
5.9	Tampilan <i>popup</i> yang ditampilkan apabila matriks yang ingin diimpor tidak sesuai dengan <i>dataset</i>	53
5.10	Tampilan <i>popup</i> peringatan apabila pengguna belum memilih <i>dataset</i> yang diinginkan untuk diacak	54
5.11	Tampilan antarmuka parameter teknik <i>Random Projection Perturbation</i>	54
5.12	Tampilan <i>popup</i> peringatan tombol “Hitung Nilai Min K”	55
5.13	Tampilan saat perangkat lunak sedang melakukan proses pengacakan	56
5.14	Tampilan <i>popup</i> untuk memberitahukan pengguna bahwa pengacakan berhasil dilakukan	57
5.15	Tampilan <i>popup</i> peringatan apabila pengguna belum memilih atau membuat matriks rotasi/proyeksi	57
5.16	Tampilan antarmuka deskripsi <i>dataset</i> setelah pengguna memilih <i>dataset</i> yang ingin diacak	58
5.17	Tampilan antarmuka deskripsi hasil pengacakan setelah perangkat berhasil melakukan pengacakan	59
5.18	Matriks rotasi dan translasi acak yang dibuat perangkat lunak dan disimpan pada satu dokumen <i>comma-separated values</i>	60

5.19	Dua puluh baris pertama <i>dataset mall_customers</i> asli	61
5.20	Dua puluh baris pertama <i>dataset mall_customers</i> setelah diacak	61
5.21	Matriks proyeksi acak yang dibuat perangkat lunak dan disimpan pada sebuah dokumen <i>comma-separated values</i>	63
5.22	Dua puluh baris terakhir <i>dataset mobile_sensor</i> yang asli	64
5.23	Dua puluh baris terakhir <i>dataset mobile_sensor</i> setelah diacak	65
5.24	Dua puluh baris pertama <i>dataset diabetes</i> asli	67
5.25	Dua puluh baris pertama <i>dataset diabetes</i> setelah diacak	67
5.26	Grafik akurasi model klasifikasi pada <i>training set</i> dan <i>test set dataset diabetes</i> . . .	70
5.27	Grafik akurasi model klasifikasi pada <i>training set</i> dan <i>test set dataset mobile_sensor</i> . . .	71
5.28	Grafik <i>Sum of Squared Error</i> model <i>clustering</i> pada <i>dataset mall_customers</i>	73
5.29	Grafik <i>Silhouette Score</i> model <i>clustering</i> pada <i>dataset mall_customers</i>	74
5.30	Visualisasi <i>cluster</i> pada <i>dataset mall_customers</i> yang asli	75
5.31	Visualisasi <i>cluster</i> pada <i>dataset mall_customers</i> yang telah diacak	76
5.32	Grafik <i>Sum of Squared Error</i> model <i>clustering</i> pada <i>dataset mobile_sensor</i>	77
5.33	Grafik <i>Silhouette Score</i> model <i>clustering</i> pada <i>dataset mobile_sensor</i>	77
5.34	Visualisasi <i>cluster</i> pada <i>dataset mobile_sensor</i>	79

DAFTAR TABEL

3.1	Tabel <i>dataset iris</i> yang digunakan sebagai contoh kasus	22
5.1	Contoh perbedaan jarak Euclidean antara <i>dataset mall_customers</i> asli dan yang telah diacak dengan <i>Random Rotation Perturbation</i>	62
5.2	Contoh perbedaan jarak Euclidean antara <i>dataset mobile_sensor</i> asli dan yang telah diacak dengan <i>Random Projection Perturbation</i>	65
5.3	Properti-properti pada <i>dataset diabetes</i> asli	68
5.4	Properti-properti pada <i>dataset diabetes</i> yang telah diacak	68
5.5	Properti-properti pada <i>dataset mobile_sensor</i> asli	68
5.6	Properti-properti pada <i>dataset mobile_sensor</i> yang telah diacak	69
5.7	Perbandingan Properti Data	69
5.8	Perbandingan Model <i>k-nearest neighbors</i> antara model yang dilatih dengan <i>dataset</i> asli dan yang telah diacak	72
5.9	Perbandingan model <i>k-means</i> antara model yang dilatih dengan <i>dataset</i> asli dan yang telah diacak	80

BAB 1

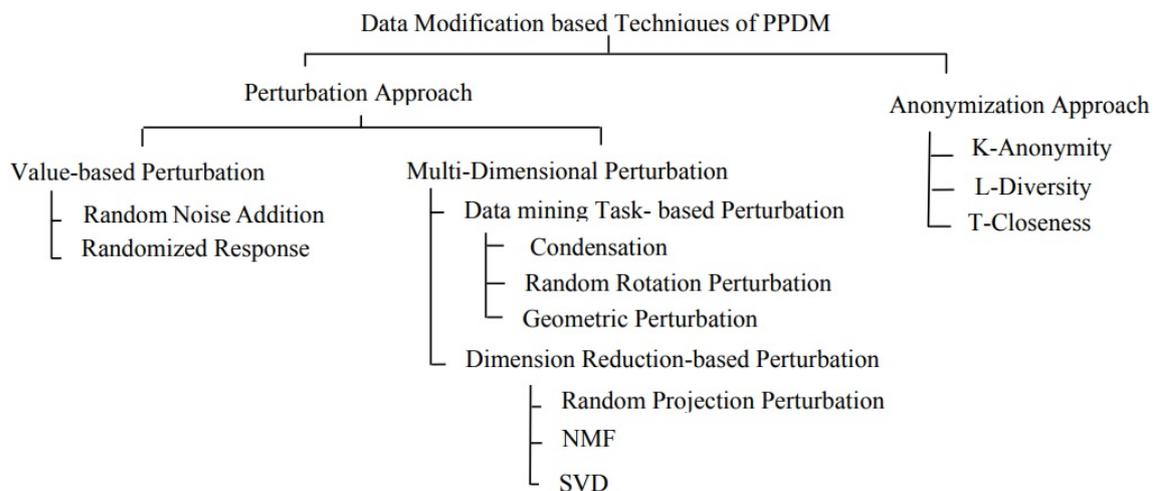
PENDAHULUAN

1.1 Latar Belakang

Semakin banyaknya penambahan data yang dilakukan dan data yang digunakan juga semakin banyak, semakin banyak juga privasi di dalam data tersebut yang tersebar kepada pihak yang melakukan penambahan data. Data privasi tersebut dapat tersebar kepada pihak yang tidak bertanggung jawab dan disalahgunakan. Oleh karena itu perlu adanya suatu cara untuk mencegah privasi tersebar pada proses penambahan data, menjaga privasi pada data tersebut. Istilah untuk hal tersebut adalah *privacy preserving data mining* [1].

Ada kesulitan dalam menentukan data seperti apa yang dapat disebut sebagai privasi. Privasi dapat dikatakan adalah sebuah informasi personal seseorang yang dapat mengidentifikasi suatu hal pada orang tersebut. Konsep yang sering kali digunakan untuk mendeskripsikan informasi personal adalah *Personally Identifiable Information* yang disingkat PII. PII [2] adalah segala informasi mengenai individu yang dikelola oleh sebuah instansi, termasuk segala informasi yang dapat digunakan untuk membedakan atau mengusut identitas seseorang dan juga segala informasi yang berhubungan atau dapat dihubungkan kepada suatu individu, seperti informasi medis, pendidikan, finansial, dan pekerjaan seseorang.

Salah satu cara untuk melakukan *privacy preserving data mining* adalah dengan melakukan modifikasi data yang ada sebelum diberikan kepada pihak lain. Ada macam-macam teknik dan algoritma yang bertujuan modifikasi data untuk *privacy preserving data mining*, 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 dapat ditambang. *Perturbation Approach* dapat dibagi menjadi dua jenis yaitu *Value-based Perturbation Techniques* dan *Multi-Dimensional Perturbation*.



Gambar 1.1: Berbagai macam teknik modifikasi data untuk *privacy preserving data mining*

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 modifikasi data sehingga properti yang bertahan pada data yang telah dimodifikasi spesifik hanya properti yang digunakan oleh suatu teknik penambangan data tertentu. Sedangkan *Dimension Reduction-based Perturbation* adalah teknik yang bekerja dengan cara modifikasi data sekaligus mengurangi dimensi dari data asli.

Dari berbagai macam teknik modifikasi data untuk *privacy preserving data mining* yang dapat dilihat pada Gambar 1.1, terdapat empat teknik yang menggunakan metode *Randomization* yaitu *Random Noise Addition*, *Randomized Response*, *Random Rotation Perturbation*, dan *Random Projection Perturbation*. Pada penelitian ini, dibuat sebuah perangkat lunak yang dapat memproses data yang akan ditambang menjadi data yang telah dimodifikasi dengan metode *Randomization* sehingga privasi pada data tersebut terlindungi, tetapi masih dapat ditambang. Dari berbagai macam teknik dengan metode *Randomization* yang ada, dipilih dua buah teknik yaitu *Random Rotation Perturbation* dan *Random Projection Perturbation* untuk diimplementasikan pada perangkat lunak serta membandingkan hasil dari kedua teknik tersebut.

Pengujian dilakukan dengan menerapkan penambangan data klasifikasi dengan algoritma *k-nearest neighbors* dan penambangan data *clustering* dengan algoritma *k-means*. Pada penambangan data klasifikasi, dihitung akurasi dan dibandingkan antara model yang dilatih dengan *dataset* asli dan *dataset* yang telah diacak dengan teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*. Tujuannya untuk melihat apakah antara kedua model tersebut memiliki akurasi yang sama. Pada penambangan data *clustering*, *dataset* asli dan *dataset* yang telah diacak masing-masing digunakan untuk membuat model *clustering*. Kemudian visualisasi hasil *clustering*-nya dibandingkan untuk melihat hasil *cluster* tersebut apakah sama. Kemiripan model *clustering* juga dihitung dengan metode *Adjusted Rand Index*. Tujuannya untuk melihat apakah antara kedua model tersebut memiliki hasil *cluster* yang sama.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana cara kerja teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* untuk *privacy preserving data mining*?
2. Bagaimana implementasi dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* pada perangkat lunak?
3. Bagaimana perbandingan antara hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*?

1.3 Tujuan

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah sebagai berikut.

1. Mempelajari cara kerja dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* untuk *privacy preserving data mining*.
2. Mengimplementasikan teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* pada perangkat lunak.
3. Melakukan analisis dan pengujian untuk membandingkan hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* dengan menggunakan penambangan data.

1.4 Batasan Masalah

Batasan-batasan masalah untuk penelitian ini adalah sebagai berikut.

1. Penelitian ini hanya menguji dengan teknik penambangan data klasifikasi dengan algoritma *k-nearest neighbors* dan *clustering* dengan algoritma *k-means*.
2. Perangkat lunak yang dibuat hanya dapat menerima masukan yang valid yaitu seluruh datanya bersifat numerik.

1.5 Metodologi

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Melakukan studi literatur dasar-dasar privasi data.
2. Melakukan studi literatur teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* untuk *privacy preserving data mining*.
3. Melakukan studi literatur teknik penambangan data yang digunakan.
4. Melakukan analisis terhadap teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* serta bagaimana penerapannya dengan teknik penambangan data yang digunakan.
5. Melakukan perancangan perangkat lunak yang mengimplementasikan teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*.
6. Membangun perangkat lunak yang mengimplementasikan teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*.
7. Menguji perangkat lunak secara fungsional dan eksperimental dengan menggunakan *real data*.
8. Menerapkan teknik penambangan data terhadap data yang telah diacak untuk menganalisis hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*.
9. Melakukan analisis dan pengujian untuk membandingkan hasil penambangan data menggunakan data yang telah diacak dengan teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*.
10. Menarik kesimpulan berdasarkan hasil eksperimen yang telah dilakukan.

1.6 Sistematika Pembahasan

Laporan penelitian tersusun ke dalam enam bab secara sistematis sebagai berikut.

- Bab 1 Pendahuluan
Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- Bab 2 Dasar Teori
Berisi dasar teori tentang dasar-dasar privasi data, metode *Randomization*, teknik penambangan data, dan bahasa pemrograman Python.
- Bab 3 Analisis
Berisi analisis masalah terkait implementasi metode *Randomization* dan pengujian dengan teknik penambangan data, studi kasus teknik *Randomization* yang diimplementasikan, dan gambaran umum perangkat lunak yang meliputi diagram aktivitas dan diagram kelas.
- Bab 4 Perancangan Perangkat Lunak
Berisi perancangan perangkat lunak yang dibangun meliputi perancangan antarmuka, diagram kelas yang lengkap, serta masukan dan keluaran perangkat lunak.
- Bab 5 Implementasi dan Pengujian
Berisi implementasi antarmuka perangkat lunak, pengujian fungsional perangkat lunak, pengujian eksperimental menggunakan teknik penambangan data, dan kesimpulan dari pengujian.
- Bab 6 Kesimpulan dan Saran
Berisi kesimpulan dari awal hingga akhir penelitian dan saran untuk pengembangan selanjutnya.

