

BAB 1

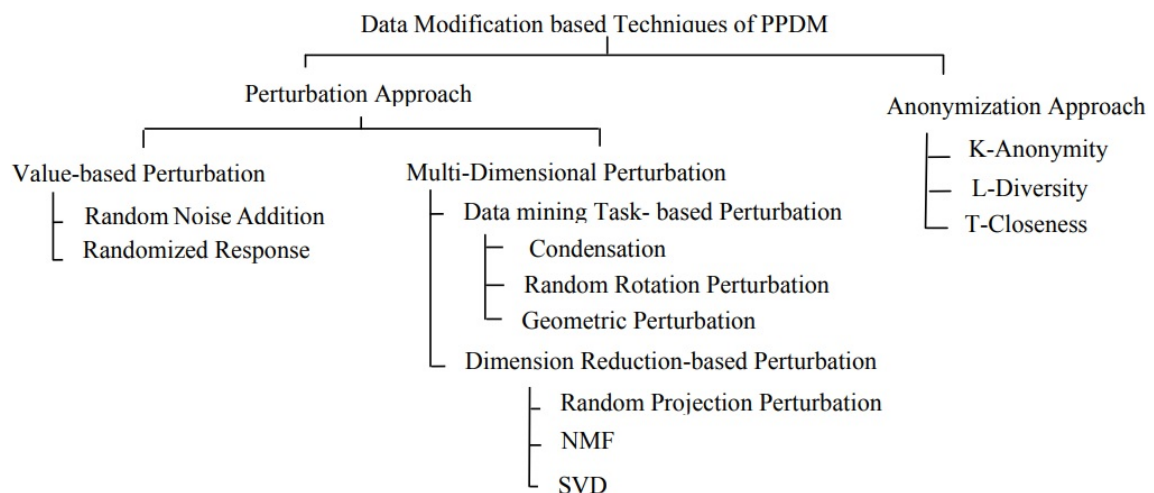
PENDAHULUAN

1.1 Latar Belakang

Dengan semakin banyaknya penambangan data yang dilakukan dan data yang digunakan juga semakin banyak, semakin banyak juga privasi di dalam data tersebut yang tersebar kepada pihak yang melakukan penambangan 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 penambangan data, menjaga privasi pada data tersebut. Istilah untuk hal tersebut adalah *privacy preserving data mining*.

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* 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 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.



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

Dari berbagai macam teknik modifikasi data untuk *privacy preserving data mining* yang dapat dilihat pada Gambar 2.1, terdapat empat teknik yang menggunakan metode *Randomization* yaitu *Random Noise Addition*, *Randomized Response*, *Random Rotation Perturbation*, dan *Random Projection Perturbation*.

Pada penelitian ini, akan dibuat sebuah perangkat lunak yang dapat memproses data yang akan ditambang menjadi 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 Rotation Perturbation* dan *Random Projection Perturbation* untuk diimplementasikan pada perangkat lunak serta membandingkan hasil dari kedua teknik tersebut.

1.2 Rumusan Masalah

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

1. Bagaimana cara kerja dari 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 dan mengukur hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*

1.4 Batasan Masalah

Batasan-batasan masalah untuk penelitian ini adalah sebagai berikut.

- 1.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

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 akan digunakan
4. Melakukan analisis terhadap teknik *Random Rotation Perturbation* dan *Random Projection Perturbation* serta bagaimana penerapannya dengan teknik penambangan data yang akan 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 diproses untuk menganalisis hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*
9. Melakukan analisis dan pengujian untuk membandingkan dan mengukur hasil dari teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*
10. Menarik kesimpulan berdasarkan hasil eksperiment 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 Landasan Teori
Berisi dasar teori tentang dasar-dasar privasi data, *Random Rotation Perturbation*, *Random Projection Perturbation*, dan teknik penambangan data.
- Bab 3 Analisis Masalah Berisi analisis masalah, studi kasus, dan diagram aliran proses.
- Bab 4 Perancangan Berisi perancangan perangkat lunak yang dibangun meliputi perancangan antarmuka dan diagram kelas yang lengkap.
- Bab 5 Implementasi dan Pengujian Berisi implementasi antarmuka perangkat lunak, pengujian fungsional, pengujian eksperimental, dan kesimpulan dari pengujian.
- Bab 6 Kesimpulan dan Saran Berisi kesimpulan dari awal hingga akhir penelitian dan saran untuk pengembangan selanjutnya.

BAB 2

LANDASAN TEORI

Dalam menjaga privasi data, perlu adanya definisi privasi yang konkrit untuk menentukan data seperti apa yang menjadi privasi. Pada penambahan data, perlu ada teknik yang baik untuk menjaga privasi tidak tersebar kepada orang yang tidak berhak. Ada beberapa teknik untuk menjaga privasi pada penambahan data antara lain modifikasi data dengan metode randomisasi yaitu teknik *Random Rotation Perturbation* dan *Random Projection Perturbation*

2.1 Privasi Data

Pada umumnya sebuah data bisa dikatakan privasi apabila data tersebut dapat dikaitkan dengan identitas seseorang. Tetapi setiap orang memiliki kepentingan privasi yang berbeda-beda sehingga definisi dari privasi sulit untuk dijelaskan secara eksak. Oleh karena itu, perlu adanya konsep privasi yang dapat menjadi acuan untuk menentukan data seperti apa yang termasuk privasi atau bukan.

2.1.1 Privasi

Dalam mendefinisikan privasi, sulit untuk mendapatkan definisi yang tepat untuk privasi karena setiap individu memiliki kepentingan yang berbeda-beda sehingga privasi pada setiap individu dapat berbeda-beda juga. Beberapa definisi privasi telah dikemukakan dan definisi tersebut bermacam-macam berdasarkan konteks, budaya, dan lingkungan. [1] Menurut Warren dan Brandeis pada papernya, mereka mendefinisikan privasi sebagai “*the right to be alone.*”, hak untuk menyendiri. Lalu pada papernya, Westin mendefinisikan privasi sebagai “*the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude, and their behavior to others*”, keinginan orang untuk memilih secara bebas dalam segala situasi dan dalam hal mengemukakan diri mereka, sikap mereka, dan tingkah laku mereka pada orang lain. Schoeman mendefinisikan privasi sebagai “*the right to determine what (personal) information is communicated to others*”, hak untuk menentukan informasi pribadi apa saja yang dikomunikasikan kepada yang lain, atau “*the control an individual has over information about himself or herself.*”, kendali seorang individu terhadap informasi tentang dirinya sendiri. Lalu baru-baru ini, Garfinkel menyatakan bahwa “*privacy is about self-possession, autonomy, and integrity.*”, privasi adalah tentang penguasaan diri sendiri, otonomi, dan integritas. Di samping itu, Rosenberg berpendapat bahwa privasi sebenarnya bukan sebuah hak tetapi sebuah rasa: “*If privacy is in the end a matter of individual taste, then seeking a moral foundation for it – beyond its role in making social institutions possible that we happen to prize – will be no more fruitful than seeking a moral foundation for the taste for truffles.*”, intinya setiap orang memiliki perhatian yang berbeda-beda terhadap privasi mereka sendiri sehingga hal tersebut tergantung apa yang dirasakan oleh setiap individu.

Dari definisi-definisi privasi yang telah disebutkan di atas, dapat disimpulkan bahwa privasi dilihat sebagai konsep sosial dan budaya. [1] Konsep privasi pada suatu lingkungan dapat berbeda dari lingkungan lainnya dan hal ini menyebabkan sulitnya menentukan apakah sebuah data termasuk privasi atau bukan. Oleh karena itu, perlu adanya sebuah standar privasi untuk menentukan data mana yang dapat disebut sebuah privasi. Organisasi National Institute of Standards and Technology

dari Amerika Serikat, membuat standar mereka sendiri untuk menentukan informasi seperti apa yang dapat disebut sebagai privasi. Mereka mengemukakan konsep *Personally Identifiable Information* sebagai informasi yang dapat dikatakan personal untuk setiap individu.

2.1.2 *Personally Identifiable Information*

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 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. [2]

Informasi yang termasuk membedakan individu adalah informasi yang dapat mengidentifikasi seorang individu. Beberapa contoh informasi yang mengidentifikasi seorang individu adalah nama, nomor KTP, tempat tanggal lahir, nama ibu kandung, atau catatan medis. Sedangkan, data yang hanya berisi misalkan saldo tabungan tanpa ada informasi lain mengenai identitas seseorang yang berkaitan tidak menyediakan informasi yang cukup untuk mengidentifikasi seorang individu.

Mengusut identitas seseorang adalah proses dari membuat perkiraan tentang aspek spesifik dari aktivitas atau status seseorang. Contohnya adalah sebuah catatan finansial seseorang dapat digunakan untuk memperkirakan aktivitas dari individu tersebut.

Informasi yang berhubungan dapat didefinisikan sebagai informasi yang berkaitan dengan seorang individu yang mana terkait secara logis dengan informasi lain tentang individu tersebut. Contohnya adalah apabila ada dua buah basis data yang memiliki data berbeda dari seorang individu, maka seseorang yang memiliki akses pada 2 basis data tersebut berpotensi dapat mengaitkan data-data tersebut lalu mengidentifikasi individu yang ada pada data tersebut.

2.2 Penambangan Data

Pada era teknologi informasi, sangat banyak data terkumpul pada basis data. Data yang masif ini dapat dimanfaatkan untuk menggali informasi penting yang berguna untuk pembuatan keputusan. Proses pada aktivitas ini secara kasar dapat disebut dengan penambangan data.

Penambangan data adalah proses mengekstrak sebuah pola atau sebuah pengetahuan dari kumpulan data yang besar, yang mana dapat direpresentasikan dan diinterpretasikan. [3] Pada penambangan data, teknik *machine learning* dan *pattern recognition* intensif digunakan untuk mendapatkan pola maupun pengetahuan baru dari data. Tujuan utama dari penambangan data adalah untuk membentuk model deskriptif dan prediktif dari suatu data. Model deskriptif berusaha untuk mengubah pola-pola yang ada pada data menjadi deskripsi yang bisa dimengerti oleh orang awam. Sedangkan model prediktif digunakan untuk memprediksi data yang tidak diketahui atau data yang berpotensi muncul di kemudian hari.

Model tersebut biasanya dibuat dengan menggunakan teknik *machine learning*, yang mana terdapat dua teknik *machine learning* yang paling sering digunakan yaitu *classification* dan *clustering*. Subbab berikutnya akan menjelaskan secara singkat kedua teknik tersebut dan contoh algoritmanya.

2.2.1 *Classification*

Tujuan utama *Classification* (klasifikasi) adalah membuat model yang dalam kasus ini disebut *classifier* yang mana dapat mengidentifikasi nilai kelas dari suatu data. [3] Dalam kata lain, sebuah *classifier* dibuat dari sebuah *training set* dan model ini digunakan untuk mengklasifikasi data tidak diketahui ke dalam salah satu kelas. Ada dua tahap dalam proses klasifikasi yaitu tahap latihan dan tahap klasifikasi.

Pada tahap latihan, model akan dibuat dengan menggunakan *training set*. *Training set* yang dimaksud adalah data yang sudah diketahui kelasnya sehingga model yang ada melatih dirinya. Setelah *classifier* terbentuk, barulah tahap klasifikasi dapat dilakukan dengan menggunakan *classifier* yang tadi sudah dibuat. *Classifier* akan memprediksi data yang kelasnya tidak diketahui. *Classifier* akan semakin baik performanya seiring dengan banyaknya tahap latihan yang dilakukan.

Teknik *machine learning* yang paling dikenal untuk klasifikasi antara lain *K-Nearest Neighbors*, *Decision Tree*, dan *Naive Bayes*. Dalam penelitian ini, hanya teknik *K-Nearest Neighbors* yang digunakan untuk pengujian sehingga berikutnya hanya akan dijelaskan teknik *K-Nearest Neighbors* saja.

Teknik *K-Nearest Neighbors* adalah «TODO»

2.2.2 Clustering

Clustering adalah proses mengelompokkan kumpulan objek ke dalam sebuah kelompok (*cluster*) sedemikian rupa sehingga objek-objek dari suatu *cluster* memiliki lebih banyak kemiripan dari pada objek-objek dari *cluster* lainnya. [3]

Salah satu contoh teknik *clustering* adalah *k-means*. Teknik *k-means* adalah «TODO»

2.3 Privacy Preserving Data Mining

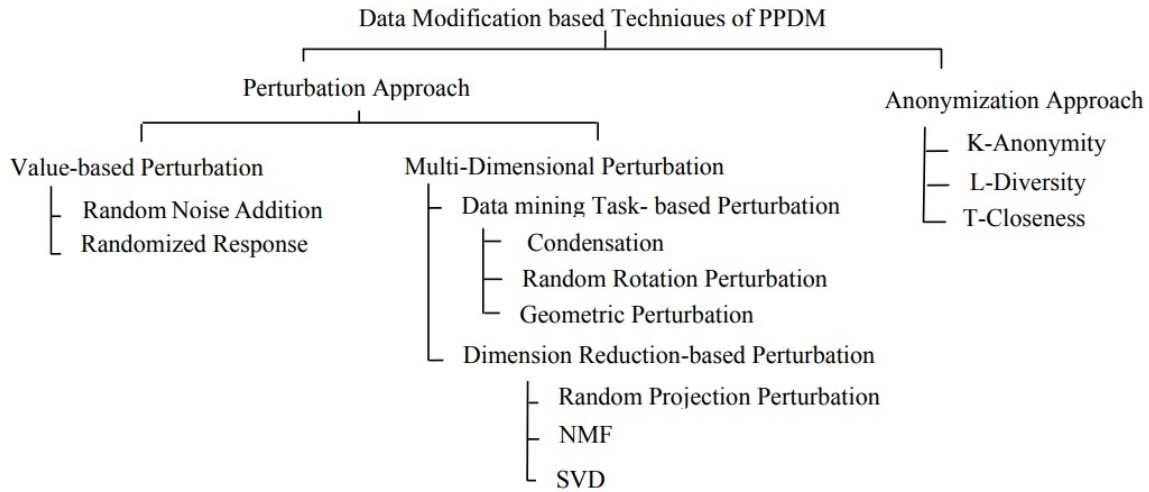
Aktivitas penambangan data melibatkan jumlah data yang sangat masif. Data-data yang digunakan memiliki privasi banyak individu di dalamnya. Hal ini berpotensi menyebabkan pelanggaran privasi dalam kasus tidak adanya proteksi yang cukup dan penyalahgunaan privasi data untuk tujuan lain. [4] Faktor utama pelanggaran privasi pada penambangan data adalah penyalahgunaan data sehingga hal ini dapat merugikan seorang individu maupun sebuah organisasi. Oleh karena itu, ada kebutuhan untuk menghindari penyebaran informasi pribadi yang rahasia maupun pengetahuan lainnya yang dapat diambil dari data yang digunakan untuk aktivitas penambangan data.

Konsep privasi sering kali lebih kompleks dari pada yang dibayangkan. Dalam kasus penambangan data, definisi dari menjaga privasi masih tidak jelas. Ada sebuah paper yang mendefinisikan *privacy preserving data mining* sebagai “getting valid data mining results without learning the underlying data values”, mendapatkan hasil penambangan data yang valid tanpa nilai pada data. Tetapi pada saat ini setiap teknik *privacy preserving data mining* yang ada memiliki definisi privasinya masing-masing.

Salah satu cara untuk melakukan *privacy preserving data mining* adalah dengan melakukan modifikasi data yang ada sebelum diberikan kepada pihak lain. Berbagai macam pendekatan modifikasi data untuk *privacy preserving data mining* telah dikembangkan antara lain *Perturbation Approach* dan *Anonymization Approach*, selengkapnya dapat dilihat pada Gambar 2.1. [4] *Perturbation Approach* adalah pendekatan untuk *privacy preserving data mining* dengan cara mengacaukan data yang ada, tetapi hasil data yang dikacaukan masih tetap bisa ditambang. Sedangkan pada *Anonymization Approach*, data diterapkan de-identifikasi di mana dataset mentah disebarluaskan setelah menghapus inti dari identitas setiap record. [4]

Perturbation Approach bisa dibagi menjadi dua jenis lagi 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 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.

Hal yang sering kali diperhatikan pada teknik-teknik *Perturbation Approach* adalah perbandingan



Gambar 2.1: Berbagai macam teknik modifikasi data untuk *privacy preserving data mining*

1 antara jumlah privasi yang hilang dan jumlah informasi yang hilang. Idealnya teknik *Perturbation*
 2 *Approach* yang baik adalah teknik yang fokus meminimalkan jumlah privasi yang hilang dan jumlah
 3 informasi yang hilang sehingga hasil penambangan dan akurasi sama baiknya dengan tanpa
 4 menerapkan teknik *Perturbation Approach*. Setiap teknik penambangan data memakai properti
 5 yang berbeda-beda pada data yang ditambah. Oleh karena itu, properti yang terjaga pun sebaiknya
 6 berdasarkan properti yang digunakan pada teknik penambangan data yang digunakan. [5] Pada
 7 saat ini, teknik modifikasi data yang ada sering kali mempunyai perbedaan pada properti-properti
 8 yang terjaga. Teknik-teknik modifikasi data tertentu sering kali mempunyai fungsi yang berbeda
 9 atau teknik penambangan data yang dapat digunakan berbeda karena properti yang terjaga pada
 10 teknik-teknik tersebut berbeda juga.

11 2.4 Metode *Randomization*

12 Dari berbagai macam teknik modifikasi data untuk *privacy preserving data mining* yang dapat dilihat
 13 pada Gambar 2.1, terdapat empat teknik yang menggunakan metode *Randomization* yaitu *Random*
 14 *Noise Addition*, *Randomized Response*, *Random Rotation Perturbation*, dan *Random Projection*
 15 *Perturbation*.

16 Berbagai macam teknik dengan metode randomisasi umumnya menerapkan perusakan nilai
 17 pada data. Salah satu teknik yang pertama kali menggunakan metode randomisasi untuk *privacy*
 18 *preserving data mining* adalah teknik *Random Noise Addition* yang dikemukakan oleh Agrawal
 19 dan Srikant pada paper berikut [6]. Teknik *Random Noise Addition* ini dilakukan dengan cara
 20 menambahkan nilai random (*noise*) pada data. Nilai random tersebut diambil dari sebuah distribusi.
 21 Untuk menambang data yang telah ditambahkan *noise* ini perlu dilakukan rekonstruksi distribusi
 22 untuk mendapatkan distribusi yang asli. Oleh karena itu, teknik *Random Noise Addition* ini
 23 hanya menjaga distribusi data asli sehingga hanya teknik penambangan data yang bergantung
 24 pada distribusi data saja yang bisa digunakan. Penyesuaian pada algoritma penambangan data
 25 yang digunakan pun perlu dilakukan agar teknik *Random Noise Addition* ini dapat digunakan
 26 dan mendapatkan hasil penambangan data yang hampir sama dengan tanpa menggunakan teknik
 27 *Random Noise Addition*.

28 Setelah teknik *Random Noise Addition* ditemukan, berbagai macam teknik lain pun dikembangkan
 29 terinspirasi dari teknik *Random Noise Addition* ini. Teknik *Random Rotation Perturbation* dan
 30 *Random Projection Perturbation* adalah teknik adalah salah satunya, tetapi teknik tersebut tidak
 31 dilakukan dengan cara menambahkan *noise* melainkan mengkalikan data asli dengan nilai random.
 32 Bagaimanapun juga, inti dari teknik-teknik randomisasi yang telah disebutkan di atas masih sama

1 yaitu merusak data sehingga data yang dirilis bukanlah data asli melainkan data yang sudah rusak
 2 sehingga data yang dirilis tidak mengandung privasi dan privasi pun terjaga. Masing-masing dari
 3 dua teknik tersebut akan dijelaskan lebih detail pada subbab berikut.

4 **2.4.1 *Random Rotation Perturbation***

5 Ide utama dari teknik *Random Rotation Perturbation* adalah jika data direpresentasikan sebagai
 6 matrix $X_{n \times d}$, *rotation perturbation* dari dataset X didefinisikan sebagai berikut.

$$G(X) = X_{n \times d} R_{d \times d} \quad (2.1)$$

7 Dimana $R_{d \times d}$ adalah *random rotation orthonormal matrix* yang mempunyai sifat *orthonormal*,
 8 yang mana matriks *orthonormal* adalah matriks yang menghasilkan matriks identitas apabila
 9 dikalikan dengan transposenya sendiri. Matriks rotasi ini dapat dibuat secara efisien mengikuti
 10 distribusi Haar. [7] Dari definisi di atas dapat disimpulkan transformasi rotasi tersebut menjaga
 11 jarak Euclidean. [5]

12 Teknik ini menjaga beberapa properti pada data antara lain yaitu jarak Euclidean, *inner*
 13 *product*, dan *geometric shape hyper* pada bidang multi-dimensi. [4] Oleh karena itu, beberapa
 14 teknik penambangan data tidak berpengaruh (dapat digunakan) terhadap teknik *Random Rotation*
 15 *Perturbation* antara lain yaitu *K-Nearest Neighbors*, *Support Vector Machines*, dan *Perceptrons*. [5]
 16 Teknik ini dipercaya dapat memberikan hasil penambangan yang maksimal, hasil penambangan
 17 data yang telah dirusak persis sama dengan hasil penambangan data aslinya. Sehingga jumlah
 18 informasi yang hilang tidak ada, tetapi jumlah privasi yang hilangnya tinggi. Walaupun demikian
 19 ada beberapa penelitian yang mengatakan bahwa karena teknik *Random Rotation Perturbation*
 20 ini mempunyai sifat demikian sehingga teknik ini dikatakan tidak aman dan dapat diserang dengan
 21 beberapa teknik untuk mendapatkan data asli yang lengkap.

22 Transformasi translasi juga perlu dilakukan agar rotasi yang dilakukan merusak data secara
 23 menyeluruh. Apabila tidak dilakukan translasi, nilai pada data yang mendekati nilai nol akan
 24 menghasilkan nilai yang mendekati nol juga setelah dirotasi. Implikasi dari hal tersebut adalah
 25 lemahnya dalam menjaga privasi. Translasi dapat dilakukan dengan cara membuat matriks translasi
 26 yang acak lalu kalikan dengan matriks data asli. Translasi dapat dilakukan karena translasi tidak
 27 mengubah properti geometris dari matriks yang ditranslasi sehingga jarak Euclidean dan properti
 28 lainnya pun terjaga dan hasil penambangan data pun tetap sama.

29 **2.4.2 *Random Projection Perturbation***

30 Ide utama dari teknik *Random Projection Perturbation* adalah mereduksi dimensi dari representasi
 31 matriks data asli dengan syarat dimensi matriks tersebut cukup besar. Dasar dari teknik *Random*
 32 *Projection Perturbation* berdiri pada *Johnson-Lindenstrauss Lemma* [8]. «TEOREMA» Inti dari
 33 Lemma ini menunjukkan bahwa titik pada bidang Euclidean d -dimensi dapat diproyeksikan ke
 34 bidang Euclidean berdimensi lebih kecil dari d , sedemikian rupa sehingga jarak antara dua titik
 35 tetap konsisten dengan *error* yang kecil tetapi dengan syarat d harus cukup besar. Oleh karena
 36 adanya *error* yang muncul, properti-properti pada data pun relatif sedikit berubah dan hal ini
 37 menyebabkan akurasi pada model yang dibuat dengan data tersebut berkurang dibandingkan data
 38 aslinya. [9]

39 *Projection perturbation* dari dataset X didefinisikan sebagai berikut. [9]

$$G(X) = X_{n \times m} R_{m \times k} \quad (2.2)$$

40 Dimana $R_{m \times k}$ adalah *random projection matrix* yang dihasilkan mengikuti distribusi normal, dengan
 41 rata-rata bernilai 0 dan standar deviasi bernilai $1/\sqrt{k}$. Ukuran matriks $R_{m \times k}$ disesuaikan dengan
 42 matriks $X_{n \times m}$ yang mana dataset asli dengan jumlah record n dan jumlah atribut m , yang mana
 43 m akan menjadi dimensi matriks. Oleh karena reduksi dimensi lah yang diinginkan maka k harus

lebih kecil dari pada m , yang mana k adalah dimensi dari matriks baru yang dihasilkan dari *Random Projection Perturbation* ini.

Jika *random projection matrix* yang dipakai dihasilkan secara acak saja, hasil dari *projection perturbation* akan terlalu merusak nilai pada data sehingga akurasi pada model yang akan dibuat kemungkinan berkurang drastis. Cara menanggulangi hal tersebut adalah membuat *random projection matrix* yang dipakai menjadi matriks *orthogonal*. Tetapi membuat matriks *orthogonal* yang berdimensi tinggi mempunyai kompleksitas yang tinggi sehingga memerlukan *cost* yang besar. Pada observasi yang dilakukan Hecht-Neilsen menunjukkan bahwa “*that in a high-dimensional space, vectors with random directions are almost orthogonal*”. Dapat disimpulkan bahwa dalam kasus matriks berdimensi tinggi apabila sebuah matriks dihasilkan secara acak mengikuti suatu distribusi, matriks tersebut akan kurang lebih hampir *orthogonal*. Oleh karena itu, matriks yang dibuat untuk *Random Projection Perturbation* cukup matriks acak yang mengikuti suatu distribusi saja.

Menurut *Johnson-Lindenstrauss Lemma*, reduksi dimensi pada matriks berdimensi tinggi minimal berdimensi k , yang mana k didefinisikan sebagai berikut.

$$k \geq 4(\epsilon^2/2 - \epsilon^3/3)^{-1} \ln n \quad (2.3)$$

Sebuah matriks yang akan diproyeksikan ke dimensi yang lebih kecil akan memiliki nilai *error* pada jarak Euclidean yang dimiliki oleh titik-titik (setiap elemen dari matriks) pada bidang Euclidean tersebut. Nilai *error* tersebut direpresentasikan dengan variabel ϵ . ϵ akan dipilih secara acak dalam rentang $[0, 1]$, yang mana ϵ menjadi ukuran seberapa baik proyeksi dilakukan. Semakin kecil nilai ϵ maka semakin besar k , yang mana k adalah dimensi minimal matriks yang dihasilkan. Semakin titik-titik pada bidang Euclidean diproyeksikan ke dimensi lebih kecil, semakin besar kerusakan yang timbul pada jarak Euclidean titik-titik tersebut.

2.5 L^AT_EX

Mengapa menggunakan L^AT_EX untuk buku skripsi dan apa keunggulan/kerugiannya bagi mahasiswa dan pembuat template.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

2.6 Template Skripsi FTIS UNPAR

Akan dipaparkan bagaimana menggunakan template ini, termasuk petunjuk singkat membuat referensi, gambar dan tabel. Juga hal-hal lain yang belum terpikir sampai saat ini.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy

magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

2.6.1 Tabel

Berikut adalah contoh pembuatan tabel. Penempatan tabel dan gambar secara umum diatur secara otomatis oleh L^AT_EX, perhatikan contoh di file bab2.tex untuk melihat bagaimana cara memaksa tabel ditempatkan sesuai keinginan kita.

Perhatikan bawa berbeda dengan penempatan judul gambar gambar, keterangan tabel harus diletakkan di atas tabel!! Lihat Tabel 2.1 berikut ini:

Tabel 2.1: Tabel contoh

	v_{start}	\mathcal{S}_1	v_{end}
τ_1	1	12	20
τ_2	1		20
τ_3	1	9	20
τ_4	1		20

Tabel 2.2 dan Tabel 2.3 berikut ini adalah tabel dengan sel yang berwarna dan ada dua tabel yang bersebelahan.

Tabel 2.2: Tabel bewarna(1)

	v_{start}	\mathcal{S}_2	\mathcal{S}_1	v_{end}
τ_1	1	5	12	20
τ_2	1	8		20
τ_3	1	2/8/17	9	20
τ_4	1			20

Tabel 2.3: Tabel bewarna(2)

	v_{start}	\mathcal{S}_1	\mathcal{S}_2	v_{end}
τ_1	1	12	5	20
τ_2	1		8	20
τ_3	1	9	2/8/17	20
τ_4	1			20

2.6.2 Kutipan

Berikut contoh kutipan dari berbagai sumber, untuk keterangan lebih lengkap, silahkan membaca file referensi.bib yang disediakan juga di template ini. Contoh kutipan:

- Buku: [10]
- Bab dalam buku: [11]
- Artikel dari Jurnal: [12]
- Artikel dari prosiding seminar/konferensi: [13]
- Skripsi/Thesis/Disertasi: [14] [15] [16]

- Technical/Scientific Report: [17]
- RFC (Request For Comments): [18]
- Technical Documentation/Technical Manual: [19] [20] [21]
- Paten: [22]
- Tidak dipublikasikan: [23] [24]
- Laman web: [25]
- Lain-lain: [26]

2.6.3 Gambar

Pada hampir semua editor, penempatan gambar di dalam dokumen \LaTeX tidak dapat dilakukan melalui proses *drag and drop*. Perhatikan contoh pada file bab2.tex untuk melihat bagaimana cara menempatkan gambar. Beberapa hal yang harus diperhatikan pada saat menempatkan gambar:

- Setiap gambar **harus** diacu di dalam teks (gunakan *field* LABEL)
- *Field* CAPTION digunakan untuk teks pengantar pada gambar. Terdapat dua bagian yaitu yang ada di antara tanda [dan] dan yang ada di antara tanda { dan }. Yang pertama akan muncul di Daftar Gambar, sedangkan yang kedua akan muncul di teks pengantar gambar. Untuk skripsi ini, samakan isi keduanya.
- Jenis file yang dapat digunakan sebagai gambar cukup banyak, tetapi yang paling populer adalah tipe PNG (lihat Gambar 2.2), tipe JPG (Gambar 2.3) dan tipe PDF (Gambar 2.4)
- Besarnya gambar dapat diatur dengan *field* SCALE.
- Penempatan gambar diatur menggunakan *placement specifier* (di antara tanda [dan] setelah deklarasi gambar. Yang umum digunakan adalah **H** untuk menempatkan gambar **sesuai** penempatannya di file .tex atau **h** yang berarti "kira-kira" di sini. Jika tidak menggunakan *placement specifier*, \LaTeX akan menempatkan gambar secara otomatis untuk menghindari bagian kosong pada dokumen anda. Walaupun cara ini sangat mudah, hindarkan terjadinya penempatan dua gambar secara berurutan.
 - Gambar 2.2 ditempatkan di bagian atas halaman, walaupun penempatannya dilakukan setelah penulisan 3 paragraf setelah penjelasan ini.
 - Gambar 2.3 dengan skala 0.5 ditempatkan di antara dua buah paragraf. Perhatikan penulisannya di dalam file bab2.tex!
 - Gambar 2.4 ditempatkan menggunakan *specifier* **h**.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

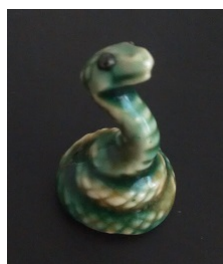
Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora

Gambar 2.2: Gambar *Serpentes* dalam format png

torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.



Gambar 2.3: Ular kecil

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros.

1 Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae
2 nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac
3 enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec
4 vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh
5 pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna
6 tincidunt congue.



Gambar 2.4: *Serpentes* jantan

DAFTAR REFERENSI

- [1] dan Osmar R. Zaiane, S. R. M. O. (2004) Towards standardization in privacy-preserving data mining. *ACM SIGKDD 3rd Workshop on Data Mining Standards*, **3**, 862–870.
- [2] NIST Special Publication 800-122 (2010) *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. National Institute of Standards and Technology, U.S. Department of Commerce, Erika McCallister, Tim Grance, Karen Scarfone. Gaithersburg, Maryland.
- [3] dan JOAO P. VILELA, R. M. (2017) Privacy-preserving data mining: Methods, metrics, and applications. *IEEE Access*, **5**, 10562–10582.
- [4] dan Somayyeh Seifi Moradi, M. K. (2011) Classification and evaluation the privacy preserving data mining techniques by using a data modification-based framework. *International Journal on Computer Science and Engineering*, **3**, 862–870.
- [5] dan L. Liu, K. C. (2005) A random rotation perturbation approach to privacy preserving data classification. Technical Report GIT-CC-05-12. Georgia Institute of Technology, Georgia.
- [6] dan R. Srikant, R. A. (2000) Privacy preserving data mining. *In Proceedings of the ACM SIGMOD*, **3**, 439–450.
- [7] STEWART, G. W. (1980) The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, **17**, 403–409.
- [8] Johnson, W. B. dan Lindenstrauss, J. (1984) Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, **26**, 189–206.
- [9] Kun Liu, d. J. R., Hillol Kargupta (2006) Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, **18**, 92–106.
- [10] de Berg, M., Cheong, O., van Kreveld, M. J., dan Overmars, M. (2008) *Computational Geometry: Algorithms and Applications*, 3rd edition. Springer-Verlag, Berlin.
- [11] van Kreveld, M. J. (2004) Geographic information systems. Bagian dari Goodman, J. E. dan O'Rourke, J. (ed.), *Handbook of Discrete and Computational Geometry*. Chapman & Hall/CRC, Boca Raton.
- [12] Buchin, K., Buchin, M., van Kreveld, M. J., Löffler, M., Silveira, R. I., Wenk, C., dan Wiratma, L. (2013) Median trajectories. *Algorithmica*, **66**, 595–614.
- [13] van Kreveld, M. J. dan Wiratma, L. (2011) Median trajectories using well-visited regions and shortest paths. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, USA, 1-4 November, pp. 241–250. ACM, New York.
- [14] Lionov (2002) Animasi algoritma sweepline untuk membangun diagram voronoi. Skripsi. Universitas Katolik Parahyangan, Indonesia.

-
- [15] Wiratma, L. (2010) Following the majority: a new algorithm for computing a median trajectory. Thesis. Utrecht University, The Netherlands.
 - [16] Wiratma, L. (2022) Coming Not Too Soon, Later, Delay, Someday, Hopefully. Disertasi. Utrecht University, The Netherlands.
 - [17] van kreveld, M., van Lankveld, T., dan Veltkamp, R. (2013) Watertight scenes from urban lidar and planar surfaces. Technical Report UU-CS-2013-007. Utrecht University, The Netherlands.
 - [18] Rekhter, Y. dan Li, T. (1994) A border gateway protocol 4 (bgp-4). RFC 1654. RFC Editor, <http://www.rfc-editor.org>.
 - [19] ITU-T Z.500 (1997) *Framework on formal methods in conformance testing*. International Telecommunications Union. Geneva, Switzerland.
 - [20] Version 9.0.0 (2016) *The Unicode Standard*. The Unicode Consortium. Mountain View, USA.
 - [21] Version 7.0 Nougat (2016) *Android API Reference Manual*. Google dan Open Handset Alliance. Mountain View, USA.
 - [22] Webb, R., Daruca, O., dan Alfadian, P. (2012) *Method of optimizing a text message communication between a server and a secure element*. Paten no. EP2479956 (A1). European Patent Organisation. Munich, Germany.
 - [23] Wiratma, L. (2009) Median trajectory. Report for GMT Experimentation Project at Utrecht University.
 - [24] Lionov (2011) Polymorphism pada C++. Catatan kuliah AKS341 Pemrograman Sistem di Universitas Katolik Parahyangan, Bandung. <http://tinyurl.com/lionov>. 30 September 2016.
 - [25] Erickson, J. (2003) CG models of computation? <http://www.computational-geometry.org/mailling-lists/compegeom-announce/2003-December/000852.html>. 30 September 2016.
 - [26] AGUNG (2012) Menjajal tango 12. Majalah HAI no 02, Januari 2012.

LAMPIRAN A

KODE PROGRAM

Listing A.1: MyCode.c

```
1 // This does not make algorithmic sense,
2 // but it shows off significant programming characters.
3
4 #include<stdio.h>
5
6 void myFunction( int input, float* output ) {
7     switch ( array[i] ) {
8         case 1: // This is silly code
9             if ( a >= 0 || b <= 3 && c != x )
10                 *output += 0.005 + 20050;
11             char = 'g';
12             b = 2^n + ~right_size - leftSize * MAX_SIZE;
13             c = (--aaa + &daa) / (bbb++ - ccc % 2 );
14             strcpy(a,"hello_$@?");
15         }
16         count = ~mask | 0x00FF00AA;
17     }
18 }
19
20 // Fonts for Displaying Program Code in LATEX
21 // Adrian P. Robson, nepsweb.co.uk
22 // 8 October 2012
23 // http://nepsweb.co.uk/docs/progfonts.pdf
```

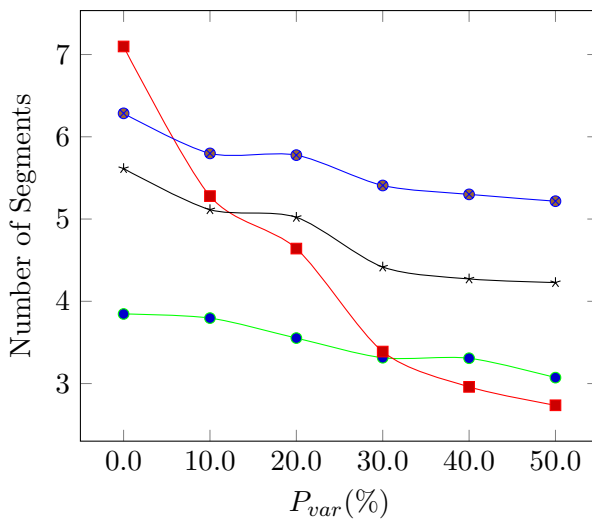
Listing A.2: MyCode.java

```
1 import java.util.ArrayList;
2 import java.util.Collections;
3 import java.util.HashSet;
4
5 //class for set of vertices close to furthest edge
6 public class MyFurSet {
7     protected int id; //id of the set
8     protected MyEdge FurthestEdge; //the furthest edge
9     protected HashSet<MyVertex> set; //set of vertices close to furthest edge
10    protected ArrayList<ArrayList<Integer>> ordered; //list of all vertices in the set for each trajectory
11    protected ArrayList<Integer> closeID; //store the ID of all vertices
12    protected ArrayList<Double> closeDist; //store the distance of all vertices
13    protected int totaltrj; //total trajectories in the set
14
15    /*
16     * Constructor
17     * @param id : id of the set
18     * @param totaltrj : total number of trajectories in the set
19     * @param FurthestEdge : the furthest edge
20     */
21    public MyFurSet(int id,int totaltrj,MyEdge FurthestEdge) {
22        this.id = id;
23        this.totaltrj = totaltrj;
24        this.FurthestEdge = FurthestEdge;
25        set = new HashSet<MyVertex>();
26        ordered = new ArrayList<ArrayList<Integer>>();
27        for (int i=0;i<totaltrj;i++) ordered.add(new ArrayList<Integer>());
28        closeID = new ArrayList<Integer>(totaltrj);
29        closeDist = new ArrayList<Double>(totaltrj);
30        for (int i = 0;i <totaltrj;i++) {
31            closeID.add(-1);
32            closeDist.add(Double.MAX_VALUE);
33        }
34    }
35
36 }
```

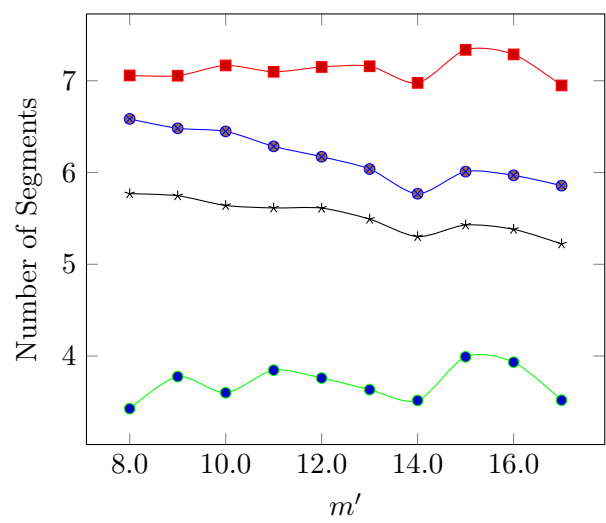

LAMPIRAN B

HASIL EKSPERIMEN

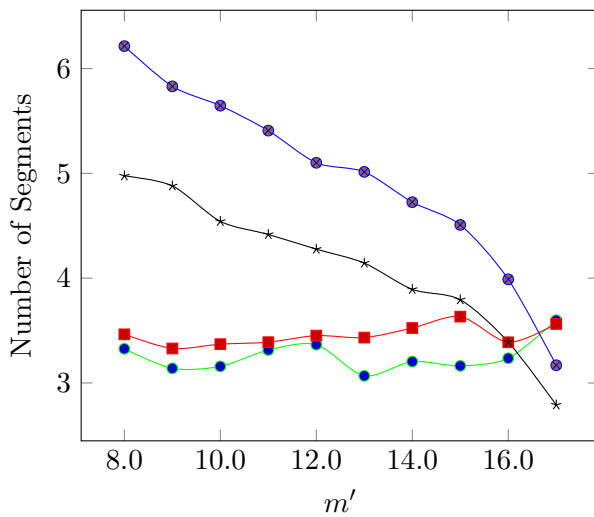
Hasil eksperimen berikut dibuat dengan menggunakan TIKZPICTURE (bukan hasil excel yg diubah ke file bitmap). Sangat berguna jika ingin menampilkan tabel (yang kuantitasnya sangat banyak) yang datanya dihasilkan dari program komputer.



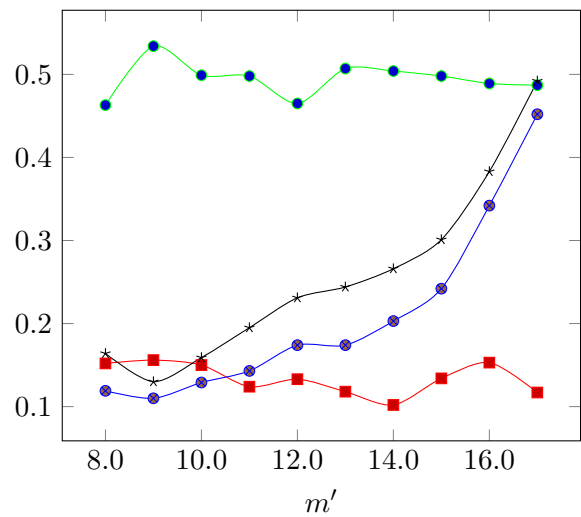
Gambar B.1: Hasil 1



Gambar B.2: Hasil 2



Gambar B.3: Hasil 3



Gambar B.4: Hasil 4