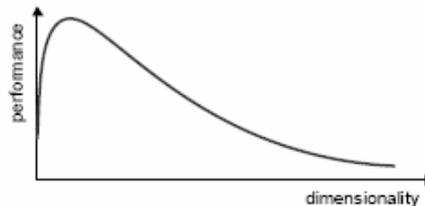


1. Pendahuluan

1.1 Latar Belakang

Data mining merupakan suatu bidang yang sangat berkembang pesat karena adanya kebutuhan untuk memanfaatkan gudang data yang dimiliki, yaitu dengan melakukan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Ada beberapa tantangan dalam *data mining* yaitu salah satunya *dimensionality* "fenomena *curse of dimensionality*".

Fenomena *curse of dimensionality* merupakan fenomena yang terjadi pada data mining dengan adanya peningkatan jumlah dimensi data sehingga dapat menyebabkan turunnya performansi. Oleh karena itu harus dilakukan *preprocessing* data yaitu dengan mengurangi dimensi tersebut [15].



Gambar 1-1: Ilustrasi Curse of Dimensionality

Tujuan utamanya antara lain menghindari *curse of dimensionality*, mengurangi jumlah waktu dan memori yang diperlukan oleh algoritma data mining, membuat data lebih mudah divisualisasikan dan membantu mengeliminasi data yang tidak relevan dengan begitu maka data akan menjadi lebih akurat.

Terdapat beberapa teknik yang sering dipakai dalam mereduksi dimensi diantaranya *principle component analysis* (PCA) dan teknik yang akan penulis analisis yaitu *independent component analysis* (ICA) yang merupakan perluasan dari PCA itu sendiri. ICA melakukan proses pengurangan dimensi dengan melihat keterhubungan antar dimensi dan independensi secara statistik. Oleh karena itu penulis tertarik untuk menganalisa sejauh mana ICA bekerja, membandingkan hasil evaluasi *clustering* dengan ICA dan PCA serta menganalisis apakah ICA dapat memberikan hasil yang lebih baik dalam mereduksi dimensi, juga sejauh mana pengaruhnya terhadap hasil *clustering*.

1.2 Perumusan masalah

Berdasarkan latar belakang terdapat permasalahan tersebut, maka akan dilakukan analisa lebih lanjut. Diantaranya bagaimana mendapatkan hasil yang terbaik pada saat melakukan mining dengan data yang memiliki dimensi tinggi. Diharapkan dengan algoritma ICA yaitu dengan menganalisa *statistical independent component* dapat mengurangi dimensi secara optimal dan mempengaruhi hasil

clustering dengan akurasi yang baik. Pengukuran tingkat akurasi dan optimal dapat diukur diantaranya dengan SSE, *overall entropy*, dan overall purity.

1.3 Tujuan

Berdasarkan rumusan masalah di atas, maka tujuan dari tugas akhir ini adalah:

1. Mengimplementasikan ICA untuk mereduksi dimensi data .
2. Menganalisis ICA terkait dengan hasil reduksi .
3. Menganalisis hasil evaluasi *clustering* setelah dilakukan pengurangan dimensi dengan ICA.
4. Membandingkan hasil evaluasi *clustering* ICA dan PCA.

1.4 Batasan Masalah

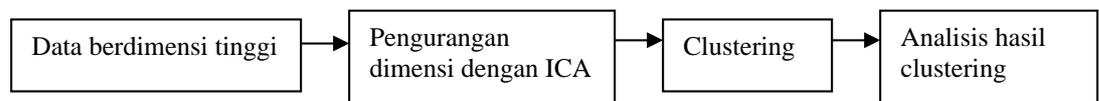
Adapun batasan masalah yang digunakan dalam penelitian ini antara lain :

1. Dataset yang digunakan berupa data numerik .
2. Menganalisis data yang telah direduksi ICA dengan menggunakan *tools* clementine 10.0 .
3. Teknik data mining yang dipakai *K-means clustering*.

1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini dengan langkah kerja sebagai berikut :

1. Studi Literatur :
Pencarian referensi dan mempelajari berbagai literatur yang layak yang berhubungan dengan *data mining* , *clustering*, *preprocessing*, *dimensional reduction*, clementine 10.0, PCA dan ICA .
2. Melakukan analisis dan perancangan perangkat lunak
Sebelum melakukan analisis terdapat beberapa tahapan secara umum, diantaranya pengurangan dimensi dengan ICA untuk data berdimensi tinggi kemudian dilakukan *clustering* dan analisis hasil *clustering* pada data tersebut. Hasil pengurangan dimensi dapat diuji setelah dilakukan *clustering* sehingga dapat dianalisis pada dimensi-dimensi tertentu akan menghasilkan klasterisasi yang optimal .



Gambar 1-2: Tahapan Umum

3. Melakukan implementasi algoritma ICA dengan matlab versi 7 .
4. Melakukan pengujian dan analisis terhadap hasil implementasi ICA dengan menggunakan clementine 10.0 .
5. Pengambilan kesimpulan dan penyusunan laporan tugas akhir.