

BAB 1 PENDAHULUAN

1.1 Latar belakang

Pada zaman sekarang bisnis berkembang menjadi lebih pesat dibandingkan dengan masa lalu, dimana tujuan perusahaan mengedepankan kebutuhan pelanggan untuk menjadi lebih kompetitif dan dinamis dibandingkan dengan perusahaan atau organisasi lainnya. Dengan demikian perkembangan Business Intelligence yang ada dimanfaatkan untuk strategi bisnis yang tepat. Business intelligence merupakan cara untuk mengumpulkan, menyimpan, mengorganisasikan, membentuk ulang, meringkas data serta menyediakan informasi baik berupa data aktifitas bisnis internal perusahaan termasuk aktifitas bisnis pesang yang mudah diakses serta dianalisis untuk berbagai kegiatan manajemen (David,2000) [1].

Business intelligence juga dapat dispesifikasi melalui empat dimensi yaitu : Strategis, Taktis, Operasional dan *Real-time* (Asghar, Fong, & Hussain, 2009) [2]. Pada Real-time BI (rt-BI) akurasi generalisasi dan waktu respon merupakan kriteria penting . Menganalisis data sesegera mungkin ketika data masuk kedalam organisasi merupakan bagian dari real-time BI.

Akurasi generalisasi dan waktu respon adalah dua kriteria penting untuk mengevaluasi adanya suatu pengklasifikasian bila diterapkan dalam real-time BI (rt-BI). Pengklasifikasian diperlukan tidak hanya untuk menggambarkan pelatihan data tetapi juga untuk memprediksi data yang tidak terlihat (Jiaqi Wang, 2005) [3].

Oleh karena itu kebutuhan perkembangan BI yang meningkat, maka diperlukan adanya analisis data sesegera mungkin ketika data masuk kedalam suatu organisasi. Proses memasukan data sesegera mungkin yang terjadi secara berulang akan menimbulkan penumpukan terhadap data sehingga mempengaruhi pencarian informasi terhadap data tersebut.

Pada karya tulis ini, salah satu metode yang digunakan untuk melakukan proses data mining dengan menggunakan stream mining . Metode ini memiliki tujuan untuk mengatasi masalah dari jumlah data yang banyak dan di proses secara terus menerus dan membutuhkan waktu yang lama . Dengan menggunakan metode ini diharapkan dapat mengeluarkan hasil analisis BI dengan tingkat keterlambatan yang rendah. Untuk metode pengelompokan data , penulis menggunakan *Ensemble Method*. Metode ini digunakan untuk mempercepat pengelompokan data dan memiliki tingkat akurasi analisis yang tinggi. Menggunakan *Ensemble Method* dikarenakan data yang dipergunakan ukurannya sangat besar sehingga satu algoritma tidak cukup untuk mengatasi masalah ini sehingga metode *Ensemble Method* yang dipilih untuk menyelesaikan masalah ini. *Ensemble Method* yang dipilih untuk membahas masalah ini adalah K-means untuk clustering dan Algoritma Naïve Bayes untuk klasifikasi. Akurasi yang didapatkan pada penelitian tugas akhir ini yaitu 93.33%.

1.2 Perumusan Masalah

Berdasarkan uraian latar belakang diatas, maka perumusan masalah yang dicapai pada penelitian ini adalah sebagai berikut :

1. Bagaimana cara mengimplementasikan ensemble method pada sistem Business Intelligence ?
2. Bagaimana metode *k-means clustering* dan algoritma *Naïve Bayes* dapat mempercepat waktu respons pengelompokan data ?
3. Bagaimana metode *k-means clustering* dan algoritma *Naïve Bayes* dapat meningkatkan tingkat akurasi pemrosesan analisis ?

1.3 Tujuan

Berdasarkan uraian latar belakang diatas, maka tujuan yang dicapai pada penelitian ini adalah sebagai berikut :

1. Mengimplementasikan algoritma K-Means clustering dan algoritma Naïve Bayes untuk proses klasifikasi pada sistem Real-time Business Intelligence
2. Menganalisa kinerja kedua algoritma yang digunakan dalam menghasilkan sebuah model clustering dan klasifikasi terbaik yang akan di implementasikan kedalam sebuah sistem
3. Menganalisa performansi sistem dalam prediksi pengelompokan pelanggan berdasarkan data sales dan pelanggan.

1.4 Batasan Masalah

Adapun batasan masalah yang dibahas pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan adalah data customer dan data sales dari sebuah perusahaan.
2. Data yang masuk kedalam sistem diasumsikan sudah melewati tahap *ekstraksi data*.
3. Data yang dianalisis memiliki atribut kategorik.
4. Metode yang digunakan adalah Clustering Kmeans dan Klasifikasi Naive Bayes.
5. Data yang akan dikelompokkan tidak dapat dilakukan melalui proses online melainkan data diproses secara offline.
6. Sistem yang dibuat tidak berorientasi objek.

1.5 Metodologi

Metodologi untuk menyelesaikan masalah pada karya tulis ini adalah sebagai berikut :

1. Identifikasi masalah

Masalah yang ada didapat dari beberapa karya ilmiah yang ada dan masalah yang didapat dalam kehidupan sehari-hari. Kemudian masalah yang ada dirumuskan menjadi permasalahan yang utama.

2. Studi literature

Melakukan pembelajaran mengenai hal-hal yang berkaitan dengan masalah yang dibahas pada karya tulis ini (*Real-time Business Intelligence*, *K-means Clustering*, dan *Data Mining*) melalui berbagai paper, jurnal, buku dan artikel di internet .

3. Pemodelan sistem

Setelah studi literature sudah dilakukan yang berikutnya melakukan pemodelan sistem dari karya tulis yang dibuat. Pemodelan sistem dilakukan untuk mempermudah proses pengerjaan dan mencegah adanya perubahan-perubahan pada saat tahap implementasi sampai dengan tahap akhir.

4. Implementasi sistem

Setelah pemodelan sistem dilakukan proses selanjutnya adalah proses abstraksi. proses pengubahan komponen komponen yang ada menjadi sebuah bahasa pemrograman.

5. Pengujian sistem

Proses ini adalah proses pengujian yang dilakukan untuk mengetahui keakuratan dari sistem yang telah dibuat. Pengujian dilakukan lebih dari satu kali agar tingkat keakuratan dapat terlihat sebagai bahan untuk analisis.

6. Pembuatan Laporan

Proses yang terakhir adalah pembuatan laporan, pada proses ini analisis dilakukan untuk mendapatkan suatu kesimpulan dari sistem yang ada apakah sesuai dengan kriteria yang ditentukan. selain itu juga pada laporan adanya suatu evaluasi dari sistem ini berupa suatu kelebihan dan kelemahan sistem sehingga nantinya dapat dikembangkan. Pada laporan semua aspek yang ada ditulis sesuai dengan ketentuan yang ada.

