

BAB I

PENDAHULUAN

Pada bab ini akan dibahas mengenai beberapa hal seputar penelitian yang dilakukan, antara lain: latar belakang penelitian, perumusan masalah, batasan masalah, keaslian penelitian, manfaat penelitian yang diharapkan, tujuan penelitian, dan sistematika penulisan.

1.1 Latar Belakang

Kemajuan teknologi, dan proses bisnis saat ini semakin meningkat, sehingga menjadi sebuah tantangan untuk meningkatkan kualitas keputusan managerial (Adhatrao dkk, 2013). Keputusan yang diambil dapat ditentukan dari data-data yang dimiliki untuk memprediksi kemungkinan yang terjadi berdasar data kejadian yang sudah terjadi. Hampir seluruh bidang memiliki kumpulan data yang kompleks dan dalam skala yang besar yang dapat digunakan untuk membantu pengambilan keputusan (Suresh dkk, 2013). Keterbatasan manusia, menjadi kendala dalam mengolah data berskala besar secara manual untuk mendapatkan suatu informasi yang tersembunyi dalam kumpulan data tersebut.

Salah satu contoh yang memiliki data besar seperti perusahaan dan rumah sakit. Data yang disimpan dalam database dapat diolah menjadi suatu informasi penting untuk meningkatkan persaingan bisnis. Hal ini dapat diatasi dengan

memanfaatkan penambangan data untuk menemukan informasi yang berguna dan membantu dalam pengambilan keputusan (Jnanamurthy dkk, 2013). Penambangan data banyak membantu perusahaan seperti asuransi, bank dan industri retail dalam menentukan tujuan pemasarannya (Patil & Sherekar, 2013). Metodologi penambangan data memiliki kontribusi yang baik bagi para peneliti untuk mengekstrak pengetahuan dan informasi tersembunyi yang telah diwariskan dalam data yang digunakan oleh peneliti (Umamaheswari & Niraimathi, 2013).

Salah satu teknik dari penambangan data yang dapat membantu mengatasi permasalahan pada banyaknya data perusahaan ialah klasifikasi. Klasifikasi digunakan untuk memprediksi kelas dari objek yang kelasnya belum diketahui (Raviya & Gajjar, 2013). Metode klasifikasi yang umum digunakan antara lain *Decision Tree*, *K-Nearest Neighbor*, *Naïve Bayes*, *neural network* dan *support vector machines* (Sahu dkk, 2011). Naïve Bayes dan KNN juga termasuk pada 10 peringkat yang sering digunakan untuk pengolahan data karena memiliki algoritma yang sederhana, cepat dan memiliki nilai akurasi tinggi (Wu dkk, 2007). Raviya & Gajjar (2013) melakukan perbandingan dari ketiga metode klasifikasi penambangan data yaitu *Decision Tree*, *KNN* dan *Naïve Bayes*. Hasil yang diperoleh pada perbandingan tersebut, KNN dan Naïve Bayes dinyatakan sebagai metode yang sederhana pada tahap implementasi dibanding dengan *Decision Tree*. Hasil akurasi metode *decision tree* dan KNN memperoleh nilai yang sama yaitu 63,7%, sedangkan Naïve Bayes 37,1%. Kecepatan pengolahan data yang digunakan metode *Decision Tree* lebih lama dibanding KNN dan Naïve Bayes yaitu, 0,78 detik. Sedangkan KNN membutuhkan waktu 0,05 detik dan Naïve

Bayes 0,01 detik. Akurasi yang dihasilkan oleh KNN dan Naïve Bayes juga telah diuji pada penelitian yang dilakukan oleh Danesh dkk, (2007) dengan hasil akurasi untuk penggunaan metode Naïve Bayes 86,7%, dan K-Nearest Neighbor (KNN) 87,57%.

Hasil perbandingan dari penelitian yang telah dikaji, menunjukkan bahwa terdapat kelemahan dan kelebihan dari setiap metode penambangan data. Beberapa peneliti mengkombinasikan metode penambangan data untuk mengatasi kelemahan dari teknik penambangan data yang sudah ada. Farid, dkk, (2010), mengkombinasikan Naïve Bayes dan Decision Tree untuk deteksi gangguan jaringan yang telah terbukti mencapai tingkat akurasi tinggi. Penelitian lain dilakukan oleh Hall (2007), dengan tujuan meningkatkan proses kinerja Naïve Bayes berdasarkan algoritma Decision Tree yang terbukti lebih baik dibandingkan dengan Naïve Bayes trees dan selective Bayes. Kombinasi algoritma tersebut dimunculkan untuk memperbaiki kelemahan dari metode penambangan data yang sudah ada sebelumnya. Kelemahan pada metode KNN berada pada proses perhitungan yang dilakukan hampir disetiap data pada tahap klasifikasi (Guo dkk, 2006). Pada penelitian ini akan dikaji modifikasi algoritma yang merupakan kombinasi metode KNN dengan Naïve Bayes untuk mengatasi kelemahan pada metode KNN. Algoritma ini diharapkan dapat meningkatkan kecepatan metode penambangan data dengan nilai keakuratan yang tinggi. Kecepatan yang dapat ditingkatkan dengan metode kombinasi yaitu waktu yang dibutuhkan untuk klasifikasi data. Perbandingan akurasi dan waktu pengolahan data dari metode

KNN, Naïve Bayes dan kombinasi KNN-Naïve Bayes diterapkan pada bahasa pemrograman C++.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan masalah, bagaimana mengkombinasikan metode KNN dan Naïve Bayes untuk klasifikasi data?

1.3 Batasan Masalah

Penelitian yang dilakukan ini memiliki beberapa poin sebagai batasan masalahnya. Adapun batasan-batasan masalah yang dimaksud adalah sebagai berikut:

1. Akurasi dan kecepatan waktu metode kombinasi KNN dan Naïve Bayes hanya akan dibandingkan dalam persentase akurasi dan waktu dalam millisecond.
2. Data yang digunakan untuk metode kombinasi KNN-Naïve Bayes ialah data yang memiliki atribut bersifat kategori.

1.4 Keaslian Penelitian

Penelitian yang ini merupakan hasil karya pribadi dan bukan kutipan atau duplikasi dari karya tulis yang telah ada sebelumnya. Adapun perbedaan penelitian ini dengan penelitian-penelitian sebelumnya, seperti yang tampak pada Tabel 1.1.

Tabel 1. 1 Perbandingan Penelitian Metode Kombinasi Penambangan data

No.	Item Perbandingan	Xie dkk, (2002)	Jiang dkk, (2005)	Hall (2007)	Farid dkk, (2010)	Sari (2015)*
1	Metode	KNN-Naïve Bayes	KNN-Naïve Bayes	Naïve Bayes-DecisionTree	Naïve Bayes-Decision Tree	KNN-Naïve Bayes
2	Algoritma	<ol style="list-style-type: none"> 1. Mencari jarak terdekat. (KNN). 2. Hasil pencarian jarak terdekat digunakan untuk mencari probabilitas terbesar dengan Naïve Bayes. 	<ol style="list-style-type: none"> 1. Mencari jarak terdekat (KNN). 2. Cloned beberapa data pelatihan dengan jarak terdekat. 3. Mencari probabilitas dengan Naïve Bayes. 	<ol style="list-style-type: none"> 1. Pembobotan atribut menggunakan Decision Tree. 2. Mencari probabilitas terbesar dengan Naïve Bayes 	<ol style="list-style-type: none"> 1. Klasifikasi data dengan mencari probabilitas. 2. Perbarui nilai setiap kelas pada D. 3. Pilih atribut terbaik dengan gain maksimum pada D. 	<ol style="list-style-type: none"> 1. Mencari probabilitas dengan Naïve Bayes. 2. Mencari jarak terdekat data hasil langkah pertama dengan KNN
3	Hasil Perbandingan	Akurasi yang dihasilkan lebih tinggi dari metode NBTree, CBA, C4.5 dan Naïve Bayes dengan lama proses cukup lama.	Secara signifikan lebih baik dari metode Naïve Bayes, C4.4, NBTree, KNN, KNNDW dan LWNB.	Lebih baik dari Naïve Bayes dan Selective Bayes.	Akurasi dari percobaan yang dilakukan dengan algoritma yang diajukan ialah 99%.	Lebih baik dari KNN atau Naïve Bayes untuk klasifikasi data kategori.

*) Penelitian yang dilakukan oleh penulis

1.5 Manfaat Penelitian

Penelitian ini akan menghasilkan algoritma modifikasi dari kombinasi KNN dan Naïve Bayes. Algoritma kombinasi KNN-Naïve Bayes ini nantinya dapat mengatasi kelemahan dari KNN yang memiliki waktu pengolahan data yang lama dan persentase akurasi naïve bayes yang lebih rendah dari KNN. Algoritma ini, nantinya dapat digunakan untuk pengolahan data kategori.

1.6 Tujuan Penelitian

Berdasar pada rumusan masalah yang tertera di atas, maka tujuan peneliti melakukan penelitian ini ialah untuk meningkatkan kecepatan waktu dan persentase akurasi dari metode penambangan data KNN dan Naïve Bayes dengan metode kombinasi KNN-Naïve Bayes.

1.7 Sistematika Penulisan

Laporan ini disusun secara sistematis berdasarkan tata cara penulisan laporan yang telah ditetapkan oleh pihak Pascasarjana Universitas Atma Jaya Yogyakarta dengan urutan penyajian sebagai berikut :

a. BAB I PENDAHULUAN

Pada bab ini akan dipaparkan masalah umum tentang penyusunan laporan tesis yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, keaslian penelitian, tujuan, serta sistematika penulisan laporan tesis.

b. BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini akan dimuat tinjauan pustaka, dasar teori terkait dengan penelitian tesis yang dilakukan dan merupakan penyempurnaan serta perluasan dari proposal tesis.

c. BAB III METODOLOGI PENELITIAN

Pada bab ini akan dimuat metodologi yang digunakan dalam penelitian tesis. Metodologi penelitian yang ada pada laporan ini juga merupakan penyempurnaan dan perluasan proposal tesis.

d. BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini akan dibahas hasil penelitian yang didapatkan. Pembahasan berisi analisis dari hasil yang diperoleh, yang ditinjau secara utuh baik secara kualitatif, kuantitatif, maupun normatif. Pada bab ini juga akan dijelaskan mengenai implementasi sistem yang dikembangkan.

e. BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir dari keseluruhan rangkaian laporan tesis yang di dalamnya berisi suatu kesimpulan atas penelitian yang dilakukan. Selain itu, pada bagian ini juga akan dimuat saran-saran dari peneliti, baik yang berupa kritik dan gagasan untuk pekerjaan di masa yang akan datang terkait dengan penelitian yang dilakukan.