

1. Pendahuluan

1.1 Latar belakang

Dalam beberapa dekade terakhir, teknologi informasi dan basis data telah berkembang dari sistem pemrosesan file primitif menjadi sistem basis data yang canggih. Namun, seiring dengan berjalannya waktu, jumlah data yang sangat besar yang dikumpulkan dan disimpan dalam gudang penyimpanan data sering kali tidak dipakai oleh para analist dalam membuat keputusan karena adanya kesulitan dalam mengekstrak informasi dari data yang jumlahnya sangat besar. Akibatnya, keputusan yang dibuat hanya berdasarkan intuisi bukan berdasarkan informasi dari data yang ada. Oleh karena itu, pembuat keputusan membutuhkan *tool* untuk mengekstrak pengetahuan berharga dari data yang sangat besar. Salah satu *tool* tersebut yaitu *data mining* yang dapat menganalisis data dan menemukan pola data yang penting untuk pengambilan keputusan.

Beberapa *task data mining* yaitu klasifikasi, regresi, asosiasi, klusterisasi, dan *anomaly detection*. Dalam Tugas Akhir ini, akan dilakukan penelitian tentang klasifikasi. Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui sebelumnya. Model yang ditemukan dapat berupa aturan “jika-maka”, *decision tree*, formula matematis atau *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, *k-nearest neighbor*, dan *bayesian*. Salah satu teknik klasifikasi *bayesian* yang sedang berkembang yaitu *Bayesian Belief Networks (BBN)*.

BBN merupakan graf asiklik berarah yang simpul-simpulnya mewakili variabel-variabel pada *dataset* dan panah-panahnya mewakili relasi ketergantungan antar variabel dan distribusi probabilitas lokal untuk masing-masing variabel yang diberikan oleh orang tuanya. Salah satu keunggulan BBN yaitu *user* dapat mengerti ketergantungan langsung dan hubungan kausal antar variabel dengan mudah.

BBN memiliki lima tipe *classifier*, yaitu : *Naïve Bayes*, *Tree Augmented Naïve Bayes (TAN)*, *Bayesian network Augmented Naïve Bayes (BAN)*, *Bayesian multi-nets*, dan *General Bayesian networks (GBN)*. Pada Tugas Akhir ini, penulis akan meneliti *Naïve Bayes classifier* dan *Tree Augmented Naïve Bayes (TAN) classifier*. *TAN classifier* merupakan pengembangan *Naive Bayes classifier* dan kedua *classifier* tersebut menggunakan *restricted structure learning*.

1.2 Perumusan masalah

Masalah yang menjadi acuan dalam pembuatan Tugas Akhir ini adalah :

1. Bagaimana membangun model *TAN classifier* berdasarkan algoritma *Forest Augmented Naive Bayes*.
2. Bagaimana kinerja *Naive Bayes classifier* dan *TAN classifier* dalam proses klasifikasi.
3. Bagaimana mengimplementasikan ke dalam perangkat lunak model klasifikasi dengan teknik BBN, khususnya *Naive Bayes classifier* dan *TAN classifier*.

Adapun batasan-batasan dalam Tugas Akhir ini adalah sebagai berikut :

1. Tidak membahas *data mining* secara keseluruhan, hanya membahas klasifikasi dengan BBN menggunakan *Naïve Bayes classifier* dan *TAN classifier*.
2. Inputan data yang dimasukkan ke dalam sistem adalah data yang berupa data *record*.
3. Tidak menangani *data preprocessing*.
4. Data yang digunakan untuk *training* dan *testing* adalah data yang sudah dalam bentuk data diskret dan tidak ada *missing value*.
5. *Dataset* yang akan digunakan sudah tersedia dalam bentuk *tabel* dengan variabel *class* berada pada urutan paling akhir.

1.3 Tujuan

Tujuan yang ingin dicapai pada Tugas Akhir ini adalah :

1. Membuat model klasifikasi *Naïve Bayes classifier* dan *TAN classifier* melalui *learning* pada *dataset*.
2. Mengimplementasikan BBN dan menganalisis perilaku *Naïve Bayes classifier* dan *TAN classifier* terhadap akurasi dan waktu pengklasifikasian data.
3. Membandingkan kinerja *Naive Bayes classifier* dan *TAN classifier*.

1.4 Metodologi penyelesaian masalah

Metodologi yang akan digunakan untuk menyelesaikan Tugas Akhir ini adalah:

1. Studi literatur
 - a. Pencarian referensi
Mencari referensi dan sumber-sumber lain yang berhubungan dengan klasifikasi data, BBN, dan hal-hal lain yang berkaitan dengan Tugas Akhir ini.
 - b. Pendalaman materi
Mempelajari dan memahami materi konsep BBN serta proses pembangunan model klasifikasi dengan BBN.
2. Mempelajari konsep BBN yang akan digunakan dalam implementasi perangkat lunak.
3. Melakukan implementasi perancangan perangkat lunak.
4. Melakukan pengujian perangkat lunak dengan memasukkan data diskret, kemudian mencatat hasil keluaran program.
5. Menganalisis hasil klasifikasi dengan BBN.
6. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir.