

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Data Mining merupakan salah satu bidang ilmu yang berupaya untuk menemukan kaidah, pola, model, maupun informasi dalam bentuk lain yang menarik dan bersifat baru dari sekumpulan data. Salah satu task dalam data mining yang menjadi pokok perhatian dalam Tugas Akhir ini adalah klasifikasi. Salah satu metode klasifikasi yang banyak digunakan karena kesederhanaannya adalah pohon keputusan.

Fase pemangkasan pohon keputusan biasanya dilakukan setelah fase pertumbuhan pohon selesai dilakukan. Hal ini memerlukan waktu pemrosesan yang lama seiring dengan meningkatnya ukuran data yang dilatihkan. Bahkan penelitian menunjukkan bahwa fase pemangkasan akan memangkas sebagian besar dari pohon keputusan yang dibangun dalam fase pertumbuhan pohon.

Algoritma PUBLIC menawarkan integrasi fase pertumbuhan dan pemangkasan pohon keputusan dalam data mining. Dalam PUBLIC, sebuah node tidak diperluas selama fase pertumbuhan jika ditemukan bahwa node tersebut akan dipangkas pada fase pemangkasan.

Pada Tugas Akhir ini dilakukan studi analisis performansi algoritma PUBLIC terhadap parameter akurasi, kecepatan, kesederhanaan pohon yang dihasilkan dan skalabilitas.

### **1.2 Rumusan Masalah**

PUBLIC adalah sebuah algoritma klasifikasi dengan metode pohon keputusan yang mengintegrasikan fase pertumbuhan dan pemangkasan pohon.

Fase pemangkasan pohon keputusan dalam PUBLIC dilakukan berdasarkan prinsip MDL (*Minimum Description Length*) karena prinsip ini

---

akurat, memungkinkan pemangkasan dilakukan dengan data pelatihan, serta menghasilkan pohon keputusan yang lebih ringkas.

Permasalahan yang menjadi obyek penulisan Tugas Akhir ini adalah:

1. Bagaimana implementasi integrasi prinsip MDL untuk pemangkasan pohon dengan pembangunan pohon keputusan untuk masalah klasifikasi?
2. Bagaimana melakukan pengujian dan analisis hasil implementasi PUBLIC?

### **1.3 Maksud dan Tujuan**

Maksud penyusunan Tugas Akhir ini:

1. Mengimplementasikan algoritma PUBLIC dalam perangkat lunak.
2. Menganalisis akurasi, kecepatan, kesederhanaan pohon yang dihasilkan, dan skalabilitas algoritma PUBLIC dalam perangkat lunak.

Tujuan penulisan Tugas Akhir ini:

1. Membangun perangkat lunak yang mengimplementasikan algoritma PUBLIC untuk menyelesaikan masalah klasifikasi dalam data mining.
2. Menganalisis pengaruh integrasi fase pertumbuhan dan pemangkasan pohon keputusan dalam algoritma PUBLIC.

### **1.4 Batasan Masalah**

Agar penulisan Tugas Akhir ini tidak terlalu luas dan menjadi lebih mudah dipahami, perlu dilakukan pembatasan masalah sebagai berikut :

1. Data yang digunakan adalah himpunan data yang ada dalam basis data dbPUBLIC di SQL Server 2000.
2. Data yang digunakan harus sudah melalui preproses dan siap digunakan.
3. Pengujian dilakukan dengan parameter akurasi, kecepatan, kesederhanaan pohon yang dihasilkan, skalabilitas, dan kompleksitas algoritma.
4. Versi PUBLIC yang diimplementasikan dalam Tugas Akhir ini adalah PUBLIC(1) karena implementasinya lebih sederhana namun akurasi tidak kurang dari dua versi lainnya yaitu PUBLIC(S) dan PUBLIC(V).

---

## 1.5 Metode Penyelesaian Masalah

Metode penyelesaian masalah dalam penulisan tugas Akhir ini disusun dalam langkah-langkah sebagai berikut:

1. Studi literatur, yang dilakukan dengan membaca dan mempelajari beberapa sumber tertulis (makalah, buku dan jurnal) yang berkaitan dengan data mining, klasifikasi, pohon keputusan, fase pertumbuhan dan pemangkasan, prinsip MDL, induksi yang scalable, serta algoritma SLIQ dan J4.8 yang digunakan sebagai pembanding dalam pengujian.
2. Pengumpulan dan analisis data yang mendukung implementasi dan analisis algoritma PUBLIC dalam perangkat, yaitu data dari UCI maupun data sintetik yang diperoleh dari generator data.
3. Analisis kebutuhan dan perancangan perangkat lunak, untuk menentukan kebutuhan pembangunan perangkat lunak, serta perancangan struktur data dan aktivitas perangkat lunak yang dibangun, dengan metode berorientasi objek.
4. Implementasi, yang merupakan langkah penerapan rancangan yang telah dibuat ke dalam perangkat lunak yang dapat digunakan untuk menyelesaikan masalah klasifikasi dengan algoritma PUBLIC.
5. Pengujian dan analisis hasil, yaitu langkah yang diambil untuk menilai performansi algoritma PUBLIC dalam perangkat lunak dengan parameter yang disebutkan dalam batasan masalah, terhadap algoritma SLIQ sebagai sebuah classifier pohon keputusan dengan post-pruning dan J4.8 dalam WEKA sebagai classifier yang biasa digunakan dalam data mining.

## 1.6 Sistematika Penulisan

Penulisan Tugas Akhir ini dibagi dalam enam bab, yang terdiri atas :

Bab I Pendahuluan

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan, metode pemecahan, serta sistematika penulisan yang digunakan dalam penyusunan Tugas Akhir ini.

**Bab II Landasan Teori**

Berisi teori dan rumusan awal yang diperoleh dari berbagai sumber tentang data mining, klasifikasi, dan algoritma PUBLIC.

**Bab III Analisis dan Perancangan Sistem**

Bab ini berisi uraian tentang perancangan perangkat lunak, meliputi spesifikasi kebutuhan, perancangan struktur data dan proses dari perangkat lunak yang dibuat.

**Bab IV Implementasi**

Berisi penjelasan tentang pelaksanaan dan penerapan algoritma PUBLIC dalam suatu perangkat lunak yang dibangun.

**Bab V Pengujian dan Analisis Hasil**

Bab ini menguraikan analisis sistem dalam menyelesaikan proses klasifikasi pada himpunan data yang digunakan serta hasil pengujian terhadap perangkat lunak.

**Bab VI Kesimpulan dan Saran**

Berisi kesimpulan dari keseluruhan sistem yang dibuat serta saran untuk pengembangan perangkat lunak agar menjadi lebih baik.