

1. Pendahuluan

1.1 Latar belakang

Dalam beberapa tahun belakangan ini, berbagai penelitian dalam data mining telah banyak dikembangkan oleh para ahli. Klasifikasi dan asosiasi merupakan teknik yang paling banyak digunakan pada data mining. Klasifikasi dan asosiasi memiliki karakteristik yang berbeda dalam hal tujuan. Klasifikasi bertujuan untuk menemukan suatu kumpulan rule yang dapat mengelompokkan data ke dalam kelas-kelas data sedangkan asosiasi bertujuan untuk menemukan semua rule asosiasi yang terdapat pada data. Klasifikasi dan asosiasi sangat dibutuhkan di berbagai aplikasi dunia nyata seperti perbankan, telekomunikasi, kesehatan dan marketing sehingga diperlukan suatu teknik yang bisa menggabungkan kedua metode tersebut.

Dalam tugas akhir ini, dibangun suatu aplikasi yang mengintegrasikan klasifikasi dan asosiasi yang dikenal sebagai *associative classification* (CPAR-Classification Based on Predictive Association Rules). Pengintegrasian ini menggunakan algoritma greedy yang dapat menghasilkan rule classifier yang bersifat asosiasi dan prediktif. Kemudian CPAR melakukan klasifikasi berdasarkan rule yang dihasilkan tersebut. Dalam melakukan klasifikasi, CPAR menggunakan *Laplace Accuracy* untuk evaluasi rule dan k rules untuk melakukan prediksi.

Dengan adanya teknik CPAR dalam menghasilkan classifier, diharapkan performansi dari classifier tersebut menjadi lebih baik dibanding classifier lain dari teknik klasifikasi yang umum digunakan. Pada akhirnya, CPAR dapat diimplementasikan diberbagai aplikasi data mining karena menghasilkan rule classifier yang lebih akurat.

1.2 Perumusan masalah

Permasalahan yang menjadi objek penelitian dalam Tugas Akhir ini adalah:

1. Bagaimana mengimplementasikan CPAR dalam mengolah suatu data sehingga menghasilkan suatu model klasifikasi.
2. Bagaimana melakukan klasifikasi pada data menggunakan CPAR.
3. Bagaimana melakukan pengujian dan analisis dari implementasi CPAR.

Batasan masalah dalam Tugas Akhir ini adalah :

1. Dataset yang digunakan adalah dataset yang berasal dari *UCI Machine Learning Repository*. Dataset yang dipakai berjumlah delapan dataset.
2. Dataset yang digunakan adalah dataset yang atributnya bertipe numerik dan nominal.

3. Pengujian dilakukan dengan membandingkan CPAR dengan tool Data Mining yaitu WEKA.
4. Pengujian dilakukan dengan parameter akurasi dan jumlah rule classifier.

1.3 Tujuan

Berdasarkan rumusan masalah diatas, tujuan yang ingin dicapai dari Tugas Akhir ini adalah :

1. Merancang dan membangun perangkat lunak yang mengimplementasikan algoritma CPAR untuk melakukan klasifikasi data.
2. Menganalisa performansi dari algoritma CPAR. Performansi yang dianalisa tersebut adalah akurasi dan jumlah rule classifier yang dihasilkan serta membandingkannya dengan algoritma klasifikasi C4.5.

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang digunakan dalam Tugas Akhir ini antara lain :

1. Studi Literatur, yang dilakukan dengan mempelajari beberapa literatur (makalah, buku ataupun jurnal) yang berhubungan dengan data mining, klasifikasi, asosiasi, *associative classification*, FOIL, PRM dan CPAR serta metode untuk menganalisa performansi algoritma CPAR tersebut.
2. Pengumpulan dan analisis data yang meliputi pengumpulan dataset yang diperlukan, memahami maksud dari data-data maupun atribut tersebut serta melakukan *preprocessing* terhadap data sehingga data siap untuk dimining.
3. Analisis kebutuhan dan perancangan perangkat lunak yang akan dibangun. Dalam hal ini digunakan metode berorientasi objek.
4. Implementasi dari rancangan perangkat lunak yang telah dibuat, sehingga dihasilkan suatu perangkat lunak yang dapat digunakan untuk melakukan klasifikasi data dengan menerapkan algoritma CPAR. Dalam pembuatan perangkat lunak ini digunakan bahasa pemrograman Java.
5. Pengujian sistem dan analisa hasil, menganalisa classifier yang dihasilkan oleh CPAR dan menguji kebenaran klasifikasi serta mengukur hasil implementasi berdasarkan parameter akurasi dan jumlah rule classifier serta membandingkannya dengan algoritma klasifikasi C4.5. Perangkat lunak WEKA digunakan untuk membandingkan dengan algoritma klasifikasi lain tersebut.
6. Penyusunan laporan Tugas Akhir dan kesimpulan akhir. Hasil penelitian akan disusun dalam laporan yang meliputi aspek-aspek penelitian teori dan implementasi.