

1. Pendahuluan

1.1 Latar belakang

Seiring dengan perkembangan teknologi dalam hal pengumpulan dan penyimpanan data, maka munculah suatu kebutuhan untuk dapat menghasilkan informasi dari data yang telah ada tersebut. Dengan adanya ukuran data yang sangat besar diharapkan informasi penting yang didapat pun akan semakin banyak.

Salah satu cara untuk mendapatkan informasi atau pola dari kumpulan data yang besar adalah dengan menggunakan teknik-teknik dalam data mining. Ada banyak teknik dalam data mining untuk menghasilkan informasi atau pola dari sekumpulan data. Salah satu teknik tersebut adalah klasifikasi.

Random Forest merupakan salah satu algoritma klasifikasi dengan tingkat akurasi yang baik [1]. *Random Forest* merupakan sebuah metode *ensemble* yang terdiri dari beberapa pohon keputusan sebagai *classifier*. Kelas yang dihasilkan dari proses klasifikasi ini diambil dari kelas terbanyak yang dihasilkan oleh pohon-pohon keputusan yang ada pada *Random Forest*. Dengan melakukan voting pada pohon-pohon keputusan yang tersedia membuat akurasi dari *Random Forest* meningkat.

1.2 Perumusan masalah

Permasalahan yang dijadikan objek penelitian dalam Tugas Akhir ini adalah bagaimana mengimplementasikan algoritma klasifikasi *Random Forest*. Selain itu masalah lainnya adalah bagaimana melakukan analisa performansi dari algoritma klasifikasi *Random Forest*. Performansi yang dianalisa tersebut adalah akurasi dari prediksi yang dihasilkan.

Ruang lingkup dari Tugas Akhir ini adalah :

1. *Dataset* yang digunakan adalah *dataset* sintetik yang dihasilkan oleh data generator ataupun *dataset* yang berasal dari *UCI Machine Learning Repository*. *Dataset* yang dipakai berjumlah delapan *dataset*.
2. *Dataset* yang digunakan adalah *dataset* yang seluruh atributnya bertipe numerik.
3. Tidak menangani tahap *pre-processing*.
4. Pengujian dilakukan dengan parameter akurasi.
5. Inputan yang diterima berupa tabel yang telah tersedia di dalam *database*.

1.3 Tujuan

Tujuan Tugas Akhir ini adalah :

1. Merancang dan membangun perangkat lunak untuk klasifikasi data dengan algoritma *Random Forest*.
2. Menganalisa performansi dari algoritma *Random Forest*. Performansi yang dianalisa tersebut adalah akurasi dari prediksi yang dihasilkan serta membandingkannya dengan beberapa algoritma klasifikasi lain, yakni Naive Bayes, AdaboostM1, RepTree dan Bagging.

1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam Tugas Akhir ini antara lain :

1. Studi Literatur, yang dilakukan dengan mempelajari beberapa literatur (makalah, buku ataupun jurnal) yang berkaitan dengan *data mining*, klasifikasi terutama yang berkaitan dengan *random forest* dan metode untuk menganalisa performansi algoritma klasifikasi tersebut.
2. Pengumpulan dan analisis data yang digunakan untuk mendukung implementasi dan analisis algoritma *random forest*.
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 perangkat lunak yang dapat digunakan untuk mengklasifikasikan data dengan menerapkan algoritma *random forest*. Dalam pembuatan perangkat lunak ini digunakan bahasa pemrograman Delphi 7, dan Microsoft Office Access.
5. Pengujian dan analisa hasil, melakukan uji kebenaran klasifikasi dan mengukur hasil implementasi berdasarkan parameter akurasi yang dihasilkan serta membandingkannya dengan beberapa algoritma klasifikasi lain yakni Naive Bayes, AdaboostM1, RepTree dan Bagging. Untuk membandingkan dengan algoritma klasifikasi lain tersebut, digunakan perangkat lunak WEKA.
6. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir.