

Abstrak

Perkembangan teknologi dan informasi tentang pemodelan data elektronik yang semakin meningkat seperti xml text, html, graph, senyawa kimia dan lain lain menyebabkan jumlah aplikasi untuk memodelkan data tersebut semakin pesat. Salah satu yang populer dan banyak dikembangkan yaitu Graph. Graph sangat powerful karena bisa memodelkan struktur yang kompleks. Salah satu penerapannya yaitu pada *risk assessment*, *toxic prediction* dan *regulatory decision*. Studi mengenai graph berbasis *classification* masih kurang dan untuk penerapannya masih jarang pada struktur molekul kimia sehingga perlu penelitian lebih lanjut guna mendapatkan pemodelan data yang baik. Berbagai penelitian telah dilakukan dengan menggunakan teknik dalam klasifikasi graph salah satunya *Graph Classification* yang bisa diterapkan untuk *chemical compound*.

Dalam penelitian Tugas Akhir ini membahas tentang metode *Graph Classification* yaitu *Graph Boosting* dengan menggunakan algoritma *gSpan* dan *Boosting* dalam melakukan klasifikasi molekul kimia dan menghitung akurasi klasifikasi yang diperoleh. Tujuannya untuk menentukan dan mengidentifikasi apakah suatu molekul kimia mengandung mutagen atau tidak berdasarkan model klasifikasi yang dibuat. Model klasifikasi ini akan membuat *prediction rule* dengan beberapa iterasi untuk mendapatkan pola. Pola ini didapatkan dengan cara mengenumerasi secara frequent kemunculan pola subgraph yang bisa digunakan sebagai feature dalam klasifikasi. Pemilihan teknik yang tepat dan rancangan sistem yang benar akan menghasilkan performansi sistem yang maksimal. molekul kimia dipilih karena keunikannya yaitu memiliki ciri vertex berlabel dan edge yang tidak berarah sehingga molekul kimia cocok jika direpresentasikan dengan graph. Metode *Graph Classification* akan mengklasifikasi graph yang mempunyai karakteristik struktural information serta menggunakan semua subgraph yang terseleksi sebagai set fitur. Hasil dari penelitian ini menunjukkan efisiensi dari algoritma *gSpan* dan *Boosting* untuk molekul kimia dengan akurasi klasifikasi tertinggi yaitu 78,18 %.

Kata kunci: *Graph Classification*, *Frequent Subgraph Mining*, Klasifikasi, *Cheminformatics*,