

BAB 1

PENDAHULUAN

1.1 Latar Belakang

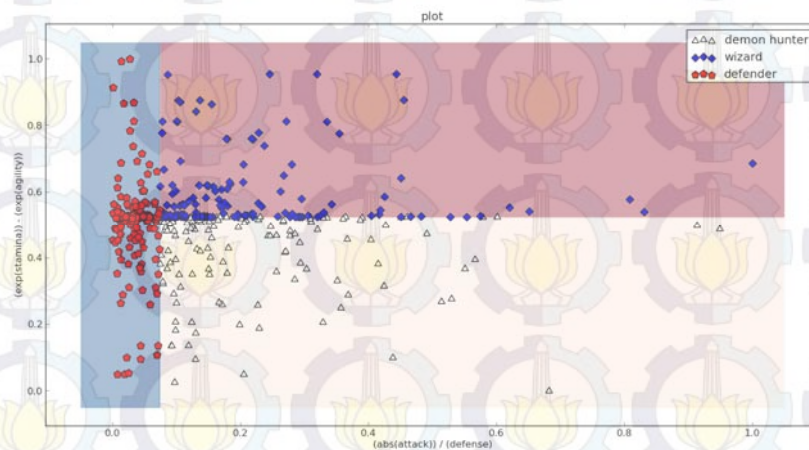
Ekstraksi fitur merupakan salah satu hal yang paling berpengaruh dalam pemecahan masalah klasifikasi. Pemilihan fitur yang tidak baik akan mengakibatkan kesulitan dalam memisahkan kelas-kelas data. Kegagalan pemisahan kelas-kelas data akan berdampak pada turunnya akurasi dalam proses klasifikasi.

Dalam penelitian sebelumnya (Gunawan, Gosaria, & Arifin, 2012), telah dicoba suatu pendekatan ekstraksi fitur dengan menggunakan *grammatical evolution*. Dalam penelitian tersebut, terdapat sebuah kelemahan dikarenakan hanya dilakukan 1 tolak ukur global untuk pengukuran *fitness value*. Hal ini mengakibatkan fitur-fitur yang sebenarnya cukup baik secara khusus, justru tersingkirkan karena nilai *fitness* globalnya rendah. Penelitian-penelitian lain seperti (Gavrilis, Tsoulous, Georgoulas, & Glavas, 2005) dan (Gavrilis, Tsoulous, & Dermatas, Selecting and Constructing Features Using Grammatical Evolution, 2008) juga menggunakan satu nilai *fitness* terhadap satu set fitur. (Li, Zhang, Tian, Mi, Liu, & Ruo, 2011) dan (Guo, Rivero, Dorado, Munteanu, & Pazos, 2011) juga melakukan hal yang hampir sama terhadap kasus yang berbeda.

Dalam penelitian ini diusulkan suatu cara baru dalam penilaian *fitness*. Penilaian *fitness* tersebut akan dilakukan dengan cara mengukur keterpisahan satu kelas terhadap kelas-kelas lain pada tiap dimensi. Metode tersebut, selanjutnya dinamakan *Tatami* karena kemiripannya dengan bentuk lantai tradisional Jepang. Dalam metode ini, untuk memisahkan n buah kelas, maka dibutuhkan maksimal $n-1$ buah fitur (dimensi).

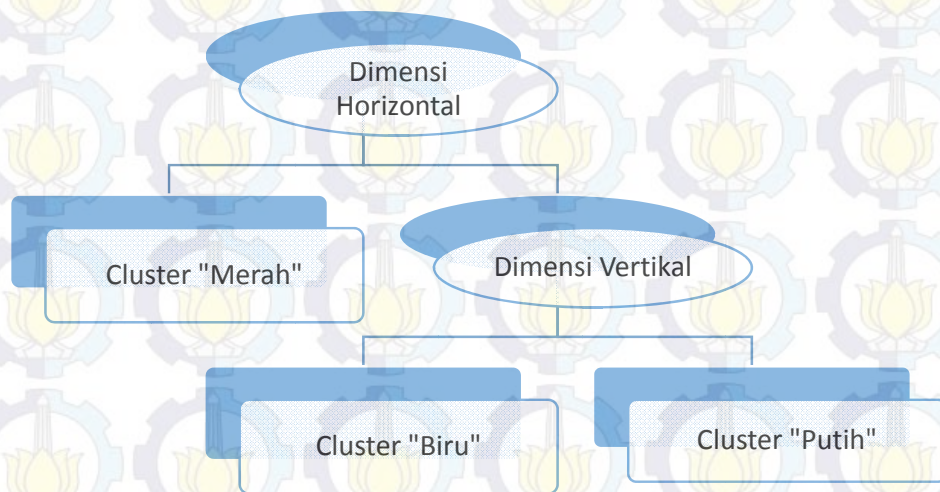
Sebagai contoh, pada Gambar 1.1 terdapat tiga buah cluster yang masing-masing direpresentasikan dengan warna merah, biru dan putih. Pemisahan ketiga cluster tersebut, dapat dilakukan dengan menggunakan 2 buah fitur (dimensi). Dimensi horizontal bertugas untuk memisahkan cluster merah dan kedua cluster lain. Pada dimensi horizontal ini, cluster biru dan putih tidak terpisahkan.

Sedangkan pada dimensi vertikal, cluster biru dan putih terpisahkan, walaupun kedua cluster tersebut tidak terpisah dari cluster merah. Dengan menggunakan kedua dimensi ini, maka akan terbentuk ruang fitur baru di mana cluster merah, biru dan putih terpisah secara linear.



Gambar 1.1 Pemisahan Cluster dengan Metode Tatami.

Adapun pada Gambar 1.1, tampak bahwa pemisahan ketiga cluster tersebut membentuk struktur hirarkikal, seperti yang ditunjukkan dalam gambar 1.2. Struktur hirarkikal tersebut mirip dengan struktur *decision tree*. Sehingga, diharapkan GE Tatami dapat menghasilkan sejumlah fitur terbaik, untuk membantu *decision tree classifier* memisahkan kelas-kelas yang ada secara optimal.



Gambar 1.2 Pemisahan Cluster Secara Hirarkikal

1.2 Perumusan Masalah

Adapun perumusan masalah berdasarkan latar belakang dari penelitian ini adalah:

- Bagaimana menentukan fitur-fitur optimal untuk masalah klasifikasi dengan menggunakan *decision treeclassifier*.
- Bagaimana menerapkan skenario pengukuran *fitness* untuk semua kelas.
- Bagaimana melakukan pengujian atas fitur-fitur yang sudah di-generate.

1.3 Tujuan dan Manfaat

Tujuan penelitian ini adalah menghasilkan dan menguji suatu metode baru (GE Tatami) yang berbasis grammatical evolution untuk mengekstrasi fitur pada data numerik, khususnya yang dimungkinkan untuk terpisah secara hirarkikal.

1.4 Manfaat Penelitian

Hasil penelitian dapat digunakan sebagai salah satu langkah *preprocessing* sebelum melakukan proses klasifikasi data numerik dengan menggunakan *decision tree*.

Dengan melakukan ekstraksi fitur sebagai bagian dalam tahap *preprocessing*, diharapkan proses klasifikasi data dengan fitur-fitur yang tidak

memiliki korelasi langsung terhadap keterpisahan kelas tetap dapat dilakukan dengan baik. Contoh penggunaan disajikan dalam bab empat.

1.5 Kontribusi Penelitian

Kontribusi penelitian ini adalah memberikan mekanisme baru untuk pengekstraksian fitur dengan menggunakan grammatical evolution, khususnya pada data numeric yang dimungkinkan untuk terpisah secara hirarkikal.

1.6 Batasan Penelitian

Batasan masalah pada penelitian ini adalah:

- a. Data yang diproses adalah data numerik. Jika ada data yang bersifat non-numerik, maka perlu dilakukan pengkodean ke dalam bentuk angka terlebih dahulu.
- b. Data yang diproses tidak memiliki *missing attribute*. Artinya, untuk setiap data, nilai semua fiturnya diketahui.
- c. *Grammar* yang digunakan hanya meliputi operator dan fungsi-fungsi matematika umum saja (+, -, *, /, exp, abs, sqr, dan sqrt).
- d. *Classifier* yang digunakan adalah *decision tree*.