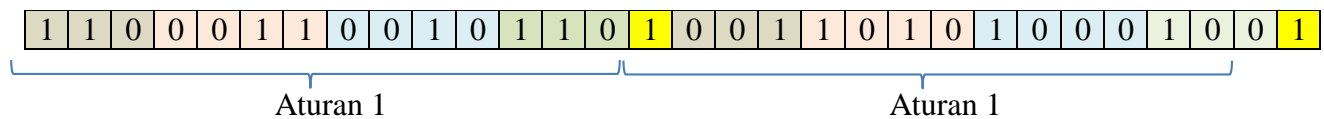


Pembangunan Decision Tree Berdasarkan Algoritma Genetika

1. Identifikasi Masalah

Himpunan data yang diberikan berisi 80 data dengan empat atribut: suhu, waktu, kondisi langit, dan kelembaban menentukan kelas terbang atau tidak. Dalam menentukan hasil akhir dari masalah klasifikasi kasus tersebut, algoritma genetika dapat digunakan untuk membangun pohon keputusan. Sebuah kromosom direpresentasikan sebagai sebuah pohon keputusan yang dapat terdiri atas beberapa aturan. Untuk menentukan kromosom tersebut ditentukan representasi kromosom, populasi, *parent selection*, *crossover*, *mutation*, dan *survivor selection*.

2. Representasi Kromosom



Secara umum, algoritma genetika mengadopsi kode bit dengan panjang tetap; penggunaan yang paling umum adalah kode biner; metode ini menggunakan string yang didasari oleh simbol {0,1} untuk menunjukkan individu. Setiap kode menanggapi atribut kondisi dan nilai atribut akan menentukan panjang penyandian. Sebagai contoh, atribut suhu, waktu, kondisi langit dan kelembaban memiliki kondisi berikut : {'Rendah', 'Normal', 'Tinggi'}, {'Pagi', 'Siang', 'Sore', 'Malam'}, {'Cerah', 'Berawan', 'Rintik', 'Hujan'}, {'Rendah', 'Normal', 'Tinggi'}. Ketika nilainya 0, itu berarti bahwa individu tidak akan mengambil nilai atribut. Ketika nilainya 1, individu akan mengambil nilai atribut. Transformasi metode ini sederhana dan setiap kromosom memiliki panjang yang tetap, yaitu 15 bit untuk satu aturan pohon keputusan. Saya mengatur panjang kromosom di generasi awal adalah 30 bit, dimana memiliki 2 aturan.

3. Parameter Algoritma Genetik

Populasi	30
Probabilitas Crossover	70%
Probabilitas Mutasi	3%
Survivor Selection	Generational Replacement

4. Fungsi Fitness

Tujuan dari setiap sistem klasifikasi adalah akurasi prediksi terbaik untuk sampel tidak berlabel baru. Untuk pohon keputusan, ukuran pohon akhir juga penting. Pohon yang terlalu besar risikonya melengkapi data pelatihan dan generalisasi yang buruk pada sampel baru. Di sisi lain pohon kecil mungkin tidak menangkap informasi penting tentang ruang sampel (Jankowski dan Jackowski 2014). Oleh karena itu dibutuhkan suatu nilai untuk menentukan apakah suatu pohon itu baik atau tidak, yaitu dengan menentukan nilai fitness suatu kromosom sebagai berikut:

$$fitness = \frac{ndata - error}{ndata}$$

5. Operasi Crossover

Crossover yang digunakan di sistem ini adalah *two-point crossover*, pertama kita membangkitkan angka random [0,1] apabila nilai yang dihasilkan kurang dari probabilitas crossover, maka kita memilih

dua titik di parent1, misalnya x_1 dan x_2 . Apabila selisih titik di parent1 melebihi 15, maka kita bisa menentukan kemungkinan titik di parent2. Yaitu, kita pertama menentukan selisih titik $x_2 - x_1$ di parent 1, lalu $m_1 = (x_2 - x_1) \bmod 15$. Dimana kemungkinan titik yang mungkin di parent2 adalah, $\{x_1, x_2\}, \{x_1, x_1 + m_1\}, \{x_2 - m_1, x_2\}$.

6. Hasil

Hasil yang diperoleh setiap running berbeda-beda, kadang semua data uji diklasifikasikan tidak terbang, kadang ya terbang. Akan tetapi fitness kedua kromosom ini adalah 1, dimana nilai akurasi terhadap data train adalah 100%.

<p>Kromosom = [0 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 1 1 0 1 0 0 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1 0 1 0 0 0 1 0 0 0 1 1 1]</p> <p>Hasil Klasifikasi =</p> <pre> 1', 'Ya'] 'Ya'] 'Ya'] 'Ya'] 'Ya'] 'Ya'] , 'Ya'] 'Ya'] 'Ya'] 'Ya'] 'Ya'] 'Ya'] , 'Ya'] , 'Ya'] 'Ya'] 'Ya'] 1', 'Ya'] 'Ya'] 'Ya'] , 'Ya'] </pre>	<p>Kromosom = [1 1 1 1 1 1 1 0 0 0 0 0 1 0 0 1 1 1 0 1 1 0 0 1 0 1 0 1 1 0 0 1 0 0 0 1 0 0 1 0 1 1 0 1 0]</p> <p>Hasil Klasifikasi =</p> <pre> h', 'Tidak'] , 'Tidak'] 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] h', 'Tidak'] , 'Tidak'] , 'Tidak'] , 'Tidak'] </pre>	<p>Kromosom = [1 1 1 1 0 0 1 0 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 1 0 1 1 1 1 0 1 1 1 1]</p> <p>Hasil Klasifikasi =</p> <pre> ['Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Ya', 'Tidak', 'Tidak', 'Ya', 'Ya', 'Tidak', 'Ya', 'Ya', 'Ya', 'Tidak'] </pre>
---	--	--

7. Kesimpulan

Hasil pohon keputusan yang dibangun masih menghasilkan ketidakpastian klasifikasi terhadap data uji. Hal ini mungkin diakibatkan karena ada beberapa aturan di dalam satu kromosom, apabila terdapat suatu aturan yang sudah *true* terhadap data uji, maka klasifikasi menggunakan rule tersebut. Dimana suatu aturan bisa saja hanya mengecek nilai 1 atribut. Selain itu bisa juga pohon yang dibangun overfit terhadap data train. Solusi yang mungkin bisa dilakukan adalah mengganti representasi kromosom yang digunakan dengan memperhatikan nilai information gain yang diperoleh dari suatu cabang seperti artikel yang ditulis oleh Liu dan Fan (2014).

REFERENSI

- Liu D dan Fan S. (2014). “*A Modified Decision Tree Algorithm Based on Genetic Algorithm for Mobile User Classification Problem*”. The Scientific World Journal.
- Dariusz Jankowski, Konrad Jackowski. “*Evolutionary Algorithm for Decision Tree Induction*”. 13th IFIP International Conference on Computer Information Systems and Industrial Management (CISIM), Nov2014, Ho Chi Minh City, Vietnam. pp.23-32.