Algoritman decision trees mengonstruksi pohon keputusan dair sebuah data training yang berupa record-record dalam basis data (Larose, 2005). Algoritma decision trees banyak digunakan karena dapat secara eksplisit menggambarkan suatu pola/pengatuan/informasi dalam bentuk pohon keputusan (Suntoro & Indah, 2017). Algoritma Decision Trees terdiri dari kumpulan node (simpul) yang dihubungkan oleh cabang, cabang tersebut bergerak ke bawah dari root (akar) node akan berakhir di leaf(daun) node. Leaf node adalah node yang sudah tidak dapat dipecah lagi, leaf node merepresentasikan prediksi jawaban dari masalah (data testing). Pohon keputusan decision trees terbentuk terbalik, dimana root node berada dipaling atas, sedangkan leaf node berada paling bawah.

Ada beberapa jenis algoritma didalam decision trees yang sering digunakan : algoritma iterative dichotomiser 3 (ID3), algoritma C4.5, dan algoritma classification and regression tree (CART).

Algoritma CART

Algoritma CART menggunakan perhitungan IndexGini untuk pembentukan cabang, sedangkan untuk pembentukan node, pada algoritma CART digunakan perhitungan GiniGain.

Dataset penerimaan karyawan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Status | Test Tertulis | Hubungan Sosial | Kinerja | Diterima |
| Menikah | Cukup | Buruk | Baik | Ya |
| Menikah | Baik | Buruk | Buruk | Tidak |
| Menikah | Cukup | Baik | Baik | Ya |
| Menikah | Baik | Baik | Baik | Ya |
| Menikah | Kurang | Baik | Baik | Tidak |
| Menikah | Cukup | Buruk | Baik | Ya |
| Menikah | Cukup | Baik | Buruk | Tidak |
| Lajang | Cukup | Baik | Baik | Ya |
| Lajang | Baik | Baik | Buruk | Tidak |
| Lajang | Cukup | Baik | Buruk | Tidak |
| Lajang | Cukup | Buruk | Baik | Tidak |
| Lajang | Cukup | Buruk | Baik | Tidak |

1. Hitung nilai IndexGini

Lakukan hal sama pada nilai dikolom hubungan social, kinerja. Yang manan nanti hasilnya akan dipakai pada gini gain.

1. Hitung nilai Gini Gain

Hasil perhitungan IndexGini

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Atribut | Kategori | Jumlah Kasus | Ya | Tidak | Index Gini |
| Diterima |  | 12 | 5 | 7 | 0,486 |
| Status | Menikah | 7 | 4 | 3 | 0,490 |
|  | Lajang | 5 | 1 | 4 | 0,320 |
| Test Tertulis | Baik | 3 | 1 | 2 | 0,444 |
|  | Cukup | 8 | 4 | 4 | 0,500 |
|  | Kurang | 1 | 0 | 1 | - |
| Hubungan Social | Baik | 7 | 3 | 4 | 0,490 |
|  | Buruk | 5 | 2 | 3 | 0,480 |
| Kinerja | Baik | 8 | 5 | 3 | 0,469 |
|  | Buruk | 4 | 0 | 4 | 0 |

1. Membuat Node dan Cang dari nilai GiniGain Maksimal

Dari nilai GiniGain tersebut dipilih nilai GiniGain maksimal di antara masing-masing atribut. Nilai GiniGain maksimal adalah GiniGain(Kinerja) dengan nilai 0,174 sehingga atribut kinerja menjadi kode akar, seperti ini :

Baik

Buruk

Pohon akar diatas menggambarkan pembentukan node akar, yaitu atribut kinerja. Cabang yang terbentuk berisi baik dan buruk. Pada kategori kinerja buruk, nilai IndexGini adalah 0 sehingga pada cangang buruk terbentuk leaf node. Atribut kinerja buruk, nilai frekuensinya tidak (sejumlah 4) lebih tinggi dibandingkan frekuensi nya (sejumlah 4) sehingga dapat disimpulkan bahwa nilai kelas pada kinerja buruk adalah tidak. Untuk kategori kinerja baik, belum terbentuk leaf node karena nilai IndexGini tidak sama dengan 0 sehingga akan dilakukan perhitungan kembali.

1. Ulangi langkah 1 sampai langkah 3 sehingga semua node terpartisi
2. Perhitungan untuk node 1.1

Node 1.1 akan dilakukan perhitungan lanjut. Untuk mempermudah perhitungan, data pada dataset sebelum difilter dengan mengambil data atribut kinerja yang memiliki nilai baik.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Status | Test Tertulis | Hubungan Sosial | Kinerja | Diterima |
| Menikah | Cukup | Buruk | Baik | Ya |
| Menikah | Cukup | Baik | Baik | Ya |
| Menikah | Baik | Baik | Baik | Ya |
| Menikah | Kurang | Baik | Baik | Tidak |
| Menikah | Cukup | Buruk | Baik | Ya |
| Lajang | Cukup | Baik | Baik | Ya |
| Lajang | Cukup | Buruk | Baik | Tidak |
| Lajang | Cukup | Buruk | Baik | Tidak |

Hitung nilai indexGini masing-masing atribut pada dataset diatas

Lakukan hal sama pada nilai dikolom hubungan social. Yang mana nanti hasilnya akan dipakai pada gini gain.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Atribut | Kategori | Jumlah Kasus | Ya | Tidak | Index Gini |
| Kinerja | Baik | 8 | 5 | 3 | 0,469 |
| Status | Menikah | 5 | 4 | 1 | 0,320 |
|  | Lajang | 3 | 1 | 2 | 0,444 |
| Test Tertulis | Baik | 1 | 1 | 0 | 0 |
|  | Cukup | 6 | 4 | 2 | 0,444 |
|  | Kurang | 1 | 0 | 1 | 0 |
| Hubungan Social | Baik | 4 | 3 | 1 | 0,375 |
|  | Buruk | 4 | 2 | 2 | 0,500 |

Selanjutnya menghitung nilai GiniGain dengan filter kategori = baik.

Baik

Buruk

Baik

Kurang

Cukup

Filter atribut kinerja = baik dan tes tertulis = cukup

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Status | Test Tertulis | Hubungan Sosial | Kinerja | Diterima |
| Menikah | Cukup | Buruk | Baik | Ya |
| Menikah | Cukup | Baik | Baik | Ya |
| Menikah | Cukup | Buruk | Baik | Ya |
| Lajang | Cukup | Baik | Baik | Ya |
| Lajang | Cukup | Buruk | Baik | Tidak |
| Lajang | Cukup | Buruk | Baik | Tidak |

1. Perhitungan untuk node 2.1

Hitung nilai indexGini masing-masing atribut pada dataset Filter atribut kinerja = baik dan tes tertulis = cukup

.

Baik

Buruk

Baik

Kurang

Cukup

Lajang

Menikah

1. Perhitungan untuk node 3.1

Didapatkan atribut hubungan social sebagai node. Pada node hubungan social, terbentuk 2 cabang, yaitu cabang baik dengan kelas ya dan cabang buruk dengan kelas tidak. Pada gambar pohon keputusan diatas terbentuk dari perhitungan dataset penerimaan karyawan lagoritma CART.

Baik

Buruk

Baik

Kurang

Cukup

Lajang

Menikah

Baik

Buruk