**Lampiran**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Charset | Server | Who\_COUNTRY | WHO\_STATEPRO |
| 1 | 1 | 75.0 | 41 | 98 |
| 1 | 3 | 7.0 | 13 | 17 |
| 0 | 6 | 7.0 | 42 | 24 |
| 0 | 1 | 148.0 | 27 | 113 |
| 0 | 3 | 214.0 | 29 | 98 |
| 0 | 3 | 184.0 | 42 | 21 |
| 0 | 3 | 7.0 | 42 | 13 |
| 0 | 6 | 124.0 | 42 | 90 |
| 0 | 6 | 224.0 | 29 | 98 |
| 0 | 4 | 73.0 | 15 | 98 |

10 Data diatas diambil dari dataset keseluruhan dengan Stratified Random Sampling. Dari data diatas akan di lakukan seleksi atribut menggunakan formula information gain yang terdapat pada algoritma C5.0 dimana nilai gain tertinggi akan menjadi node pilihan saat ini. Entropi dari Type nya adalah

Dimana P(Type = 1) = 2 dan P(Type = 0) = 8. Kemudian akan dihitung information gain dari ke 4 atribut tersebut

|  |  |  |
| --- | --- | --- |
| Charset | Type = 1 | Type = 0 |
| 1 | 1 | 1 |
| 3 | 1 | 3 |
| 4 | 0 | 1 |
| 6 | 0 | 3 |

Lakukan perhitungan yang sama juga untuk 3 atribut lainnya, sehingga didapatkan nilai Gain keseluruhan sebagai berikut :

Didapatkan Gain terbesar dari ke 4 atribut tersebut yaitu Who\_Country. Gain tertinggi akan dipilih sebagai node pertama(root) .sehingga Root node saat ini menjadi Who Country. Yang kemudian secara berulang diisi dengan atribut atribut yang gainnya lebih rendah hingga semua leaf terisi oleh Type 1 atau Type 0, atau jika ke 4 atribut telah digunakan sebagai node.