Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

Team project ©2017
Dony Pratidana S. Hum | Bima Agus Setyawan S. IIP

BAB II

LANDASAN TEORI

2.1 Sistem Pendukung Keputusan

Sistem pendukung keputusan adalah sebuah sistem yang dapat membantu manusia untuk mengambil keputusan secara objektif. Konsep sistem seperti ini pertama kali dicetuskan pada tahun 1970-an oleh Michael S.Scott Morton, Michael pertama kali menyebut sistem seperti ini dengan nama *Management Decision System.* Maksud dan tujuan dari adanya sistem pengambil keputusan adalah mendukung pengambil keputusan untuk memilih alternatif keputusan menggunakan model – model pengambilan keputusan dan untuk menyelesaikan masalah yang bersifat terstruktur, semi terstruktur, maupun tidak terstruktur (Mengkepe, 2004).

Menurut Mengkepe (2004) kerangka dasar pengambilan keputusan Manajerial dalam tipe keputusan dibagi menjadi beberapa bagian, yaitu :

1. Terstruktur: Berisi masalah yang sering terjadi, solusinya dapat berupa standar dan baku.
2. Tidak Terstruktur: Berisi masalah kompleks menggunakan pemecahan masalah yang tidak Standard, solusinya melibatkan intuisi manusia sebagai dasar pengambil keputusan.
3. Semi Terstruktur: Gabungan antara keputusan terstruktur dengan tidak terstruktur, solusinya merupakan gabungan antara prosedur solusi standar dengan kemampuan individu manusia.

Implementasi algortima..., Fandy Ferdian Harryanto, FTI UMN, 2016

2.2 Pohon Keputusan

Manusia selalu dihadapkan dengan berbagai macam masalah dari berbagai bidang, masalah ini juga memiliki variasi tingkat kesulitannya. Untuk menghadapi masalah ini manusia mulai mengembangkan sebuah sistem untuk membantu mereka menyelesaikan masalah–masalah ini, salah satu sistem tersebut adalah pohon keputusan (Jefri, 2013).

Pohon keputusan adalah metode klasifikasi dan prediksi yang sudah terbukti *powerfull* dan sangat terkenal. Metode ini berfungsi untuk mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang dapat mudah dimengerti dengan bahasa alami. Proses dari pohon keputusan ini dimulai dari *node* akar hingga *node* daun yang dilakukan secara rekursif dimana setiap percabangan menyatakan kondisi dan setiap ujung pohon akan menyatakan keputusan (Jefri, 2013).

Arsitektur pohon keputusan dibuat sedemikian rupa agar menyerupai pohon asli, dimana terdapat beberapa bagian yaitu (Triisant, 2015):

- *Root Node: Node* ini terletak pada bagian paling atas dari pohon keputusan.
- *Internal Node: Node* ini merupakan percabangan dimana membutuhkan satu input dan mengeluarkan maksimal dua *output*.
- *Leaf Node*: *Node* ini merupakan *node* yang terletak pada ujung pohon. *Node* ini hanya memiliki satu *input* dan tidak memiliki *output.*

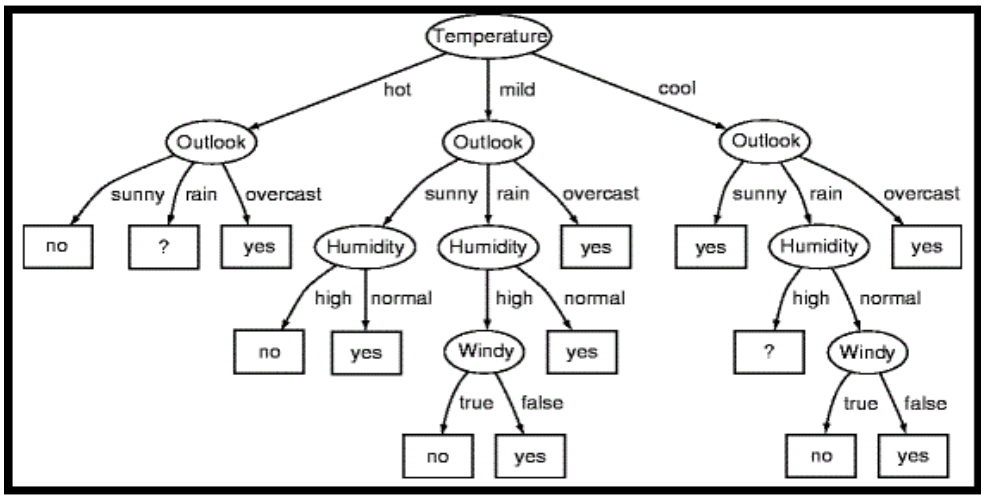
Menurut Triisant (2015) pohon keputusan ini memiliki beberapa kekurangan dan kelebihan, yakni:

1. Kelebihan Pohon Keputusan:

- Daerah pengambilan keputusan yang kompleks dapat diubah menjadi sederhana.
- Dapat menghilangkan perhitungan yang tidak penting karena proses pengujian hanya berdasarkan kriteria yang diperlukan saja.
- Proses pemilihan fitur dari *internal node* yang berbeda lebih fleksibel. Fitur yang telah dipilih ini akan menjadi pembeda antara kriteria yang satu dengan kriteria lainnya.
- Metode ini dapat menghindari munculnya permasalahan dengan cara menggunakan kriteria dengan jumlah yang sedikit pada *node internal* tanpa mengurangi kualitas keputusan yang dihasilkan

2. Kekurangan Pohon Keputusan :
   - Dapat terjadi *overlap* apabila hasil keputusan dan kriteria yang digunakan jumlahnya sangat banyak. Hal ini juga dapat berakibat bertambahnya waktu yang digunakan untuk pengambilan keputusan dan jumlah memori yang dibutuhkan semakin tinggi.
   - Akumulasi jumlah *error* dari setiap tingkat pohon keputusan besar.
   - Mendesain pohon keputusan yang optimal sulit.
   - Kualitas keputusan yang didapatkan sangat tergantung dengan bagaimana pohon tersebut didesain.

Implementasi algortima..., Fandy Ferdian Harryanto, FTI UMN, 2016



Gambar 2.1 Contoh Pohon Keputusan (sumber : https://static-content.springer.com)

2.3 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan teknik pohon keputusan. Algoritma C4.5 merupakan ekstensi dari algoritma ID3 dan menggunakan prinsip *decision tree* yang mirip (Triisant, 2015). Algoritma ini sudah sangat terkenal dan disukai karena memiliki banyak kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan – aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain (Triisant, 2015).

Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi (Jiandi, 2016). Kemudian secara rekursif cabang–cabang pohon diperluas sehingga seluruh pohon terbentuk. Menurut kamus IGI Global (*International Publisher of Progressive Academic), entropy* adalah jumlah data yang tidak relevan terhadap informasi dari suatu kumpulan data. *Gain* adalah

informasi yang didapatkan dari perubahan *entropy* pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu *set* data (Jiandi, 2016).

Berdasarkan apa yang ditulis oleh Jefri (2013), terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu :

1. Memilih atribut sebagai akar
2. Membuat cabang untuk masing – masing nilai
3. Membagi setiap kasus dalam cabang
4. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Menurut Jiandi (2016) data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain.*

$$Entropy(S) = \sum_{i=0}^{n} - pi * \log_2 pi \qquad \text{. . .Rumus (2.1)}$$

Rumus (2.1) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya (Jefri, 2013) :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

pi : Jumlah kasus pada partisi ke-i

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \quad \text{. . .Rumus (2.2)}$$

Rumus (2.2) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy.* Berikut keterangannya (Jefri,2013):

S : Himpunan Kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Dengan mengetahui rumus–rumus diatas, data yang telah diperoleh dapat dimasukkan ke dalam algoritma C4.5 untuk proses pembuatan *decision tree.*

```
FormTree(T)
        (1) ComputeClassFrequency(T);
        (2) if OneClass or FewCases
                Return a leaf;
            Create a decision node N;
        (3) ForEach Attribute A
                ComputeGain(A);
        (4)  N.test = AttributeWithBestGain;
        (5) if N.test is Continuous
                Find Threshold;
        (6) ForEach T1 in the Splitting of T
        (7)     if T1 is Empty
                        Child of N is a leaf
                Else
                        Child of N = FormTree(T1);
        (8) ComputeErrors of N;
            Return N;
```

Gambar 2.2 *Pseudocode* Algoritma C4.5 (Ruggieri, 2002)

Gambar 2.2 merupakan *pseudocode* dari algoritma C4.5 yang berfungsi untuk pembentukan pohon keputusan. Perhitungan dimulai dari menghitung banyaknya jumlah atribut dan menentukan atribut mana yang akan digunakan sebagai akar dari pohon keputusan. Selanjutnya akan dilakukan perhitungan *entropy* dan *gain* untuk menentukan *leaf* dari pohon keputusan tersebut. Setelah semua perhitungan selesai dilakukan, pohon keputusan dapat dibentuk berdasarkan nilai *gain* yang telah dihitung sebelumnya. Atribut dengan nilai *gain*

tertinggi akan terletak pada prioritas yang lebih tinggi dan memiliki kedudukan yang lebih tinggi juga pada pohon keputusan.

2.4 Validasi Silang

Validasi silang merupakan metode statistika yang digunakan untuk melakukan evaluasi dan komparasi terhadap sebuah satu *set* data dengan cara membagikan data tersebut menjadi dua bagian yaitu *data training* dan *data testing*. Salah satu jenis validasi silang adalah *ten-fold cross validation* (Refaeilzadeh dkk, 2008).Validasi ini dilakukan dengan cara membagi suatu *set* data menjadi sepuluh segmen d_0 – d_{10} yang berukuran sama besar dengan cara melakukan pengacakan data. Kemudian d_0 akan digunakan terlebih dahulu untuk proses *training* dan dilakukan validasi menggunakan sisa dari data selain d_0. Setelah itu d_1 akan digunakan untuk *training* sisa data selain d_1 digunakan untuk validasi, dan seterusnya. Dengan melakukan validasi seperti ini maka akurasi yang akan didapatkan akan lebih tinggi (Refaeilzadeh dkk, 2008).