



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.

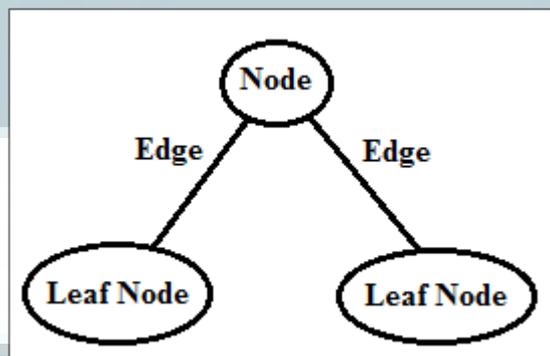
BAB II

LANDASAN TEORI

2.1 Pohon Keputusan

Pohon keputusan atau *decision tree* adalah model prediksi menggunakan struktur pohon atau hirarki. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami sehingga *decision tree* ini merupakan metode klasifikasi yang paling populer digunakan (Ginting, 2014).

Tipikal dari *decision tree* adalah gabungan dari *internal nodes*, *edges*, dan *leaf nodes*. Setiap *internal node* dalam *decision tree* merepresentasikan sebuah *test* dari atribut atau subset atribut, dan setiap *edges* menandakan nilai spesifik atau *range* nilai dari *input* atributnya. Setiap *leaf node* adalah sebuah sambungan *node* dari *tree* berupa *class label* (Dai, 2014).



Gambar 2. 1 Contoh *Tree*

2.2 Data Mining

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis yang menemukan keteraturan, pola, dan hubungan dalam *dataset* berukuran besar. Maksud dari pengertian ini adalah proses pencarian informasi

yang tidak diketahui sebelumnya dari sekumpulan data yang besar. Karakteristik *data mining* adalah sebagai berikut.

- a. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. *Data mining* menggunakan data yang besar. Data yang besar digunakan untuk membuat hasil lebih dipercaya.
- c. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Secara umum ada dua jenis metode pada *data mining*, metode *predictive* dan metode *descriptive*. Metode *predictive* adalah proses untuk menentukan pola dari data yang menggunakan beberapa variabel untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya. Teknik yang termasuk dalam *predictive mining* antara lain klasifikasi, regresi, dan deviasi. Metode *descriptive* adalah proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Teknik *data mining* yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining* (Ginting, 2014).

2.3 Algoritma C4.5

Algoritma C4.5 digunakan untuk membangun pohon keputusan (Widiarto, 2011). Bening (2014) berpendapat bahwa algoritma C4.5 merupakan salah satu algoritma *machine learning*. Dengan algoritma ini, mesin (komputer) akan memberikan sekelompok data untuk dipelajari yang disebut *learning dataset*. Kemudian hasil dari pembelajaran tersebut akan digunakan untuk mengolah data-data yang baru disebut *test dataset*. Karena algoritma C4.5 digunakan untuk

melakukan klasifikasi, jadi hasil dari pengolahan *test dataset* berupa pengelompokan data ke dalam kelas-kelasnya.

Andriani (2014) berpendapat bahwa ada 4 tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5.

1. Mempersiapkan *data training*. *Data training* biasanya diambil dari data *history* yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus 2.1.

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad \text{..Rumus 2.1}$$

Keterangan :

Entropy : Pengukuran yang berdasarkan kemungkinan yang digunakan untuk menghitung jumlah ketidaktentuan.

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi S_i terhadap S

Kemudian hitung nilai *gain* menggunakan rumus 2.2.

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

Keterangan :

Gain : Salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*.

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S

3. Ulangi langkah ke-2 dan langkah ke-3 hingga semua *record* terpartisi.
4. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.

Marwana (2013) berpendapat bahwa kunci pencarian *entropy* ada 2 yaitu sebagai berikut.

1. Jika di antara kolom “benar” atau “tidak” ada yang bernilai 0 (nol) maka *entropy*-nya dipastikan juga bernilai 0 (nol).
2. Jika kolom “benar” dan “tidak” mempunyai nilai yang sama maka *entropy*-nya dipastikan juga bernilai 1 (satu).

2.4 Penjurusan SMA

Kurikulum yang digunakan pada SMA Tarakanita Gading Serpong adalah Kurikulum Tingkat Satuan Pendidikan (KTSP) 2006. Kurikulum Tingkat Satuan Pendidikan (KTSP) adalah kurikulum operasional yang disusun oleh dan dilaksanakan di masing-masing satuan pendidikan, yang berfungsi sebagai

pedoman penyelenggaraan kegiatan pembelajaran untuk mencapai kompetensi yang mencakup tiga domain sikap, pengetahuan, dan keterampilan yang harus terintegrasi, serta dapat menggambarkan kesesuaian dan kekhasan kondisi dan potensi daerah, satuan pendidikan dan peserta didik.

Kriteria penjurusan dan peminatan, lintas minat, dan pendalaman minat adalah berisi tentang kriteria penjurusan yang diberlakukan oleh sekolah yang melaksanakan kurikulum 2006, dengan mengacu kepada panduan penjurusan yang disusun oleh Direktorat Pembinaan SMA (Pengembara, 2013).



UMN