

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Penentuan Jurusan**

Sekolah Menengah Atas (SMA), adalah jenjang pendidikan menengah pada pendidikan formal di Indonesia setelah lulus Sekolah menengah pertama (atau sederajat). Kurikulum yang digunakan pada Sekolah Menengah Atas (SMA) saat ini adalah kurikulum 2013. Penjurusan merupakan upaya untuk membantu siswa dalam memilih jenis sekolah atau program pengajaran khusus atau program studi yang akan diikuti oleh siswa dalam pendidikan lanjutannya. Tujuan dari penjurusan siswa adalah agar siswa dapat memperoleh informasi yang lengkap dan jelas tentang berbagai kemungkinan pilihan yang ada bagi kelanjutan pendidikannya. Sehingga dengan upaya tersebut peserta didik dapat memilih dengan tepat jenis sekolah atau program pengajaran khusus, atau program studi yang ada itu sesuai dengan kemampuan dasar umum (kecerdasan), bakat, minat, kecenderungan pribadi dan hal-hal yang dapat mempengaruhi kelanjutan pendidikannya itu.

Perbedaan individual antara siswa di sekolah di antaranya meliputi perbedaan kemampuan kognitif, motivasi berprestasi, minat dan kreativitas (Snow 1986). Lebih lanjut Snow mengemukakan bahwa oleh karena adanya perbedaan individu tersebut, maka fungsi pendidikan tidak hanya dalam proses belajar mengajar, tetapi juga meliputi bimbingan/konseling, pemilihan dan penempatan siswa sesuai dengan kapasitas individual yang dimiliki, rancangan sistem pengajaran yang sesuai dan strategi mengajar yang disesuaikan dengan karakteristik individu siswa.

Oleh karena itu, sekolah memegang peranan penting untuk dapat mengembangkan potensi diri yang dimiliki siswa. Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penjurusan adalah rendahnya prestasi belajar siswa atau dapat menyebabkan terjadinya kegamangan dalam aktualisasi diri. Tak jarang siswa tidak mengerti alasan pemilihan jurusan tersebut, hendak kemana setelah tamat sekolah dan apa cita-citanya.

Psikolog UI, Indri Savitri, mengemukakan bahwa penjurusan siswa di sekolah menengah tidak saja ditentukan oleh kemampuan akademik tetapi juga harus didukung oleh faktor minat, karena karakteristik suatu ilmu menuntut karakteristik yang sama dari yang mempelajarinya. Dengan demikian, siswa yang mempelajari suatu ilmu yang sesuai dengan karakteristik kepribadiannya (minat terhadap suatu ilmu tertentu) akan merasa senang ketika mempelajari ilmu tersebut (Gupta ML, 2006). Sekolah Menengah Atas (SMA) adalah jenjang pendidikan menengah pada pendidikan formal di Indonesia (Hertyana, 2018).

## 2.2 Tujuan Penjurusan

Penjurusan di SMA ini diadakan atas dasar bahwa para siswa merupakan individu-individu yang mandiri dengan keanerakagamannya (perbedaan individu). Para siswa dijuruskan untuk :

- a) Mengelompokkan para siswa yang mempunyai kecakapan, kemampuan, bakat, dan minat yang relatif sama.
- b) Membantu mempersiapkan para siswa dalam melanjutkan studi dan memilih dunia kerjanya.
- c) Membantu meramalkan keberhasilan untuk mencapai prestasi yang baik dalam kelanjutan studi dan dunia kerjanya.
- d) Membantu memperkokoh keberhasilan dan kecocokan atas prestasi yang akan dicapai di waktu mendatang (kelanjutan studi dan dunia kerjanya).

## 2.3 Decision Tree

Secara singkat bahwa *Decision Tree* merupakan salah satu metode klasifikasi pada *Text Mining*. Klasifikasi adalah “proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu (Han, 2001).

Pohon Keputusan atau dikenal dengan *Decision Tree* adalah salah satu metode klasifikasi yang menggunakan representasi suatu struktur pohon yang berisi alternatif-alternatif untuk pemecahan suatu masalah. Pohon ini juga menunjukkan faktor-faktor yang mempengaruhi hasil alternatif dari keputusan tersebut disertai dengan estimasi hasil akhir bila kita mengambil keputusan

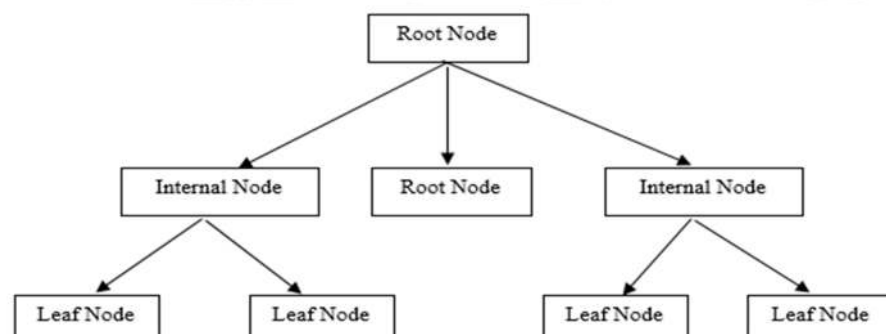
tersebut. Peranan pohon keputusan ini adalah sebagai Decision Support Tool untuk membantu manusia dalam mengambil suatu keputusan.

Manfaat dari *Decision Tree* adalah melakukan break down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga orang yang mengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Konsep yang digunakan oleh *Decision Tree* adalah mengubah data menjadi suatu keputusan pohon dan aturan-aturan keputusan(*rule*).

*Decision Tree* menggunakan struktur hierarki untuk pembelajaran supervised. Proses dari *Decision Tree* dimulai dari *root node* hingga *leaf node* yang dilakukan secara rekursif. Di mana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan kelas dari suatu data.

Pada *Decision Tree* terdiri dari tiga bagian yaitu:

- a. *Root node* *Node* ini merupakan *node* yang terletak paling atas dari suatu pohon.
- b. *Internal node* *Node* ini merupakan *node* percabangan, hanya terdapat satu *input* serta mempunyai minimal dua output.
- c. *Leaf node*. *Node* ini merupakan *node* akhir, hanya memiliki satu *input*, dan tidak memiliki output.



*Decision Tree* merupakan salah satu teknik klasifikasi terhadap objek atau record. Teknik ini terdiri dari kumpulan *decision node*, dan dihubungkan oleh cabang, bergerak ke bawah dari *root node* sampai berakhir di *leaf node* (W, 2007).

## 2.4 Data Mining

Data *mining* mempunyai beberapa metode yang dilakukan pengguna untuk meningkatkan proses *mining* supaya lebih efektif. Oleh karena itu, data *mining* dibagi menjadi beberapa kelompok berdasarkan metodenya, yaitu (Larose D, 2005).

### a. Estimasi

Variabel target pada proses estimasi lebih condong ke arah numerik daripada ke arah kategorikal (*nominal*). Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target dibuat berdasarkan pada nilai prediksi.

### b. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, namun pada prediksi data yang digunakan adalah data rentet waktu (*data time series*) dan hasil nilai akhirnya digunakan untuk beberapa waktu mendatang.

### c. Klasifikasi

Pada klasifikasi terdapat variabel target yang berupa nilai kategorikal (*nominal*). Contoh dari klasifikasi adalah pendapatan masyarakat digolongkan ke dalam tiga kelompok, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Algoritma klasifikasi yang biasa digunakan adalah *Naive Bayes*, *K-Nearest Neighbor*, dan *C4.5*.

### d. Klastering

*Klastering* merupakan pengelompokan data atau pembentukan data ke dalam jenis yang sama. *Klastering* tidak untuk mengklasifikasi, mengestimasi, atau memprediksi nilai, tetapi membagi seluruh data menjadi kelompok-kelompok yang relatif sama (*homogen*). Perbedaan algoritma *klastering* dengan algoritma klasifikasi adalah *klastering* tidak memiliki target/class/label, jadi *klastering* termasuk *unsupervised learning*. Contoh algoritma *klastering* *K-Means* dan *Fuzzy C-Means*.

e. Aturan Asosiasi

Aturan asosiasi digunakan untuk menemukan atribut yang muncul dalam waktu yang bersamaan dan untuk mencari hubungan antara dua atau lebih data dalam sekumpulan data.

## 2.5 Algoritma C45

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 akan membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.

Beberapa kelebihan dari pohon keputusan yaitu :

- a. Hasil analisa berupa diagram pohon yang sangat mudah dimengerti.
- b. Mudah untuk dibangun, serta membutuhkan data percobaan yang lebih sedikit dibandingkan algoritma klasifikasi lainnya.
- c. Mampu mengolah data nominal dan kontinyu.
- d. Model yang dihasilkan dapat dengan mudah dimengerti, beberapa teknik klasifikasi yang berbeda seperti *neural network* menyajikan model dengan informasi logis yang tersirat, sehingga perlu dipelajari.
- e. Menggunakan teknik statistik sehingga dapat divalidasikan.
- f. Cepat dan handal dalam mengolah dataset besar.
- g. Akurasi yang dihasilkan mampu menandingi teknik klasifikasi yang lainnya.

Pada tahap pembelajaran algoritma C4.5 memiliki 2 prinsip kerja :

1. Pembuatan pohon keputusan

Tujuan dari algoritma penginduksi pohon keputusan adalah mengkontruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau record baru yang belum memiliki kelas. C4.5 melakukan konstruksi pohon keputusan dengan metode *divide and conquer*.

Pada awalnya hanya dibuat *node* akar dengan menerapkan algoritma *divide and conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan *gain ratio*, kemudian *node-node* yang

terbentuk di level berikutnya, algoritma *divide and conquer* akan diterapkan lagi sampai terbentuk daun-daun.

## 2. Pembuatan aturan- aturan (*rule set*)

Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan-aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*. Sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*.

Yang menjadi hal penting dalam induksi *Decision Tree* adalah bagaimana menyatakan syarat pengujian pada *node*. Ada 3 kelompok penting dalam syarat pengujian *node* :

### a. Fitur biner

Adalah fitur yang hanya mempunyai dua nilai berbeda. Syarat pengujian kaetika fitur ini menajdi *node* (akar maupun interval) hanya punya dua pilihan cabang.

### b. Fitur kategorikal

Fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Secara umum ada 2 pemecahan yaitu pemecahan biner (*biner splitting*) dan (*multi splitting*).

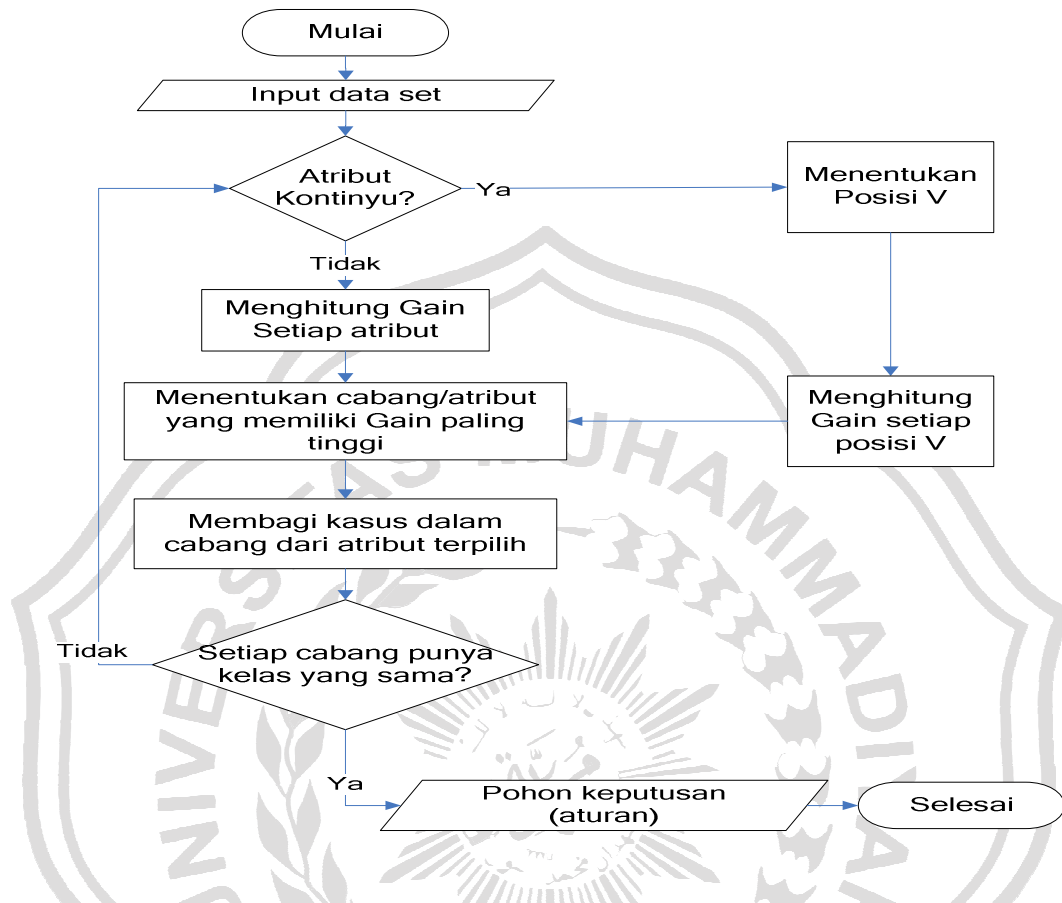
### c. Fitur Numerik

Untuk fitur bertipe numerik, syarat pengujian dalam *node* (akar maupun internal) dinyatakan dengan pengujian perbandingan ( $A \leq V$ ) atau ( $A > V$ ) dengan hasil biner.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut sebagai akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Berikut ini akan dijelaskan secara lebih detail algoritma C4.5 menggunakan *flowcart*



**Gambar 2. 1** Algoritma C45

Untuk memilih atribut sebagai simpul akar (*root node*) atau simpul dalam (*internal node*), *Entropy* merupakan distribusi probabilitas dalam teori informasi dan diadopsi kedalam algoritma C4.5 untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data (*data set*). Semakin tinggi tingkat *entropy* dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Perhitungan *information gain* menggunakan rumus 2.1, sedangkan *entropy* menggunakan rumus 2.2.

$$Gain = Entropy (S) = \sum_i^n = 1 \frac{|S_i|}{s} * Entropy (S_i) \dots\dots(2.1)$$



Keterangan :

S = Himpunan kasus

A = Fitur = jumlah partisi atribut A

$|S_i|$  = proporsi  $S_i$  terhadap S

$|S|$  = jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i} \quad (2.2)$$

Keterangan :

S: himpunan kasus

n: jumlah partisi S

$p_i$ : proporsi dari  $S_i$  terhadap S

Selain *Information Gain* kriteria yang lain untuk memilih atribut sebagai pemecah adalah *Rasio Gain*. Perhitungan rasio *gain* menggunakan rumus 2.3

$$GainRasio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (2.3)$$

Sedangkan untuk perhitungan *split information* menggunakan rumus 2.4.

$$SplitInformation(S, A) = - \sum_{i=1}^c \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (2.4)$$

Dimana  $S_1$  sampai  $S_c$  adalah c subset yang dihasilkan dari pemecahan S dengan menggunakan atribut A yang mempunyai sebanyak c nilai.

## 2.6 Penelitian Sebelumnya

Penelitian sebelumnya dilakukan oleh Nuris Wahidatun Nisak, Mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik yang dilakukan tahun 2016. Penelitian ini berjudul “Prediksi Siswa SD Negeri 2 Sukomulyo Untuk Diterima Di SMP Negeri Menggunakan Metode *Naive Bayes*. Penelitian ini bertujuan untuk menentukan siswa SD 2 Sukomulyo untuk diterima di SMP Negeri atau Tidak. Atribut yang digunakan dalam penelitian ini yaitu nilai rapor siswa, prestasi akademik dan prestasi non akademik. Dengan menggunakan metode *Naive Bayes* diperoleh hasil presentasi sebesar 96,1% dengan jumlah 120 data latih dan 77 data uji.

Sedangkan penelitian lain yang digunakan sebagai rujukan dalam tugas akhir ini adalah penelitian yang dilakukan oleh lailatul Qomariyah, Mahasiswa



Teknik Informatika Universitas Muhammadiyah Gresik yang dilakukan tahun 2016. Penelitian ini berjudul “Klasifikasi Calon Pendorong Darah Dengan Metode *Decision Tree* C4.5 di Kabupaten Gresik (Studi Kasus : PMI Kabupaten Gresik). Penelitian ini bertujuan untuk menentukan apakah calon pendonor darah dapat melakukan donor darah atau tidak. Atribut yang digunakan dalam penelitian ini yaitu usia, kadar hemoglobin, berat badan dan tekanan darah. Data yang digunakan sebanyak 60 data yang diperoleh dari PMI Kabupaten Gresik untuk melakukan pengujian data dilakukan penghitungan nilai akurasi, laju error, sensitivitas dan spesifitas, pengujian dilakukan sebanyak 2 kali dalam 3 variasi data. Variasi data pertama yaitu 30 data latih dan 30 data uji, variasi kedua yaitu 36 data latih dan 24 data uji, dan variasi ketiga sebanyak 48 data latih dan 12 data uji. Dari ketiga variasi percobaan tersebut diperoleh rata-rata nilai akurasi sebesar 86,80%, rata-rata laju error sebesar 13,19%, rata-rata sensitivitas sebesar 91,39%, dan rata-rata spesifitas sebesar 82,22% dari ketiga variasi percobaan tersebut diketahui nilai akurasi tertinggi terjadi pada percobaan dengan 48 data latih dan 12 data uji dengan nilai akurasi mencapai 91,67%.