

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian dengan algoritma C4.5 sudah beberapa kali digunakan pada penelitian sebelumnya, baik di bidang kesehatan, pendidikan dan bidang-bidang lainnya. Pohon keputusan dengan C4.5 sebagai algoritma merupakan suatu metode penambangan data bertujuan memprediksi pemilihan penerima beasiswa untuk siswa dan siswi Sekolah Menengah Atas. Hal tersebut dipandang dari faktor internal dan eksternal siswa, masalah dalam mendapatkan beasiswa adalah beberapa kriteria seperti pendapatan orang tua, tanggung jawab orang tua, dan jurusan, prestasi akademik dan non-akademik. Hasil dan analisis algoritma Algoritma Keputusan C4.5 diterapkan secara akurat untuk prediksi dengan akurasi 94,74% dari nilai akhir siswa sekolah menengah atas (Sugiyarti dkk., 2018).

Algoritma pohon keputusan C4.5 digunakan untuk membangun pohon keputusan sesuai dengan situasi aktual dan faktor non-intelijen pada prestasi belajar bahasa Inggris mahasiswa. Hasil tes menunjukkan bahwa aturan prediksi dapat secara akurat memprediksi rata-rata siswa dan siswa yang kurang berprestasi dalam prestasi belajar bahasa Inggris. Ini dapat digunakan sebagai prediktor yang baik untuk memenuhi persyaratan bantuan pengajaran dan digunakan untuk siswa dengan prestasi belajar Bahasa Inggris dengan nilai rata-rata atau buruk (Li, 2018).

Metode pohon keputusan dengan algoritma C4.5 dibangun metodologi seleksi penalaran kategoris dengan meningkatkan akuisisi pada koefisien keseimbangan menggunakan elemen pengetahuan yang temporer. Metode ini merupakan kunci manajemen temporer dibagi dengan pemilihan model dua tingkat melalui penggunaan model hirarkis. Metode ini merupakan proses konstruksi model teratas berdasarkan penalaran kategoris dan proses pemilihan model yang mendasarinya berdasarkan algoritma pohon keputusan C4.5. Dari hasil tersebut, metode ini meningkatkan akurasi dan ketepatan waktu (Han dkk., 2019).

Algoritma C4.5 dapat digunakan untuk mendeteksi dan mendiagnosis kesalahan dalam sistem *photovoltaic* yang terhubung jaringan. Kumpulan data telah dibagi dalam dua bagian, di mana dalam penerapannya 66% *training dataset* dan 34% sebagai *testing dataset*. Selanjutnya, data baru dicatat untuk mengevaluasi ketahanan, efektivitas dan efisiensi kedua model. Hasil pengujian membuktikan bahwa model memiliki kinerja prediksi yang tinggi dalam deteksi dengan akurasi tinggi sedangkan model diagnosis memiliki akurasi sebesar 99,80% (Benkercha dan Moulahoum, 2018).

Implementasi C4.5 dapat digunakan dalam sebuah proses untuk mendeteksi dan mengklasifikasikan aritmia VT (*Ventricular Tachycardia*) dan VF (*Ventricular Fibrillation*) menggunakan fitur temporal, spektral, dan statistik. Penilaian proses dengan algoritma ini dilakukan pada 57 *data testing* rekaman sinyal elektrokardiogram (EKG) dan hasilnya menunjukkan bahwa metode yang diusulkan mencapai sensitivitas 90,97%, spesifisitas 97,86% dan akurasi 97,02% dalam klasifikasi C4.5, yang lebih baik daripada hasil yang diperoleh algoritma *Support Vector Machine* yang memiliki akurasi 92,23%. Fitur informatif dan algoritma C4.5 dapat menjadi bantuan bagi dokter untuk deteksi secara tepat (Mohanty dkk., 2018).

2.2 Dasar Teori

2.2.1 Sistem Informasi

Sistem informasi merupakan sumber daya data dibuat oleh manusia, mencakup *brainware* atau orangnya, *hardware* atau perangkat kerasnya dan *software* atau perangkat lunaknya. Dengan tujuan memproses, mengumpulkan, menyimpan, mengirim, dan memberikan informasi yang efektif dan efisien untuk memperoleh regulasi/aturan dan manajemen kegiatan suatu organisasi atau instansi. Sistem keputusan merupakan sistem yang terkait secara langsung dengan informasi yang disediakan oleh pendukung keputusan.

Sistem informasi dalam sistem manajemen keputusan membagi ke dalam 2 sistem, yaitu *decision information system* dan *decision support system*. Sistem informasi keputusan menyediakan sistem pengambilan keputusan dengan semua

jenis informasi yang dibutuhkan dengan menggunakan teknologi informasi yang canggih dan modern, seperti komputer dan jaringan.

2.2.2 Kartu Indonesia Pintar (KIP)

Kartu Indonesia Pintar merupakan kartu yang diberikan dari pemerintah dalam bentuk dana pendidikan kepada siswadimana secara ekonomi tidak mampu untuk dapat memperoleh pendidikan di sekolah. Program Indonesia Pintar yang disingkat PIP melalui wujud yang berupa KIP dari pemerintah merupakan program yang digunakan secara khusus dengan memberikan bantuan kepada anak yang tidak mampu secara ekonomi atau dari keluarga miskin agar dapat bersekolah dan tetap mendapatkan layanan pendidikan berupa dana bantuan pendidikan sampai tamat sekolah di tingkat menengah atau sederajat. Selain itu KIP juga digunakan untuk memberikan motivasi siswa yang putus sekolah agar bisa berada di satuan pendidikan secara resmi/formal dan tidak resmi/non formal untuk meneruskan pendidikan, serta mencegah siswa tidak berhenti sekolah (Direktorat Pembinaan Sekolah Menengah Atas, 2015).

Program KIP diberikan kepada anak-anak dengan kriteria umur yang sesuai dengan persyaratan dengan kriteria-kriteria lainnya yang menunjang sebagai penerima program tersebut, sehingga anak sebagai peserta didik akan memperoleh banyak manfaat dari Program Kartu Indonesia Pintar. Menteri Pendidikan dan Kebudayaan menurut instruksi dari Presiden dengan nomor 7 tahun 2014 menyampaikan bahwa KIP sangat penting untuk mencegah siswa putus sekolah, selaras dengan program bantuan pendidikan. Dengan adanya KIP, akses pendidikan bagi anak-anak Indonesia akan diputus kesenjangannya, sehingga alasan untuk tidak bersekolah dapat ditemukan solusi permasalahannya. Program KIP bertujuan untuk menghilangkan hambatan ekonomi yang selama ini menjadi kendala siswa untuk bersekolah, sehingga siswa yang secara ekonomi tidak mampu, tidak lagi akan berpikir untuk berhenti bersekolah. Selain mendorong siswa untuk sekolah, siswa yang telah putus sekolah ditarik kembali agar kembali bersekolah sehingga siswa dapat melanjutkan pendidikan sampai lulus (Direktorat Pembinaan Sekolah Menengah Atas, 2015).

Kriteria penerima KIP dari program PIP adalah anak berusia 6 sampai 21 tahun (Direktorat Pembinaan Sekolah Menengah Atas, 2015), sebagai berikut:

1. Siswa/anak miskin yang terancam putus sekolah, serta keluarga dengan kategori miskin dan kategori rentan miskin.
2. Siswa/anak yang tidak bersekolah dengan usia 6 sampai usia 21 tahun dan putus sekolah.
3. Siswa pemegang KPS yang menerima BSM 2014
4. Siswa/anak dari keluarga peserta PKH yang merupakan program keluarga harapan dari pemerintah.
5. Siswa/anak dari panti asuhan yang memiliki status yatim, status piatu dan status yatim piatu.
6. Siswa/anak dari keluarga yang terkena dampak bencana alam.

Tahap pembuatan klasifikasi ini terlebih dahulu menentukan dan merencanakan kriteria - kriteria dalam golongan KIP. Dalam kriteria ini ada beberapa atribut yang dibutuhkan, yaitu pendidikan yang ditempuh ayah dan ibu, pekerjaan yang dimiliki ayah dan ibu, atribut jumlah tanggungan dalam keluarga, tagihan PBB dan tagihan listrik. Berikut untuk memudahkan dalam proses perhitungan sistem, kriteria dan sub kriteria yang akan dikonversikan dan ditransformasikan dalam kategori/kelas serta dalam bentuk notasi atau simbol, keterangannya ada di tabel 2.1 dan tabel 2.2.

Berikut ini adalah keterangan dari transformasi data siswa ke dalam bentuk data yang berisi atribut dan kategori, yaitu:

1. Pendidikan Ayah

Pendidikan merupakan proses kegiatan belajar dan mengajar yang didapatkan oleh peserta didik dalam mengembangkan bakat dirinya agar mempunyai ilmu, skill, kecerdasan, akhlak mulia yang secara terencana dan sadar dalam mewujudkan terciptanya suasana edukatif yang diperlukan bagi dirinya, masyarakat, bangsa dan negara. Pendidikan dibagi dalam kategori, sebagai berikut: pendidikan dasar/tingkat rendah (SD/MI) dan (SMP/MTs), pendidikan tingkat menengah (SMA, SMK) dan (MA), pendidikan tingkat tinggi (Diploma, Sarjana, Magister, Doktor dan Profesor) (Undang-undang nomor 20 tahun 2003).

2. Pekerjaan Ayah

Pekerjaan merupakan suatu ikatan tugas seseorang yang wajib dikerjakan, supaya mendapat kompensasi berupa upah dan gaji menurut kualifikasi pekerjaannya. Pekerjaan dibagi dalam kategori sebagai berikut: pekerjaan rendah (buruh tani, pedagang kecil, wiraswasta kecil), pekerjaan sedang (karyawan menengah, wiraswasta menengah), pekerjaan tinggi (PNS/TNI/POLRI, pedagang besar, karyawan tinggi) (Sub Direktorat Klasifikasi dan Pembakuan Statistik, Direktorat Metodologi Statistik, 2002).

3. Pendidikan Ibu

Pendidikan merupakan proses kegiatan belajar dan mengajar agar peserta didik secara aktif dalam mengembangkan bakat dirinya agar mempunyai ilmu, skill, kecerdasan, akhlak mulia yang secara terencana dan sadar dalam mewujudkan terciptanya suasana edukatif yang diperlukan bagi dirinya, masyarakat, bangsa dan negara. Pendidikan dibagi dalam kategori, sebagai berikut: pendidikan dasar/tingkat rendah (SD/MI) dan (SMP/MTs), pendidikan tingkat menengah (SMA, SMK) dan (MA), pendidikan tingkat tinggi (Diploma, Sarjana, Magister, Doktor dan Profesor) (Undang-undang nomor 20 tahun 2003).

4. Pekerjaan Ibu

Pekerjaan merupakan suatu ikatan tugas seseorang yang wajib dikerjakan, supaya mendapat kompensasi berupa upah dan gaji menurut kualifikasi pekerjaannya. Pekerjaan dibagi dalam kategori sebagai berikut: pekerjaan rendah (ibu rumah tangga, buruh tani, pedagang kecil, wiraswasta kecil), pekerjaan sedang (karyawan menengah, wiraswasta menengah), pekerjaan tinggi (PNS/TNI/POLRI, pedagang besar, karyawan tinggi) (Sub Direktorat Klasifikasi dan Pembakuan Statistik, Direktorat Metodologi Statistik, 2002).

5. Jumlah Tanggungan Keluarga

Jumlah tanggungan dalam keluarga merupakan tanggungan di dalam keluarga yang terdiri dari istri, anak dan kerabat yang tinggal bersama dalam satu rumah pada satu keluarga yang menjadi banyaknya anggota keluarga. Jumlah keluarga kecil terdiri dari 1 sampai 2 orang, jumlah keluarga sedang terdiri dari 3 orang dan jumlah keluarga besar terdiri dari lebih dari 3 orang (BKKBN, 2015).

6. Tagihan PBB

Tagihan PBB adalah tagihan dalam bentuk pajak yang dibayarkan atas bangunan dan tanah karena adanya keuntungan dan manfaat bagi seseorang. Tagihan PBB dibagi dalam kategori tagihan sebagai berikut: kategori rendah (Rp.10.500 – 50.000), kategori sedang (Rp.50.500 – 150.000), kategori mahal (Rp.150.500 – 300.000) (Direktorat Jenderal Perimbangan Keuangan, 2014).

7. Tagihan Listrik

Tagihan listrik adalah biaya yang wajib dibayar oleh pelanggan listrik setiap bulan. Tagihan listrik dibagi dalam kategori sebagai berikut: kategori rendah (Rp.10.500 – 60.000), kategori sedang (Rp.60.500 – 160.00), kategori mahal (Rp.160.500 – 300.000) (Direktorat Jenderal Ketenagalistrikan, Kementerian Energi dan SDM, 2016).

Tabel 2.1. Kriteria Dalam Penentuan Variabel Siswa Penerima KIP

No	Kode	Kriteria	Keterangan
1	A	Pendidikan Ayah	Variabel Input
2	B	Pekerjaan Ayah	Variabel Input
3	C	Pendidikan Ibu	Variabel Input
4	D	Pekerjaan Ibu	Variabel Input
5	E	Jumlah Tanggungan Keluarga	Variabel Input
6	F	Tagihan PBB	Variabel Input
7	G	Tagihan Listrik	Variabel Input
8	H	Berhak Menerima KIP Tidak Berhak Menerima KIP	Variabel Input

Tabel 2.2.Kriteria dan Sub Kriteria Siswa Penerima KIP

No	Kriteria	Kode	Sub Kriteria	Kategori / Kelas	Simbol
1	Pendidikan Ayah	A1 A2 A3	Putus SD, SD/MI, SMP SMA, SMK, MA D1, D2, D3, D4, S1, S2, S3	Rendah Sedang Tinggi	0 1 2
2	Pekerjaan Ayah	B1 B2 B3	Buruh, Pedagang Kecil, Karyawan Rendah, Wiraswasta Kecil Petani Pemilik Lahan, Wiraswasta Menengah, Karyawan Menengah PNS/TNI/POLRI, Pedagang Besar, Karyawan Atas	Rendah Sedang Tinggi	0 1 2
3	Pendidikan Ibu	C1 C2 C3	Putus SD, SD/MI, SMP SMA, SMK, MA D1, D2, D3, D4, S1, S2, S3	Rendah Menengah Tinggi	0 1 2
4	Pekerjaan Ibu	D1 D2	Ibu Rumah Tangga, Buruh, Pedagang Kecil, Karyawan Rendah, Wiraswasta Kecil Petani pemilik lahan, Wiraswasta	Rendah Menengah	0 1

		D3	Menengah, Karyawan Menengah PNS/TNI/POLRI, Pedagang Besar, Karyawan Atas	Tinggi	2
5	Jumlah Tanggungan Keluarga	E1 E2 E3	>3 Anggota Keluarga 3 Anggota Keluarga 1 - 2 Anggota Keluarga	Banyak Sedang Sedikit	0 1 2
6	Tagihan PBB	H1 H2 H3	Rp.10.500 – 50.000 Rp.50.500 – 150.000 Rp.150.500 – 300.000	Murah Sedang Mahal	0 1 2
7	Tagihan Listrik	I1 I2 I3	Rp.10.500 – 60.000 Rp.60.500 – 160.000 Rp.160.500 – 300.000	Murah Sedang Mahal	0 1 2

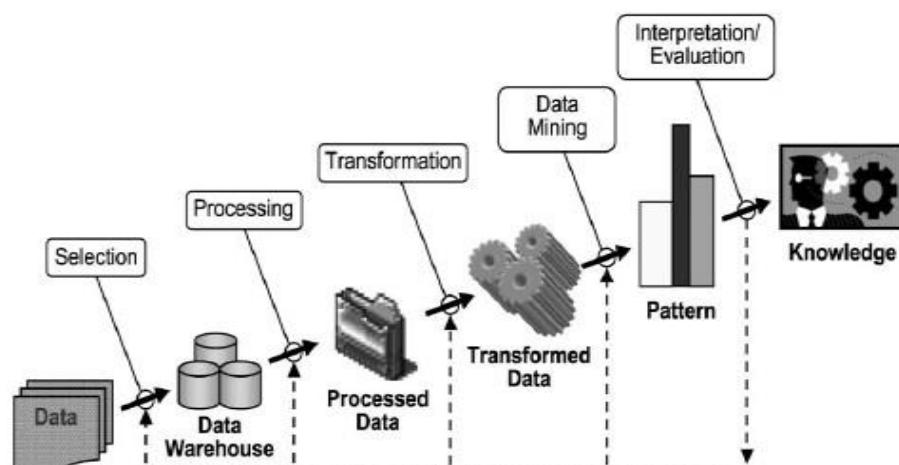
Berdasarkan tabel 2.1 dan 2.2, maka dibentuk tabel klasifikasi siswa penerima Kartu Indonesia Pintar (KIP) yang berisi atribut-atribut yang terdiri dari beberapa atribut sebagai berikut, antara lain atribut pendidikan yang ditempuh oleh ayah dan ibu, atribut pekerjaan yang dimiliki ayah dan ibu, atribut jumlah tanggungan dalam keluarga, atribut tagihan PBB dan atribut tagihan listrik, dengan atribut keputusannya yaitu berhak dan tidak berhak. Keterangan dari penjelasan klasifikasi dengan atribut – atribut yang sudah dibuat dapat ditunjukkan pada tabel 4.1.

2.2.3 Data Mining

Penambangan data atau *data mining* merupakan suatu proses menentukan suatu patron dan informasi dalam jumlah yang besar dari suatu data. Data sumbernya berupa basis data, gudang data, data *internet* yang sistemnya bersifat dinamis dan sebagai tempat penyimpanan informasi lainnya (Han dkk., 2011).

Data mining adalah pengamatan *database* dengan suatu metode baru yang bermanfaat dan bisa dipahami oleh orang yang memiliki data secara analisis untuk meringkas data. Pada dasarnya, *data mining* melakukan proses implementasinya sebagai berikut: prediksi, deskripsi, klasifikasi, estimasi, *clustering* dan asosiasi (Issad dkk., 2019).

Tahapan dalam *data mining* prosesnya dimulai dari menyeleksi data, dari data *target* yang berasal dari data sumber, tahap seleksi dengan perbaikan kualitas data atau *preprocessing*, transformasi untuk menghasilkan data yang berkualitas, serta evaluasi yang akan menghasilkan pengetahuan baru agar dapat memberikan data untuk berkontribusi yang lebih baik. Tahapan dijelaskan dalam gambar prosesnya pada gambar 2.1 (Han dkk., 2011).



Gambar 2.1 Proses *Data Mining*

2.2.4 Klasifikasi

Klasifikasi merupakan suatu proses penemuan suatu model yang berfungsi membedakan konsep atau kelas data berdasarkan data analisis *training dataset* dari data yang kelasnya sudah diketahui objeknya. Mempunyai tujuan agar objek yang tidak diketahui dari label kelasnya akan diprediksi kelasnya. Jenis-jenis

algoritma untuk proses klasifikasi, antara lain algoritma C4.5, *naïve bayes*, algoritma genetika, *k-nearest neighbor*, *rule based method*, , dan *Support Vector Machine (SVM)* (Meng dan Shi, 2020).

Proses pada klasifikasi berdasarkan dari empat komponen, yaitu:

1. *Class* atau Kelas

Kelas merupakan label yang direpresentasikan oleh kategori variabel bersifat dependen yang terdapat pada objek. Contoh: penerima KIP, penerima beasiswa, bahaya penyakit jantung, bahaya penyakit diabetes dan risiko kredit.

2. *Predictor* atau Pemprediksi

Pemprediksi merupakan atribut data yang direpresentasikan oleh variabel bersifat independen. Contoh: jumlah tanggungan dalam keluarga, tagihan PBB, tagihan listrik, besar tekanan darah dan jumlah tabungan.

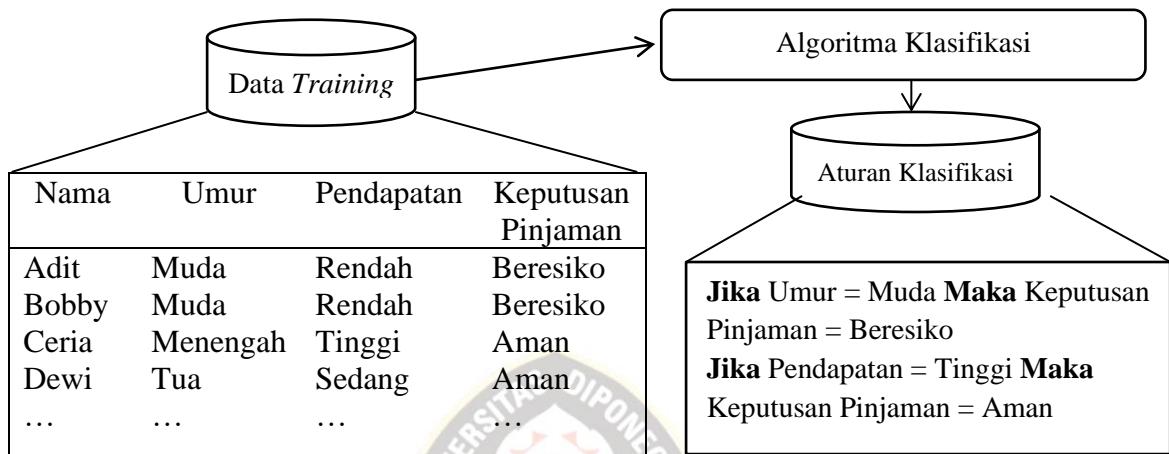
3. *Training Dataset* atau Data Latih

Data latih merupakan data yang mencakup nilai atribut dan kelas yang berfungsi dalam menentukan *predictor* berdasarkan sasaran kelas keputusan atau *decision* yang sesuai.

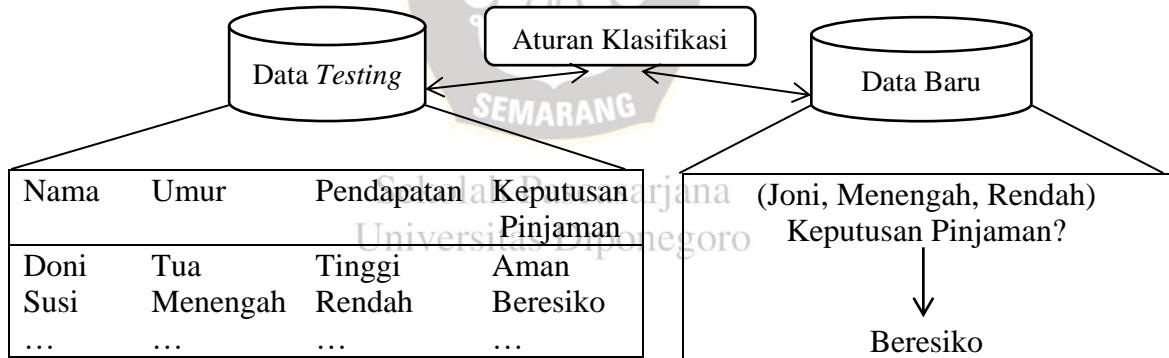
4. *Testing Dataset* atau Data Uji

Data uji merupakan data baru yang akan diklasifikasikan, diuji dan dievaluasi oleh model dari data yang sudah dibuat sebelumnya (Gorunescu, 2011).

Klasifikasi dalam penerapan prosesnya dicontohkan dan dijelaskan di gambar 2.2 dan gambar 2.3. Gambar 2.2 merupakan proses latihan atau *training klasifikasi* dimana data *training* atau data latih dianalisis menggunakan algoritma C4.5. Label kelas dari atribut keputusan kredit, dan mempresentasikan model pengklasifikasian dari pembelajaran sehingga menghasilkan aturan-aturan klasifikasi (*classification rules*). Gambar 2.3 merupakan *testing classification process* atau proses *testing klasifikasi*. Proses data uji klasifikasi digunakan untuk mengestimasi keakurasiannya yang menghasilkan aturan klasifikasi. Apabila akurasi dapat diterima dari penerapan data *training* yang dihitung dengan menggunakan algoritma C4.5 sehingga didapatkan aturan-aturan yang bisa diimplementasikan pada klasifikasi data *testing* atau data uji yang merupakan data baru (Meng dan Shi, 2020).



Gambar 2.2 Proses *Training* Klasifikasi

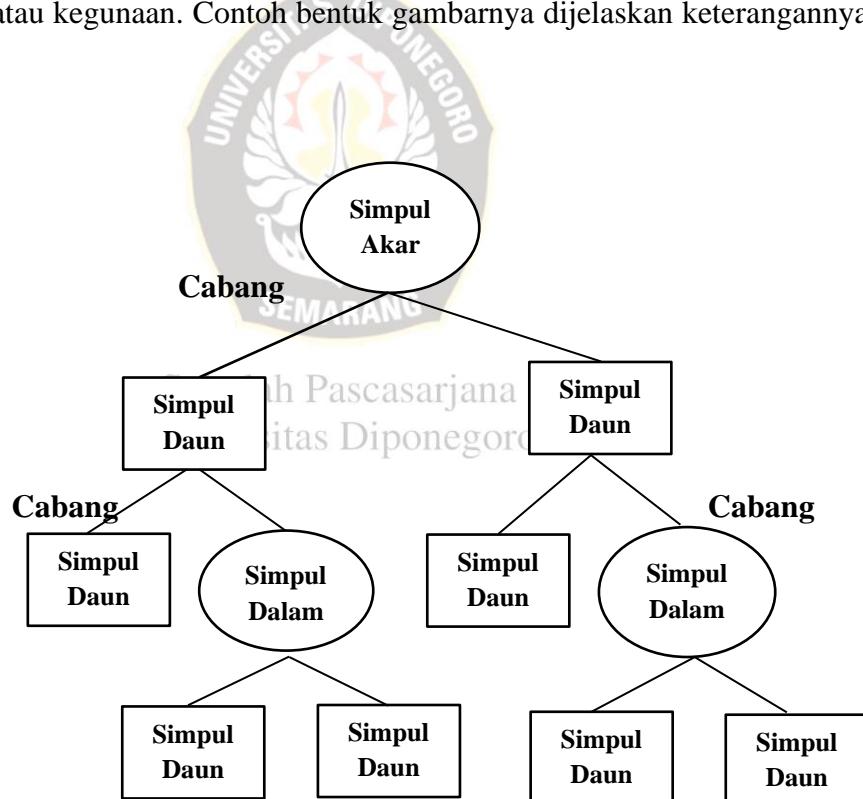


Gambar 2.3 Proses *Testing* Klasifikasi

2.2.5 *Decision Tree*

Decision tree atau pohon keputusan merupakan metode pendukung keputusan dengan menggunakan model keputusan berbentuk seperti pohon atau *tree*. *Decision tree* dapat mengatasi suatu masalah dengan cara memetakan berbagai alternatif yang mungkin, dan terdapat juga kemungkinan faktor-faktor yang dapat dipengaruhi oleh alternatif tersebut beserta estimasi yang ada. Pohon keputusan penggunaannya dapat menampilkan algoritma dimana hanya berisi pernyataan kontrol bersyarat, menjelaskan urutan proses didahului bertahap dari bagian atas sampai bagian bawah yang akan menghasilkan nilai secara bertahap sehingga menghasilkan keputusan yang akurat (Han dkk., 2019).

Decision tree seperti bagan alur (*flowchart*) terstruktur dimana pada setiap cabang dari atribut yang merepresentasikan nilai dari kemungkinan tersebut diwakili oleh setiap simpul dalam, label kelas atau keputusan diambil setelah menghitung semua atribut diwakili oleh setiap simpul daun, kemudian aturan klasifikasi yang telah di tentukan diwakili oleh jalur dari akar ke daun. Dalam analisis keputusan, *decision tree* menggunakan keputusan analitis dan visual sebagai alat pendukung keputusan, diharapkan dari alternatif yang ada akan dihitung nilai atau kegunaan. Contoh bentuk gambarnya dijelaskan keterangannya di gambar 2.4.



Gambar 2.4 *Decision tree*

Pada gambar 2.4, dijelaskan bahwa untuk membuat pohon keputusan/*decision tree*, pertama kali menentukan *root node* (simpul akar), dimana *root node* (simpul akar) adalah simpul paling atas di dalam pohon keputusan. Kemudian membuat *branch* (cabang), dimana tiap – tiap cabang mewakili hasil pengujian dari *training dataset* (data latih). Setelah itu membuat *internal node* (simpul bagian dalam), dimana *internal node* menunjukkan sebuah pengujian di tiap atribut/variabel. Langkah terakhir menentukan *leaf node* (simpul daun, dimana *leaf node* menentukan prediksi klasifikasi (aturan klasifikasi) (Meng dkk., 2020).

2.2.6 Algoritma C4.5

Algoritma C4.5 merupakan jenis algoritma klasifikasi dan prediksi dengan menggunakan teknik *decision tree* dengan kelebihan-kelebihan yang dimilikinya yang terkenal dan sangat disukai di banding teknik klasifikasi dan prediksi yang lain, seperti *naive bayes*, *random forest* dan *key nearest neighbour (KNN)*. Kelebihan – kelebihan yang dimiliki algoritma C4.5 dengan teknik *decision tree* ini, misalnya; nilai atribut yang hilang dapat ditangani, dapat mengolah data diskret dan numerik (kontinyu) pada komputer yang menggunakan memori utama dan mempunyai kemampuan paling cepat di antara algoritma-algoritma lainnya dengan menghasilkan adanya aturan-aturan (*rules*) yang akurat dan mudah untuk diimplementasikan (Hssina dkk., 2014).

Proses klasifikasi membagi dalam dua tahap pengolahan data, yaitu tahap belajar atau pelatihan yang dapat menghasilkan suatu model pelatihan dari data lampau dan tahap klasifikasi dari data *testing* (data uji) yang dapat menghasilkan suatu model untuk digunakan sebagai prediksi klasifikasi kelas. Pada tahap data *training* (pelatihan), dari data pelatihan yang ada berupa *record – record* data, kemudian dibangun metode *decision tree C4.5* untuk menentukan *decision*. Nilai atribut – atribut untuk sebuah kelas berisikan dari kasus yang ada. Setiap atribut atau variabel dapat berupa data diskret atau numerik (kontinyu). Selain itu ketika ada kasus yang tidak mempunyai nilai atribut atau lebih dari data yang ada dapat ditangani oleh algoritma C4.5.

Secara umum urutan-urutan untuk membangun *decision tree C4.5* (Quinlan, 2014) sebagai berikut:

1. Atribut yang akan dijadikan sebagai akar (*root*) dipilih terlebih dahulu.
2. Masing – masing nilai atribut dibuat cabang (*branch*).
3. Proses dari kasus dibagi ke dalam cabang.
4. Tiap-tiap cabang diulangi prosesnya sampai mendapatkan semua kasus pada cabang memiliki kelas yang sama.

Tahap awal yang dilakukan pertama kali untuk membentuk pohon keputusan adalah dengan cara menentukan variabel/atribut mana yang dijadikan akar dari pohon keputusan (*decision tree*) tersebut. Cara menentukan variabel/atribut yang

menjadi akar adalah dengan menggunakan *entropy total*, *entropy each attribute and gain*. Dalam menentukan atribut sebagai akar (*root node*) melalui kriteria atribut didasarkan pada nilai *gain* tertinggi. Untuk menghitung *gain*, nilai *entropy* total dihitung terlebih dahulu, kemudian dihitung *entropy* tiap atribut.

Entropy adalah ukuran untuk bisa mengetahui jenis karakteristik dari *impurity* dan *homogeneity* suatu kumpulan data dari teori informasi (Han dkk., 2011). Rumus *entropy* didefinisikan pada persamaan (2.1), sebagai berikut:

$$E(S) = \sum_{i=1}^n (-p_i) * \log_2 (p_i) \quad (2.1)$$



Keterangan:

E : Entropy

S : Himpunan Kasus

n : Jumlah Partisi S

p_i : Jumlah Sampel Untuk Kelas i / Proporsi Dari Si Terhadap S
Universitas Diponegoro

Kemudian setelah menghitung nilai entropy total dan entropy tiap atribut, maka langkah selanjutnya mencari nilai *gain* dari tiap – tiap atribut yang hasil selanjutnya dicari nilai atribut *gain* yang paling besar untuk dijadikan *root node*.

Gain yaitu ukuran efektivitas data yang diklasifikasikan dari atribut, setelah didapat nilai *entropy* untuk suatu kumpulan data (Hssina dkk., 2014). Rumus *gain* didefinisikan dari persamaan (2.2), sebagai berikut:

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

Keterangan:

G : Gain

S : Himpunan Kasus

A : Atribut

n : JumlahPartisiAtribut A

$|S_i|$: Jumlah Sampel Pada Partisi ke – i

$|S|$: Jumlah Sampel Dalam S

Dengan S menyatakan sebagai himpunan kasus, total kasus pada partisi ke- i dinyatakan oleh S_i , atribut diwakili oleh A, sedangkan n sebagai jumlah partisi. Untuk mencari nilai *information gain* dilakukan dengan menentukan terlebih dahulu nilai *entropy*.

Berikut adalah contoh sebuah data klasifikasi yang ditunjukkan pada tabel 2.3 untuk menentukan suatu *decision* (keputusan) bermain bulutangkis atau tidak dari seseorang, ditentukan oleh beberapa atribut diantaranya yaitu atribut cuaca, atribut suhu, atribut kelembaban, dan atribut angin. Penghitungan data ini akan digunakan metodenya sebagai contoh untuk menghitung klasifikasi siswa penerima Kartu Indonesia Pintar (KIP) yang digunakan di dalam penelitian.

Tabel 2.3 Contoh Data Klasifikasi

No	Cuaca	Suhu	Kelembaban	Angin	Keputusan Bermain
1	Hujan	Sejuk	Tinggi	Kencang	Tidak
2	Cerah	Panas	Tinggi	Tidak Kencang	Tidak
3	Berawan	Panas	Tinggi	Kencang	Ya
4	Hujan	Sejuk	Tinggi	Kencang	Ya
5	Hujan	Dingin	Sedang	Kencang	Ya
6	Hujan	Dingin	Sedang	Tidak Kencang	Ya
7	Berawan	Dingin	Sedang	Tidak Kencang	Ya
8	Cerah	Sejuk	Tinggi	Kencang	Tidak
9	Cerah	Dingin	Sedang	Kencang	Ya
10	Hujan	Sejuk	Sedang	Kencang	Ya
11	Cerah	Sejuk	Sedang	Kencang	Ya
12	Berawan	Sejuk	Tinggi	Kencang	Ya
13	Berawan	Panas	Sedang	Tidak Kencang	Ya
14	Cerah	Panas	Tinggi	Kencang	Tidak

Berdasarkan tabel 2.3, akan dijadikan model/pola dari proses pembuatannya, yang fungsinya memprediksi klasifikasi sebagai kategori/klasifikasi yang belum dimiliki oleh data yang baru (Hssina, 2014). Adapun tahapan dari prosesnya sebagai berikut: pertama hitung nilai *gain* dari tiap-tiap atribut (Cuaca, Suhu, Kelembaban, Angin) kemudian dicari dulu *gain* yang terbesar, yang nanti

dijadikan akar (*root node*). Langkah awal dihitung nilai *entropy* total dan berikutnya menghitung nilai dari *entropy* masing-masing kategori dalam setiap atribut untuk menghitung *gain* dengan rumus (2.1), sebagai berikut:

$$E(S) = \sum_{i=1}^n (-p_i) * \log_2(p_i)$$

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

$$\log_2(x) = \frac{\ln(x)}{\ln(2)}$$



Kemudian menghitung nilai dari *gain* memakai rumus (2.2), sebagai berikut:

$$G(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Sekolah Pascasarjana

Universitas Diponegoro

Berdasarkan hasil penghitungan dari nilai *gain*, kemudian membuat *root node* (simpul akar) dengan cara memilih nilai *gain* terbesar untuk tahapan proses membuat *decision tree* atau pohon keputusan.