

BAB II

LANDASAN TEORI

2.1 Konsep Beras Miskin

Beras Miskin adalah bagian dari program penanggulangan kemiskinan yang berada pada kluster I, yaitu kegiatan perlindungan sosial berbasis keluarga dalam pemenuhan kebutuhan pangan pokok bagi masyarakat kurang mampu. Raskin mempunyai multi fungsi, yaitu memperkuat ketahanan pangan keluarga miskin, sebagai pendukung bagi peningkatan kualitas Sumber Daya Manusia (SDM), pendukung usaha tani padi dan sektor lainya dan peningkatan pemberdayaan ekonomi daerah. Disamping itu Raskin berdampak langsung pada stabilitas harga beras, dan juga berperan dalam menjaga stabilitas ekonomi nasional.

Program Raskin dalam penelitian ini adalah adalah program penanggulangan kemiskinan kluster satu, termasuk program bantuan sosial berbasis keluarga yang sudah berjalan secara rutin sejak tahun 1998. Melalui program ini Pemerintah memberikan bantuan kepada masyarakat miskin untuk mendapatkan hak atas pangan. Jika rata-rata konsumsi beras nasional sekitar 113,7 kg/per kapita/tahun dan setiap RTS-PM terdiri atas empat jiwa, maka program Raskin telah memberikan kontribusi sebesar 39,6 % dari kebutuhan beras setiap bulannya bagi setiap RTS.

Program Raskin tergolong program nasional. Program ini melibatkan berbagai pihak baik vertikal maupun horzontal. Secara horizontal semua sektor terkait memiliki tanggung jawab dan wewenang sesuai dengan tupoksinya masing-masing dalam pelaksanaan program Raskin. Issue aktual yang terkait secara horizontal adalah penetapan data dari RTS. Program Raskin mengacu pada data RTS hasil PPLS-2012 BPS, yang di tetapkan oleh TNP2K menggunakan sistem *basic* data terpadu perlindungan sosial. Secara vertikal program Raskin bukan semata-mata program pusat semata, dikutip dari Pedoman Umum Penyaluran RASKIN yang diterbitkan Kementerian Koordinator Bidang

Kesejahteraan Rakyat Republik Indonesia Tahun 2012 menerangkan bahwa Pemerintah daerah memiliki tanggung jawab secara proporsional. Dalam hal ini Pemerintah pusat berperan dalam membuat kebijakan nasional, sedangkan dalam pelaksanaan dan penyalurannya sangat tergantung pada peran Pemerintah daerah. Pemerintah pusat memberikan subsidi pembelian beras yang dilaksanakan oleh Perum BULOG untuk disalurkan sampai titik distribusi (TD). Untuk selanjutnya Pemerintah daerah menyampaikan beras tersebut kepada RTS-PM dengan 6 Tepat (Tepat Sasaran, Tepat Jumlah, Tepat Mutu, Tepat Waktu, Tepat Harga, dan Tepat Administrasi). Oleh karena itu pelaksanaan program Raskin sangat tergantung pada peran Pemerintah daerah seperti sosialisasi, pengawasan mutu, angkutan, dan biaya operasional dll.

Syarat untuk penerima Raskin yang digunakan di Desa Sidomulyo Kecamatan Deket Kabupaten Lamongan yaitu mengacu pada hasil klarifikasi data dari penelitian sebelumnya oleh Friday Aries L W, dengan mencakup kriteria yaitu penghasilan, harta benda (kendaraan), kondisi rumah, dan jumlah keluarga.

2.2 Pengertian

- a. Rumah Tangga Sasaran Penerima Manfaat (RTS-PM) adalah rumah tangga miskin di Desa/Kelurahan yang berhak menerima Raskin dan terdaftar dalam Daftar Penerima Manfaat (DPM-1) yang ditetapkan oleh Kepala Desa/Lurah sebagai hasil dari Musyawarah Desa/Kelurahan dan disahkan oleh Camat sesuai dengan pendataan PPLS-12 BPS tahun 2012.
- b. Musyawarah Desa/Kelurahan merupakan forum pertemuan musyawarah di tingkat Desa/Kelurahan yang melibatkan aparat Desa/Kelurahan, kelompok masyarakat Desa/Kelurahan dan perwakilan RTS-PM Raskin dari setiap Satuan Lingkungan Setempat (SLS) setingkat Dusun/RW untuk menetapkan daftar nama RTS-PM.
- c. Titik distribusi (TD) adalah tempat atau lokasi penyerahan beras Raskin dari Satker Raskin kepada pelaksana distribusi Raskin di tingkat Desa/Kelurahan, atau lokasi yang disepakati secara tertulis oleh

Pemerintahan Kabupaten/Kota dengan Divre/Subdivre/Kansilog Perum BULOG.

- d. Titik bagi (TB) adalah tempat atau lokasi penyerahan beras Raskin dari pelaksana distribusi Raskin kepada RTS-PM.
- e. Pelaksana Distribusi Raskin adalah kelompok kerja (Pokja) di TD atay Warung Desa (Wardes) atau Kelompok Masyarakat (Pokmas) yang ditetapkan oleh kepala Desa/Lurah.
- f. Kelompok Kerja (Pokja) adalah sekelompok masyarakat Desa/Kelurahan yang terdiri dari aparat Desa/Kelurahan, ketua RT/RW/RK dan beberapa orang yang ditunjuk dan ditetapkan oleh kepala Desa/Lurah sebagai pelaksana distribusi Raskin.
- g. Warung Desa (Wardes) adalah lembaga ekonomi di Desa/Kelurahan, baik milik masyarakat, koperasi maupun pemerintah Desa/Kelurahan yang memiliki fasilitas bangunan/tempat penjualan bahan pangan dan barang lainnya yang ditetapkan oleh Tim Koordinasi Raskin Kabupaten/Kota sebagai tempat penyerahan beras Raskin dari Satker Raskin.
- h. Kelompok Masyarakat (Pokmas) adalah lembaga masyarakat dan/atau kelompok masyarakat di Desa/Kelurahan yang ditetapkan oleh Kepala Desa/Lurah sebagai Pelaksana Distribusi Raskin.
- i. Padat Karya Raskin adalah sistem penyaluran Raskin kepada RTS-PM yang dikaitkan dengan pemberdayaan masyarakat dimana para RTS-PM diwajibkan bekerja dan meningkatkan produktivitas daerah dengan diberikan kompensasi pembayaran HBD Raskin oleh Pemerintah daerah melalui APBD.
- j. Satker Raskin adalah satuan kerja pelaksana penyaluran Raskin yang dibentuk oleh Divisi Regional (Divre) / Sub Divisi Regional (Subdivre) / Kantor Seksi Logistik (Kansilog) Perum BULOG terdiri dari ketua dan anggota yang diangkat dengan surat perintah (SP) Kadivre/Kasub Divre/Kakansilog.

- k. Kualitas Beras adalah beras medium kondisi baik sesuai dengan persyaratan kualitas beras yang diatur dalam Inpres Kebijakan Perberasan yang berlaku.
- l. SPA adalah surat Pemerintahan alokasi yang di buat oleh Bupati/Walikota atau ketua tim Koordinasi Raskin Kab/Kota atau pejabat yang ditunjuk oleh Bupati/Walikota kepada Kadivre/Kasubdivre/Kakansilog berdasarkan alokasi Pagu Raskin dan rincian masing-masing Kecamatan dan Desa/Kelurahan.
- m. Surat Perintah Penyerahan Barang (SPPB)/Delively Order (DO) adalah perintah tertulis yang diterbitkan oleh Kadivre/Kasubdivre/Kakansilog atau pejabat lain yang berwenang kepada kepala gudang untuk mengeluarkan dan menyerahkan barang kepada pihak lain.
- n. BAST adalah berita acara serah terima beras Raskin berdasarkan SPA dari Bupati/Walikota dan ditandatangani antara Perum BULOG dan pelaksana distribusi.
- o. DPM-1 adalah model daftar penerima manfaat Raskin di Desa/Kelurahan.
- p. DPM-2 adalah model daftar penjualan Raskin di Desa/Kelurahan.
- q. HPB adalah harga penjualan beras secara tunai sebesar Rp. 1600/kg netto di TD.
- r. MBA-0 adalah model rekap BAST ditingkat Kecamatan.
- s. MBA-1 adalah model rekap MBA-0 ditingkat Kabupaten/Kota.
- t. MBA-2 adalah model rekap MBA-1 ditingkat Provinsi.
- u. TT-HP Raskin adalah model tanda terima uang hasil penjualan Raskin dari pelaksana distribusi kepada Satker Raskin.
- v. UPM adalah unit pengaduan masyarakat.
- w. PPLS-12 adalah pendataan program perlindungan sosial tahun 2012 yang dilakukan oleh Badan Pusat Statistik (BPS).

2.3 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang

menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk. 2005).

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Daniel T. Larose, 2005).

2.3.1 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Daniel T. Larose, 2005):

1. Deskripsi

Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning*.

3. Estimasi

Estimasi hampir sama dengan klasifikasi, perbedaannya adalah variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

4. Prediksi

Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, perbedaannya adalah hasil dari prediksi akan ada dimasa mendatang.

Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi.

5. Klastering

Klastering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Klastering menggunakan *unsupervised learning*.

6. Asosiasi

Tugas asosiasi atau sering disebut juga sebagai *market basket analysis* dalam data mining adalah menemukan relasi atau korelasi diantara himpunan item-item dan menemukan atribut yang muncul dalam satu waktu. Asosiasi menggunakan *unsupervised learning*. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* dan *confidence*.

Metode yang akan digunakan pada penelitian ini termasuk kedalam kelompok prediksi, karena menggunakan teknik klasifikasi yang hasilnya akan ada dimasa mendatang.

2.3.2 Tahap – Tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar, Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung dengan perantaraan *knowledge base*. Tahap – tahap data mining ada 6 yaitu :

1. Pembersihan data (*data Cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Pada umumnya data diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data

mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Intregasi data (*data integraton*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asoisasi dan clustering hanya bisa menerima input data kategorikal, Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*pattren evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining yang lebih sesuai, atau menerima hasil ini sebagai suatu hal yang diluar dugaan yang mungkin bermanfaat.

2.4 *Decision Tree* (Pohon Keputusan)

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Selain itu dapat diekspresikan dalam bentuk bahasa basis data seperti *Structure Query Language* untuk mencari *record* pada kategori tertentu (Kusrini dan Emha, 2009).

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan variabel target.

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan, dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain.

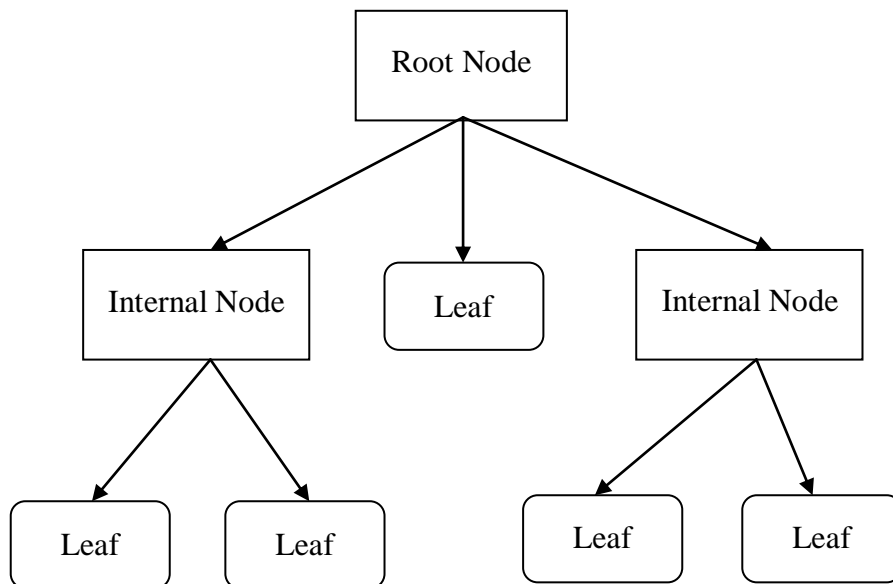
2.4.1 Model *Decision Tree*

Decision tree adalah *flow-chart* seperti *struktur tree*, dimana tiap *internal node* menunjukkan sebuah test pada sebuah atribut, tiap cabang menunjukkan hasil dari test, dan *leaf node* menunjukkan *class-class* atau *class distribution*.

Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

- Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Contoh dari model pohon keputusan yaitu seperti pada **gambar 2.1** berikut:



Gambar 2.1 Model *Decision Tree*

2.4.2 Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3. Dalam ID3, induksi decision tree hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan (Eko Prasetyo, 2014).

Yang menjadi hal penting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node :

1. Fitur biner

Adalah Fitur yang hanya mempunyai dua nilai berbeda. Syarat pengujian ketika fitur ini menjadi node (akar maupun internal) hanya punya dua pilihan cabang.

2. Fitur kategorikal

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

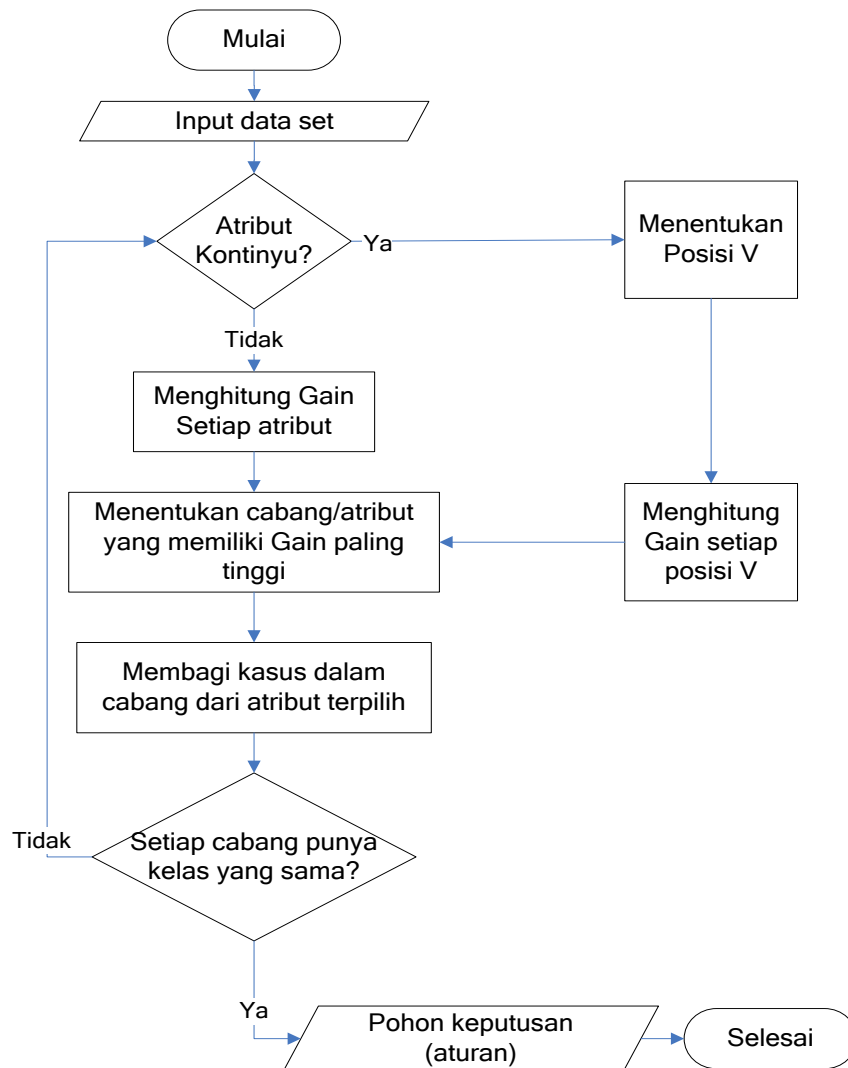
3. Fitur numerik

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:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. 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* yang disajikan pada **gambar 2.2**



Gambar 2.2 Flowchart algoritma Decision Tree C4.5

Untuk memilih atribut sebagai simpul akar (*root node*) atau simpul dalam (*internal node*), didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada. Sebelum perhitungan *information gain*, akan dilakukan perhitungan *entropy*. *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.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.1)$$

dimana,

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (2.2)$$

dimana,

S : Himpunan kasus

A : Fitur

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, sedangkan *split information* menggunakan rumus 2.4.

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

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

dimana S₁ sampai S_c adalah c subset yang dihasilkan dari pemecahan S dengan menggunakan atribut A yang mempunyai sebanyak c nilai.

Untuk mengukur nilai akurasi yang didapat dari hasil pengujian, menggunakan rumus 2.5. Sedangkan untuk mengukur tingkat kesalahannya menggunakan rumus 2.6.

$$Akurasi = \frac{Jumlah\ h\ data\ yang\ diprediksi\ secara\ benar}{Jumlah\ h\ prediksi\ yang\ dilakukan} \quad (2.5)$$

$$Laju\ error = \frac{Jumlah\ h\ data\ yang\ diprediksi\ secara\ salah}{Jumlah\ h\ prediksi\ yang\ dilakukan} \quad (2.6)$$

Sensitivitas akan mengukur proporsi positif asli yang dikenali (diprediksi) secara benar sebagai positif asli. Rumus perhitungannya menggunakan rumus 2.7. Sedangkan spesifisitas akan mengukur proporsi negatif asli yang dikenali (diprediksi) secara benar sebagai negatif asli. Rumus perhitungannya menggunakan rumus 2.8.

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \quad (2.7)$$

Keterangan :

TP : Keterangan terima yang diprediksi secara benar sebagai Keterangan terima

FN : Keterangan terima yang diprediksi secara salah sebagai Keterangan tidak terima

$$\text{Spesifisitas} = \frac{TN}{FP + TN} \quad (2.8)$$

Keterangan :

TN : Keterangan tidak terima yang diprediksi secara benar sebagai Keterangan tidak terima

FP : Keterangan tidak terima yang diprediksi secara salah sebagai Keterangan terima

2.4.3 Contoh Perhitungan

Berikut ini akan dijelaskan ilustrasi dari alur proses perhitungan algoritma *Decision Tree C4.5*. Data set yang digunakan pada contoh ini adalah data untuk menentukan *Play* atau *Don't Play* dengan beberapa atribut yaitu atribut *outlook*, *temperature*, *humidity*, dan *windy*. Dimana atribut *temperature* dan *humidity* bertipe kontinyu sedangkan *outlook* dan *windy* bertipe kategorikal. Sedangkan kolom *Class* adalah kelas tujuannya atau label kelas-nya.

Tabel 2.1 Contoh data set

Outlook	Temperature	Humidity	Windy	Class
sunny	75	70	TRUE	Play
sunny	80	90	TRUE	Don't Play

sunny	85	85	FALSE	Don't Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
overcast	72	90	TRUE	Play
overcast	83	78	FALSE	Play
overcast	64	65	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play
rain	65	70	TRUE	Don't Play
rain	75	80	FALSE	Play
rain	68	80	FALSE	Play
rain	70	96	FALSE	Play

Pada contoh ini rumus yang digunakan untuk memilih atribut sebagai *node* adalah rumus *information gain*. Proses pertama adalah menghitung *entropy* untuk semua data.

Jumlah class play = 9

Jumlah class don't play = 5

Berikut adalah perhitungan *entropy* untuk semua data:

$$\begin{aligned} Entropy(S) &= -\frac{9}{14} * \log_2\left(\frac{9}{14}\right) - \frac{5}{14} * \log_2\left(\frac{5}{14}\right) \\ &= 0.940 \end{aligned}$$

Selanjutnya menghitung *gain* untuk setiap atribut. Berikut adalah contoh perhitungan gain untuk atribut *outlook*:

Tabel 2.2 Distribusi jumlah atribut *outlook*

Nilai Outlook	Σ Play	Σ Don't Play	Total
Sunny	2	3	5
Overcast	4	0	4
Rain	3	2	5

Berdasarkan tabel 2.2, maka nilai *information gain* untuk atribut *outlook* adalah sebagai berikut:

$$Gain(outlook) = 0.940 - \left(\frac{5}{14} * \left(-\frac{2}{5} * \log_2\left(\frac{2}{5}\right) - \frac{3}{5} * \log_2\left(\frac{3}{5}\right) \right) \right)$$

$$\begin{aligned}
& + \frac{4}{14} * \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4} \right) - \frac{0}{4} * \log_2 \left(\frac{0}{4} \right) \right) \\
& + \frac{5}{14} * \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} * \log_2 \left(\frac{2}{5} \right) \right) \\
& = 0.940 - 0.694 \\
& = 0.246
\end{aligned}$$

Untuk perhitungan atribut yang bertipe kontinyu, harus menentukan *posisi* V terbaik yang dinyatakan dalam perbandingan ($A \leq V$) atau ($A > V$). Berikut akan dijelaskan contoh perhitungan dari atribut *temperature*.

Misal posisi V yang akan digunakan pada atribut *temperature* adalah 65,70,75,dan 80, kemudian dihitung nilai *information gain*-nya.

Contoh perhitungan *temperature* posisi $v=65$:

$$\begin{aligned}
Gain(temp) & = 0.940 - \left(\frac{2}{14} * \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) \right. \\
& \quad \left. + \frac{12}{14} * \left(-\frac{8}{12} * \log_2 \left(\frac{8}{12} \right) - \frac{4}{12} * \log_2 \left(\frac{4}{12} \right) \right) \right) \\
& = 0.940 - 0.930 \\
& = 0.010
\end{aligned}$$

Berikut hasil perhitungan atribut numerik untuk setiap posisi yang telah ditentukan :

Tabel 2.3 Hasil perhitungan posisi V untuk atribut *temperature*

Temperature	65		70		75		80	
	\leq	$>$	\leq	$>$	\leq	$>$	\leq	$>$
Play	1	8	4	5	7	2	7	2
Don't Play	1	4	1	4	3	2	4	1
Jumlah	2	12	5	9	10	4	11	3
Entropy	1.000	0.918	0.722	0.991	0.881	1.000	0.946	0.918
Gain	0.010		0.045		0.025		0.0005	

Berdasarkan tabel 2.3, nilai gain tertinggi adalah 70, maka nilai *information gain* pada atribut *temperature* adalah 0.045. Hasil perhitungan pada setiap atribut disajikan pada tabel 2.4.

Tabel 2.4 Hasil perhitungan *Information gain* untuk setiap atribut

		Jumlah	Play	Don't Play	Entropy	Gain
Total		14	9	5	0.940	
Outlook	Sunny	5	2	3	0.971	0.247
	Overcast	4	4	0	0.000	
	Rain	5	3	2	0.971	
Temperature	≤ 70	5	4	1	0.722	0.045
	> 70	9	5	4	0.991	
Humidity	≤ 80	9	7	2	0.764	0.102
	> 80	5	2	3	0.971	
Windy	TRUE	6	3	3	1.000	0.048
	FALSE	8	6	2	0.811	

Berdasarkan tabel 2.4 menunjukkan bahwa atribut *outlook* memiliki nilai gain tertinggi, maka atribut *outlook* akan menjadi *node*. Karena atribut *outlook* memiliki tiga nilai atribut atau lebih dari dua, maka dilakukan perhitungan rasio gain untuk memilih pilihan percabangan terbaik. Berikut adalah contoh perhitungan rasio gain untuk pilihan percabangan {sunny, overcast, rain}.

$$\begin{aligned}
 \text{Split info}(\text{Semua}, \text{overcast}) &= \left(-\frac{5}{14} * \log_2 \left(\frac{5}{14} \right) \right) + \left(-\frac{4}{14} * \log_2 \left(\frac{4}{14} \right) \right) \\
 &\quad + \left(-\frac{5}{14} * \log_2 \left(\frac{5}{14} \right) \right) \\
 &= 0.531 + 0.516 + 0.531 = 1.577
 \end{aligned}$$

$$\begin{aligned}
 \text{Rasio Gain}(\text{Semua}, \text{overcast}) &= \frac{0.247}{1.577} \\
 &= 0.156
 \end{aligned}$$

Hasil untuk perhitungan *rasio gain* lainnya ada pada tabel 2.5.

Tabel 2.5 Hasil perhitungan *Rasio gain* untuk setiap pilihan cabang

		Jumlah	Split Inf	Gain	Rasio Gain
Total		14		0.247	
Pilihan 1	sunny	5	1.577		0.156
	overcast	4			
	rain	5			

Pilihan 2	sunny	5	0.940		0.262
	overcast rain	9			
Pilihan 3	sunny overcast	9	0.940		0.262
	rain	5			
Pilihan 4	sunny rain	10	0.863		0.286
	overcast	4			

Dari tabel 2.5 pilihan 4 yaitu {*sunny, rain*} dan {*overcast*} memiliki nilai *rasio gain* tertinggi, maka atribut terpilih (*outlook*) akan dibagi menjadi dua cabang. Pembagian cabang disajikan pada tabel 2.6 dan tabel 2.7.

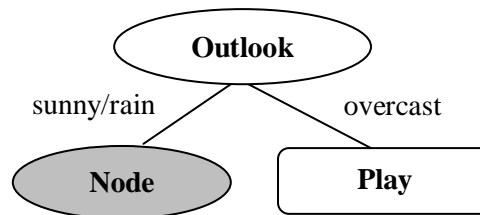
Tabel 2.6 Pembagian cabang (*sunny, rain*)

Outlook	Temperature	Humidity	Windy	Class
sunny	75	70	TRUE	Play
sunny	80	90	TRUE	Don't Play
sunny	85	85	FALSE	Don't Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	71	80	TRUE	Don't Play
rain	65	70	TRUE	Don't Play
rain	75	80	FALSE	Play
rain	68	80	FALSE	Play
rain	70	96	FALSE	Play

Tabel 2.7 Pembagian cabang (*overcast*)

Outlook	Temperature	Humidity	Windy	Class
overcast	72	90	TRUE	Play
overcast	83	78	FALSE	Play
overcast	64	65	TRUE	Play
overcast	81	75	FALSE	Play

Pada cabang *overcast* memiliki kelas yang sama yaitu *Play*, maka *node* ini akan menjadi daun dengan nilai *Play*. Sedangkan cabang *sunny* dan *rain* masih ada kelas yang berbeda, maka akan memilih atribut sebagai *node* seperti ditunjukkan pada gambar 2.3. Proses tersebut akan berulang sampai semua kasus pada cabang memiliki kelas yang sama atau menjadi daun (*leaf*).



Gambar 2.3 Hasil pembentukan cabang pada node akar

Berikut aturan IF THEN untuk *decision tree* pada penjelasan pada gambar 2.3.

IF Outlook = overcast THEN Class = Play

2.5 Penelitian Sebelumnya

Penelitian sebelumnya yang menggunakan metode Naïve Bayes dilakukan oleh Friday Aries L. W (Universitas Muhammadiyah Gresik, 2014) melakukan penelitian sistem klasifikasi penentuan penerimaan beras miskin (Raskin). Algoritma yang digunakan adalah *naïve bayes*. Data yang digunakan terdiri dari 600 kepala keluarga yang diperoleh dari Desa Sidomulyo Kec. Deket Kab. Lamongan tahun 2012 s/d 2013. Atribut yang digunakan dalam penelitian ini adalah pekerjaan, penghasilan, harta benda (kendaraan), kondisi rumah, dan jumlah keluarga. Penelitian ini diuji sebanyak 4 kali pengujian dengan rata-rata akurasi sebesar 74,93 % dan sensitivitas sebesar 81,23 %.

Penelitian lainnya yang terkait dengan Sistem penentuan penerimaan beras miskin (Raskin) yaitu Rahmawati, Syahriyatur, (2010) lulusan Universitas Muhammadiyah Gresik, Penelitiannya berjudul penerapan metode Fuzzy C-means untuk pengelompokan keluarga penerima beras miskin dikelurahan pekauman Kec. Gresik. Atribut yang digunakan dalam penelitian ini adalah pekerjaan, pendidikan, jumlah keluarga dan indikator (kesehatan, pendidikan, sosial budaya dll). Dari 20 data uji yang digunakan didapatkan hasil label keluarga prasejahtera sebanyak 7, label keluarga sejahtera 1 sebanyak 3, label keluarga sejahtera 2 sebanyak 5, label keluarga sejahtera 3 sebanyak 3, label keluarga sejahtera 3 sebanyak 2.

Penelitian selanjutnya adalah tentang metode *Decision Tree C4.5* dalam penelitian yang berjudul “*System prediksi prestasi akademik mahasiswa*

menggunakan metode decision tree C4.5 (Studi kasus:Jurusan Teknik informatika UNMUH GRESIK)”, dibuat oleh Aunur Rasyid (Universitas Muhammadiyah Gresik, 2014). Tujuan dari penelitian tersebut adalah untuk menghasilkan informasi perkiraan kategori prestasi mahasiswa menggunakan metode *Decision Tree C4.5* sebagai peringatan dini dan motivasi mahasiswa dalam mendapatkan prestasi yang maksimal. Atribut-atribut yang digunakan adalah instansi sekolah asal (SMK,SMA atau MA), satatus sekolah asal (Negri atau Swasta), jurusan sekolah asal (IPA,IPS,Bahasa,Teknik,Administrasi), nilai rata-rata UN, status kerja (Sudah atau Belum), dan pihak yang mempengaruhi mahasiswa dalam memilih kuliah (Sendiri,Orang tua atau Orang lain). Hasil dari penelitian tersebut, Sistim Prediksi mahasiswa yang dirancang menggunakan algoritma C4.5 dapat memprediksi prestasi mahasiswa agar mampu mempertahankan kondisinya atau melakukan perbaikan utuk mencapai prestasi yang maksimal. Hasil akurasi dari penelitian tersebut adalah 90%.

Penelitian lainnya yang terkait metode *Decision Tree C4.5* juga dilakukan oleh Liliana swistina (Universitas STMIK Indonesia, 2013), yaitu penelitian tentang penerapan algoritma C4.5 untuk menentukan jurusan mahasiswa. Evaluasi pengukuran *rapid miner* yaitu membandingkan nilai akurasi antara algoritma C4.5 dan Naïve Bayes. Data yang digunakan adalah data mahasiswa baru STMIK Indonesia Banjarmasin tahun 2008/2009. Data sampel terdiri dari atribut nama, jenis kelamin, umur, asal sekolah, nilai uan, ipk semester 1 dan semester 2. Hasil akurasi yang didapatkan adalah 93,31% untuk metode C4.5 dan 89,02% untuk Naïve Bayes.