

BAB II

LANDASAN TEORI

2.1 Sistem Pendukung Keputusan

Dalam bukunya yang berjudul “*Komputerisasi Pengambilan Keputusan*”, Dadan Umar Daihani (2011) menjelaskan bahwa konsep sistem penunjang keputusan (SPK) / *Decision Support System (DSS)* pertama kali diungkapkan pada tahun 1970-an oleh Michael S. Scott Marton dengan istilah *management decision system*. Terdapat banyak definisi mengenai SPK, antara lain :

a. Mann dan Waston.

Sistem berbasis komputer interaktif, yang membantu para pengambil keputusan untuk menggunakan data dan berbagai model untuk memecahkan masalah-masalah semi terstruktur dan tidak terstruktur.

b. Maran Alavi dan H. Albert Napier.

Suatu kumpulan prosedur pemrosesan data dan informasi yang berorientasi pada pengguna model untuk menghasilkan berbagai jawaban yang dapat membantu manajemen dalam pengambilan keputusan. Sistem ini harus sederhana, mudah dan adaptif.

c. Little.

Suatu sistem informasi berbasis komputer yang menghasilkan berbagai alternatif keputusan untuk membantu manajemen dalam menangani berbagai permasalahan yang terstruktur ataupun yang tidak terstruktur dengan menggunakan data dan model.

d. Raymond McLeod, Jr.

Sistem penghasil informasi spesifik yang ditujukan untuk memecahkan suatu masalah tertentu yang harus dipecahkan oleh manajer pada berbagai tingkatan.

Dari berbagai definisi diatas dapat dikatakan bahwa sistem pendukung keputusan adalah suatu sistem informasi spesifik yang ditujukan untuk membantu manajemen dalam mengambil keputusan yang berkaitan dengan persoalan yang

bersifat semi struktur. Sistem ini memiliki fasilitas untuk menghasilkan berbagai alternatif yang secara interaktif dapat digunakan oleh pemakai.

2.1.1 Konsep Sistem Pendukung Keputusan

Konsep Sistem Pendukung Keputusan (SPK) atau *Decision Support Systems* (DSS) pertama kali diperkenalkan oleh Michael S. Scott Morton pada awal tahun 1970-an, yang selanjutnya dikenal dengan istilah *Management Decision Systems*. Konsep SPK ditandai dengan sistem interaktif berbasis komputer yang membantu pengambilan keputusan dengan memanfaatkan data dan model untuk menyelesaikan masalah yang bersifat tidak terstruktur dan semi terstruktur.

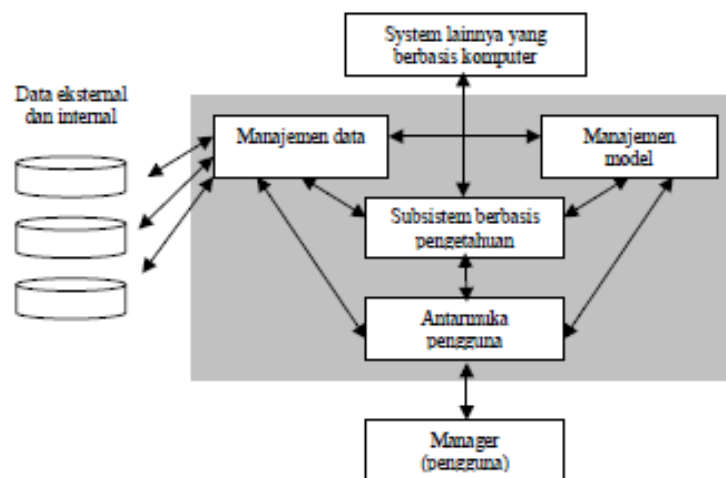
Pada proses pengambilan keputusan, pengolahan data dan informasi yang dilakukan bertujuan untuk menghasilkan berbagai alternatif keputusan yang dapat diambil. SPK yang merupakan penerapan dari sistem informasi ditujukan hanya sebagai alat bantu manajemen dalam pengambilan keputusan. SPK tidak dimaksudkan untuk menggantikan fungsi pengambil keputusan dalam membuat keputusan, melainkan hanyalah sebagai alat bantu pengambil keputusan dalam melaksanakan tugasnya. SPK dirancang untuk menghasilkan berbagai alternatif yang ditawarkan kepada para pengambil keputusan dalam melaksanakan tugasnya. Sehingga dapat dikatakan bahwa SPK memberikan manfaat bagi manajemen dalam hal meningkatkan efektivitas dan efisiensi kerjanya terutama dalam proses pengambilan keputusan. Di samping itu, SPK menyatukan kemampuan komputer dalam pelayanan interaktif terhadap penggunanya dengan adanya proses pengolahan atau pemanipulasian data yang memanfaatkan model atau aturan yang tidak terstruktur sehingga menghasilkan alternatif keputusan yang situasional.

2.1.2 Komponen Sistem Pendukung Keputusan

Adapun komponen-komponen dari Sistem Pendukung Keputusan adalah sebagai berikut (Kosasi, Sandy. 2002):

1. **Manajemen Data**, mencakup *database* yang mengandung data yang relevan dan diatur oleh sistem yang disebut *Database Management System* (DBMS).
2. **Manajemen Model**, merupakan paket perangkat lunak yang memasukkan model-model finansial, statistik, ilmu manajemen, atau model kuantitatif yang lain yang menyediakan kemampuan analisis sistem dan *management software* yang terkait.
3. **Antarmuka Pengguna**, media interaksi antara sistem dengan pengguna, sehingga pengguna dapat berkomunikasi dan memberikan perintah pada SPK melalui subsistem ini.
4. **Subsistem Berbasis Pengetahuan**, subsistem yang dapat mendukung subsistem lain atau bertindak sebagai komponen yang berdiri sendiri.

Untuk dapat lebih jelas memahami model konseptual SPK, perhatikan gambar 2.1:



Gambar 2.1 Model Konseptual SPK

2.1.3 Proses Pengambilan Keputusan

Menurut Simon, proses pengambilan keputusan meliputi tiga tahapan utama yaitu tahap inteligensi, desain, dan pemilihan. Namun kemudian ditambahkan dengan tahap keempat yaitu tahap implementasi (Kosasi, Sandy. 2002). Keempat tahapan tersebut dapat dijelaskan sebagai berikut:

1. Tahap Penelusuran (*Intelligence*)

Merupakan tahap pendefinisian masalah serta identifikasi informasi yang dibutuhkan yang berkaitan dengan persoalan yang dihadapi serta keputusan yang akan diambil. Langkah ini sangat penting karena sebelum suatu tindakan diambil, tentunya persoalan yang dihadapi harus dirumuskan secara jelas terlebih dahulu.

2. Perancangan (*Design*)

Merupakan tahap analisa dalam kaitan mencari atau merumuskan alternatif-alternatif pemecahan masalah. Setelah permasalahan dirumuskan dengan baik, maka tahap berikutnya adalah merancang atau membangun model pemecahan masalahnya dan menyusun berbagai alternatif pemecahan masalah.

3. Pemilihan (*Choice*)

Dengan mengacu pada rumusan tujuan serta hasil yang diharapkan, selanjutnya manajemen memilih alternatif solusi yang diperkirakan paling sesuai. Pemilihan alternatif ini akan mudah dilakukan kalau hasil yang diinginkan terukur atau memiliki nilai kuantitas tertentu.

4. Implementasi (*Implementation*)

Merupakan tahap pelaksanaan dari keputusan yang telah diambil. Pada tahap ini perlu disusun serangkaian tindakan yang terencana, sehingga hasil keputusan dapat dipantau dan disesuaikan apabila diperlukan perbaikan-perbaikan.

2.1.4 Karakteristik dan Nilai Guna Sistem Pendukung Keputusan

Beberapa karakteristik dari Sistem Pendukung Keputusan menurut Turban adalah sebagai berikut (Suryadi, Kadarsah. 1998) :

1. Sistem Pendukung Keputusan dirancang untuk membantu pengambil keputusan dalam memecahkan masalah yang sifatnya semi terstruktur ataupun tidak terstruktur.
2. Dalam proses pengolahannya, sistem pendukung keputusan mengombinasikan penggunaan model-model/teknik-teknik analisis

dengan teknik pemasukan data konvensional serta fungsi-fungsi pencari/interogasi informasi.

3. Sistem Pendukung Keputusan, dirancang sedemikian rupa sehingga dapat digunakan/dioperasikan dengan mudah oleh orang-orang yang tidak memiliki dasar kemampuan yang tinggi. Oleh karena itu pendekatan yang digunakan biasanya model interaktif.
4. Sistem Pendukung Keputusan dirancang dengan menekankan pada aspek fleksibilitas serta kemampuan adaptasi yang tinggi. Sehingga mudah disesuaikan dengan berbagai perubahan lingkungan yang terjadi dan kebutuhan pemakai.

Dengan berbagai karakter khusus seperti yang dikemukakan di atas, sistem pendukung keputusan dapat memberikan berbagai manfaat atau keuntungan bagi pemakainya. Keuntungan yang dimaksud di antaranya meliputi:

1. Sistem Pendukung Keputusan memperluas kemampuan pengambil keputusan dalam memproses data/informasi bagi pemakainya.
2. Sistem Pendukung Keputusan membantu pengambil keputusan dalam hal penghematan waktu yang dibutuhkan untuk memecahkan masalah terutama berbagai masalah yang sangat kompleks dan tidak terstruktur.
3. Sistem Pendukung Keputusan dapat menghasilkan solusi dengan lebih cepat serta hasilnya dapat diandalkan.
4. Walaupun suatu Sistem Pendukung Keputusan, mungkin saja tidak mampu memecahkan masalah yang dihadapi oleh pengambil keputusan, namun dapat dijadikan stimulan bagi pengambil keputusan dalam memahami persoalannya. Karena sistem ini mampu menyajikan berbagai alternatif.
5. Sistem Pendukung Keputusan dapat menyediakan bukti tambahan untuk memberikan pembenaran sehingga dapat memperkuat posisi pengambil keputusan.

Di samping berbagai keuntungan dan manfaat yang dikemukakan di atas, Sistem Pendukung Keputusan juga memiliki keterbatasan diantaranya adalah sebagai berikut:

1. Ada beberapa kemampuan manajemen dan bakat manusia yang tidak dapat dimodelkan, sehingga model yang ada dalam sistem tidak semuanya mencerminkan persoalan sebenarnya.
2. Kemampuan suatu SPK terbatas pada pembendaharaan pengetahuan yang dimilikinya (pengetahuan dasar serta model dasar).
3. Proses-proses yang dapat dilakukan oleh SPK biasanya tergantung juga pada kemampuan perangkat lunak yang digunakannya.
4. SPK tidak memiliki kemampuan intuisi seperti yang dimiliki oleh manusia. Karena walau bagaimanapun canggihnya suatu SPK, tetap saja berupa kumpulan dari perangkat keras, perangkat lunak dan sistem operasi yang tidak dilengkapi dengan kemampuan berpikir.

2.2 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).

2.2.1 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.

2.2.2 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 asosiasi 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.

2.3 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.

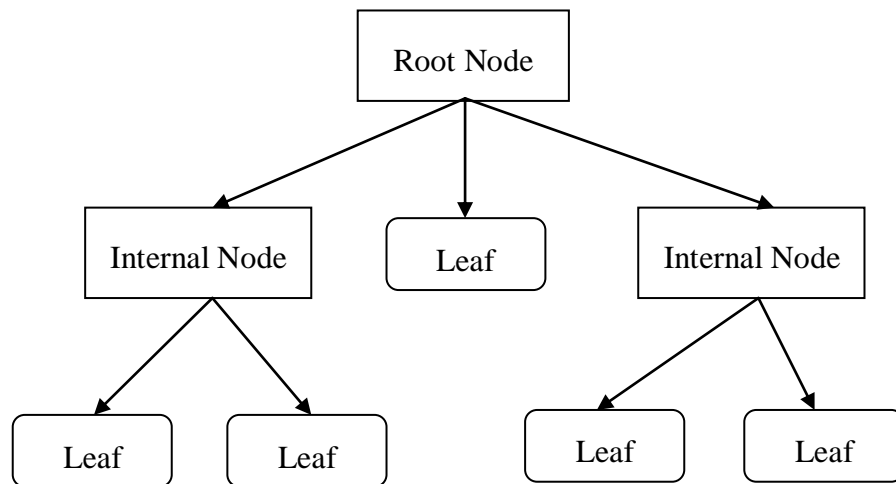
2.3.1 Model Decision Tree

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:

- a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- c. *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 2.2.



Gambar 2.2 Model *Decision Tree*

2.3.2 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 interval) 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*).

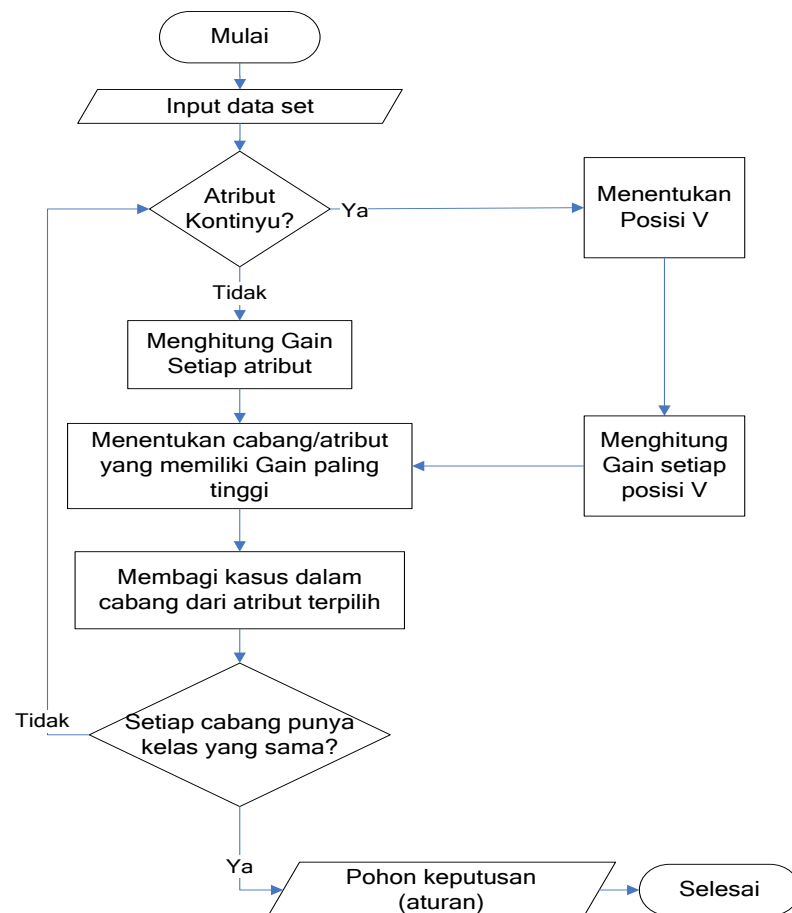
3. Fitur numerik

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 *flowchart* yang disajikan pada gambar 2.3.



Gambar 2.3 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.

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

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

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.

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

Keterangan :

TP : Kelas acc yang diprediksi secara benar sebagai Kelas acc

FN : Kelas acc yang diprediksi secara salah sebagai Kelas tolak

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

Keterangan :

TN : Kelas tolak yang diprediksi secara benar sebagai Kelas tolak

FP : Kelas tolak yang diprediksi secara salah sebagai Kelas acc

2.4 Penelitian Terdahulu

Penelitian terdahulu yang menggunakan 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 (Negeri 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.