



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

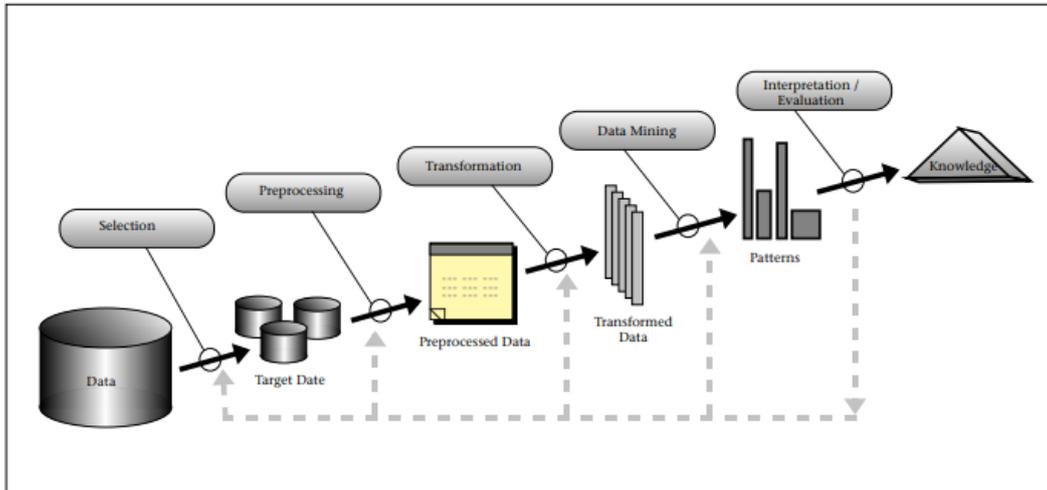
LANDASAN TEORI

2.1 Data Mining

Data mining adalah proses untuk menemukan pola yang menarik, berguna, dan dapat dipahami dari kumpulan data yang berjumlah besar. *Data mining* juga berkaitan dengan bidang-bidang lain seperti *machine learning*, *database*, statistika, dan *pattern recognition* (Zaki dan Meira, 2014).

Dalam keseluruhan proses *Knowledge Discovery in Databases* (KDD), *data mining* merupakan salah satu tahap inti yang mencari dan menentukan suatu pola baru dari suatu *data set* yang besar. Tahapan-tahapan lain dari KDD antara lain (Fayyad dkk., 1996):

1. *Selection*: Tahapan dimana dilakukan penentuan *data set* mana yang akan digunakan dan menentukan variabel atau data sampel utama yang akan digunakan.
2. *Preprocessing*: Tahapan dimana dilakukan penghilangan *noise*, penanggulangan terhadap data yang hilang, dan pengelolaan perubahan yang terjadi secara berkala.
3. *Transformation*: Tahapan dimana dilakukan reduksi data yang berguna untuk mengurangi variabel yang masih diperhitungkan.
4. *Data Mining*: Tahapan dimana dicari pola dari suatu *data set* berdasarkan metode-metode *classification*, *tree*, *regression*, atau *clustering*.
5. *Interpretation/Evaluation*: Tahapan dimana dilakukan evaluasi dan interpretasi secara keseluruhan dari data yang telah diproses pada tahap-tahap sebelumnya.



Gambar 2.1 Tahapan-tahapan dalam proses KDD (Sumber: Fayyad dkk., 1996)

2.2 Sistem Pendukung Keputusan

Sistem Pendukung Keputusan atau *Decision Support System* (DSS) adalah suatu *Computer Based Information System* (CBIS) yang interaktif, mampu beradaptasi dan secara fleksibel saling mempengaruhi, dimana sistem ini menggunakan aturan-aturan keputusan, model-model, dan penggabungan model dasar yang meliputi basis data dan pengetahuan pengambil keputusan berada didalamnya menuju suatu hasil tertentu yang merupakan keputusan yang dapat diterapkan dalam menyelesaikan masalah (Turban, 1988).

Kerangka dasar pengambilan keputusan manajerial dalam tipe keputusan dibagi menjadi tiga, antara lain (Mengkepe, 2004):

1. Terstruktur: Berisi masalah rutin yang sering terjadi, solusinya adalah *standard* dan baku. Teknologi yang digunakan adalah Sistem Informasi Manajemen dan Penelitian Operasional.
2. Tidak Terstruktur: Berisi masalah kompleks menggunakan pemecahan masalah yang tidak *standard*, dan mengandalkan intuisi manusia sebagai basis pembuat keputusan. Teknologi yang digunakan adalah Sistem Pakar.

3. Semi Terstruktur: Adalah gabungan antara keputusan struktur dengan tidak terstruktur. Pengambilan keputusan ini tidak hanya memberikan solusi tunggal tapi juga memberikan alternatif solusi. Teknologi yang digunakan adalah Sistem Pendukung Keputusan.

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) merupakan algoritma yang memanfaatkan teori probabilitas, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Klasifikasi-klasifikasi *Bayes* adalah adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* (Astuti dkk., 2015).

Persamaan umum dari *Bayes* dijelaskan pada rumus berikut (Han dkk., 2012).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad \dots (2.1)$$

Keterangan:

X : Suatu kejadian atau *event*.

H : Suatu hipotesis.

$P(H)$: Peluang hipotesis H .

$P(X)$: Peluang kejadian X .

$P(X|H)$: Peluang kejadian X jika H terpenuhi.

$P(H|X)$: Peluang kejadian H jika X terpenuhi.

Persamaan umum *Bayes* tersebut kemudian disesuaikan agar dapat digunakan dalam klasifikasi. Perubahan tersebut dijelaskan pada persamaan berikut (Han dkk., 2012):

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \text{ dimana } 1 \leq i \leq m \quad \dots (2.2)$$

Keterangan:

X : Suatu kejadian atau *event*.

C_i : Suatu *class* klasifikasi.

$P(C_i)$: Peluang *class* C_i .

$P(X)$: Peluang suatu kejadian X .

$P(X|C_i)$: Peluang kejadian X jika C_i terpenuhi.

$P(C_i|X)$: Peluang kejadian C_i jika X terpenuhi.

Karena nilai $P(X)$ dianggap konstan disebabkan oleh jumlah sampel yang tetap, maka $P(X)$ tidak dimasukkan ke persamaan turunan selanjutnya. Selain itu, untuk menyederhanakan proses perhitungan, maka dibuat asumsi independensi antar atribut yang tinggi. Dengan demikian, persamaan *Naïve Bayes Classifier* yang telah dimodifikasi menjadi sebagai berikut (Han dkk., 2012):

$$P(C_i|X) = P(X|C_i)P(C_i) \quad \dots (2.3)$$

dimana

$$P(X|C_i) = \sum_{k=1}^n P(x_k|C_i)$$

Berikut adalah contoh penerapan algoritma *Naïve Bayes Classifier* untuk memprediksi apakah seorang siswa berpotensi untuk memiliki perilaku yang bermasalah atau tidak.

Tabel 2.1 Contoh *Dataset* (Shalihah, 2016)

jurusan	nilai_iq	stabilitas emosi	kepercayaan diri	hubungan sosial	potensi_perilaku
RPL	nilai>800	400	600	400	Bermasalah
RPL	nilai>800	800	600	400	Tidak Bermasalah
RPL	700-799	600	400	400	Tidak Bermasalah
RPL	700-799	400	600	600	Tidak Bermasalah
RPL	700-799	600	600	600	Tidak Bermasalah

Tabel 2.1 Contoh *Dataset* (Shalihah, 2016) (lanjutan)

jurusan	nilai_iq	stabilitas emosi	kepercayaan diri	hubungan sosial	potensi_perilaku
RPL	700-799	400	600	400	Bermasalah
RPL	700-799	600	600	600	Tidak Bermasalah
RPL	700-799	600	600	600	Tidak Bermasalah
RPL	500-699	600	600	600	Tidak Bermasalah
RPL	500-699	400	600	800	Tidak Bermasalah
RPL	500-699	600	600	600	Tidak Bermasalah
RPL	500-699	800	400	600	Tidak Bermasalah
RPL	500-699	600	600	600	Tidak Bermasalah
RPL	500-699	600	600	600	Tidak Bermasalah

Berdasarkan Tabel 2.1, terdapat dua *class* atau kondisi yang dituju yaitu siswa memiliki potensi untuk memiliki perilaku bermasalah atau tidak. Misalkan terdapat contoh data yang belum diketahui *class* tujuannya:

Tabel 2.2 Contoh Data Prediksi (Shalihah, 2016)

jurusan	nilai iq	stabilitas emosi	kepercayaan diri	hubungan sosial	potensi_perilaku
RPL	700-799	600	400	400	

Pada Tabel 2.2 terdapat data yang menunjukkan siswa dari jurusan RPL, nilai IQ berkisar 700-799, nilai stabilitas emosi 600, nilai kepercayaan diri 400, dan nilai hubungan sosial 400. Dari data ini akan dilakukan prediksi apakah siswa yang bersangkutan memiliki potensi untuk memiliki perilaku bermasalah atau tidak. Perhitungan *Naïve Bayes* dilakukan untuk mengetahui *class* tujuan dengan notasi $P(C_i)$. Kemudian atribut dari *class* tersebut akan ditentukan dengan ketentuan sebagai berikut:

$$C_1 = (\text{potensi_perilaku} = \text{"Bermasalah"})$$

$$C_2 = (\text{potensi_perilaku} = \text{"Tidak Bermasalah"})$$

$$x_1 = (\text{jurusan} = \text{"RPL"})$$

$$x_2 = (\text{nilai_iq} = \text{"700-799"})$$

$$x_3 = (\text{stabilitas_emosi} = \text{"600"})$$

$$x_4 = (\text{kepercayaan_diri} = \text{"400"})$$

$x_5 = (\text{hubungan_sosial} = \text{"400"})$

Langkah-langkah perhitungan *Naïve Bayes* untuk menghitung probabilitas *class* C_i untuk $i = 1,2$ berdasarkan jumlah nilai atribut pada tabel 1 adalah sebagai berikut:

1. Pada *class* *potensi_perilaku* terdapat dua data yang bernilai "Bermasalah" dan dua belas data yang bernilai "Tidak Bermasalah". Setelah itu akan dihitung nilai $P(C_i)$.

$$P(C_1) = P(\text{potensi_perilaku} = \text{"Bermasalah"}) = \frac{2}{14} = 0,142$$

$$P(C_2) = P(\text{potensi_perilaku} = \text{"Tidak Bermasalah"}) = \frac{12}{14} = 0,857$$

2. Menghitung probabilitas dari tiap kejadian terhadap suatu *class*. Dalam contoh ini, perhitungan yang akan dilakukan adalah sebagai berikut.

$$\begin{aligned} P(x_1|C_1) &= P(\text{jurusan} = \text{"RPL"} \mid \text{potensi_perilaku} = \text{"Bermasalah"}) \\ &= \frac{2}{2} = 1 \end{aligned}$$

$$\begin{aligned} P(x_1|C_2) &= P(\text{jurusan} = \text{"RPL"} \mid \text{potensi_perilaku} = \text{"Tidak Bermasalah"}) \\ &= \frac{12}{12} = 1 \end{aligned}$$

$$\begin{aligned} P(x_2|C_1) &= P(\text{nilai_iq} = \text{"700-799"} \mid \text{potensi_perilaku} = \text{"Bermasalah"}) \\ &= \frac{1}{2} = 0,5 \end{aligned}$$

$$\begin{aligned} P(x_2|C_2) &= P(\text{nilai_iq} = \text{"700-799"} \mid \text{potensi_perilaku} = \text{"Tidak Bermasalah"}) \\ &= \frac{5}{12} = 0,416 \end{aligned}$$

$$\begin{aligned} P(x_3|C_1) &= P(\text{stabilitas emosi} = \text{"600"} \mid \text{potensi_perilaku} = \text{"Bermasalah"}) \\ &= \frac{0}{2} = 0 \end{aligned}$$

$$\begin{aligned} P(x_3|C_2) &= P(\text{stabilitas emosi} = \text{"600"} \mid \text{potensi_perilaku} = \text{"Tidak Bermasalah"}) \\ &= \frac{8}{12} = 0,666 \end{aligned}$$

$$P(x_4|C_1) = P(\text{kepercayaan_diri} = "400" \mid \text{potensi_perilaku} = "Bermasalah")$$

$$= \frac{0}{2} = 0$$

$$P(x_4|C_2) = P(\text{kepercayaan_diri} = "400" \mid \text{potensi_perilaku} = "Tidak Bermasalah") = \frac{2}{12} = 0,166$$

$$P(x_5|C_1) = P(\text{hubungan_sosial} = "400" \mid \text{potensi_perilaku} = "Bermasalah")$$

$$= \frac{2}{2} = 1$$

$$P(x_5|C_2) = P(\text{hubungan_sosial} = "400" \mid \text{potensi_perilaku} = "Tidak Bermasalah") = \frac{2}{12} = 0,166$$

3. Menghitung probabilitas suatu *class* berdasarkan kejadian x_1, x_2, \dots, x_n . Terdapat 2 pilihan *class* yaitu $\text{potensi_perilaku} = "Bermasalah"$ dan $\text{potensi_perilaku} = "Tidak Bermasalah"$.

Untuk $C_1 = (\text{potensi_perilaku} = "Bermasalah")$

$$P(C_1) \times P(x_1|C_1) \times P(x_2|C_1) \times P(x_3|C_1) \times P(x_4|C_1) \times P(x_5|C_1)$$

$$= 0,142 \times 1 \times 0,5 \times 0 \times 0 \times 0$$

$$= 0$$

Untuk $C_2 = (\text{potensi_perilaku} = "Tidak Bermasalah")$

$$P(C_2) \times P(x_1|C_2) \times P(x_2|C_2) \times P(x_3|C_2) \times P(x_4|C_2) \times P(x_5|C_2)$$

$$= 0,857 \times 1 \times 0,416 \times 0,666 \times 0,166 \times 0,166$$

$$= 0,006$$

Berdasarkan hasil perhitungan tersebut, maka diketahui $C_1 < C_2$ dengan nilai $0 < 0,006$. Dapat disimpulkan bahwa potensi perilaku dari siswa yang bersangkutan adalah "Tidak Bermasalah".

2.4 Inventori Tugas Perkembangan

Inventori Tugas Perkembangan (ITP) adalah salah satu instrumen yang dapat digunakan untuk mengukur tingkat perkembangan peserta didik. Dengan menggunakan ITP, pembimbing dapat memahami tingkat perkembangan individu maupun kelompok, mengidentifikasi masalah yang menghambat perkembangan, dan membantu peserta didik yang bermasalah (Kartadinata dkk., 2003).

ITP mengukur tingkat perkembangan sebelas aspek, yaitu: (1) landasan hidup religius, (2) landasaan perilaku etis, (3) kematangan emosional, (4) kematangan intelektual, (5) kesadaran tanggung jawab, (6) peran sosial sebagai pria atau wanita, (7) penerimaan diri dan pengembangannya, (8) kemandirian perilaku ekonomi, (9) wawasan dan persiapan karir, (10) kematangan hubungan dengan teman sebaya, dan (11) persiapan diri untuk pernikahan dan hidup berkeluarga. ITP berbentuk angket yang terdiri atas kumpulan pertanyaan yang harus dijawab oleh peserta didik. Tingkat perkembangan peserta didik dapat dilihat dari skor yang diperoleh pada setiap aspek.

ITP memiliki tujuh tingkat perkembangan siswa yang digunakan untuk mengklasifikasikan tingkat perkembangan siswa yang dimaksud. Tujuh tingkatan ini dibagi berdasarkan jenjang pendidikan SD, SLTP, SLTA, dan PT (Kartadinata dkk, 2003).

Tabel 2.3 Tingkatan ITP Jenjang Sekolah Dasar

Skor	Kode	Tingkat Perkembangan Siswa
1	IMP	Tahap Impulsif
2	PLD	Tahap Perlindungan Diri
3	KOM	Tahap Konformitas
4	SDI	Tahap Sadar Diri

Tabel 2.4 Tingkatan ITP Jenjang Sekolah Lanjutan Tingkat Pertama

Skor	Kode	Tingkat Perkembangan Siswa
2	PLD	Tahap Perlindungan Diri

Tabel 2.4 Tingkatan ITP Jenjang Sekolah Lanjutan Tingkat Pertama (lanjutan)

Skor	Kode	Tingkat Perkembangan Siswa
3	KOM	Tahap Konformitas
4	SDI	Tahap Sadar Diri
5	SAK	Tahap Saksama

Tabel 2.5 Tingkatan ITP Jenjang Sekolah Lanjutan Tingkat Atas

Skor	Kode	Tingkat Perkembangan Siswa
3	KOM	Tahap Konformitas
4	SDI	Tahap Sadar Diri
5	SAK	Tahap Saksama
6	IND	Tahap Individualitas

Tabel 2.6 Tingkatan ITP Tingkat Perguruan Tinggi

Skor	Kode	Tingkat Perkembangan Siswa
4	SDI	Tahap Sadar Diri
5	SAK	Tahap Saksama
6	IND	Tahap Individualitas
7	OTO	Tahap Otonomi

Kategorisasi tingkat perkembangan pada ITP diukur dengan jumlah skor yang dicapai oleh siswa yang melakukan tes. Semakin tinggi skor yang dicapai, maka pada umumnya siswa tersebut akan lebih dewasa dan lebih bijaksana dalam kehidupan sehari-hari (Kartadinata dkk, 2003).

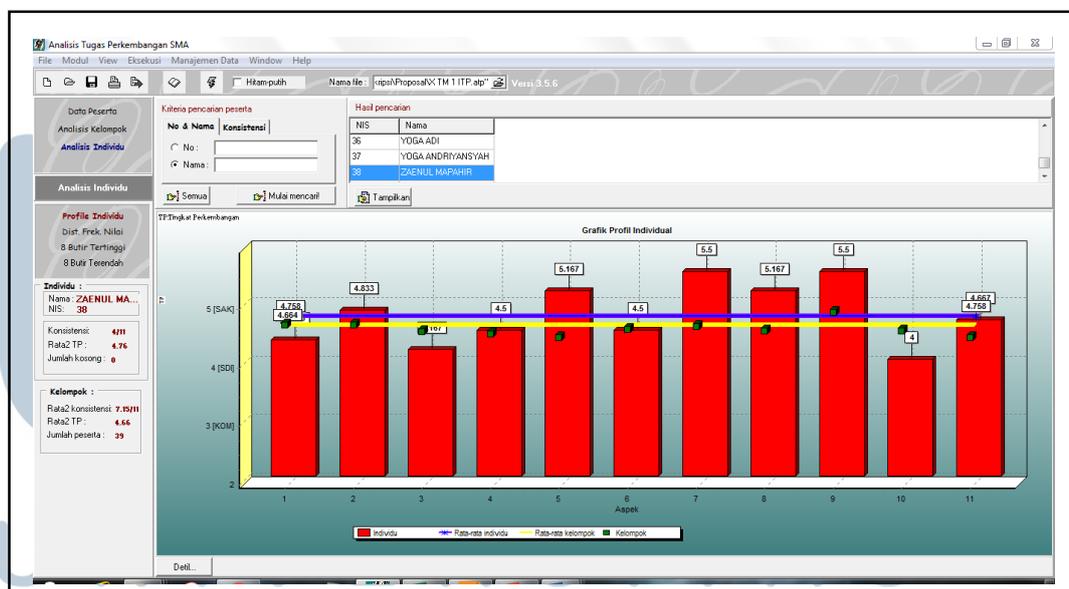
2.5 Analisis Tugas Perkembangan

Analisis Tugas Perkembangan (ATP) adalah perangkat lunak yang dibuat khusus untuk membantu guru konseling dalam mengolah dan menerjemahkan data ITP. Dengan ATP, identifikasi perkembangan siswa dapat dilakukan dengan mudah, cepat, dan menyenangkan (Kartadinata dkk., 2003).

ATP menyediakan berbagai fasilitas dan kemampuan yang dapat digunakan oleh guru konseling dalam melakukan analisis terhadap perkembangan peserta didik, diantaranya yaitu:

1. Pengolahan data mentah secara cepat. Pada komputer Pentium 4, hanya dibutuhkan waktu 1 detik untuk mengolah data dari 100 peserta.

2. Analisis kelompok yang terdiri atas profil kelompok, grafik distribusi frekuensi untuk setiap aspek, grafik distribusi frekuensi konsistensi, delapan butir tertinggi dan terendah.
3. Analisis per individu yang terdiri atas profil individual, distribusi frekuensi nilai, delapan butir tertinggi dan terendah untuk individu tersebut.
4. Visualisasi hasil pengolahan skor dalam bentuk grafik akan memudahkan dan mempercepat guru konseling dalam analisis.
5. Manajemen data, terdiri atas pengelompokan siswa berdasarkan kriteria tertentu, dan penggabungan kelompok.
6. *Export* hasil pengolahan data ke Microsoft Excel®.
7. *Import* data dari Microsoft Excel®.
8. *Multi window*, beberapa *window* bisa dibuka sekaligus untuk membandingkan hasil pengolahan.



Gambar 2.2 Contoh Tampilan Utama ATP (Sumber: Kartadinata dkk., 2003)

MULTIMEDIA
NUSANTARA

2.6 Skala Likert

Skala Likert adalah sebuah tipe skala psikometri yang menggunakan angket dan menggunakan skala yang lebih luas dalam penelitian survei. Metode *rating* yang di jumlahkan (*summated rating*) ini sering disebut penskalaan model Likert. Metode Likert merupakan metode penskalaan pernyataan sikap yang menggunakan distribusi respons sebagai dasar penentuan nilai skalanya. Dalam pendekatan ini tidak diperlukan adanya kelompok panel penilai (*Judging Group*) dikarenakan nilai skala setiap pernyataan tidak akan ditentukan oleh derajat favorabilitasnya masing-masing, akan tetapi ditentukan oleh distribusi respons setuju atau tidak setuju dari sekelompok responden yang bertindak sebagai kelompok uji coba (Risnita, 2012).

Pada dasarnya, penskalaan menggunakan skala Likert memiliki dua asumsi sebagai berikut (Risnita, 2012):

1. Setiap pernyataan sikap yang telah ditulis dapat disepakati sebagai pernyataan yang *favorable* atau pernyataan yang tidak *favorable*.
2. Untuk pernyataan positif, jawaban yang diberikan oleh individu yang memiliki sikap positif harus diberi bobot atau nilai yang lebih tinggi dari jawaban yang diberikan oleh responden yang mempunyai sikap negatif. Demikian sebaliknya untuk pernyataan negatif, jawaban yang diberikan oleh individu yang memiliki sikap negatif harus diberi bobot atau nilai yang lebih tinggi dari jawaban yang diberikan oleh responden yang mempunyai sikap positif.

Ketika merespon angket Likert, responden secara spesifik menyatakan tingkat pernyataan mereka. Bentuk tes/kuesioner pada skala Likert adalah bentuk pernyataan. Responden mengindikasikan tingkat keyakinan mereka dengan pernyataan atau evaluasi objektif/subjektif. Biasanya dalam skala Likert terbagi dalam lima

kategori yang digunakan, tetapi banyak pakar psikometri menggunakan tujuh sampai sembilan kategori. Kelima kategori tersebut dijabarkan sebagai berikut:

Tabel 2.7 Kategori Pernyataan Skala Likert

Pernyataan Positif		Pernyataan Negatif	
5.	Sangat setuju	5.	Sangat tidak setuju
4.	Setuju	4.	Tidak setuju
3.	Ragu-ragu	3.	Ragu-ragu
2.	Tidak setuju	2.	Setuju
1.	Sangat tidak setuju	1.	Sangat setuju

2.7 End-User Computing Satisfaction

Doll dan Torkzadeh (1988) menyatakan bahwa *end-user computing satisfaction* merupakan salah satu indikator keberhasilan penerapan sistem informasi. Jika pengguna sistem merasa puas dengan sistem yang telah dibuat, berarti ada manfaat tertentu yang dirasakan oleh pengguna sistem tersebut yang membantu dan mempermudah para pengguna ketika sedang bekerja. Berdasarkan hal tersebut, maka akan dilakukan pengujian terhadap kepuasan sistem informasi untuk membuktikan apakah sistem yang dirancang bermanfaat bagi pihak yang menggunakan sistem ini.

Uji kepuasan pengguna dapat dilakukan menggunakan kuesioner berupa pertanyaan yang mengandung aspek bentuk, kemudahan, isi, akurasi dan ketepatan waktu sesuai dengan konsep *End-User Computing Satisfaction* (EUCS) yang diusulkan oleh Doll dan Torkzadeh (1988).

2.8 Information Gain

Information gain adalah suatu pengukuran untuk melakukan seleksi terhadap atribut-atribut sehingga dapat disimpulkan atribut apa saja yang akan digunakan. Penelitian yang dilakukan oleh Astuti (2015) menggunakan nilai

information gain untuk mereduksi atribut pada sebelum melakukan pengukuran akurasi. Rumus *entropy* adalah sebagai berikut (Astuti, 2015):

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad \dots (2.4)$$

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*

Setelah mendapatkan nilai *entropy*, selanjutnya adalah mengukur efektivitas suatu atribut dalam mengklasifikasi data. Rumus *information gain* adalah sebagai berikut (Astuti, 2015):

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad \dots (2.5)$$

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-1

|S| : Jumlah kasus dalam *S*