

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Penelitian Pendahuluan**

Sistem Pendukung Keputusan (SPK) sudah banyak diterapkan hampir dalam setiap pekerjaan manusia. Sistem ini bertujuan membantu manusia dalam mengambil keputusan dengan menggunakan data dan model untuk menyelesaikan masalah-masalah yang tak terstruktur. Dengan kata lain SPK merupakan salah satu jenis sistem informasi yang bertujuan untuk menyediakan informasi, membimbing, memberikan prediksi serta mengarahkan kepada pengguna informasi agar dapat melakukan pengambilan keputusan dengan lebih baik dan berbasis *evidence* (Hadiyanti, 2008).

Pengembangan SPK menggunakan teknik penambahan data dengan Algoritma ID3 sebelumnya sudah dilakukan dalam Perancangan Sistem Pendukung Keputusan (SPK) untuk menentukan Kelaiklautan Kapal (Setiawan, 2010). Ada tiga aspek utama yang dianalisis yaitu: Nautis (Dokumentasi, Keselamatan pelayaran, Alat-alat keselamatan, Alat-alat pemadam, Peralatan pemadam akomodasi & pelayanan), Teknis (Sumber listrik darurat, Sistem pencegahan kebakaran akibat listrik, Mesin kemudi, Instalasi kemudi dan hidraulik, Sistem keselamatan dari kebakaran, Permesinan) dan Radio (Sumber tenaga listrik, Instalasi radio, Alat-alat radio, Dokumentasi). Hasilnya dikategorikan kedalam dua hasil

yaitu Laik dan Tidak Laik. Dalam penelitian ini didapatkan hasil sebagai berikut: untuk kriteria nautis faktor yang paling mempengaruhi adalah dokumentasi, untuk kriteria teknis adalah instalasi kemudi dan hidraulik, sedangkan untuk kriteria radio adalah alat-alat radio.

Penelitian lain yang pernah membahas mengenai pohon keputusan dan algoritma ID3 adalah metode *iterative dichotomizer 3* (ID3) untuk penerimaan mahasiswa baru (Wahyudin, 2009). Dalam tulisan ini pengambilan keputusan penerimaan mahasiswa baru ini dibatasi dengan memperlihatkan tiga atribut yaitu SPMB, UAN, dan Psikotest. Hasilnya pohon keputusan dengan Algoritma ID3 dapat digunakan untuk memperoleh pengetahuan pada bidang pendidikan khususnya memberikan keputusan dalam hal penerimaan mahasiswa baru (Ya dan Tidak).

SPK untuk menentukan penerima beasiswa Bank BRI menggunakan FMADM (studi kasus: mahasiswa Fakultas Teknologi Industry Universitas Islam Indonesia) telah dilakukan oleh Wibowo. dkk, 2009. Sistem pendukung keputusan ini menggunakan *Fuzzy MADM (Multiple Attribute Decision Making)* dengan metode SAW (*Simple Additive Weighting*). Hasilnya berdasarkan kriteria-kriteria yang telah ditentukan, dimana kriteria tersebut diterjemahkan dari bilangan fuzzy kedalam bentuk sebuah bilangan crisp. Sehingga nilainya bisa dilakukan proses perhitungan untuk mencari alternatif terbaik. Dalam penelitian ini diketahui bahwa semakin banyak sampel yang dipunyai, maka tingkat validitasnya akan cenderung naik.

SPK untuk menentukan pemenang tender proyek oleh Guterres (2012) menggunakan Algoritma C4.5. Ada beberapa kriteria yang digunakan dalam pengambilan keputusan antara lain: Evaluasi Administrasi, Evaluasi Teknis, Evaluasi Harga dan Evaluasi Kualifikasi dengan kategori Tinggi, Sedang, dan Rendah. Aplikasi yang dibangun menggunakan algoritma ini mengklasifikasikan pemenang tender ke dalam dua kelas (diterima dan ditolak). Dari hasil uji keakurasian, didapatkan hasil sebesar 88,79 %.

## **2.2. Landasan Teori**

### **2.2.1. Bantuan Khusus Murid Miskin**

Bantuan Khusus Murid Miskin (BKMM) adalah program pemerintah yang di implementasikan melalui kebijakan yang berpihak kepada masyarakat miskin (*affirmative and pro poor policy*), yaitu mensubsidi biaya pendidikan (Pedoman BKMM 2011) bagi masyarakat yang kurang mampu. Di tahun 2011 ini dianggarkan dana sebesar Rp. 243 Milyar untuk seluruh siswa miskin khususnya Sekolah Menengah Atas di 33 Propinsi di Indonesia. Sehingga jumlah siswa menengah atas yang terancam putus sekolah semakin berkurang.

Tujuan dari program ini antara lain (Panduan BKMM 2011):

1. Membantu siswa miskin memenuhi kebutuhan biaya pendidikan selama duduk di bangku SMA.
2. Mencegah siswa miskin dari kemungkinan putus sekolah akibat kesulitan biaya pendidikan.
3. Memberikan peluang dan kesempatan yang lebih besar kepada siswa miskin untuk terus besekolah hingga menyelesaikan pendidikan.
4. Membantu kelancaran program sekolah.

Landasan hukum dalam pelaksanaan Program BKMM berdasarkan pada peraturan perundang-undangan yang berlaku antara lain:

1. Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah Nomor 28 Tahun 1990 tentang Pendidikan Menengah.
3. Peraturan Pemerintah Nomor 108 Tahun 2000 tentang Pengelolaan dan Pertanggung-jawaban Keuangan dalam Pelaksanaan Dekonsentrasi dan Tugas Pembantuan.
4. Peraturan Pemerintah Nomor 7 Tahun 2008 tentang Dekonsentrasi dan Tugas Pembantuan.
5. Peraturan Pemerintah Nomor 48 Tahun 2008 tentang Pendanaan Pendidikan.

6. Peraturan Pemerintah Nomor 17 Tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan.
7. Instruksi Presiden Nomor 1 Tahun 2010 tentang Percepatan Pelaksanaan Prioritas Pembangunan Nasional Tahun 2010.
8. Keputusan Menteri Pendidikan Nasional Nomor 044/U/2002 tentang Dewan Pendidikan dan Komite Sekolah.
9. Keputusan Menteri Pendidikan Nasional Nomor 060/U/2002 tentang Pedoman Pendirian Sekolah.

## **2.2.2. Sistem Pendukung Keputusan**

### **2.2.2.1. Definisi Sistem Pendukung Keputusan**

Ada beberapa pengertian tentang Sistem Pendukung Keputusan (SPK):

- a. Dalam buku *Information Technology for Management* (Turban dan Volonio, 2010) sebuah Sistem Pendukung Keputusan (SPK) menggabungkan model dan data untuk menyelesaikan masalah semiterstruktur dan tidak terstruktur dengan melibatkan pengguna. SPK bisa dilihat sebagai sebuah pencapaian atau sebuah filosofi daripada sebuah metodologi yang tepat.

- b. Little (1970) Mendefinisikan SPK sebagai “serangkaian model dari prosedur-prosedur dalam pengolahan data dan penilaian untuk membantu manajer dalam pengambilan keputusan”. Ia berpendapat untuk sukses, sistem harus sederhana, kuat, mudah untuk kontrol, adaptif, dapat menyelesaikan masalah-masalah penting, dan mudah untuk berkomunikasi.
- c. Bonczek, dkk (1980) mendefinisikan SPK sebagai sebuah sistem berbasis komputer yang terdiri dari tiga komponen yang berinteraksi: sistem bahasa (mekanisme untuk menyediakan komunikasi antara pengguna dan komponen-komponen lain dari SPK), sistem pengetahuan (menghimpun pengetahuan dasar yang terkandung dalam SPK sebagai salah satu data atau prosedur) , dan sistem pemrosesan masalah (hubungan antara dua komponen yang berlainan, yang mengandung lebih dari satu masalah umum yang dimanipulasi untuk pengambilan keputusan).

- d. Keen (1980), menjelaskan istilah SPK untuk situasi di mana sistem akhir dapat dikembangkan hanya melalui proses belajar yang adaptif dan evolusi. Dengan demikian, beliau mendefinisikan SPK sebagai produk dari proses perkembangan di mana pengguna SPK, pembangun SPK, dan SPK itu sendiri mampu mempengaruhi satu sama lain, mengakibatkan evolusi sistem dan penggunaan pola.
- e. Dunham (2002) mendefinisikan SPK adalah sistem komputer yang komprehensif dan alat-alat yang saling terkait yang membantu manajer dalam membuat keputusan dan pemecahan masalah. Tujuannya adalah untuk meningkatkan proses pengambilan keputusan dengan menyediakan informasi spesifik yang diperlukan oleh manajemen.

Dari beberapa definisi di atas dapat ditarik satu definisi tentang SPK yaitu sebuah sistem berbasis komputer yang adaptif, fleksibel, dan interaktif yang digunakan untuk memecahkan masalah-

masalah tidak terstruktur sehingga meningkatkan nilai keputusan yang diambil (Khoiruddin, 2008).

#### **2.2.2.2. Manfaat Sistem Pendukung Keputusan**

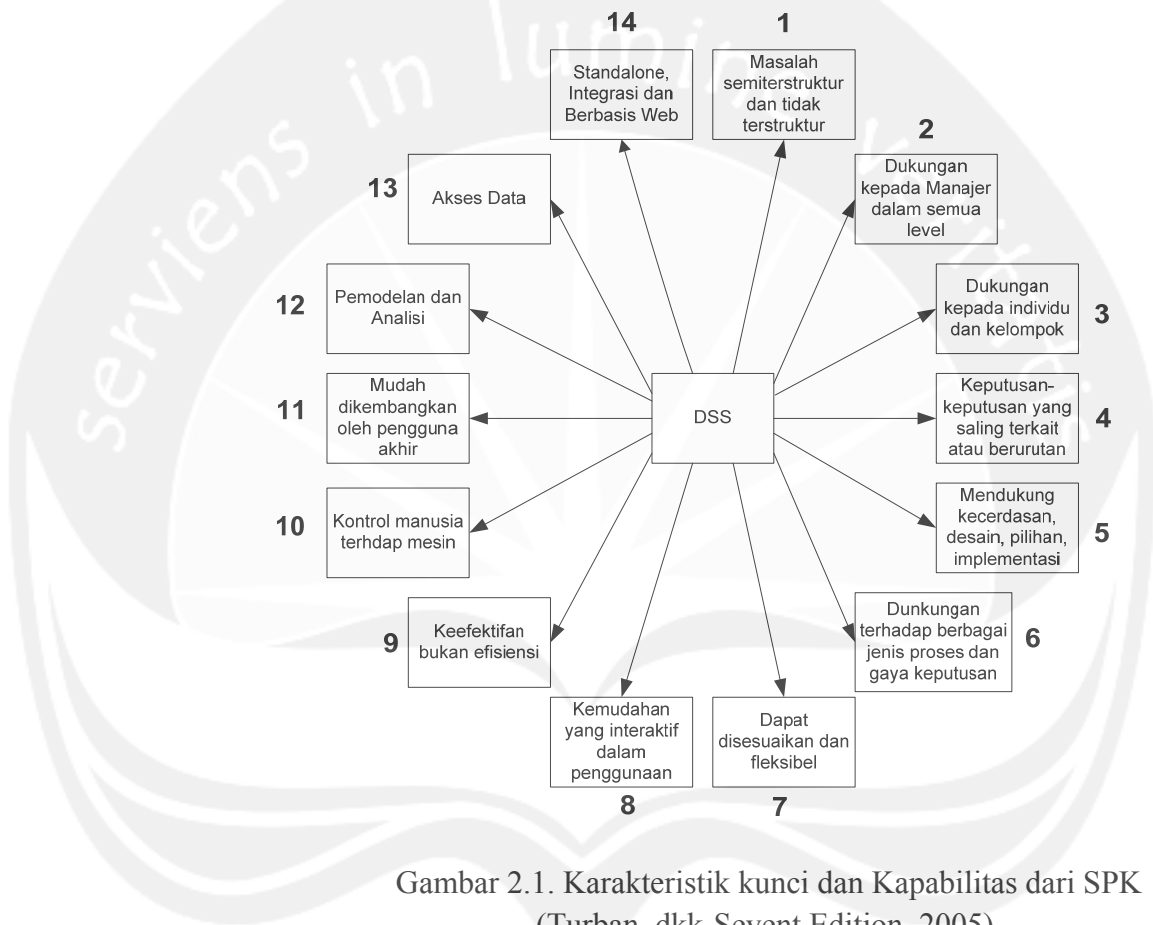
Menurut Tromp dan Wegner (1990), ada berbagai manfaat dalam penggunaan SPK, yaitu:

- a. Keuntungan administrasi (penghematan waktu dan tenaga).
- b. Efisiensi dalam penggunaan manajemen waktu.
- c. Pembuatan keputusan yang lebih baik dalam menangani masalah yang lebih kompleks.
- d. Pengenalan terhadap masalah yang lebih baik sehingga ada kemantapan yang tinggi dalam membuat keputusan.
- e. Penggunaan data yang lebih baik (aktual, akurasi, akses dan ketersediaan)
- f. Perencanaan dan kontrol yang lebih baik.
- g. Pencarian berbagai alternatif yang lebih baik lagi.
- h. Keuntungan komunikasi.



### 2.2.2.3. Karakteristik Dan Kapabilitas Dari Sistem Pendukung Keputusan

Karakteristik dan kapabilitas dari SPK dapat dilihat dalam gambar 2.1 dibawah ini.



Gambar 2.1. Karakteristik kunci dan Kapabilitas dari SPK (Turban, dkk-Sevent Edition, 2005).

1. Dukungan bagi para pengambil keputusan, terutama pada situasi semiterstruktur dan tidak terstruktur, dengan menyertakan penilaian manusia dan informasi terkomputerisasi. Masalah-masalah tersebut tidak dapat dipecahkan (atau tidak dapat

dipecahkan dengan mudah) oleh sistem komputerisasi lain atau alat-alat atau metode-metode kuantitatif yang digunakan.

2. Dukungan untuk semua level manajerial, dari eksekutif atas sampai manajer lini.
3. Dukungan untuk individu dan kelompok. Masalah yang kurang terstruktur sering memerlukan keterlibatan individu dari departemen dan tingkat organisasional yang berbeda atau bahkan dari organisasi lain. Sistem Pendukung Keputusan (SPK) mendukung tim-tim virtual melalui alat-alat web yang kolaboratif.
4. Dukungan untuk keputusan independen dan atau sekuensial. Keputusan dapat dibuat satu kali, beberapa kali, atau berulang.
5. Dukungan dalam semua fase proses pengambilan keputusan: intelegensi, desain, pilihan dan implementasi.
6. Dukungan untuk di berbagai proses dan gaya pengambilan keputusan.
7. Adoptivitas sepanjang waktu. Pengambilan keputusan seharusnya reaktif, dapat

menghadapi perubahan kondisi secara cepat, dan dapat mengadaptasikan SPK untuk memenuhi perubahan tersebut. SPK bersifat fleksibel dan karena itu pengguna dapat menambahkan, menghapus, menggabungkan, mengubah, atau menyusun kembali elemen-elemen dasar. SPK juga fleksibel dalam hal dapat dimodifikasi untuk memecahkan masalah lain yang sejenis.

8. Pengguna merasa seperti di rumah. Ramah bagi pengguna, kapabilitas-kapabilitas grafis yang kuat, dan antarmuka manusia mesin yang interaktif dengan bahasa alami dapat meningkatkan keefektifan dari SPK. Kebanyakan aplikasi SPK yang baru menggunakan antarmuka berbasis web.
9. Peningkatan terhadap keefektifan pengambilan keputusan (akurasi, timeliness, kualitas) daripada efisiensinya (biaya pengambilan keputusan). Ketika SPK digunakan, pengambilan keputusan sering membutuhkan waktu lebih lama, namun keputusannya lebih baik.

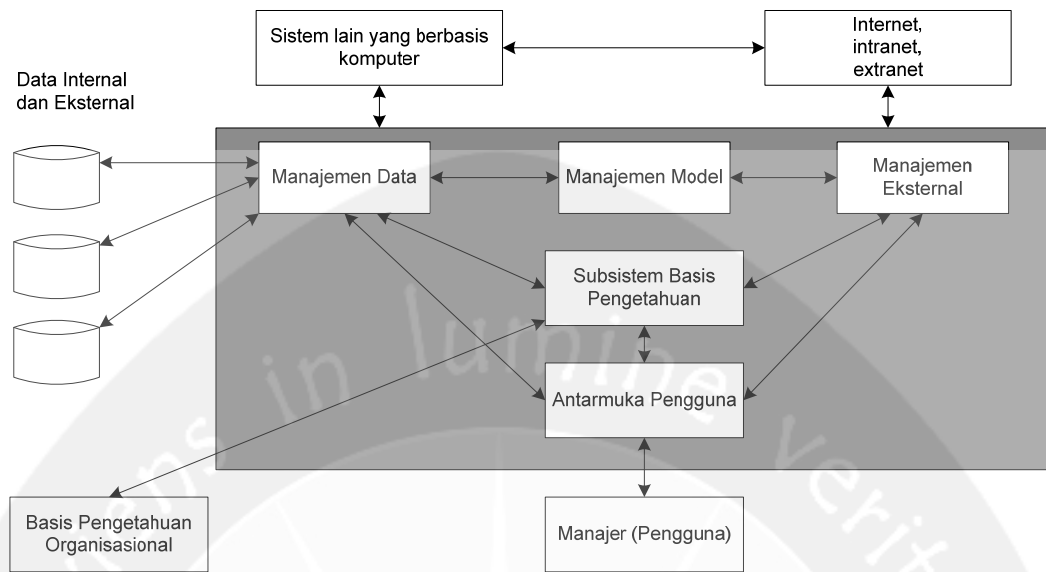
10. Kontrol penuh oleh pengambil keputusan terhadap semua langkah proses pengambilan keputusan dalam memecahkan suatu masalah. SPK secara khusus bertujuan untuk mendukung bukan menggantikan pengambil keputusan.
11. Pengguna akhir dapat mengembangkan dan memodifikasi sendiri sistem yang sederhana. Sistem yang lebih besar dapat dibangun dengan bantuan ahli sistem informasi. OLAP (*online analytical processing*) perangkat lunak dalam hubungannya dengan data *warehouse* (gudang data) memungkinkan pengguna untuk membangun sebuah sistem yang cukup besar, sebuah SPK yang kompleks.
12. Biasanya model-model digunakan untuk menganalisis situasi-situasi pengambilan keputusan. Kapabilitas pemodelan memungkinkan eksperimen dengan berbagai strategi yang berbeda di bawah konfigurasi yang berbeda. Sebenarnya, model-model

tersebut membuat SPK berbeda dari kebanyakan MIS.

13. Akses disediakan untuk berbagai sumber data, format, dan tipe, mulai dari Sistem Informasi Geografis (SIG) sampai sistem berorientasi objek.
14. Dapat dipakai sebagai alat standalone (berdiri sendiri) yang digunakan oleh seorang pengambil keputusan pada satu lokasi atau didistribusikan di satu organisasi keseluruhan dan di berbagai organisasi sepanjang rantai pasok. Dapat diintegrasikan dengan SPK lain dan atau aplikasi lain, dan dapat didistribusikan secara internal dan eksternal, menggunakan jaringan dan teknologi web.

#### **2.2.2.4. Bagan dari Sistem Pendukung Keputusan**

Sebuah SPK dapat terdiri dari subsistem-subsistem yang dapat dilihat dalam gambar 2.2 dibawah ini.



Gambar 2.2. Bagan dari SPK (Turban, dkk-Sevent Edition, 2005).

**Subsistem Manajemen Data.** Subsistem manajemen data meliputi basis data yang berisi data yang berkaitan dengan sebuah kondisi yang dikelola oleh perangkat lunak yang disebut sistem manajemen basis data (*Database Manajemen System*) atau DBMS. Subsistem manajemen basis data bisa saling berhubungan dengan gudang data perusahaan, *repository* data untuk pengambilan keputusan perusahaan untuk masalah yang terkait. Biasanya data tersimpan atau diakses melalui web server.

**Subsistem Manajemen Model.** Merupakan paket perangkat lunak yang meliputi keuangan, statistik, ilmu manajemen, atau model kuantitatif

lain yang memberikan kemampuan analitis dan manajemen perangkat lunak yang sesuai. Bahasa pemodelan untuk membangun model model khusus juga disertakan. Perangkat lunak ini sering disebut basis model manajemen sistem (*model base management system*) atau MBMS. Komponen ini dapat dihubungkan ke model penyimpanan eksternal perusahaan. Metode-metode solusi model dan sistem-sistem manajemen sudah diimplementasikan dalam sistem pengembangan web (seperti Java) untuk dijalankan dalam server-server aplikasi.

#### **Subsistem Antarmuka Pengguna.**

Pengguna berkomunikasi dan memberikan perintah kepada SPK melalui subsistem ini. Pengguna dianggap bagian dari sistem ini. Para peneliti menyatakan bahwa beberapa kontribusi yang unik dari SPK berasal dari interaksi yang intensif antara komputer dan pembuat keputusan. *Browser web* semakin familiar, dengan grafik struktur antarmuka yang konsisten untuk kebanyakan SPK.

**Subsistem Manajemen Berbasis Pengetahuan.** Subsistem ini dapat mendukung

berbagai subsistem-subsistem lain atau bertindak sebagai komponen yang bebas. Subsistem ini memberikan kecerdasan tambahan bagi si pembuat keputusan. Dapat dihubungkan dengan *repository* pengetahuan organisasi (bagian dari sistem manajemen pengetahuan), kadang disebut basis pengetahuan organisasi. Pengetahuan bisa disediakan lewat server-server web. Banyak metode-metode kecerdasan buatan yang sudah diimplementasikan dalam sistem-sistem pengembangan web seperti Java, dan mudah diintegrasikan kedalam komponen-komponen SPK yang lain.

Menurut definisi, SPK harus mencakup tiga komponen utama dari DBMS, MBMS, dan antarmuka pengguna. Subsistem manajemen berbasis pengetahuan berdifat opsional, namun dapat memberikan banyak manfaat dengan menyediakan kecerdasan kedalam tiga komponen utama. Seperti dalam kebanyakan sistem informasi manajemen, pengguna dapat dianggap sebagai komponendari SPK.



## 2.2.3. Penambangan Data

### 2.2.3.1. Definisi Penambangan Data

Penambangan data atau *Data Mining* adalah suatu istilah untuk menguraikan penemuan pengetahuan dalam basis data (Kusrini dan Luthfi, 2009). Istilah penambangan data memiliki beberapa padanan, seperti *knowledge discovery* (penemuan pengetahuan) yaitu bertujuan untuk menemukan pengetahuan yang masih tersembunyi di dalam bongkahan data besar (Al-Hegami, 2007), ataupun *pattern recognition* (pengenalan pola) yaitu pengetahuan yang hendak digali memang berbentuk pola-pola yang juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi (Susanto dan Suryadi, 2010). Tujuan digunakan teknik penambangan data adalah menghasilkan informasi baru bagi para pembuat keputusan dari basis data yang sangat besar (Rastegari dan Noor, 2008). Teknik, metode, atau algoritma dalam penambangan data sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD (*knowledge discovery in*

*databases*) secara keseluruhan (Elmande dan Widodo, 2012).

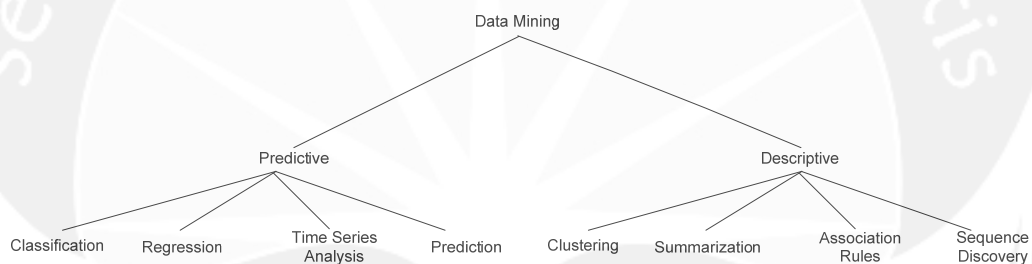
Berikut adalah beberapa pengertian penambangan data:

1. Menurut Gartner Group (Larose, 2005) adalah proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan data besar yang tersimpan dalam penyimpanan dengan menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika.
2. Tsipsis dan Chorianopoulos (2009) mendefinisikan Penambangan data bertujuan untuk mengambil (mengeksktrak) pengetahuan dan pengetahuan analisis dari sejumlah besar data menggunakan teknik pemodelan yang canggih. Penambangan data mengkonversi data menjadi pengetahuan dan informasi yang bisa ditindaklanjuti.
3. Dunham (2002), Penambangan data sering didefenisikan sebagai penemuan informasi yang tersembunyi dalam basis data. Penambangan data menggunakan algoritma

untuk mengekstrak informasi dan pola yang berasal dari proses penemuan pengetahuan dari basis data.

### 2.2.3.2. Model dan Tugas Dalam Penambangan Data

Mengacu pada Dunham (2002), model-model dalam penambangan data dapat dikelompokkan menjadi dua (Gambar 2.1) bagian yaitu:



Gambar 2.3. Model dan Tugas Penambangan Data (Dunham, 2002).

#### 1. Model Prediktif (*Predictive*)

Model ini berfungsi melakukan prediksi nilai dari data dengan menggunakan hasil yang sudah diketahui dari data yang berbeda. Model prediksi dapat dibuat berdasarkan pada penggunaan data histori.

Yang termasuk dalam model prediksi adalah:

- a. *Classification* (Klasifikasi),

- b. *Regression* (Regeresi),
- c. *Time Series Analysis* (Analisis Runtun Waktu),
- d. *Prediction* (Prediksi).

2. Model Deskriptif (*Descriptive*)

Model ini mengidentifikasi pola atau hubungan dalam data. Berbeda dengan model prediksi, model deskripsi digunakan sebagai jalan untuk mengeksplorasi sifat-sifat dari data yang sudah diperiksa, tidak untuk memprediksi sifat-sifat baru.

Yang termasuk dalam model prediksi adalah:

- a. *Clustering* (Pengelompokan),
- b. *Summarization* sering kali juga disebut *Characterization* (Karakterisasi) atau *Generalization* (Generalisasi),
- c. *Association Rules* (Aturan-Aturan Asosiasi),
- d. *Sequence Discovery*.

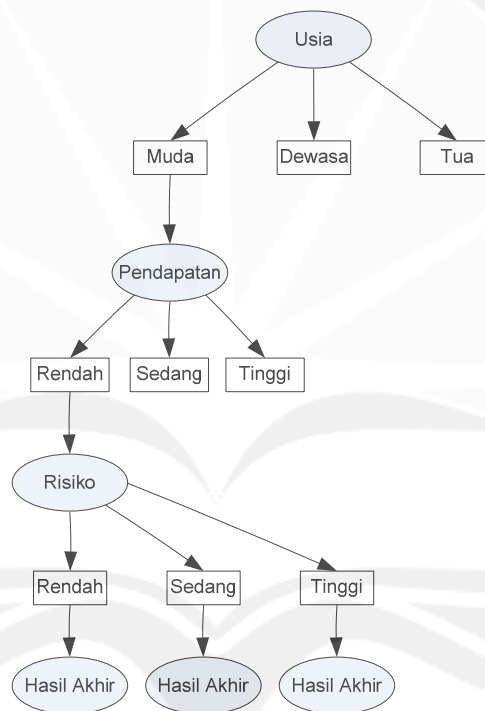
#### 2.2.4. Klasifikasi

Klasifikasi merupakan salah satu bentuk dari model prediksi. Klasifikasi memetakan data kedalam kelompok-kelompok atau kelas-kelas yang sudah diketahui. Sering disebut sebagai pembelajaran yang diawasi karena kelas sudah ditentukan sebelum melakukan pemeriksaan data (Dunham, 2002). Pengelompokan data ke dalam kelas-kelas biasanya dengan melihat pada kesamaan karakteristik dari data tersebut.

#### 2.2.5. Pohon Keputusan

Pohon keputusan adalah salah satu metode belajar yang sangat populer dan banyak digunakan secara praktis (Wahyudin, 2009). Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit (Mitchell, 1997) dan tahan terhadap data-data yang terdapat kesalahan (*noisy data*) serta mampu mempelajari ekspresi-ekspresi *disjunctive* (ekspresi OR). Pohon keputusan memiliki struktur seperti pohon, dimana setiap *node* pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan *node* daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah pohon keputusan adalah *node* akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya pohon

keputusan melakukan strategi pencarian secara *top-down* (atas-bawah) untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai node akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu. Bentuk pohon keputusan seperti terlihat dalam gambar 2.4 dibawah ini menurut Olson dan Shi, 2008.



Gambar 2.4. Pohon Keputusan.

Gambar diatas menunjukkan bahwa kriteria pada simpul yang paling atas adalah akar dari pohon tersebut, kemudian dicari akar selanjutnya menggunakan cabang (kriteria) yang ada dalam simpul akar. Gambar 2.4 menggambarkan pohon keputusan data permohonan pinjaman di bank.

Pada pohon keputusan setiap simpul daun menandai label kelas (Setiawan, 2010). Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*).

#### **2.2.6. Algoritma *Iterative Dichotomiser 3* (ID3)**

*Iterative Dichotomiser 3* (ID3) adalah algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang paling dasar (Wahyudin, 2009). Tujuan Algoritma ID3 adalah untuk membangun pohon keputusan berdasarkan pada teori informasi dan upaya untuk meminimalkan jumlah yang diharapkan dari perbandingan (Dunham, 2002). Algoritma ini melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma ID3 berusaha membangun pohon keputusan (*decision tree*) secara *top-down* (dari atas ke bawah), mulai dengan pertanyaan: “atribut mana yang pertama kali harus dicek dan diletakkan pada *root* (akar)?” pertanyaan ini dijawab dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistik (yang banyak digunakan adalah *information gain*) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data. Atribut yang

terbaik akan dipilih dan digunakan sebagai akar dari pohon (Mitchell, 1997). Karakteristik ID3 adalah membangun pohon keputusan dari simpul akar ke daun, sementara *divide and conquer* artinya *training data* secara rekursif dipartisi ke dalam bagian-bagian yang lebih kecil saat membangun pohon.

#### 2.2.6.1. Entropy

Entropy (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas.

**Rumus mencari Entropy :**

$$\text{Entropy}(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

Keterangan

S : ruang (data) *sample* yang digunakan untuk training.

P<sub>+</sub> : jumlah yang bersolusi positif (mendukung) pada data *sample* untuk kriteria tertentu.

P<sub>-</sub> : jumlah yang berolusi negatif (tidak



mendukung) pada data *sample* untuk kriteria tertentu.

### 2.2.6.2. Information Gain

Setelah mendapat nilai *entropy* untuk suatu kumpulan data, maka kita dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut *information gain*. Secara matematis, *information gain* dari suatu atribut A, dituliskan sebagai berikut :

**Rumus mencari *Information Gain* :**

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan:

A: atribut.

V: menyatakan suatu nilai yang mungkin untuk atribut A.

Values (A): himpunan yang mungkin untuk atribut A.

|S<sub>v</sub>|: jumlah sampel untuk nilai v.

|S|: jumlah seluruh sampel data.

Entropy (S<sub>v</sub>): entropy untuk sampel-sampel yang memiliki nilai v.

### 2.2.6.3. Pembentukan Pohon Keputusan Menggunakan Algoritma ID3

Berikut ini adalah langkah-langkah dalam pembentukan pohon keputusan menggunakan Algoritma ID3.

ID3 (Sampel Training, Label Training, Atribut)

- Membuat simpul akar untuk pohon yang dibuat.
- Jika semua sampel positif, berhenti dengan satu simpul akar, dengan label +.
- Jika semua sampel negatif, berhenti dengan satu simpul akar, dengan label -.
- Jika atribut kosong, berhenti dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label training.
  - Untuk yang lain, mulai
  - $A \leftarrow$  atribut yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan gain ratio).
  - Atribut pohon keputusan untuk akar  $\leftarrow A$ .
  - Untuk setiap nilai yang mungkin,  $v_i$  dari  $A$ ,
    - Tambah cabang dibawah akar yang berhubungan dengan  $A = v_i$ .
    - Tentukan sampel ( $S_i$ ), sebagai subset dari sampel yang mempunyai nilai  $v_i$  untuk  $A$ .
    - Jika sampel ( $S_i$ ) kosong,
      - Maka dibawah cabang tambahkan simpul daun dengan label = nilai terbanyak yang ada dalam label training.
      - Yang lain, tambah cabang baru dibawah cabang yang sekarang ID3 (sampel training, label training, atribut-[A]).
    - Berhenti.
  - Kembali ke Akar.