BAB 2
TINJAUAN PUSTAKA

2.1 Tinjauan Sekolah

Berikut ini adalah tinjauan SMA Pasundan 2 Bandung dari sejarah, visi dan misi, struktur organisasi serta deskripsi pekerjaannya.

2.1.1 Sejarah Sekolah

SMA Pasundan 2 Bandung berdiri sejak tahun 1980. Keberadaan SMA Pasundan 2 Bandung tidak lepas dari tujuan Yayasan Pendidikan dasar dan Menengah Pasundan dalam upaya mencerdaskan kehidupan bangsa Indonesia

Sekarang SMA Pasundan 2 Bandung telah berkembang menjadi sekolah yang besar, yaitu sekolah yang menjadi pilihan masyarakat, hal ini terbukti dari adanya minat siswa baru kelas X yang mendaftar ke SMA Pasundan 2 Bandung dari tahun ke tahun mengalami peningkatan dan kepercayaan Badan Akreditasi Sekolah yang memberikan status TERAKREDITASI Peringkat A (amat baik) berturut-turut selama 3 periode (2003 dan 2007 dan 2012).

Kurikulum yang berlaku mengacu pada Kurikulum Nasional dan ditunjang dengan muatan lokal yaitu yang terkait dengan visi, misi dan tujuan yayasan pendidikan dasar dan menengah Pasundan, yaitu mengagungkan syiar Islam dan nilai-nilai budaya sunda.

Selain kelas REGULER, di SMA Pasundan 2 Bandung telah dibuka kelas UNGGULAN yang telah dipersiapkan untuk Lolos Perguruan Tinggi Negeri (PTN) tanpa tes, dan sekarang sedang dipersiapkan menuju Sekolah Mandiri oleh Dinas Pendidikan Kota Bandung.

Kegiatan belajar dilaksanakan oleh staf pengajar yang memiliki kualifikasi S1 dan S-2 dan S3 berlangsung pada PAGI HARI, kegiatan belajar akan berlangsung lima hari kerja dengan sistem *full day*.

Sampai dengan tahun 20015, SMA Pasundan 2 Bandung telah menghasilkan lulusan yang diterima di berbagai Perguruan Tinggi favorit, dan terdapat juga alumni yang berprestasi dalam berbagai *event*, seperti Akademi

Fantasi Indosiar (AFI) a.n. Indri, *Super Star Show* a.n Rina Nose, Adriansyah sebagai Tim Futsal Nasional Dll

Berbekal kepercayaan, kerendahan hati, komitmen, kesungguhan dan keyakinan untuk berbuat lebih baik dimasa yang akan datang, pada tahun 2015 ini SMA Pasundan 2 Bandung kembali membuka pendaftaran Peserta Didik Baru untuk dididik menjadi siswa yang pengekuh agamana (ta'at beragama), luhung elmuna (memiliki ilmu pengetahuan), jembar budayana (melestarikan nilai-nilai budaya sunda).

2.1.2 Visi dan Misi Sekolah

Berikut ini adalah visi dan misi dari SMA Pasundan 2 Bandung:

1. Visi

   SMA Pasundan 2 Bandung mengembangkan pendidikan berkualitas unggul dalam bidang keilmuan, moralitas, mentalitas berdasarkan keisaman dan budaya sunda yang mampu bersaing di tingkat lokal, nasional dan internasional.

2. Misi

   Mendidik sumber daya manusia unggul yang menguasai, memahami, menghayati bidang keilmuan yang ditekuni dengan dilandasi nilai keimanan dan budaya sunda.

   Memberi kontribusi terhadap peningkatan kualitas yang mampu mengaplikasikan bidang keilmuan, nilai-nilai keislaman dan budaya sunda.

   Mengembangkan bidang keilmuan dan tekhnologi informasi yang disertai nilai-nilai kehidupan masyarakat serta paham aktualisasi nilai-nilai budaya sunda dan agama islam sebagai implementasi perwujudan ibadah pada Allah SWT.

   Pengembangan keilmuan dan nilai budaya sunda yang dilandasi nilai-nilai keislaman dalam implementasi perwujudan puncak budaya nasional yan berakar pada budaya daerah

Memberi kontribusi dalam pembangunan nasional yang peka dan dinamis terhadap perubahan globalisasi.

Melaksanakan pengembangan keislaman, budaya sunda, keilmuan dan teknologi melalui pedekatan Silih Asah, Silih Asih, Silih, Silih Asuh.

2.1.3 Landasan Teori

Landasan teori akan membahas tentang berbagai teori yang menjadi landasan untuk pembangunan tugas akhir ini. Pembahasan teori tersebut bertujuan sebagai penjelas teori-teori yang dipakai pada saat penelitian tugas akhir ini disusun. Teori-teori yang akan dibahas segala hal yang terlibat dalam pembangunan aplikasi ini seperti *data mining, algoritma ID3*, beberapa perangkat lunak yang terlibat dan lain-lain yang bersangkutan dengan pembangunan aplikasi ini.

2.1.4 *Data Mining*

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 data yang besar [10]. *Data mining* dapat juga didefinisikan sebagai analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga untuk dapat dipahami dan bermanfaat bagi pemilik data [11]. *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dilakukan, yaitu :

1. Klasifikasi (*Classification*)

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek. Selain itu, klasifikasi digunakan untuk melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan dari nilai yang kontinu. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi datayang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi[45].

2. Prediksi (*Prediction*)

Model dengan tipe regresi ini menggunakan serangkaian nilai-nilai atribut yang ada untuk memprediksi suatu nilai kontinu.

3. Asosiasi (Association)

Asosiasi adalah teknik *data mining* untuk menemukan aturan asosiatif antara kombinasi *item*. Metode aturan asosiasi membuat aturan berdasarkan kondisi yang sering terjadi. Aturan yang dibuat untuk data dalam kondisi X, diharapkan dapat diterapkan untuk data-data lainnya dalam kondisi Y. Analisis korelasi juga termasuk ke dalam kelompok ini, dimana tujuan dari analisis ini adalah menemukan hubungan atau tingkat kedekatan antar variabel.

4. Deskriptif

Deskriptif bertujuan untuk menggambarkan pola dan kecenderungan yang terdapat dalam data serta memberikan penjelasan pola atau kecenderungan tersebut.

5. Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan data tanpa berdasarkan kelas data tertentu. *Clustering* dapat digunakan untuk memberikan label pada kelas data yang belum diketahui.

6. Estimasi

Variabel yang digunakan estimasi berupa numerik. Tujuan dari estimasi adalah menghasilkan suatu model yang menyediakan nilai variabel dari target sebagai nilai prediksi, selanjutnya peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

2.1.5 Penggolongan Atribut

2.1.5.1 Atribut Katagorikal (*Categorical Attribute*)

Adalah salah satu tipe atribut yang domainnya merupakan sebuah himpunan symbol berhingga (memiliki batas) Contoh:

- **a.** Jenis kelamin = {L,P}
- **b.** Status = {menikah, belum menikah}
- **c.** Pendidikan = {SD,SMP,SMA,D3,S1,S2,S3,lainnya}

2.1.5.2 Atribut Numerik (*Numeric Attribute*)

Adalah salah satu tipe atribut yang domainnya berupa bilangan riil atau integer. Contoh : umur dan gaji yang merupakan bilangan riil positif. Atribut numeric dibedakan menjadi 2, yaitu:

a. Interval : atribut ini mempunyai sifat bahwa pebedaan antara nilai-nilainya sangat berarti. Contoh: tanggal, suhu dalam celcius atau Fahrenheit, karena tidak ada bedanya jika kita menyatakan bahwa 20°C = dua kali dinginnya 10°C

b. Rasio : dalam atibut jenis ini, baik beda maupun rasio sangat berarti. Contoh : suhu dalam Kelvin, panjang, waktu, jumlah. Kita dapat menyatakan bahwa orang berumur 20 tahun dua kali lebih tua dari yang berusia 10 tahun.

2.1.5.3 Atribut Diskrit (*Discrete Attribute*)

Yaitu atribut yang hanya menggunakan sebuah himpunan nilai berhingga atau himpunan nilai tak berhingga yang dapat dihitung. Contoh : jumlah atau himpunan kata dalam kumpulan dokumen. Sering dinyatakan sebagai variable bilangan bulat (integer).

2.1.5.4 Atribut Kontinu (*Continuous Attribute*)

Yaitu atribut yang menggunakan bilangan riil sebagai nilai atribut. Contoh: suhu, ketinggian atau berat.

Pada kenyataannya nilai riil hanya dapat diukur dan dinyatakan menggunakan sejumlah digit yang berhingga. Atribut kontinyu secara khusus dinyatakan sebagai variable decimal (floating-point)

2.1.6 Klasifikasi Variabel

Variabel-variabel yang telah diidentifikasikan perlu diklasifikasikan, sesuai dengan jenis dan peranannya dalam penelitian. Klasifikasi ini sangat perlu untuk penentuan alat pengambilan data apa yang akan digunakan dan metode analisis mana yang sesuai untuk diterapkan.

Berkaitan dengan proses kuantifikasi data biasa digolongkan menjadi 4 jenis yaitu Data Nominal, Data Ordinal, Data Interval, Data Ratio. Demikianlah pula variabel, kalau dilihat dari segi ini biasa dibedakan dengan cara yang sama[13][14][15][16]:

1. Variabel Nominal, yaitu variabel yang ditetapkan berdasar atas proses penggolongan; variabel ini bersifat diskret dan saling pilah (*mutually exclusive*) antara kategori yang satu dan kategori yang lain; contoh: jenis kelamin, status perkawinan, jenis pekerjaan
2. Variabel Ordinal, yaitu variabel yang disusun berdasarkan atas jenjang dalam atribut tertentu. Jenjang tertinggi biasa diberi angka 1, jenjang di bawahnya diberi angka 2, lalu di bawahnya di beri angka 3 dan seterusnya. (ranking)
3. Variabel Interval, yaitu variabel yang dihasilkan dari pengukuran, yang di dalam pengukuran itu diasaumsikan terdapat satuan (unit) pengukuran yang sama. Contoh: variabel interval misalnya prestasi belajar, sikap terhadap sesuatu program dinyatakan dalam skor, penghasilan dan sebagainya.
4. Variabel ratio, adalah variabel yang dalam kuantifikasinya mempunyai nol mutlak.

Menurut Fungsinya variabel dapat dibedakan :

a. Variabel Tergantung (*Dependen*t Variabel)

Yaitu kondisi atau karakteristik yang berubah atau muncul ketika penelitian mengintroduksi, pengubah atau mengganti variabel bebas. Menurut fungsinya variabel ini dipengaruhi oleh variabel lain, karenanya juga sering disebut variabel yang dipengaruhi atau variabel terpengaruhi. Variabel ini sering disebut sebagai variabel output, Kriteria, Konsekuen. Atau dalam bahasa Indonesia sering disebut Variabel terikat. Dalam SEM (Structural Equation Modeling) variabel dependen disebut variabel Indogen.*

b. Variabel Bebas (*Independen*t Variabel)

Adalah kondisi-kondisi atau karakteristik-karakteristik yang oleh peneliti dimanipulasi dalam rangka untuk menerangkan hubungannya dengan fenomena yang diobservasi. Karena fungsi ini sering disebut variabel

pengaruh, sebab berfungsi mempengaruhi variabel lain, jadi secara bebas berpengaruh terhadap variabel lain. Variabel ini juga sering disebut sebgai variabel Stimulus, Prediktor, antecendent. Dalam SEM(*Structural Equation Modeling*) variabel independen disebut variabel eksogen.

c. Variabel *Intervening*

Variabel intervenig adalah variabel yang secara teoritis mempengaruhi hubungan antara variabel independen dengan Variabel dependen menjadi hubungan yang tidak langsung dan tidak dapat diamati dan diukur. Variabel ini merupakan variabel penyela/antara yang terletak di antara variabel independen dan dependen, sehingga variabel independen tidak langsung mempengaruhi berubahnya atau timbulnya variabel dependen. Variabel *Intervening* juga merupakan variabel yang berfungsi menghubungkan variabel satu dengan variabel yang lain. Hubungan itu dapat menyangkut sebab akibat atau hubungan pengaruh dan terpengaruh.

d. Variabel Moderator

Dalam mengidentifikasi variabel moderator dimaksud adalah variabel yang karena fungsinya ikut mempengaruhi variabel tergantung serta meperjelas hubungan bebas dengan variabel tergantung.

e. Variabel kendali

Yaitu yang membatasi (sebagai kendali) atau mewarnai variabel mederator. Variabel ini berfungsi sebagai kontrol terhadap variabel lain terutama berkaitan dengan variabel moderator jadi juga seperti variabel moderator dan bebas ia juga ikut berpengaruh terhadap variabel tergantung

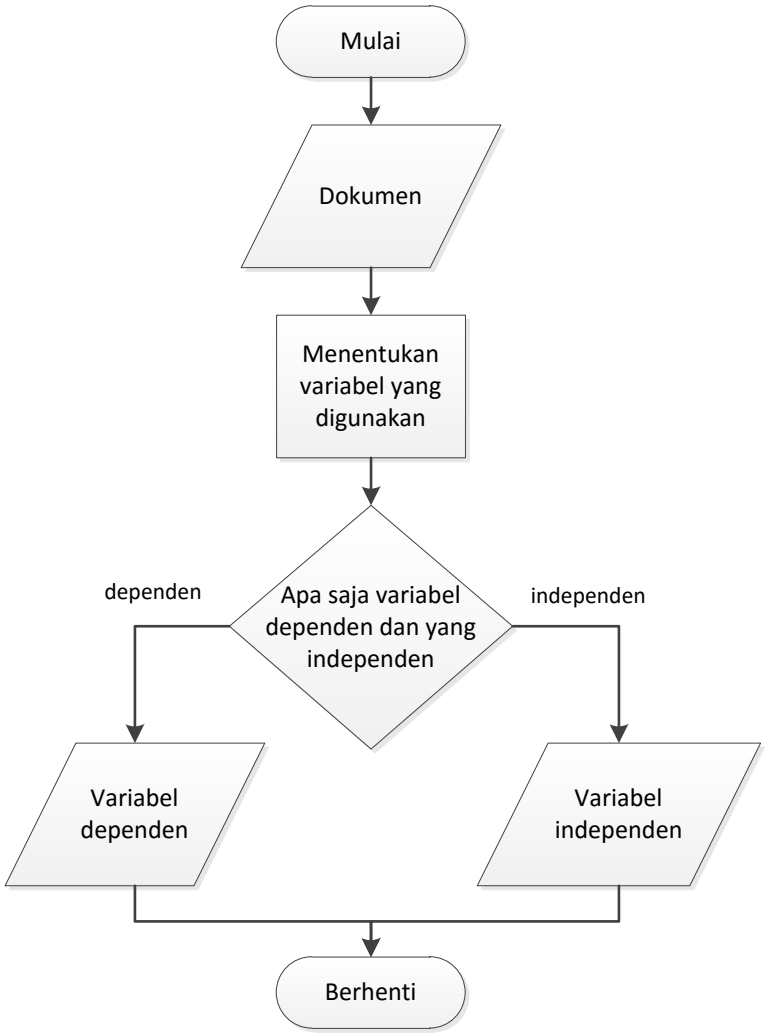
f. Variabel Rambang

Berlainan dengan variabel bebas, yaitu fungsinya sangat diperhatikan dalam penelitian. Variabel rambang yaitu variabel yang fungsinya dapat diabaikan atau pengaruhnya hampir tidak diperhatikan terhadap variabel bebas maupun tergantung.

2.1.7 Representasi Data

Data yang ingin diteliti terlebih dahulu dipersiapkan. Kemudian menentukan variabel-variabel yang akan digunakan. Setelah itu, variabel-variabel

dikelompokkan ke variabel *dependen* atau *independen*. Pada Gambar 2.1 digambarkan alur proses penentuan variabel *dependen* dan *independen*.



Gambar 2.1 *Flowchart* Menentukan Variabel *Dependen* dan *Independen*

2.1.8 Bidang Ilmu *Data Mining*

Data mining bukanlah ilmu yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu.

Istilah data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut

memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut [1].

2.1.8.1 *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.

2.1.8.2 Pre-Processing

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pres-processing pada data yang menjadi fokus KDD.

- *Cleaning*

Proses *cleaning* mencakup pemeriksaan data yang inkonsisten, dan memperbaiki kesalahan pada data noise data. Beberapa algoritma data mining tidak dapat mengatasi data noise sehinga diperlukan penanganan terlebih dahulu. Beberapa cara untuk menangani data noise dapat dilakukan dengan mengisi:

[a] Nilai yang paling sering mucul. Tekhnik ini cocok untuk data nominal atau ordinal

[b] Nilai rata-rata seluruh data atau nilai rata-rata per kelompok. Tekhnik ini cocok untuk data ratio dan interval.

[c] Nilai perhitungan regresi. Tekhnik ini cocok untuk data ratio dan interval.

2.1.8.3 Transformation

Coding adalah proses transformasi pada tahap yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

2.1.8.4 *Data Mining*

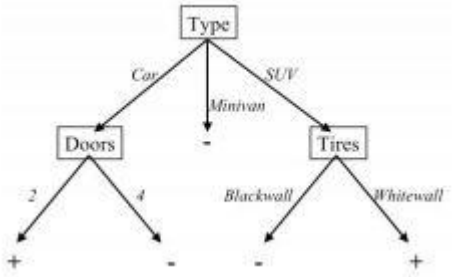
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

2.1.8.5 Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *Interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditentukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.1.9 *Decision Tree*

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Contoh dari pohon keputusan dapat dilihat di **Gambar II-2**. Di sini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklasifikasikan *record* yang belum ada kelasnya. Dimulai dari *node root*, menggunakan tes terhadap atribut dari *record* yang belum ada kelasnya tersebut lalu mengikuti cabang yang sesuai dengan hasil dari tes tersebut, yang akan membawa kepada internal *node* (*node* yang memiliki satu cabang masuk dan dua atau lebih cabang yang keluar), dengan cara harus melakukan tes lagi terhadap atribut atau *node* daun. *Record* yang kelasnya tidak diketahui kemudian diberikan kelas yang sesuai dengan kelas yang ada pada *node* daun. Pada pohon keputusan setiap simpul daun menandai label kelas



Gambar 2.2 Contoh Pohon Keputusan

Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*)[11].

2.1.10 *ID3*

Iterative Dichotomiser 3 (*ID3*) merupakan sebuah metode yang digunakan untuk membangkitkan pohon keputusan. Algoritma pada metode ini berbasis pada Occam's razor: lebih memilih pohon keputusan yang lebih kecil (teori sederhana) dibanding yang lebih besar. Tetapi tidak dapat selalu menghasilkan pohon keputusan yang paling kecil dan karena itu Occam's razor bersifat *heuristik*. Occam's razor diformalisasi menggunakan konsep dari *entropi* informasi. Berikut algoritma dari *ID3* :

Pemillihan atribut pada *ID3* dilakukan dengan properti statistik, yang disebut dengan *information Gain*. *Gain* mengukur seberapa baik suatu atribut memisahkan *training example* ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan *Gain*, pertama-tama digunakanlah ide dari teori informasi yang disebut *entropi*. *Entropi* mengukur jumlah dari informasi yang ada pada atribut.

Rumus menghitung *entropi* informasi adalah:

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

Rumus untuk menghitung *Gain* adalah:

$$\text{Gain}(S,A) \equiv \text{Entropy}(S) - \sum \frac{|Sv|}{|S|} \text{Entropy}(Sv)$$

[11].

Input : Data Training

• Membuat Root node dari hasil perhitungan Gain

• Membuat simpul akar untuk pohon yang dibuat

• Jika semua sampel positif, berhenti dengan suatu pohon dengan satu simpul akar, beri label (+)

• Jika semua sampel negatif, berhenti dengan suatu pohon dengan satu simpul akar beri label (-)

• Jika atribut kosong. berhenti dengan suatu pohon dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label *training*

• Untuk yang lain, Mulai

• A <- atribut yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan *Gain ratio*)

• Atribut keputusan untuk simpul akar <- A

• Untuk setiap nilai. v,, yang mungkin untuk A,

- Tambahkan cabang di bawah akar yang berhubungan dengan A =v,

- Tentukan sampel Sv, sebagai *subset* dari sampel yang mempunyai nilai v, untuk atribut A

- Jika sampel Sv, kosong,

• Di bawah cabang tambahkan simpul daun dengan label = nilai yang terbanyak yang ada pada label *training*

• Yang lain, tambah cabang baru di bawah cabang yang sekarang C4.5(sampel *training*. label *training*, atribut-[A])

• Berhenti

Adapun *sample* data yang digunakan oleh *ID3* memiliki beberapa syarat, yaitu:

1. Deskripsi atribut-nilai. Atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan.
2. Kelas yang sudah didefinisikan sebelumnya. Suatu atribut contoh harus sudah didefinisikan, karena mereka tidak dipelajari oleh *ID3*.
3. Kelas-kelas yang diskrit. Kelas harus digambarkan dengan jelas. Kelas yang kontinu dipecah-pecah menjadi kategori-kategori yang relatif, misalnya saja metal dikategorikan menjadi "*hard, quite hard, fleXIIble, soft, quite soft*".
4. Jumlah contoh (*example*) yang cukup. Karena pembangkitan induktif digunakan, maka dibutuhkan *test case* yang cukup untuk membedakan pola yang valid dari peluang suatu kejadian.

2.1.11 Basis Data

Database atau basis data adalah kumpulan data yang disimpan secara sistematis di dalam komputer dan dapat diolah atau dimanipulasi menggunakan perangkat lunak untuk menghasilkan informasi. Basis data merupakan aspek yang

sangat penting dalam sistem informasi dimana basis data merupakan gudang penyimpanan data yang akan diolah lebih lanjut [13].

Proses memasukan dan mengambil data ke dan dari media penyimpanan data memerlukan perangkat lunak yang disebut sistem manajemen basis data (DBMS). DBMS merupakan sistem perangkat lunak yang memungkinkan pengguna untuk memelihara, mengontrol dan mengakses data secara data secara praktis dan efisien.