

**CONCEPT LEARNING PADA ALGORITMA FIND-S  
DAN ALGORITMA ELIMINASI CALON  
(Kajian Teoritik)**

Skripsi

Sebagai Salah Satu Syarat Untuk Memperoleh  
Gelar Sarjana Sains  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Syarif Hidayatullah Jakarta

Oleh :

**MAHMUD DZULZALALI**

**103094029737**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI  
SYARIF HIDAYATULLAH  
JAKARTA  
2008 M / 1428 H**

**CONCEPT LEARNING PADA ALGORITMA FIND-S  
DAN ALGORITMA ELIMINASI CALON  
(Kajian Teoritik)**

Skripsi  
Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Sains  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Syarif Hidayatullah Jakarta

**Oleh :**

**MAHMUD DZULZALALI**

**103094029737**

Pembimbing I

Menyetujui,

Pembimbing II

Taufik Edy Sutanto, M.ScTech  
NIP. 150 377 447

Nina Fitriyati, M.Kom  
NIP. 150 377 448

Mengetahui,  
Ketua Program Studi Matematika

Nur Inayah, M.Si  
NIP.150 326 911

## PENGESAHAN UJIAN

Skripsi yang berjudul ”*Concept Learning* pada Algoritma *Find-S* dan Algoritma Eliminasi Calon (Kajian Teoritik)”, telah diuji dan dinyatakan lulus dalam sidang Munaqosah Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta, pada hari Selasa tanggal 8 Januari 2008. Skripsi ini telah diterima sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Program Studi Matematika.

Jakarta, 8 Januari 2008

Tim Penguji,

Penguji I

Penguji II

Ir. Alimuddin, M.M, M.T  
NIDS. 901704197201

Yanne Irene, M.Si  
NIP. 150 368 744

Mengetahui,

Dekan Fakultas Sains dan Teknologi

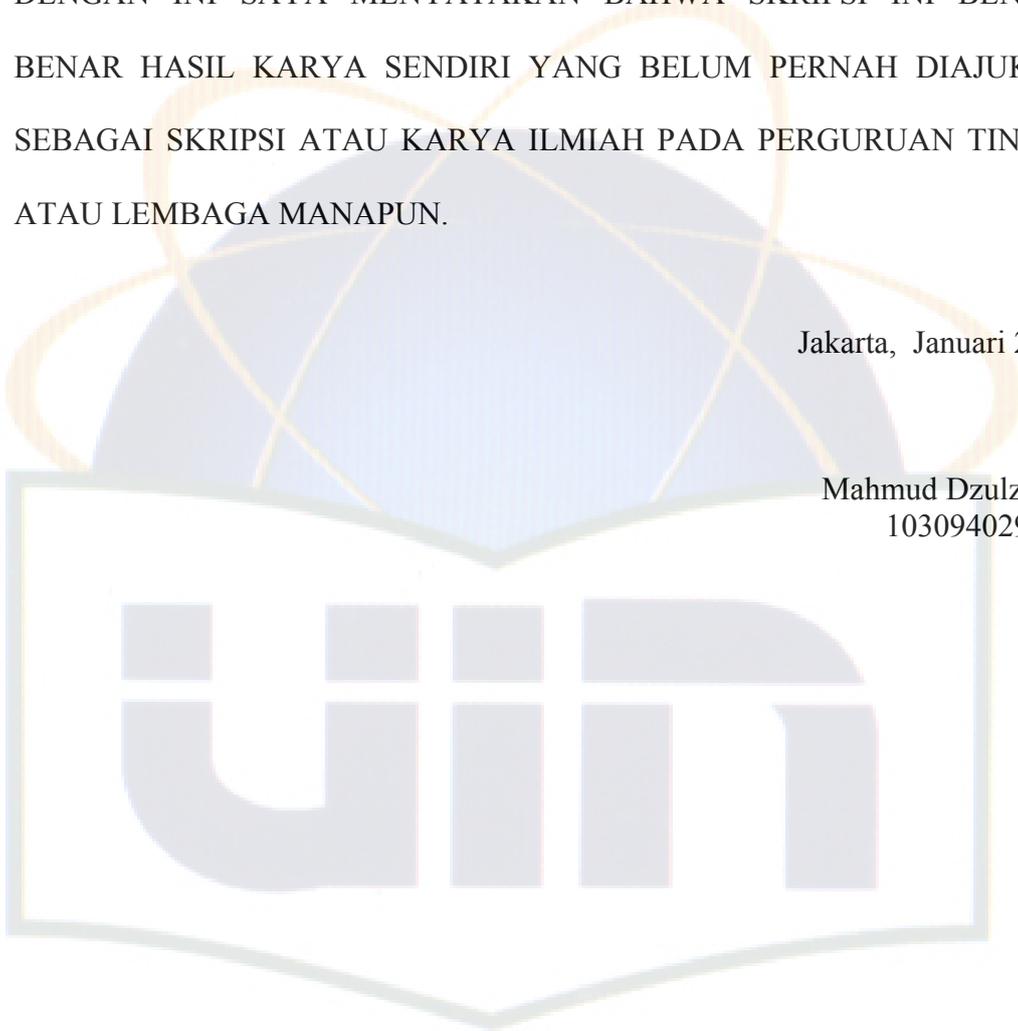
Dr. Syopiansyah Jaya Putra, M.Sis  
NIP. 150 317 956

## **PERNYATAAN**

DENGAN INI SAYA MENYATAKAN BAHWA SKRIPSI INI BENAR-BENAR HASIL KARYA SENDIRI YANG BELUM PERNAH DIAJUKAN SEBAGAI SKRIPSI ATAU KARYA ILMIAH PADA PERGURUAN TINGGI ATAU LEMBAGA MANAPUN.

Jakarta, Januari 2008

Mahmud Dzulzalali  
103094029737



## ABSTRAK

*Concept learning* merupakan sebuah pemahaman tentang pembelajaran komputer yang bertujuan untuk menentukan sebuah fungsi *boolean* dari data yang terdiri dari variabel prediktor dan variabel target. Sebuah studi mengenai *concept learning* dilakukan dengan memperhatikan ruang hipotesis, induktif bias, dan sistem pencarian hipotesis dalam ruang hipotesis.

Algoritma *Find-S* dan Algoritma Eliminasi Calon digunakan dalam skripsi ini sebagai ilustrasi dari pengaplikasian *concept learning*. Kedua algoritma ini menghasilkan hipotesis yang dapat mengklasifikasikan kejadian pada data sesuai dengan nilai variabel target. Induktif bias dari Algoritma *Find-S* dan Algoritma Eliminasi Calon dapat dilakukan apabila konsep target terdapat pada ruang hipotesis. Sehingga hipotesis yang dihasilkan dapat digeneralisasikan ke populasi.

Kata Kunci : *concept learning*, hipotesis, ruang hipotesis, induktif bias, dan konsep target



## ABSTRACT

Concept Learning inferring a boolean-value function from training examples of its input and output. Study about concept learning concerns about hypothesis space, inductive bias, and concept learning as search at hypothesis space.

Find-S Algorithm and Candidate-Eliminate Algorithm are used as illustration of concept learning application. Both of these algorithms produce hypotheses that can easily classify instances based on the value of its variable target. The Inductive Bias of Find-S Algorithm and Candidate-Eliminate Algorithm is the existence of concept target in the given hypothesis-space. Therefore the hypothesis can be generalized to population.

Key word : concept learning, hypothesis, hypothesis-space, inductive bias, and concept target.



## KATA PENGANTAR

Alhamdulillahirobbil'alamin, puji syukur penulis panjatkan kehadiran Allah S.W.T atas rahmat dan hidayah-Nya yang telah diberikan, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Salawat dan salam marilah kita curahkan kepada Nabi Muhammad saw.

Penulisan skripsi ini tidak akan terlaksana tanpa adanya dukungan, bantuan dan kerja sama dari berbagai pihak berupa bimbingan, saran, pengarahan dan fasilitas lainnya. Untuk ini penulis ingin mengucapkan terima kasih kepada :

1. Dr. Syopiansyah Jaya Putra, M.Sis, Dekan Fakultas Sains dan Teknologi.
2. Nur Inayah, M.Si, Ketua Program Studi Matematika, Nina Fitriyati, M.Kom., Sekretaris Program Studi Matematika.
3. Taufik Edy Susanto, M.ScTech., Pembimbing I
4. Nina Fitriyati, M.Kom., Pembimbing II
5. Ibunda yang telah memberikan dukungan baik secara moril, materiil dan spiritual.
6. Ayahanda, alm. Semoga Damai selalu di sisi-Nya.
7. Rekan-rekan Guru Al-Azhar Syifa Budi Jakarta yang selalu menyemangati dan memberikan masukan dalam proses belajar-mengajar.
8. Para sahabatku, Yosy, Rina, Retno, Dini, Mimi, Citra, Retno, Farida, Ida, Iin, Iqbal, Riyadi, Denis, Suparno, Sandhi, Din-Din, Selamat Menempuh Hidup Baru! Tetap Semangat! Always remember our sweet memories...

9. Ka Sopi, ka Hata, ka Abub, ka Andi, ka Bambang, ka nana, ka khaidar, dan rekan-rekan mahasiswa angkatan 2002 lainnya, thanks for everything
10. Adik-adik kelas angkatan 2004, 2005, 2006, 2007 jadilah generasi yang lebih unggul.
11. Semua pihak yang tidak bisa saya sebutkan namanya satu persatu.

Saya menyadari bahwa masih banyak kekurangan dalam penyajian materi maupun penulisan skripsi ini. Untuk itu penulis mengharapkan saran dan kritik yang bersifat membangun untuk penulisan yang akan datang.

Akhir kata, penulis mengharapkan kiranya skripsi ini dapat memberikan manfaat bagi pembaca pada umumnya dan penulis pada khususnya. Semoga Allah S.W.T senantiasa memberikan karunia-Nya yang telah diberikan kepada penulis dan pembaca. Amin.

Jakarta, Januari 2008

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>PENGESAHAN PEMBIMBING</b> .....	ii
<b>PENGESAHAN UJIAN</b> .....	iii
<b>PERNYATAAN</b> .....	iv
<b>PERSEMBAHAN DAN MOTTO</b>	
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>KATA PENGANTAR</b> .....	vii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penulisan .....	3
1.5 Manfaat Penulisan .....	4
<b>BAB II LANDASAN TEORI</b> .....	5
2.1 Himpunan .....	5
2.2 Logika Matematika .....	6
2.2.1 Konjungsi .....	7
2.2.2 Disjungsi .....	8

2.2.3	Ingkaran .....	8
2.3	Fungsi Boolean .....	9
2.4	<i>Concept Learning</i> .....	11
2.4.1	Hipotesis .....	13
2.4.2	Notasi pada <i>Concept Learning</i> .....	14
2.4.3	Keterurutan Hipotesis Umum ke Khusus .....	16
2.4.4	Pencarian Hipotesis dalam <i>Concept Learning</i> .....	18
2.4.5	Ruang Versi .....	21
2.4.6	Operator Dasar <i>Concept Learning</i> .....	21
<b>BAB III CONCEPT LEARNING PADA ALGORITMA</b>		
<b>DAN ALGORITMA ELIMINASI-CALON</b> .....		23
3.1	<i>Concept Learning</i> Algoritma FIND-S .....	23
3.2	<i>Concept Learning</i> Algoritma Eliminasi-Calon .....	26
3.3	Bias pada Ruang Hipotesis .....	36
<b>BAB IV KESIMPULAN DAN SARAN</b> .....		39
4.1	Kesimpulan .....	39
4.2	Saran .....	40
<b>REFERENSI</b> .....		41

## DAFTAR TABEL

Tabel 2.1 : Tabel Kebenaran Konjungsi .....	8
Tabel 2.2 : Tabel Kebenaran Disjungsi .....	8
Tabel 2.3 : Tabel Kebenaran Ingkaran .....	9
Tabel 2.4 : Data Jenis Kontak Lensa .....	12
Tabel 3.3 : Contoh Data Alternatif .....	37



## DAFTAR GAMBAR

Gambar 2.1 : Keterurutan Hipotesis Umum ke Spesifik .....	17
Gambar 3.1 : Algoritma <i>Find-S</i> .....	23
Gambar 3.2 : Langkah-langkah Iterasi Algoritma <i>Find-S</i> .....	26
Gambar 3.3 : Algoritma Eliminasi-Calon .....	27
Gambar 3.4 : Algoritma <i>List-Then-Eliminate</i> .....	28
Gambar 3.5 : Iterasi Pertama Algoritma Eliminasi-Calon .....	29
Gambar 3.6 : Iterasi Kedua Algoritma Eliminasi-Calon .....	30
Gambar 3.7 : Iterasi keempat Algoritma Eliminasi-Calon .....	31
Gambar 3.8 : Iterasi keenam Algoritma Eliminasi-Calon .....	32
Gambar 3.9 : Hasil Akhir Algoritma Eliminasi-Calon .....	34

# **BAB I**

## **PENDAHULUAN**

Perkembangan teknologi semakin mempermudah manusia dalam menyelesaikan pekerjaan. Salah satu contoh perkembangan teknologi tersebut adalah perkembangan piranti keras dan program komputer. Program komputer senantiasa dikembangkan agar performanya menjadi lebih baik dan dapat mengefisienkan pekerjaan manusia. Lebih jauh lagi komputer juga dapat menjadi sarana hiburan dan sumber informasi melalui internet. Perkembangan pemanfaatan komputer ini dipacu oleh kemampuan komputer dalam melakukan suatu proses pembelajaran (kecerdasan buatan). Skripsi ini akan menjelaskan bagaimana konsep dasar komputer melakukan proses pembelajaran.

### **1.1 Latar Belakang**

Keberhasilan manusia dalam membuat komputer melakukan proses pembelajaran menjadikan komputer semakin bermanfaat di berbagai bidang aplikasi dalam kehidupan sehari-hari. Sebuah komputer menurut [6] dikatakan melakukan proses pembelajaran jika ia mengubah struktur, program, atau datanya berdasarkan input tertentu yang diberikan dan diharapkan memiliki performa yang lebih baik ketika mendapatkan input yang lebih banyak. Contoh aplikasi pembelajaran komputer adalah di bidang astronomi yaitu pengklasifikasian struktur astronomi yang baru dengan menggunakan *decision tree* (Pohon Keputusan).

Manusia membutuhkan komputer yang dapat melakukan proses pembelajaran. Hal ini dikarenakan lingkungan manusia senantiasa berubah dari waktu ke waktu sehingga dibutuhkan komputer yang dapat beradaptasi dengan perubahan tersebut. Contohnya adalah seseorang yang hendak meramal cuaca berdasarkan data pada hari-hari sebelumnya. Peramalan ini dapat dilaksanakan menggunakan program komputer yang algoritmanya melakukan proses pembelajaran. Kemudian dari algoritma dan data tersebut dibentuklah model (*learner*) yang akan mempunyai kemampuan untuk menganalisa data yang baru.

Kemampuan pembelajaran komputer belum sebaik kemampuan manusia. Program komputer senantiasa dikembangkan agar performanya menjadi semakin baik untuk menjawab berbagai tambahan permasalahan yang semakin berkembang. Apabila dibandingkan dengan manusia, komputer memiliki beberapa kelebihan diantaranya kemampuan dalam mengolah data dengan cepat, tidak lelah, akurat, dan dapat menyimpan data relatif lebih banyak.

Proses pembelajaran komputer membutuhkan algoritma tertentu agar dapat menganalisis data dengan baik. Algoritma ini akan senantiasa dikembangkan oleh perancang model agar program komputer dapat beradaptasi dengan data baru. Salah satu contoh algoritma yang digunakan dalam proses pembelajaran komputer adalah Algoritma Find-s dan Algoritma Eliminasi-Calon yang akan dibahas lebih lanjut dalam skripsi ini.

## 1.2 Perumusan Masalah

Pada penulisan ini, penulis merumuskan masalah sebagai berikut :

1. Bagaimana Algoritma Find-S menemukan hipotesis terbaik dalam ruang hipotesis?
2. Bagaimana Algoritma Eliminasi-Calon menemukan hipotesis terbaik dalam ruang hipotesis?
3. Apakah induktif bias pada Algoritma Find-S dan Algoritma Eliminasi-Calon ?

## 1.3 Batasan Masalah

Dalam penulisan skripsi ini, penulis membatasi permasalahan sebagai berikut :

1. Data yang digunakan adalah data yang bersifat kategorik.
2. Klasifikasi yang digunakan adalah klasifikasi biner.
3. Hipotesis yang dibentuk adalah hipotesis konjungtif.
4. Algoritma yang digunakan adalah Algoritma Find-S dan Algoritma Eliminasi Calon.

## 1.4 Tujuan Penulisan

Tujuan yang akan dicapai dalam penulisan skripsi ini antara lain:

1. Menemukan hipotesis yang terbaik menggunakan Algoritma *Find-S*.
2. Menemukan hipotesis yang terbaik menggunakan Algoritma Eliminasi Calon.

3. Mengetahui induktif bias pada Algoritma *Find-S* dan Algoritma Eliminasi-Calon.

### 1.5 Manfaat Penulisan

Manfaat yang dapat diperoleh dari skripsi ini bagi penulis dan mahasiswa matematika pada khususnya, serta bagi pembaca pada umumnya adalah

- 1) Menambah wawasan mengenai *Machine Learning* (pembelajaran komputer) khususnya tentang *concept learning* (pembelajaran konsep model komputer).
- 2) Sebagai referensi untuk mengembangkan *learner* (model) dan algoritma pembelajaran yang baru.
- 3) Sebagai sebuah konsep untuk menyelesaikan permasalahan klasifikasi biner data sederhana, dimana datanya terdiri atas variabel prediktor dan variabel target yang datanya bersifat kategorik.

## BAB II

### LANDASAN TEORI

Pada bab ini, akan dijelaskan tentang *concept learning*, hipotesis, ruang hipotesis, ruang versi dan beberapa teori yang mendukungnya. Kemudian teori-teori dasar yang mendukung dasar *concept learning* tersebut seperti himpunan, logika matematika, fungsi boolean, dan operator dasar *concept learning* secara singkat dijelaskan

#### 2.1 Himpunan

Dalam buku Retno Hendrowati dan Bambang Hariyanto [1] himpunan didefinisikan sebagai sekumpulan objek yang terdefinisi dengan jelas. Unsur anggota dalam himpunan tersebut disebut elemen. Ada dua bentuk penulisan suatu himpunan, yaitu :

1. Bentuk pendaftaran (*Tabular-Form*) yaitu dengan menuliskan semua elemen himpunan tersebut di dalam kurung kurawal. Sebagai contoh :

Himpunan  $A = \{ 2, 3, 5, 7 \}$  dan

Himpunan  $B = \{ \text{Jakarta, Bogor, Bandung} \}$ .

2. Bentuk pencirian (*Set-Builder Form*) yaitu dengan menuliskan sifat atau ketentuan mengenai elemen himpunan tersebut. Sebagai contoh :

Himpunan  $M = \{x | x \text{ adalah bilangan prima} \}$

Suatu himpunan disebut berhingga bila banyak anggotanya (yang berbeda) berhingga. Jika banyak anggotanya tak berhingga maka disebut

himpunan tak berhingga. Himpunan yang tidak mempunyai anggota disebut himpunan hampa (kosong) dinyatakan dengan  $\emptyset$ . Himpunan A dan B dikatakan sama  $A = B$  apabila A dan B mempunyai anggota-anggota yang sama,

**Definisi 2.1 Subset.** *Himpunan A dikatakan himpunan bagian (Subset) dari himpunan B, bila setiap anggota dari A juga merupakan anggota dari himpunan B, ditulis  $A \subset B$  (atau B merupakan himpunan super set dari A).*

Notasi " $\subseteq$ " digunakan juga untuk menyatakan pernyataan "Subset atau Sama Dengan". Jadi  $A \subseteq B$  berarti A subset B atau  $A = B$ .

Keluarga semua subset dari suatu himpunan S disebut himpunan kuasa (*Power Set*) dari S ditulis  $2^S$ . Banyaknya anggota himpunan dari  $2^S$  adalah  $2^n$  di mana n adalah jumlah anggota dari S. Dari keseluruhan anggota tersebut  $\emptyset$  termasuk didalamnya, karena  $\emptyset$  merupakan subset dari himpunan manapun.

### **Contoh 2.1. Himpunan Kuasa.**

Misal Himpunan  $S = \{a, b\}$ , subset-subset dari S adalah  $\emptyset$ ,  $\{a\}$ ,  $\{b\}$ , dan  $\{a, b\}$ . Jadi  $2^S = \{ \emptyset, \{a\}, \{b\}, \{a, b\} \}$ . Banyaknya anggota dari  $2^S$  adalah  $2^2 = 4$ .

## **2.2 Logika Matematika**

Di dalam matematika, hukum-hukum logika menspesifikasikan makna dari pernyataan matematis. Hukum-hukum logika tersebut

membantu manusia untuk membedakan antara argumen yang valid dan tidak valid. Logika juga digunakan untuk membuktikan teorema-teorema di dalam matematika.

Logika pertama kali dikembangkan oleh filsuf Yunani, Aristoteles, sekitar 2300 tahun yang lalu. Saat ini, logika mempunyai aplikasi yang luas di dalam ilmu komputer, misalnya dalam bidang pemrograman, analisis kebenaran algoritma, kecerdasan buatan (*artificial intelligence*), perancangan komputer, dan pembelajaran komputer (*machine learning*) [3].

Pada bab ini, penulis hanya membahas beberapa operasi dalam logika matematika yang mendukung *concept learning* yaitu konjungsi, disjungsi, dan ingkaran (negasi) yang dikutip dari [4].

### 2.2.1 Konjungsi

Operasi konjungsi merupakan operasi biner (operasi yang dikenakan pada dua pernyataan) yang dilambangkan dengan tanda “ $\wedge$ ”. Dengan operasi ini dua pernyataan dihubungkan dengan kata “ dan ”.

**Definisi 2.2** *Jika  $p$  dan  $q$  dua pernyataan, maka  $p \wedge q$  bernilai benar, jika  $p$  dan  $q$  keduanya bernilai benar, sebaliknya  $p \wedge q$  bernilai salah jika salah satu dari  $p$  atau  $q$  bernilai salah atau keduanya salah .*

Tabel 2.1 menampilkan tabel kebenaran yang menentukan nilai kebenaran pada operasi konjungsi.

Tabel 2.1. Tabel kebenaran konjungsi.

P	Q	$p \wedge q$
B	B	B
B	S	S
S	B	S
S	S	S

### 2.2.2 Disjungsi

Operasi disjungsi juga merupakan operasi biner, operasi ini dilambangkan dengan tanda " $\vee$ ". Operasi ini menggabungkan dua pernyataan menjadi satu dengan kata hubung "atau".

**Definisi 2.3** *Jika  $p$  dan  $q$  dua pernyataan maka  $p \vee q$  bernilai benar, jika  $p$  dan  $q$  keduanya bernilai benar atau salah satu dari  $p$  atau  $q$  bernilai benar, sebaliknya  $p \vee q$  bernilai salah jika keduanya bernilai salah.*

Tabel 2.2. Tabel kebenaran Disjungsi

p	q	$p \vee q$
B	B	B
B	S	B
S	B	B
S	S	S

### 2.2.3 Ingkaran

**Definisi 2.4** *Ingkaran atau (negasi) dari  $p$ , dinyatakan dengan notasi  $\sim p$ , adalah proposisi 'tidak  $p$ '.*

Tabel 2.3 menampilkan tabel kebenaran yang menentukan nilai kebenaran pada ingkaran.

Tabel 2.3 Tabel kebenaran Ingkaran

P	$\sim p$
B	S
S	B

### 2.3 Fungsi Boolean

Menurut [3] Fungsi *Boolean* (disebut juga fungsi biner) adalah pemetaan dari  $B^n$  ke  $B$  melalui ekspresi Boolean, dinotasikan sebagai:  $f: B^n \rightarrow B$  dalam hal ini  $B^n$  adalah himpunan yang beranggotakan pasangan terurut ganda- $n$  (*ordered n-tuple*) di dalam daerah asal himpunan  $B$ .

Setiap ekspresi *Boolean* tidak lain merupakan fungsi *Boolean*.

Misalkan sebuah fungsi *Boolean* adalah  $f(x, y, z) = xyz + x'y + y'z$ .

Fungsi  $f$  memetakan nilai-nilai pasangan terurut ganda-3  $(x, y, z)$  ke himpunan  $\{0, 1\}$ . Contohnya,  $(1, 0, 1)$  yang berarti  $x = 1, y = 0,$  dan  $z = 1$

sehingga  $f(1, 0, 1) = 1 \cdot 0 \cdot 1 + 1' \cdot 0 + 0' \cdot 1 = 0 + 0 + 1 = 1$ .

Aljabar *boolean* menggunakan notasi  $\cdot, +, -, \bar{\phantom{x}}$ . Sebagai contoh, fungsi 'dan' dari dua variabel ditulis  $x_1 \cdot x_2$ . Biasanya notasi ' $\cdot$ ' tidak ditulis, maka penulisan fungsi 'dan' menjadi  $x_1 x_2$ . Fungsi  $x_1 x_2$  bernilai 1 jika hanya jika  $x_1$  dan  $x_2$  masing-masing bernilai 1. Jika salah satunya bernilai 0 maka  $x_1 x_2$  bernilai 0. Fungsi 'atau' (inklusif) dari dua variabel dinotasikan  $x_1 + x_2$ . Fungsi  $x_1 + x_2$  bernilai 1 jika salah satu variabelnya bernilai 1 atau kedua variabelnya bernilai 1. Jika kedua variabel masing-masing bernilai 0 maka  $x_1 + x_2$  bernilai 0. Negasi atau komplemen dari

variabel  $x$  dinotasikan  $\bar{x}$ . Jika  $x$  bernilai 0 maka  $\bar{x}$  bernilai 1. Jika  $x$  bernilai 1 maka  $\bar{x}$  bernilai 0.

Berikut ini adalah aturan aljabar *boolean* :

$$1 + 1 = 0; \quad 1 + 0 = 0; \quad 0 + 0 = 0;$$

$$1 \cdot 1 = 1; \quad 1 \cdot 0 = 0; \quad 0 \cdot 0 = 0; \text{ dan } \bar{1} = 0; \quad \bar{0} = 1$$

Setiap peubah di dalam fungsi Boolean, termasuk dalam bentuk komplementnya, disebut *literal*. Contoh: Fungsi  $h(x, y, z) = xyz'$  pada contoh di atas terdiri dari 3 buah *literal*, yaitu  $x, y$ , dan  $z'$ .

Operasi boolean sering dibutuhkan dalam pemrograman komputer.

Operasi boolean dinyatakan dalam ekspresi logika (atau dinamakan juga ekspresi boolean). Operator boolean yang digunakan adalah *AND*, *OR*, *XOR*, dan *NOT*. Ekspresi boolean tersebut hanya menghasilkan salah satu dari dua nilai, *true* atau *false* [2]. Misalkan  $x_1, x_2, x_3$ , dan  $x_4$  adalah peubah boolean dalam Bahasa Pascal, maka ekspresi *boolean* di bawah ini adalah valid:

$$x_1 \text{ and } x_2$$

$$x_1 \text{ or } (\text{not}(x_2 \text{ and } x_3))$$

yang bersesuaian dengan ekspresi logika

$$x_1 \wedge x_2$$

$$x_1 \vee \sim(x_2 \wedge x_3)$$

Operasi lain dalam pemrograman yang bersesuaian dengan operasi logika adalah operasi bit. Komputer merepresentasikan informasi dengan menggunakan bit. Sebuah bit hanya mempunyai dua nilai, yaitu 1 atau 0.

Sebuah bit dapat digunakan untuk merepresentasikan nilai kebenaran, yaitu 1 untuk merepresentasikan *true* (T) dan 0 untuk merepresentasikan *false* (F). Notasi  $\sim$ ,  $\wedge$ ,  $\vee$ , dan  $\oplus$  masing-masing digunakan untuk melambangkan operator *NOT*, *AND*, *OR*, dan *XOR*. Dengan demikian, operasi bit

$$\sim 0, 1 \wedge 0, 0 \vee 0, 1 \oplus 0$$

Secara berurutan bersesuaian dengan operasi logika

$$\sim F, T \wedge F, F \vee F, T \oplus F$$

#### 2.4 *Concept Learning*

*Concept learning* (konsep pembelajaran) menurut [6] adalah suatu metode untuk menentukan fungsi *boolean* dari data yang terdiri atas masukan nilai variabel prediktor dan keluarannya (variabel target). Tujuan dari *concept learning* adalah memperoleh konsep target yaitu sebuah fungsi (model) yang dapat menjelaskan secara ideal data yang sudah ada maupun data yang baru (dalam populasi). Namun, tujuan ini seringkali tidak dapat tercapai karena beberapa alasan, antara lain jumlah data input yang terlalu sedikit dibandingkan data populasi atau tidak adanya pengetahuan apriori tentang konsep target tersebut. Oleh karena itu, perancang model berupaya membentuk sebuah fungsi yang sedekat mungkin dengan konsep targetnya.

Untuk memudahkan memahami *concept learning*, perhatikanlah contoh 2.2 berikut ini.

## Contoh 2.2. Permasalahan Kontak Lensa

Ada sebuah permasalahan sederhana yaitu bagaimana menentukan kriteria seseorang cocok menggunakan kontak lensa mata jenis *hard*. Permasalahan ini dapat dijawab berdasarkan data pada tabel 2.4.

Tabel 2.4 Data Jenis Kontak Lensa

No	Usia	Rabun	Kelainan Mata	Air Mata	Contact Lenses
1	Muda	Jauh	Tidak	Normal	<i>Soft</i>
2	Muda	Jauh	Ya	Normal	<i>Hard</i>
3	Muda	Dekat	Tidak	Normal	<i>Soft</i>
4	Muda	Dekat	Ya	Normal	<i>Hard</i>
5	Pre-presbyopia	Jauh	Tidak	Normal	<i>Soft</i>
6	Pre-presbyopia	Jauh	Ya	Normal	<i>Hard</i>
7	Pre-presbyopia	Dekat	Tidak	Normal	<i>Soft</i>
8	Presbyopia	Jauh	Ya	Normal	<i>Hard</i>
9	Presbyopia	Dekat	Tidak	Normal	<i>Soft</i>

Tabel 2.4 menjelaskan faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses* atau *soft contact lenses*. Variabel pada data ini terdiri dari beberapa variabel prediktor dan sebuah variabel target. Faktor-faktor yang termasuk dalam variabel prediktor adalah *usia*, *jenis rabun mata*, *kelainan mata*, dan *tingkat produksi air mata*. Gabungan dari setiap nilai dari variabel prediktor tersebut (kejadian) akan memberikan nilai pada variabel targetnya yaitu *jenis lensa kontak (Contact Lenses)* yang cocok digunakan.

Nilai-nilai variabel prediktor pada tabel 2.4 adalah sebagai berikut. Faktor-faktor yang berpengaruh pada variabel usia adalah muda, pre-presbyopia, atau presbyopia (kehilangan kemampuan mata untuk melihat

secara fokus karena penurunan keelastisan lensa mata). Variabel rabun menentukan jenis rabun jauh atau rabun dekat. Variabel kelainan mata menentukan seseorang yang mempunyai kelainan mata (*Yes*) atau tidak (*No*). Variabel air mata menjelaskan tingkat produksi air mata berkurang atau normal. Variabel *Contact Lenses* adalah nilai variabel target yang menentukan seseorang cocok menggunakan *Hard Contact Lenses* atau *Soft Contact Lenses*.

#### 2.4.1 Hipotesis

Makna hipotesis yang digunakan dalam *concept learning* berbeda dengan hipotesis dalam ilmu statistik pada umumnya. Hipotesis dalam *concept learning* adalah serangkaian nilai-nilai variabel prediktor yang dihubungkan dengan logika konjungsi dan akan memberikan nilai positif pada outputnya (variabel target). Pada setiap variabel prediktor, ada hipotesis yang memuat simbol “ $\emptyset$ ”. Simbol ini menyatakan bahwa sembarang nilai pada variabel prediktor akan bernilai negatif pada variabel targetnya. Pada hipotesis lainnya, ada kemungkinan terdapat simbol “?”. Simbol ini menyatakan bahwa nilai pada variabel prediktor belum ditentukan dan dapat diisi oleh sembarang nilai menurut domain variabelnya.

Tom M. Mitchell [6], mengungkapkan dalam bukunya yang berjudul *Machine Learning* menyatakan bahwa hipotesis yang menjelaskan data dan juga menjelaskan data yang belum ada pada

populasi disebut hipotesis induktif. Hipotesis ini dapat diperoleh karena satu-satunya informasi mengenai konsep target adalah nilai yang terdapat pada variabel target di dalam data. Nilai pada variabel target akan digunakan sebagai panduan pembentukan hipotesis dari model pembelajaran komputer.

### **Contoh 2.3. Hipotesis Untuk Contoh 2.2**

Hipotesis yang memprediksikan bahwa seseorang yang cocok menggunakan jenis kontak lensa yang soft adalah seseorang yang usianya muda dan tidak memiliki kelainan mata, dapat diekspresikan dengan notasi (Muda,?,Tidak,?). Hipotesis yang paling umum menyatakan bahwa dalam kondisi apa pun seseorang akan cocok menggunakan jenis kontak lensa *hard*, diekspresikan dengan notasi (?,?,?,?). Sedangkan hipotesis dengan ekspresi ( $\emptyset,\emptyset,\emptyset,\emptyset$ ) menyatakan bahwa pada kondisi apapun seseorang tidak akan cocok menggunakan jenis kontak lensa *hard*.

#### **2.4.2 Notasi Pada Concept Learning.**

Pada concept learning beberapa notasi khusus yang digunakan adalah sebagai berikut:

- 1) Himpunan dari seluruh kejadian disebut  $X$ , dimana  $X = \{x_1, x_2, x_3, \dots, x_m\}$  dengan  $m$  banyaknya kejadian.
- 2) Elemen dari  $X$  adalah  $x$  (sebuah kejadian) .

- 3) Nilai yang mungkin pada setiap variabel prediktor disebut  $a_i$
- 4) Sebuah konsep atau fungsi (model/*learner*) yang paling baik menjelaskan contoh data (*training data*) dan data pada populasi disebut konsep target  $c$ . Secara umum,  $c$  adalah fungsi boolean yang didefinisikan atas  $X$ , yaitu  $c : X \rightarrow \{0,1\}$ .
- 5) Kejadian dimana  $c(x)=1$  disebut contoh positif sedangkan  $c(x)=0$  disebut contoh negatif.
- 6) Penulisan  $(x, c(x))$  menggambarkan pasangan dari kejadian  $x$  dengan nilai fungsi targetnya  $c(x)$ .
- 7) Himpunan semua hipotesis yang mungkin adalah  $H$  dan biasa disebut ruang hipotesis, dimana  $H = \{h_1, h_2, h_3, \dots, h_n\}$  dengan  $n$  adalah banyaknya hipotesis yang mungkin.
- 8) Hipotesis  $h$  adalah elemen dari  $H$ . Secara umum, setiap hipotesis  $h$  di  $H$  merepresentasikan fungsi boolean yang didefinisikan atas  $X$ ; yaitu,  $h : X \rightarrow \{0,1\}$ . Tujuan dari perancang model adalah menemukan hipotesis  $h$  yang paling mendekati nilai dari fungsi target  $c$ . Hal ini dinotasikan dengan  $h(x) = c(x)$  untuk semua kejadian  $x$  di  $X$ .

#### **Contoh 2.4. Notasi Penulisan Untuk Contoh 2.2**

- 1) Berdasarkan tabel 2.4 himpunan kejadian  $X$  adalah orang-orang yang memiliki perbedaan kondisi fisik dimana

kondisinya dijelaskan oleh variabel prediktor yaitu usia, jenis rabun mata, kelainan mata, dan tingkat produksi air mata.

- 2) Seseorang yang usianya muda, matanya rabun jauh, tidak memiliki kelainan mata, dan produksi air matanya normal menggunakan *soft contact lenses* disimbolkan dengan  $x$
- 3) Salah satu nilai kemungkinan ( $a_i$ ) pada variabel *usia* adalah *muda*.
- 4) Sebuah fungsi yang dapat menentukan seseorang cocok menggunakan kontak lensa jenis *hard* disimbolkan oleh  $c$ .
- 5) Sebuah kejadian yang menunjukkan bahwa seseorang cocok menggunakan *hard contact lenses* dinotasikan oleh  $c(x) = 1$ , sedangkan  $c(x) = 0$  jika sebuah kejadian yang menunjukkan bahwa seseorang cocok menggunakan *soft contact lenses*.
- 6) Ada seorang pasien yang usianya muda, matanya rabun jauh, tidak memiliki kelainan mata, dan produksi air matanya normal, ia menggunakan *soft contact lenses*. Kejadian ini dapat dinotasikan  $(x, c(x))$ .
- 7)  $H : \{(Muda, ?, ?, Normal), (Muda, ?, ?, ?), (Muda, ?, Ya, ?), \dots\}$ .
- 8) Hipotesis  $h$  adalah  $(Muda, ?, ?, Normal)$ .

### 2.4.3 Keterurutan Hipotesis Umum ke Khusus

**Definisi 2.4 :** Misalkan  $h_j$  dan  $h_k$  adalah fungsi boolean yang didefinisikan pada  $X$ . Maka  $h_j$  lebih umum atau sama dengan  $h_k$  (dinotasikan  $h_j \geq_g h_k$ ) jika hanya jika

$$(\forall x \in X)[(h_k(x) = 1) \rightarrow (h_j(x) = 1)]$$

**Contoh 2.4. Ilustrasi Keterurutan Hipotesis**

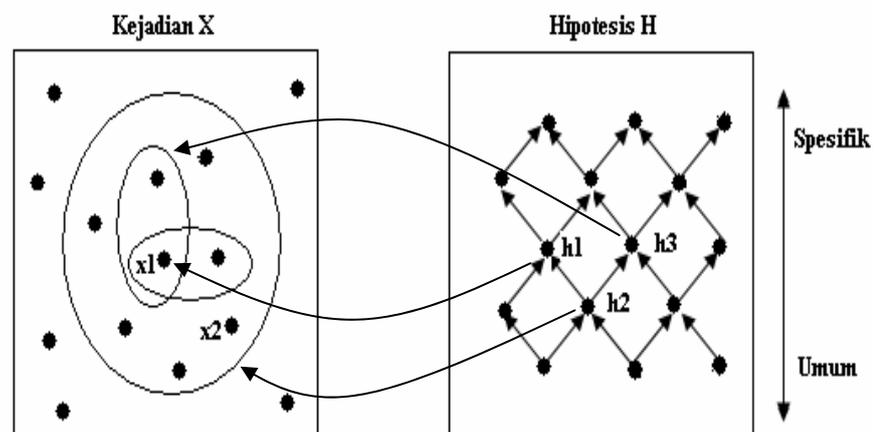
Untuk mengilustrasikan keterurutan hipotesis umum ke khusus, berikut merupakan sebuah contoh sederhana :

$$h_1 = (\text{Muda, ?, ?, Normal}),$$

$$h_2 = (\text{Muda, ?, ?, ?}), \text{ dan}$$

$$h_3 = (\text{Muda, ?, Ya, ?})$$

Karena pada  $h_2$  hanya terdapat satu variabel yang nilainya tetap Muda, maka  $h_2$  lebih banyak mengklasifikasikan kejadian  $x_1, x_2, \dots, x_n$  sebagai hipotesis positif. Kejadian-kejadian yang diklasifikasikan positif oleh  $h_1$  dan  $h_3$  juga akan diklasifikasikan positif oleh  $h_2$ . Tetapi hal ini tidak berlaku sebaliknya. Oleh karena itu  $h_2$  lebih umum daripada  $h_1$  dan  $h_3$ . Untuk lebih jelasnya dapat dilihat pada gambar 2.1.



x1=(Muda,Jauh,Tidak,Normal)	h1=(Muda,?,?,Normal)
x2=(Muda,Jauh,Ya,Normal)	h2=(Muda,?,?,?)
	h3=(Muda,?,Ya,?)

Gambar 2.1. Keterurutan hipotesis umum ke spesifik

#### 2.4.4 Pencarian Hipotesis dalam Concept Learning

*Concept Learning* digunakan dalam mencari hipotesis yang terbaik dalam ruang hipotesis. Tujuan dari pencarian ini adalah menemukan hipotesis yang sedekat mungkin dengan konsep target. Sebelum mencari hipotesis yang terbaik, perancang model harus dapat membuat ruang hipotesis berdasarkan data yang tersedia.

Pada permasalahan kontak lensa di contoh 2.2, untuk mengetahui banyaknya kejadian yang mungkin ada pada ruang kejadian  $X$  adalah dengan menghitung semua kombinasi dari setiap variabel prediktor. Secara umum, dapat digunakan persamaan 2.1 untuk menghitung banyaknya kejadian yang mungkin ada pada ruang kejadian  $X$  yaitu

$$X = V_1 * V_2 * V_3 * \dots * V_n \tag{2.1}$$

dengan :  $X$  = Himpunan kejadian  $x$  yang mungkin.

$V_n$  = banyaknya nilai kemungkinan pada variabel ke- $n$

Perancang model dapat menetapkan sendiri aturan-aturan (syntax) dari pemodelannya. Sebagai contoh, dalam pemodelan tersebut diperkenalkan lagi dua notasi ( ? ) dan (  $\emptyset$  ) pada setiap variabel prediktor. Sehingga variabel *usia* memiliki lima nilai

kemungkinan yaitu muda, pre-presbyopia, presbyopia, '?', atau 'Ø'. Demikian pula penambahan '?' dan 'Ø' pada variabel prediktor lainnya. Sehingga variabel *rabun* menjadi terdiri atas empat kemungkinan yaitu Jauh, Dekat, '?', atau 'Ø'. Variabel *kelainan mata* terdiri atas ya, tidak, '?', atau 'Ø'. Variabel *air mata* terdiri dari tingkat produksi air mata berkurang, normal, '?', atau 'Ø'. Secara umum, dapat digunakan persamaan 2.2 untuk menghitung banyaknya kejadian yang mungkin secara syntax ada pada ruang kejadian X yaitu

$$X = (V_1 + 2) * (V_2 + 2) * \dots * (V_n + 2) \quad (2.2)$$

dengan,

angka 2 menyatakan '?' dan 'Ø'.

Hipotesis yang mengandung satu atau lebih simbol Ø akan diklasifikasikan sebagai kejadian yang negatif. Oleh karena itu, semua hipotesis yang mengandung sebuah atau lebih komponen simbol Ø akan dianggap sama yaitu sebagai hipotesis yang bernilai negatif pada variabel targetnya. Jadi, nilai yang mungkin pada variabel *usia* menjadi memiliki empat nilai kemungkinan yaitu muda, pre-presbyopia, presbyopia, atau, '?'. Variabel *rabun* menjadi terdiri atas tiga kemungkinan yaitu Jauh, Dekat, atau '?'. Variabel *kelainan mata* menjadi terdiri atas ya, tidak, atau '?'. Variabel *air mata* menjadi terdiri dari tingkat produksi air mata berkurang, normal, atau '?'. Secara umum, dapat digunakan

persamaan 2.3 untuk menghitung banyaknya kejadian yang mungkin secara simantik ada pada ruang kejadian  $X$  yaitu

$$X = 1 + [(V_1 + 1) * (V_2 + 1) * \dots * (V_n + 1)] \quad (2.3)$$

dengan :

$(V_n + 1)$  = banyaknya nilai kemungkinan pada variabel ke- $n$  ditambah dengan simbol '?'

1 = menyatakan semua hipotesis yang memuat simbol 'Ø'.

#### **Contoh 2.5. Menghitung Jumlah Kejadian**

Berdasarkan data pada tabel 2.4, dengan menggunakan persamaan 2.1 maka jumlah kejadian yang mungkin adalah  $3 * 2 * 2 * 2 = 24$  kejadian. Sehingga total terdapat 24 kejadian yang berbeda yang mungkin ada dalam  $X$ .

#### **Contoh 2.6. Menghitung jumlah kejadian yang secara syntax**

Berdasarkan data pada tabel 2.4, dengan menggunakan persamaan 2.2 maka jumlah kejadian yang mungkin secara syntax adalah  $5 * 4 * 4 * 4 = 320$  kejadian. Sehingga total terdapat 320 kejadian yang secara syntax berbeda dalam  $X$ .

#### **Contoh 2.7. Menghitung jumlah kejadian yang secara simantik**

Berdasarkan data pada tabel 2.4, dengan menggunakan persamaan 2.3 maka jumlah kejadian yang mungkin secara simantik adalah  $1 + (4*3*3*3) = 109$  kejadian. Sehingga total terdapat 109 kejadian yang secara simantik berbeda dalam  $X$ .

Berdasarkan penjelasan di atas, dapat diambil kesimpulan bahwa *concept learning* digunakan untuk mencari hipotesis yang terbaik diantara seluruh hipotesis yang ada di ruang hipotesis. Untuk mendapatkan hipotesis terbaik, maka dibutuhkan algoritma.

#### 2.4.5 Ruang Versi

Berdasarkan data yang tersedia, ada banyak hipotesis yang dapat dibuat. Hipotesis-hipotesis tersebut membentuk suatu himpunan yang disebut ruang hipotesis. Namun, tidak semua hipotesis konsisten dengan data. Himpunan bagian dari ruang hipotesis yang konsisten dengan data  $D$  disebut ruang versi.

Sebuah hipotesis  $h$  dikatakan konsisten dengan data  $D$  jika hanya jika  $h(x) = c(x)$  untuk setiap  $(x, c(x))$  di  $D$ , dinotasikan sebagai berikut.

$$\text{Konsisten}(h, D) \equiv (\forall(x, c(x) \in D)h(x) = c(x)).$$

#### 2.4.6 Operator Dasar Concept Learning

Dalam *concept learning* terdapat beberapa operator khusus untuk menyatakan keterurutan hipotesis, operator-operator tersebut didefinisikan sebagai berikut:

**Definisi 2.5** : Batas umum hipotesis  $G$  adalah himpunan dari hipotesis-hipotesis di  $H$  yang umum secara maksimal dan konsisten dengan  $D$ .

$$G \equiv \{g \in H \mid \text{Konsisten}(g, D) \wedge ((g \succ_g g') \vee \neg \text{Konsisten}(g', D))\}$$

**Definisi 2.6** : Batas khusus hipotesis  $S$  adalah himpunan dari hipotesis-hipotesis di  $H$  yang umum secara minimal atau maksimal spesifik dari  $H$  dan konsisten dengan  $D$ .

$$S \equiv \{s \in H \mid \text{Konsisten}(s, D) \wedge ((s \succ_g s') \vee \neg \text{Konsisten}(s', D))\}$$

Simbol ' $\succ_g$ ' menyatakan hipotesis yang lebih umum.

Contohnya, penulisan  $(g \succ_g g')$  menyatakan bahwa hipotesis  $g$  lebih umum daripada hipotesis  $g'$ . Sedangkan  $(s \succ_g s')$  menyatakan bahwa hipotesis  $s$  lebih umum daripada hipotesis  $s'$ .

$D_c = \{(x, c(x))\}$  adalah sebuah data dan  $L$  adalah algoritma.  $L(x_i, D_c)$  bermakna bahwa algoritma akan mengklasifikasikan  $x_i$  pada data  $D_c$  menjadi positif atau negatif.

$$(D_c \wedge x_i) \succ L(x_i, D_c)$$

Notasi ini dibaca sebagai berikut :  $L(x_i, D_c)$  secara induktif diinferensi dari  $(D_c \wedge x_i)$ .

Penulisan  $y \succ z$  menunjukkan bahwa  $z$  secara induktif diinferensi dari  $y$ . Sebagai contoh, misalkan  $L$  adalah algoritma Eliminasi-Calon. Maka  $L(x_i, D_c)$  akan mengklasifikasikan  $x_i$  pada  $D_c$  menjadi positif atau negatif.

## BAB III

### CONCEPT LEARNING PADA ALGORITMA *FIND-S* DAN ALGORITMA ELIMINASI-CALON

Untuk menemukan hipotesis terbaik dibutuhkan algoritma. Dalam skripsi ini, algoritma yang digunakan untuk menemukan hipotesis terbaik adalah Algoritma *Find-S* dan Algoritma Eliminasi Calon. Kedua algoritma ini digunakan dikarenakan kesederhanaannya dalam mencari hipotesis terbaik, namun tetap efektif dalam melakukan proses pembelajaran.

Data yang digunakan oleh algoritma dalam skripsi ini menggunakan beberapa variabel prediktor dan sebuah variabel target. Jenis model pembelajaran tersebut disebut sebagai model pembelajaran *supervised*. Sedangkan untuk data yang hanya terdiri atas beberapa variabel prediktor, perancang model menggunakan model pembelajaran *unsupervised*, contohnya adalah analisis *cluster* dalam bidang ilmu statistik.

#### 3.1 Concept Learning Pada Algoritma *FIND-S*

Keterurutan hipotesis Algoritma *Find-S* akan mendapatkan hipotesis yang paling spesifik dan konsisten dengan data. Menurut Tom M. Mitchell [6], Algoritma *Find-S* digambarkan sebagai berikut.

- 
1. Initialize  $h$  to the most specific hypothesis in  $H$
  2. For each positive training instance  $x$ 
    - For each attribute constraint  $a_i$  in  $h$ 
      - If the constraint  $a_i$  is satisfied by  $x$
      - Then do nothing
      - Else replace  $a_i$  in  $h$  by the next more general constraint that is satisfied by  $x$
  3. Output hypothesis  $h$
- 

Gambar 3.1 Algoritma *Find-S*

Berdasarkan data pada tabel 2.4, analisa awal Algoritma *Find-S* digunakan untuk menemukan hipotesis yang terbaik dalam ruang hipotesisnya. Model yang akan ditemukan berupa hipotesis yang akan mengklasifikasikan jenis lensa kontak berdasarkan nilai dari variabel prediktornya. Variabel target dengan nilai lensa kontak *hard* akan dimisalkan sebagai kelas positif dan lensa kontak *soft* sebagai kelas negatif.

Langkah pertama dari algoritma *Find-S* adalah menginisialisasi  $h_0$  sebagai hipotesis kosong:

$$h_0 \leftarrow (\emptyset, \emptyset, \emptyset, \emptyset).$$

Hipotesis  $h_0$  memuat ' $\emptyset$ ' pada setiap variabel prediktornya, maka hipotesis ini memetakan negatif pada outputnya (variabel target). Oleh karena itu, setiap nilai ' $\emptyset$ ' harus diganti dengan nilai yang lebih umum. Kejadian pertama pada tabel 2.4 menentukan seseorang harus menggunakan *soft contact lenses*. Oleh karena itu hipotesis  $h_1$  sama dengan hipotesis  $h_0$ .

$$h_0 = h_1 \leftarrow (\emptyset, \emptyset, \emptyset, \emptyset).$$

Hipotesis  $h_1$  ini juga masih terlalu spesifik karena hipotesis ini menyatakan semua kondisi yang lain pada variabel prediktor akan menjadi negatif. Pada iterasi kedua, kejadian kedua memberikan informasi tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian  $h_1$  diganti dengan  $h_2$  yang lebih umum.

$$h_2 \leftarrow (\text{Muda}, \text{Jauh}, \text{Ya}, \text{Normal})$$

Langkah berikutnya, memperhatikan kejadian yang ketiga. Ternyata kejadian ketiga memberikan informasi tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *soft contact lenses*. Jadi, hipotesis  $h_3$  sama dengan hipotesis  $h_2$ . Kejadian keempat memberikan informasi baru tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian hipotesis  $h_3$  diganti oleh hipotesis  $h_4$  yang memiliki informasi baru pada variabel jauh.

$$h_4 \leftarrow (\text{Muda}, ?, \text{Ya}, \text{Normal})$$

Kejadian kelima menjelaskan faktor-faktor seseorang yang harus menggunakan *soft contact lenses*, maka hipotesis  $h_5$  sama dengan hipotesis  $h_4$ . Kejadian keenam memberikan informasi baru tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian hipotesis  $h_5$  diganti oleh hipotesis  $h_6$  yang memiliki informasi baru pada variabel usia. Sehingga hipotesis  $h_5$  menjadi

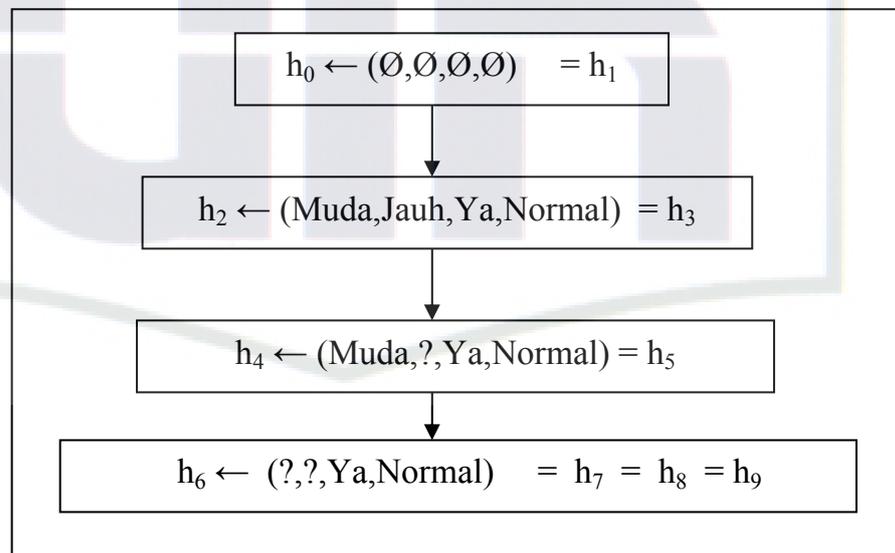
$$h_6 \leftarrow (?, ?, \text{Ya}, \text{Normal})$$

Kejadian ketujuh menjelaskan faktor-faktor seseorang yang harus menggunakan *soft contact lenses*, karena faktor-faktor tersebut tidak memberikan pengaruh maka hipotesis  $h_6$  sama dengan hipotesis  $h_7$ . Kejadian kedelapan memberikan informasi baru faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Tapi hipotesis  $h_7$  sudah dapat menjelaskan kejadian yang kedelapan sehingga hipotesis  $h_8$  sama dengan hipotesis  $h_7$  yaitu  $(?, ?, \text{Ya}, \text{Normal})$ . Kejadian

yang terakhir menjelaskan faktor-faktor yang membuat seseorang cocok dengan *soft contact lenses*. Oleh karena itu, hipotesis  $h_8$  tidak diubah.

$$h_9 \leftarrow (?, ?, Ya, Normal)$$

Hipotesis terakhir yang dihasilkan oleh Algoritma *Find-S* adalah  $(?, ?, Ya, Normal)$ . Hipotesis ini menyatakan bahwa Seseorang akan cocok menggunakan *hard contact lenses* apabila ia memiliki kelainan mata dan tingkat produksi air matanya normal, dimana faktor usia dan kerabunan mata tidak berpengaruh. Pencarian hipotesis dengan algoritma *Find-S* dilakukan dari suatu hipotesis ke hipotesis lainnya, mencari hipotesis yang lebih spesifik dari hipotesis yang lebih umum. Gambar 3.2 mengilustrasikan pencarian ini.



Gambar 3.2. Langkah-langkah Iterasi Algoritma Find-S

### 3.2 *Concept Learning* Pada Algoritma Eliminasi-Calon

Algoritma Eliminasi-Calon menutupi kekurangan pada Algoritma *Find-S*, dimana Algoritma *Find-S* hanya menghasilkan satu hipotesis yang terbaik spesifik dan konsisten dengan contoh data. Sedangkan Algoritma Eliminasi-Calon menghasilkan sebuah deskripsi dari hipotesis-hipotesis yang konsisten dengan data. Jadi, Algoritma Eliminasi-Calon merepresentasikan himpunan dari semua hipotesis yang konsisten dengan contoh data. Menurut Tom M. Mitchell [6], Algoritma Eliminasi Calon digambarkan sebagai berikut.

---

```

Initialize G to the set of maximally general hypotheses in H
Initialize S to the set of maximally specific hypotheses in H
For each training example d, do
• If d is a positive example
  • Remove from G any hypothesis inconsistent with d ,
  • For each hypothesis s in S that is not consistent with d ,-
    • Remove s from S
    • Add to S all minimal generalizations h of s such that
      • h is consistent with d, and some member of G is more general than h
    • Remove from S any hypothesis that is more general than another hypothesis in S
• If d is a negative example
  • Remove from S any hypothesis inconsistent with d
  • For each hypothesis g in G that is not consistent with d
    • Remove g from G
    • Add to G all minimal specializations h of g such that
      • h is consistent with d, and some member of S is more specific than h
    • Remove from G any hypothesis that is less general than another hypothesis in G
  
```

---

Gambar 3.3 Algoritma Eliminasi Calon.

Algoritma Eliminasi-Calon membutuhkan ruang versi yang dapat dibentuk oleh Algoritma *List-Then-Eliminate*. Langkah awal dari algoritma ini yaitu dengan membentuk terlebih dahulu ruang versi yang mengandung semua hipotesis H, lalu hipotesis yang tidak konsisten dengan data dieliminasi. Sisa dari hasil eliminasi ini hipotesis-hipotesis yang konsisten dengan data.

Algoritma *List-Then-Eliminate* hanya dapat diaplikasikan pada ruang hipotesis yang terbatas. Namun demikian, Algoritma ini mempunyai beberapa kelebihan, salah satunya yaitu Algoritma *List-Then-Eliminate* menjamin akan menghasilkan semua hipotesis yang konsisten dengan data. Sedangkan kekurangannya adalah algoritma ini harus mendaftarkan seluruh hipotesis di  $H$ , sehingga apabila datanya besar atau bahkan tak berhingga maka algoritma ini menjadi tidak reliabel. Menurut Tom M. Mitchell [6], Algoritma Eliminasi Calon digambarkan sebagai berikut.

---

---

**The LIST-THEN-ELIMINATE Algorithm**

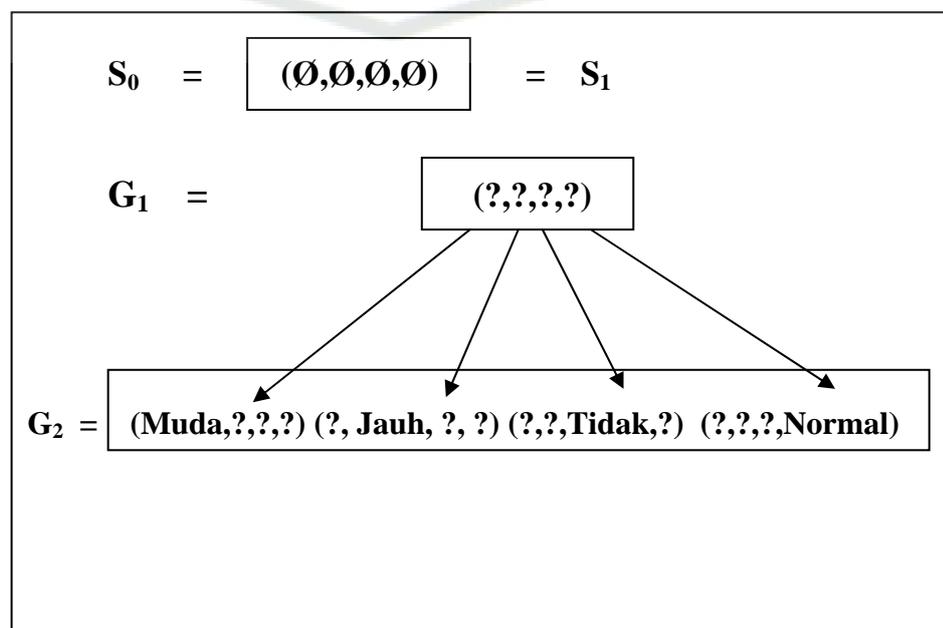
1.  $VersionSpace \leftarrow$  a list containing every hypothesis in  $H$
  2. For each training example,  $(x, c(x))$   
remove from  $VersionSpace$  any hypothesis  $h$  for which  $h(x) \neq c(x)$
  3. Output the list of hypotheses in  $VersionSpace$
- 
- 

Gambar 3.4 Algoritma *List-Then-Eliminate*.

Ada beberapa hal yang harus dipenuhi dalam menggunakan ruang versi dan Algoritma Eliminasi-Calon, yaitu tidak ada error dalam data dan ada beberapa hipotesis di  $H$  yang dapat menggambarkan konsep target secara benar. Apabila terdapat eror pada data, maka terdapat kemungkinan Algoritma Eliminasi-Calon membuang konsep target yang benar karena algoritma ini akan langsung menghilangkan semua hipotesis yang tidak konsisten dengan data sehingga mungkin saja konsep target yang seharusnya benar akan tereliminasi. Agar terhindar dari masalah ini, perancang model dapat menambahkan variasi kejadian baru. Sehingga jumlah hipotesis yang dapat ditemukan menjadi lebih banyak dan peluang terjadinya kesalahan menjadi kecil.

Algoritma Eliminasi-Calon menggunakan ruang versi yang mengandung semua hipotesis dari  $H$  dan konsisten dengan data. Gambar 3.2 merupakan langkah awal dari Algoritma Eliminasi-Calon yang diaplikasikan pada dua kejadian pertama pada contoh data tabel 2.4. Algoritma ini dimulai dengan menginisialisasikan hipotesis  $G_0$  sebagai hipotesis yang paling umum di  $H$ , yaitu  $(?,?,?,?)$ . kemudian hipotesis  $S_0$  sebagai hipotesis kosong, yaitu  $(\emptyset,\emptyset,\emptyset,\emptyset)$ .

Berdasarkan tabel 2.4, kejadian pertama menunjukkan faktor-faktor yang membuat seseorang cocok dengan *soft contact lenses*. Algoritma Eliminasi-Calon memeriksa batasan  $S_0$  karena hipotesis  $S_0$  mewakili kejadian pertama maka  $S_0$  tidak diganti. Gambar 3.5 menunjukkan penggantian  $S_0$  oleh  $S_1$ . Batas Umum  $G_0$  diubah karena hipotesis ini sudah tidak sesuai dengan kejadian pertama, oleh karena itu setiap variabel pada  $G_0$  digantikan oleh nilai-nilai hipotesis yang lebih umum. Sehingga diperoleh hipotesis  $G_1$  yaitu  $(Muda,?,?,?),(?, Jauh, ?, ?),(?,?,Tidak,?)$  dan  $(?,?,?,Normal)$

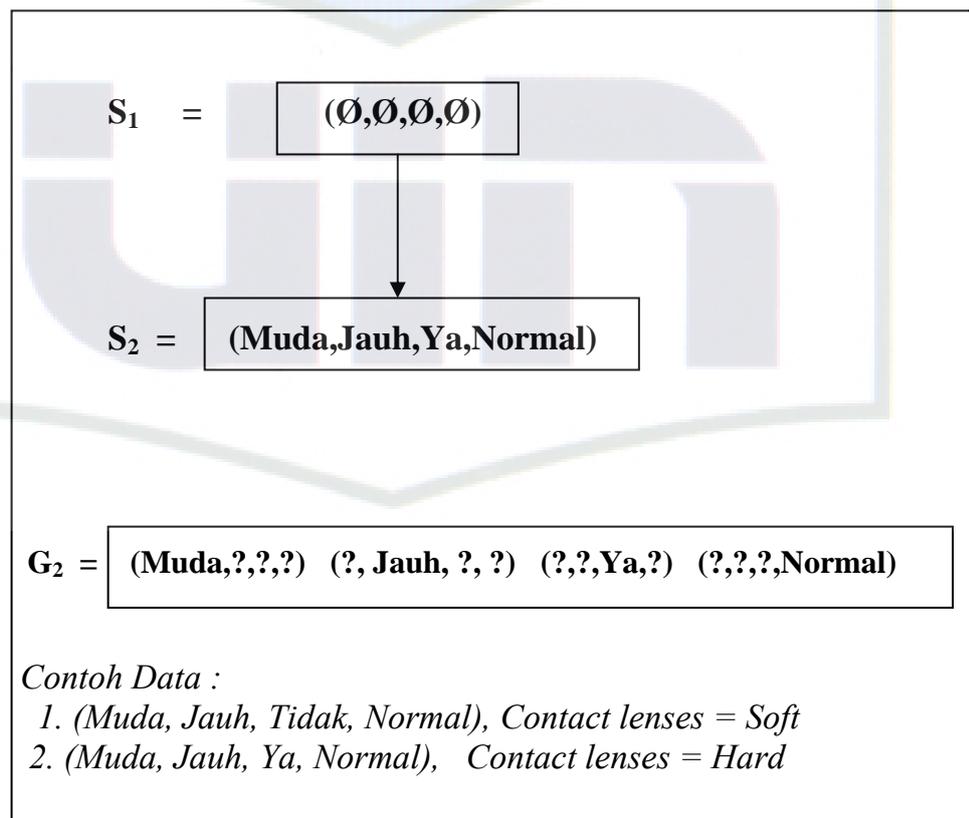


*Contoh Data :*

1. (Muda, Jauh, Tidak Normal), Contact lenses = Soft

Gambar 3.5. Iterasi Pertama Algoritma Eliminasi-Calon

Pada iterasi kedua, Kejadian kedua memberikan informasi tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian Hipotesis  $S_1$  berubah menjadi hipotesis  $S_2$  yaitu (Muda,Jauh,Ya,Normal). Sedangkan pada  $G_1$ , hipotesis  $(?,?,Tidak,?)$  berubah menjadi  $(?,?,Ya,?)$ , maka hipotesis pada  $G_2$  adalah (Muda,?,?,?), (? ,Jauh,?,?,?), (?,?,Ya,?) dan (?,?,?,Normal). Gambar 3.6 menunjukkan perubahan hipotesis  $S_1$  menjadi hipotesis  $S_2$ .

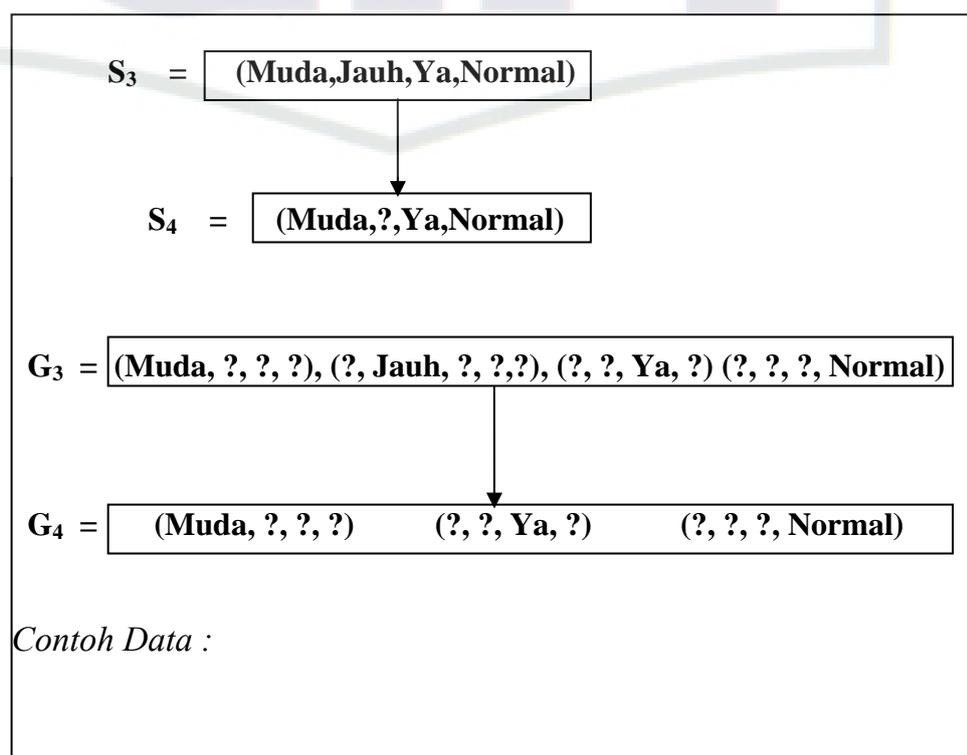


Gambar 3.6. Iterasi Kedua Algoritma Eliminasi-Calon

Pada kejadian ketiga dimana nilai variabel targetnya negatif maka hipotesis  $S_2$  tidak diubah karena kejadian ketiga memberikan informasi

tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *soft contact lenses*. Jadi, hipotesis  $S_3$  sama dengan hipotesis  $S_2$  yaitu (Muda,Jauh,Ya,Normal). Demikian juga hipotesis  $G_3$  sama dengan hipotesis  $G_2$  yaitu (Muda, ?, ?, ?), (?, Jauh, ?, ?,?), (?, ?, Ya, ?) dan (?, ?, ?, Normal).

Kejadian keempat bernilai positif karena memberikan informasi baru tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian hipotesis  $S_3$  diganti oleh hipotesis  $S_4$  yang memiliki informasi baru pada variabel jauh. Sehingga hipotesis  $S_4$  berubah menjadi (Muda,?,Ya,Normal). Hipotesis kedua pada  $G_3$  tidak sesuai dengan  $S_4$ , maka hipotesis (?,Jauh,?,?,?) dihilangkan, sehingga diperoleh hipotesis  $G_4$  yaitu (Muda,?,?,?), (?, ?,Ya,?) dan (?,,?,Normal). Gambar 3.7 mengilustrasikan perubahan ini.

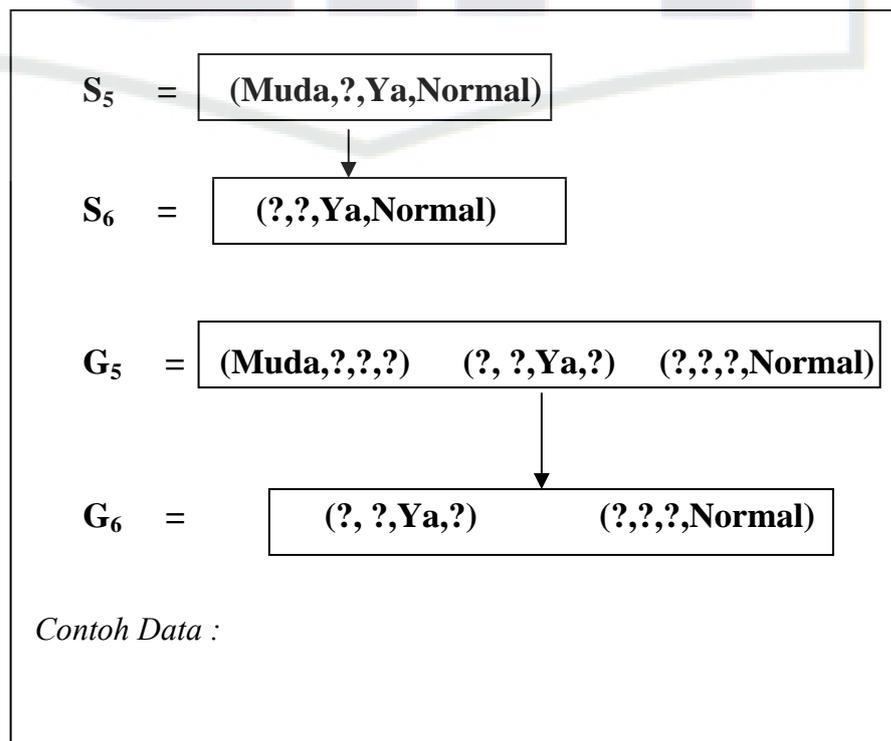


3. (*Muda, Dekat, Tidak, Normal*), *Contact lenses = Soft*
4. (*Muda, Dekat, Ya, Normal*), *Contact lenses = Hard*

Gambar 3.7. Iterasi keempat Algoritma Eliminasi-Calon

Kejadian kelima bernilai negatif karena menjelaskan faktor-faktor seseorang yang harus menggunakan *soft contact lenses*, maka hipotesis  $S_5$  sama dengan hipotesis  $S_4$ . Hipotesis  $G_4$  masih dapat merepresentasikan  $S_4$ , maka  $G_5$  sama dengan  $G_4$ .

Kejadian keenam memberikan informasi baru tentang faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Dengan demikian hipotesis  $S_5$  diganti oleh hipotesis  $S_6$  yang memiliki informasi baru pada variabel usia. Sehingga hipotesis  $S_6$  menjadi  $(?, ?, Ya, Normal)$ . Hipotesis  $(Muda, ?, ?, ?)$  pada  $G_5$  tidak sesuai dengan  $S_6$ , maka hipotesis tersebut dapat dihilangkan. Hipotesis  $G_5$  menjadi hipotesis  $G_6$  yaitu  $(?, ?, Ya, ?)$  dan  $(?, ?, ?, Normal)$ . Gambar 3.8 mengilustrasikan perubahan hipotesis  $S_5$  dan hipotesis  $G_5$ .



5. (*Pre-presbyopia, Jauh, Tidak, Normal*), *Contact lenses = Soft*

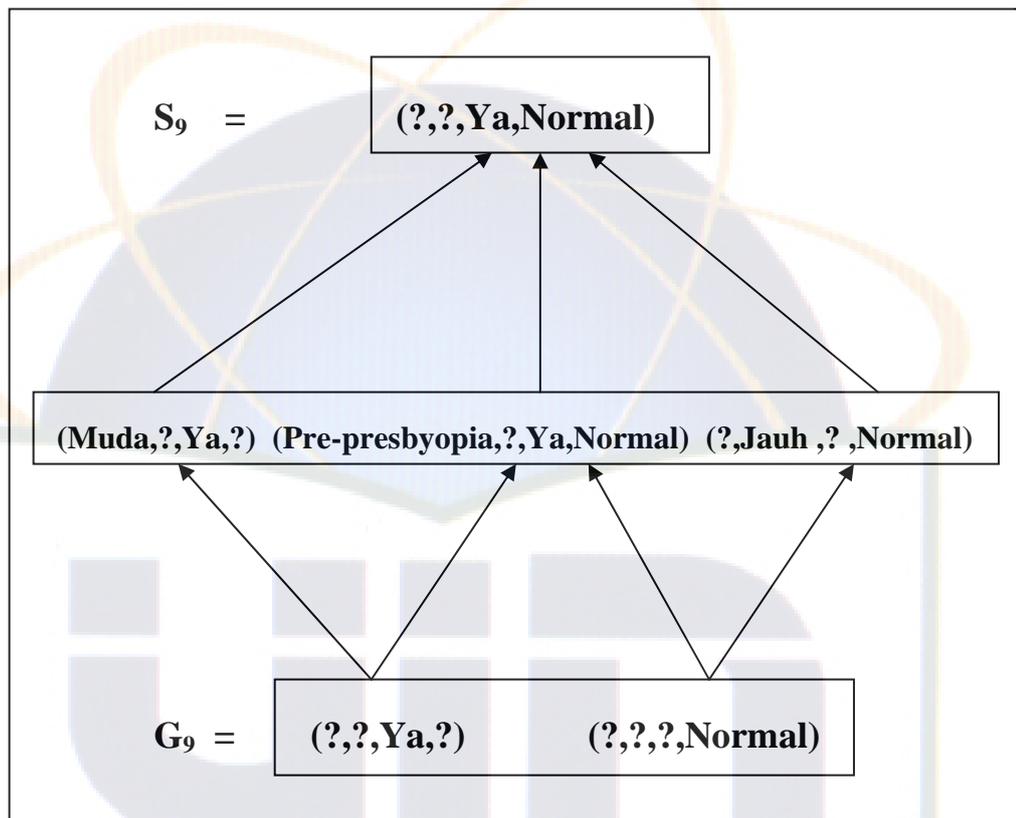
6. (*Pre-presbyopia, Jauh, Ya, Normal*), *Contact lenses = Hard*

Gambar 3.8 Iterasi keenam Algoritma Eliminasi Calon

Kejadian ketujuh menjelaskan faktor-faktor seseorang yang harus menggunakan *soft contact lenses*, karena faktor-faktor tersebut tidak memberikan pengaruh maka hipotesis  $S_7$  sama dengan hipotesis  $S_6$  yaitu ( $?, ?, Ya, Normal$ ). Hipotesis  $G_6$  tidak diubah karena masih dapat merepresentasikan  $S_7$  dan kejadian ketujuh. Jadi, hipotesis  $G_6$  sama dengan hipotesis  $G_7$  yaitu ( $?, ?, Ya, ?$ ) dan ( $?, ?, ?, Normal$ ).

Kejadian kedelapan bernilai positif pada variabel target karena memberikan informasi baru faktor-faktor yang dapat menyebabkan seseorang cocok menggunakan *hard contact lenses*. Tapi hipotesis  $S_7$  masih dapat menjelaskan kejadian yang kedelapan sehingga hipotesis  $S_8$  sama dengan hipotesis  $h_7$  yaitu ( $?, ?, Ya, Normal$ ). Demikian juga Hipotesis  $G_8$  sama dengan hipotesis  $G_7$ .

Kejadian yang terakhir menjelaskan faktor-faktor yang membuat seseorang cocok dengan *soft contact lenses*. Oleh karena itu hipotesis  $S_9$  sama dengan hipotesis  $S_8$ . Hipotesis  $G_8$  tidak berubah karena masih dapat merepresentasikan  $S_9$  dan kejadian terakhir. Dengan demikian, batas hipotesis spesifik  $S_9$  adalah ( $?, ?, Ya, Normal$ ) sedangkan batas hipotesis umum  $G_9$  adalah ( $?, ?, Ya, ?$ ) dan ( $?, ?, ?, Normal$ ). Diantara kedua hipotesis tersebut dapat dibentuk hipotesis yang konsisten dengan datanya. Gambar 3.9 mengilustrasikan hasil akhir dari algoritma ini.



Gambar 3.9 Hasil Akhir Algoritma Eliminasi Calon.

Untuk mengetahui besarnya error pada hipotesis yang dihasilkan oleh algoritma Eliminasi Calon dan Algoritma Find-S, perancang model dapat menguji hipotesis yang dihasilkan oleh masing-masing algoritma. Pengujian hipotesis dilakukan dengan mensubsidikan setiap kejadian pada data ke hipotesis, kemudian jika hasil prediksi hipotesis dari subsidi tersebut sama dengan nilai variabel pada data maka hipotesis mengklasifikasikan kejadian tersebut dengan benar.

### **Contoh 3.1. Uji Hipotesis Hasil dari Algoritma *Find-S***

Hipotesis yang dihasilkan oleh Algoritma *Find-S* adalah  $(?, ?, Ya, Normal)$ . Hipotesis ini akan mengklasifikasikan setiap kejadian yang memprediksikan bahwa seseorang hanya cocok menggunakan kontak lensa jenis hard. Untuk menguji kebenaran tersebut, subsidikan kejadian pertama yaitu  $(Muda, Jauh, Tidak, Normal)$  ke  $(?, ?, Ya, Normal)$ . Ternyata nilai pada variabel kelainan mata dikejadian pertama tidak dapat direpresentasikan oleh hipotesis. Maka hipotesis *Find-S* akan mengklasifikasikan kejadian pertama sebagai seseorang yang cocok dengan kontak lensa yang soft. Pengujian dilanjutkan dengan kejadian kedua yaitu  $(Muda, Jauh, Ya, Normal)$ . Ternyata hipotesis  $(?, ?, Ya, Normal)$  merepresentasikan kejadian ini sebagai kejadian yang positif yaitu memprediksikan seseorang cocok dengan kontak lensa berjenis hard. Demikian pengujian ini terus dilakukan sampai kejadian terakhir.

### **Contoh 3.2. Uji Hipotesis Algoritma Eliminasi Calon.**

Pengujian Hipotesis yang dihasilkan oleh Algoritma Eliminasi Calon dilakukan dengan memeriksa apakah setiap kejadian dapat diklasifikasikan dengan benar oleh hipotesis yang berada diantara batas hipotesis spesifik dan batas hipotesis umum.

Seperti yang telah dibahas sebelumnya, Algoritma Eliminasi-Calon akan menghasilkan hipotesis yang benar apabila diberikan data yang akurat dan terdapat konsep target pada ruang hipotesisnya. Tapi,

bagaimana jika konsep targetnya tidak terdapat pada ruang hipotesis? Dapatkah perancang model menghindari masalah ini dengan menambahkan hipotesis-hipotesis yang mungkin pada ruang hipotesis? Bagaimana ukuran dari ruang hipotesis mempengaruhi kemampuan algoritma untuk mengeneralisasikan kejadian yang belum diamati? Bagaimana ukuran dari ruang hipotesis mempengaruhi sejumlah data yang harus diamati? Ini adalah beberapa pertanyaan seputar induktif bias secara umum yang akan dibahas lebih lanjut pada subbab berikutnya

### **3.3 Bias pada Ruang Hipotesis**

Apabila pada ruang hipotesis tidak memuat konsep targetnya, maka salah satu solusinya adalah memperluas ruang hipotesis dengan menambahkan hipotesis-hipotesis yang mungkin. Dengan demikian diharapkan tidak terjadi penyimpangan (bias) pada ruang hipotesis.

#### **Contoh 3.3. Bias Pada Ruang Hipotesis.**

Pada tabel 3.3, Algoritma *Find-S* tidak akan mendapatkan hipotesis yang terbaik pada ruang versinya dan hipotesisnya tidak konsisten dengan data. Hipotesis yang konsisten dengan dua kejadian pertama adalah  $S_2$  (?, Jauh, Ya, Normal). Hipotesis ini tidak benar untuk kejadian yang ketiga karena seharusnya dengan menggunakan hipotesis  $S_2$  variabel target pada data ketiga bernilai positif, tapi ternyata pada data ketiga variabel targetnya bernilai negatif. Hal ini menunjukkan bahwa modelnya bias karena terbatas pada hipotesis yang hanya memuat logika penghubung

konjungtif. Oleh karena itu, perancang model harus membuat ruang hipotesis yang lebih variatif.

Tabel 3.3 Contoh Data Alternative

No	Usia	Rabun	Kelainan Mata	Air Mata	Contact Lenses
1	Muda	Jauh	Ya	Normal	Hard
2	Presbyopia	Jauh	Ya	Normal	Hard
3	Pre-presbyopia	Jauh	Ya	Normal	Soft

Sebagai ilustrasi, pada permasalahan *contact lenses*, ruang hipotesisnya hanya terdiri atas hipotesis yang hanya menggabungkan setiap variabel prediktor dengan logika konjungsi. Karena keterbatasan ini, ruang hipotesisnya tidak bisa menampilkan hipotesis yang disjungtif seperti "Usia = Muda atau Usia = Presbyopia"

Solusi untuk menjamin bahwa konsep targetnya ada di ruang hipotesis  $H$  adalah membuat ruang hipotesis yang mampu menampilkan setiap himpunan bagian yang mungkin dari kejadian  $X$ . Secara umum, himpunan dari semua himpunan bagian dari himpunan  $X$  disebut himpunan kuasa dari  $X$ .

### Contoh 3.3. Himpunan Kuasa.

Pada permasalahan *Contact Lenses* terdapat 24 kejadian yang berbeda. Seberapa banyakkah hipotesis yang bisa didefinisikan oleh kejadian ini? Berapa banyakkah himpunan kuasanya? Secara umum, himpunan kuasanya dapat dihitung dengan rumus:  $2^n$  dimana  $n$  adalah banyaknya elemen dalam himpunan tersebut. Hal ini berarti ada  $2^{24}$

hipotesis berbeda yang dapat ditampilkan di ruang hipotesis dan dapat dipelajari oleh perancang model. Sedangkan pada ruang hipotesis yang konjungtif hanya mampu menampilkan 103 hipotesis. Perbandingan antara  $2^{24}$  dan 103 sangat besar. Jadi pada ruang hipotesis yang konjungtif hanya terdapat sedikit sekali hipotesis dari jumlah seluruh hipotesis yang bisa dirancang. Hal ini yang menyebabkan terjadinya bias pada ruang hipotesis sedemikian hingga kemungkinan besar konsep targetnya tidak terdapat pada ruang hipotesis tersebut. Jika konsep target tidak terdapat pada ruang hipotesis maka algoritma Find-S maupun Algoritma Eliminasi Calon tidak akan pernah mendapatkan hipotesis yang terbaik yang mendekati konsep targetnya.

## BAB IV

### KESIMPULAN DAN SARAN

#### 4.1 Kesimpulan

Sebuah studi tentang konsep pembelajaran komputer dilakukan dengan memperhatikan ruang hipotesis, induktif bias, dan sistem pencarian hipotesis dalam ruang hipotesis. Dua buah algoritma dasar dalam sistem pembelajaran komputer diaplikasikan dalam sebuah data permasalahan sederhana yang diambil dari machine learning database yang dapat diakses secara umum lewat media internet.

Algoritma Find-S hanya menghasilkan satu hipotesis yang diharapkan dapat mendekati konsep target sehingga hipotesis tersebut mampu digeneralisasikan ke populasi. Berikut ini adalah langkah-langkah Algoritma Find-S:

- 
1. Initialize  $h$  to the most specific hypothesis in  $H$
  2. For each positive training instance  $x$ 
    - For each attribute constraint  $a_i$  in  $h$ 
      - If the constraint  $a_i$  is satisfied by  $x$ 
        - Then do nothing
        - Else replace  $a_i$  in  $h$  by the next more general constraint that is satisfied by  $x$
  3. Output hypothesis  $h$
- 

Algoritma kedua yang dapat digunakan dalam concept learning adalah Algoritma Eliminasi Calon. Algoritma ini menghasilkan deskripsi dari hipotesis yang konsisten dengan data. Deskripsi dari hipotesis ini dibatasi oleh hipotesis yang spesifik dan batas hipotesis umum. Berikut ini adalah gambar dari langkah-langkah Algoritma Eliminasi Calon:

---

Initialize  $G$  to the set of maximally general hypotheses in  $H$   
Initialize  $S$  to the set of maximally specific hypotheses in  $H$   
For each training example  $d$ , do

- If  $d$  is a positive example
  - Remove from  $G$  any hypothesis inconsistent with  $d$ ,
  - For each hypothesis  $s$  in  $S$  that is not consistent with  $d$ , -
    - Remove  $s$  from  $S$
    - Add to  $S$  all minimal generalizations  $h$  of  $s$  such that
      - $h$  is consistent with  $d$ , and some member of  $G$  is more general than  $h$
      - Remove from  $S$  any hypothesis that is more general than another hypothesis in  $S$
- If  $d$  is a negative example
  - Remove from  $S$  any hypothesis inconsistent with  $d$
  - For each hypothesis  $g$  in  $G$  that is not consistent with  $d$ 
    - Remove  $g$  from  $G$
    - Add to  $G$  all minimal specializations  $h$  of  $g$  such that
      - $h$  is consistent with  $d$ , and some member of  $S$  is more specific than  $h$
    - Remove from  $G$  any hypothesis that is less general than another hypothesis in  $G$

---

Induktif bias pada Algoritma Eliminasi-Calon adalah apabila konsep target terdapat di ruang hipotesis. Sehingga algoritma ini mampu menghasilkan hipotesis yang dapat digeneralisasikan ke populasi.

#### 4.2. Saran

Skripsi ini menitikberatkan pada konsep dasar *concept learning* secara umum, sehingga kompleksitas hipotesis yang digunakan bukanlah titik berat dari penelitian ini. Pengembangan dapat dilakukan dengan menerapkan ruang hipotesis yang lebih besar dan hipotesis yang lebih kompleks sehingga akan diperoleh gambaran lengkap tentang *concept learning* yang lebih menyeluruh.

## REFERENSI

- [1] Hendrowati, Retno & Bambang Hariyanto, *Buku Teks Ilmu Komputer, Logika Matematika*, Penerbit Informatika, Bandung, 2000.
- [2] <http://mail.informatika.org/~rinaldi/Matdis/2005-2006/Aljabar%20Boolean>, [23/12/2007 09.30 WIB].
- [3] [http://informatika.org/~rinaldi/Buku/Matematika%20Diskrit/Bab01%20Logika\\_edisi%203.pdf](http://informatika.org/~rinaldi/Buku/Matematika%20Diskrit/Bab01%20Logika_edisi%203.pdf), [23/12/2007 10.05 WIB].
- [4] <http://www.parjono.files.wordpress.com/2007/09/rumus-matematika-logika-matematika.doc>, [13/11/2007 09.00 WIB].
- [5] Irvine D.W., & Murphy P.M., *Machine Learning Database*, <http://www.ics.uci.edu/~mlearn/MLRepository.html>, University of California, 2004.
- [6] Mitchell, Tom. M, *Machine Learning*, McGraw. Hill, 1997.
- [7] Nilsson, Nils J., *Intoduction Machine Learning*, Department of Computer Science Stanford University Stanford. CA.94305, 1997.
- [8] Yahya, Yusuf. dkk, *Matematika Dasar*, Dhalia Indonesia, Jakarta, 1990.

