

# Machine Learning

## Concept Learning

**Ali Ridho Barakbah**

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Politeknik Elektronika Negeri Surabaya



**Politeknik Elektronika Negeri Surabaya**  
**Departemen Teknik Informatika dan Komputer**

# Konten

- Ide Mesin Pembelajaran
- Learning Proses
- Find-S
- Candidate Elimination

# Tujuan Instruksi Umum

Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode mesin pembelajaran yang tepat berdasarkan supervised, unsupervised dan reinforcement learning, baik secara individu maupun berkelompok/kerjasama tim.

# Tujuan Instruksi Khusus

- Memahami konsep machine learning dan metode ML yang sederhana
- Mampu menerapkan Find-S, Candidate Elimination

# Ide Mesin Pembelajaran

Fakta harian dalam 6 hari dan keputusan untuk berolah-raga sebagai berikut:

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Tinggi	Pelan	Tidak
4	Cerah	Normal	Kencang	Ya
5	Hujan	Tinggi	Kencang	Tidak
6	Cerah	Normal	Pelan	Ya

(1) *Ketika cuaca cerah, apakah akan berolah-raga?*

(2) *Ketika cuaca cerah dan temperatur normal, apakah akan berolah-raga?*

Penyajian keputusan berdasarkan fakta inilah yang mengilhami konsep dari mesin pembelajaran

# Data Training

Key ↓	Attribut ←-----→			Target ↓
Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D3	Hujan	Tinggi	Pelan	Tidak
D4	Cerah	Normal	Kencang	Ya
D5	Hujan	Tinggi	Kencang	Tidak
D6	Cerah	Normal	Pelan	Ya

- Attribut adalah kolom data, ada atribut dan target
- Instance adalah isi dari atribut sebagai contoh atribut cuaca mempunyai instance “cerah” dan “hujan”, sering ditulis dengan cuaca={cerah,hujan}
- Record/tuple adalah baris data

# Ide Mesin Pembelajaran

Pada dasarnya semua algoritma yang dikembangkan dalam mesin pembelajaran adalah algoritma yang menghasilkan hipotesa dari suatu keputusan berdasarkan data pembelajaran yang diberikan.



# Fact

Data	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

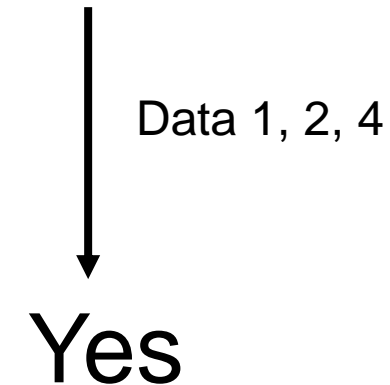


# Problem description

<?, Cold, High, ?, ?, ?>

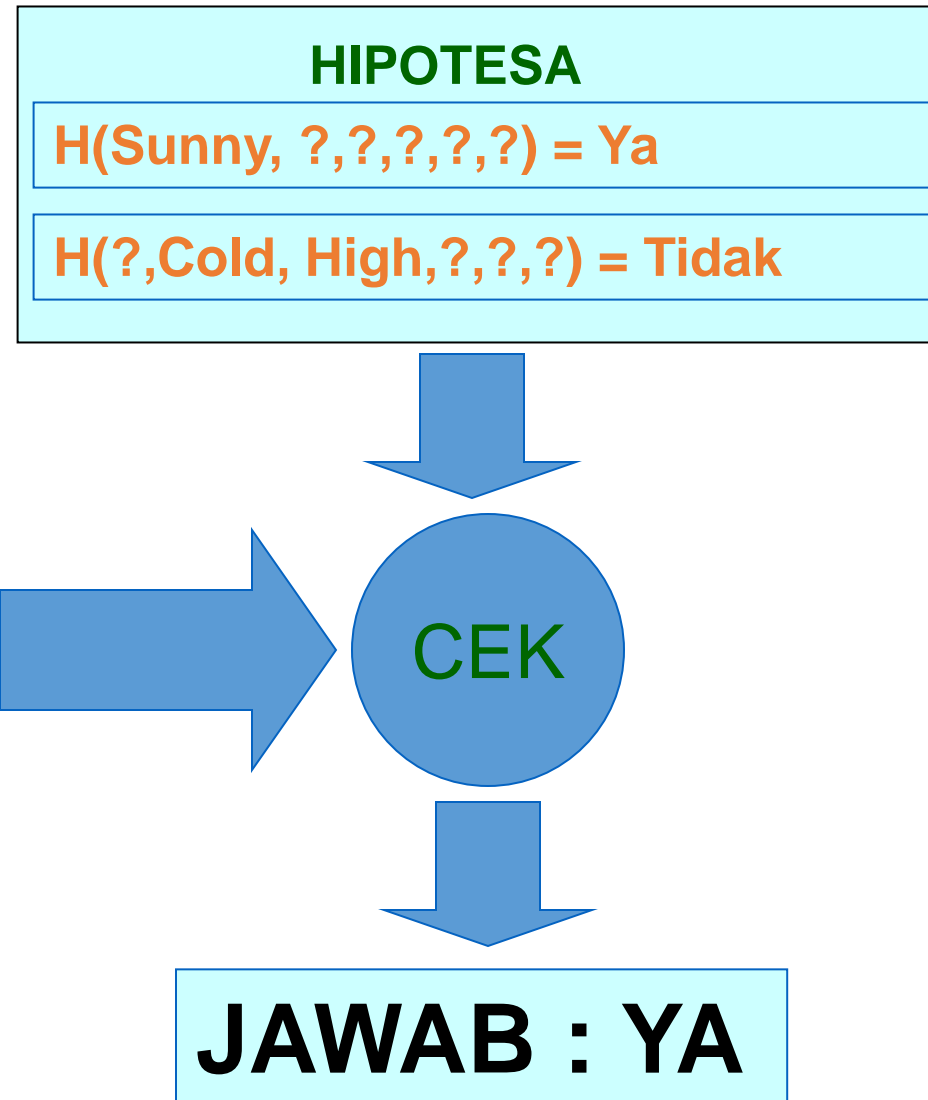


<Sunny, Warm, ?, ?, ?, ?>

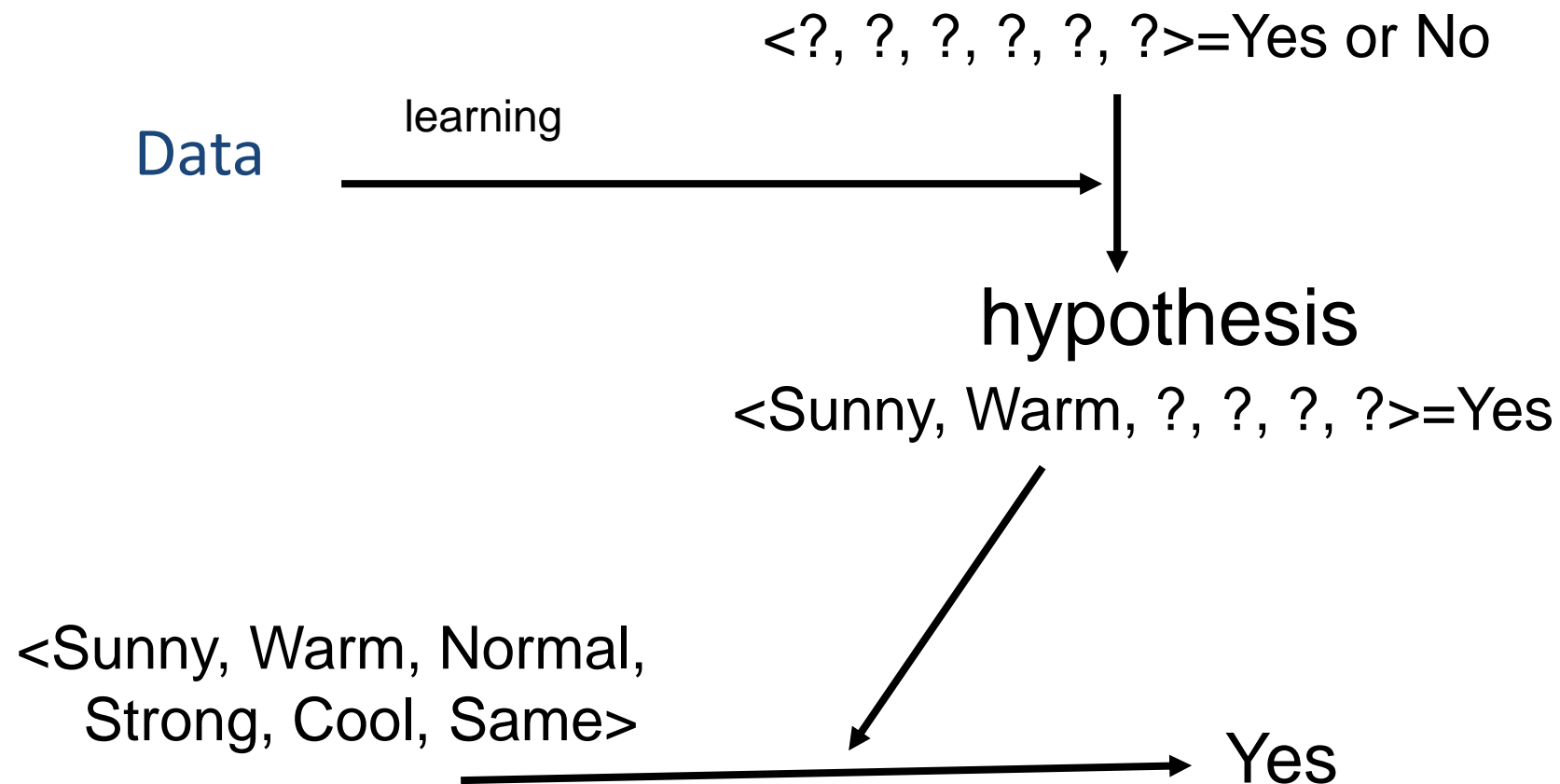


**Our human brain can answer these questions.  
But how the machine can answer?**

# Contoh Keputusan Dari Hipotesa



# Learning Process



# Find-S

- Find-S adalah suatu metode paling sederhana yang dapat digunakan untuk mendapatkan suatu hipotesa berdasarkan data.
- Find-S mencari kesamaan nilai atribut untuk memperoleh suatu hipotesa
- Kelemahan dari Find-S adalah data yang digunakan harus bersifat konsisten dan tidak bias ??? (Terlalu sulit untuk dapat memperoleh data semacam ini pada persoalan nyata)



# Find-S

<  $\phi$ ,  $\phi$ ,  $\phi$ ,  $\phi$ ,  $\phi$ ,  $\phi$  >

<Sunny, Warm, Normal, Strong, Warm, Same>  $\longrightarrow$  <Sunny, Warm, Normal, Strong, Warm, Same>

<Sunny, Warm, High, Strong, Warm, Same>  $\longrightarrow$  <Sunny, Warm, ? , Strong, Warm, Same>

<Sunny, Warm, High, Strong, Cool, Change>  $\longrightarrow$  <Sunny, Warm, ? , Strong, ? , ? >



- Advantage
  - Very simple
- Disadvantage
  - Ignores the negative data

# Candidate-Elimination

$S_0$   $\langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$

?

$G_0$   $\langle ?, ?, ?, ?, ?, ? \rangle$

# Candidate-Elimination

$S_0$   $\langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$



$S_1$   $\langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle$

?

$\langle \text{Sunny, Warm, Normal, Strong, Warm, Same} \rangle$   
=Yes

$G_1$   $\langle ?, ?, ?, ?, ?, ? \rangle$



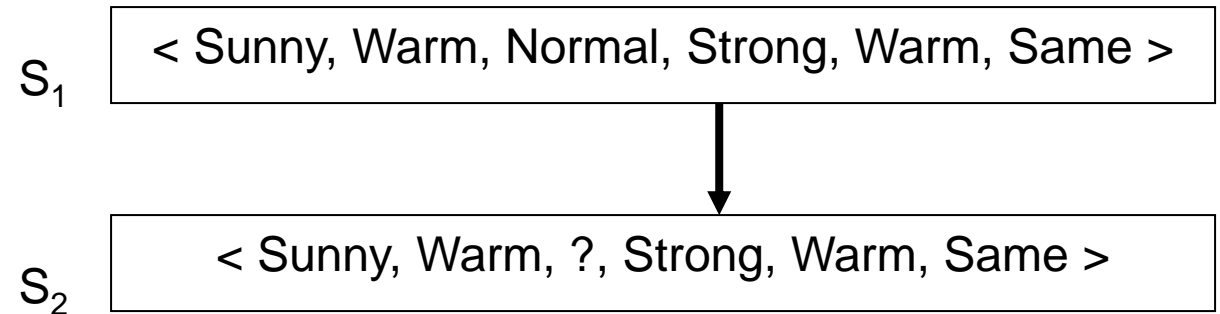
$G_0$   $\langle ?, ?, ?, ?, ?, ? \rangle$



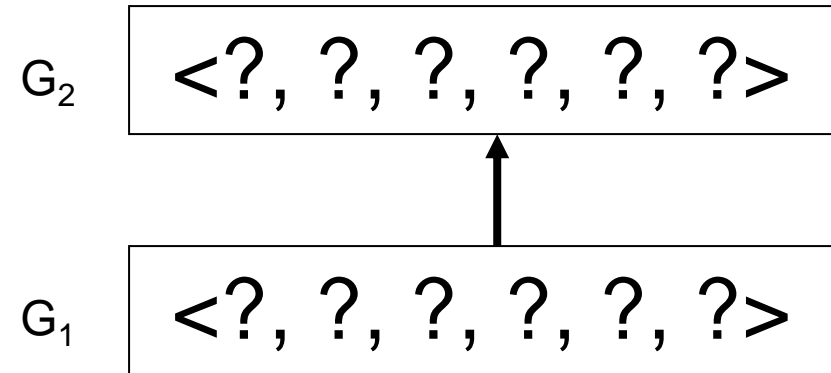


# Candidate-Elimination

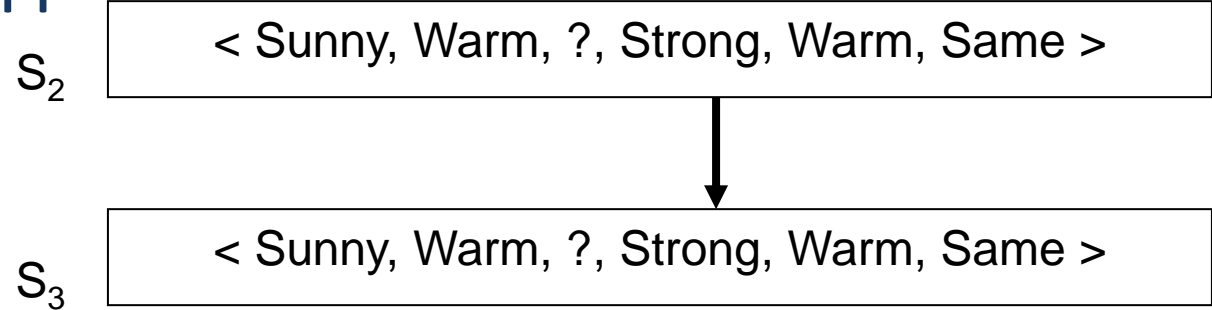
<Sunny, Warm, High,  
Strong, Warm, Same>  
=Yes



?

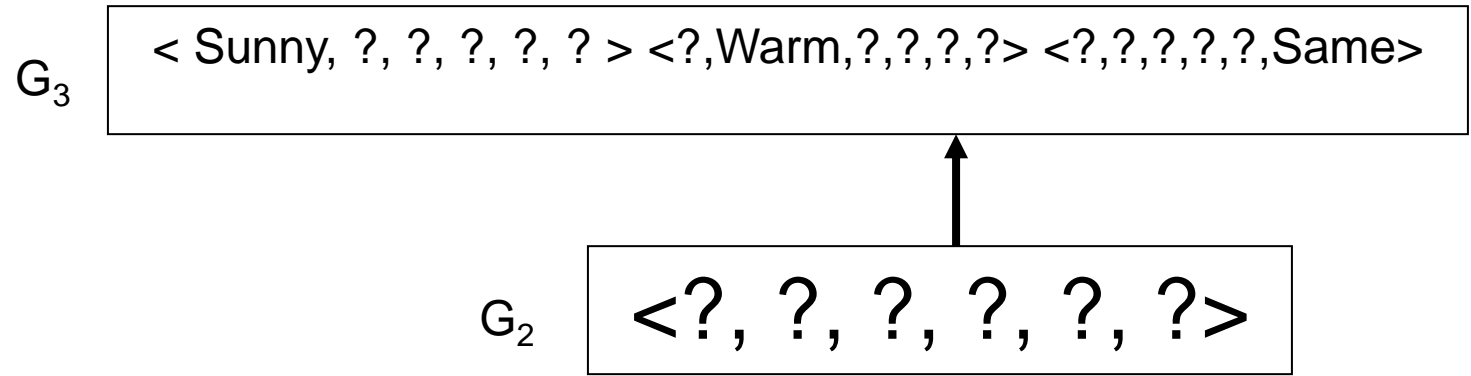


# Candidate-Elimination

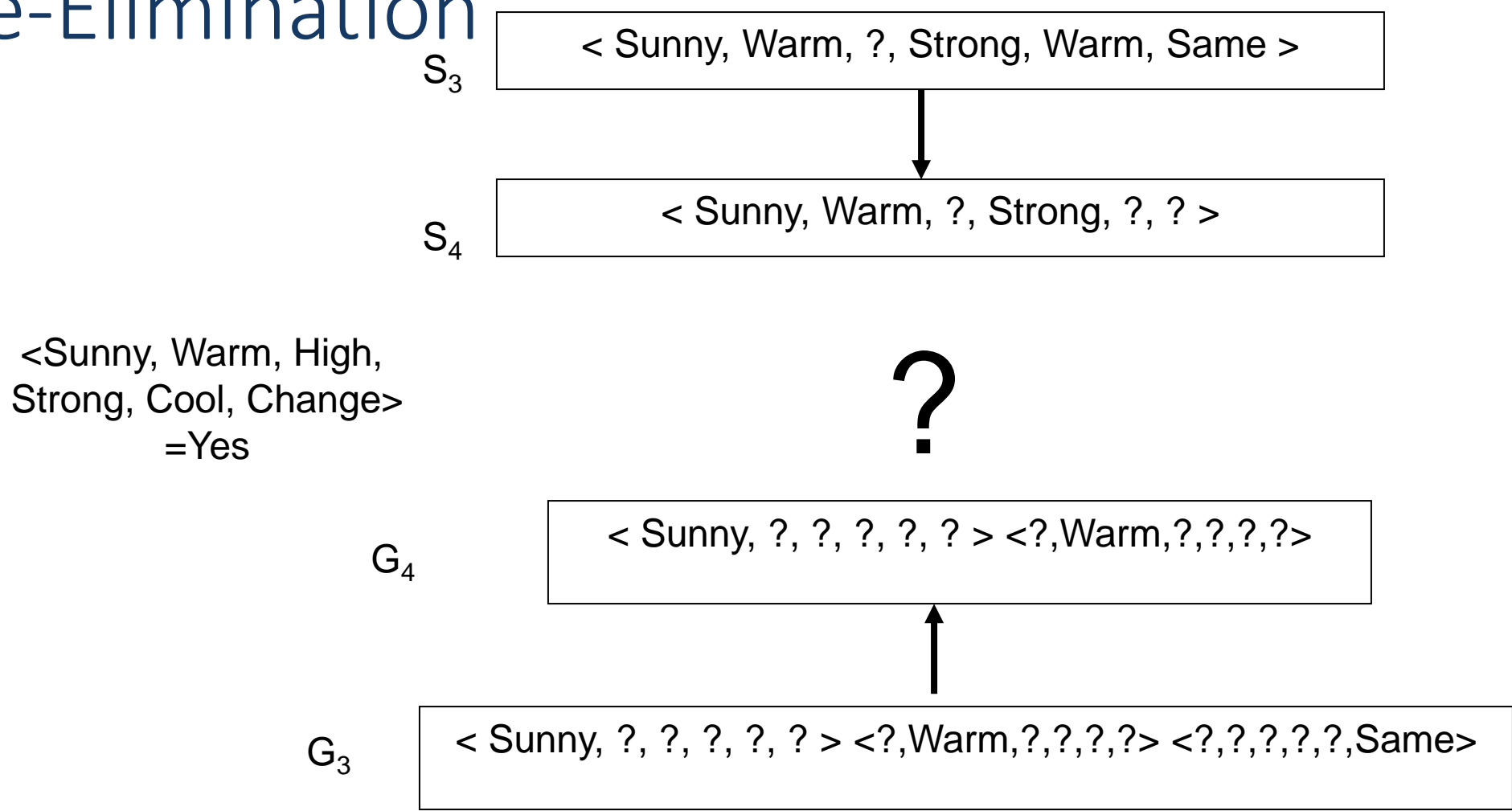


<Rainy, Cold, High,  
Strong, Warm, Change>  
=No

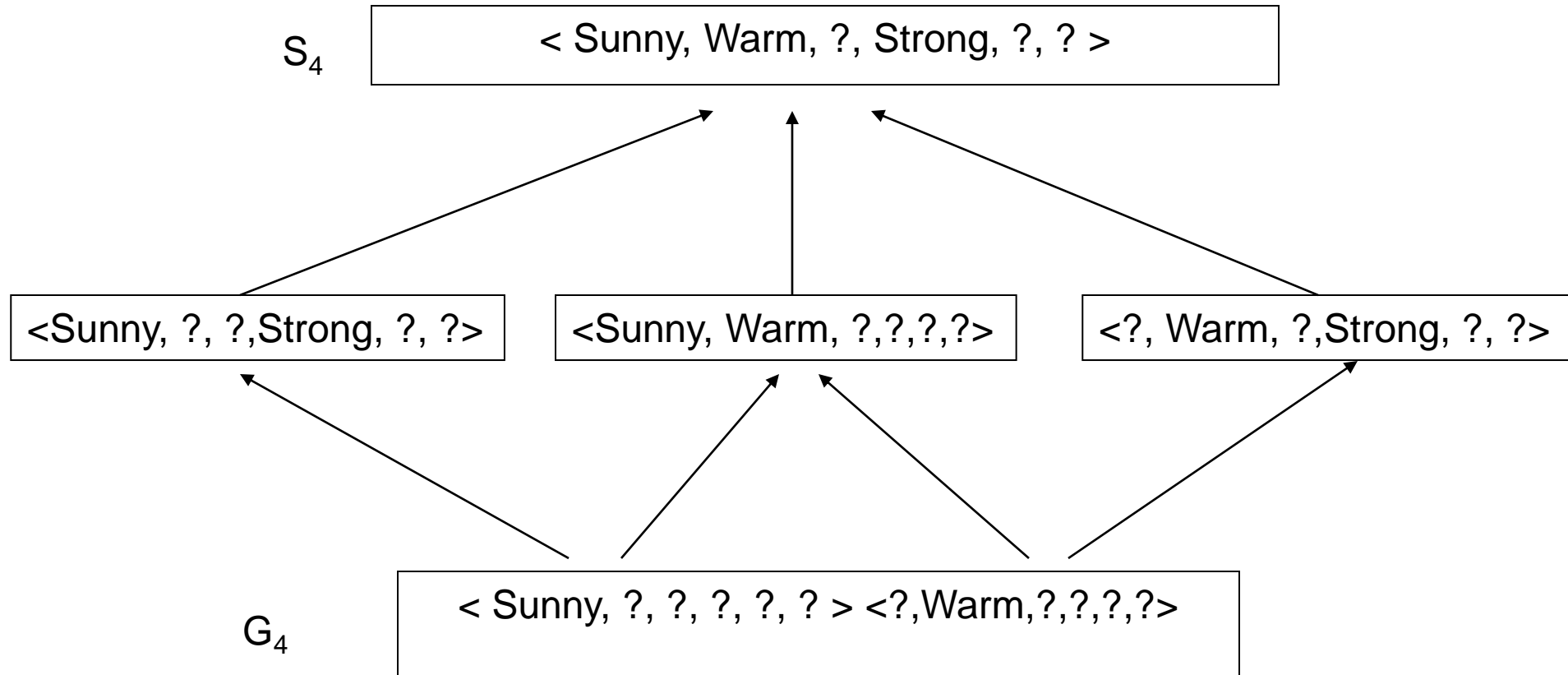
?



# Candidate-Elimination



# Candidate-Elimination



- Advantage
  - Consider the negative data to strengthen the hypothesis
- Disadvantage
  - If the data is not consistent, S and G can not match
  - Difficult to implement in the programming

## Data Konsisten

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
3	Hujan	Tinggi	Pelan	Tidak
5	Hujan	Tinggi	Kencang	Tidak

*Attribut cuaca dan temperatur mempunyai nilai yang sama dalam satu keputusan (berolah-raga), maka data ini adalah data yang konsisten.*

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Tinggi	Pelan	Ya
4	Hujan	Normal	Kencang	Ya
6	Cerah	Normal	Pelan	Ya

*Tidak satupun attribut yang mempunyai nilai yang sama dalam satu keputusan (berolah-raga), maka data ini adalah data yang tidak konsisten.*

# Data Bias

#	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
1	Cerah	Normal	Pelan	Ya
2	Cerah	Normal	Pelan	Ya
3	Hujan	Normal	Pelan	Ya
4	Cerah	Normal	Pelan	Tidak

*Perhatikan data ke 4, data ini mempunyai keputusan yang berbeda dengan data 1 dan data 2, tetapi instance pada semua atributnya sama, sehingga data ini disebut dengan data bias.*

# Latihan Soal

1. Implementasikan Find-S dalam program! Buatlah dengan data training dan data tes yang dinamis!
2. Implementasikan Candidate Elimination dalam program!



# Referensi

- Modul Ajar Machine Learning, Entin Martiana, Ali Ridho Barakbah, Nur Rosyid Mubtadaí, Politeknik Elektronika Negeri Surabaya, 2013.
- Machine Learning, Tom Mitchell, McGraw-Hill. 2008.



# bridge to the future

<http://www.eepis-its.edu>