

Mesin Pembelajaran



Achmad Basuki

PENS – ITS 2006



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 attribut sebagai contoh atribut cuaca mempunyai instance “cerah” dan “hujan”, sering ditulis dengan cuaca={cerah,hujan}
- Record/tuple adalah baris data



Pengisian Data Dalam Matlab

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

`dtTraining(1,:) = {'D1', 'cerah', 'normal', 'pelan', 'ya'}`

`dtTraining(2,:) = {'D2', 'cerah', 'normal', 'pelan', 'ya'}`

`dtTraining(3,:) = {'D3', 'hujan', 'tinggi', 'pelan', 'tidak'}`

`dtTraining(4,:) = {'D4', 'cerah', 'normal', 'kencang', 'ya'}`

`dtTraining(5,:) = {'D5', 'hujan', 'tinggi', 'kencang', 'tidak'}`

`dtTraining(6,:) = {'D6', 'cerah', 'normal', 'pelan', 'ya'}`



Menyimpan dan Memanggil Data

- Untuk menyimpan data training yang sudah dibuat:

`save dataTraining dtTraining`

↓
nama file

↓
nama variabel yang disimpan

- Untuk memanggil data training yang sudah disimpan:

`load dataTraining`

↓
nama file



Mengambil Ukuran data

Setiap data di dalam MATLAB dinyatakan sebagai matrik, sehingga dapat diambil ukuran data yang berupa jumlah baris data dan jumlah kolom (attribut).

```
nData=size(dtTraining,1);  
nAttribut=size(dtTraining,2);
```

nData adalah jumlah data (baris)
nAttribut adalah jumlah attribut (kolom)



Membaca Instance Keputusan

```
% Instance pertama diambil dari data pertama
nInstance=1;
instance{1}=dtTraining{1,nAtribut};

% Membaca semua baris data
for i=2:nData

    % Cek apa instance data ke I adalah instance baru
    sw=0;
    for k=1:nInstance
        if strcmp(dtTraining{i,nAtribut},instance{k})==1
            sw=1;
        end
    end

    % Jika instance baru, maka tambahkan data instance
    if (sw==0)
        nInstance=nInstance+1;
        instance(nInstance)=dtTraining(i,nAtribut);
    end
end
end
```



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.



Hipotesa

Hipotesa dituliskan dengan:

$$H(\text{attribut}_1, \text{attribut}_2, \dots, \text{attribut}_n) = \text{keputusan}$$

$$\underline{H(\text{cuaca}=\text{cerah}, \text{temperatur}=\text{normal}, \text{kec.angin}=\text{pelan}) = \text{ya}}$$

Hipotesa ini menunjukkan bahwa keputusan untuk berolahraga bila cuaca=cerah, temperatur=normal dan kec.angin=pelan, untuk singkatnya dituliskan hanya instance pada setiap attribut dengan $H(\text{cerah}, \text{normal}, \text{pelan}) = \text{ya}$

$$\underline{H(\text{cuaca}=\text{cerah}, \text{kec.angin}=\text{pelan}) = \text{ya}}$$

Hipotesa ini menunjukkan bahwa keputusan untuk berolahraga bila cuaca=cerah, dan kec.angin=pelan, untuk singkatnya dituliskan hanya instance pada setiap attribut dengan $H(\text{cerah}, *, \text{pelan}) = \text{ya}$

$$\underline{H(\text{cuaca}=\text{cerah}) = \text{ya}}$$

Hipotesa ini menunjukkan bahwa keputusan untuk berolahraga bila cuaca=cerah, untuk singkatnya dituliskan hanya instance pada setiap attribut dengan $H(\text{cerah}, *, *) = \text{ya}$



Penulisan Hipotesa Dalam Matlab

Penulisan untuk hipotesa H(cerah, normal, pelan)=ya dalam MATLAB adalah:

$$H(1,:) = \text{'cerah', 'normal', 'pelan' }$$

Penulisan ini menggunakan format cell-array dengan array 1 baris dan 3 kolom (sesuai jumlah atribut yang dimasukkan). Nilai 1 berarti keputusan untuk YA.

Untuk hipotesa dengan keputusan tidak seperti H(hujan)=tidak dituliskan dengan indeks 2 sebagai berikut:

$$H(2,:) = \text{'hujan', '*', '*'} }$$



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.





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)



(1) Langkah Find-S

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

Untuk memperoleh hipotesa dengan Find-S dari data training di atas, langkah pertama yang harus dilakukan adalah memcah data berdasarkan keputusannya, sehingga akan diperoleh 2 data: pertama untuk keputusan=ya dan kedua untuk keputusan=tidak



(2) Langkah Find-S

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D4	Cerah	Normal	Kencang	Ya
D6	Cerah	Normal	Pelan	Ya

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D3	Hujan	Tinggi	Pelan	Tidak
D5	Hujan	Tinggi	Kencang	Tidak

Dari hasil pemisahan tersebut, terlihat bahwa data training tersebut Konsisten dan tidak bias. Langkah berikutnya adalah membuat hipotesa untuk masing-masing keputusan. Pembuatan hipotesa ini dilakukan dengan mengambil data pertama sebagai hipotesa awal yang dianggap sebagai hipotesa spesifik, dan diteruskan hingga data terakhir dengan memperhatikan kesamaan sampai didapatkan hipotesa umum.



[3] Langkah Find-S

Langkah ini digunakan untuk menentukan hipotesa dari keputusan=ya

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D4	Cerah	Normal	Kencang	Ya
D6	Cerah	Normal	Pelan	Ya

1

Hipotesa awal disamakan dengan data pertama:

$$H(\text{Cerah, Normal, Pelan}) = \text{Ya}$$

2

Data kedua, tidak ada perubahan karena semua nilai instancenya sama:

$$H(\text{Cerah, Normal, Pelan}) = \text{Ya}$$

3

Data ketiga, ada perubahan di atribut kec angin, sehingga :

$$H(\text{Cerah, Normal, *}) = \text{Ya}$$

4

Data keenam, ada perubahan di atribut kec angin, sehingga :

$$H(\text{Cerah, Normal, *}) = \text{Ya}$$



[4] Langkah Find-S

Langkah ini digunakan untuk menentukan hipotesa dari keputusan=tidak

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D3	Hujan	Tinggi	Pelan	Tidak
D5	Hujan	Tinggi	Kencang	Tidak

1

Hipotesa awal disamakan dengan data pertama (D3):
 $H(\text{Hujan, Tinggi, Pelan}) = \text{Tidak}$

2

Data kelima, ada perubahan di kec.angin
 $H(\text{Cerah, Normal, *}) = \text{Tidak}$



Hasil Find-S

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

Hasil Find-S adalah:

$$H(\text{Cerah, Normal, *}) = \text{Ya}$$

$$H(\text{Hujan, Tinggi, *}) = \text{Tidak}$$



Contoh Keputusan Dari Hipotesa

Bila cuaca cerah,
apakah akan
berolahraga?



HIPOTESA

$H(\text{Cerah, Normal, *}) = \text{Ya}$

$H(\text{Hujan, Tinggi, *}) = \text{Tidak}$

CEK

JAWAB : YA



Proses Dilakukan Untuk Setiap Instance

```
for n=1:nInstance
    % Menyeleksi data sesuai instance ke-l
    nh=0;
    for i=1:nData
        if strcmp(dtTraining{i,nAttribut},instance{n})==1
            nh=nh+1;
            dataH(nh,:)=dtTraining(i,:);
        end
    end
    % Proses Find-S untuk menentukan hipotesa
    % dari keputusan ke-l
    .....
end
```



Proses Find-S

```
% Hipotesa awal diambil dari data pertama dari data terseleksi
hipotesis(n,:)=dataH(1,2:nAttribut-1);
% Proses pembentukan hipotesa berdasarkan semua
% data terseleksi
for i=2:nh
    for j=2:nAttribut-1
        % Penentuan apakah ada instance yang berbeda
        % pada atribut yang sama
        % Bila ada maka hipotesa dijadikan *
        % Proses ini hanya dilakukan untuk hipotesa bukan *
        if strcmp(hipotesis(n,j-1),'')==0
            if strcmp(hipotesis{n,j-1},dataH{i,j})==0
                hipotesis{n,j-1}='*';
            end
        end
    end
end
end
end
```



Algoritma Berikutnya

**BAGAIMANA BILA DATA
TIDAK KONSISTEN DAN BIAS?**



METODE BAYES

ALGORITMA GENETIKA

JARINGAN SYARAF TIRUAN