Feature Selection (Practice)

Ali Ridho Barakbah

Soft Computing Laboratory

Department of Information and Computer Engineering
Electronic Engineering Polytechnic Institute of Surabaya





Bunga Iris





Iris Setosa



Iris Versicolor



Iris Virginica





Extracted Features



- Petal Length
- Petal Width
- Sepal Length
- Sepal Width



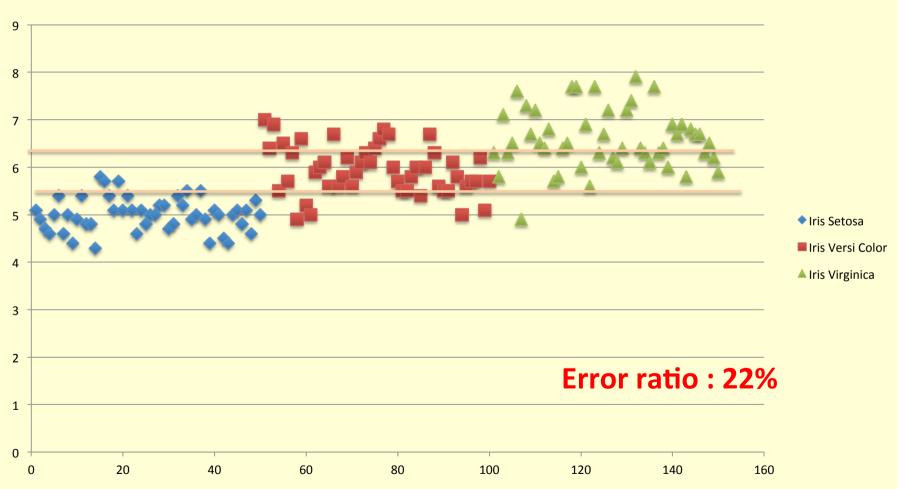
Informasi Dataset

- Source : UCI Repository
- Number of attributes: 4
- Number of instances: 150
- Number of classes: 3
 - Iris Setosa (50 instances)
 - Iris Versicolour (50 instances)
 - Iris Virginica (50 instances)





Sepal Length







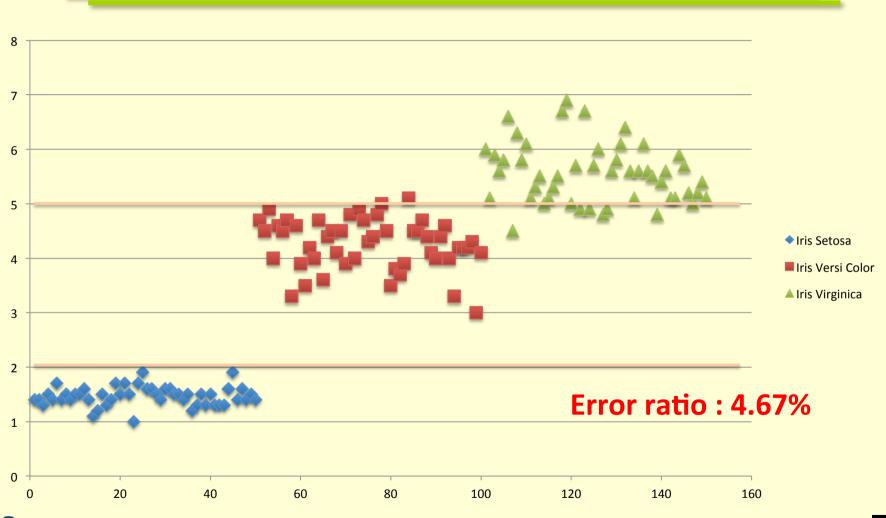
Sepal Width







Petal Length

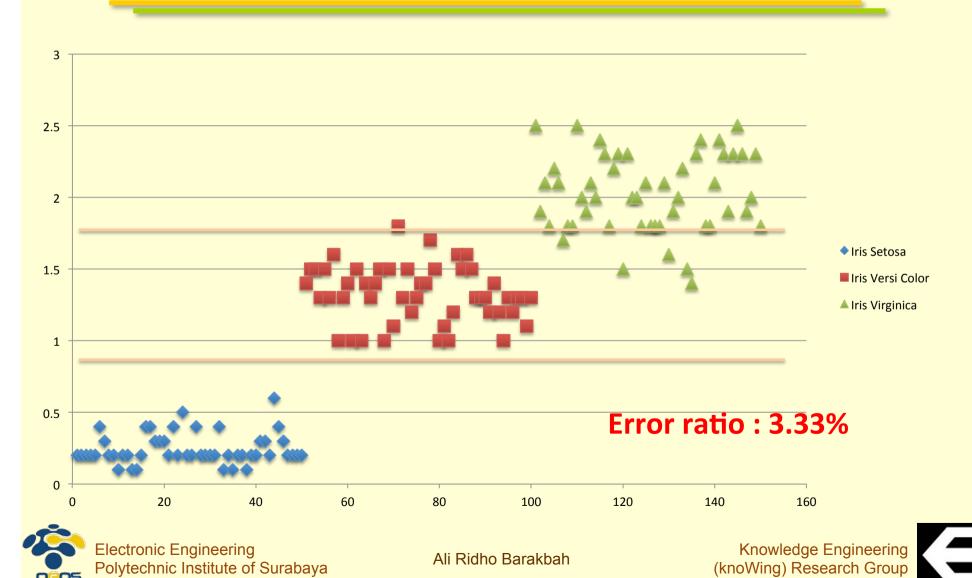


Ali Ridho Barakbah





Petal Width (Salah satu hasil terbaik)



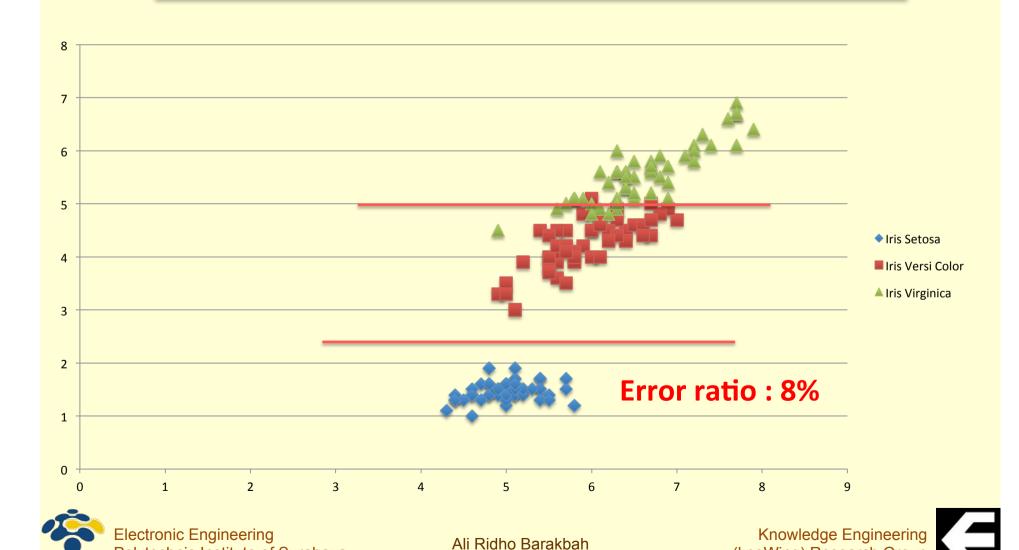
Sepal Length & Width



Ali Ridho Barakbah

(knoWing) Research Group

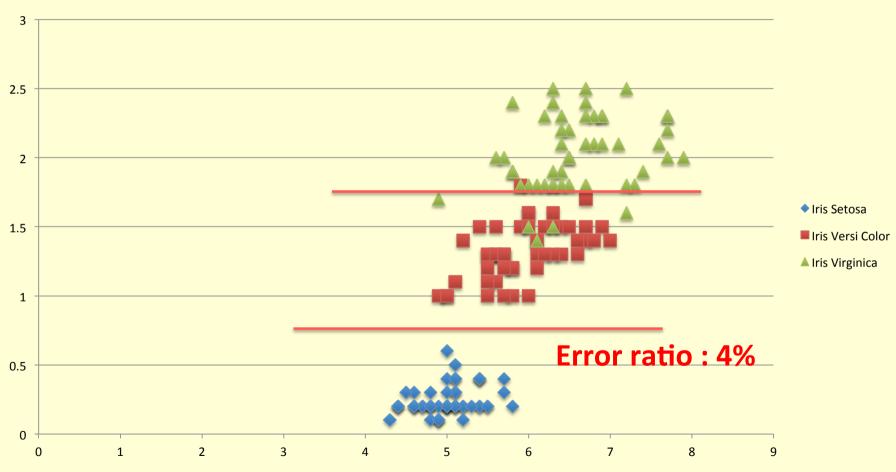
Sepal Length & Petal Length



(knoWing) Research Group

Polytechnic Institute of Surabaya

Sepal Length & Petal Width







Sepal Width & Petal Length



Ali Ridho Barakbah



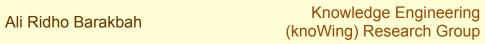




Sepal Width & Petal Width

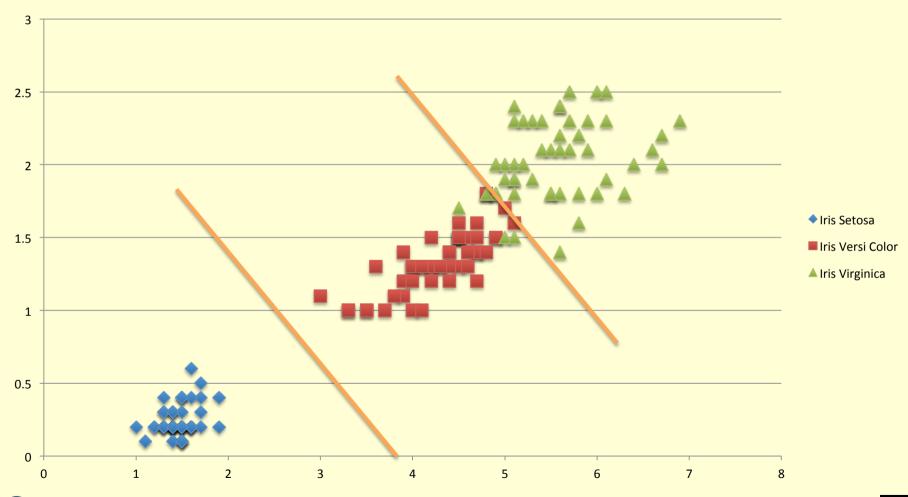








Petal Width & Petal Length (Salah satu hasil terbaik)



Ali Ridho Barakbah





Kesimpulan

- Pembeda yang terbaik didapat dengan mengekstraksi fitur Petal Width, dimana error terkecil muncul di angka 3.33%
- Atau bisa juga dengan menggunakan dua fitur yaitu Petal Width & Petal Length dimana error juga mencapai angka yang sama 3.33%



Tugas

- Studi kasus: Pengenalan Pola dari data survei nasional Kontrasepsi tahun 1987
- Tujuan: Pengembangan kebijakan dan strategi nasional untuk program pengendalian jumlah penduduk
- Jumlah data: 1473
- Jumlah atribut: 9





Extracted Features

```
    Wife's age (numerical)
    Wife's education (categorical) 1=low, 2, 3, 4=high
    Husband's education (categorical) 1=low, 2, 3, 4=high
    Number of children ever born (numerical)
    Wife's religion (binary) 0=Non-Islam, 1=Islam
    Wife's now working? (binary) 0=Yes, 1=No
    Husband's occupation (categorical) 1, 2, 3, 4
    Standard-of-living index (categorical) 1=low, 2, 3, 4=high
    Media exposure (binary) 0=Good, 1=Not good
```



Informasi Dataset

- Source : UCI Repository
- Number of attributes: 9
- Number of instances: 1473
- Number of classes: 3
 - No-use
 - Long-term
 - Short-term





Tugas

- Amatilah distribusi data untuk tiap fitur dan gabungan fitur
- Lakukan seleksi fitur-fitur manakah yang terbaik untuk klasifikasi. Berikan alasannya?

