

**Nama : Widiya Lis Susana**

**NIM : 211001002**

### **Dataset Iris**

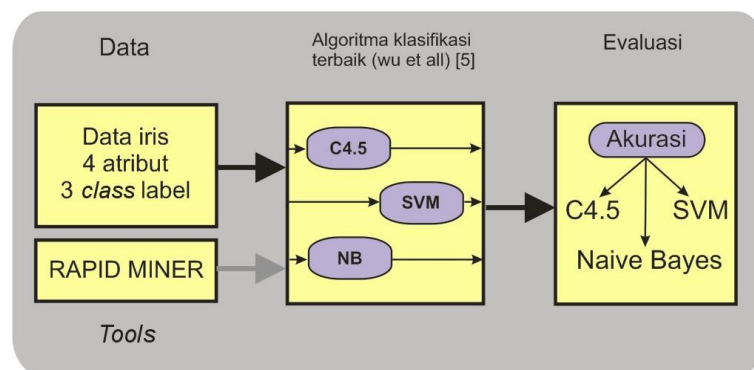
Dataset yang digunakan adalah data iris. Data iris merupakan salah satu dataset yang banyak digunakan dalam proses klasifikasi. Data iris merupakan data publik yang dikeluarkan oleh uci repository. dalam data iris ada 4 atribut yang dapat mempengaruhi klasifikasi yaitu: *sepal length*, *sepal width*, *petal length*, serta *petal width*. Atribut tujuan atau label dari data iris memiliki 3 kelas yaitu: iris setosa, iris versicolour, serta iris virginica. Tabel 1 menunjukkan potongan dataset data iris.

Tabel 1 Data iris

No.record	A1	A2	A3	A4	Class
1	5.100	3.500	1.400	0.200	Iris setosa
2	4.900	3.000	1.400	0.200	Iris setosa
3	4.700	3.200	1.300	0.200	Iris setosa
4	4.600	3.100	1.500	0.200	Iris setosa
...	...	...	...	...	...

...	...	...	...	...	...
147	6.300	2.500	5.000	1.900	Iris virginica
148	6.500	3.000	5.200	2.000	Iris virginica
149	6.200	3.400	5.400	2.300	Iris virginica
150	5.900	3.000	5.100	1.800	Iris virginica

Dataset yang digunakan adalah data iris yang nantinya akan dilakukan perhitungan dengan menggunakan *tools* rapid miner.



Gambar 1 Kerangka pemikiran

### ➤ Evaluasi hasil

Dalam setiap penelitian klasifikasi data mining pasti terdapat evaluasi untuk mengetahui tingkat akurasi dari algoritma klasifikasi. Dalam sebuah klasifikasi terdapat pembagian data menjadi 2, yaitu data *training* dan data *testing*. Data *training* merupakan bagian dari data yang digunakan untuk membuat suatu pola atau pengetahuan baru. Sedangkan data *testing* merupakan bagian data yang akan dipakai untuk mencoba pola tersebut guna mengetahui akurasi dari algoritma. Umumnya percobaan dilakukan secara berulang guna mendapatkan hasil akurasi yang lebih kuat. Evaluasi yang digunakan dalam penelitian ini adalah dengan menghitung rata-rata dari keseluruhan percobaan yang dilakukan.

### ➤ Hasil dan Pembahasan

Dari hasil diatas menunjukkan bahwa algoritma C4.5 merupakan algoritma dengan performa terbaik dengan mendapatkan akurasi sebesar 98.67%. Sedangkan Naïve Bayes mempunyai akurasi sebesar 96%. SVM dengan menggunakan data iris ternyata hanya mendapatkan akurasi sebesar 91.33%. Keseluruhan hasil dari penelitian ini dapat dilihat pada tabel 3.1 untuk C4.5, tabel 3.2 untuk Naïve Bayes, serta tabel 3.3 untuk SVM.

**Tabel 2 Performa algoritma C4.5 untuk dataset iris**

	<i>True Iris Setosa</i>	<i>True Iris Versicolour</i>	<i>True Iris Virginica</i>	<i>Class precission</i>
Pred. Iris Setosa	50	0	0	100%
Pred. Iris Versicolour	0	49	1	98.00%
Pred. Iris Virginica	0	1	49	98.00%
<i>Class recall</i>	100%	98.00%	98.00%	

**Tabel 3 Performa algoritma Naïve Bayes untuk dataset iris**

	<i>True Iris Setosa</i>	<i>True Iris Versicolour</i>	<i>True Iris Virginica</i>	<i>Class precission</i>
Pred. Iris Setosa	50	0	0	100%
Pred. Iris Versicolour	0	47	3	94.00%
Pred. Iris Virginica	0	3	47	94.00%
<i>Class recall</i>	100%	94.00%	94.00%	

**Tabel 4 Performa algoritma SVM untuk dataset iris**

	<i>True Iris Setosa</i>	<i>True Iris Versicolour</i>	<i>True Iris Virginica</i>	<i>Class precission</i>
Pred. Iris Setosa	49	0	0	100%
Pred. Iris Versicolour	1	40	2	93.02%
Pred. Iris Virginica	0	10	48	82.76%
<i>Class recall</i>	98.00%	80.00%	96.00%	

### ➤ Kesimpulan

Dari beberapa algoritma klasifikasi terbaik menurut Wu et al (2007) yaitu: C4.5, SVM, serta Naïve Bayes. Kesemuanya merupakan algoritma dengan golongan *best classification*. Karena akurasi dari kesemuanya menunjukkan angka diatas 90%. Dalam klasifikasi data iris C4.5 merupakan algoritma terbaik dengan tingkat akurasi 98,67%. Sedangkan Naïve Bayes 96% dan SVM 91,33%.