MEMBANDINGKAN ANALISA DATA SET MENGGUNAKAN APLIKASI WEKA DENGAN ALGORITMA ID3 DAN NAÏVE BAYES SECARA MANUAL

DISUSUN OLEH :

AFFNI SYAVIERA NOVA (5235111827)

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

UNIVERSITAS NEGERI JAKARTA

2015

I. Tentang ID3 dan Naïve Bayes

I.1 ID3

ID3 atau *Induction of Decision "3"*, atau bisa dikenal *Iterative Dichotomiser* 3, merupakan sebuah algoritma yang digunakan untuk mengklasifikasikan data. Dalam proses training ID3 akan menghasilkan decision tree, sebuah model yang digunakan untuk memprediksi label data *unknown*. ID3 dikembangkan oleh Ross Quinlan pada tahun 1970. Kemudian algoritma ini dikembangkan lagi oleh Quinlan pada tahun 1993 menjadi C4.5. ID3 menggunakan pendekatan *greedy search* dalam memilih atribut yang akan dijadikan sebagai *node*, sehingga ID3 tidak menjamin untuk mencapai solusi optimal, karena *greedy search* merupakan pendekatan yang menggunakan konsep *cost minimal*.

ID3 dimulai dengan menghitung Gain gain(S, V_n) setiap atribut, atribut dengan nilai Gain tertinggi akan dipilih untuk mensplit data set. Proses di atas akan dilakukan secara rekursif menggunakan atribut yang tersisa pada dataset.

$$\text{gain}(\text{S}, Vn) = E(s) - \sum \left|\frac{Sv}{S}\right| E(Sv)$$

$$E(s) = \sum \quad - pi \log 2\, pi$$

I.2 Naïve Bayes

Naïve Bayes merupakan *probabilistic classifier* yang mengasumsikan bahwa setiap komponen pada fitur adalah *independent* secara statistik. Latar belakang dari Naive Bayes adalah untuk menyelesaikan permasalahan *curse dimensionality* yang dialami oleh *Probability Density Function* (PDF) dalam mendapatkan hasil akurasi yang tinggi, di mana N data *points* dibutuhkan untuk mencapai performa yang maksimal dalam kasus *one-dimensional space*, dan N^l data *points* untuk kasus *l-dimensional space.*

$$\text{class} = \max \text{P (class k)} \prod_{\text{j}}^{\text{M}} p\ (xj|class\ k)$$

KESIMPULAN ANALISA:

1. **HASIL ANALISA MENGGUNAKAN WEKA**
   - **Analisa 1 menggunakan menu "use training set"**
   - **Analisa 2 menggunakan "percentage split 70%"**

Data yang di analisa:

parents	has_nurs	form	children	housing	finance	social	health	class
usual	critical	completed	2	critical	inconv	problematic	not_recom	not_recom
usual	critical	incomplete	2	less_conv	inconv	problematic	not_recom	not_recom
usual	critical	complete	2	convenient	convenient	nonprob	not_recom	not_recom
usual	critical	complete	2	convenient	convenient	slightly_prob	not_recom	not_recom
usual	critical	complete	1	convenient	convenient	nonprob	not_recom	not_recom
usual	critical	complete	2	less_conv	convenient	problematic	not_recom	not_recom
usual	critical	complete	3	convenient	inconv	problematic	not_recom	not_recom
usual	critical	complete	2	less_conv	inconv	problematic	not_recom	not_recom
usual	critical	complete	1	critical	inconv	problematic	not_recom	not_recom
usual	improper	complete	3	critical	convenient	problematic	not_recom	not_recom
usual	improper	incomplete	more	critical	convenient	problematic	not_recom	not_recom
pretentious	improper	completed	2	critical	inconv	problematic	not_recom	not_recom
pretentious	improper	complete	3	convenient	inconv	problematic	not_recom	not_recom
pretentious	improper	complete	2	less_conv	inconv	problematic	not_recom	not_recom
pretentious	improper	incomplete	2	critical	inconv	problematic	not_recom	not_recom
pretentious	improper	completed	2	critical	inconv	problematic	not_recom	not_recom
pretentious	less_proper	completed	3	less_conv	inconv	problematic	not_recom	not_recom
pretentious	less_proper	complete	3	convenient	inconv	nonprob	not_recom	not_recom
great_pret	less_proper	complete	more	critical	inconv	problematic	not_recom	not_recom
great_pret	less_proper	incomplete	more	critical	convenient	slightly_prob	not_recom	not_recom
great_pret	less_proper	foster	3	critical	inconv	problematic	not_recom	not_recom
great_pret	less_proper	complete	1	convenient	convenient	problematic	not_recom	not_recom

great_pret	proper	completed	more	critical	convenient	problematic	not_recom	not_recom
great_pret	proper	foster	3	less_conv	convenient	nonprob	not_recom	not_recom
great_pret	proper	incomplete	2	critical	inconv	nonprob	not_recom	not_recom
usual	critical	foster	1	critical	inconv	problematic	recommended	priority
usual	critical	complete	2	convenient	convenient	slightly_prob	priority	priority
usual	critical	complete	2	convenient	convenient	problematic	recommended	priority
usual	critical	complete	1	convenient	convenient	nonprob	priority	priority
usual	critical	complete	1	convenient	convenient	problematic	recommended	priority
usual	critical	complete	1	convenient	convenient	nonprob	priority	priority
usual	critical	complete	2	less_conv	inconv	nonprob	priority	priority
usual	critical	complete	3	convenient	convenient	nonprob	recommended	priority
usual	critical	complete	2	convenient	convenient	nonprob	recommended	priority
pretentious	improper	complete	more	less_conv	inconv	slightly_prob	recommended	priority
pretentious	improper	completed	1	convenient	convenient	nonprob	priority	priority
pretentious	improper	complete	3	less_conv	convenient	nonprob	recommended	priority
pretentious	improper	complete	1	critical	inconv	nonprob	priority	priority
pretentious	improper	complete	more	critical	inconv	slightly_prob	recommended	priority
pretentious	improper	completed	more	critical	convenient	nonprob	recommended	priority
pretentious	improper	incomplete	1	less_conv	inconv	nonprob	priority	priority
pretentious	improper	incomplete	3	convenient	convenient	nonprob	recommended	priority
great_pret	less_proper	completed	2	convenient	convenient	nonprob	recommended	priority
great_pret	less_proper	foster	2	less_conv	inconv	nonprob	recommended	priority
great_pret	less_proper	complete	2	convenient	inconv	nonprob	recommended	priority
great_pret	less_proper	foster	more	convenient	convenient	nonprob	recommended	priority
great_pret	proper	complete	1	critical	inconv	slightly_prob	recommended	priority
great_pret	proper	foster	2	convenient	convenient	nonprob	recommended	priority
usual	proper	complete	1	convenient	convenient	nonprob	recommended	recommend
usual	critical	incomplete	more	critical	inconv	nonprob	recommended	spec_prior
pretentious	improper	foster	3	critical	inconv	problematic	recommended	spec_prior
pretentious	improper	completed	2	critical	inconv	problematic	priority	spec_prior

pretentious	less_proper	completed	3	critical	convenient	nonprob	recommended	spec_prior
pretentious	less_proper	completed	1	critical	inconv	nonprob	priority	spec_prior
pretentious	less_proper	complete	more	convenient	inconv	nonprob	recommended	spec_prior
great_pret	very_crit	complete	more	critical	convenient	nonprob	priority	spec_prior
great_pret	very_crit	complete	more	convenient	inconv	problematic	priority	spec_prior
great_pret	very_crit	foster	2	less_conv	convenient	slightly_prob	priority	spec_prior
great_pret	very_crit	foster	more	less_conv	convenient	slightly_prob	recommended	spec_prior
usual	very_crit	completed	1	critical	inconv	nonprob	recommended	spec_prior
pretentious	very_crit	foster	more	less_conv	convenient	nonprob	recommended	spec_prior
great_pret	proper	incomplete	2	critical	inconv	slightly_prob	priority	spec_prior
great_pret	proper	foster	2	convenient	inconv	nonprob	recommended	spec_prior
great_pret	proper	completed	2	convenient	inconv	nonprob	priority	spec_prior
great_pret	proper	completed	3	critical	inconv	problematic	recommended	spec_prior
usual	critical	completed	3	convenient	convenient	nonprob	recommended	very_recom
usual	critical	complete	2	convenient	convenient	slightly_prob	recommended	very_recom
usual	critical	complete	1	convenient	convenient	nonprob	recommended	very_recom
pretentious	improper	completed	1	convenient	convenient	nonprob	recommended	very_recom
pretentious	improper	completed	3	convenient	convenient	nonprob	recommended	very_recom
great_pret	proper	incomplete	1	less_conv	convenient	nonprob	not_recom	not_recom
great_pret	proper	completed	3	critical	inconv	problematic	not_recom	not_recom
usual	proper	complete	2	less_conv	inconv	nonprob	not_recom	not_recom
usual	proper	incomplete	2	critical	inconv	problematic	not_recom	not_recom
pretentious	very_crit	foster	1	convenient	inconv	problematic	not_recom	not_recom
usual	improper	foster	2	convenient	convenient	nonprob	recommended	very_recom
usual	less_proper	complete	more	convenient	convenient	nonprob	recommended	very_recom
pretentious	very_crit	complete	2	less_conv	inconv	slightly_prob	recommended	very_recom
pretentious	less_proper	completed	2	convenient	inconv	nonprob	recommended	very_recom
usual	proper	completed	2	critical	inconv	problematic	priority	priority
usual	proper	complete	more	convenient	inconv	nonprob	recommended	priority
great_pret	proper	incomplete	1	less_conv	convenient	nonprob	priority	priority

usual	proper	incomplete	3	critical	convenient	slightly_prob	priority	priority
usual	proper	completed	more	critical	convenient	slightly_prob	recommended	priority
great_pret	very_crit	complete	1	critical	inconv	slightly_prob	recommended	priority
great_pret	very_crit	incomplete	2	convenient	convenient	nonprob	recommended	priority
pretentious	very_crit	complete	1	less_conv	inconv	nonprob	recommended	priority
great_pret	proper	completed	more	critical	inconv	problematic	priority	spec_prior
great_pret	proper	foster	2	less_conv	convenient	nonprob	priority	spec_prior
usual	very_crit	foster	1	critical	inconv	problematic	priority	spec_prior
pretentious	very_crit	completed	3	critical	inconv	nonprob	recommended	spec_prior
pretentious	very_crit	completed	3	critical	inconv	problematic	recommended	spec_prior
great_pret	very_crit	incomplete	2	convenient	inconv	problematic	priority	spec_prior
great_pret	very_crit	foster	3	less_conv	inconv	problematic	priority	spec_prior
pretentious	less_proper	complete	3	convenient	inconv	slightly_prob	recommended	spec_prior
pretentious	less_proper	complete	2	critical	convenient	slightly_prob	recommended	spec_prior
pretentious	less_proper	foster	1	critical	convenient	nonprob	priority	spec_prior
great_pret	less_proper	foster	more	convenient	inconv	slightly_prob	priority	spec_prior
great_pret	less_proper	completed	1	critical	convenient	nonprob	recommended	spec_prior
great_pret	less_proper	incomplete	3	less_conv	convenient	nonprob	recommended	spec_prior

- **ID3:**
  - **Nilai Akurasi**

Analisa 1:

```
Correctly Classified Instances         100               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              100
```

Analisa 2:

```
Correctly Classified Instances          22               73.3333 %
Incorrectly Classified Instances         6               20      %
Kappa statistic                          0.6917
Mean absolute error                      0.0857
Root mean squared error                  0.2928
Relative absolute error                 31.7827 %
Root relative squared error             80.7948 %
UnClassified Instances                   2                6.6667 %
Total Number of Instances               30
```

  - **Confussion Matrix**

Analisa 1:

```
a  b  c  d  e   <-- classified as
1  0  0  0  0 |  a = recommend
0 31  0  0  0 |  b = priority
0  0 30  0  0 |  c = not_recom
0  0  0  9  0 |  d = very_recom
0  0  0  0 29 |  e = spec_prior
```

Analisa 2:

```
a b c d e   <-- classified as
0 0 0 0 0 | a = recommend
0 8 0 1 1 | b = priority
0 0 8 0 0 | c = not_recom
0 1 0 0 0 | d = very_recom
0 2 0 1 6 | e = spec_prior
```

- **Rules Analisa 1:**

```
health = recommended
|  housing = convenient
|  |  has_nurs = proper
|  |  |  children = 1: recommend
|  |  |  children = 2
|  |  |  |  finance = convenient: priority
|  |  |  |  finance = inconv: spec_prior
|  |  |  children = 3: null
|  |  |  children = more: priority
|  |  has_nurs = less_proper
|  |  |  parents = usual: very_recom
|  |  |  parents = pretentious
|  |  |  |  form = complete: spec_prior
|  |  |  |  form = completed: very_recom
|  |  |  |  form = incomplete: null
|  |  |  |  form = foster: null
|  |  |  parents = great_pret: priority
|  |  has_nurs = improper
|  |  |  form = complete: null
|  |  |  form = completed: very_recom
|  |  |  form = incomplete: priority
|  |  |  form = foster: very_recom
|  |  has_nurs = critical
|  |  |  social = nonprob
|  |  |  |  children = 1: very_recom
|  |  |  |  children = 2: priority
|  |  |  |  children = 3
|  |  |  |  |  form = complete: priority
|  |  |  |  |  form = completed: very_recom
|  |  |  |  |  form = incomplete: null
|  |  |  |  |  form = foster: null
|  |  |  |  children = more: null
|  |  |  social = slightly_prob: very_recom
|  |  |  social = problematic: priority
|  |  has_nurs = very_crit: priority
|  housing = less_conv
|  |  form = complete
|  |  |  children = 1: priority
|  |  |  children = 2: very_recom
|  |  |  children = 3: priority
|  |  |  children = more: priority
|  |  form = completed: null
|  |  form = incomplete: spec_prior
|  |  form = foster
|  |  |  has_nurs = proper: null
|  |  |  has_nurs = less_proper: priority
|  |  |  has_nurs = improper: null
|  |  |  has_nurs = critical: null
|  |  |  has_nurs = very_crit: spec_prior
|  housing = critical
|  |  children = 1
|  |  |  form = complete: priority
|  |  |  form = completed: spec_prior
|  |  |  form = incomplete: null
|  |  |  form = foster: priority
|  |  children = 2: spec_prior
|  |  children = 3: spec_prior
|  |  children = more
|  |  |  has_nurs = proper: priority
|  |  |  has_nurs = less_proper: null
|  |  |  has_nurs = improper: priority
|  |  |  has_nurs = critical: spec_prior
|  |  |  has_nurs = very_crit: null
health = priority
|  has_nurs = proper
|  |  children = 1: priority
```

```
|  |  children = 2
|  |  |  parents = usual: priority
|  |  |  parents = pretentious: null
|  |  |  parents = great_pret: spec_prior
|  |  children = 3: priority
|  |  children = more: spec_prior
|  has_nurs = less_proper: spec_prior
|  has_nurs = improper
|  |  children = 1: priority
|  |  children = 2: spec_prior
|  |  children = 3: null
|  |  children = more: null
|  has_nurs = critical: priority
|  has_nurs = very_crit: spec_prior
health = not_recom: not_recom
```

Analisa 2:

```
health = recommended
|  housing = convenient
|  |  has_nurs = proper
|  |  |  children = 1: recommend
|  |  |  children = 2
|  |  |  |  finance = convenient: priority
|  |  |  |  finance = inconv: spec_prior
|  |  |  children = 3: null
|  |  |  children = more: priority
|  |  has_nurs = less_proper
|  |  |  parents = usual: very_recom
|  |  |  parents = pretentious
|  |  |  |  form = complete: spec_prior
|  |  |  |  form = completed: very_recom
|  |  |  |  form = incomplete: null
|  |  |  |  form = foster: null
|  |  |  parents = great_pret: priority
|  |  has_nurs = improper
|  |  |  form = complete: null
|  |  |  form = completed: very_recom
|  |  |  form = incomplete: priority
|  |  |  form = foster: very_recom
|  |  has_nurs = critical
|  |  |  social = nonprob
|  |  |  |  children = 1: very_recom
|  |  |  |  children = 2: priority
|  |  |  |  children = 3
|  |  |  |  |  form = complete: priority
|  |  |  |  |  form = completed: very_recom
|  |  |  |  |  form = incomplete: null
|  |  |  |  |  form = foster: null
|  |  |  |  children = more: null
|  |  |  social = slightly_prob: very_recom
|  |  |  social = problematic: priority
|  |  has_nurs = very_crit: priority
|  housing = less_conv
|  |  form = complete
|  |  |  children = 1: priority
|  |  |  children = 2: very_recom
|  |  |  children = 3: priority
|  |  |  children = more: priority
|  |  form = completed: null
|  |  form = incomplete: spec_prior
|  |  form = foster
|  |  |  has_nurs = proper: null
|  |  |  has_nurs = less_proper: priority
|  |  |  has_nurs = improper: null
|  |  |  has_nurs = critical: null
|  |  |  has_nurs = very_crit: spec_prior
|  housing = critical
|  |  children = 1
|  |  |  form = complete: priority
|  |  |  form = completed: spec_prior
|  |  |  form = incomplete: null
|  |  |  form = foster: priority
|  |  children = 2: spec_prior
|  |  children = 3: spec_prior
|  |  children = more
|  |  |  has_nurs = proper: priority
|  |  |  has_nurs = less_proper: null
|  |  |  has_nurs = improper: priority
|  |  |  has_nurs = critical: spec_prior
|  |  |  has_nurs = very_crit: null
health = priority
|  has_nurs = proper
```

```
|  |  children = 1: priority
|  |  children = 2
|  |  |  parents = usual: priority
|  |  |  parents = pretentious: null
|  |  |  parents = great_pret: spec_prior
|  |  children = 3: priority
|  |  children = more: spec_prior
|  has_nurs = less_proper: spec_prior
|  has_nurs = improper
|  |  children = 1: priority
|  |  children = 2: spec_prior
|  |  children = 3: null
|  |  children = more: null
|  has_nurs = critical: priority
|  has_nurs = very_crit: spec_prior
health = not_recom: not_recom
```

- **NAIVE BAYES:**
  - **Nilai Akurasi**

Analisa 1:

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances          84               84      %
Incorrectly Classified Instances        16               16      %
Kappa statistic                          0.7768
Mean absolute error                      0.1199
Root mean squared error                  0.2225
Relative absolute error                 41.3384 %
Root relative squared error             58.5583 %
Total Number of Instances              100
```

Analisa 2:

```
=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances          20               66.6667 %
Incorrectly Classified Instances        10               33.3333 %
Kappa statistic                          0.5313
Mean absolute error                      0.1774
Root mean squared error                  0.3008
Relative absolute error                 61.3597 %
Root relative squared error             80.1698 %
Total Number of Instances               30
```

- **Confussion Matrix**

Analisa 1:

```
a  b  c  d  e   <-- classified as
0  1  0  0  0 |  a = recommend
0 24  0  2  5 |  b = priority
0  0 30  0  0 |  c = not_recom
0  3  0  5  1 |  d = very_recom
0  3  0  1 25 |  e = spec_prior
```

Analisa 2:

```
a b c d e   <-- classified as
0 0 0 0 0 | a = recommend
0 8 1 2 0 | b = priority
0 0 7 1 0 | c = not_recom
0 0 0 0 1 | d = very_recom
0 4 0 1 5 | e = spec_prior
```