

TUGAS 6 MELAKUKAN PROSES KLASIFIKASI TERHADAP DATASET NURSERY MENGGUNAKAN METODE NAÏVE BAYESIAN

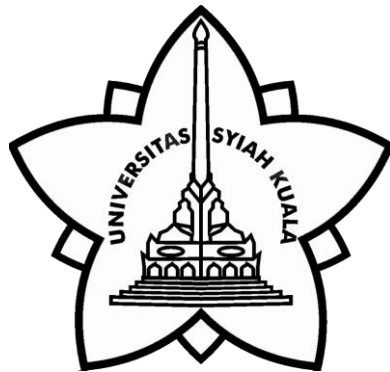
Disusun Untuk Memenuhi

Tugas Data Mining

Oleh:

Khairul Umam Albi

2008107010072



**JURUSAN INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SYIAH KUALA
DARUSSALAM, BANDA ACEH
2022**

MELAKUKAN PROSES KLASIFIKASI TERHADAP DATASET NURSERY MENGGUNAKAN METODE NAÏVE BAYESIAN.

1. Tools yang digunakan

- Aplikasi Weka
- Text editor(notepad dan vscode)
- Kalkulator

2. Langkah Proses Klasifikasi dataset Nursery

- Dataset dapat diunduh melalui link berikut :
<http://archive.ics.uci.edu/ml/datasets/Nursery>
- Menganti dan mengubah file menjadi bentuk arrf
- Buka applikasi weka dan pilih opsi open didalam applikasi tersebut.
- Pilih tab classifikasi dan pilih metode naïve bayes dari modul bayes sebagai classifier
- Pilih atribut
- Pilih opsi test option dengan cross-validation 5 dan 10
- Pilih atribut dan cross-validation dan tekan start.

A. Cross-validation folds 5

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	not_recom
	0.000	0.000	?	0.000	?	?	0.913	0.001	recommend
	0.061	0.000	0.870	0.061	0.114	0.227	0.995	0.829	very_recom
	0.902	0.097	0.820	0.902	0.859	0.787	0.966	0.937	priority
	0.867	0.047	0.894	0.867	0.880	0.828	0.978	0.942	spec_prior
Weighted Avg.	0.902	0.047	?	0.902	?	?	0.982	0.957	

```
=== Confusion Matrix ===
```

a	b	c	d	e	<-- classified as
4320	0	0	0	0	a = not_recom
0	0	2	0	0	b = recommend
0	0	20	308	0	c = very_recom
0	0	1	3848	417	d = priority
0	0	0	536	3508	e = spec_prior

akurasi per Class

a. not_recom

- $P = TP / (FP + TP)$
 $P = 4320 / (0+0+0+0+ 4320) = 1$
- $R = TP / (FN + TP)$
 $R = 4320 / (0+0+0+0+ 4320) = 1$
- $F = (2*P*R) / (P + R)$
 $F = (2*1*1) / (1 + 1) = 1$

b. recommend

- $P = TP / (FP + TP)$
 $P = 0 / (0+0+0+0+0) = \text{indeterminate}$
- $R = TP / (FN + TP)$
 $R = 0 / (0+2+0+0 + 0) = 0$
- $F = (2*P*R) / (P + R)$
 $F = (2*0*2) / (0 + 2) = \text{indeterminate}$

c. very_recom

- $P = TP / (FP + TP)$
 $P = 20 / (0 + 2 + 1 + 0 + 20) = 0,87$
- $R = TP / (FN + TP)$
 $R = 20 / (0 + 0 + 308 + 0 + 20) = 0,061$
- $F = (2*P*R) / (P + R)$
 $F = (2*0,87*0,061) / (0,87 + 0,061) = 0,1140$

d. priority

- $P = TP / (FP + TP)$
 $P = 3848 / (0 + 0 + 308 + 536 + 3848) = 0.820$
- $R = TP / (FN + TP)$
 $R = 3848 / (0 + 1 + 417 + 3848) = 0.902$
- $F = (2*P*R) / (P + R)$
 $F = (2*0.820*0.902) / (0.820+0.902) = 0.859$

e. spec_prior

- $P = TP / (FP + TP)$
 $P = 3508 / (0 + 0 + 0 + 417 + 3508) = 0.894$
- $R = TP / (FN + TP)$
 $R = 3508 / (0 + 0 + 536 + 3508) = 0.867$
- $F = (2*P*R) / (P + R)$
 $F = (2 * 0.894 * 0.867) / (0.894 + 0.867) = 0.880$

B. Cross-validation folds 10

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	not_recom
	0.000	0.000	?	0.000	?	?	0.906	0.001	recommend
	0.058	0.000	0.905	0.058	0.109	0.226	0.995	0.827	very_recom
	0.903	0.096	0.821	0.903	0.860	0.789	0.966	0.937	priority
	0.869	0.047	0.894	0.869	0.882	0.829	0.978	0.943	spec_prior
Weighted Avg.	0.903	0.046	?	0.903	?	?	0.982	0.957	

```

=== Confusion Matrix ===

```

	a	b	c	d	e	<-- classified as
a	4320	0	0	0	0	a = not_recom
b	0	0	2	0	0	b = recommend
c	0	0	19	309	0	c = very_recom
d	0	0	0	3851	415	d = priority
e	0	0	0	528	3516	e = spec_prior

Akurasi per class

a. not_recom

- $P = TP / (FP + TP)$
 $P = 4320 / (0+0+0+0+ 4320) = 1$
- $R = TP / (FN + TP)$
 $R = 4320 / (0+0+0+0+ 4320) = 1$
- $F = (2*P*R) / (P + R)$
 $F = (2*1*1) / (1 + 1) = 1$

b. recommend

- $P = TP / (FP + TP)$
 $P = 0 / (0+0+0+0+0) = \text{indeterminate}$
- $R = TP / (FN + TP)$
 $R = 0 / (0+2+0+0 + 0) = 0$
- $F = (2*P*R) / (P + R)$
 $F = (2*0*2) / (0 + 2) = \text{indeterminate}$

c. very_recom

- $P = TP / (FP + TP)$
 $P = 19 / (0+ 2 + 0 + 0 + 20) = 0.905$
- $R = TP / (FN + TP)$
 $R = 19 / (0 + 0 + 309 +0 +19) = 0.058$
- $F = (2*P*R) / (P + R)$
 $F = (2*0.905*0.058) / (0.905+ 0.058) = 0.109$

d. priority

- $P = TP / (FP + TP)$
 $P = 3851 / (0 + 0 + 309 + 528 + 3851) = 0.821$
- $R = TP / (FN + TP)$
 $R = 3851 / (0 + 1 + 415+ 3851) = 0.096$
- $F = (2*P*R) / (P + R)$
 $F = (2*0.821*0.096) / (0.821+0.096) = 0.860$

e. spec_prior

- $P = TP / (FP + TP)$

$$P = 3516 / (0 + 0 + 0 + 415 + 3516) = 0.894$$

- $R = TP / (FN + TP)$

$$R = 3508 / (0 + 0 + 528 + 3508) = 0.869$$

- $F = (2 * P * R) / (P + R)$

$$F = (2 * 0.894 * 0.869) / (0.894 + 0.869) = 0.882$$

C. Kesimpulan

Dapat disimpulkan bahwa :

1. Cross-validation membolehkan penggunaan testing langsung dari dataset yang digunakan untuk training. dataset dibagi menjadi beberapa bagian yang dihitung melalui kelipatannya, disebut folds.
2. Pada folds 5 ditemukan bahwa presision yang mendekati angka 1 adalah class spec_prior dengan presision 0.894 , recall yang mendekati angka 1 adalah class priority dengan recall 0.902 , dan f-measure yang mendakati angka 1 adalah class spec_prior dengan f-measure 0.880.
3. Pada folds 10 ditemukan bahwa presision yang mendekati angka 1 adalah class very_recom dengan presision 0.905 , recall yang mendekati angka 1 adalah class priority dengan recall 0.903 , dan f-measure yang mendakati angka 1 adalah class spec_prior dengan f-measure 0.882
4. Dengan menggunakan folds 10 daripada folds 5 ditemukan bahwa nilai dari hasil presision, recal,dan f-measure meningkat Ketika menggunakan folds 10.
5. Dan yang mendapat nilai sempurna dalam presision , recall, dan f-measure adalah class not_recom dengan nilai 1 pada folds 5 dan 10 dan rata-rata nilai f-measure tertinggi diperoleh oleh spec_prior dengan nilai rata-rata adalah 0.880 pada folds 5 dan 0.882 pada folds 10.