TUGAS BESAR II

IF3170 – INTELEGENSIA BUATAN

KELOMPOK 10 :

Thea Olivia 13511001

Pratama Nugraha Damanik 13513001

Ignatius Alriana 13513051

Lucky Chayadi 13513061

Mahesa Gandakusuma 13513091

PROGRAM STUDI TEKNIK INFORMATIKA

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

INSTITUT TEKNOLOGI BANDUNG

BANDUNG

2015

Hasil simulasi WEKA

k-Nearest Neighbor

10-fold validation

```
=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Relation:     car
Instances:    1728
Attributes:   7
              buying
              maint
              doors
              persons
              lug_boot
              safety
              class
Test mode:evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier

                Class
Attribute       unacc    acc  vgood   good
                (0.7) (0.22) (0.04) (0.04)
============================================
buying
  vhigh         361.0   73.0    1.0    1.0
  high          325.0  109.0    1.0    1.0
  med           269.0  116.0   27.0   24.0
  low           259.0   90.0   40.0   47.0
  [total]      1214.0  388.0   69.0   73.0

maint
  vhigh         361.0   73.0    1.0    1.0
  high          315.0  106.0   14.0    1.0
  med           269.0  116.0   27.0   24.0
  low           269.0   93.0   27.0   47.0
  [total]      1214.0  388.0   69.0   73.0

doors
```

2.0	327.0	82.0	11.0	16.0
3.0	301.0	100.0	16.0	19.0
4.0	293.0	103.0	21.0	19.0
5more	293.0	103.0	21.0	19.0
[total]	1214.0	388.0	69.0	73.0

persons

2.0	577.0	1.0	1.0	1.0
4.0	313.0	199.0	31.0	37.0
more	323.0	187.0	36.0	34.0
[total]	1213.0	387.0	68.0	72.0

lug_boot

small	451.0	106.0	1.0	22.0
med	393.0	136.0	26.0	25.0
big	369.0	145.0	41.0	25.0
[total]	1213.0	387.0	68.0	72.0

safety

low	577.0	1.0	1.0	1.0
med	358.0	181.0	1.0	40.0
high	278.0	205.0	66.0	31.0
[total]	1213.0	387.0	68.0	72.0

Time taken to build model: 0.05 seconds

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

Correctly Classified Instances	1505	87.0949 %
Incorrectly Classified Instances	223	12.9051 %
Kappa statistic	0.7065	
Mean absolute error	0.1112	
Root mean squared error	0.2218	
Relative absolute error	48.5842 %	
Root relative squared error	65.5935 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.96	0.168	0.93	0.96	0.945	0.985	unacc
	0.747	0.091	0.702	0.747	0.724	0.959	acc
	0.538	0.001	0.946	0.538	0.686	0.999	vgood
	0.304	0.007	0.636	0.304	0.412	0.986	good
Weighted Avg.	0.871	0.138	0.868	0.871	0.865	0.979	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1162	46	0	2	a = unacc
87	287	0	10	b = acc
0	30	35	0	c = vgood
0	46	2	21	d = good

full training

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Relation: car
Instances: 1728
Attributes: 7
buying
maint
doors
persons
lug_boot
safety
class
Test mode:evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier

	Class			
Attribute	unacc	acc	vgood	good
	(0.7)	(0.22)	(0.04)	(0.04)
==========				
buying				
vhigh	361.0	73.0	1.0	1.0

high	325.0	109.0	1.0	1.0
med	269.0	116.0	27.0	24.0
low	259.0	90.0	40.0	47.0
[total]	1214.0	388.0	69.0	73.0

maint

vhigh	361.0	73.0	1.0	1.0
high	315.0	106.0	14.0	1.0
med	269.0	116.0	27.0	24.0
low	269.0	93.0	27.0	47.0
[total]	1214.0	388.0	69.0	73.0

doors

2.0	327.0	82.0	11.0	16.0
3.0	301.0	100.0	16.0	19.0
4.0	293.0	103.0	21.0	19.0
5more	293.0	103.0	21.0	19.0
[total]	1214.0	388.0	69.0	73.0

persons

2.0	577.0	1.0	1.0	1.0
4.0	313.0	199.0	31.0	37.0
more	323.0	187.0	36.0	34.0
[total]	1213.0	387.0	68.0	72.0

lug_boot

small	451.0	106.0	1.0	22.0
med	393.0	136.0	26.0	25.0
big	369.0	145.0	41.0	25.0
[total]	1213.0	387.0	68.0	72.0

safety

low	577.0	1.0	1.0	1.0
med	358.0	181.0	1.0	40.0
high	278.0	205.0	66.0	31.0
[total]	1213.0	387.0	68.0	72.0

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	1505	87.0949 %
Incorrectly Classified Instances	223	12.9051 %
Kappa statistic	0.7065	
Mean absolute error	0.1112	
Root mean squared error	0.2218	
Relative absolute error	48.5842 %	
Root relative squared error	65.5935 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.96	0.168	0.93	0.96	0.945	0.985	unacc
	0.747	0.091	0.702	0.747	0.724	0.959	acc
	0.538	0.001	0.946	0.538	0.686	0.999	vgood
	0.304	0.007	0.636	0.304	0.412	0.986	good
Weighted Avg.	0.871	0.138	0.868	0.871	0.865	0.979	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1162	46	0	2	a = unacc
87	287	0	10	b = acc
0	30	35	0	c = vgood
0	46	2	21	d = good

Naïve Bayes

10-fold Cross Validation

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Relation: car
Instances: 1728
Attributes: 7
buying
maint
doors
persons
lug_boot
safety

class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

	Class			
Attribute	unacc	acc	vgood	good
	(0.7)	(0.22)	(0.04)	(0.04)
buying				
vhigh	361.0	73.0	1.0	1.0
high	325.0	109.0	1.0	1.0
med	269.0	116.0	27.0	24.0
low	259.0	90.0	40.0	47.0
[total]	1214.0	388.0	69.0	73.0
maint				
vhigh	361.0	73.0	1.0	1.0
high	315.0	106.0	14.0	1.0
med	269.0	116.0	27.0	24.0
low	269.0	93.0	27.0	47.0
[total]	1214.0	388.0	69.0	73.0
doors				
2.0	327.0	82.0	11.0	16.0
3.0	301.0	100.0	16.0	19.0
4.0	293.0	103.0	21.0	19.0
5more	293.0	103.0	21.0	19.0
[total]	1214.0	388.0	69.0	73.0
persons				
2.0	577.0	1.0	1.0	1.0
4.0	313.0	199.0	31.0	37.0
more	323.0	187.0	36.0	34.0
[total]	1213.0	387.0	68.0	72.0
lug_boot				
small	451.0	106.0	1.0	22.0
med	393.0	136.0	26.0	25.0
big	369.0	145.0	41.0	25.0

[total]	1213.0	387.0	68.0	72.0

safety

low	577.0	1.0	1.0	1.0
med	358.0	181.0	1.0	40.0
high	278.0	205.0	66.0	31.0
[total]	1213.0	387.0	68.0	72.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	1478	85.5324 %
Incorrectly Classified Instances	250	14.4676 %
Kappa statistic	0.6665	
Mean absolute error	0.1137	
Root mean squared error	0.2262	
Relative absolute error	49.6626 %	
Root relative squared error	66.9048 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.96	0.203	0.917	0.96	0.938	0.982	unacc
	0.706	0.098	0.672	0.706	0.689	0.95	acc
	0.415	0.001	0.931	0.415	0.574	0.998	vgood
	0.275	0.007	0.633	0.275	0.384	0.98	good
Weighted Avg.	0.855	0.164	0.852	0.855	0.847	0.976	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1161	48	0	1	a = unacc
104	271	0	9	b = acc
0	37	27	1	c = vgood
1	47	2	19	d = good

full training

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Relation: car
Instances: 1728
Attributes: 7
buying
maint
doors
persons
lug_boot
safety
class
Test mode:evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier

	Class			
Attribute	unacc	acc	vgood	good
	(0.7)	(0.22)	(0.04)	(0.04)
buying				
vhigh	361.0	73.0	1.0	1.0
high	325.0	109.0	1.0	1.0
med	269.0	116.0	27.0	24.0
low	259.0	90.0	40.0	47.0
[total]	1214.0	388.0	69.0	73.0
maint				
vhigh	361.0	73.0	1.0	1.0
high	315.0	106.0	14.0	1.0
med	269.0	116.0	27.0	24.0
low	269.0	93.0	27.0	47.0
[total]	1214.0	388.0	69.0	73.0
doors				
2.0	327.0	82.0	11.0	16.0
3.0	301.0	100.0	16.0	19.0
4.0	293.0	103.0	21.0	19.0
5more	293.0	103.0	21.0	19.0

[total]	1214.0	388.0	69.0	73.0

persons

2.0	577.0	1.0	1.0	1.0
4.0	313.0	199.0	31.0	37.0
more	323.0	187.0	36.0	34.0
[total]	1213.0	387.0	68.0	72.0

lug_boot

small	451.0	106.0	1.0	22.0
med	393.0	136.0	26.0	25.0
big	369.0	145.0	41.0	25.0
[total]	1213.0	387.0	68.0	72.0

safety

low	577.0	1.0	1.0	1.0
med	358.0	181.0	1.0	40.0
high	278.0	205.0	66.0	31.0
[total]	1213.0	387.0	68.0	72.0

Time taken to build model: 0.02 seconds

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

Correctly Classified Instances	1505	87.0949 %
Incorrectly Classified Instances	223	12.9051 %
Kappa statistic	0.7065	
Mean absolute error	0.1112	
Root mean squared error	0.2218	
Relative absolute error	48.5842 %	
Root relative squared error	65.5935 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.96	0.168	0.93	0.96	0.945	0.985	unacc
0.747	0.091	0.702	0.747	0.724	0.959	acc
0.538	0.001	0.946	0.538	0.686	0.999	vgood

	0.304	0.007	0.636	0.304	0.412	0.986	good
Weighted Avg.	0.871	0.138	0.868	0.871	0.865	0.979	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1162	46	0	2	a = unacc
87	287	0	10	b = acc
0	30	35	0	c = vgood
0	46	2	21	d = good

Hasil percobaan implementasi naïve bayes dan KNN :

- Data set yang digunakan “car.arff”
  1. Naïve Bayes
     a. Full Training
        Akurasi dari weka : 87.049 %
        Akurasi dari program : 87.384 %

```
=> Jumlah data yang match : 1510
=> Jumlah yang data tidak match : 218
=> Akurasi : 87.38425925925925
========================================================================
BUILD SUCCESSFUL (total time: 4 seconds)
```

     b. 10 fold cross validation
        Akurasi dari weka : 85.532 %
        Akurasi dari program : 82.648 %

```
Accuracy uji ke -1 : 66.86046511627907
Accuracy uji ke -2 : 84.88372093023256
Accuracy uji ke -3 : 83.13953488372093
Accuracy uji ke -4 : 94.76744186046511
Accuracy uji ke -5 : 93.02325581395348
Accuracy uji ke -6 : 86.62790697674419
Accuracy uji ke -7 : 90.11627906976744
Accuracy uji ke -8 : 69.18604651162791
Accuracy uji ke -9 : 77.32558139534885
Accuracy uji ke -10 : 80.55555555555556
Accuracy rata rata : 82.6485788113695
BUILD SUCCESSFUL (total time: 4 seconds)
```

2. KNN (2NN)
   a. Full Training
      Akurasi weka : 96.29 %
      Akurasi program : 89.47 %

```
=> Jumlah data yang match : 1546
=> Jumlah yang data tidak match : 182
=> Akurasi : 89.4675925925926
========================================================================
BUILD SUCCESSFUL (total time: 3 seconds)
```

   b. 10 fold cross validation :
      Akurasi weka : 93.58 %
      Akurasi program :

```
Accuracy uji ke -1 : 60.46511627906976
Accuracy uji ke -2 : 55.23255813953488
Accuracy uji ke -3 : 56.395348837209305
Accuracy uji ke -4 : 72.67441860465115
Accuracy uji ke -5 : 54.06976744186046
Accuracy uji ke -6 : 54.06976744186046
Accuracy uji ke -7 : 86.04651162790698
Accuracy uji ke -8 : 79.06976744186046
Accuracy uji ke -9 : 76.74418604651163
Accuracy uji ke -10 : 75.5813953488372
Accuracy rata rata : 74.48320413436693
BUILD SUCCESSFUL (total time: 2 seconds)
```

KESIMPULAN

Akurasi untuk KNN pada weka cukup berbeda jauh dengan implementasi yang kami lakukan. Hal ini mungkin terjadi karena pada KNN terdapat aturan dimana jika terdapat nilai yang sama, maka yang digunakan adalah data paling awal. Pada program yang kami buat, data yang dipilih belum tentu data paling atas. Karena, dalam proses pencarian kelas, kami memasukkan ke dalam tabel untuk menampung hasil penentuan jarak secara terurut. Jadi jika ada data baru dengan jarak yang sama dengan data sebelumnya, maka data yang lama akan ditimpa. Oleh karena itu, kedepannya akan dilakukan perbaikan untuk implementasi KNN.

Sedangkan pada algoritma Naïve Bayes untuk klasifikasi 10 fold, hasil yang didapat juga berbeda . Hal ini mungkin terjadi, karena saat pembagian data, jika jumlah data ketika di bagi 10 bersisa, sisa data tersebut akan dimasukkan ke data set 10. Sedangkan pada weka, jika ada data yang lebih, maka data tersebut akan disebar secara merata ke dataset yang lain.