

**LAPORAN UTS**  
**PRAKTIKUM DATA WAREHOUSING DAN DATA MINING**



**Oleh:**

**NAMA : Daffa Putra Alwansyah**  
**NIM : L200190031**  
**KELAS : B**  
**PRODI : INFORMATIKA**

**Fakultas Komunikasi dan Informatika**  
**Universitas Muhammadiyah Surakarta**

- Gunakan tool Weka dan RapidMiner untuk mendapatkan hasil prediksi keputusan pemilihan lokasi berdasarkan dataset di atas menggunakan algoritma Naïve Bayes! Bandingkan hasilnya menggunakan kedua tool tersebut! Lampirkan hasilnya!

**Weka.**

ARFF-Viewer - C:\Kuliah\SMT 5\Praktikum DWDM (Data Warehousing)\UTS\Coba\HasilPrediksi.arff

File Edit View

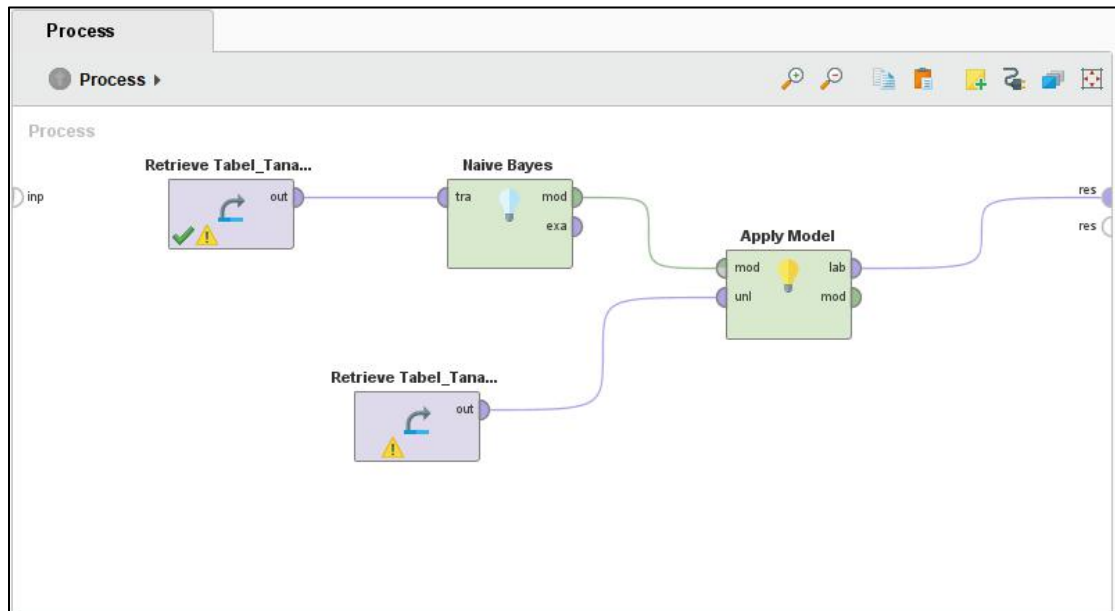
HasilPrediksi.arff \*

Relation: TanahTesting\_predicted

No. 1: Lokasi 2: Jarak Dari Pemukiman (meter) 3: Akses Jalan 4: Luas Tanah (m2) 5: Harga Tanah (juta) 6: PBB (juta) 7: prediction margin 8: predicted Keputusan Dipilih 9: Keputusan Dipilih

	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	A	600.0	mudah	10800.0	780.0	3.0	-0.138481	TIDAK	
2	B	603.0	mudah	10000.0	885.0	3.0	-0.603794	TIDAK	
3	C	711.0	sulit	11300.0	824.0	2.0	0.418453	YA	
4	D	801.0	sulit	10400.0	867.0	3.0	-0.263795	TIDAK	
5	E	611.0	mudah	10200.0	626.0	4.0	-0.492461	TIDAK	
6	F	899.0	sedang	12900.0	638.0	4.0	0.668007	YA	
7	G	777.0	sulit	11250.0	640.0	2.0	0.543797	YA	
8	H	612.0	mudah	10200.0	540.0	5.0	-0.515028	TIDAK	
9	I	851.0	sedang	12900.0	892.0	5.0	0.546236	YA	
10	J	757.0	sulit	11250.0	890.0	2.0	0.362369	YA	

**Rapid Miner.**



<new process> - RapidMiner Studio Educational 9.10.001 @ DESKTOP-4Q38BK1

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Find data, operators...etc. All Studio

Result History

ExampleSet (//Local Repository/Tabel\_Tanah\_Testing) ExampleSet (//Local Repository/Tabel\_Tanah) ExampleSet (Apply Model) ExampleSet (//Local Repository/Tabel\_Tanah\_Testing) ExampleSet (//Local Repository/Tabel\_Tanah\_Training)

Open in Turbo Prep Auto Model

Filter (10 / 10 examples): all

Row No.	prediction(K...	confidence[...	confidence[...	Lokasi	Jarak Dari P...	Akses Jalan	Luas Tanah ...	Harga Tana...	PBB (juta)
1	TIDAK	0.345	0.655	A	600	mudah	10800	780	3
2	TIDAK	0.154	0.846	B	603	mudah	10000	885	3
3	YA	0.606	0.394	C	711	sedang	11300	824	2
4	TIDAK	0.273	0.727	D	801	sedang	10400	867	3
5	TIDAK	0.199	0.801	E	611	mudah	10200	626	4
6	YA	0.770	0.230	F	899	sulit	12900	638	4
7	YA	0.679	0.321	G	777	sedang	11250	640	2
8	TIDAK	0.189	0.811	H	612	mudah	10200	540	5
9	YA	0.682	0.318	I	851	sulit	12900	892	5
10	YA	0.561	0.439	J	757	sedang	11250	890	2

ExampleSet (10 examples, 3 special attributes, 6 regular attributes)

**Kesimpulan:** hasil prediksi keputusan pemilihan lokasi pada kedua tool tersebut menghasilkan hasil prediksi yang sama dengan 5 Tidak & 5 Ya.

2. Berdasarkan Tabel Data Training pada soal no. 1, gantilah algoritma naïve bayes menjadi decision tree kemudian tuliskan aturan / rule (IF-THEN) berdasarkan pohon keputusan yang terbentuk!

**Weka.**

```
Classifier output

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Tanah
Instances:    25
Attributes:   7
              Lokasi
              Jarak Dari Pemukiman (meter)
              Akses Jalan
              Luas Tanah (m2)
              Harga Tanah (juta)
              PBB (juta)
              Keputusan Dipilih
Test mode:    evaluate on training data

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

J48 pruned tree
-----

PBB (juta) <= 2: YA (6.0)
PBB (juta) > 2
|   PBB (juta) <= 3: TIDAK (7.0/1.0)
|   PBB (juta) > 3: YA (12.0/4.0)

Number of Leaves :    3
Size of the tree :    5

Time taken to build model: 0 seconds
```

```
Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

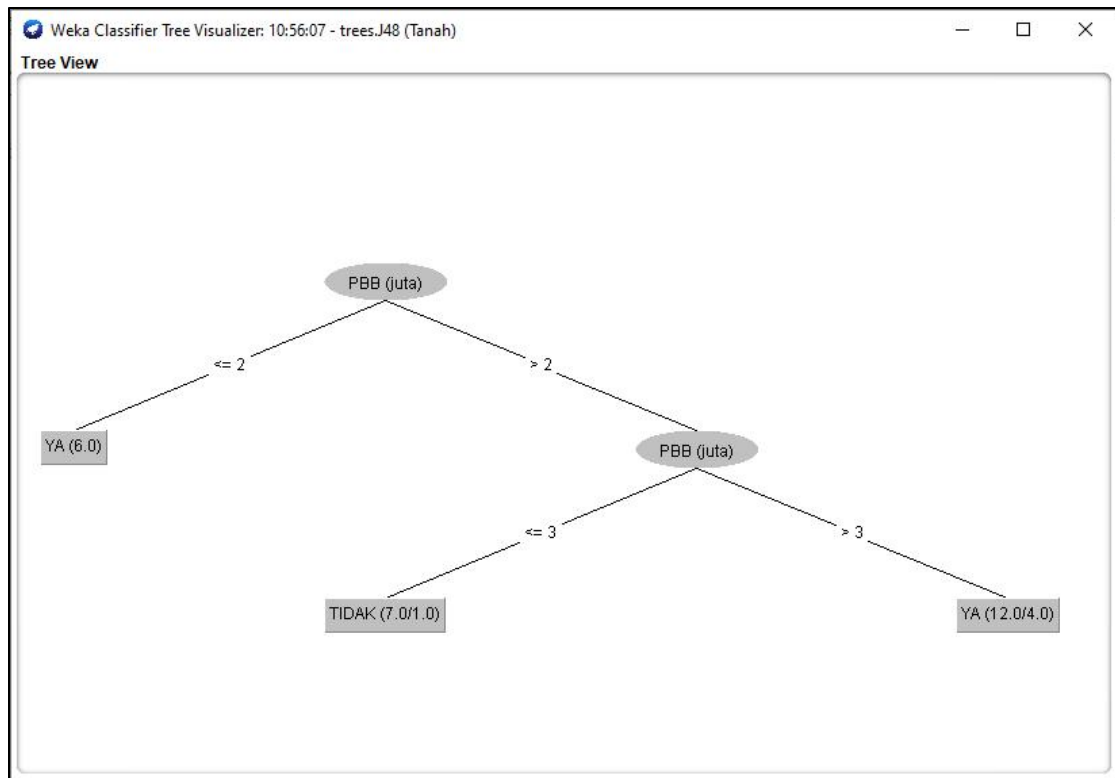
Correctly Classified Instances      20          80      %
Incorrectly Classified Instances     5          20      %
Kappa statistic                    0.5614
Mean absolute error                  0.2819
Root mean squared error              0.3754
Relative absolute error              58.5495 %
Root relative squared error          76.6268 %
Total Number of Instances           25

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              0.933    0.400    0.778      0.933    0.848      0.582    0.847     0.855     YA
              0.600    0.067    0.857      0.600    0.706      0.582    0.847     0.725     TIDAK
Weighted Avg.   0.800    0.267    0.810      0.800    0.791      0.582    0.847     0.803

=== Confusion Matrix ===

 a b  <-- classified as
14 1 | a = YA
 4 6 | b = TIDAK
```

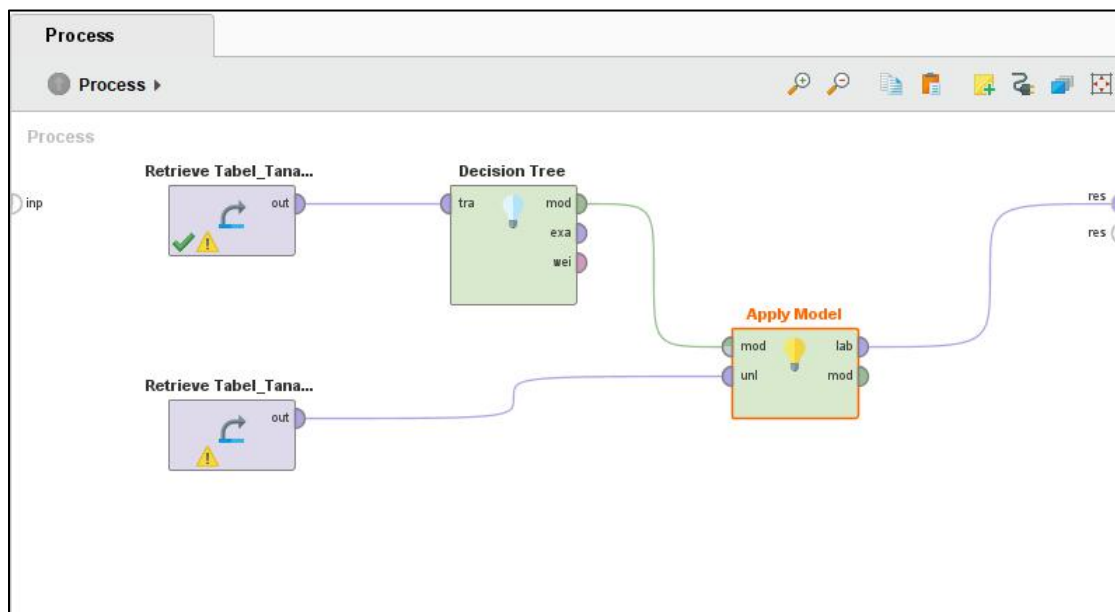


Berdasarkan pohon keputusan tersebut, dapat dilihat Keputusan sebagai berikut:

- Seseorang akan mempunyai keputusan YA jika kondisi seperti berikut :
  - PBB  $\leq 2$
  - PBB  $> 2$ , PBB  $> 3$
- Seseorang akan mempunyai keputusan TIDAK jika kondisi seperti berikut :
  - PBB  $> 2$ , PBB  $\leq 3$

### Rapid Miner.

Mengganti dengan Decision Tree pada no1.



<new process> - RapidMiner Studio Educational 9.10.001 @ DESKTOP-4Q38BK1

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

apply model X All Studio

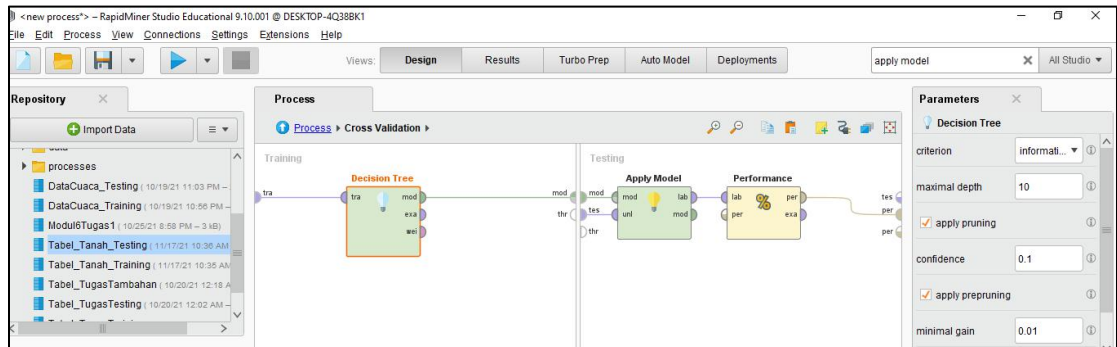
Result History

ExampleSet (Apply Model) X ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) X ExampleSet (/Local Repository/Tabel\_Tanah\_Training) X

Open in Turbo Prep Auto Model

Filter (10 / 10 examples): all

Row No.	prediction(K...	confidence(...	confidence(...	Lokasi	Jarak Dari P...	Akses Jalan	Luas Tanah ...	Harga Tana...	PBB (juta)
1	TIDAK	0	1	A	600	mudah	10800	780	3
2	TIDAK	0	1	B	603	mudah	10000	885	3
3	YA	1	0	C	711	sedang	11300	824	2
4	YA	0.500	0.500	D	801	sedang	10400	867	3
5	YA	0.800	0.200	E	611	mudah	10200	626	4
6	YA	1	0	F	899	sulit	12900	638	4
7	YA	1	0	G	777	sedang	11250	640	2
8	YA	0.500	0.500	H	612	mudah	10200	540	5
9	YA	1	0	I	851	sulit	12900	892	5
10	YA	1	0	J	757	sedang	11250	890	2



<new process> - RapidMiner Studio Educational 9.10.001 @ DESKTOP-4Q38BK1

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Find data, operators, etc. X All Studio

Result History

ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) X ExampleSet (/Local Repository/Tabel\_Tanah\_Training) X

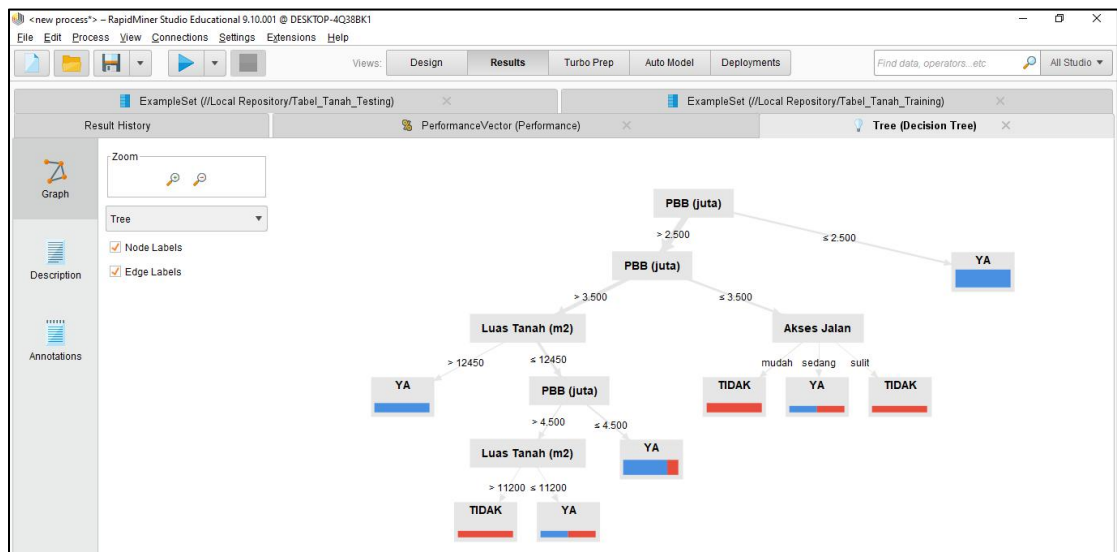
PerformanceVector (Performance)

Table View Plot View

Criterion: accuracy

accuracy: 43.33% +/- 25.09% (micro average: 48.00%)

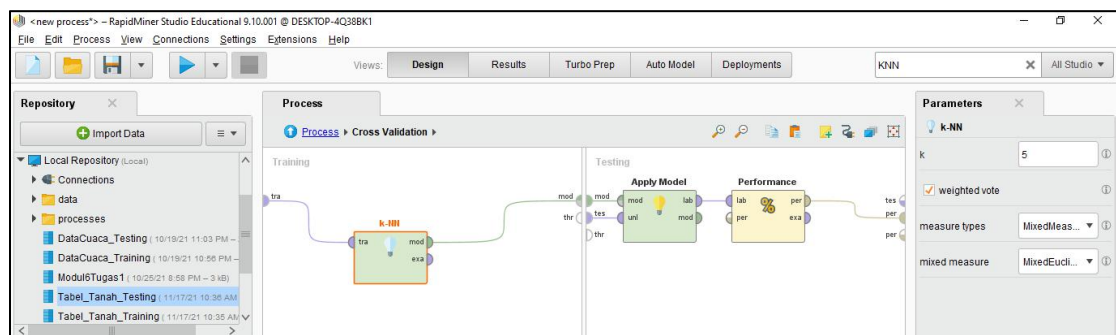
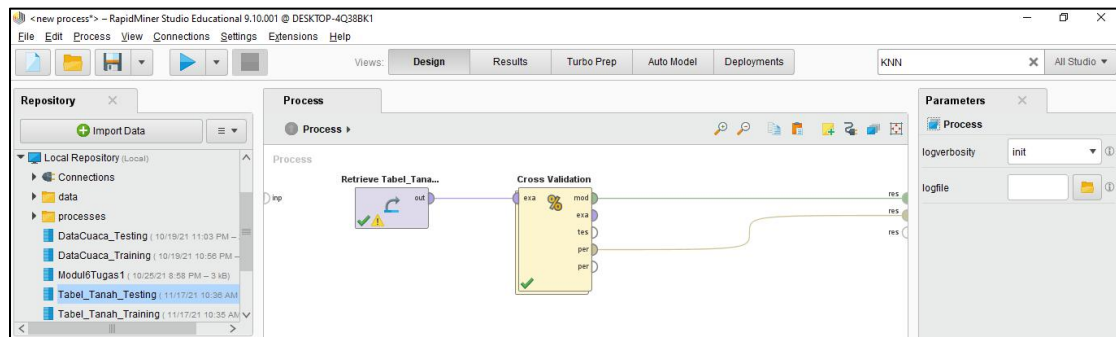
	true YA	true TIDAK	class precision
pred. YA	8	8	50.00%
pred. TIDAK	7	2	22.22%
class recall	53.33%	20.00%	



Berdasarkan pohon keputusan tersebut, dapat dilihat Keputusan sebagai berikut :  
sebagai berikut:

- PBB  $\leq$  2500
- PBB  $<$  2500, PBB  $<$  3500, Luas Tanah  $>$  12450
- PBB  $<$  12450, PBB  $<$  3500, Luas Tanah  $\leq$  12450, PBB  $>$  4500, Luas Tanah  $\leq$  11200
- PBB  $<$  12450, PBB  $<$  3500, Luas Tanah  $\leq$  12450, PBB  $\leq$  4500
- PBB  $>$  2500, PBB  $\leq$  3500, Akses Jalan (sedang)
- Keputusan dipilih (TIDAK) jika kondisi sebagai berikut :
  - PBB  $>$  2500, PBB  $\leq$  3500, Akses Jalan (mudah)
  - PBB  $>$  2500, PBB  $\leq$  3500, Akses Jalan (sulit)
  - PBB  $<$  12450, PBB  $<$  3500, Luas Tanah  $\leq$  12450, PBB  $>$  4500, Luas Tanah  $>$  11200

3. Masih menggunakan Tabel Data Training pada soal no. 1, ubahlah menjadi algoritma k-NN untuk mengolah dataset tersebut! Pada nilai k ke berapakah akurasi optimal didapatkan? Lampirkan hasilnya!



ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) x ExampleSet (/Local Repository/Tabel\_Tanah\_Training) x KNNClassification (k-KNN) x

Result History PerformanceVector (Performance) x

Criterion accuracy

Table View Plot View

accuracy: 68.33% +/- 24.15% (micro average: 68.00%)

	true YA	true TIDAK	class precision
pred. YA	13	6	68.42%
pred. TIDAK	2	4	66.67%
class recall	86.67%	40.00%	

ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) x ExampleSet (/Local Repository/Tabel\_Tanah\_Training) x

Result History PerformanceVector (Performance) x

Performance

PerformanceVector:

accuracy: 68.33% +/- 24.15% (micro average: 68.00%)

ConfusionMatrix:

True: YA TIDAK

YA: 13 6

TIDAK: 2 4

**Kesimpulan:** Nilai K = 5 adalah yang paling tinggi untuk k-NN dengan akurasi 68.33%

4. Berdasarkan hasil pekerjaan anda pada soal no. 1 dan no. 2, pada tools RapidMiner. Adakah perbedaan hasil prediksi antara naïve bayes VS decision tree? Tuliskan analisis anda!

## Naïve Bayes

ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) x ExampleSet (/Local Repository/Tabel\_Tanah\_Training) x

Result History ExampleSet (Apply Model) x ExampleSet (/Local Repository/Tabel\_Tanah\_Testing) x ExampleSet (/Local Repository/Tabel\_Tanah\_Training) x

Open in Turbo Prep Auto Model

Filter (10 / 10 examples): all

Row No.	prediction(K...	confidence(...	confidence(...	Lokasi	Jarak Dari P...	Akses Jalan	Luas Tanah ...	Harga Tana...	PBB (juta)
1	TIDAK	0.345	0.655	A	600	mudah	10800	780	3
2	TIDAK	0.154	0.846	B	603	mudah	10000	885	3
3	YA	0.606	0.394	C	711	sedang	11300	824	2
4	TIDAK	0.273	0.727	D	801	sedang	10400	867	3
5	TIDAK	0.199	0.801	E	611	mudah	10200	626	4
6	YA	0.770	0.230	F	899	sulit	12900	638	4
7	YA	0.679	0.321	G	777	sedang	11250	640	2
8	TIDAK	0.189	0.811	H	612	mudah	10200	540	5
9	YA	0.682	0.318	I	851	sulit	12900	892	5
10	YA	0.561	0.439	J	757	sedang	11250	890	2

ExampleSet (10 examples, 3 special attributes, 6 regular attributes)



## Decision Tree

Row No.	prediction(K...	confidence(...	confidence(...	Lokasi	Jarak Dari P...	Akses Jalan	Luas Tanah ...	Harga Tana...	PBB (juta)
1	TIDAK	0	1	A	600	mudah	10800	780	3
2	TIDAK	0	1	B	603	mudah	10000	885	3
3	YA	1	0	C	711	sedang	11300	824	2
4	YA	0.500	0.500	D	801	sedang	10400	867	3
5	YA	0.800	0.200	E	611	mudah	10200	626	4
6	YA	1	0	F	899	sulit	12900	638	4
7	YA	1	0	G	777	sedang	11250	640	2
8	YA	0.500	0.500	H	612	mudah	10200	540	5
9	YA	1	0	I	851	sulit	12900	892	5
10	YA	1	0	J	757	sedang	11250	890	2

**Kesimpulan:** Terdapat perbedaan pada prediksi

- Naïve Bayes, 5 Tidak & 5 Ya
- Decision Tree, 2 Tidak & 8 Ya

5. Buatlah perbandingan tingkat akurasi data menggunakan algoritma klasifikasi naïve bayes, decision tree dan k-NN pada dataset Tabel Data Training pada soal no. 1! Manakah yang memiliki hasil tertinggi? Lampirkan hasilnya!

## Naive bayes

Criterion	accuracy
accuracy	51.67% +/- 12.30% (micro average: 52.00%)

## Decision Tree

Criterion	accuracy
accuracy	43.33% +/- 25.09% (micro average: 40.00%)



## K-NN

The screenshot displays the RapidMiner Studio interface with the 'Results' tab selected. The main window shows the performance metrics for a K-NN model. The 'PerformanceVector (Performance)' view is active, displaying a table of results. The table includes columns for 'true YA', 'true TIDAK', and 'class precision'. The overall accuracy is 68.33% with a micro average of 68.00%.

	true YA	true TIDAK	class precision
pred. YA	13	6	68.42%
pred. TIDAK	2	4	66.67%
class recall	86.67%	40.00%	

**Kesimpulan:** K-NN merupakan akurasi tertinggi dengan rincian:

- Naive Bayes memiliki tingkat akurasi 51.67%
- Decision Tree memiliki tingkat akurasi 43.33%
- k-NN memiliki tingkat akurasi 68.33%