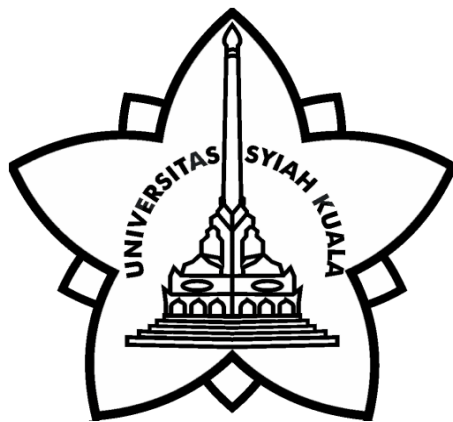


**LANGKAH – LANGKAH MENGUBAH HIMPUNAN
DATA WDBC, ABALONE DAN BANK MARKETING
SERTA MEMBAGINYA MENJADI DATA TRAINING
DAN TESTING**

disusun untuk memenuhi
tugas mata kuliah Data Mining

oleh :

Khairul Umam Albi
2008107010072



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

A. Dataset WDBC

- Tools yang digunakan adalah WEKA explorer

1. Langkah-langkah untuk memisahkan data WDBC

- Buka WEKA Explorer dan buka file WDBC.arff.
- Kemudian pada bagian filter pilih Randomize pada package weka => filters => unsupervised=>instance => Randomize.
- Tanpa melakukan konfigurasi tambahan maka dataset tersebut telah teracak.
- Selanjutnya Kembali ke bagian filter pilih RemovePercentage pada package weka => filters =>unsupervised => instance => RemovePercentage.
- Pada bagian tersebut mengantikan presentase menjadi 80.0 untuk data testing maka data terhapus sebanyak 80 persen.
- Selanjutnya tekan apply dan simpan data tersebut dengan menekan save as di WEKA menjadi wdbc-testing.arff.
- Kembali ke halaman utama weka dan klik undo, selanjutnya ganti nilai invertSelection menjadi True, dengan percentage masih sama maka data tersimpan sebanyak 80 persen.

2. Uraikan hasil klasifikasi yang diperoleh.

Proses klasifikasi menggunakan Classifier IBk (Lazy/KNN) dengan k=3, k=5, k=7, k=9,dan k=11 serta pembobotan dalam voting penentuan class adalah distanceWeighting =Weight by 1/distance.

a. Supplied test set WDBC-testing.arff pada k = 3

```
=== Summary ===
Correctly Classified Instances      110          96.4912 %
Incorrectly Classified Instances     4           3.5088 %
Kappa statistic                     0.923
Mean absolute error                  0.0562
Root mean squared error              0.1738
Relative absolute error              12.1191 %
Root relative squared error          36.3571 %
Total Number of Instances           114

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC       ROC Area  PRC Area  Class
               0.950    0.027    0.950     0.950    0.950     0.923     0.983     0.975     M
               0.973    0.050    0.973     0.973    0.973     0.923     0.983     0.982     B
Weighted Avg.   0.965    0.042    0.965     0.965    0.965     0.923     0.983     0.980

=== Confusion Matrix ===
  a  b  <-- classified as
38  2  |  a = M
 2 72  |  b = B
```

Precision bernilai 0.965, Recall bernilai 0.965, dan F-Measure bernilai 0.965

b. Supplied test set WDBC-testing.arff pada $k = 5$

```

=== Summary ===

Correctly Classified Instances      111          97.3684 %
Incorrectly Classified Instances     3           2.6316 %
Kappa statistic                    0.9426
Mean absolute error                 0.0495
Root mean squared error             0.1402
Relative absolute error             10.6772 %
Root relative squared error         29.3291 %
Total Number of Instances          114

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.975    0.027    0.951     0.975    0.963      0.943    0.997    0.993     M
                0.973    0.025    0.986     0.973    0.980      0.943    0.997    0.998     B
Weighted Avg.   0.974    0.026    0.974     0.974    0.974      0.943    0.997    0.996

=== Confusion Matrix ===

  a  b  <-- classified as
39  1  |  a = M
 2 72  |  b = B

```

Precision bernilai 0.974, Recall bernilai 0.974, dan F-Measure bernilai 0.974

c. Supplied test set WDBC-testing.arff pada $k = 7$

```

=== Summary ===

Correctly Classified Instances      113          99.1228 %
Incorrectly Classified Instances     1           0.8772 %
Kappa statistic                    0.9809
Mean absolute error                 0.0529
Root mean squared error             0.1311
Relative absolute error             11.4089 %
Root relative squared error         27.4322 %
Total Number of Instances          114

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1.000    0.014    0.976     1.000    0.988      0.981    0.999    0.998     M
                0.986    0.000    1.000     0.986    0.993      0.981    0.999    0.999     B
Weighted Avg.   0.991    0.005    0.991     0.991    0.991      0.981    0.999    0.999

=== Confusion Matrix ===

  a  b  <-- classified as
40  0  |  a = M
 1 73  |  b = B

```

Precision bernilai 0.991, Recall bernilai 0.991, dan F-Measure bernilai 0.991.

d. Supplied test set WDBC-testing.arff pada k = 9

```

Correctly Classified Instances      113          99.1228 %
Incorrectly Classified Instances    1           0.8772 %
Kappa statistic                    0.9809
Mean absolute error                0.0616
Root mean squared error            0.1385
Relative absolute error            13.286 %
Root relative squared error        28.9701 %
Total Number of Instances         114

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1.000    0.014    0.976     1.000    0.988      0.981    0.999    0.998     M
                0.986    0.000    1.000     0.986    0.993      0.981    0.999    0.999     B
Weighted Avg.   0.991    0.005    0.991     0.991    0.991      0.981    0.999    0.999

=== Confusion Matrix ===

  a  b  <-- classified as
40  0  |  a = M
 1 73 |  b = B

```

Precision bernilai 0.991, Recall bernilai 0.991, dan F-Measure bernilai 0.991.

e. Supplied test set WDBC-testing.arff pada k = 11

```

=== Summary ===

Correctly Classified Instances      112          98.2456 %
Incorrectly Classified Instances    2           1.7544 %
Kappa statistic                    0.961
Mean absolute error                0.068
Root mean squared error            0.1473
Relative absolute error            14.6526 %
Root relative squared error        30.8048 %
Total Number of Instances         114

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.950    0.000    1.000     0.950    0.974      0.962    0.999    0.998     M
                1.000    0.050    0.974     1.000    0.987      0.962    0.999    0.999     B
Weighted Avg.   0.982    0.032    0.983     0.982    0.982      0.962    0.999    0.999

=== Confusion Matrix ===

  a  b  <-- classified as
38  2  |  a = M
 0 74 |  b = B

```

Precision bernilai 0.982, Recall bernilai 0.982, dan F-Measure bernilai 0.982.

f. Cross-validation dengan file WDBC.arff pada k = 3

```

=== Summary ===

Correctly Classified Instances      551          96.8366 %
Incorrectly Classified Instances    18           3.1634 %
Kappa statistic                    0.9318
Mean absolute error                0.0487
Root mean squared error            0.1711
Relative absolute error            10.4128 %
Root relative squared error        35.3883 %
Total Number of Instances         569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.939    0.014    0.975     0.939    0.957      0.932    0.980    0.973     M
                0.986    0.061    0.964     0.986    0.975      0.932    0.980    0.978     B
Weighted Avg.   0.968    0.044    0.969     0.968    0.968      0.932    0.980    0.976

=== Confusion Matrix ===

  a  b  <-- classified as
199 13 |  a = M
 5 352 |  b = B

```

Precision bernilai 0.969, Recall bernilai 0.968, dan F-Measure bernilai 0.968.

g. Cross-validation dengan file WDBC.arff pada k = 5

```

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

Correctly Classified Instances      552          97.0123 %
Incorrectly Classified Instances    17           2.9877 %
Kappa statistic                    0.9355
Mean absolute error                0.0503
Root mean squared error            0.1621
Relative absolute error            10.7568 %
Root relative squared error        33.5346 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.939   0.011   0.980     0.939   0.959      0.936   0.987    0.982     M
                0.989   0.061   0.964     0.989   0.976      0.936   0.987    0.987     B
Weighted Avg.   0.970   0.043   0.970     0.970   0.970      0.936   0.987    0.985

=== Confusion Matrix ===

  a  b  <-- classified as
199 13 |  a = M
  4 353 |  b = B

```

Precision bernilai 0.970, Recall bernilai 0.970, dan F-Measure bernilai 0.970

h. Cross-validation dengan file WDBC.arff pada k = 7

```

=== Summary ===

Correctly Classified Instances      553          97.188 %
Incorrectly Classified Instances    16           2.812 %
Kappa statistic                    0.9394
Mean absolute error                0.0568
Root mean squared error            0.1639
Relative absolute error            12.1383 %
Root relative squared error        33.9003 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.943   0.011   0.980     0.943   0.962      0.940   0.989    0.983     M
                0.989   0.057   0.967     0.989   0.978      0.940   0.989    0.989     B
Weighted Avg.   0.972   0.040   0.972     0.972   0.972      0.940   0.989    0.987

=== Confusion Matrix ===

  a  b  <-- classified as
200 12 |  a = M
  4 353 |  b = B

```

Precision bernilai 0.927, Recall bernilai 0.972, dan F-Measure bernilai 0.972

i. Cross-validation dengan file WDBC.arff pada k = 9

```

=== Summary ===

Correctly Classified Instances      553          97.188 %
Incorrectly Classified Instances    16           2.812 %
Kappa statistic                    0.9392
Mean absolute error                0.0611
Root mean squared error            0.1643
Relative absolute error            13.073 %
Root relative squared error        33.9814 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.934   0.006   0.990     0.934   0.961      0.940   0.991    0.989     M
                0.994   0.066   0.962     0.994   0.978      0.940   0.991    0.991     B
Weighted Avg.   0.972   0.044   0.972     0.972   0.972      0.940   0.991    0.990

=== Confusion Matrix ===

  a  b  <-- classified as
198 14 |  a = M
  2 355 |  b = B

```

Precision bernilai 0.927, Recall bernilai 0.972, dan F-Measure bernilai 0.972

j. Cross-validation dengan file WDBC.arff pada k = 11

```

=== Summary ===

Correctly Classified Instances      550          96.6608 %
Incorrectly Classified Instances    19           3.3392 %
Kappa statistic                    0.9278
Mean absolute error                 0.067
Root mean squared error             0.1722
Relative absolute error             14.3168 %
Root relative squared error         35.6207 %
Total Number of Instances          569

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0.929    0.011    0.980     0.929    0.954     0.929    0.990    0.987     M
                0.989    0.071    0.959     0.989    0.974     0.929    0.990    0.990     B
Weighted Avg.   0.967    0.049    0.967     0.967    0.966     0.929    0.990    0.989

=== Confusion Matrix ===

  a    b  <-- classified as
197  15 |  a = M
  4 353 |  b = B

```

Precision bernilai 0.967, Recall bernilai 0.967 dan F-Measure bernilai 0.966

3. Kesimpulan

Precision terbesar yang di dapat dari uji coba adalah pada supplied test set dengan k = 7 dan k = 9 berada pada angka 0.991. Recall terbesar yang di dapat dari uji coba adalah pada supplied test set dengan k = 7 dan k = 9 berada pada angka 0.991. F-Measure terbesar yang di dapat dari uji coba adalah pada supplied test set dengan k = 7 dan k = 9 berada pada angka 0.991. Pada kasus dataset WDBC, menggunakan ke-5 k menunjukkan pengujian supplied test dapat menghasilkan akurasi yang lebih baik daripada cross-validation, selain itu dari ke-5 k menunjukkan penggunaan k = 7 dan k = 9 lebih baik digunakan untuk dataset ini.

B. Abalone

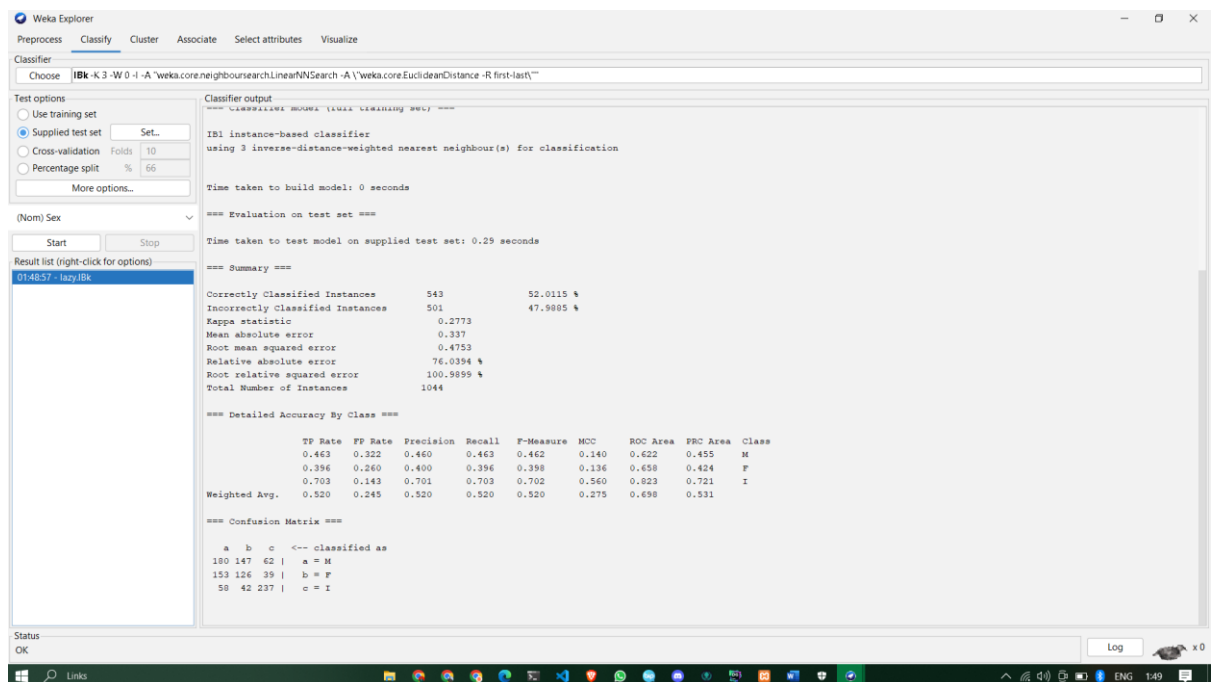
- Tools yang digunakan adalah WEKA explorer
1. Langkah-langkah untuk memisahkan data Abalone
 - Buka WEKA Explorer dan buka file Abalone.arff.
 - Kemudian pada bagian filter pilih Randomize pada package weka => filters => unsupervised=>instance => Randomize.
 - Tanpa melakukan konfigurasi tambahan maka dataset tersebut telah teracak.
 - Selanjutnya Kembali ke bagian filter pilih RemovePercentage pada package weka => filters =>unsupervised => instance => RemovePercentage.
 - Pada bagian tersebut mengantikan presentase menjadi 75.0 untuk data testing maka data terhapus sebanyak 75 persen.
 - Selanjutnya tekan apply dan simpan data tersebut dengan menekan save as di WEKA menjadi Abalone-testing.arff.

- Kembali ke halaman utama weka dan klik undo, selanjutnya ganti nilai invertSelection menjadi True, dengan percentage masih sama maka data tersimpan sebanyak 75 persen.

2. Uraikan hasil klasifikasi yang diperoleh.

Proses klasifikasi menggunakan Classifier IBk (Lazy/KNN) dengan $k=3$, $k=5$, $k=7$, $k=9$, dan $k=11$ serta pembobotan dalam voting penentuan class adalah distanceWeighting = Weight by $1/\text{distance}$.

a. Supplied test set Abalone-testing.arff pada $k = 3$



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'IBk - K 3 - W 0 - I - A' with the distance weighting set to 'weka.core.neighboursearch.LinearNNSearch -A' and the distance metric set to 'weka.core.EuclideanDistance -R first-last'. The test options are set to 'Supplied test set' with a 'Set...' button. The evaluation results are displayed in the 'Classifier output' pane.

Classifier output

```
==== CLASSIFIER MODEL (Lazy Learning Set) ====
IBk instance-based classifier
using 3 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===
Time taken to test model on supplied test set: 0.29 seconds

=== Summary ===
Correctly Classified Instances      543      52.0115 %
Incorrectly Classified Instances    501      47.9885 %
Kappa statistic                     0.2773
Mean absolute error                 0.337
Root mean squared error             0.4753
Relative absolute error             76.0394 %
Root relative squared error         100.9899 %
Total Number of Instances          1044

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.463    0.322    0.460     0.463    0.462     0.140    0.622    0.455    M
      0.396    0.260    0.400     0.396    0.398     0.136    0.658    0.424    F
      0.703    0.143    0.701     0.703    0.702     0.560    0.823    0.721    I
Weighted Avg.    0.520    0.245    0.520     0.520    0.520     0.275    0.698    0.531

=== Confusion Matrix ===
      a  b  c  <-- classified as
180 147  62 | a = M
153 126  39 | b = F
 58  42 237 | c = I
```

b. Supplied test set Abalone-testing.arff pada $k = 5$

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **IBk -K 5 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options:
☐ Use training set
☒ Supplied test set (Set...)
☐ Cross-validation (Folds: 10)
☐ Percentage split (%: 66)
 More options...

(Nom) Sex
 Start Stop
 Result list (right-click for options):
 014857 - lazyJbk
 014936 - lazyJbk

Classifier output
 IBk instance-based classifier
 using 5 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 Time taken to test model on supplied test set: 0.25 seconds
 Evaluation on test set: 0.25 seconds
 Summary:
 Correctly Classified Instances: 553 (52.9693 %)
 Incorrectly Classified Instances: 491 (47.0307 %)
 Kappa statistic: 0.2931
 Mean absolute error: 0.3402
 Root mean squared error: 0.4522
 Relative absolute error: 76.7569 %
 Root relative squared error: 56.0855 %
 Total Number of Instances: 1044
 Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
0.450	0.295	0.476	0.450	0.462	0.157	0.623	0.456	M
0.418	0.264	0.409	0.418	0.414	0.153	0.670	0.433	F
0.727	0.150	0.698	0.727	0.712	0.571	0.837	0.738	I
Weighted Avg.	0.530	0.239	0.527	0.530	0.528	0.289	0.706	

 Confusion Matrix:

a	b	c	<-- classified as
175	150	64	a = M
143	133	42	b = F
50	42	245	c = I

c. Supplied test set Abalone-testing.arff pada k = 7

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **IBk -K 7 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options:
☐ Use training set
☒ Supplied test set (Set...)
☐ Cross-validation (Folds: 10)
☐ Percentage split (%: 66)
 More options...

(Nom) Sex
 Start Stop
 Result list (right-click for options):
 014857 - lazyJbk
 014936 - lazyJbk

Classifier output
 IBk instance-based classifier
 using 7 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 Time taken to test model on supplied test set: 0.24 seconds
 Evaluation on test set: 0.24 seconds
 Summary:
 Correctly Classified Instances: 567 (54.3103 %)
 Incorrectly Classified Instances: 477 (45.6897 %)
 Kappa statistic: 0.3116
 Mean absolute error: 0.3387
 Root mean squared error: 0.4395
 Relative absolute error: 76.4197 %
 Root relative squared error: 53.3862 %
 Total Number of Instances: 1044
 Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
0.476	0.313	0.474	0.476	0.475	0.163	0.633	0.467	M
0.403	0.227	0.437	0.403	0.419	0.180	0.684	0.447	F
0.754	0.151	0.704	0.754	0.728	0.592	0.851	0.752	I
Weighted Avg.	0.543	0.235	0.537	0.543	0.540	0.306	0.719	

 Confusion Matrix:

a	b	c	<-- classified as
185	135	69	a = M
152	128	38	b = F
53	30	254	c = I

d. Supplied test set Abalone-testing.arff pada k = 9

Classifier
Choose: IBk - K 9 - W 0 - I - A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options
☐ Use training set
☒ Supplied test set (Set...)
☐ Cross-validation (Folds: 10)
☐ Percentage split (%: 66)
 More options...

(Nom) Sex

Result list (right-click for options)
 014857 - lazyJbk
 014936 - lazyJbk
 014946 - lazyJbk
 015000 - lazyJbk (selected)
 015020 - lazyJbk

Classifier output
 === CLASSIFIER MODEL (lazy learning set) ===
 IB1 instance-based classifier
 using 9 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 === Evaluation on test set ===
 Time taken to test model on supplied test set: 0.17 seconds
 === Summary ===
 Correctly Classified Instances 554 53.0651 %
 Incorrectly Classified Instances 490 46.9349 %
 Kappa statistic 0.2936
 Mean absolute error 0.3385
 Root mean squared error 0.4329
 Relative absolute error 76.372 %
 Root relative squared error 51.9903 %
 Total Number of Instances 1044
 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
	0.445	0.313	0.458	0.445	0.451	0.133	0.643	M
	0.396	0.242	0.417	0.396	0.406	0.156	0.694	F
	0.757	0.154	0.701	0.757	0.728	0.591	0.862	I
Weighted Avg.	0.531	0.240	0.524	0.531	0.527	0.288	0.729	

 === Confusion Matrix ===

a	b	c	<-- classified as
173	146	70	a = M
153	126	35	b = F
52	20	255	c = I

e. Supplied test set Abalone-testing.arff pada k = 11

Classifier
Choose: IBk - K 11 - W 0 - I - A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options
☐ Use training set
☒ Supplied test set (Set...)
☐ Cross-validation (Folds: 10)
☐ Percentage split (%: 66)
 More options...

(Nom) Sex

Result list (right-click for options)
 014857 - lazyJbk
 014936 - lazyJbk
 014946 - lazyJbk
 015000 - lazyJbk
 015020 - lazyJbk (selected)

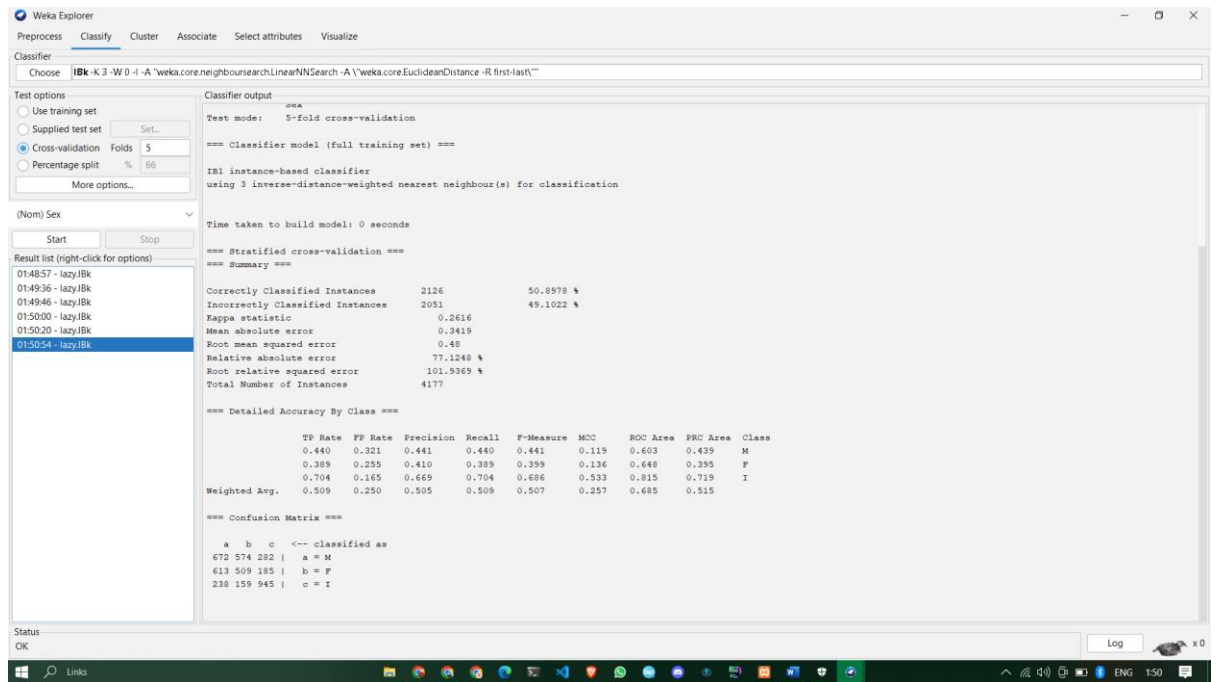
Classifier output
 === CLASSIFIER MODEL (lazy learning set) ===
 IB1 instance-based classifier
 using 11 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 === Evaluation on test set ===
 Time taken to test model on supplied test set: 0.14 seconds
 === Summary ===
 Correctly Classified Instances 565 54.5019 %
 Incorrectly Classified Instances 475 45.4981 %
 Kappa statistic 0.3144
 Mean absolute error 0.3376
 Root mean squared error 0.4284
 Relative absolute error 76.1703 %
 Root relative squared error 51.028 %
 Total Number of Instances 1044
 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
	0.478	0.311	0.477	0.478	0.478	0.167	0.649	M
	0.399	0.219	0.444	0.399	0.421	0.186	0.704	F
	0.760	0.158	0.696	0.760	0.726	0.588	0.866	I
Weighted Avg.	0.545	0.234	0.538	0.545	0.540	0.309	0.736	

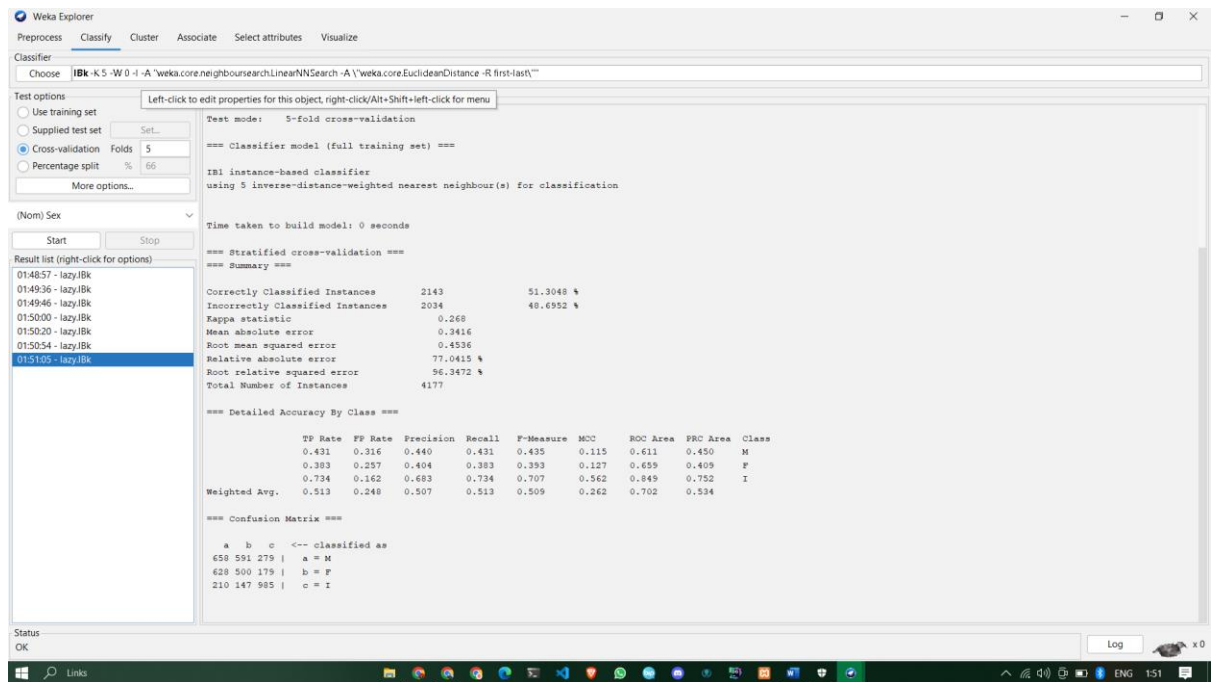
 === Confusion Matrix ===

a	b	c	<-- classified as
186	130	73	a = M
152	127	35	b = F
52	29	256	c = I

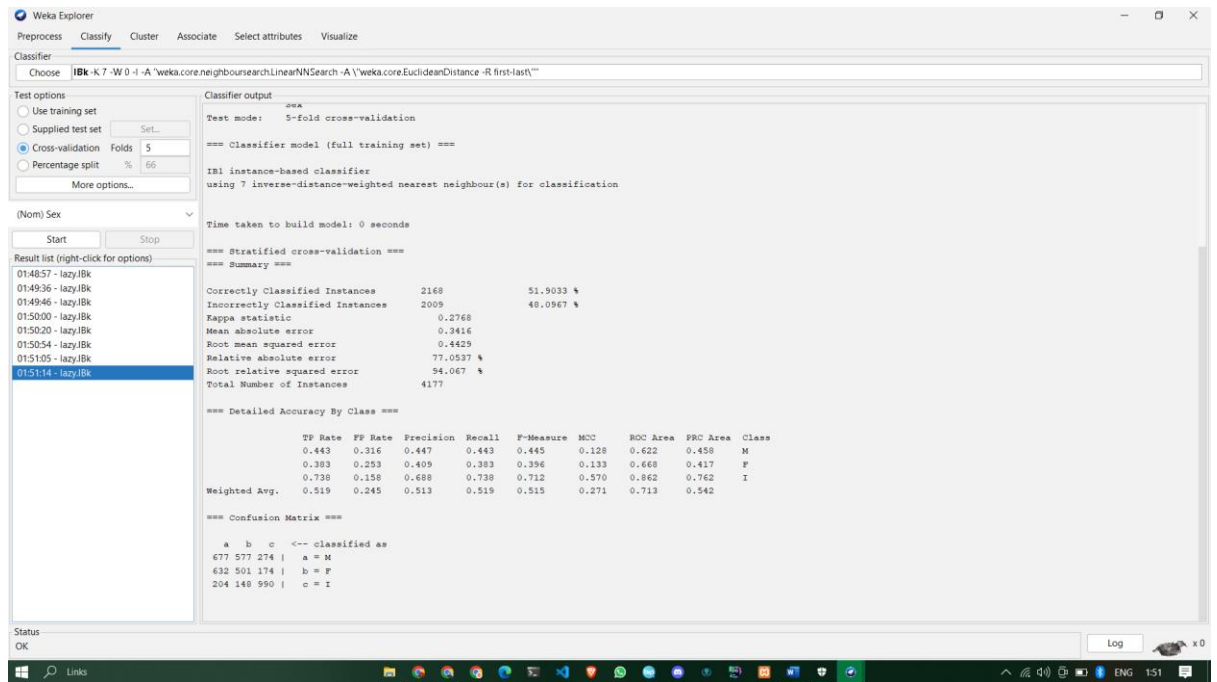
f. Cross-validation dengan file Abalone.arff pada k = 3



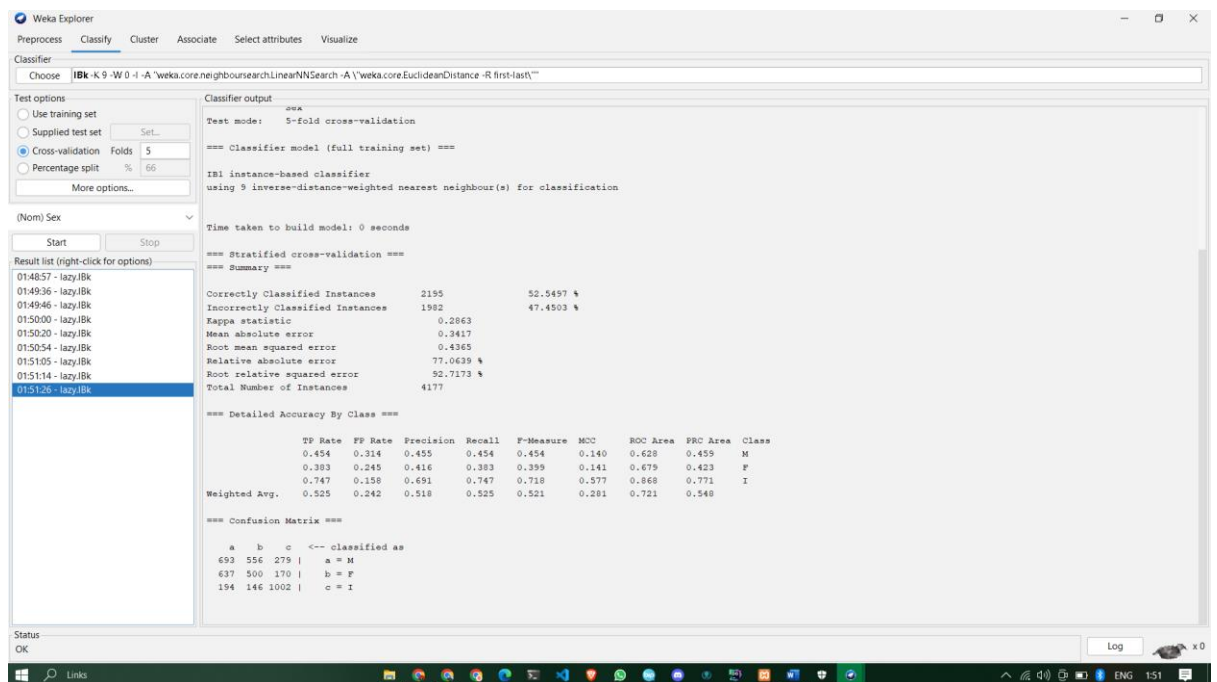
g. Cross-validation dengan file Abalone.arff pada k = 5



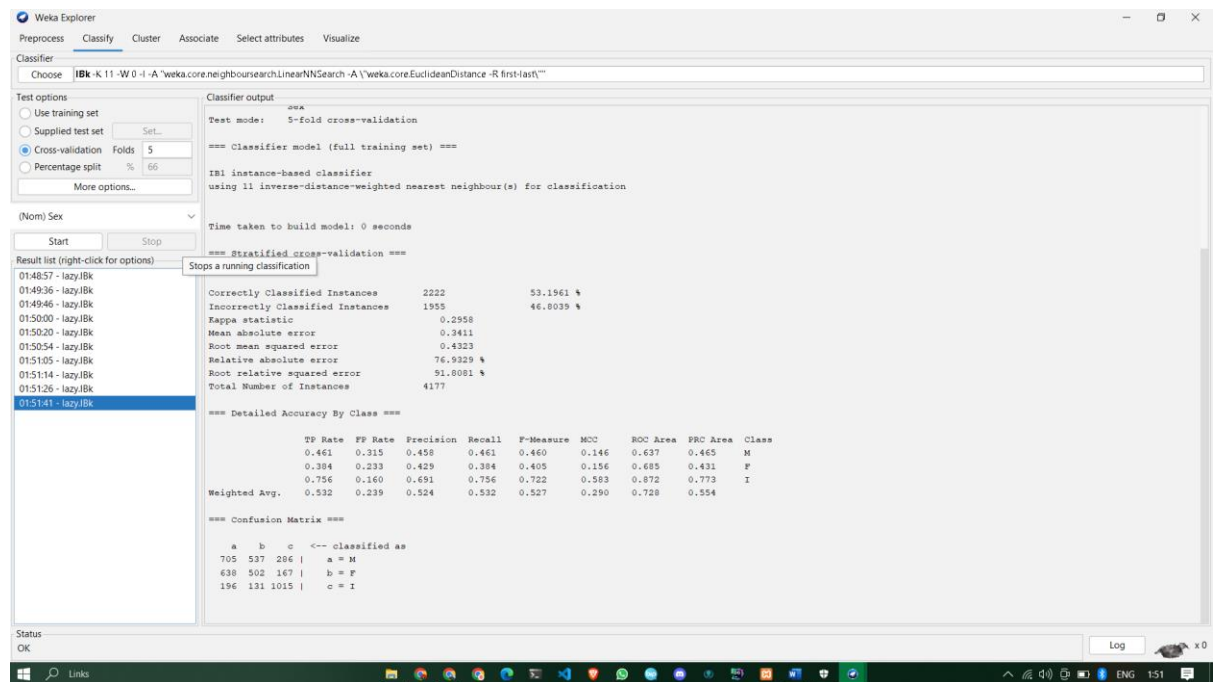
h. Cross-validation dengan file Abalone.arff pada k = 7



i. Cross-validation dengan file Abalone.arff pada k = 9



j. Cross-validation dengan file Abalone.arff pada k = 11



3. Kesimpulan.

Precision terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 7$ berada pada angka 0.537. Recall terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 11$ berada pada angka 0.545. F-Measure terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 7$ dan $k = 11$ berada pada angka 0.540. Pada kasus dataset Abalone, menggunakan ke-5 k menunjukkan pengujian supplied test dapat menghasilkan akurasi yang lebih baik daripada cross-validation, selain itu dari ke-5 k menunjukkan penggunaan $k = 7$ dan $k = 11$ lebih baik digunakan untuk dataset ini.

C. Bank Data

1. Langkah-langkah untuk memisahkan data Bank.

- Buka WEKA Explorer dan buka file Bank.arff.
- Kemudian pada bagian filter pilih Randomize pada package weka => filters => unsupervised=>instance => Randomize.
- Tanpa melakukan konfigurasi tambahan maka dataset tersebut telah teracak.
- Selanjutnya Kembali ke bagian filter pilih RemovePercentage pada package weka => filters =>unsupervised => instance => RemovePercentage.
- Pada bagian tersebut mengantikan presentase menjadi 75.0 untuk data testing maka data terhapus sebanyak 75 persen.
- Selanjutnya tekan apply dan simpan data tersebut dengan menekan save as di WEKA menjadi bank-testing.arff.

- Kembali ke halaman utama weka dan klik undo, selanjutnya ganti nilai invertSelection menjadi True, dengan percentage masih sama maka data tersimpan sebanyak 75 persen.

2. Uraikan hasil klasifikasi yang diperoleh.

Proses klasifikasi menggunakan Classifier IBk (Lazy/KNN) dengan $k=3$, $k=5$, $k=7$, $k=9$, dan $k=11$ serta pembobotan dalam voting penentuan class adalah distanceWeighting = Weight by $1/\text{distance}$.

a. Supplied test set bank-testing.arff pada $k = 3$

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'IBk - K 3 - W 0 - I - A' with the distance weighting set to 'weka.core.neighboursearch.LinearNNSearch - A' and the distance metric set to 'weka.core.EuclideanDistance - R first-last'. The test options are set to 'Supplied test set' with a percentage split of 66%. The classifier output shows that the model was built successfully and evaluated on the test set. The evaluation results are as follows:

Time taken to build model: 0 seconds
 Time taken to test model on supplied test set: 0.3 seconds

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	984	87.0796 %
Incorrectly Classified Instances	146	12.9204 %
Kappa statistic	0.1328	
Mean absolute error	0.1527	
Root mean squared error	0.329	
Relative absolute error	75.024 %	
Root relative squared error	103.468 %	
Total Number of Instances	1130	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.132	0.034	0.333	0.132	0.189	0.150	0.706	0.240	yes
	0.966	0.868	0.896	0.966	0.930	0.150	0.706	0.945	no

=== Confusion Matrix ===

	a	b	<-- classified as
17 112	a = yes		
34 967	b = no		

b. Supplied test set bank-testing.arff pada $k = 5$

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 5 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds 5

☐ Percentage split % 66

More options...

(Nom) y

Start Stop

Result list (right-click for options)

01:48:57 - lazy.IBk

01:49:36 - lazy.IBk

01:49:46 - lazy.IBk

01:50:00 - lazy.IBk

01:50:20 - lazy.IBk

01:50:54 - lazy.IBk

01:51:05 - lazy.IBk

01:51:14 - lazy.IBk

01:51:26 - lazy.IBk

01:51:41 - lazy.IBk

01:55:39 - lazy.IBk

01:56:14 - lazy.IBk

01:56:46 - lazy.IBk

01:57:19 - lazy.IBk

01:57:51 - lazy.IBk

01:58:39 - lazy.IBk

01:58:54 - lazy.IBk

01:59:06 - lazy.IBk

01:59:20 - lazy.IBk

01:59:34 - lazy.IBk

02:01:22 - lazy.IBk

02:01:40 - lazy.IBk

Classifier output

test model: user supplied test set: size unknown (loading incrementally)

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

IBk instance-based classifier
using 5 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.32 seconds

=== Summary ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Correctly Classified Instances	988				87.4336 %				yes
Incorrectly Classified Instances	142				12.5664 %				no
Kappa statistic			0.0714						
Mean absolute error			0.1553						
Root mean squared error			0.3175						
Relative absolute error			76.3196 %						
Root relative squared error			99.0357 %						
Total Number of Instances	1130								

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.070	0.022	0.290	0.070	0.113	0.093	0.712	0.253	yes
	0.978	0.930	0.891	0.978	0.932	0.093	0.712	0.944	no

=== Confusion Matrix ===

	a	b	<-- classified as
5 120	a = yes		
22 979	b = no		

Status OK

c. Supplied test set bank-testing.arff pada k = 7

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 7 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"

Test options

☐ Use training set

☒ Supplied test set Set...

☐ Cross-validation Folds 5

☐ Percentage split % 66

More options...

(Nom) y

Start Stop

Result list (right-click for options)

01:48:57 - lazy.IBk

01:49:36 - lazy.IBk

01:49:46 - lazy.IBk

01:50:00 - lazy.IBk

01:50:20 - lazy.IBk

01:50:54 - lazy.IBk

01:51:05 - lazy.IBk

01:51:14 - lazy.IBk

01:51:26 - lazy.IBk

01:51:41 - lazy.IBk

01:55:39 - lazy.IBk

01:56:14 - lazy.IBk

01:56:46 - lazy.IBk

01:57:19 - lazy.IBk

01:57:51 - lazy.IBk

01:58:39 - lazy.IBk

01:58:54 - lazy.IBk

01:59:06 - lazy.IBk

01:59:20 - lazy.IBk

01:59:34 - lazy.IBk

02:01:22 - lazy.IBk

02:01:40 - lazy.IBk

02:01:51 - lazy.IBk

Classifier output

test model: user supplied test set: size unknown (loading incrementally)

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

IBk instance-based classifier
using 7 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.23 seconds

=== Summary ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Correctly Classified Instances	990				87.6106 %				yes
Incorrectly Classified Instances	140				12.3894 %				no
Kappa statistic			0.0358						
Mean absolute error			0.1571						
Root mean squared error			0.3114						
Relative absolute error			77.1757 %						
Root relative squared error			97.913 %						
Total Number of Instances	1130								

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.035	0.016	0.238	0.035	0.067	0.054	0.723	0.260	yes
	0.984	0.961	0.888	0.984	0.934	0.054	0.723	0.946	no

=== Confusion Matrix ===

	a	b	<-- classified as
5 124	a = yes		
16 985	b = no		

Status OK

d. Supplied test set bank-testing.arff pada k = 9

Classifier output

```

test model: user supplied test set: size unknown (loading incrementally)
=== Classifier model (full training set) ===

IB1 instance-based classifier
using 9 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.28 seconds

=== Summary ===

Correctly Classified Instances      993      87.8761 %
Incorrectly Classified Instances    137      12.1239 %
Kappa statistic                    0.8201
Mean absolute error                 0.1584
Root mean squared error             0.3101
Relative absolute error              77.8287 %
Root relative squared error         97.5123 %
Total Number of Instances          1130

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.023   0.011   0.214   0.023   0.042   0.035   0.720   0.258   yes
0.989   0.977   0.887   0.989   0.935   0.935   0.720   0.944   no
Weighted Avg.   0.879   0.866   0.810   0.875   0.833   0.035   0.720   0.866

=== Confusion Matrix ===

  a  b  <-- classified as
3 126 | a = yes
11 990 | b = no

```

e. Supplied test set bank -testing.arff pada k = 11

Classifier output

```

test model: user supplied test set: size unknown (loading incrementally)
=== Classifier model (full training set) ===

IB1 instance-based classifier
using 11 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.34 seconds

=== Summary ===

Correctly Classified Instances      997      88.2301 %
Incorrectly Classified Instances    133      11.7699 %
Kappa statistic                    0.8272
Mean absolute error                 0.1592
Root mean squared error             0.3094
Relative absolute error              78.2102 %
Root relative squared error         97.3015 %
Total Number of Instances          1130

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.023   0.007   0.300   0.023   0.043   0.055   0.724   0.255   yes
0.993   0.977   0.888   0.993   0.937   0.955   0.724   0.946   no
Weighted Avg.   0.882   0.866   0.820   0.882   0.835   0.055   0.724   0.867

=== Confusion Matrix ===

  a  b  <-- classified as
3 126 | a = yes
7 994 | b = no

```

f. Cross-validation dengan file bank.arff pada k = 3

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **IBk** -K 3 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **5**
☐ Percentage split % 66
 More options...

(Nom) y

Start Stop

Result list (right-click for options):

- 01:49:36 - lazyJbk
- 01:49:46 - lazyJbk
- 01:50:00 - lazyJbk
- 01:50:20 - lazyJbk
- 01:50:34 - lazyJbk
- 01:51:05 - lazyJbk
- 01:51:14 - lazyJbk
- 01:51:26 - lazyJbk
- 01:51:41 - lazyJbk
- 01:55:39 - lazyJbk
- 01:56:14 - lazyJbk
- 01:56:46 - lazyJbk
- 01:57:19 - lazyJbk
- 01:58:39 - lazyJbk
- 01:58:54 - lazyJbk
- 01:59:06 - lazyJbk
- 01:59:20 - lazyJbk
- 01:59:34 - lazyJbk
- 02:01:22 - lazyJbk
- 02:01:40 - lazyJbk
- 02:01:51 - lazyJbk
- 02:02:01 - lazyJbk
- 02:02:10 - lazyJbk
- 02:02:35 - lazyJbk**

Status OK

Classifier output:

```

previous
postoutcome
y
Test mode: 5-fold cross-validation

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

IBk instance-based classifier
using 3 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      3966      87.724 %
Incorrectly Classified Instances    555      12.276 %
Kappa statistic                    0.1978
Mean absolute error                 0.1469
Root mean squared error             0.324
Relative absolute error             71.9886 %
Root relative squared error        101.4604 %
Total Number of Instances          4521

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.877   0.729   0.846    0.877   0.855    0.219   0.734    0.875    no

=== Confusion Matrix ===

  a    b  <-- classified as
54  427 |  a = yes
128 3872 |  b = no

```

g. Cross-validation dengan file bank.arff pada k = 5

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier: Choose **IBk** -K 5 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last""

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **5**
☐ Percentage split % 66
 More options...

(Nom) y

Start Stop

Result list (right-click for options):

- 01:49:46 - lazyJbk
- 01:50:00 - lazyJbk
- 01:50:20 - lazyJbk
- 01:50:34 - lazyJbk
- 01:51:05 - lazyJbk
- 01:51:14 - lazyJbk
- 01:51:26 - lazyJbk
- 01:51:41 - lazyJbk
- 01:55:39 - lazyJbk
- 01:56:14 - lazyJbk
- 01:56:46 - lazyJbk
- 01:57:19 - lazyJbk
- 01:57:51 - lazyJbk
- 01:58:39 - lazyJbk
- 01:58:54 - lazyJbk
- 01:59:06 - lazyJbk
- 01:59:20 - lazyJbk
- 01:59:34 - lazyJbk
- 02:01:22 - lazyJbk
- 02:01:40 - lazyJbk
- 02:01:51 - lazyJbk
- 02:02:01 - lazyJbk
- 02:02:10 - lazyJbk
- 02:02:35 - lazyJbk
- 02:02:47 - lazyJbk**

Status OK

Classifier output:

```

previous
postoutcome
y
Test mode: 5-fold cross-validation

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

IBk instance-based classifier
using 5 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      3987      88.188 %
Incorrectly Classified Instances    534      11.811 %
Kappa statistic                    0.1538
Mean absolute error                 0.1511
Root mean squared error             0.3137
Relative absolute error             74.0244 %
Root relative squared error        98.2402 %
Total Number of Instances          4521

=== Detailed Accuracy By Class ===

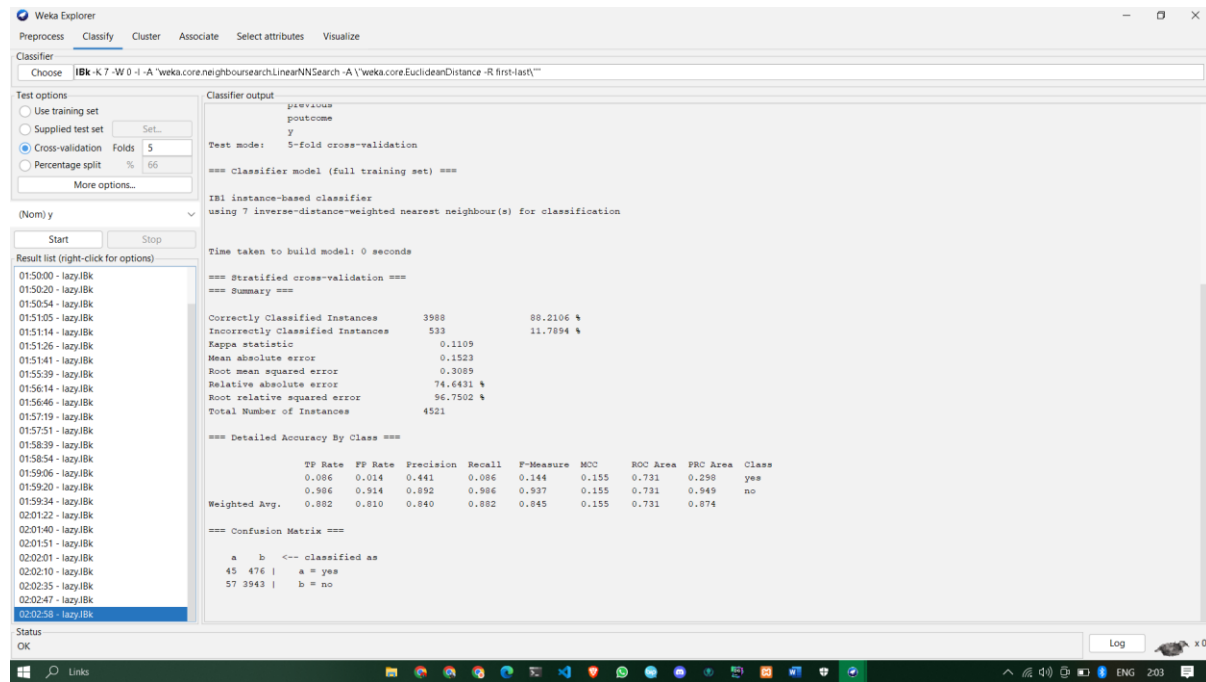
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.882   0.777   0.845    0.882   0.851    0.192   0.729    0.873    no

=== Confusion Matrix ===

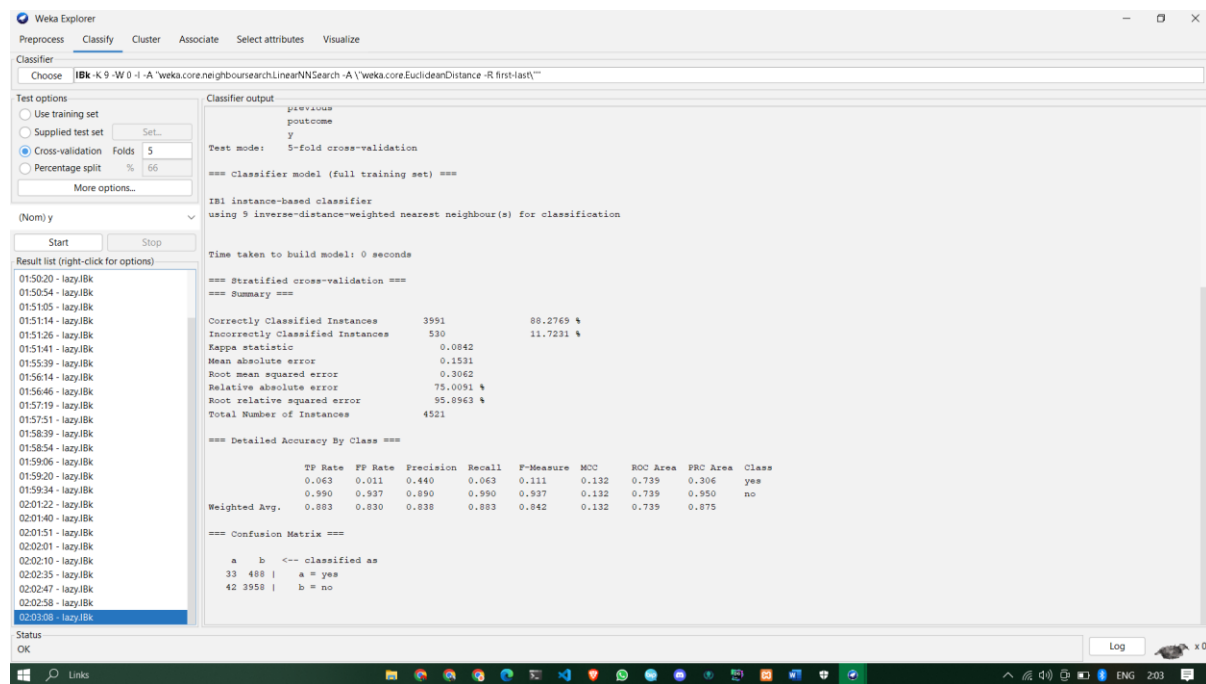
  a    b  <-- classified as
65  456 |  a = yes
78 3922 |  b = no

```

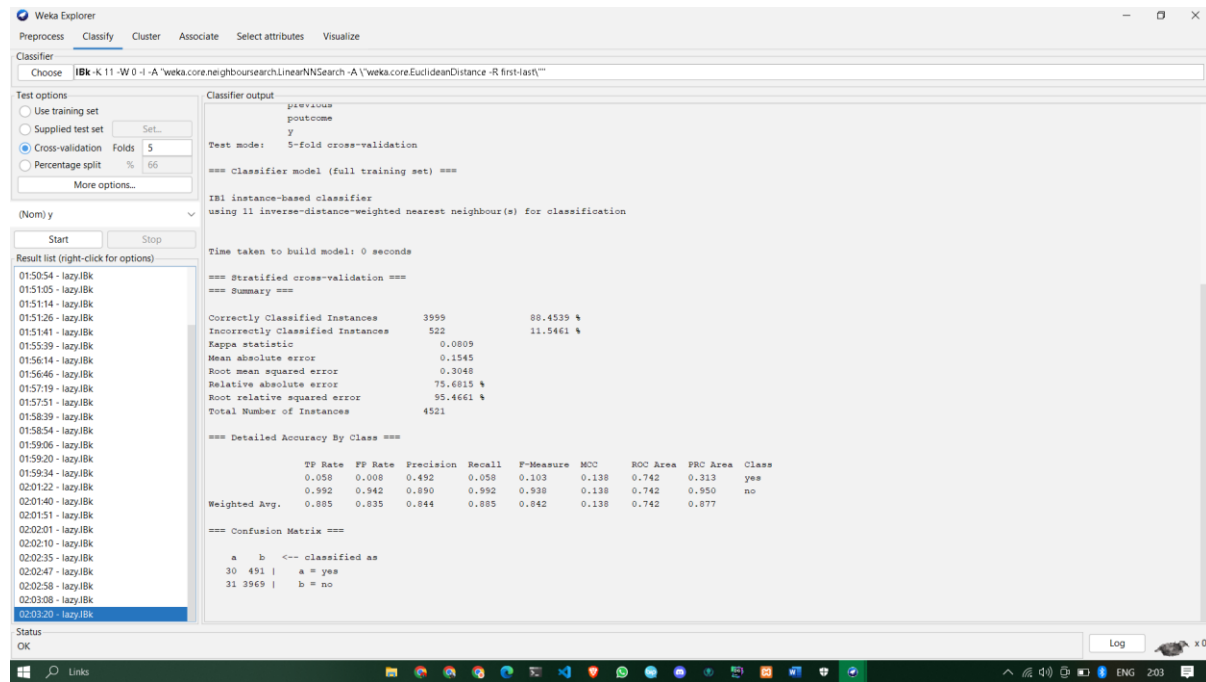
h. Cross-validation dengan file bank.arff pada k = 7



i. Cross-validation dengan file bank.arff pada k = 9



j. Cross-validation dengan file bank.arff pada k = 11



3. Kesimpulan.

Precision terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 3$ berada pada angka 0.832. Recall terbesar yang di dapat dari uji coba adalah pada cross-validation dengan $k = 11$ berada pada angka 0.885. F-Measure terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 3$ berada pada angka 0.845. Pada kasus dataset Bank Marketing, menggunakan ke-5 k menunjukkan hasil pengujian yang tidak memiliki selisih yang cukup besar baik pada pengujian supplied test maupun cross-validation. Nilai k yang digunakan pada supplied test lebih baik digunakan pada $k = 3$, lebih dari itu menghasilkan akurasi yang lebih rendah, namun pada cross-validation cenderung lebih baik menggunakan k yang lebih besar seperti 11.

D. Bank-full

1. Langkah-langkah untuk memisahkan data Bank-full.

- Buka WEKA Explorer dan buka file Bank-full.arff.
- Kemudian pada bagian filter pilih Randomize pada package weka => filters => unsupervised=>instance => Randomize.
- Tanpa melakukan konfigurasi tambahan maka dataset tersebut telah teracak.
- Selanjutnya Kembali ke bagian filter pilih RemovePercentage pada package weka => filters =>unsupervised => instance => RemovePercentage.
- Pada bagian tersebut mengantikan presentase menjadi 75.0 untuk data testing maka data terhapus sebanyak 75 persen.

- Selanjutnya tekan apply dan simpan data tersebut dengan menekan save as di WEKA menjadi bank-full-testing.arff.
- Kembali ke halaman utama weka dan klik undo, selanjutnya ganti nilai invertSelection menjadi True, dengan percentage masih sama maka data tersimpan sebanyak 75 persen.

2. Uraikan hasil klasifikasi yang diperoleh.

Proses klasifikasi menggunakan Classifier IBk (Lazy/KNN) dengan $k=3$, $k=5$, $k=7$, $k=9$, dan $k=11$ serta pembobotan dalam voting penentuan class adalah distanceWeighting = Weight by $1/\text{distance}$.

a. Supplied test set bank-full-testing.arff pada $k = 3$

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'IBk: K 3 -W 0.4 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"'. The test options are set to 'Supplied test set' with 'Set...' selected. The evaluation results are displayed on the right side of the window.

Classifier output
 test model: weka-supplied-test-set: size unknown (loading incrementally)
 === Classifier model (full training set) ===
 IBk instance-based classifier
 using 3 inverse-distance-weighted nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 20.64 seconds

=== Summary ===

Metric	Value
Correctly Classified Instances	9988
Incorrectly Classified Instances	1315
Kappa statistic	0.3173
Mean absolute error	0.1327
Root mean squared error	0.3088
Relative absolute error	64.2858 %
Root relative squared error	96.1711 %
Total Number of Instances	11303

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.804	0.621	0.865	0.884	0.871	0.330	0.804	0.898	no

=== Confusion Matrix ===

	a	b	<-- classified as
398	921		a = yes
394	9590		b = no

The result list on the left shows various time-stamped entries for the 'lazy:IBk' classifier, with the entry '02:04:09 - lazy:IBk' selected.

b. Supplied test set bank-full-testing.arff pada $k = 5$

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier
Choose **IBk -K 5 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options
☐ Use training set
☒ Supplied test set **Set...**
☐ Cross-validation Folds **5**
☐ Percentage split % **66**
 More options...

(Nom) y **y**

Start Stop

Result list (right-click for options)

- 01:51:14 - lazy.IBk
- 01:51:26 - lazy.IBk
- 01:51:41 - lazy.IBk
- 01:55:39 - lazy.IBk
- 01:56:14 - lazy.IBk
- 01:56:46 - lazy.IBk
- 01:57:19 - lazy.IBk
- 01:57:51 - lazy.IBk
- 01:58:39 - lazy.IBk
- 01:58:54 - lazy.IBk
- 01:59:06 - lazy.IBk
- 01:59:20 - lazy.IBk
- 01:59:34 - lazy.IBk
- 02:01:22 - lazy.IBk
- 02:01:40 - lazy.IBk
- 02:01:51 - lazy.IBk
- 02:02:01 - lazy.IBk
- 02:02:10 - lazy.IBk
- 02:02:35 - lazy.IBk
- 02:02:47 - lazy.IBk
- 02:02:58 - lazy.IBk
- 02:03:08 - lazy.IBk
- 02:03:20 - lazy.IBk
- 02:04:09 - lazy.IBk
- 02:04:44 - lazy.IBk**

Status
OK

Classifier output
 test model: user supplied test set: size unknown (loading incrementally)
 === Classifier model (full training set) ===
 IBk instance-based classifier
 using 5 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 === Evaluation on test set ===
 Time taken to test model on supplied test set: 21.21 seconds
 === Summary ===

Correctly Classified Instances	10075	89.1356 %
Incorrectly Classified Instances	1228	10.8644 %
Kappa statistic	0.3302	
Mean absolute error	0.1347	
Root mean squared error	0.2977	
Relative absolute error	65.2208 %	
Root relative squared error	92.7173 %	
Total Number of Instances	11303	

 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.891	0.632	0.872	0.891	0.875	0.353	0.811	0.902	no
	0.289	0.029	0.568	0.289	0.383	0.353	0.811	0.412	yes

 === Confusion Matrix ===

a	b	<-- classified as
381	938	a = yes
250	5694	b = no

c. Supplied test set bank-full-testing.arff pada k = 7

Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize

Classifier
Choose **IBk -K 7 -W 0 -I -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options
☐ Use training set
☒ Supplied test set **Set...**
☐ Cross-validation Folds **5**
☐ Percentage split % **66**
 More options...

(Nom) y **y**

Start Stop

Result list (right-click for options)

- 01:51:26 - lazy.IBk
- 01:51:41 - lazy.IBk
- 01:55:39 - lazy.IBk
- 01:56:14 - lazy.IBk
- 01:56:46 - lazy.IBk
- 01:57:19 - lazy.IBk
- 01:57:51 - lazy.IBk
- 01:58:39 - lazy.IBk
- 01:58:54 - lazy.IBk
- 01:59:06 - lazy.IBk
- 01:59:20 - lazy.IBk
- 01:59:34 - lazy.IBk
- 02:01:22 - lazy.IBk
- 02:01:40 - lazy.IBk
- 02:01:51 - lazy.IBk
- 02:02:01 - lazy.IBk
- 02:02:10 - lazy.IBk
- 02:02:35 - lazy.IBk
- 02:02:47 - lazy.IBk
- 02:02:58 - lazy.IBk
- 02:03:08 - lazy.IBk
- 02:03:20 - lazy.IBk
- 02:04:09 - lazy.IBk
- 02:04:44 - lazy.IBk
- 02:05:15 - lazy.IBk**

Status
OK

Classifier output
 test model: user supplied test set: size unknown (loading incrementally)
 === Classifier model (full training set) ===
 IBk instance-based classifier
 using 7 inverse-distance-weighted nearest neighbour(s) for classification
 Time taken to build model: 0 seconds
 === Evaluation on test set ===
 Time taken to test model on supplied test set: 21.55 seconds
 === Summary ===

Correctly Classified Instances	10066	89.056 %
Incorrectly Classified Instances	1237	10.944 %
Kappa statistic	0.31	
Mean absolute error	0.1362	
Root mean squared error	0.2926	
Relative absolute error	65.9458 %	
Root relative squared error	91.1406 %	
Total Number of Instances	11303	

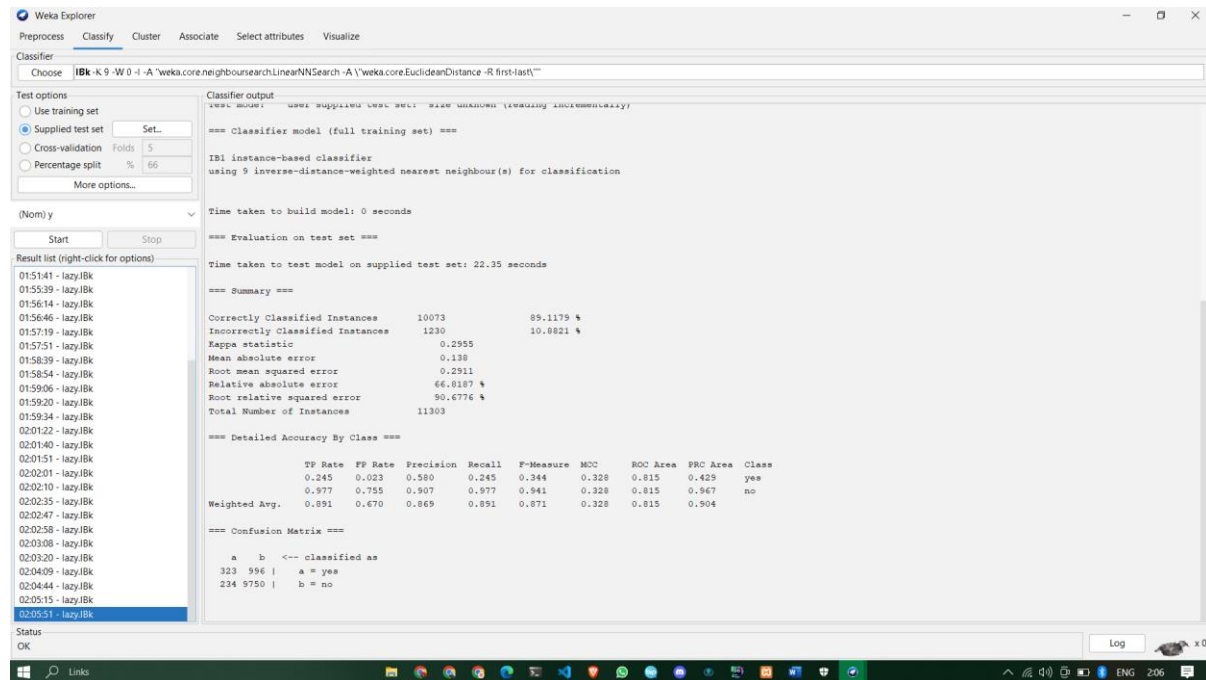
 === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.891	0.652	0.869	0.891	0.873	0.337	0.815	0.904	no
	0.265	0.027	0.566	0.265	0.361	0.337	0.815	0.426	yes

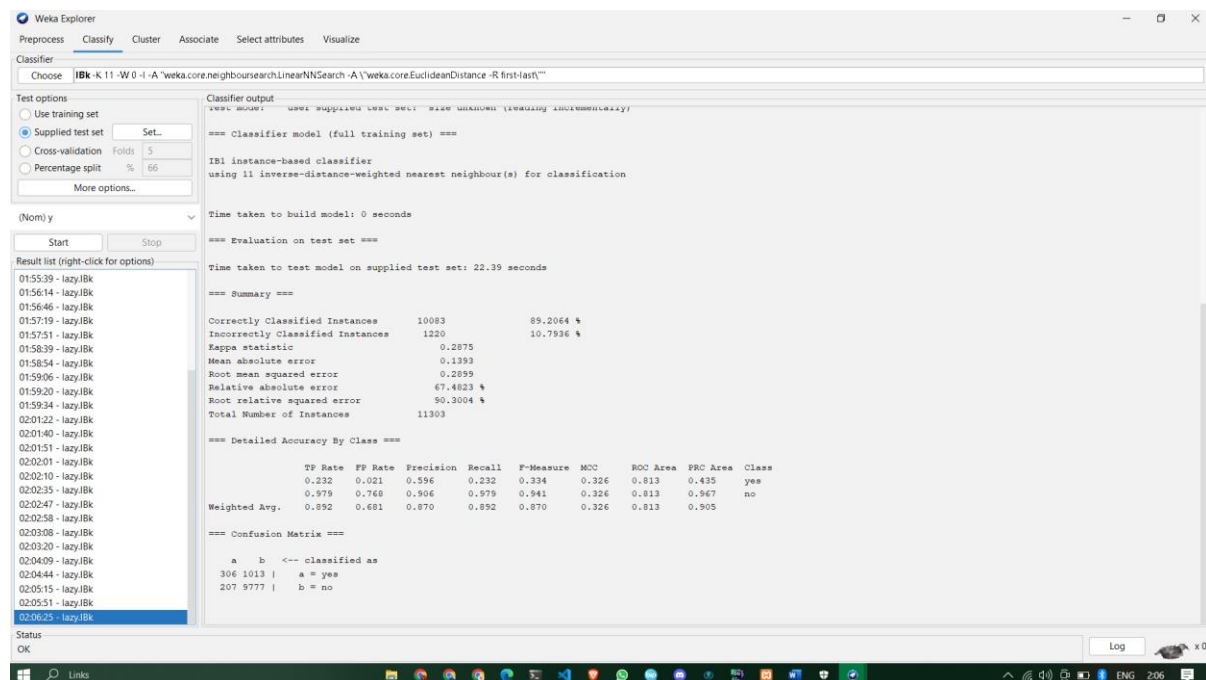
 === Confusion Matrix ===

a	b	<-- classified as
350	569	a = yes
268	5716	b = no

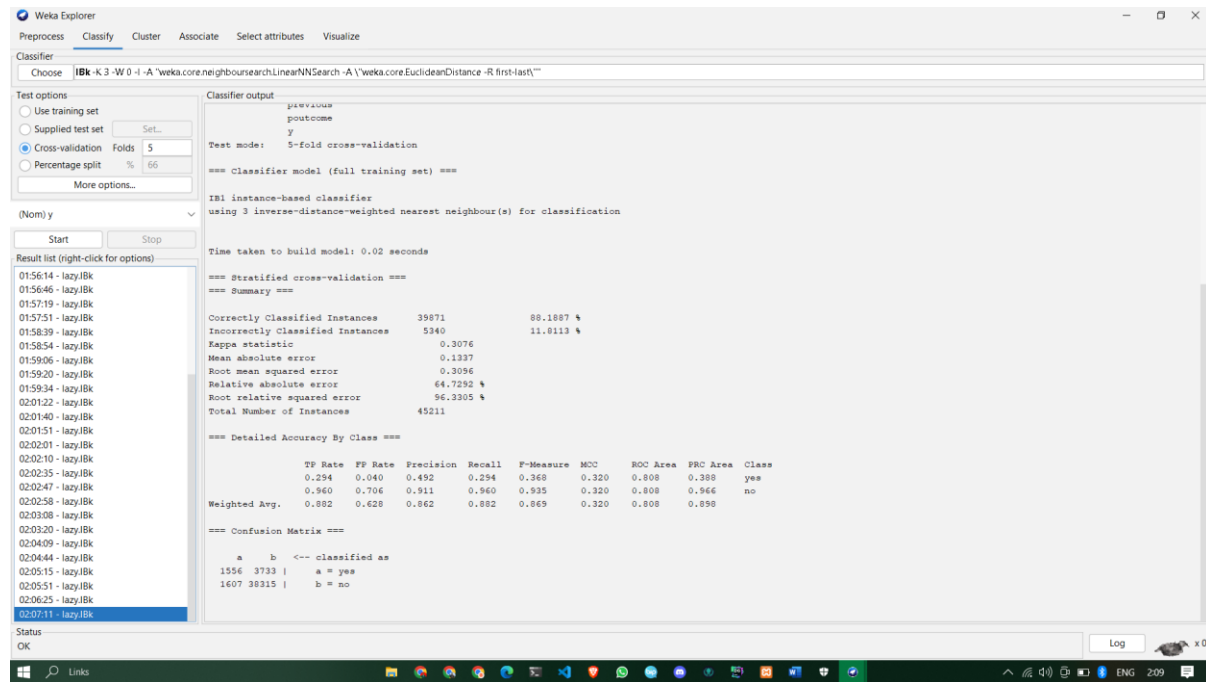
d. Supplied test set bank-full-testing.arff pada k = 9



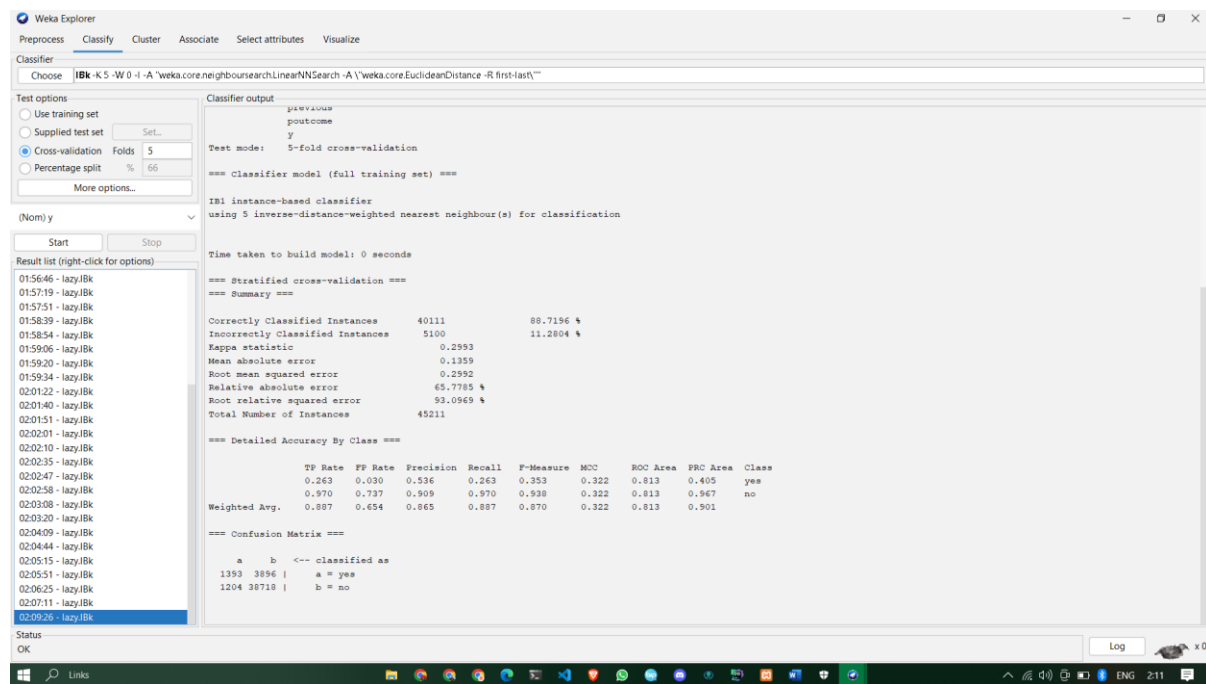
e. Supplied test set bank-full-testing.arff pada k = 11



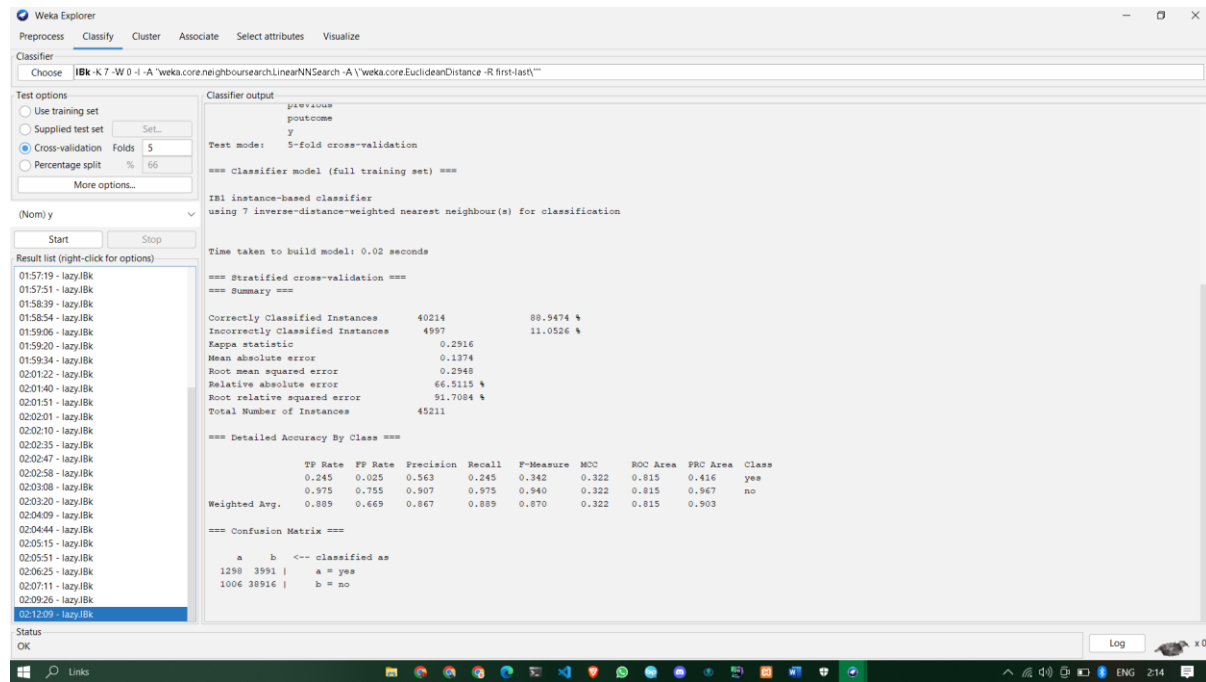
f. Cross-validation dengan file bank-full.arff pada k = 3



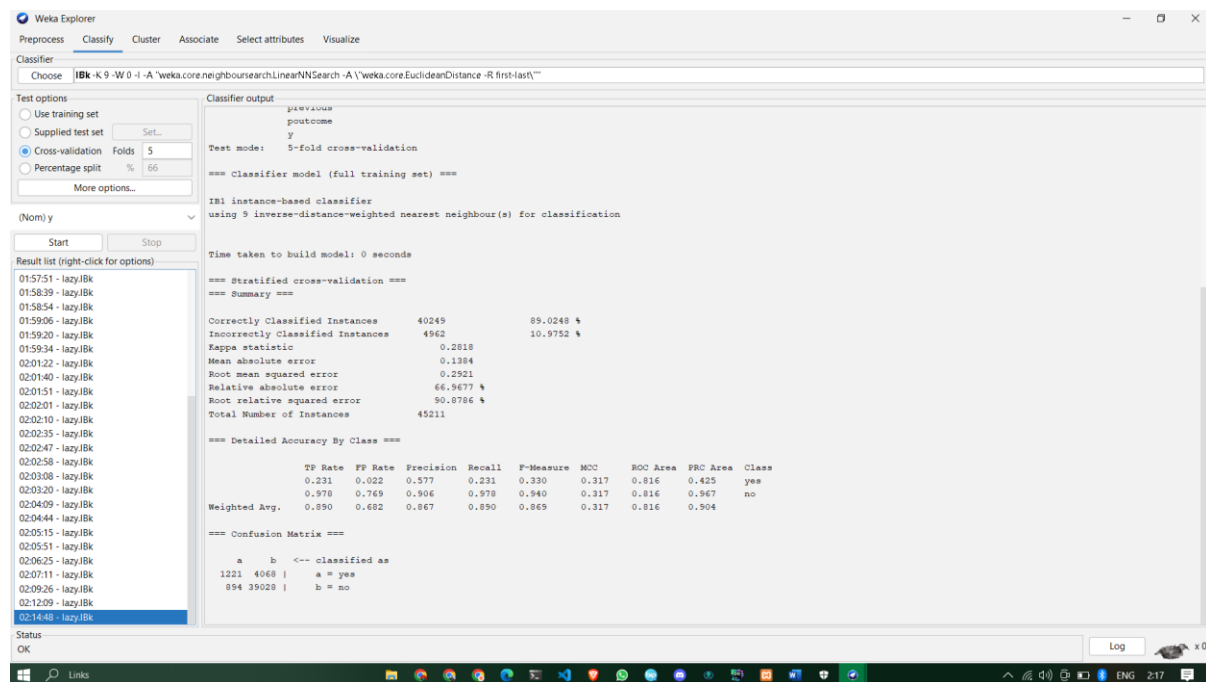
g. Cross-validation dengan file bank-full.arff pada k = 5



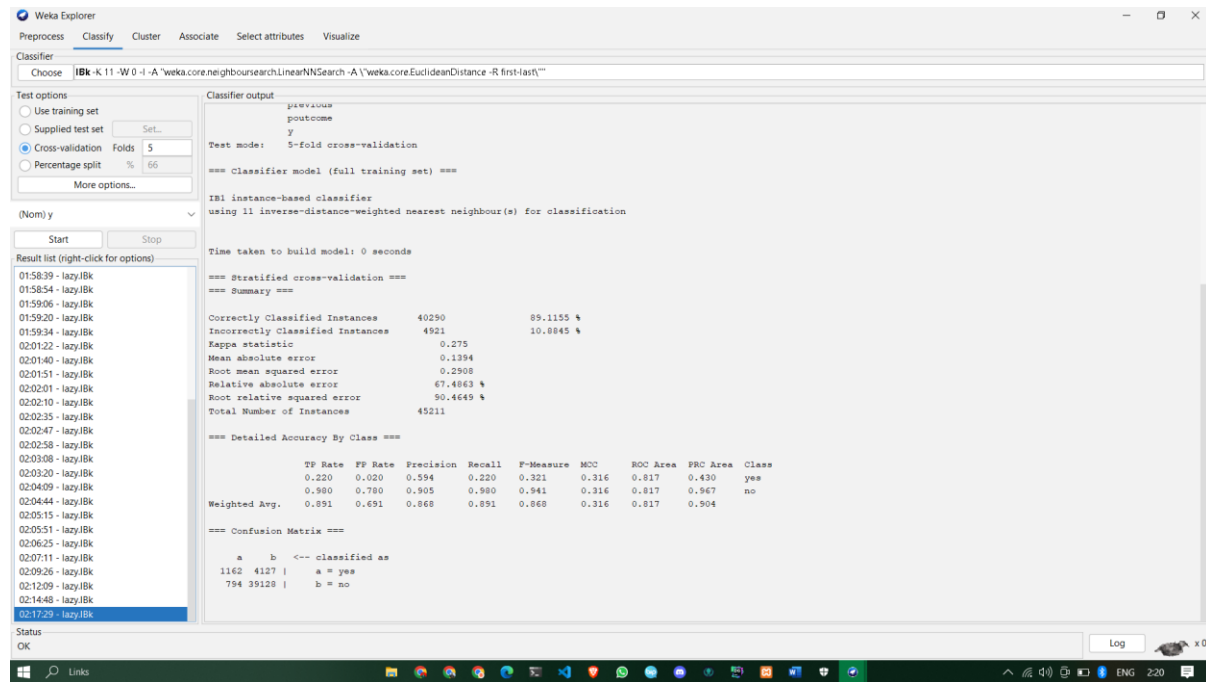
h. Cross-validation dengan file bank-full.arff pada k = 7



i. Cross-validation dengan file bank-full.arff pada k = 9



j. Cross-validation dengan file bank-full.arff pada k = 11



3. Kesimpulan.

Precision terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 5$ berada pada angka 0.872. Recall terbesar yang di dapat dari uji coba adalah pada cross-validation dengan $k = 11$ berada pada angka 0.892. F-Measure terbesar yang di dapat dari uji coba adalah pada supplied test set dengan $k = 5$ berada pada angka 0.875. Pada kasus dataset Bank Marketing, menggunakan ke-5 k menunjukkan hasil pengujian yang tidak memiliki selisih yang cukup besar baik pada pengujian supplied test maupun cross-validation. Nilai k yang digunakan pada supplied test lebih baik digunakan pada $k = 5$, lebih dari itu menghasilkan akurasi yang lebih rendah, namun pada cross-validation cenderung lebih baik menggunakan k yang lebih besar seperti 9 atau 11.